Host-symbiont population genomics provide insights into partner fidelity, transmission mode and habitat adaptation in deep-sea hydrothermal vent snails

Corinna Breusing[1*], Maximilian Genetti[2], Shelbi L. Russell[3], Russell B. Corbett-Detig[2], Roxanne A. Beinart[1*]

[1]University of Rhode Island, Graduate School of Oceanography, Narragansett, RI 02882, USA
[2]University of California Santa Cruz, Jack Baskin School of Engineering, Santa Cruz, CA 95064, USA
[3]University of California Santa Cruz, Department of Molecular, Cell, and Developmental Biology, Santa Cruz, CA 95064, USA

*Corresponding author: corinnabreusing@gmail.com, rbeinart@uri.edu

Running title: Host-symbiont population genomics in deep-sea hydrothermal vent snails

Keywords: Chemosynthetic symbiosis, hydrothermal vents, host-symbiont population genomics, habitat adaptation, symbiont transmission, symbiont specificity

Abstract

Symbiont specificity, both at the phylotype and strain level, can have profound consequences for host ecology and evolution. However, except for insights from a few model symbiosis systems, the degree of partner fidelity and the influence of host versus environmental factors on symbiont composition are still poorly understood. Nutritional symbioses between invertebrate animals and chemosynthetic bacteria at deep-sea hydrothermal vents are examples of relatively selective associations, where hosts affiliate only with particular phylotypes of gammaproteobacterial or campylobacterial symbionts. In hydrothermal vent snails of the sister genera *Alviniconcha* and *Ifremeria* this phylotype specificity has been shown to play a role in habitat distribution and partitioning among different holobiont species. However, it is currently unknown if fidelity goes beyond species level associations that might influence genetic structuring, connectivity and habitat adaptation of holobiont populations. We used metagenomic analyses to assess sequence variation in hosts and symbionts and identify correlations with geographic and environmental factors. Our analyses indicate that host populations are not differentiated across a ~800 km gradient, while symbiont populations are clearly structured between vent locations due to a combination of neutral and selective processes. Overall, these results suggest that host individuals flexibly associate with local strains of their specific symbiont phylotypes, which confirms a model of horizontal transmission in both *Ifremeria* and *Alviniconcha*. Strain flexibility in these snails likely enables host populations to exploit a range of habitat conditions, which might favor wide-spread genetic connectivity and ecological resilience unless physical dispersal barriers are present.

Introduction

Mutualistic relationships between eukaryotes and bacterial microbes are ubiquitous in nature. Symbionts enable hosts to gain access to novel resources and habitats, provide protection against pathogens and predators, and can be essential for the host's diet (Fisher et al. 2017; Russell 2019). For symbiotic associations to persist over evolutionary time, hosts must successfully transmit their symbionts from one generation to the other, either through symbiont acquisition from the environment (horizontal transmission), direct inheritance of symbiont lineages through the host germline (vertical transmission) or a combination of both mechanisms (mixed transmission) (Vrijenhoek 2010; Bright and Bulgheresi 2013). The mode of transmission has

50 significant implications for the composition and variation of symbionts within and between host
51 individuals. Vertical transmission typically results in strong genetic coupling between host and
52 symbiont lineages and an accompanied reduction in intra-host symbiont diversity (Vrijenhoek
53 2010). By contrast, horizontal transmission exposes aposymbiotic hosts to a potentially
54 heterogenous pool of free-living symbiont lineages (Vrijenhoek 2010), which can promote the
55 formation of generalist partnerships, where multiple hosts and symbionts associate with each
56 other, to more specialized associations between only one or a few potential partners (Fisher et al.
57 2017; Chomicki et al. 2020). This range is often referred to as host-symbiont specificity, which
58 can vary in its taxonomic level for both partners depending on the symbiotic system.
59 The degree of partner fidelity and its effect on symbiont composition can have dramatic
60 impacts on holobiont functioning (Ginete and Goodrich-Blair 2021). For example, host-symbiont
61 specificity, both at the species and genotype level, is crucial for light production in
62 bioluminescent squid (Verma and Miyashiro 2013), while shifts in microbiome assemblages
63 have been shown to affect phytoplankton growth rates (Jackrel et al. 2021) and the efficiency of
64 nitrogen fixation in nodule-forming legumes (Wang et al. 2018; Allito et al. 2020; Walker et al.
65 2020). Despite its importance for host biology, the taxonomic level of partner specificity as well
66 as the relative contributions of host genetic, environmental and geographical factors to symbiont
67 composition remain understudied in the majority of horizontally transmitted symbioses.
68 Chemosynthetic animal-microbe symbioses are globally significant phenomena that
69 dominate hydrothermal vent and hydrocarbon seep ecosystems in the deep sea. In these
70 associations, the bacterial partner uses chemical energy from the oxidation of reduced fluid
71 compounds, such as hydrogen, sulfide or methane, to convert carbon to organic matter, which
72 serves as primary nutrition for the host (Dubilier et al. 2008). Due to the nutritional dependence
73 on their bacterial symbionts, animal hosts are typically highly selective in their partner choice: In
74 the predominant number of cases host individuals associate with only 1–2 phylotypes of
75 gammaproteobacterial or campylobacterial symbionts (Dubilier et al. 2008), whereas symbionts
76 can exhibit a comparatively broad host range. Chemosynthetic endosymbioses in deep-sea snails
77 of the sister genera *Alviniconcha* and *Ifremeria* are examples of reciprocally relatively specific
78 partnerships, which dominate hydrothermal vent systems in the Southwest Pacific back-arc
79 basins. Given the absence of host-symbiont phylogenetic concordance, the symbionts are
80 assumed to be environmentally acquired (Suzuki et al. 2006), although a pseudo-vertical

81 transmission component is possible in *Ifremeria* given its brooding reproductive mode (Reynolds
82 et al. 2010). In the Eastern Lau Spreading Center (ELSC), previous work suggested that
83 specificity to functionally distinct symbiont phylotypes drives local and regional-scale habitat
84 partitioning among four co-occurring *Alviniconcha* and *Ifremeria* species (Beinart et al. 2012,
85 2015; Breusing et al. 2020a, b). *Alviniconcha boucheti* contains a campylobacterial phylotype
86 (Epsilon) and is usually found at northern vent sites with high concentrations of sulfide and
87 hydrogen, while *A. kojimai* and *A. strummeri* associate with different gammaproteobacterial
88 phylotypes (Gamma1 and Gamma1/GammaLau, respectively) and usually occupy mid-latitude
89 to southern vent sites where the concentrations of these chemical reductants are lower (Beinart et
90 al. 2012, 2019; Breusing et al. 2020a, b). *Ifremeria nautilei* establishes dual symbioses with
91 thiotrophic and methanotrophic gammaproteobacterial endosymbionts (Beinart et al. 2019) and
92 is co-distributed with *Alviniconcha* across their geographical range, although it typically
93 segregates into habitat patches with reduced fluid flow relative to its sister genus (Podowski et al.
94 2010). While it has been well-established that niche differentiation across hydrothermal vents is
95 mediated by symbiont phylotype specificity in *Alviniconcha* and *Ifremeria* host species, nothing
96 is known about the fidelity of these associations at the population level, nor how regional
97 adaptation is conferred functionally.
98 In this study we applied population genomic methods to assess symbiont strain-level
99 genetic variation and patterns of host-symbiont genetic subdivision in *Alviniconcha* and
100 *Ifremeria* species from the ELSC and Tonga Volcanic Arc. Using multivariate statistical
101 analyses, we evaluated the impact of host traits, environment and geography on symbiont
102 composition in populations of both genera and identified potential candidate genes for local
103 adaptation.
104
105 **Materials and Methods**
106 *Sample collection, nucleic acid extraction and sequencing*
107 Samples of *Alviniconcha* and *Ifremeria* were collected from six vent sites (1164–2722 m) of the
108 Lau Basin and Tonga Volcanic Arc in 2009 and 2016 using remotely operated vehicles (Fig. 1;
109 Table 1). Upon recovery, animal samples were dissected, placed in RNALater™ Stabilization
110 Solution (Thermo Fisher Scientific, Inc.) and frozen at –80°C until further analysis. DNA was
111 extracted with the Quick-DNA 96 Plus extraction kit (Zymo Research, Inc.) and further purified

112 with the MO BIO PowerClean DNA Pro Clean-Up kit (Qiagen, Inc.). High molecular weight
113 (HMW) DNA for long-read sequencing was isolated with Qiagen Genomic-tips following
114 manufacturer's instructions.
115
116 *Host transcriptome sequencing and assembly*
117 Illumina RNAseq reads for host transcriptome assemblies were obtained from sequencing
118 experiments performed in Seston et al. (2016) and Breusing et al. (2020b). Adapter clipping,
119 quality trimming, and removal of rRNA and symbiont reads was performed as in Breusing et al.
120 (2020b). Cleaned host reads were error corrected with RCORRECTOR (Song and Florea 2015) and
121 filtered for uncorrectable and overrepresented sequences with the
122 TRANSCRIPTOMEASSEMBLYTOOLS package
123 (https://github.com/harvardinformatics/TranscriptomeAssemblyTools). Host transcriptome co-
124 assemblies for each *Alviniconcha* and *Ifremeria* species were performed with TRINITY (Grabherr
125 et al. 2011) using the PASAFLY algorithm. For each *Alviniconcha* species, additional transcripts
126 were reconstructed from 454 reads obtained from Sanders et al. (2013). Assembled contigs for
127 each species were clustered with CD-HIT-EST (Li and Godzik 2006) at a 95% identity threshold
128 to reduce transcript redundancies. Open reading frames (ORFs) were predicted with
129 TRANSDECODER (https://github.com/TransDecoder/TransDecoder) considering homologies to
130 known proteins (UniRef90) and protein domains (Pfam) as ORF retention criteria. Transcripts
131 that did not contain any ORF or had a non-eukaryotic origin based on taxonomy classifications
132 with BLOBTOOLS (Laetsch and Blaxter 2017) were removed from the assembly. Final
133 transcriptome assemblies were evaluated for quality and completeness with BUSCO (Seppey et al.
134 2019).
135
136 *Symbiont genome sequencing and re-assembly*
137 Illumina reads for the *A. boucheti*, *A. kojimai* and *Ifremeria* holobionts were obtained from
138 previous sequencing runs performed in Beinart et al. (2019). Raw reads were trimmed with
139 TRIMMOMATIC (Bolger et al. 2014) and filtered for sequence contaminants through mapping
140 against the human (GRCh38) and PhiX reference genomes. Decontaminated reads were grouped
141 into symbiont and host reads by assessing best matches against draft symbiont genomes (Beinart
142 et al. 2019) with BBSPLIT (https://sourceforge.net/projects/bbmap/). To improve contiguity of the

143 genome assemblies we conducted 3–4 Nanopore sequencing runs for each symbiont phylotype
144 on a MinION device (Oxford Nanopore Technologies) using the SQK-LSK109 ligation kit after
145 HMW DNA enrichment with the Circulomics Short Read Eliminator kit. Basecalling of the
146 Nanopore reads was done with ALBACORE (Oxford Nanopore Technologies) and adapters were
147 clipped with PORECHOP (https://github.com/rrwick/Porechop). Hybrid assemblies of Illumina
148 and Nanopore reads were constructed for each symbiont genome with SPADES (Gamma1)
149 (Bankevich et al. 2012) or METASPADES (all others) (Nurk et al. 2017) choosing k-mers between
150 21 and 91 in 10 step increments. Raw assemblies for the Epsilon, Gamma1 and Ifr-SOX
151 symbionts were binned with GBTOOLS (Seah and Gruber-Vodicka 2015) and incrementally gap
152 filled, corrected and scaffolded with LR_GAPCLOSER (Xu et al. 2019), ra2.py (Brown et al.
153 2015), and SLR (Luo et al. 2019), respectively, following recommendations by Chen et al.
154 (2020). The Ifr-MOX symbiont assembly was automatically binned with METABAT2 (Kang et
155 al. 2019) for contigs ≧ 1500 bp. Shorter contigs (≧ 500 bp) were binned with MAXBIN (Wu et al.
156 2016). Scaffolds < 200 bp were excluded from all assemblies. Final assemblies were polished
157 with PILON (Walker et al. 2014), functionally annotated with RAST-TK (Brettin et al. 2015) and
158 quality-checked with CHECKM (Parks et al. 2015) and QUAST (Gurevich et al. 2013).
159
160 *Population-level metagenomic sequencing and variant identification in hosts and symbionts*
161 192 barcoded high-throughput DNA sequencing libraries were prepared with a Tn5 transposase-
162 based protocol after Picelli et al. (2014) at the University of California Santa Cruz and then sent
163 for 150 bp paired-end sequencing on a NovaSeq 6000 instrument at the University of California
164 Davis. However, due to low read allocation 66 of the libraries were excluded from further
165 analysis. Raw sequence reads were trimmed, filtered and sorted by host species and symbiont
166 phylotype as described above. Optical duplicates were removed with PICARD's
167 MARKDUPLICATES tool (https://github.com/broadinstitute/picard). To resolve common alignment
168 errors and improve base call accuracy, we locally realigned reads around indels and recalibrated
169 base quality scores with LOFREQ (Wilm et al. 2012).
170 Host population genomic variation was assessed in ANGSD (Korneliussen et al. 2014) by
171 inferring genotype likelihoods based on Hardy-Weinberg equilibrium considering individual
172 inbreeding coefficients. To increase accuracy of the analyses, variant sites with mapping
173 qualities < 30 (minMapQ = 30), base qualities < 20 (minQ = 20), and minimum minor allele

174 frequencies < 0.01 (minMaf = 0.01) were excluded. We further filtered sites based on strand bias
175 (sb_pval = 0.05), heterozygote bias (hetbias_pval = 0.05) and probability of being variable
176 (SNP_pval = 1e-6). In addition, we removed spurious and improperly paired reads, adjusted
177 mapping qualities for excessive mismatches (C = 50) and computed per-base alignment qualities
178 (BAQ = 1) to disregard variants close to indel regions. Putative paralogous variants were
179 excluded by discarding reads with multiple mappings and by considering only sites that had a
180 maximum depth of 40–80. Genetic distances between individuals were inferred by calculating
181 pairwise covariance matrices.
182 Symbiont population genomic variation was determined with FREEBAYES (Garrison and
183 Marth 2012) using input parameters adjusted for the analysis of metagenomic data (-F 0.01 -C 1
184 -p 1 --pooled-continuous --haplotype-length 0 --report-monomorphic). Variant calls were
185 restricted to sites with a minimum base quality of 20, a minimum mapping quality of 30, and a
186 minimum coverage of 10. To eliminate bias in variant identification and other downstream
187 analyses due to uneven read depth between samples, we normalized all samples to the lowest
188 amount of coverage found in a sample for a particular symbiont phylotype (> 10X coverage).
189 Variants were further filtered based on strand bias (SRP > 5 && SAP > 5 && EPP > 5),
190 proximity to indels (5 bp) and maximum mean depth with BCFTOOLS (Li et al. 2009) and
191 VCFTOOLS (Danecek et al. 2011). In addition, sites and individuals with more than 25% missing
192 data were excluded from the analysis. Allele counts (= symbiont strain abundances) and
193 consensus haplotypes (= dominant symbiont strains) were extracted with GATK's
194 VARIANTSTOTABLE tool (Van der Auwera and O'Connor 2020).
195
196 *Genomic structure and differentiation*
197 We performed ordination analyses with the APE and STATS packages in R (Paradis and Schliep
198 2019; R Core Team 2020) to assess genetic variation between symbiont and host populations.
199 Host genetic structure was determined through principal component analyses based on genetic
200 covariance matrices, while symbiont genetic structure was inferred through principal coordinate
201 analyses based on both consensus haplotype and allele count data transformed into Euclidean
202 distances and Bray-Curtis dissimilarities, respectively. For each sample, absolute allele counts
203 were normalized to relative counts prior to analysis. Negative eigenvalues were corrected using
204 the method by Cailliez (1983) and final ordination plots were produced with GGPLOT2

205 (Wickham 2016). F_{ST} values between host and symbiont populations were calculated in ANGSD
206 and SCIKIT-ALLEL (https://github.com/cggh/scikit-allel), respectively, following the procedure
207 by Bhatia et al. (2013). For the host species, potential outlier loci under selective pressures were
208 inferred with OUTFLANK (Whitlock and Lotterhos 2015) based on neutral F_{ST} distributions that
209 were obtained from quasi-independent subsets of single nucleotide polymorphisms (SNPs)
210 determined with PLINK (Purcell et al. 2007).
211
212 *Assessment of gene content variation*
213 We used PANPHLAN (Beghini et al. 2020) to investigate potential differences in gene
214 composition between symbiont strains from different vent localities. Downsampled symbiont
215 reads for each species were mapped against the corresponding symbiont reference genome using
216 custom RAST-TK annotations for functional categorizations. Samples were profiled for gene
217 presence/absence based on the following parameter thresholds: --min_coverage 1 --left_max
218 1.70 --right_min 0.30. Gene content variation between symbiont strains was assessed by
219 identifying genes that were present/absent in 90% of samples from one or the other geographic
220 region. To account for gene content variation between different strains within hosts we further
221 quantified gene abundances with SALMON (Patro et al. 2017) using the following parameters: -l
222 IU --meta --rangeFactorizationBins 4 --numBootstraps 1000 --seqBias --gcBias -s -u. Abundance
223 values were normalized with the Trimmed Mean of M method using TRINITY's
224 *abundance_estimates_to_matrix.pl* script (Robinson and Oshlack 2010; Grabherr et al. 2011).
225 Gene presence/absence and differential abundance heatmaps were produced with the
226 COMPLEXHEATMAP package in R (Gu et al. 2016).
227
228 *Assessment of symbiont variation based on environment and host genetics*
229 We conducted partial Mantel tests with the VEGAN package in R (Oksanen et al. 2020) to assess
230 relationships between host and symbiont genetic distances and geography based on Spearman
231 rank correlations. As no population genetic structure could be observed in any host species based
232 on the full SNP datasets, we estimated host covariance matrices using only moderately to highly
233 differentiated SNP subsets for this analysis (*Alviniconcha*: $F_{ST} > 0.15$; *Ifremeria*: $F_{ST} > 0.10$).
234 Geographic distances were determined by calculating the geodesics between vent sites with the
235 GEOSPHERE package (Hijmans 2019). We used redundancy analyses (RDA) following the

236 approach in Forester et al. (2018) (https://popgen.nescent.org/2018-03-27_RDA_GEA.html) to
237 evaluate the influence of hydrothermal fluid composition and depth on the genetic structure of
238 each symbiont phylotype. We further included the effect of sampling year, which encompasses
239 changes in hydrothermal circulation within the ELSC as indicated by the cessation in fluid flow
240 at the Kilo Moana vent field between 2009 and 2016. Endmember concentrations for
241 geochemical compounds were obtained from the literature or unpublished data (Mottl et al. 2011;
242 Beinart et al. 2012; Flores et al. 2012; Seewald, unpubl.; Diehl, unpubl.; Table 1). Due to multi-
243 collinearity among the chemical species, we used the first eigenvector from principal coordinate
244 analyses (corresponding to the eigenvalue with the largest explanatory power) as composite
245 value. For each symbiont phylotype, we further assessed the strength of correlation with other
246 predictors to exclude variables that were highly collinear. Unless geographic location was
247 strongly linked to environmental predictors, we used latitude as a conditioning factor in the
248 analyses to correct for isolation by distance. Based on these collinearity evaluations, we tested
249 the effect of all factors on the two *Ifremeria* symbionts, the effect of depth and fluid composition
250 on the *A. boucheti* symbiont and the effect of fluid composition and year (uncorrected for
251 geography) on the *A. kojimai* and *A. strummeri* symbionts. SNPs were considered candidates for
252 local adaptation if their loadings on significant constrained RDA axes deviated more than 2.5
253 standard deviations from the mean of the distribution. For each symbiont phylotype we further
254 calculated the ratio of non-synonymous to synonymous polymorphisms (pN/pS) with SNPEFF
255 (Cingolani et al. 2012) to infer candidate genes evolving under natural selection.
256
257 **Results**
258 *Host transcriptome assemblies and population genomic structure*
259 Host transcriptome assemblies consisted of 24,176–35,654 transcripts (totaling 20.68–28.68 Mb)
260 and were approximately 30.30–55.40% complete (Table S1). Mapping of host reads against the
261 transcriptome references and subsequent filtering of variant sites yielded 1,655–9,185 SNPs per
262 species for population genetic analyses. Irrespective of host taxon, F_{ST} and ordination analyses
263 revealed low genetic differentiation among host populations sampled from different vent
264 localities (Fig. 2; Table S2). With the exception of four SNP sites in *A. kojimai*, no F_{ST} outliers
265 could be detected in any host species (Table S3). However, all species contained a number of
266 SNPs that were moderately to highly divergent among host populations. When analyses were

267 constrained to these SNP subsets (*Alviniconcha*: $F_{ST} > 0.15$; *Ifremeria*: $F_{ST} > 0.10$), population
268 genetic structuring by vent site was observed in all *Alviniconcha* species, but not *Ifremeria* (Fig.
269 S1). For both *A. kojimai* and *A. strummeri*, the respective SNP subsets further indicated genetic
270 differences between Tui Malila populations that were sampled in different years. The effect of
271 geography on host population genetic structuring was not significant after correction for
272 symbiont genetic distance in partial Mantel correlations (Table S4).
273
274 *Symbiont genome assemblies and population genomic structure*
275 Symbiont genome assemblies varied in size from 2.10–4.86 Mb and contained 2,006–6,660
276 predicted protein-coding genes, with GC content ranging from 34.30–59.00% (Table S2). All
277 assemblies were characterized by a high level of completeness (92.79–99.25%) and low amount
278 of contamination (0.74–5.48%) (Table S5). Symbiont populations for the Epsilon, Gamma1, Ifr-
279 SOX and Ifr-MOX phylotypes were largely structured by vent field or broader geographic region
280 based on 239–7,057 variant sites (Fig. 3, S2; Table S6). These associations were significant even
281 when corrected for host genetic variation based on highly differentiated SNP markers (Table S4;
282 $p \leqq 0.0244$, $r = 0.2665–0.8705$). Genetic differentiation between symbiont populations typically
283 increased with geographic distance between vent fields (Table S6). Although F_{ST} values for
284 GammaLau populations of *A. strummeri* were moderate when calculated between distinct vent
285 sites (Table S6), ordination analyses and Mantel tests provided no evidence for genetic
286 differentiation of this symbiont phylotype across geographic locations based on 192 marker loci
287 (Fig. 3, S2). The Gamma1 phylotype is associated with both *A. kojimai* and *A. strummeri* and we
288 therefore investigated whether populations of this phylotype varied between host species.
289 Ordination analyses indicated a clear clustering by host taxon that superseded the effect of
290 geography (Fig. 4, S3), suggesting that different Gamma1 strains associate selectively with either
291 *A. kojimai* or *A. strummeri*. By contrast, the effect of host genetics on symbiont genetic variation
292 within species appeared to be weak. Although partial Mantel tests suggested significant
293 associations of host genotype with strain composition or dominant symbiont type for the *A.*
294 *boucheti* – Epsilon, *A. kojimai* – Gamma1, *A. strummeri* – Gamma1, and *I. nautilei* – Ifr-SOX
295 pairs, *r* statistics were relatively low especially compared to the effect of geography, indicating
296 limited biological relevance (Table S4; $p \leqq 0.0364$, $r = 0.0510–0.2270$).
297

298 *Gene content variation between symbiont strains*

299 We assessed variation in gene content between symbiont strains from different vent localities or
300 broader geographic regions (Fig. 5, S4; Table S7). For the Gamma1 symbiont, we further
301 determined gene content variation between strains from different host species given that this
302 symbiont phylotype occurs in both *A. strummeri* and *A. kojimai* (Fig. 4, S3; Table S8).
303 Differentially preserved or abundant genes between symbiont strains comprised about 1–5% of
304 protein-coding regions. Independent of symbiont phylotype, the most common differences
305 among geographic or host-specific strains concerned genes of unknown function as well as a
306 smaller number of genes related to mobilome and anti-viral defense (Fig. 4–5, S3–4; Table S7,
307 S8). Geographic strains of the *A. boucheti* Epsilon symbiont further differed in the presence of an
308 ABC transporter, a GDP-L-fucose synthetase, a NAD(FAD)-utilizing hydrogenase and the DNA
309 repair protein RecN, which were conserved in strains from Kilo Moana and Tow Cam but were
310 absent or very lowly abundant in strains from ABE (Table S7). Within the *A. kojimai* Gamma1
311 phylotype, strains from ABE (but rarely Tui Malila) contained a few genes involved in DNA,
312 protein and cell wall metabolism as well as maturation and regulation of uptake [NiFe]
313 hydrogenases (*hyaC*, *hyaD, hoxJ*). Differences in hydrogenase-related genes were also observed
314 in the Gamma1 symbiont of *A. strummeri*, where an operon for a hydrogen-sensing hydrogenase
315 (*hupUV*/*hoxBC*) and genes for hydrogenase maturation and assembly proteins (*hypF*, *hyaF*,
316 *hoxV*/*hupK*, *hypD*, *hypE*, *hoxX*) were largely missing in strains from Tui Malila (but not Tahi
317 Moana) (Table S7). Between host species, Gamma1 strains notably differed in the presence of
318 genes for a sulfite dehydrogenase complex (*soeABC*), which was abundant in strains specific to
319 *A. kojimai* but not *A. strummeri* (Table S8). Within the *A. strummeri* GammaLau phylotype,
320 strains from Tui Malila contained a broad range of metabolic genes that were absent or
321 infrequent in strains from Tahi Moana, including genes related to macronutrient metabolism,
322 stress response, membrane transport and several other functional processes (Table S7). The Ifr-
323 SOX strains of *I. nautilei* differed in several genes related to DNA, nitrogen, lipid and amino
324 acid metabolism, membrane transport, cell regulation and detoxification, which were present in
325 strains from the ELSC but mostly missing in strains from Niua South. Within the Ifr-MOX
326 phylotype, strains from Tahi Moana and Tui Malila differed in genes involved in sulfide
327 respiration and oxidative phosphorylation, macronutrient metabolism, cation transport, motility,

328 cofactor, cell wall, DNA and nucleotide metabolism as well as cell regulation and stress response
329 (Table S7).
330
331 *Adaptive variation within symbiont phylotypes*
332 A subset of genetic variants exhibited significant associations with environmental factors in all
333 symbiont phylotypes except for GammaLau, with constrained ordinations explaining between
334 5.39% and 19.76% of the variance in the respective RDA models (Fig. 6; Table S9). Candidate
335 adaptive loci for each symbiont phylotype encompassed a variety of metabolic categories,
336 although in many cases no particular function could be assigned (Table S9). In the *A. boucheti*
337 Epsilon symbiont, 129 variants were significantly associated with fluid composition, while
338 another 47 were correlated with depth. These variants were mostly located in genes related to
339 protein, amino acid, and cell wall metabolism, followed by cofactor, DNA, carbon and
340 nucleotide metabolism, as well as membrane transport (Table S9). Several other variants were
341 linked to nitrogen, iron and RNA metabolism, virulence, motility, cell signaling, stress response,
342 respiration, cell cycle, as well as hydrogen and sulfur metabolism. In the *A. kojimai* Gamma1
343 symbiont, 91 variants were correlated with fluid composition and 9 were linked to year. 29 of
344 these variants were classified into functional categories. The most represented categories were
345 mobilome, cell wall and DNA metabolism, anti-viral defense, membrane transport, cofactor
346 metabolism and respiration. A few variants were associated with amino acid, hydrogen, protein
347 and RNA metabolism, and cell regulation. In the *A. strummeri* Gamma1 symbiont, 95 variants
348 showed associations with fluid composition, while 3 were correlated with year. 54 of these
349 variants could be functionally annotated and assigned to the following metabolic categories:
350 protein, amino acid and cell wall metabolism, membrane transport, virulence, carbon, nitrogen,
351 DNA/RNA and nucleoside metabolism, cofactor, fatty acid and sulfur metabolism, stress
352 response, cell signaling and mobilome (Table S9). The *I. nautilei* Ifr-SOX and Ifr-MOX
353 symbionts contained 94 and 11 putatively adaptive variants, respectively. In both symbionts,
354 these variants were mostly linked to depth, followed by fluid composition and year. While the
355 majority of variants were located in genes with unknown functions, a number of loci was linked
356 to mobile elements, cell cycle, anti-viral defense, DNA/RNA metabolism, and membrane
357 transport. In each symbiont phylotype, virtually all candidate variants were characterized by
358 elevated F_{ST} and gene-wide *p*N/*p*S values, further supporting their potential role in habitat

359 adaptation (Table S9). Several other genes exhibited increased ratios of non-synonymous to
360 synonymous substitution, although sites within these genes did not show significant associations
361 with environment (Table S10). It is possible that sites within these genes covary with other
362 ecological factors that could not be tested in this study or that these patterns are caused by recent
363 slightly deleterious mutations that were not yet purged by natural selection.
364
365 **Discussion**
366 Strain-level variation within microbial symbionts is increasingly recognized as an important
367 driver of host ecology and evolution (Ginete and Goodrich-Blair 2021), yet its patterns,
368 determining factors and functional implications remain poorly investigated in many non-model
369 symbioses. In this study, we used metagenomic analyses to assess sequence and gene content
370 differences between chemosynthetic symbionts associated with four co-occurring species of
371 deep-sea hydrothermal vent snails and determined the impact of host genetic, geographical and
372 environmental factors on symbiont strain composition and variation.
373 Despite fidelity between hosts and symbionts at the species level (Beinart et al. 2012;
374 Breusing et al. 2020a), our results indicate that specificity between host genotypes and symbiont
375 strains in *Alviniconcha* and *Ifremeria* is weak: Host populations were not partitioned across a
376 ~800 km gradient, whereas symbiont populations were clearly structured between vent locations
377 or broader geographic regions. Even when we correlated symbiont and host genetic distances
378 based on highly differentiated markers, test statistics for significant associations remained low,
379 suggesting that host genetics has a minor impact compared to environment or geography on
380 strain composition within both *Alviniconcha* and *Ifremeria*.
381 These findings qualitatively agree with observations in deep-sea mussel hybrids that
382 appear to associate with locally available symbiont strains (Ücker et al. 2021), but contrast
383 markedly with patterns in other horizontally transmitted partnerships, such as the squid-*Vibrio*
384 symbiosis and some legume-rhizobia associations, where hosts exhibit strong strain specificity
385 (reviewed in Ginete and Goodrich-Blair 2021). Environmental uptake of locally adapted
386 symbiont strains provides host organisms with the opportunity to optimally exploit novel
387 habitats, but carries the risk of unsuccessful symbiont acquisition and infection by cheaters
388 (Vrijenhoek 2010). Depending on the amount of partner reliance, chances of symbiont encounter
389 and fitness variation among symbiont strains, holobionts likely find different tradeoffs between

390 these opposing factors. Associations with chemosynthetic bacteria are obligate for hydrothermal
391 vent animals, but are restricted to relatively ephemeral habitats that are characterized by large
392 temporal and spatial fluctuations in environmental conditions and associated shifts in microbial
393 communities (Tivey 2007; Dick et al. 2013; Dick 2019). Strong nutritional dependency in vent
394 symbioses combined with the uncertainty of habitat (and thus symbiont) encounter might
395 promote specificity towards a mutualistic symbiont phylotype, while enabling enough flexibility
396 towards different local strains of that phylotype to maximize recruitment success of host larvae at
397 a new vent site. By contrast, bobtail squids inoculate their environment with symbiotic bacteria
398 and thereby ensure symbiont availability for their offspring (reviewed in Nyholm and McFall-
399 Ngai 2004), which might favor increased strain selectivity in this association. Similarly, some
400 legume species appear to preferentially associate with certain rhizobial strains that are abundant
401 in the native host range (Andrews and Andrews 2017). Although the factors underlying strain
402 specificity are not fully understood, it is possible that the relative fitness advantages provided by
403 these locally available strains (Andrews and Andrews 2017) coupled with the benefit of
404 decreased cheater invasion (Walker et al. 2020) might have selected for strong partner fidelity in
405 these symbioses.
406 *Alviniconcha kojimai* and *A. strummeri* represent notable exceptions to the observed
407 patterns, as they share identical phylotypes of one gammaproteobacterial symbiont (Gamma1),
408 but appear to take up different strains even in habitats where these host species co-occur. The
409 phylogenetic divergence between *A. kojimai* and *A. strummeri* (~25 MYA) is more recent than
410 that between either species and *A. boucheti* (~38 MYA) or *I. nautilei* (~113 MYA) (Breusing et
411 al. 2020a). Perhaps the closer evolutionary relationship between these two species favors
412 associations with the same symbiont phylotype, as has been suggested for *Bathymodiolus*
413 *azoricus* and *B. puteoserpentis* on the Mid-Atlantic Ridge as well as *B. thermophilus* and *B.*
414 *antarcticus* on the East Pacific Rise (Ücker et al. 2021). However, compared to the
415 bathymodiolin mussel system, the split between *A. kojimai* and *A. strummeri* is significantly
416 older (>17 MYR), which could indicate that the timing of phylotype specificity evolution varies
417 among taxonomic groups, possibly as a result of contrasting ecological or evolutionary contexts.
418 The fact that *A. kojimai* and *A. strummeri* nevertheless associate with distinct symbiont strains in
419 sympatry is potentially a mechanism to avoid competition for niche space. Gamma1 strains of
420 these two species notably differed in the presence of a molybdenum-containing sulfite

421 dehydrogenase (SoeABC), which was preserved in strains of *A. kojimai* but not *A. strummeri*.
422 SoeABC is the predominant enzyme for sulfite oxidation in many purple sulfur bacteria (Dahl et
423 al. 2013) and is further involved in taurine and dimethylsulfoniopropionate degradation in
424 *Roseobacter* clade bacteria (Rinta-Kanto et al. 2011; Lehmann et al. 2012; Lenk et al. 2012).
425 Although the physiological role of SoeABC in *A. kojimai*'s Gamma1 strain is unknown, it is
426 possible that it contributes to partitioning of sulfur resources among co-occurring snail
427 holobionts.
428 Our data support a model of horizontal symbiont transmission in both *Alviniconcha* and
429 *Ifremeria*. While these results are expected for *Alviniconcha* which produce free-swimming
430 planktotrophic larvae and do not invest in their young (Warén and Bouchet 1993), they are rather
431 surprising for *Ifremeria* which brood their offspring in a modified pouch in the female's foot
432 (Reynolds et al. 2010) and would thus have the opportunity to pseudo-vertically transmit their
433 symbionts. It is possible that a pseudo-vertical transmission component exists in *Ifremeria*, but
434 that maternally acquired symbionts get replaced or complemented by more competitive strains in
435 the habitat where the snail larvae settle. Our current dataset cannot distinguish this possibility
436 from a strict horizontal transmission mode, though future studies assessing symbiont
437 composition in different developmental stages of *Ifremeria* would be helpful to address this
438 hypothesis.
439 While the majority of genetic variants in the symbionts did not deviate from neutral
440 expectations, a subset of loci showed evidence for natural selection, implying a role of both
441 genetic drift and local adaptation in shaping symbiont population structure. Candidate adaptive
442 loci spanned a surprisingly broad range of metabolic functions, many of which were probably
443 not causally linked to the investigated environmental predictors but other correlated variables
444 that we could not account for in this study. For example, in almost all symbiont phylotypes we
445 observed adaptive variation in genes that were related to anti-viral defense (e.g., CRISPR-
446 associated proteins, restriction-modification systems) or mobile elements. Although differences
447 in depth or geochemistry might contribute to these patterns, it is more likely that they reflect
448 exposure of symbiont strains to distinct viral assemblages that might covary with local habitat
449 conditions, as has been suggested for vestimentiferan tubeworm symbionts (Perez and Juniper
450 2016). Polymorphisms in other genes, by contrast, are likely directly explained by variation in
451 the analyzed environmental factors. Geographic strains of *A. boucheti*'s Epsilon symbiont, for

452 instance, contained several variants under positive selection that were located in genes involved
453 in hydrogenase assembly, iron transport, sulfur oxidation and respiration. Likewise, the Gamma1
454 strains of *A. kojimai* and *A. strummeri* showed adaptive differences in variants related to
455 hydrogen and sulfur metabolism, respectively. Niche-specific differences in hydrogen metabolic
456 genes were also observed in comparisons of gene content among symbiont strains. In both *A.*
457 *kojimai* and *A. strummeri*, Gamma1 strains from Tui Malila lacked some subunits of uptake or
458 hydrogen-sensing hydrogenases as well as various genes for hydrogenase maturation, synthesis
459 and regulation. Given that H_2 concentrations at Tui Malila can drop to 35 μM in endmember
460 fluids and are probably lower in diffuse flow habitats (Beinart et al. 2012), it is possible that
461 hydrogen does not constitute a major energy source for Gamma1 strains from this locality. This
462 hypothesis is in agreement with previous physiological experiments that revealed strikingly low
463 hydrogen oxidation and associated carbon fixation rates in Gamma1 symbionts from Tui Malila
464 (Breusing et al. 2020b). An alternative though mutually non-exclusive explanation for the loss of
465 hydrogenase-related genes at least in the *A. strummeri* strain from Tui Malila could be avoidance
466 of intra-host competition with co-occurring GammaLau strains, which often co-dominate in *A.*
467 *strummeri* individuals at this vent site (Beinart et al. 2012). Such functional diversity is predicted
468 to enable symbiont coexistence in a variety of hydrothermal vent symbioses, including
469 bathymodiolin mussels (Ikuta et al. 2016; Ansorge et al. 2019), alvinocaridid shrimp (Cambon-
470 Bonavita et al. 2021) and vestimentiferan tubeworms (Zimmermann et al. 2014).
471 Compared to their symbionts, host populations were markedly less structured across the
472 same spatial scales. Although a proportion of genetic markers was differentiated across vent
473 sites, we did not find strong evidence for local adaptation in any host species, suggesting that
474 these patterns likely reflect random variation among localities. The limited genetic subdivision
475 between host populations agrees with predictions from biophysical models that indicate absence
476 of physical dispersal barriers for vent larvae within the Lau Back-Arc Basin (Mitarai et al. 2016;
477 Breusing et al. in review). While symbiont population structure, by contrast, appeared to be at
478 least partly driven by natural selection, it is possible that symbionts also experience stronger
479 dispersal limitations than their hosts, as has been hypothesized in some coral-algae symbioses
480 (Davies et al. 2020). The environmental distributions and life cycles of the free-living stages of
481 *Alviniconcha* and *Ifremeria* symbionts are currently unknown and a better understanding of these

482 aspects will be necessary to evaluate the relative importance of dispersal barriers on symbiont
483 biogeography.
484 Overall, our findings reveal a lack of strain-level specificity in *Alviniconcha* and
485 *Ifremeria* symbioses, which possibly reflects an evolutionary strategy to cope with the transient
486 and dynamic nature of hydrothermal vent habitats. Strain flexibility in these associations likely
487 contributes to the wide-spread genetic connectivity observed among host populations, which in
488 turn could favor ecological resilience to natural but also anthropogenic environmental
489 disturbances. Considering the impact of symbiont availability and specificity on host distribution
490 in a variety of vent symbioses, these results could be important for future conservation
491 management given the increasing human pressures on hydrothermal ecosystems worldwide
492 (Levin et al. 2020).
493
494 **Acknowledgements**
495 We thank the captains, crews and ROV pilots of the R/V *Thomas G. Thompson* (ROV *Jason II*)
496 and R/V *Falkor* (ROV *Ropos*) for supporting the sample collections that made this study
497 possible. We thank Pete Girguis for his contribution of the 2009 samples to this project,
498 Alexander Diehl and Jeff Seewald for providing geochemical data, Michelle Hauer and Erin
499 Frates for their assistance with sample preparation and Brown University for providing access to
500 their high-performance computing cluster where the bioinformatic analyses were performed. We
501 further thank the technical staff at the UC Davis Genome Center for sequencing our Illumina
502 metagenomic libraries. This work was supported by the National Science Foundation (grant
503 numbers OCE-1536331, 1819530 and 1736932 to R.A.B.) and the National Institute of Health
504 (grant numbers 5K99GM135583-02 to S.L.R. and 5R35GM128932-03 to R.C.D.).
505
506 **Data Accessibility**
507 Raw Nanopore reads, host transcriptomes and symbiont genomes have been deposited in the
508 National Center for Biotechnology Information under BioProjects PRJNA523619,
509 PRJNA526236 and PRJNA741492. Annotations for new symbiont genomes are available on the
510 RAST webserver (https://rast.nmpdr.org/rast.cgi) through the guest access (login: guest,
511 password: guest) under job IDs 6666666.685764 (Epsilon), 6666666.685765 (Ifr-SOX),

512 6666666.686795 (Gamma1), and 6666666.687466 (Ifr-MOX). Scripts for bioinformatic analyses
513 are available on GitHub under https://github.com/cbreusing/Provannid_host-symbiont_popgen.
514

515 **Author's Contributions**

516 C.B. and R.A.B. designed the study with advice from S.L.R. and R.C.D. C.B. performed data
517 analyses and wrote the first draft of the paper. M.G. prepared Illumina sequencing libraries. All
518 authors contributed to data interpretation and writing of the manuscript.
519

520 **Competing Interests**

521 The authors declare no competing interests.
522

523 **References**

524 Allito BB, Ewusi-Mensah N, Logah V. Legume-*Rhizobium* strain specificity enhances nutrition
525 and nitrogen fixation in faba bean (*Vicia faba* L.). Agronomy. 2020; 10(6): 826.
526 Andrews M, Andrews ME. Specificity in Legume-Rhizobia Symbioses. Int J Mol Sci. 2017;
527 18(4): 705.
528 Ansorge R, Romano S, Sayavedra L, Porras MÁG, Kupczok A, Tegetmeyer HE *et al.* Functional
529 diversity enables multiple symbiont strains to coexist in deep-sea mussels. Nat Microbiol.
530 2019; 4: 2487–2497.
531 Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M *et al.* SPAdes: a new genome
532 assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;
533 19(5): 455–477.
534 Beghini F, McIver LJ, Blanco-Míguez A, Dubois L, Asnicar F, Maharjan S *et al.* Integrating
535 taxonomic, functional, and strain-level profiling of diverse microbial communities with
536 bioBakery 3. Elife. 2021; 10: e65088.
537 Beinart RA, Sanders JG, Faure B, Sylva SP, Lee RW, Becker EL *et al.* Evidence for the role of
538 endosymbionts in regional-scale habitat partitioning by hydrothermal vent symbioses.
539 PNAS 2012; 109: E3241–3250.
540 Beinart RA, Gartman A, Sanders JG, Luther GW III, Girguis PR. The uptake and excretion of
541 partially oxidized sulfur expands the repertoire of the energy resources metabolized by
542 hydrothermal vent symbioses. Proc Biol Sci. 2015; 282: 20142811.
543 Beinart RA, Luo C, Konstantinidis K, Stewart FJ, Girguis PR. The bacterial symbionts of closely
544 related hydrothermal vent snails with distinct geochemical habitats show broad similarity
545 in chemoautotrophic gene content. Front Microbiol. 2019; 10: 1818.
546 Bhatia G, Patterson N, Sankararaman S, Price AL. Estimating and interpreting FST: the impact
547 of rare variants. Genome Res. 2013; 23(9): 1514–1521.
548 Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data.
549 Bioinformatics 2014; 30: 2114–2120.

550 Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ *et al.* RASTtk: a modular and
551 extensible implementation of the RAST algorithm for building custom annotation pipelines
552 and annotating batches of genomes. Sci Rep. 2015; 5: 8365.
553 Breusing C, Johnson SB, Tunnicliffe V, Clague DA, Vrijenhoek RC, Beinart RA. Allopatric and
554 sympatric drivers of speciation in *Alviniconcha* hydrothermal vent snails. Mol Biol Evol.
555 2020a; 37(12); 3469–3484.
556 Breusing C, Mitchell J, Delaney J, Sylva SP, Seewald JS, Girguis PR *et al.* Physiological
557 dynamics of chemosynthetic symbionts in hydrothermal vent snails. ISME J. 2020b; 14:
558 2568–2579.
559 Bright M, Bulgheresi S. A complex journey: transmission of microbial symbionts. Nat Rev
560 Microbiol. 2010 Mar; 8(3): 218–230.
561 Brown CT, Hug LA, Thomas BC, Sharon I, Castelle CJ *et al.* Unusual biology across a group
562 comprising more than 15% of domain Bacteria. Nature 2015; 523: 208–211.
563 Cailliez F. The analytical solution of the additive constant problem. Psychometrika 1983; 48:
564 305–308.
565 Cambon-Bonavita MA, Aubé J, Cueff-Gauchard V, Reveillaud J. Niche partitioning in the
566 *Rimicaris exoculata* holobiont: the case of the first symbiotic Zetaproteobacteria.
567 Microbiome 2021; 9: 87.
568 Chen LX, Anantharaman K, Shaiber A, Eren AM, Banfield JF. Accurate and complete genomes
569 from metagenomes. Genome Res. 2020; 30(3): 315–333.
570 Chomicki G, Kiers ET, Renner SS. The evolution of mutualistic dependence. Annu Rev Ecol
571 Evol Syst. 2020; 51: 409–432.
572 Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L *et al.* A program for annotating
573 and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the
574 genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly (Austin) 2012; 6(2):
575 80–92.
576 Dahl C, Franz B, Hensen D, Kesselheim A, Zigann R. Sulfite oxidation in the purple sulfur
577 bacterium Allochromatium vinosum: identification of SoeABC as a major player and
578 relevance of SoxYZ in the process. Microbiology (Reading). 2013; 159(Pt 12): 2626–2638.
579 Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA *et al.* The variant call
580 format and VCFtools. Bioinformatics. 2011; 27(15): 2156–2158.
581 Davies SW, Moreland KN, Wham DC, Kanke MR, Matz MV. *Cladocopium* community
582 divergence in two *Acropora* coral hosts across multiple spatial scales. Mol Ecol. 2020; 29:
583 4559–4572.
584 Dick GJ, Anantharaman K, Baker BJ, Li M, Reed DC, Sheik CS. The microbiology of deep-sea
585 hydrothermal vent plumes: ecological and biogeographic linkages to seafloor and water
586 column habitats. Front Microbiol. 2013; 4: 124.
587 Dick GJ. The microbiomes of deep-sea hydrothermal vents: distributed globally, shaped locally.
588 Nat Rev Microbiol. 2019; 17(5): 271–283.
589 Dubilier N, Bergin C, Lott C. Symbiotic diversity in marine animals: the art of harnessing
590 chemosynthesis. Nat Rev Microbiol. 2008; 6: 725–740.
591 Fisher RM, Henry LM, Cornwallis CK, Kiers ET, West SA. The evolution of host-symbiont
592 dependence. Nat Commun. 2017; 8: 15973.
593 Flores GE, Shakya M, Meneghin J, Yang ZK, Seewald JS, Wheat GC *et al.* Inter-field variability
594 in the microbial communities of hydrothermal vent deposits from a back-arc basin.
595 Geobiology. 2012; 10(4): 333–346.

596 Forester BR, Lasky JR, Wagner HH, Urban DL. Comparing methods for detecting multilocus
597 adaptation with multivariate genotype–environment associations. Mol Ecol. 2018; 27:
598 2215–2233.
599 Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv
600 preprint 2012; arXiv: 1207.3907v2 [q-bio.GN].
601 Ginete DR, Goodrich-Blair H. From binary model systems to the human microbiome: factors
602 that drive strain specificity in host-symbiont associations. Front Ecol Evol. 2021; 9: 38.
603 Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I *et al.* Full-length
604 transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol.
605 2011; 29: 644–652.
606 Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in
607 multidimensional genomic data. Bioinformatics. 2016; 32(18): 2847–2849.
608 Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome
609 assemblies. Bioinformatics 2013; 29: 1072–1075.
610 Hijmans RJ. geosphere: Spherical Trigonometry. R package version 1.5-10 2019.
611 https://CRAN.R-project.org/package=geosphere
612 Ikuta T, Takaki Y, Nagai Y, Shimamura S, Tsuda M, Kawagucci S *et al.* Heterogeneous
613 composition of key metabolic gene clusters in a vent mussel symbiont population. ISME J
614 2016; 10: 990–1001.
615 Jackrel SL, Yang JW, Schmidt KC, Denef VJ. Host specificity of microbiome assembly and its
616 fitness effects in phytoplankton. ISME J. 2021; 15(3): 774–788.
617 Kang DD, Li F, Kirton E, Thomas A, Egan R, An H *et al*. MetaBAT 2: an adaptive binning
618 algorithm for robust and efficient genome reconstruction from metagenome assemblies.
619 PeerJ. 2019; 7: e7359.
620 Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: Analysis of Next Generation Sequencing
621 Data. BMC Bioinformatics. 2014; 15(1): 356.
622 Laetsch DR, Blaxter ML. BlobTools: Interrogation of genome assemblies [version 1; peer
623 review: 2 approved with reservations]. F1000Research 2017; 6: 1287.
624 Lehmann S, Johnston AWB, Curson ARJ, Todd JD, Cook AM. SoeABC, a novel sulfite
625 dehydrogenase in Roseobacters. In: Muyzer G, Stams AJM (eds). EMBO Workshop on
626 Microbial Sulfur Metabolism. (Sieca Repro, Delft, 2012) pp 29.
627 Lenk S, Moraru C, Hahnke S, Arnds J, Richter M, Kube M *et al. Roseobacter* clade bacteria are
628 abundant in coastal sediments and encode a novel combination of sulfur oxidation genes.
629 ISME J. 2012; 6(12): 2178–2187.
630 Levin LA, Amon DJ, Lily H. Challenges to the sustainability of deep-seabed mining. Nat Sustain
631 2020; 3: 784–794.
632 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G *et al.* The Sequence
633 alignment/map (SAM) format and SAMtools. Bioinformatics 2009; 25: 2078–2079.
634 Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or
635 nucleotide sequences. Bioinformatics. 2006; 22(13): 1658–1659.
636 Luo J, Lyu M, Chen R, Zhang X, Luo H, Yan C. SLR: a scaffolding algorithm based on long
637 reads and contig classification. BMC Bioinformatics 2019; 20: 539.
638 Mitarai S, Watanabe H, Nakajima Y, Shchepetkin AF, McWilliams JC. Quantifying dispersal
639 from hydrothermal vent fields in the western Pacific Ocean. PNAS 2016; 113(11): 2976–
640 2981.

641 Mottl MJ, Seewald JS, Wheat CG, Tivey MK, Michael PJ, Proskurowski G *et al.* Chemistry of
642 hot springs along the Eastern Lau Spreading Center. Geochim Cosmochim Acta. 2011;
643 75(4): 1013–1038.
644 Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic
645 assembler. Genome Res. 2017; 27(5): 824–834.
646 Nyholm SV, McFall-Ngai MJ. The winnowing: establishing the squid-*Vibrio* symbiosis. Nat Rev
647 Microbiol. 2004; 2: 632–642.
648 Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D *et al.* vegan: Community
649 Ecology Package. R package version 2.5-7 2020. https://CRAN.R-
650 project.org/package=vegan
651 Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary
652 analyses in R. Bioinformatics. 2019; 35(3): 526–528.
653 Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality
654 of microbial genomes recovered from isolates, single cells, and metagenomes. Genome
655 Res. 2015; 25(7):1043– 1055.
656 Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware
657 quantification of transcript expression. Nat Methods. 2017; 14: 417–419.
658 Perez M, Juniper SK. Insights into symbiont population structure among three vestimentiferan
659 tubeworm host species at Eastern Pacific spreading centers. Appl Environ Microbiol. 2016;
660 82(17): 5197–5205.
661 Picelli S, Björklund AK, Reinius B, Sagasser S, Winberg G, Sandberg R. Tn5 transposase and
662 tagmentation procedures for massively scaled sequencing projects. Genome Res. 2014;
663 24(12): 2033–2040.
664 Podowski EL, Ma S, Luther GW III, Wardrop D, Fisher CR. Biotic and abiotic factors affecting
665 distributions of megafauna in diffuse flow on andesite and basalt along the Eastern Lau
666 Spreading Center, Tonga. Mar Ecol Prog Ser. 2010; 418: 25–45.
667 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D *et al.* PLINK: a toolset
668 for whole-genome association and population-based linkage analysis. Am J Hum Genet.
669 2007; 81(3): 559–575.
670 R Core Team. R: A language and environment for statistical computing. (R Foundation for
671 Statistical Computing, Vienna, Austria, 2020) https://www.R-project.org/.
672 Reynolds KC, Watanabe H, Strong EE, Sasaki T, Uematsu K, Miyake H *et al.* New molluscan
673 larval form: brooding and development in a hydrothermal vent gastropod, *Ifremeria*
674 *nautilei* (Provannidae). Biol Bull. 2010; 219(1): 7–11.
675 Rinta-Kanto JM, Bürgmann H, Gifford SM, Sun S, Sharma S, del Valle DA *et al.* Analysis of
676 sulfur-related transcription by *Roseobacter* communities using a taxon-specific functional
677 gene microarray. Environ Microbiol. 2011; 13(2): 453–467.
678 Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of
679 RNA-seq data. Genome Biol. 2010; 11: R25.
680 Russell SL. Transmission mode is associated with environment type and taxa across bacteria-
681 eukaryote symbioses: a systematic review and meta-analysis. FEMS Microbiol Lett. 2019;
682 366(3): fnz013.
683 Sanders J, Beinart R, Stewart F, Delong EF, Girguis PR. Metatranscriptomics reveal differences
684 in in situ energy and nitrogen metabolism among hydrothermal vent snail symbionts. ISME
685 J. 2013; 7: 1556–1567.

686 Seah BK, Gruber-Vodicka HR. gbtools: interactive visualization of metagenome bins in R. Front
687 Microbiol. 2015; 6: 1451.
688 Seppey M, Manni M, Zdobnov EM. BUSCO: Assessing genome assembly and annotation
689 completeness. Methods Mol Biol. 2019; 1962: 227–245.
690 Seston SL, Beinart RA, Sarode N, Shockey AC, Ranjan P, Ganesh S *et al.* Metatranscriptional
691 response of chemoautotrophic *Ifremeria nautilei* endosymbionts to differing sulfur
692 regimes. Front Microbiol. 2016; 7: 1074.
693 Song L, Florea L. Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads.
694 Gigascience. 2015; 4: 48.
695 Suzuki Y, Kojima S, Sasaki T, Suzuki M, Utsumi T, Watanabe H *et al.* Host-symbiont
696 relationships in hydrothermal vent gastropods of the genus *Alviniconcha* from the
697 Southwest Pacific. Appl Environ Microbiol. 2006; 72: 1388–1393.
698 Tivey MK. Generation of seafloor hydrothermal vent fluids and associated mineral deposits.
699 Oceanography 2007; 20: 50–65.
700 Ücker M, Ansorge R, Sato Y, Sayavedra L, Breusing C, Dubilier N. Deep-sea mussels from a
701 hybrid zone on the Mid-Atlantic Ridge host genetically indistinguishable symbionts. ISME
702 J. 2021; doi: 10.1038/s41396-021-00927-9.
703 Van der Auwera GA, O'Connor BD. Genomics in the Cloud. (O'Reilly Media, Inc., 2020) pp
704 834.
705 Verma SC, Miyashiro T. Quorum sensing in the squid-Vibrio symbiosis. Int J Mol Sci. 2013;
706 14(8): 16386–16401.
707 Vrijenhoek RC. Genetics and evolution of deep-sea chemosynthetic bacteria and their
708 invertebrate hosts. In: Kiel S (ed). The Vent and Seep Biota. (Springer, Dordrecht,
709 Netherlands, 2010) pp 15–49.
710 Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S *et al.* Pilon: an integrated
711 tool for comprehensive microbial variant detection and genome assembly improvement.
712 PLoS ONE 2014; 9(11): e112963.
713 Walker L, Lagunas B, Gifford ML. Determinants of host range specificity in legume-rhizobia
714 symbiosis. Front Microbiol. 2020; 11: 585749.
715 Wang Q, Liu J, Zhu H. Genetic and molecular mechanisms underlying symbiotic specificity in
716 legume-rhizobium interactions. Front Plant Sci. 2018; 9: 313.
717 Warén A, Bouchet P. New records, species, genera, and a new family of gastropods from
718 hydrothermal vents and hydrocarbon seeps. Zool Scr. 1993; 22: 1–90.
719 Whitlock MC, Lotterhos KE. Reliable detection of loci responsible for local adaptation:
720 inference of a null model through trimming the distribution of F(ST). Am Nat. 2015;
721 186(Suppl 1): S24–36.
722 Wickham H. ggplot2: Elegant Graphics for Data Analysis. (Springer, New York, 2016) pp 260.
723 Wilm A, Aw PP, Bertrand D, Yeo GH, Ong SH, Wong CH *et al.* LoFreq: a sequence-quality
724 aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from
725 high-throughput sequencing datasets. Nucleic Acids Res. 2012; 40(22): 11189–11201.
726 Wu Y-W, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover
727 genomes from multiple metagenomic datasets, Bioinformatics 2016; 32(4): 605–607.
728 Xu GC, Xu TJ, Zhu R, Zhang Y, Li SQ, Wang HW *et al.* LR_Gapcloser: a tiling path-based gap
729 closer that uses long reads to complete genome assembly. Gigascience. 2019; 8(1): giy157.

730 Zimmermann J, Lott C, Weber M, Ramette A, Bright M, Dubilier N *et al.* Dual symbiosis with
731 co-occurring sulfur-oxidizing symbionts in vestimentiferan tubeworms from a
732 Mediterranean hydrothermal vent. Environ Microbiol. 2014; 16(12): 3638–3656.
733

734 **Figure Captions**

735 **Fig. 1** Sampling map for *Alviniconcha* and *Ifremeria* species in the Eastern Lau Spreading
736 Center and Tonga Volcanic Arc.

737

738 **Fig. 2** Principal component plots for *Alviniconcha* and *Ifremeria* host species based on genetic
739 covariance matrices.

740

741 **Fig. 3** Principal coordinate plots for *Alviniconcha* and *Ifremeria* symbionts based on relative
742 allele counts transformed into Bray-Curtis dissimilarities. Allele counts approximate the relative
743 proportions of different symbiont strains within host individuals.

744

745 **Fig. 4** Principal coordinate plot based on Bray-Curtis dissimilarities (A) and presence/absence
746 heatmap of differentially preserved genes (B) for the Gamma1 symbiont of *A. kojimai* and *A.*
747 *strummeri*. Strains of this symbiont are clearly distinct between the two host species, even when
748 these taxa co-occur (see Fig. S3). Segregation of symbiont populations by host species along the
749 first ordination axis suggests that host affinity is a stronger predictor than geography for
750 symbiont composition between species. Dark and light blue colors in the heatmap indicate
751 presence and absence of genes, respectively. Dendrograms show similarities between samples
752 based on their gene content profiles.

753

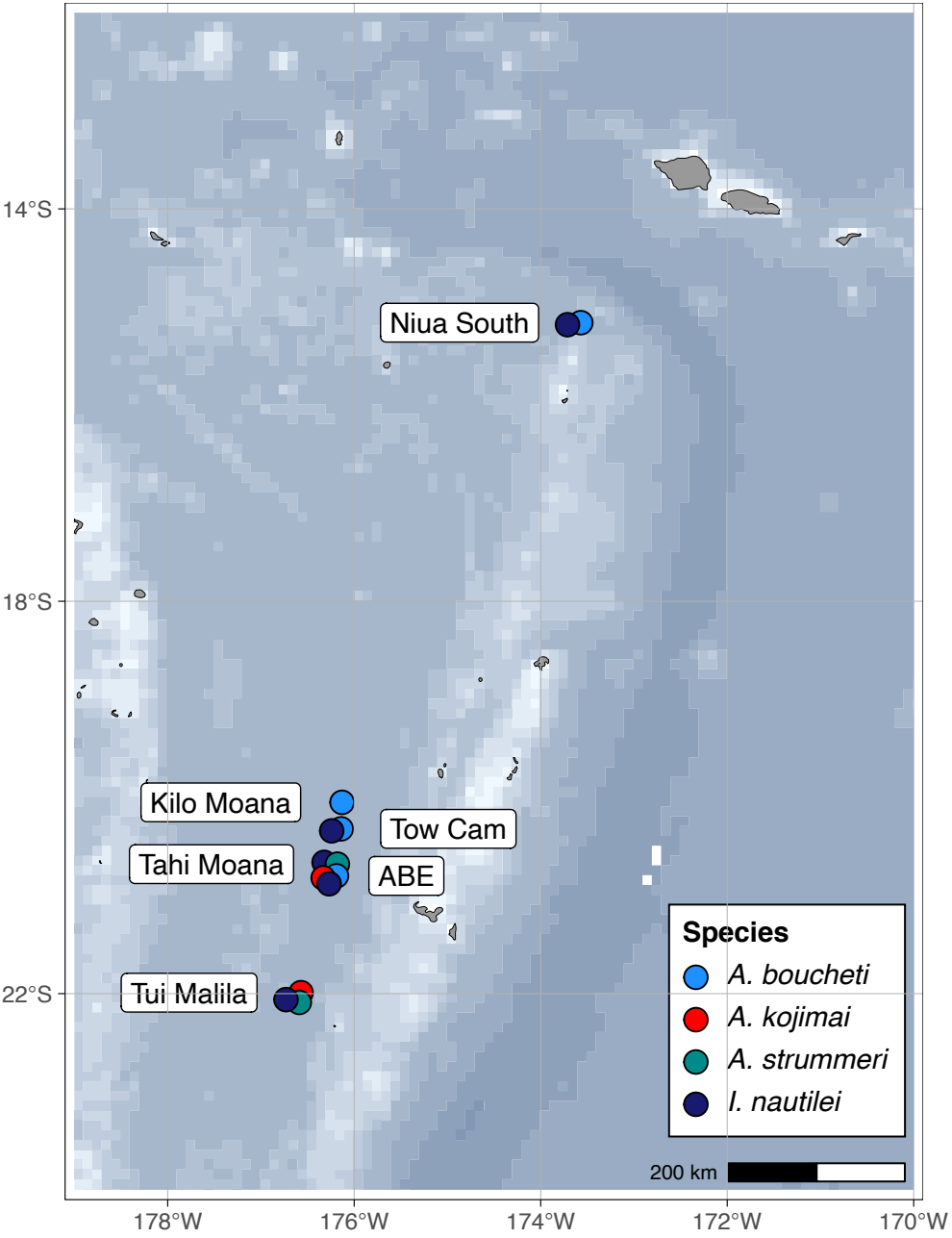
754 **Fig. 5** Presence/absence heatmap for differentially preserved genes.

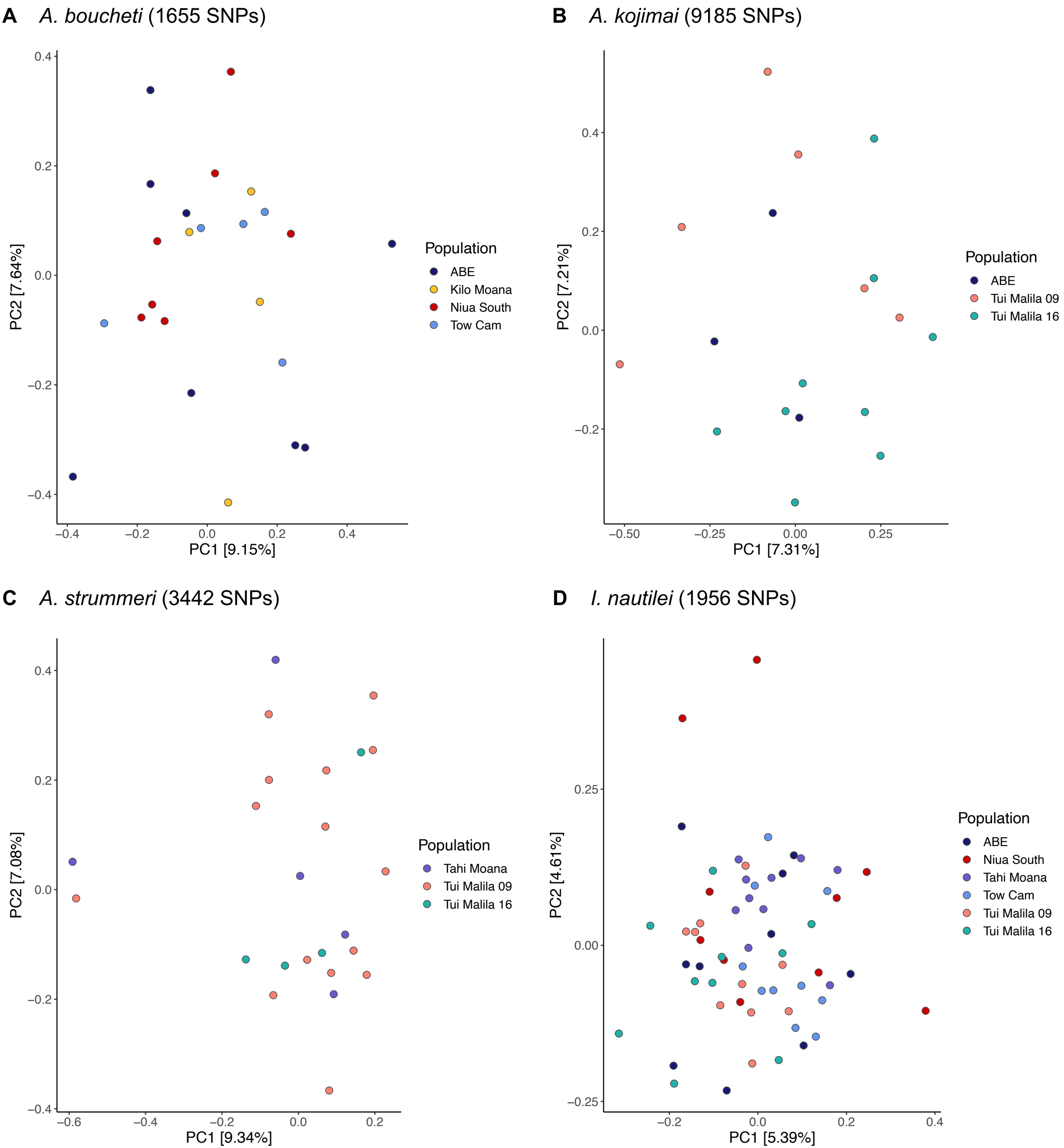
755

756 **Fig. 6** Redundancy analysis plots of the first two ordination axes for *Alviniconcha* and *Ifremeria*
757 symbionts. Small grey dots in the center of the plot indicate genetic variants, while large colored
758 dots represent intra-host symbiont populations sampled from different vent localities. Vectors
759 show the environmental predictors. Redundancy analyses were conditioned by geography for all
760 symbionts except for Gamma1 and GammaLau. Genotype-environment associations were not
761 significant for GammaLau.

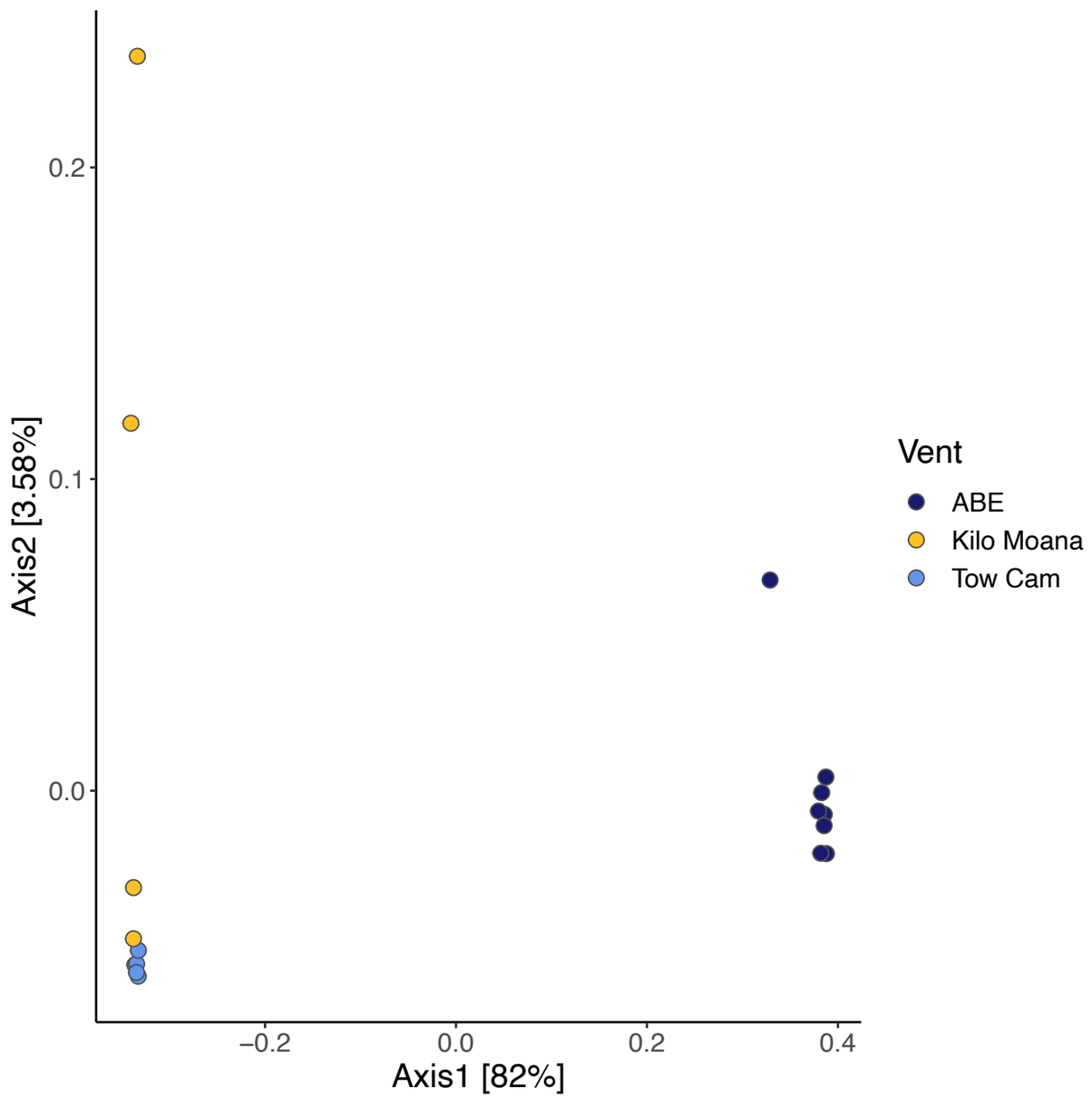
Table 1 Sampling information for *Alviniconcha* and *Ifremeria* species from the Eastern Lau Spreading Center and Tonga Volcanic Arc. Chemical concentrations are mean endmember values for each vent field obtained from Mottl et al. (2011), Beinart et al. (2012), Flores et al. (2012), and unpublished data provided by J. Seewald (Woods Hole Oceanographic Institution) and A. Diehl (MARUM). n.a. = not available.

Vent site	Species	Coordinates	Year	Depth [m]	H_2 [mM]	CO_2 [mM]	CH_4 [mM]	H_2S [mM]	Fe [mM]	Mn [mM]
Tonga Volcanic Arc										
Niua South	*A. boucheti* *I. nautilei*	15°09.79'S 173°34.48'W	2016	1164	0.037	n.a.	0.005	1.911	n.a.	n.a.
Eastern Lau Spreading Center										
Kilo Moana	*A. boucheti*	20°03.23'S 176°08.01'W	2009	2614	0.382	8.227	0.031	5.432	3.062	0.643
Tow Cam	*A. boucheti* *I. nautilei*	20°18.97'S 176°08.19'W	2009	2711–2722	0.150	10.648	0.045	4.795	0.316	0.389
Tahi Moana	*A. strummeri* *I. nautilei*	20°40.94'S 176°11.01'W	2016	2234–2273	0.097	7.191	0.038	3.548	0.286	0.525
ABE	*A. boucheti* *A. kojimai* *I. nautilei*	20°45.79'S 176°11.47'W	2009 2016	2130–2152	0.079	5.980	0.048	3.060	0.245	0.330
Tu'i Malila	*A. kojimai* *A. strummeri* *I. nautilei*	21°59.36'S 176°34.10'W	2009 2016	1860–1887	0.084	12.912	0.037	2.294	0.203	0.406

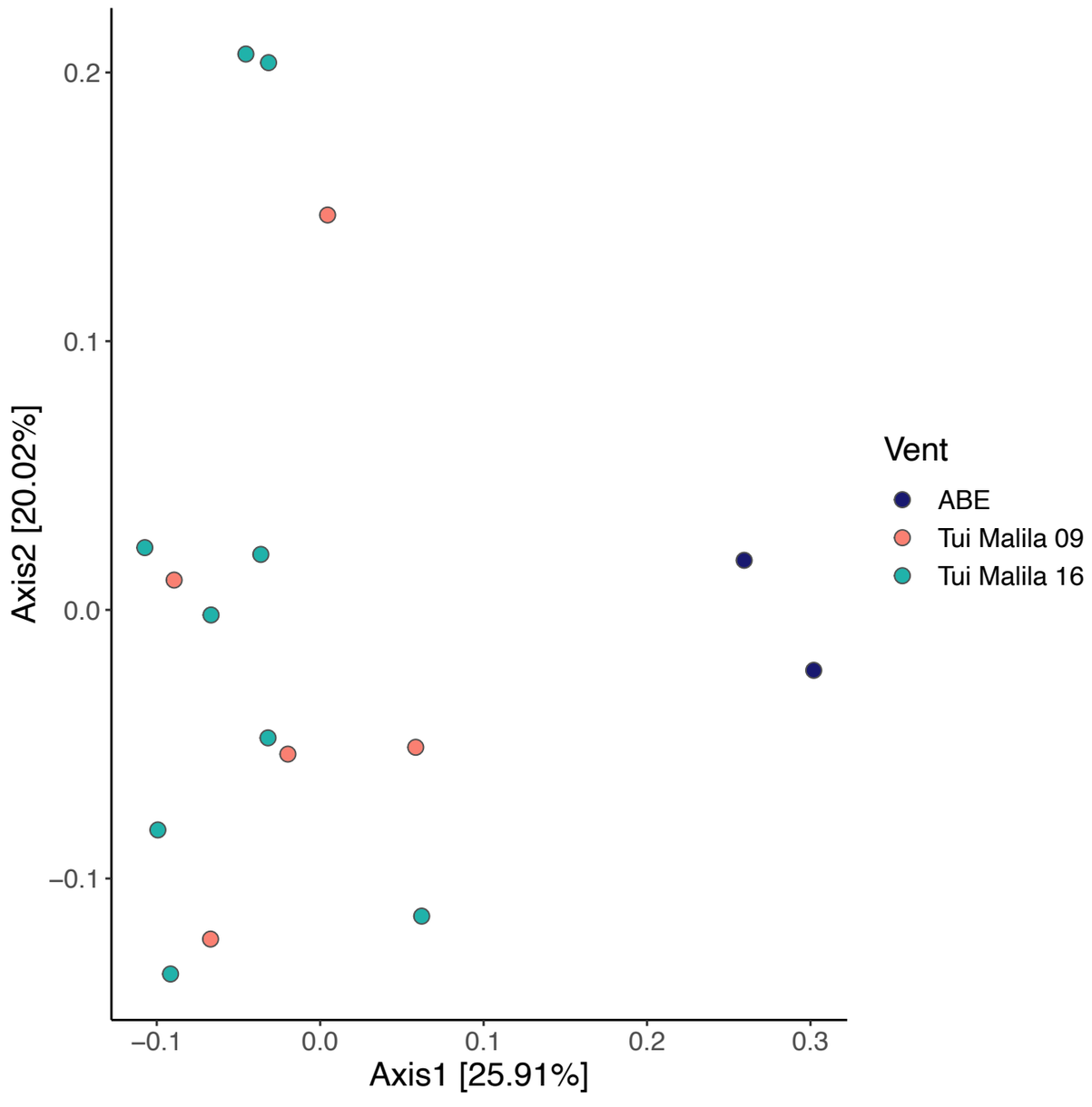

14°S
18°S
22°S
178°W
176°W
174°W
172°W
170°W
Niua South
Kilo Moana
Tow Cam
Tahi Moana
ABE
Tui Malila
Species
A. boucheti
A. kojimai
A. strummeri
I. nautilei
200 km


A
A. boucheti (1655 SNPs)
PC2 [7.64%]
PC1 [9.15%]
Population
ABE
Kilo Moana
Niua South
Tow Cam
B
A. kojimai (9185 SNPs)
PC2 [7.21%]
PC1 [7.31%]
Population
ABE
Tui Malila 09
Tui Malila 16
C
A. strummeri (3442 SNPs)
PC2 [7.08%]
PC1 [9.34%]
Population
Tahi Moana
Tui Malila 09
Tui Malila 16
D
I. nautilei (1956 SNPs)
PC2 [4.61%]
PC1 [5.39%]
Population
ABE
Niua South
Tahi Moana
Tow Cam
Tui Malila 09
Tui Malila 16

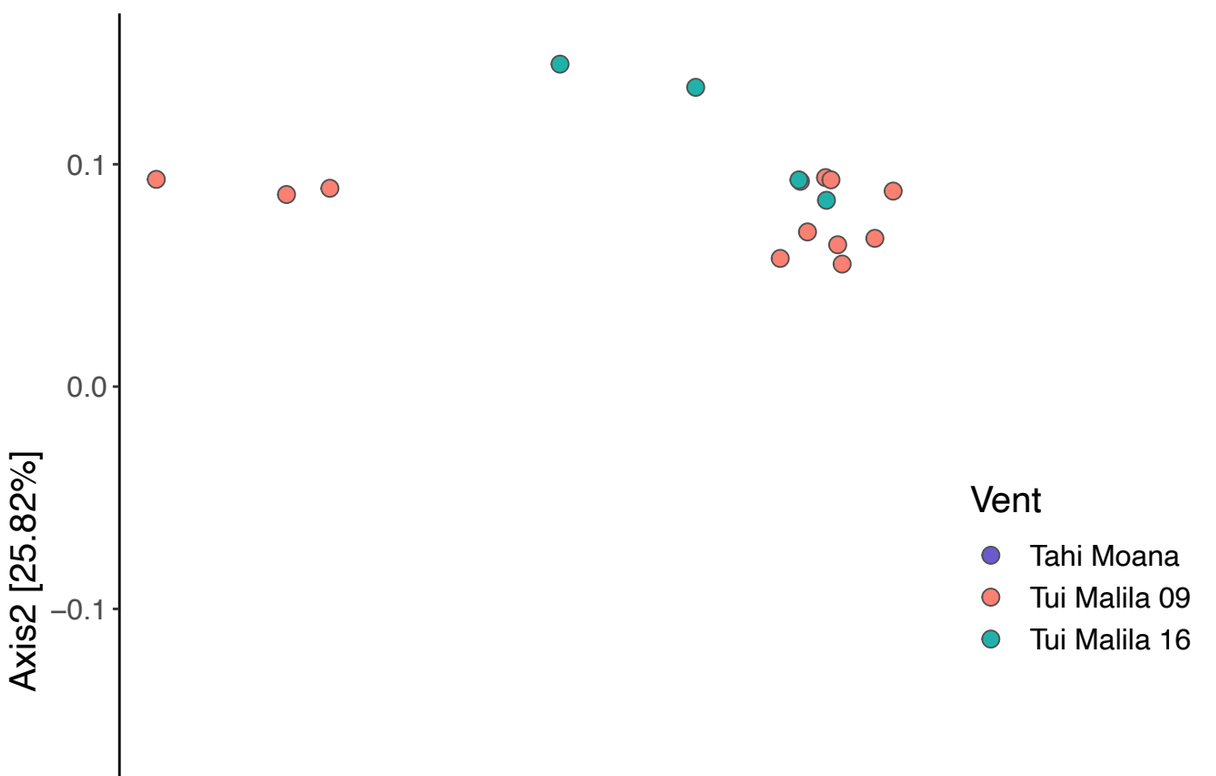
A Epsilon (*A. boucheti*; 7057 variants)



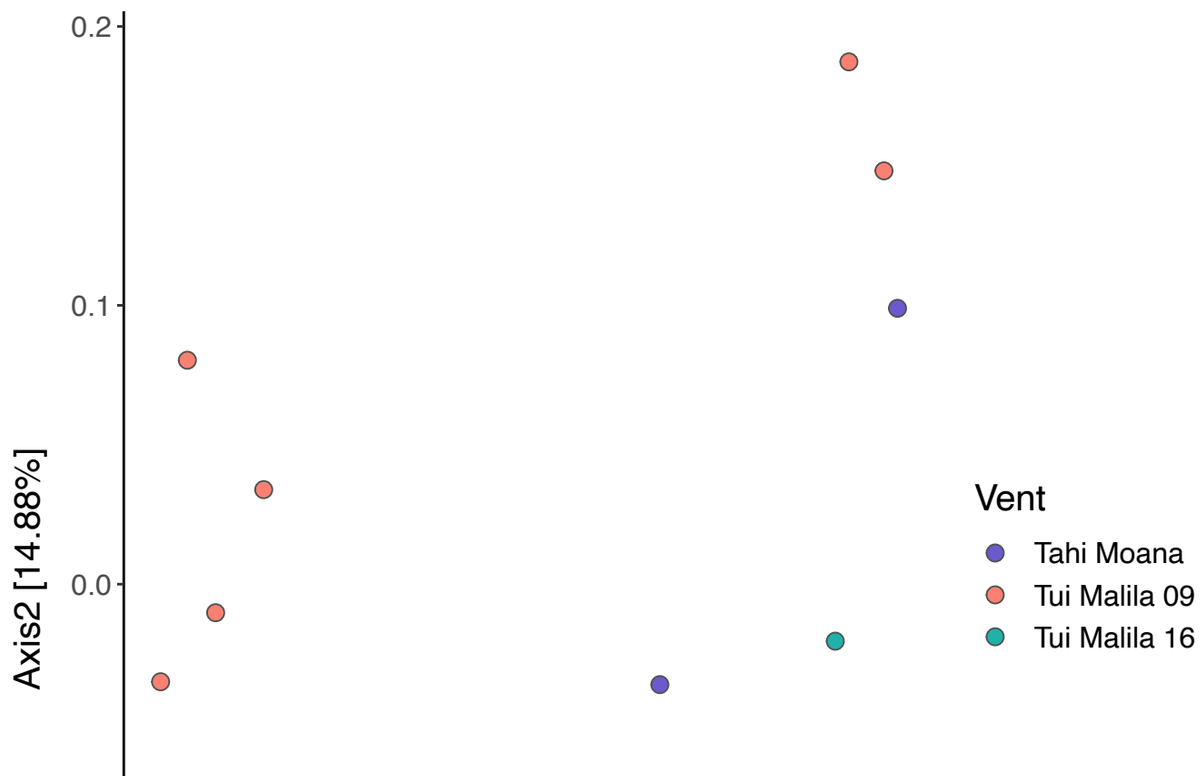
B Gamma1 (*A. kojimai*; 2730 variants)



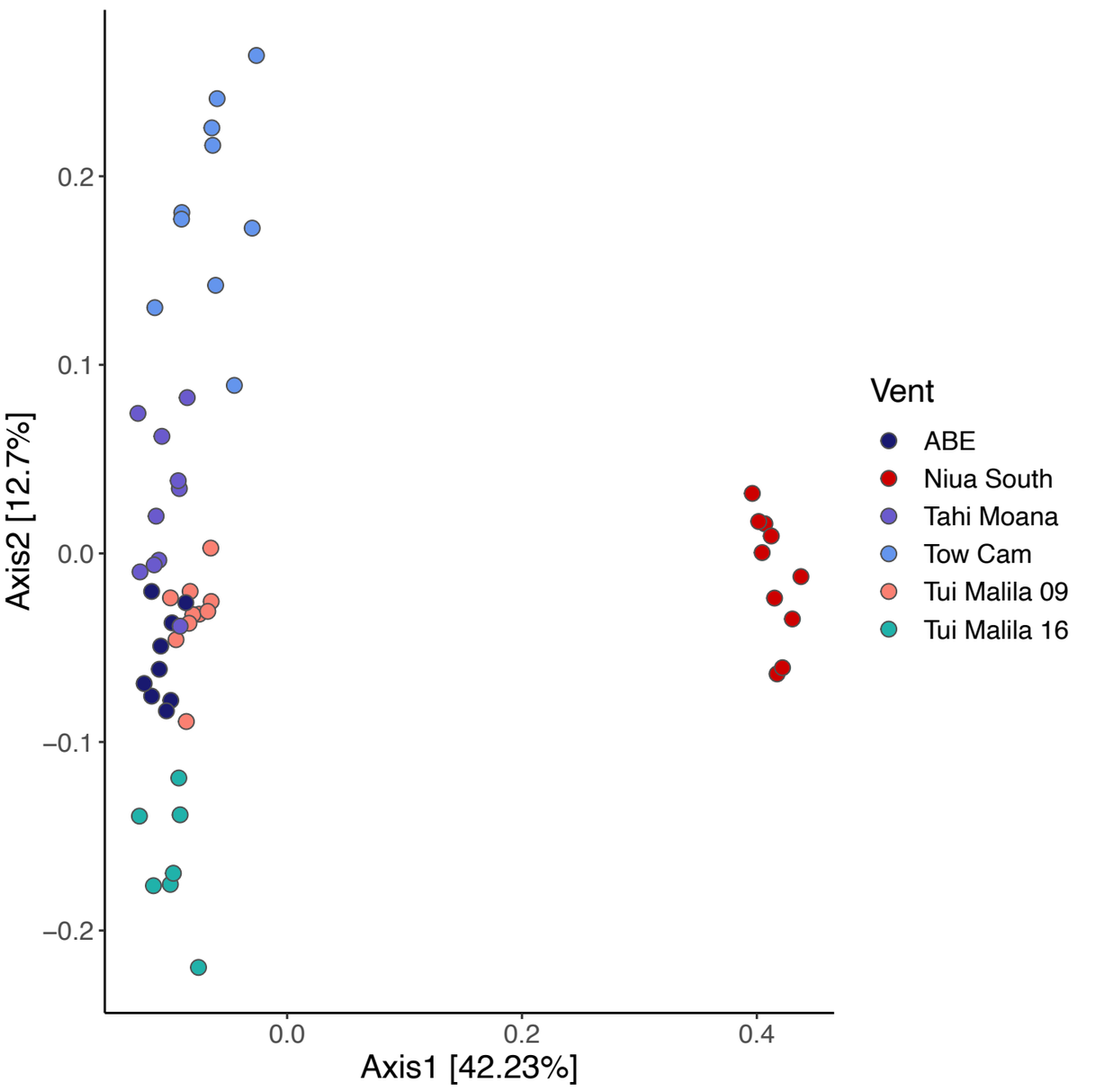
C Gamma1 (*A. strummeri*; 2545 variants)



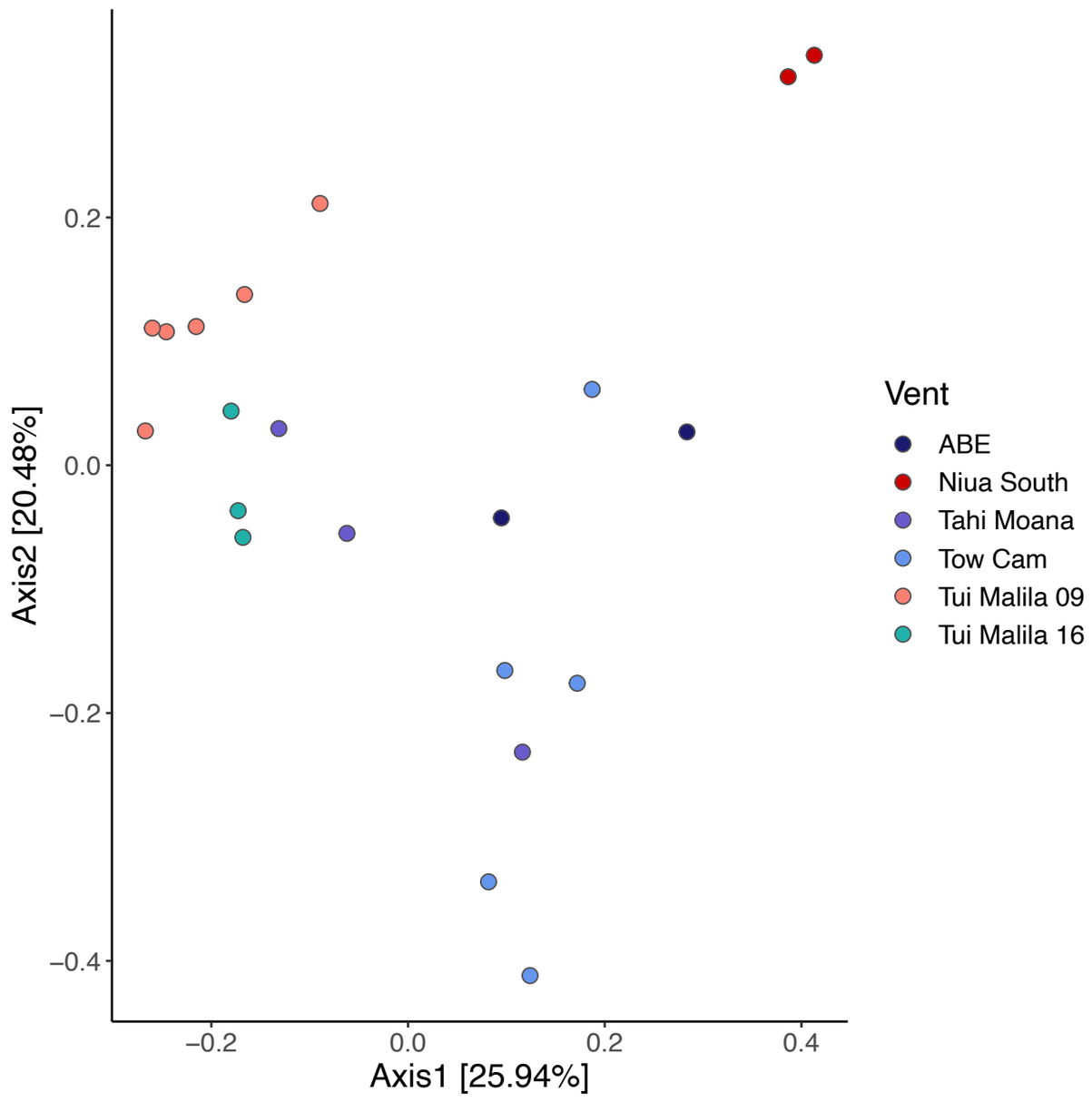
D GammaLau (*A. strummeri*; 192 variants)

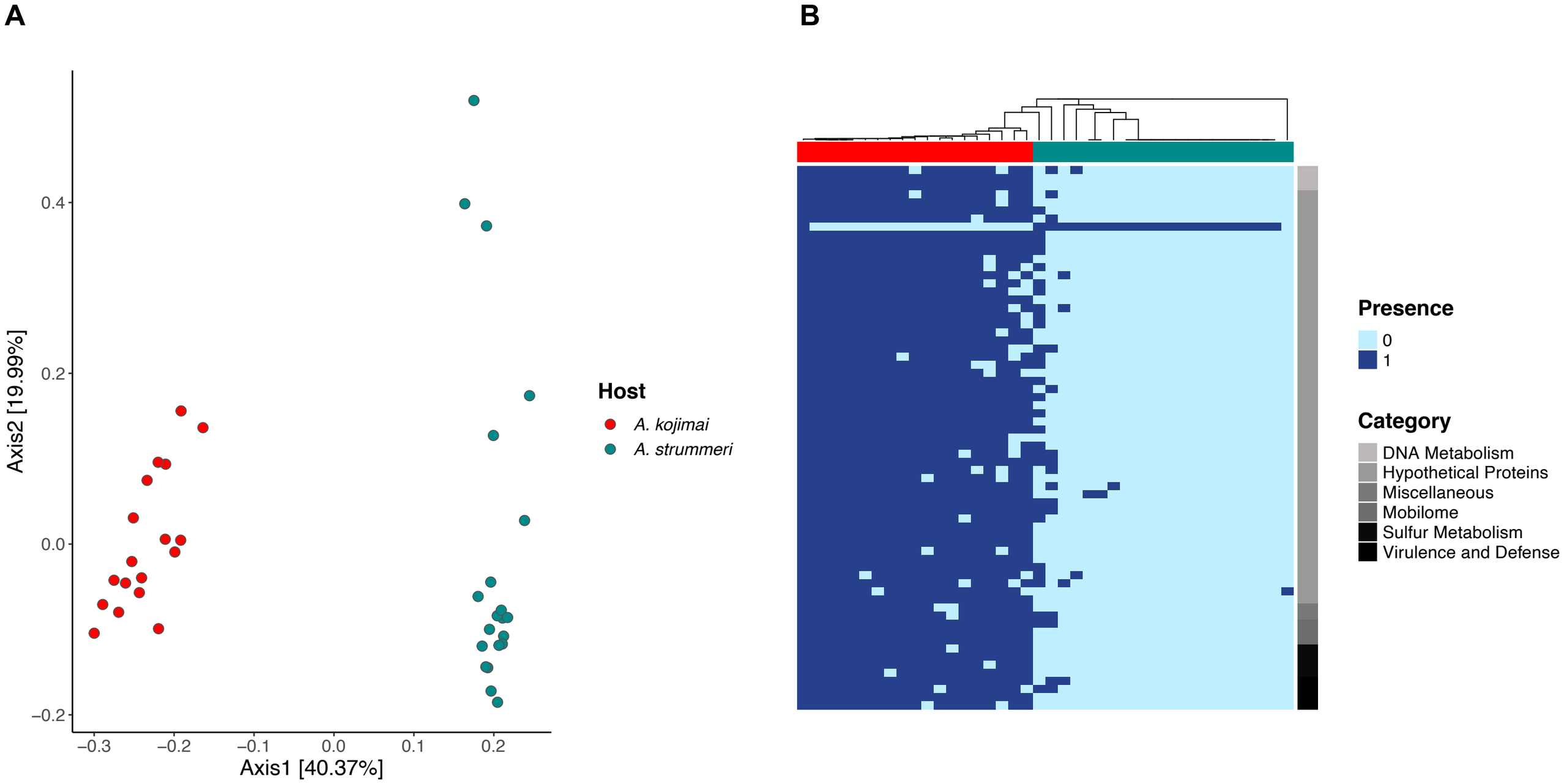


E Ifr-SOX (*I. nautilei*; 1244 variants)

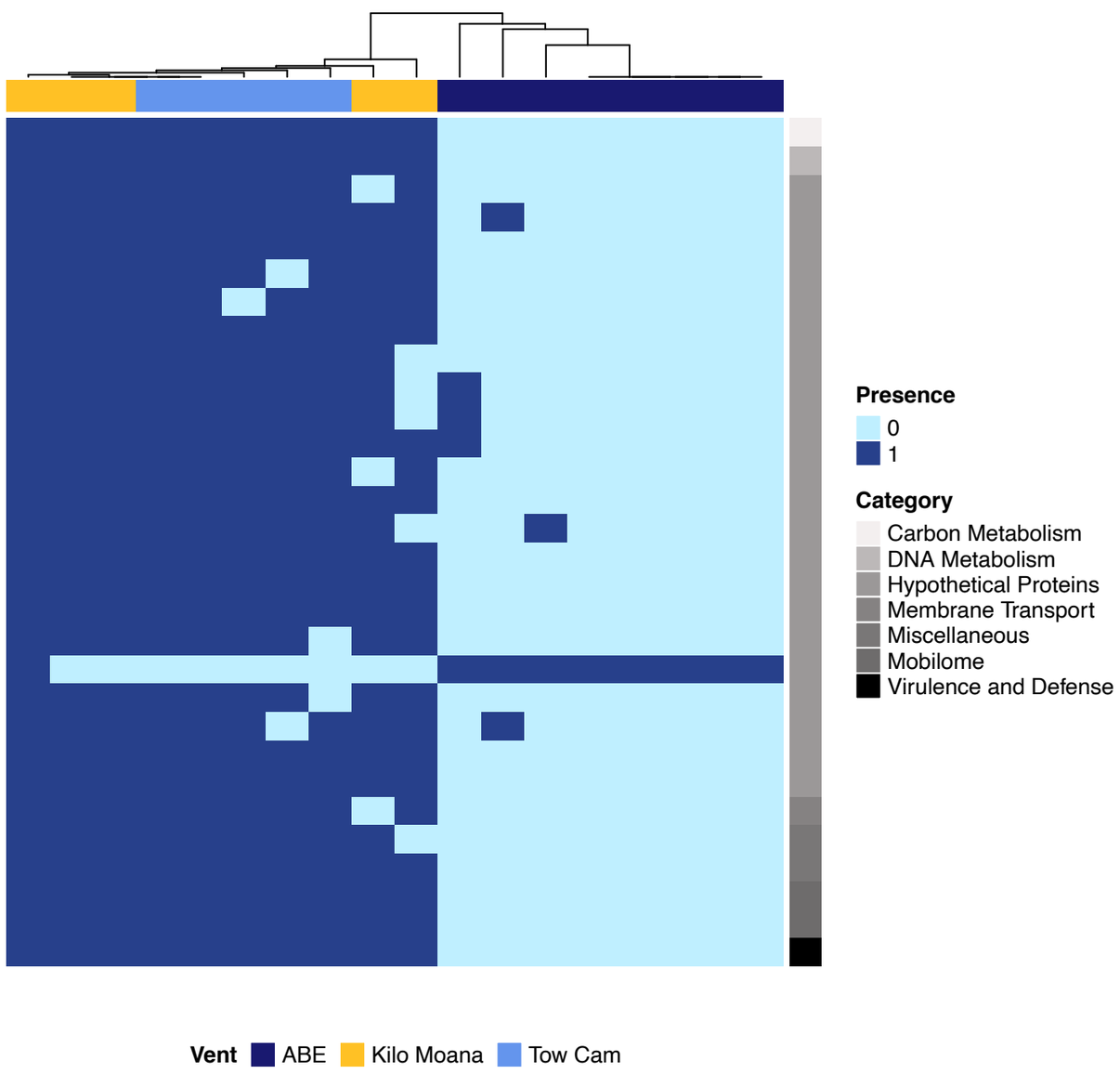


F Ifr-MOX (*I. nautilei*; 239 variants)

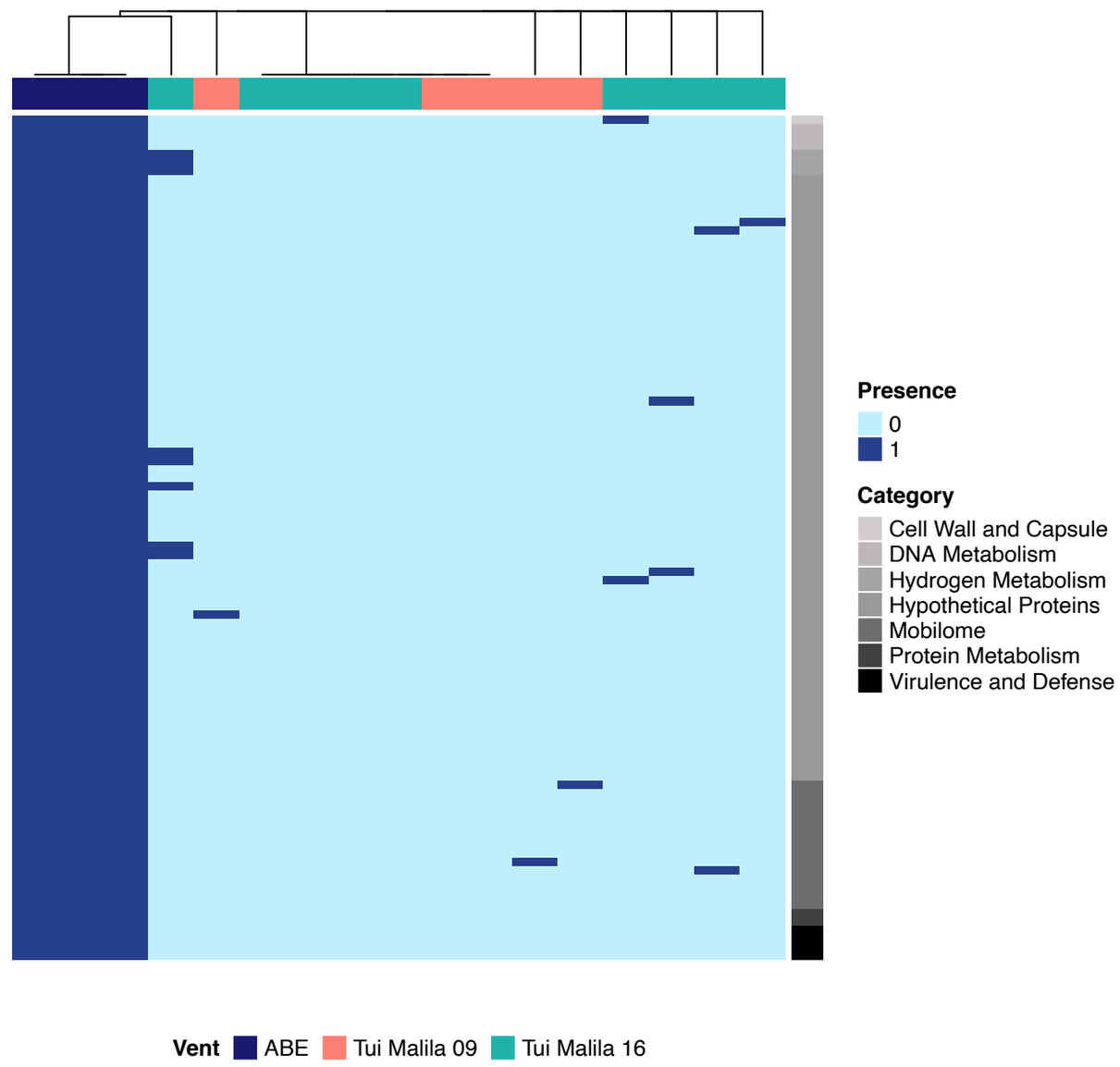



A
B
Axis2 [19.99%]
Axis1 [40.37%]
0.4
0.2
0.0
−0.2
−0.3
−0.2
−0.1
0.0
0.1
0.2
Host
A. kojimai
A. strummeri
Presence
0
1
Category
DNA Metabolism
Hypothetical Proteins
Miscellaneous
Mobilome
Sulfur Metabolism
Virulence and Defense

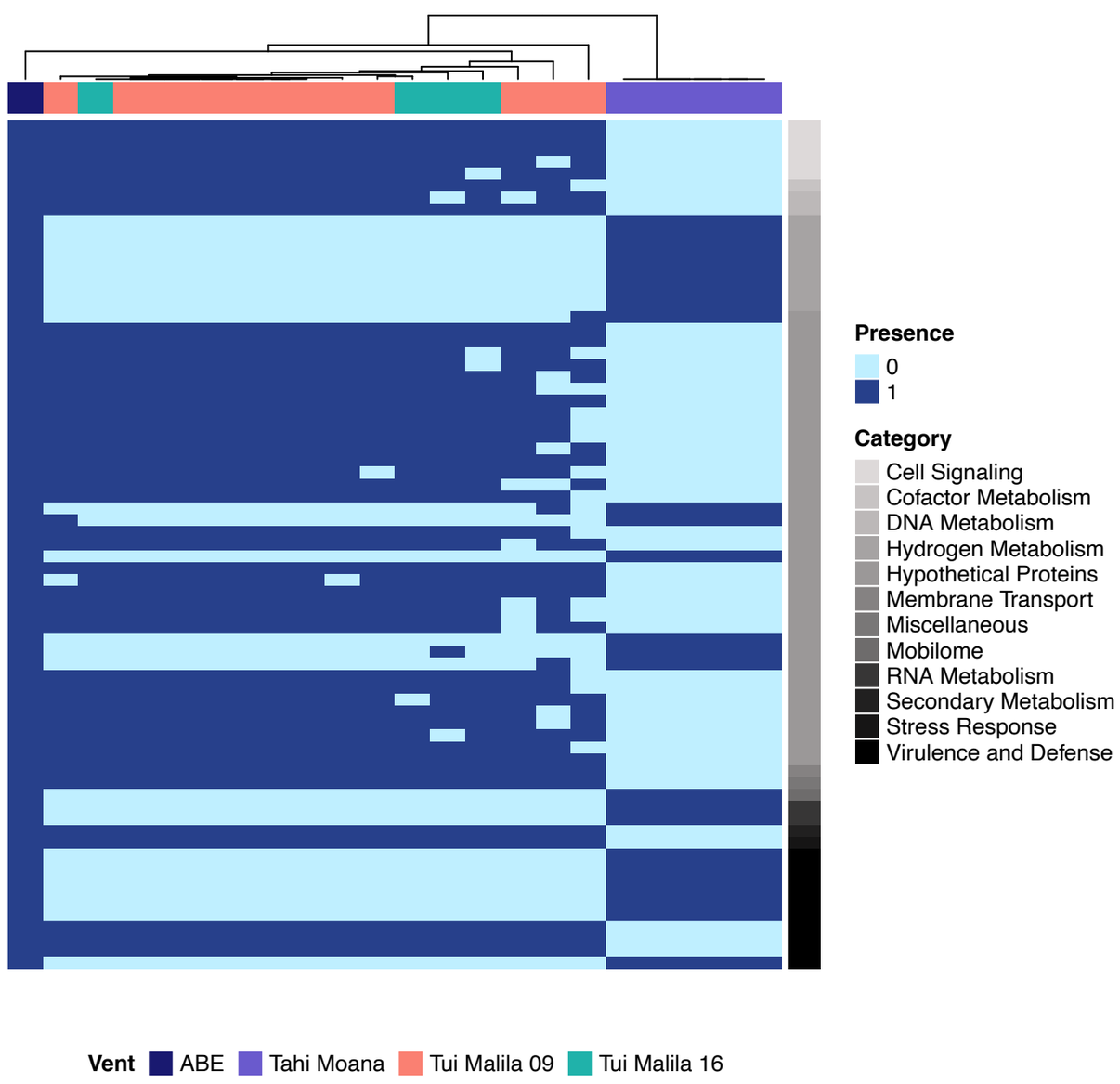
A Epsilon (*A. boucheti*; 30 genes)



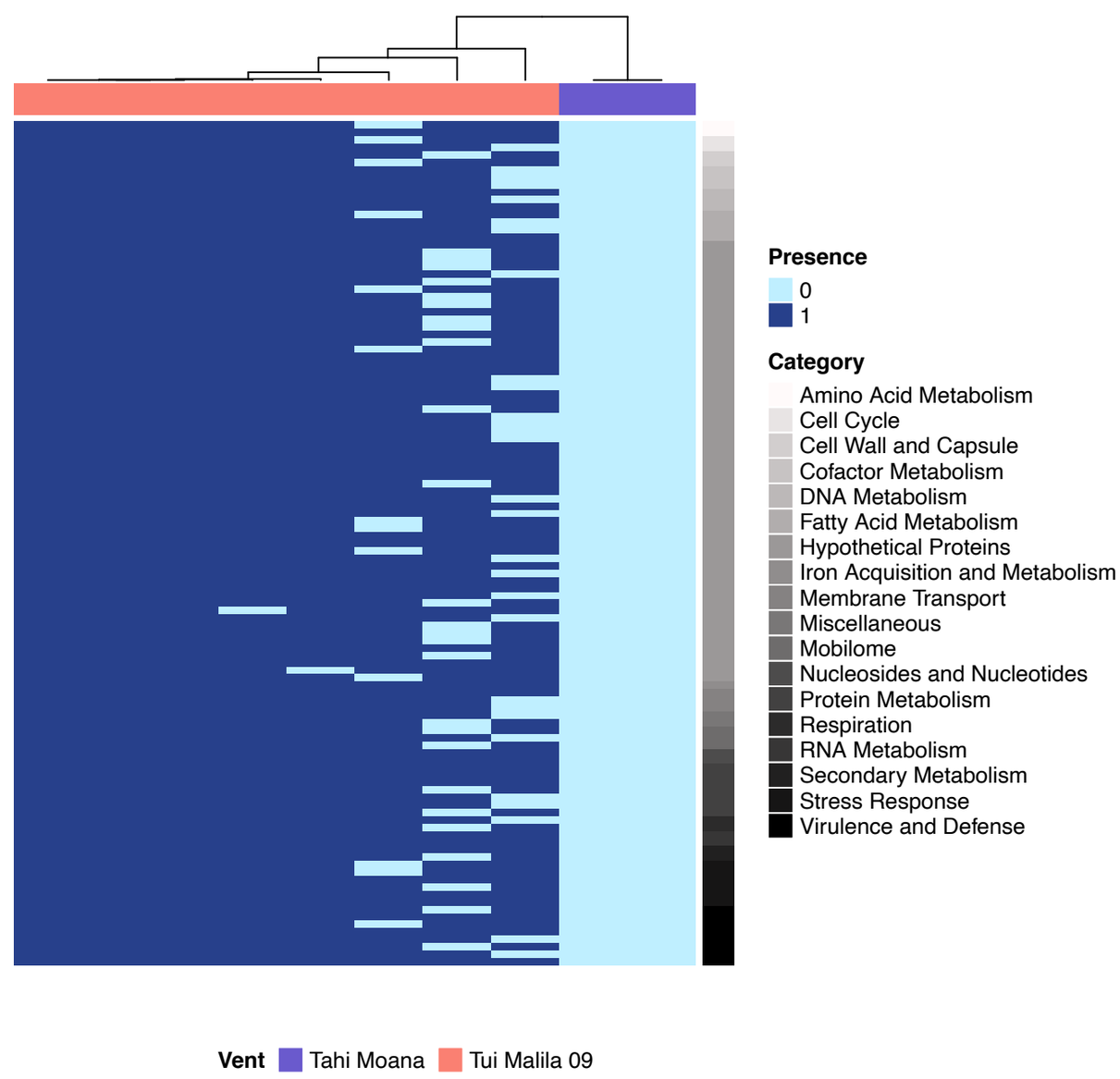
B Gamma1 (*A. kojimai*; 99 genes)



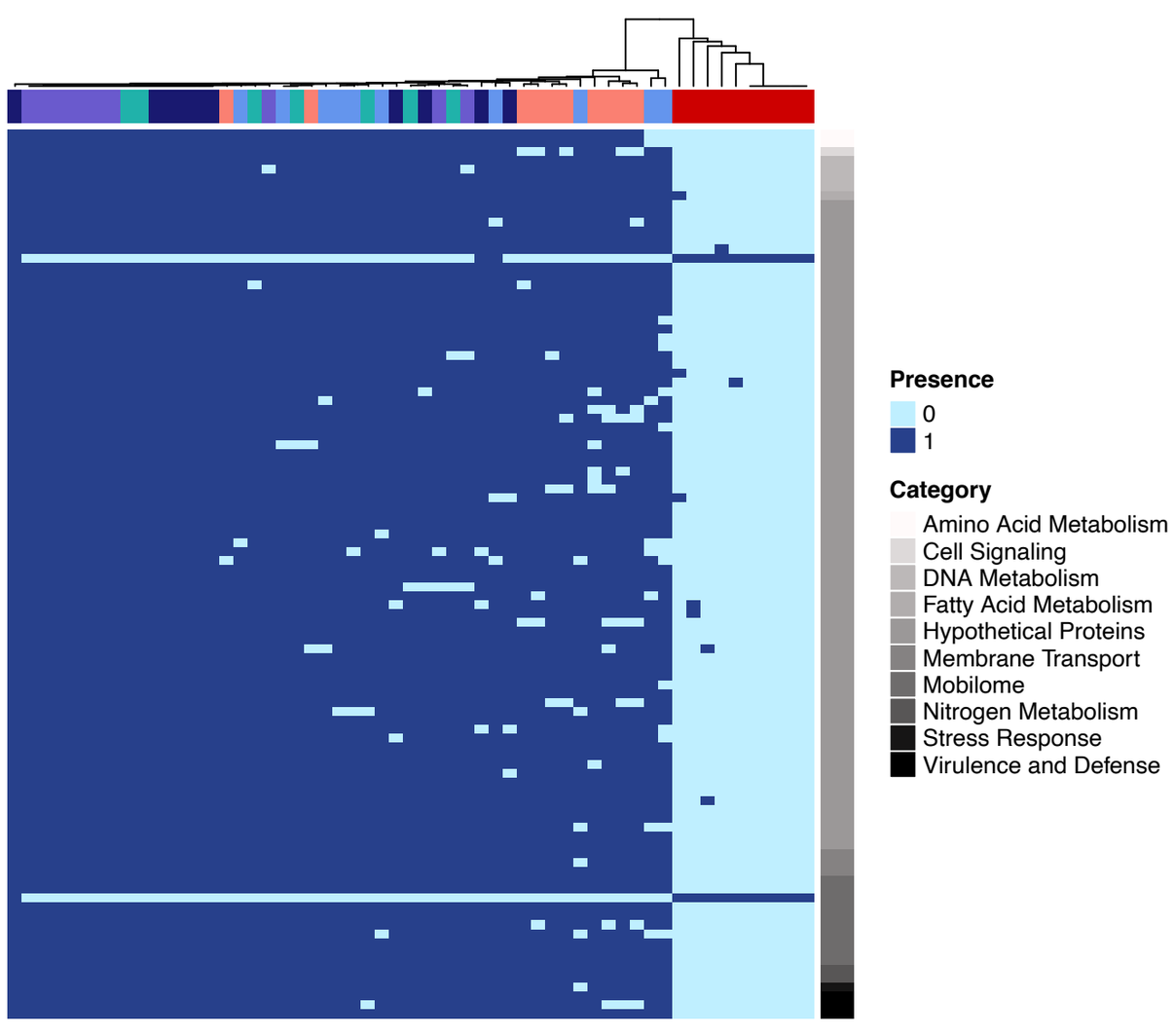
C Gamma1 (*A. strummeri*; 71 genes)



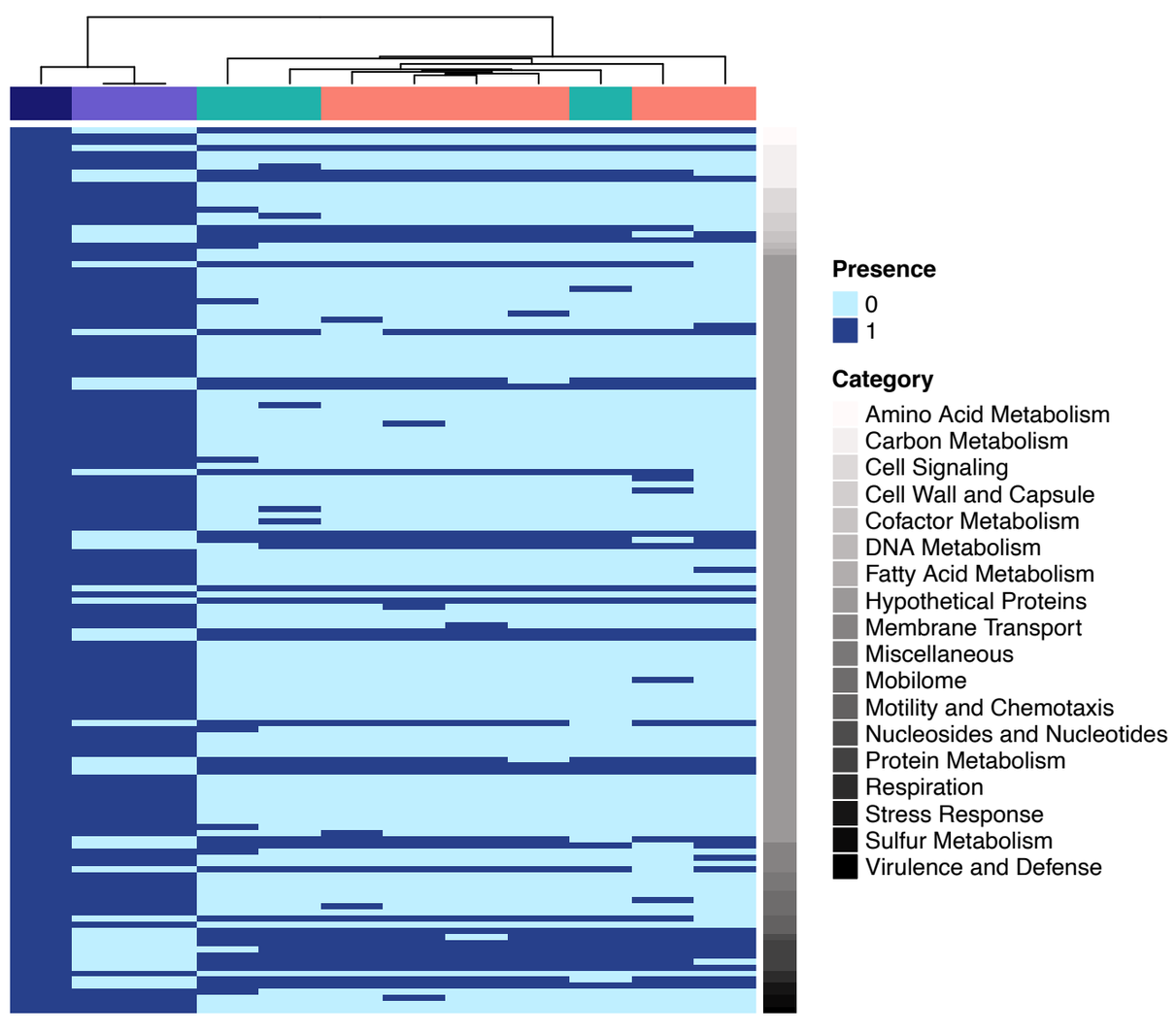
D GammaLau (*A. strummeri*; 113 genes)



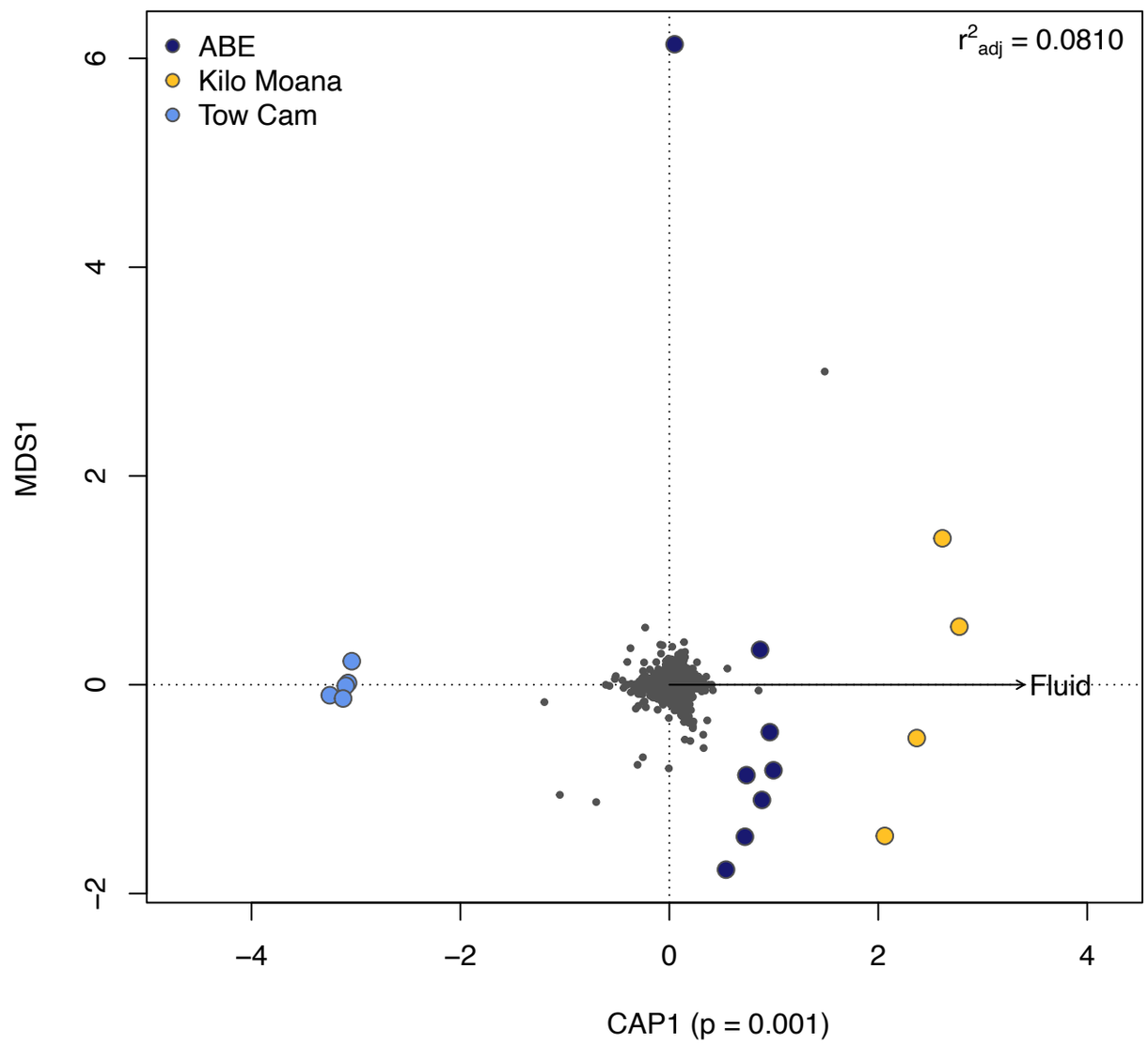
E Ifr-SOX (*I. nautilei*; 100 genes)



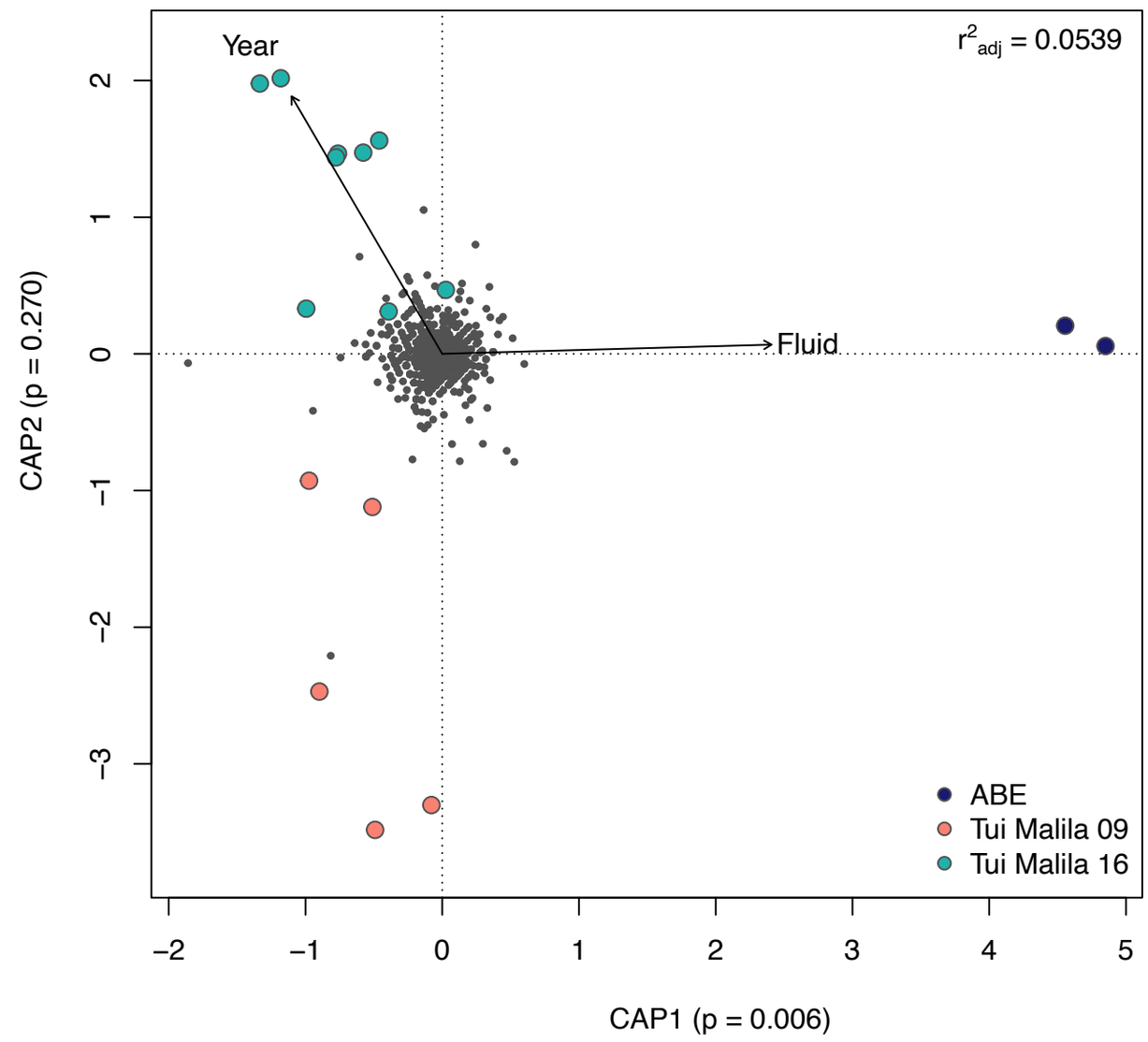
E Ifr-MOX (*I. nautilei*; 145 genes)



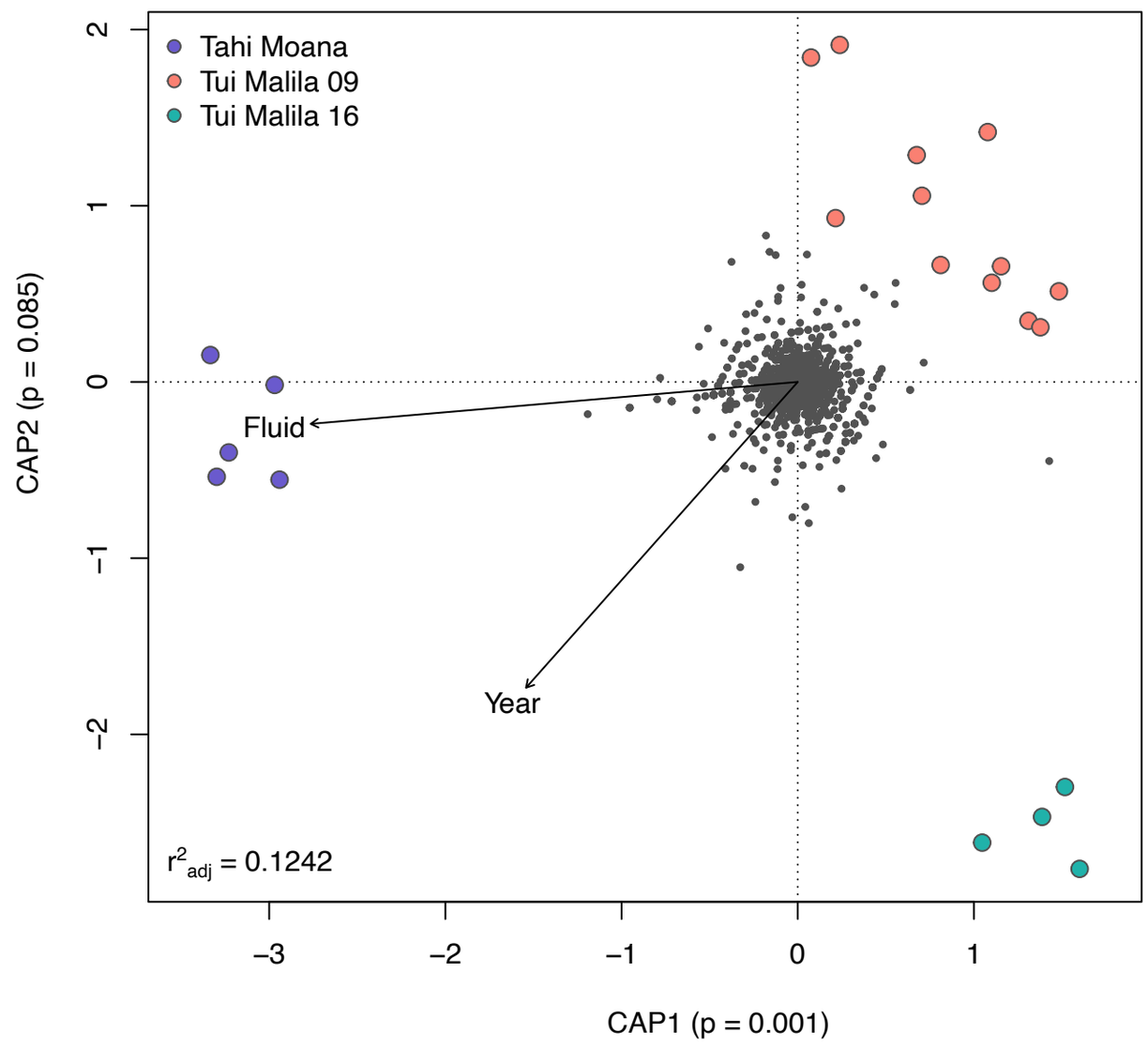
A Epsilon (*A. boucheti*)



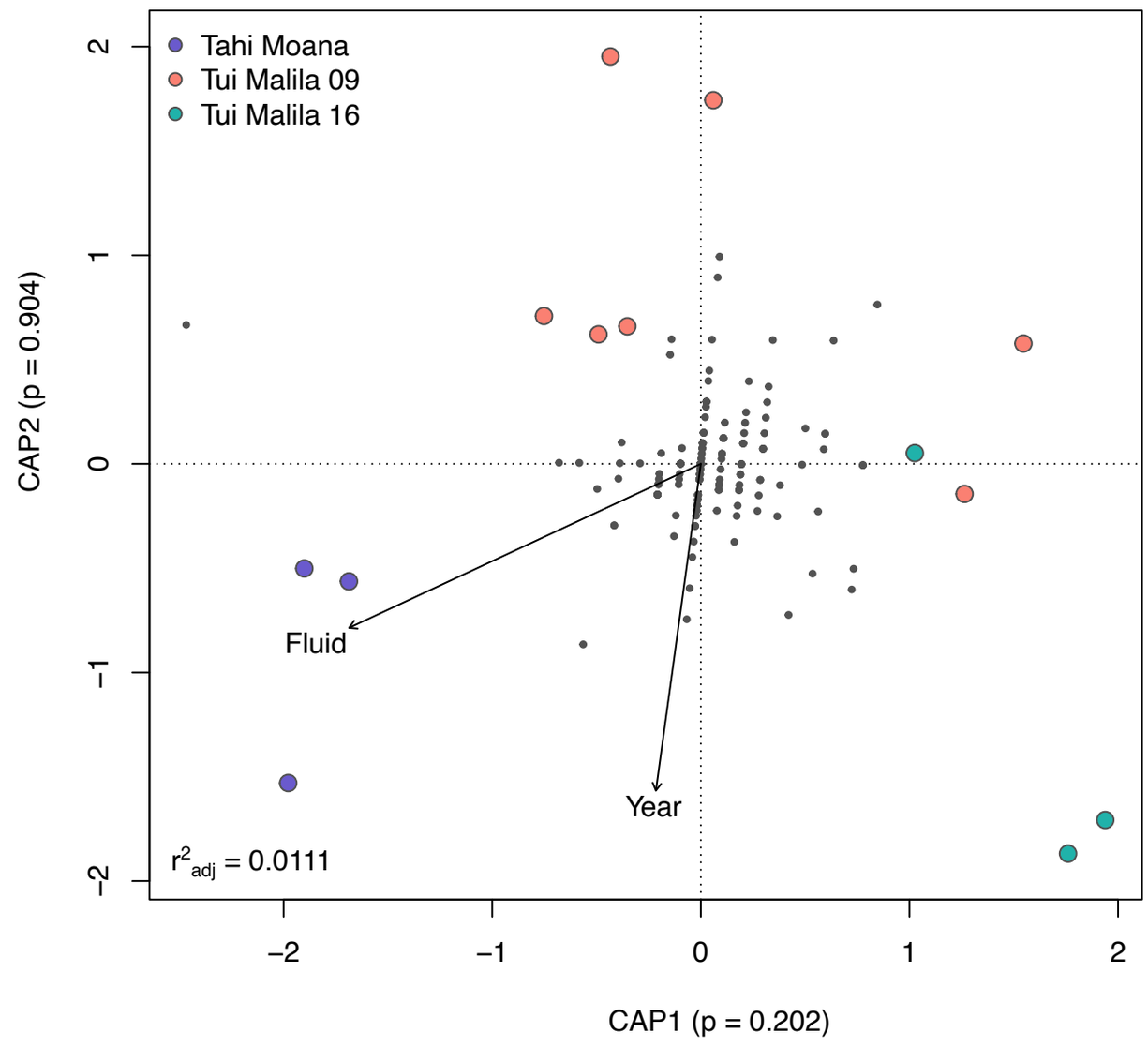
B Gamma1 (*A. kojimai*)



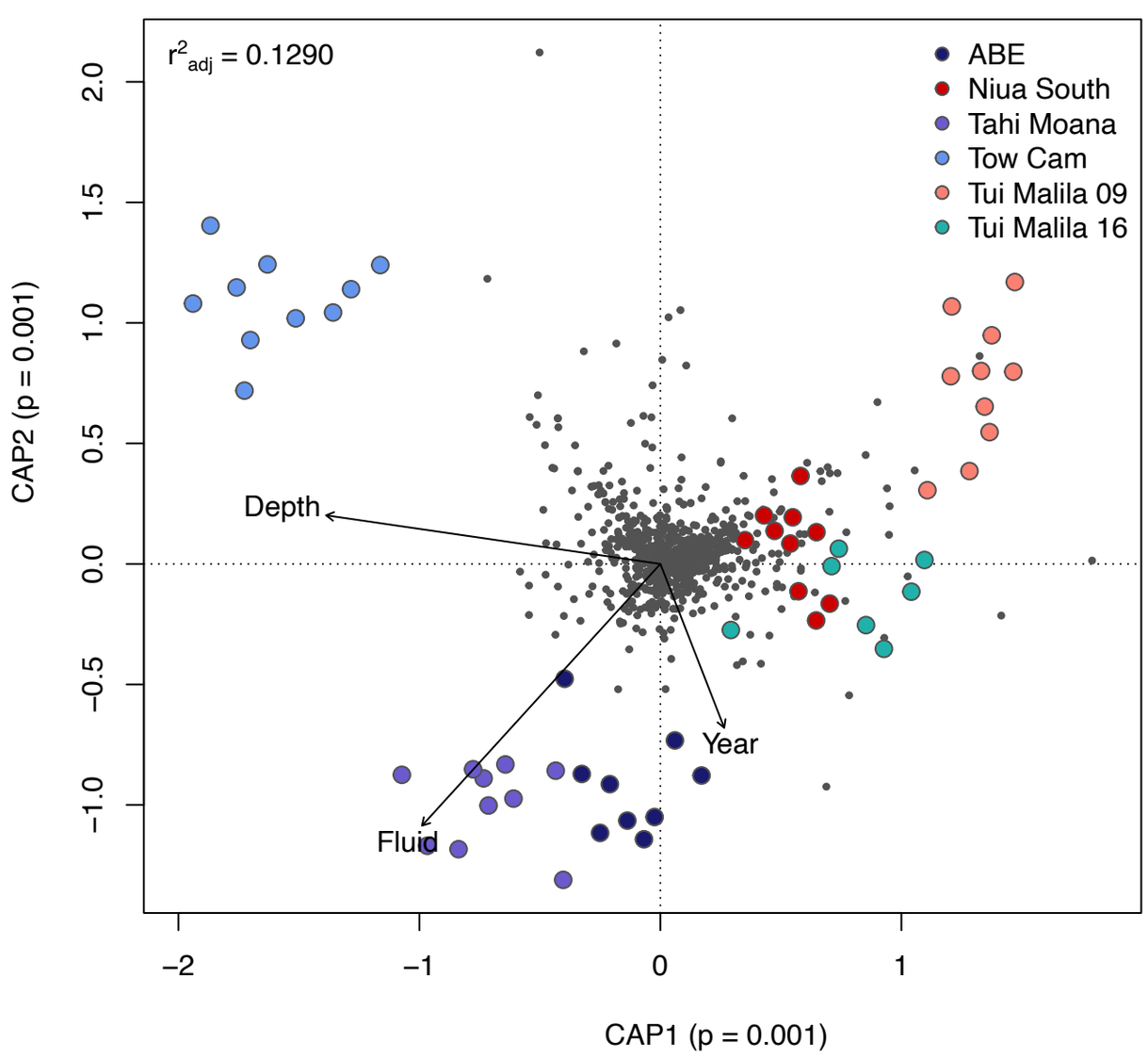
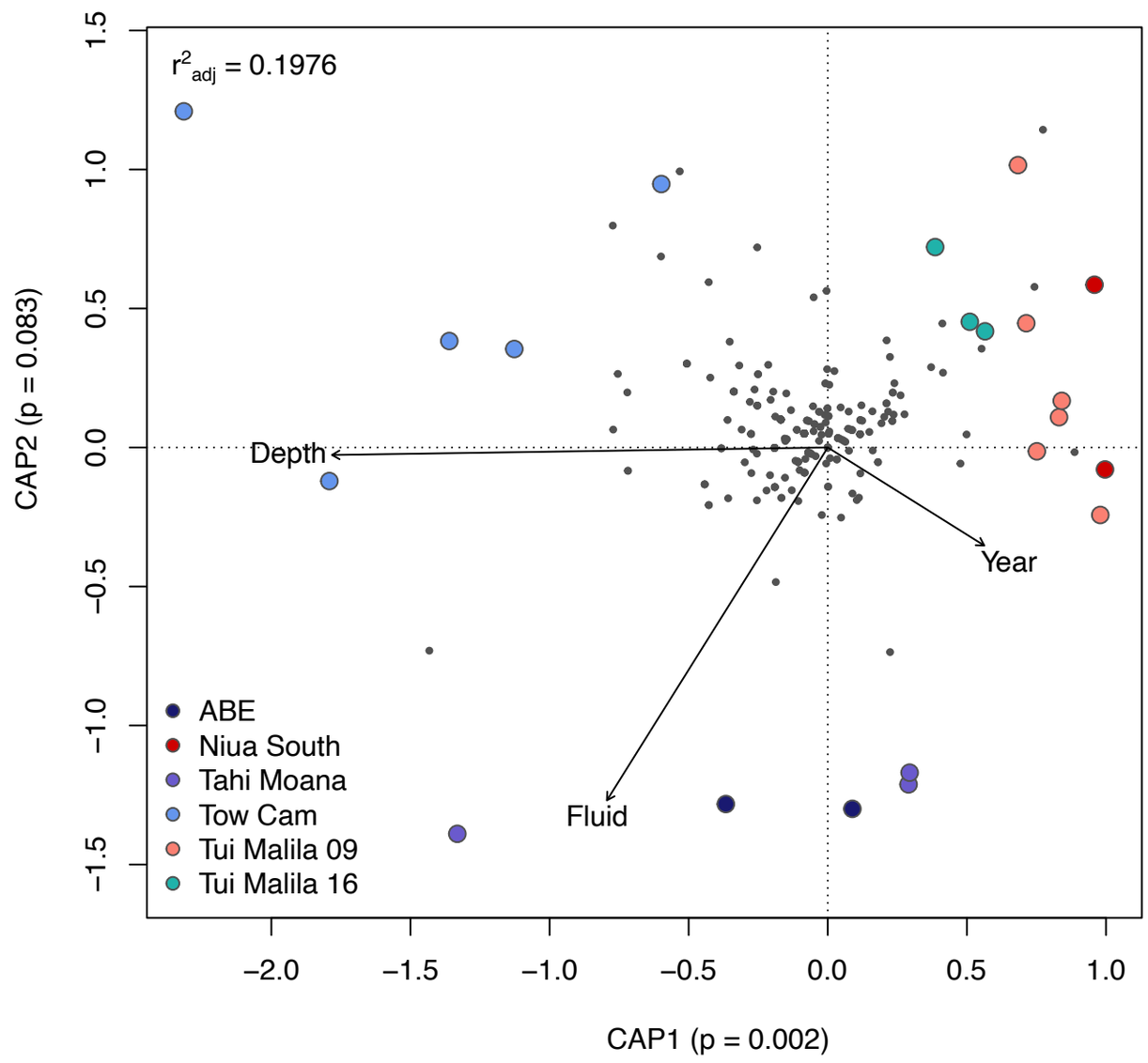
C Gamma1 (*A. strummeri*)



D GammaLau (*A. strummeri*)



E Ifr-SOX (*I. nautilei*)



F Ifr-MOX (*I. nautilei*)