Host mixotrophy, habitat stability and recombination may all decrease geographic differentiation between symbiont populations in Western Pacific deep-sea mussels

Corinna Breusing[1*], Yao Xiao[2,3], Shelbi L. Russell[4], Russell B. Corbett-Detig[4], Sixuan Li[1], Jin Sun[5], Chong Chen[6], Yi Lan[2,3], Pei-Yuan Qian[2,3], Roxanne A. Beinart[1*]

[1]Graduate School of Oceanography, University of Rhode Island, Narragansett, RI 02882, USA

[2]Department of Ocean Science, Division of Life Science and Hong Kong Branch of the Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), The Hong Kong University of Science and Technology, Hong Kong, China

[3]The Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Nansha, Guangzhou, China

[4]Department of Biomolecular Engineering, University of California Santa Cruz, Santa Cruz, CA 95064, USA

[5]Institute of Evolution & Marine Biodiversity, Ocean University of China, Qingdao, China

[6]X-STAR, Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Yokosuka, Kanagawa Prefecture, Japan

*Corinna Breusing, Roxanne A. Beinart

Email: corinnabreusing@gmail.com, rbeinart@uri.edu

Keywords: Chemosynthetic symbiosis, hydrothermal vents, symbiont strain composition, mixotrophy, homologous recombination, local adaptation, *Bathymodiolus septemdierum*

Competing interests: The authors declare no competing interests.

Abstract

The intra-host composition of horizontally transmitted microbial symbionts can vary across host populations due to interactive effects of host genetic, environmental and geographic factors. While adaptation to local habitat conditions can drive geographic subdivision of symbiont strains in autotrophic symbioses, little is known about the processes shaping symbiont population structure in mixotrophic host-microbe associations where hosts are less dependent on symbiont function and fitness. To address this question, we sequenced metagenomes of different populations of the deep-sea mussel *Bathymodiolus septemdierum*, which belongs to the foundation fauna at deep-sea hydrothermal vents in the Western Pacific Ocean. *Bathymodiolus septemdierum* lives in close symbiotic association with sulfur-oxidizing chemosynthetic bacteria but supplements its symbiotrophic diet through filter-feeding, enabling it to occupy ecological niches with less exposure to geochemical reductants. Our analyses indicate that symbiont populations associated with *B. septemdierum* are only weakly structured by geochemical habitat or geography and show elevated rates of homologous recombination, in contrast to co-occurring gastropod symbioses that exhibit greater symbiont nutritional dependence and occupy habitats with higher spatial variability in environmental conditions. These results suggest that relative habitat stability combined with sufficient symbiont dispersal and genomic mixing might promote persistence of similar symbiont strains across geographic locations, while mixotrophy might decrease selective pressures on the host to affiliate with locally adapted symbiont strains. Overall, these data contribute to our understanding of the potential mechanisms influencing symbiont population structure across the spectrum of obligate to facultative marine microbial symbioses.

Introduction

In the ocean, mutualistic symbioses between animals and bacteria are predominantly established through symbiont acquisition from the environment [1]. While some hosts develop complete dependence on their symbionts, for example through loss of essential metabolic or physiological functions, others can survive without their microbial partners, although they can gain benefits from the symbiosis under at least some conditions [2]. The composition of environmentally acquired symbionts within and across host populations is typically influenced by a combination of host genetic makeup, environmental selection, physical barriers to symbiont dispersal, symbiont competition and homologous recombination (i.e., the exchange of homologous genetic material between bacterial strains or species) [3–5]. For instance, while host specificity, ecological partitioning or dispersal limitations may increase genetic divergence among symbiont populations, frequent recombination may promote genetic cohesiveness within a symbiont species.

In marine horizontally transmitted symbioses, host specificity [6, 7] and adaptation to local habitat conditions [8, 9] can significantly shape symbiont population structure across geographic locations. However, how these patterns are influenced by the relative dependence of hosts on their bacterial symbionts is poorly understood. For example, host organisms whose fitness is less impacted by their symbiont may not require a locally adapted strain to persist in a particular habitat. On the other hand, the lower dependence on a certain symbiont type might allow hosts to be choosier in their affiliations [10].

Chemosynthetic animal-bacteria symbioses at deep-sea hydrothermal vents provide an opportunity to gain insights into the effects of host dependence on symbiont composition, given that they range from nutritionally fully obligate associations, as in siboglinid tubeworms [11], to

less dependent partnerships, as in bathymodioline mussels [12]. In contrast to most symbiotic vent invertebrates, species within the sub-family Bathymodiolinae retain a functional albeit reduced digestive tract and can complement their symbiont-derived nutrition with particulate and dissolved organic matter from their surroundings [12, 13], thereby being able to survive without their symbionts for months to possibly years, at least in a laboratory setting [14]. Nevertheless, a minimal exposure to venting fluids for chemosynthetic primary production appears to be necessary for the symbiosis to remain healthy [15].

Owing to their ecological and evolutionary success, bathymodioline mussels belong to the dominant fauna at deep-sea hydrothermal vents worldwide [16], with the genus *Bathymodiolus* representing the most prominent member of the group. In the majority of cases, *Bathymodiolus* species host thiotrophic gammaproteobacterial symbionts that use reduced sulfur compounds as energy sources for chemosynthesis [17]. *Bathymodiolus septemdierum* is widespread in the Indo-Pacific Ocean and although various morphotypes have been described, including *B. septemdierum, B. brevior,* and *B. marisindicus*, genetic data have revealed that they represent populations of the same species [18, 19]. In the Western Pacific, this mussel forms expansive beds that are typically arranged in concentric rings around patches of sympatric provannid gastropods of the genera *Alviniconcha* and *Ifremeria* [20]. Fine-scale habitat partitioning among these holobionts is likely driven by contrasting thermal tolerances and sulfide requirements for chemosynthesis, resulting in differential exposure of hosts and symbionts to available geochemical reductants: *Alviniconcha* holobionts usually encounter the highest and most variable sulfur concentrations, while *Bathymodiolus* holobionts experience the lowest and most stable levels of sulfur compounds [20, 21]. The ability of *Bathymodiolus* symbioses to thrive under diminished sulfide availability is hypothesized to be partly related to the hosts'

mixotrophic feeding mode [15, 20], which would decrease dependence on optimal chemosynthetic production rates. Although *Alviniconcha* and *Ifremeria* snails still retain a reduced and functional gut [22–24], there is no evidence that they can survive for prolonged periods of time without their endosymbionts.

In this study we sequenced metagenomes of *B. septemdierum* populations from three different vent sites in the Lau Basin to assess the population genetic structure and strain-level functional variation of its thiotrophic symbiont, *Candidatus* Thiodubiliella endoseptemdiera [5], in relation to host genotype, environmental factors and geographic location. To infer if the relative intensity of host dependence and homologous recombination might play a role in symbiont geographic partitioning, we subsequently compared the results to patterns in co-occurring *Alviniconcha* and *Ifremeria* symbioses [8] where hosts are more reliant on their symbionts for nutrition.

Materials and Methods

Sample collection, DNA/RNA extraction and host species identification

Mussel samples were collected from the Tu'i Malila (n = 27), ABE (n = 13) and Tahi Moana (n = 13) vent sites with the remotely operated vehicle (ROV) *Ropos* during R/V *Falkor* cruise FK160407 to the Eastern Lau Spreading Center (ELSC), Southwest Pacific, in April/May 2016 (Fig. 1A; Table S1). Upon ROV recovery, symbiont-bearing gill tissue of each sample was dissected on board ship and stored in RNALater™ (Thermo Fisher Scientific, Inc., Waltham, MA, USA) at –80°C until further analysis. DNA was extracted with the Qiagen DNeasy Blood & Tissue kit (Qiagen, Hilden, Germany), while RNA was extracted with the Zymo Direct-zol RNA Miniprep kit (Zymo Research, Inc., Irvine, CA, USA). Host species identities were determined

through unidirectional Sanger sequencing of the mitochondrial *COI* gene using *Bathymodiolus*-specific primers [25].

Host transcriptome assembly and population genomic analyses

To determine host population genomic structure, we *de novo* assembled the gill transcriptome of *B. septemdierum* and used the resulting transcripts as reference for variant identification. PolyA-enriched, 150-bp paired-end RNA-seq libraries were prepared with the TruSeq Stranded mRNA kit (Illumina, Inc., San Diego, CA, USA) for one individual from each vent site and sequenced on a NovaSeq 6000 instrument to a depth of 100 M reads per library by Psomagen, Inc. (Rockville, MD, USA). After initial quality checks with FASTQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), raw reads were trimmed with TRIMMOMATIC [26], error-corrected with RCORRECTOR [27], filtered with TRANSCRIPTOMEASSEMBLYTOOLS (https://github.com/harvardinformatics/TranscriptomeAssemblyTools), and assembled with TRINITY in PASAFLY mode [28]. Transcript annotation and further curation, including removal of transcript redundancies, followed the procedure in Breusing et al. [9]. Single nucleotide polymorphisms (SNPs) among individuals with respect to the reference transcriptome were subsequently estimated with ANGSD [29] based on a model correcting for deviations from Hardy-Weinberg equilibrium (see ref. [9] for details). To evaluate host population genomic differentiation between vent sites we determined pairwise fixation indices (F_{ST}s) using the method described in Bhatia et al. [30] and performed principal component analyses of host covariance matrices with the STATS and GGPLOT2 packages in R [31, 32].

Symbiont pangenome reconstruction and population genomic analyses

To assess symbiont population genomic structure, we reconstructed a pangenome for *Ca.* Thiodubiliella endoseptemdiera out of individual metagenome-assembled genomes (MAGs) that we used as reference for read mapping and variant calling. Metagenomic 2 x 150 bp libraries for 25 samples were prepared and sequenced at Novogene, Co. (Beijing, China) on a NovaSeq 6000 platform. In addition, *B. septemdierum* metagenomic reads and symbiont assemblies from Russell et al. [5] were downloaded from the National Center for Biotechnology Information. Raw reads for each individual were adaptor-clipped and quality-trimmed with TRIMMOMATIC and then assembled with METASPADES [33] using kmers from 21 to 121. Binning of MAGs was done following the METAWRAP [34] pipeline and procedures outlined in Breusing et al. [9]. MAGs were taxonomically assigned with GTDB-TK [35] and evaluated for species-level genomic similarity with FASTANI [36]. Only *Ca.* T. endoseptemdiera MAGs with CHECKM [37] completeness scores > 85% and contamination scores < 10% were considered further for pangenome reconstruction with PANAROO [38]. Genetic variants between symbiont populations were identified by comparing metagenomic reads against the *Ca.* Thiodubiliella pangenome with FREEBAYES [39] as in Breusing et al. [9]. The identified variants were subsequently used to infer symbiont population structure through calculation of pairwise F_{ST}s with SCIKIT-ALLEL (https://github.com/cggh/scikit-allel) and principal coordinate analyses with the APE package in R [40]. F_{ST} values between symbiont populations of individual mussel hosts were calculated after Romero Picazo et al. [41]. Associations of symbiont haplotypes with host genotypes, geography and environment were assessed through redundancy analyses and Mantel tests with the VEGAN and NCF packages [42, 43]. Pairwise gene content differences between symbiont populations

from different vent sites were evaluated with PANPHLAN [44] by determining genes that were present/absent in ~75% of samples from a given location.

Symbiont strain decomposition and analysis

To quantify the abundance and number of distinct symbiont strains within each mussel host, we inferred strain haplotypes on co-assembly graphs with STRONG [45]. All samples were normalized to the same number of reads before analysis (~5.4 M reads/sample). STRONG was run by evaluating multiple kmers for assembly (33, 55, 77) and using METABAT2 [46] for binning. MAGs with qualities > 0.85 were considered for strain analysis. The BAYESPATHS algorithm for strain resolution was run in five iterations, using five rounds of gene filtering and allowing up to 15 strains to be detected. To determine the influence of read and sample coverage on the amount of inferred strains, we repeated the STRONG analyses three times with different subsets of reads and samples. Strain abundance heatmaps, phylogenetic trees and beta diversity estimates based on weighted UniFrac distances were produced with the COMPLEXHEATMAP [47] GGTREE [48] and GUNIFRAC [49] packages in R.

For comparative purposes, we ran DESMAN [50] on the metagenomic data (parameters: --nb_haplotypes 12 --nb_repeat 5 --min_cov 1) and determined amplicon sequence variants (ASVs) of the *Ca.* T. endoseptemdiera 16S rRNA V4-V5 hypervariable region. 250-bp paired-end amplicon libraries were prepared with degenerate primers after Walters et al. [51], sequenced on an Illumina MiSeq platform at the Argonne National Laboratory (Lemont, IL, USA) and assessed for sequence variation with the USEARCH-UNOISE [52] and OLIGOTYPING [53] pipelines as in Breusing et al. [54]. Fractional abundance plots of the final *Ca.* T. endoseptemdiera ASVs were produced in PHYLOSEQ [55].

To confirm that strain detection across samples was not markedly influenced by potential sample cross-contamination during processing or sequencing, we evaluated the mitochondrial heterogeneity in each sample. Due to bottleneck effects during vertical transmission, intra-individual mitochondrial diversity should be naturally low, though elevated levels are possible in bivalves in the case doubly uniparental inheritance where two divergent lineages of maternal (F) and paternal (M) mitogenomes are present [56]. Individual host mitochondrial genomes were assembled with MITOBIM [57] using a previously assembled *B. septemdierum* (AP014562) mitogenome as bait for mitochondrial reconstruction. The mitogenome of sample BS_463 could not be resolved through this method and was reassembled using the sample's *COI* sequence as seed. Low frequency variants were then called with LOFREQ [58] using default filters for coverage and strand bias, a minimum mapping quality of 30 and a minimum base quality of 20.

Homologous recombination among symbiont genomes

To assess how homologous recombination might affect patterns of population genetic structure in *Ca.* T. endoseptemdiera in comparison to co-occurring symbionts of *Alviniconcha* and *Ifremeria* snails [9], we used the maximum likelihood approach implemented in CLONALFRAMEML [59] following Osvatic et al. [60]. MAGs of each symbiont species were first aligned in PROGRESSIVEMAUVE [61] with default parameters. Core locally colinear blocks (LCBs) ≧ 500 bp were then extracted with *stripSubsetLCBs* [61], concatenated, and realigned with MAFFT [62] in auto mode. The core LCB alignment was trimmed with TRIMAL [63] (parameters: -resoverlap 0.75 -seqoverlap 0.80) before phylogenetic reconstruction in RAxML [64] (parameters: -f a -m GTRGAMMA -p 12345 -x 12345 -# 100). The best RAxML tree and the trimmed core LCB alignment were subsequently used in CLONALFRAMEML to infer

homologous recombination events among symbiont MAGs. All CLONALFRAMEML analyses were run with 100 pseudo-bootstrap replicates and applying the transition/transversion rate ratios from RAxML as estimates for the kappa parameter. To evaluate the algorithm's performance, we calculated the compound parameter Rδ [59]. Rδ values greater than 1 typically imply that the relative rate of recombination to mutation R/θ is underestimated due to repeated recombination events at any given genomic position that are not accounted for by the program [59]. R values for the calculation of Rδ were inferred by multiplying R/θ with an estimate of Watterson's θ according to the following formula: $\theta = L \times G/a_n$, where L is the sum of branch lengths in the recombination-corrected phylogeny, G is the alignment length, and a_n is the (n-1)th harmonic number with n being the number of MAGs.

Results

Host populations are undifferentiated across geographic localities

Co-assembly of representative mussel gill samples from Tu'i Malila, ABE and Tahi Moana resulted in 35,060 transcripts that provided a reference for SNP estimation among host populations (Table S2). The curated *B. septemdierum* gill transcriptome assembly comprised an overall size of 57.71 Mb and was 88.90% complete. Assessment of variant sites based on mappings of 117,310,394 average metagenomic reads per sample against the reference transcriptome produced 75,146 SNPs for analysis of host population genomic subdivision. Both principal component analyses based on host covariance matrices and pairwise F_{ST}s between vent localities (<0.0022) indicated an absence of geographic structuring in *B. septemdierum* and a more individual-based partitioning of genomic diversity (Fig. 1B; Table 1). Although Mantel tests between host identity-by-state and geographic distance matrices were significant, r values

were close to 0 (0.0766), suggesting correlations that are likely not very relevant biologically (Table 2).

Symbiont populations show low genetic structure across vent locations

Thirty species-level (ANI > 95%), moderate- to high-quality MAGs were combined to reconstruct a 3.54 Mb pangenome for *Ca.* T. endoseptemdiera, which consisted of 1042 core and 4776 accessory genes (Table S3, S4). Read mapping against the pangenome reference yielded 2,952 variant sites (SNPs and indels) for symbiont population genomic analyses. Principal coordinate analyses and F_{ST} calculations based on these variant sites indicated weak partitioning of symbiont populations among vent localities (Fig. 1C; Table 1: F_{ST} = 0.0235–0.1292). Maximum differentiation was observed between Tahi Moana and Tu'i Malila without evidence for isolation-by-distance. These patterns are in contrast to those in co-occurring *Alviniconcha* (F_{ST} = 0.2892–0.5589) and *Ifremeria* (F_{ST} = 0.0359–0.2242) symbioses for the same pairwise comparisons, as symbiont populations in these associations are notably differentiated by geography except in cases of physical proximity among sites (e.g., ABE and Tahi Moana) [9]. Inter-sample differentiation between symbiont populations based on pairwise F_{ST}s and beta diversities was higher than differentiation among vent sites (Fig. 1D; Table S5: F_{ST} = 0.1857–0.5745, β = 0.0195–0.7021), suggesting that symbiont genetic diversity was predominantly partitioned among individual hosts. Despite overall low population genetic divergence on a geographic level, the association of symbiont nucleotide and strain composition with vent locality was significant and higher than that of the host based on traditional Mantel tests (Table 2; $p < 0.0013$; r = 0.1925–0.3444). By contrast, the dominant symbiont haplotype appeared to be

unrelated to geography. Likewise, there was no correlation between symbiont and host genetic distance (Table 2).

Symbiont strains are weakly associated with habitat geochemistry or geography

To get more detailed insights into strain composition within and among individual mussel hosts, we decomposed strain haplotypes from the metagenomic dataset with BAYESPATHS. Ten different strains were resolved across samples via this method and the settings chosen, with all strains being detectable in individual hosts – though with drastically varying coverage (Fig. 2). Although we cannot fully exclude that the presence of some lowly abundant strains in individual samples might be due to sample cross-contamination during processing or sequencing, analyses of mitochondrial heterogeneity suggested that intra-sample mitochondrial diversity was in general low (0.06–1.32 variants/Kbp) and relatively characteristic of vertically transmitted symbionts and organelles [65, 66]. Four mussel individuals from ABE exhibited slightly elevated mitochondrial variant densities (2.54–3.54 variants/Kbp). The diversity in these samples might largely reflect natural variability, especially as none of the samples were outliers in ordination or other statistical analyses. For example, these observations might be explained by intra-host mutations of the mitochondrial populations, possibly due to increased oxidative stress in gill tissue, or – in the case of doubly uniparental inheritance – potential presence of lowly abundant M genomes that leaked into somatic tissue during transmission [56]. Overall, these results indicate that potential sample cross-contamination is likely not an important factor influencing the observed patterns (Table S6). Comparative strain quantifications based on DESMAN and, in particular, 16S rRNA oligotyping were lower than those by BAYESPATHS, recovering only seven and three strains, respectively (Fig. S1). Given that the detection of symbiont strains tended to

increase with both read depth and number of samples (Fig. S2), the amount of recovered strains presented here likely provides a snapshot of the true strain diversity within *B. septemdierum*. In addition, due to the high amount of recombination within *Ca.* Thiodubiliella (see below), the detected strains should be regarded as a continuum of possible haplotypes rather than discrete ones. Of the ten strains identified by BAYESPATHS, six showed a slight affinity to vent site, though the overall variation in strain abundances that could be explained by differences in fluid geochemistry and correlated factors was low (7.75%; Fig. 2A, B). Based on redundancy analyses and assessments of strain abundances within host individuals, strains 4, 5, 7 and 9 were often relatively frequent in host individuals from Tahi Moana, whereas strains 3 and 6 were typically dominant in host individuals from Tu'i Malila. The association of different strains with geographic location seemed to be independent of their phylogenetic relationships as more closely related strains were not necessarily abundant in the same habitats (Fig. 2B, C). Given the weak association of symbiont strains with vent site, no evidence for local adaptation of the dominant symbiont haplotypes could be found, as correlations between consensus variants and environment were insignificant. Redundancy analyses based on nucleotide counts, however, revealed 172 genetic variants within the overall symbiont populations that might be involved in adaptation to local habitat conditions (Fig. S3; Table S7). Most of these candidate variants were present in genes with uncharacterized or poorly constrained functions or genes involved in anti-viral defense, DNA metabolism and cell signaling (Table S7).

Symbiont populations from different vent sites exhibit some variation in gene content

We profiled the metagenomic samples using pangenome-based phylogenomic analyses to determine potential gene content variation among symbiont populations from distinct vent sites.

At a presence/absence threshold of ~75%, the main differences among symbiont populations were observed for genes with unknown function, followed by genes involved in DNA replication and transposition. A small subset of differentially conserved genes was associated with other metabolic categories (Fig. 3; Table S8). For example, *Ca.* T. endoseptemdiera strains from Tu'i Malila typically lacked a particular hydrogenase maturation factor (HybG), whereas symbiont strains from Tahi Moana missed several genes coding for ATPases and enzymes involved in aerobic respiration (Table S8). Symbiont populations from ABE were similar in gene content to those from Tu'i Malila, but more frequently contained genes for a CRISPR-associated endonuclease (Cas9), an ATP-dependent RNA helicase (DeaD), a filamentous hemagglutinin, a SAM-dependent methyltransferase as well as a RelA/SpoT domain protein (Table S8). No fixed or nearly fixed differences among symbiont populations were observed, further supporting the similarity of *Ca.* T. endoseptemdiera strains between geographic locations.

Homologous recombination contributes to cohesiveness of Ca. *T. endoseptemdiera populations*

Intra-specific homologous recombination plays a critical role in maintaining genetic connectivity of bacterial populations [67]. To evaluate if homologous recombination might have a homogenizing effect on *Ca.* T. endoseptemdiera population structure compared to co-occurring *Alviniconcha* and *Ifremeria* symbiont species, we calculated the relative rates (R/θ) and importance (r/m) of recombination to mutation events within the core genome of each symbiont group. R/θ ratios in core LCB alignments (726,156–1,152,639 bp) were at least six times higher in *Ca.* T. endoseptemdiera (0.62–0.65) than in thiotrophic symbionts associated with *Alviniconcha* and *Ifremeria*, while r/m ratios were at least four times larger (1.94–2.09), suggesting an increased relevance of recombination for the evolution of *Ca.* T. endoseptemdiera

(Table 3; Fig 4). Although CLONALFRAMEML is likely to underestimate R/θ within the inferred parameter ranges given that all Rδ values were markedly larger than 1 (Table 3) [59], the results obtained are correlated with true values and we therefore expect that the relative R/θ ratios among symbiont populations produce a useful comparison.

Discussion

Symbiont strain composition in horizontally transmitted symbioses is determined by a variety of interacting factors, including host genetics, environment and geography [3]. At deep-sea hydrothermal vent sites, local adaptation to geographically variable fluid geochemistry plays a significant role in shaping strain-level variation within two provannid snail genera that are obligately dependent on their symbionts for nutrition [9]. How the composition of symbiont strains is affected in host species with less symbiont dependence and/or lower fluctuations in habitat conditions is currently poorly understood. To address this question, we assessed symbiont strain composition in mixotrophic populations of the deep-sea mussel *Bathymodiolus septemdierum*, which inhabits environmentally relatively stable diffuse flow areas around hydrothermal vents in the Western Pacific Ocean [20].

In agreement with patterns observed in co-occurring *Alviniconcha* and *Ifremeria* symbioses, symbiont strain composition in *B. septemdierum* was unrelated to host genetics but appeared to vary to some degree with vent habitat and geography. Compared to the provannid snail symbionts, however, mussel symbiont populations were strikingly less genetically differentiated, with notable overlaps in the geographic distributions of different *Ca.* Thiodubiliella endoseptemdiera strains and absence of habitat affiliation for the dominant symbiont haplotypes. Leydet and Hellberg [10] suggested that symbiont genotypes in facultative

host-microbe associations may not be strongly linked to host genetic traits or environmental conditions since host fitness is likely less dependent on partnerships with locally adapted symbiont strains. Additionally, the lower thermal and geochemical variation in the microhabitat patches that *B. septemdierum* symbioses occupy [20, 21] might favor similar *Ca.* T. endoseptemdiera strains across the symbiont's distributional range, thus allowing host individuals to affiliate with the same symbiont genotypes at different vent locations.

Despite the overall weak correlation of *Ca.* T. endoseptemdiera strains with habitat, symbiont populations exhibited some differences in gene content and strain abundances, indicating that environmental conditions might at least partially shape the population genetic structure of *Ca.* T. endoseptemdiera. For example, the strains from Tahi Moana lacked specific genes involved in aerobic respiration, whereas the strains from Tu'i Malila frequently missed a hydrogenase maturation factor. These discrepancies could indicate certain adaptations to low oxygen and hydrogen concentrations at the respective vent sites or resource partitioning with the host and co-existing symbiont strains. We found that up to ten distinct strains co-occurred within a single mussel individual, although the actual number of strains is likely higher, given that strain detection depended notably on read coverage, number of samples, sequencing method and bioinformatic algorithm assumptions. In general, however, these estimates are within the range of variation reported for thiotrophic symbionts of other *Bathymodiolus* species, including *B. brooksi* from cold seeps in the Gulf of Mexico (up to 9 strains) [41] and *B. puteoserpentis* from hydrothermal vents on the Mid-Atlantic Ridge (up to 16 strains) [68]. Symbiont diversity in bathymodioline mussels has previously been shown to be maintained by functional heterogeneity among symbiont subpopulations and is thought to confer metabolic flexibility of the symbiosis to changing environmental conditions [68, 69].

Our data indicate that ecological factors other than fluid geochemistry might be relevant for structuring *Ca.* T. endoseptemdiera populations. Functionally, candidate genetic variants under natural selection were not clearly linked to differences in geochemical conditions but appeared to be involved in DNA metabolic processes, cell regulation and immunity against bacteriophages. This implies that the composition of viral communities might be an important and possibly common mediator of symbiont genetic differentiation at hydrothermal vents, as similar impacts have been reported for symbiont populations associated with siboglinid tubeworms [70, 71] and provannid snails [9].

Altogether, the relatively weak differentiation of mussel symbiont populations by vent habitat is in agreement with studies in other *Bathymodiolus* species, where environmental factors appear to have limited influence on symbiont population structure [72, 73]. By contrast, symbionts in other marine partnerships that are characterized by low host dependence have been found to be ecologically subdivided. For example, facultative zooxanthellae symbionts of *Oculina* corals are genetically partitioned according to sea surface temperature, at least in the Mediterranean Sea [10]. These discrepancies indicate that associations with locally adapted symbiont strains in facultative horizontally transmitted symbioses are not universal, and that the strength of these partnerships might depend on the type of selective pressures and corresponding environmental filtering experienced by free-living symbiont populations across habitats.

The low genetic subdivision across geographic locations is in stark contrast to the degree of symbiont population divergence between individual mussels, which was on average 4.7 times higher. The detected levels of inter-host symbiont population differentiation agree with the patterns observed in other sulfur-oxidizing *Bathymodiolus* symbionts and were similarly unrelated to host genotype [41, 73], suggesting that variation in symbiont composition between

host individuals is driven by neutral evolutionary processes [41]. Although *Bathymodiolus* mussels have been shown to remain competent for symbiont uptake throughout their lifetime [74], Romero Picazo et al. [41] suggested that strong genetic isolation among individual symbiont populations can occur if environmental symbiont acquisition is restricted in later host developmental stages and re-colonization happens increasingly through self-infection. This model of symbiont transmission might explain how divergence of symbiont populations is maintained on an individual but not geographic level. Founder events combined with possible priority effects [75] during symbiont infection may increase inter-host genetic differentiation of symbiont populations. By contrast, occasional reshuffling of symbiont strains between host and environment combined with adequate symbiont dispersal may promote homologous recombination and mixing of the free-living symbiont pools, thereby diminishing geographic and ecological isolation. *Candidatus* T. endoseptemdiera has previously been shown to have some of the highest effective recombination rates observed for bacteria [5]. Our data indicate that, in contrast to sympatric symbionts of *Alviniconcha* and *Ifremeria* snails, the relative rates and effects of recombination to mutation in *Ca.* T. endoseptemdiera surpass an estimated threshold beyond which recombination is sufficient to prevent population differentiation [76]. Similar observations have been made in chemosynthetic lucinid clam symbioses, where increased levels of homologous recombination are hypothesized to maintain genetic cohesiveness within a cosmopolitan symbiont species [60].

At high enough recombination rates, even limited dispersal may promote homogenization among distant symbiont populations. Recent biophysical models for the Lau Basin have implied that dispersal barriers within this region are mostly absent [77], while evidence from a variety of biogeographic studies suggests that vent-associated microbial communities are not significantly

shaped by dispersal limitations [78]. However, the dispersal mechanisms of bacterial symbionts are not well understood, making it difficult to assess how well dispersal patterns are represented by model simulations. In addition, several population genetic studies on the symbionts of vent-associated animals, including those of *Bathymodiolus* species from mid-ocean ridge systems and other mollusk taxa from the Lau Basin, have highlighted the effect of topological barriers and geographic distance on symbiont biogeography [9, 72, 73]. Despite contrasting evidence from these systems, it is possible that *Ca.* T. endoseptemdiera faces less dispersal restrictions than co-occurring symbionts by employing different dispersal strategies. For example, the bacterial symbionts of lucinid clams have been hypothesized to be transported over long distances with seagrass fragments [60], whereas some algal symbionts of gorgonian octocorals are known to hitchhike with the corals' larvae [79] – a strategy that has also been suggested for the symbionts of bathymodioline mussels [80]. Considering the broad distribution of *B. septemdierum* in the Western Pacific and Indian Ocean [18], vectored transport, if it can be confirmed, would greatly increase the dispersal ability of *Ca.* T. endoseptemdiera, thereby reducing geographic partitioning among disparate populations. It is currently unknown how widely distributed *Ca.* T. endoseptemdiera is, as sequence information for bacterial symbionts of other *B. septemdierum* populations is scarce. Based on comparisons of publicly available full-length 16S rRNA sequences, *Ca.* T. endoseptemdiera from the Lau Basin is >97% identical to other *B. septemdierum* symbionts from the Izu-Ogasawara Arc vents in the Northwest Pacific. In addition, our sequence searches within the Integrated Microbial NGS platform (https://www.imngs.org) at an identity cutoff of 100% suggest that *Ca.* T. endoseptemdiera might occur at various locations in the Northeast Pacific Ocean and around the Solomon Islands. Although further genomic analyses will be needed to confirm the species-level designations of

the recovered sequences, these results support a possible widespread distribution of *Ca.* T. endoseptemdiera.

Conclusions

Our data provide evidence that symbiont populations associated with *Bathymodiolus septemdierum* are weakly structured among vent locations in the Lau Basin, in sharp contrast to thiotrophic symbionts of co-occurring *Alviniconcha* and *Ifremeria* snails. While the lower dependence of *B. septemdierum* on symbiont-derived nutrition might contribute to these patterns, observations in facultative symbioses from other marine systems suggest that environmental selection does influence symbiont population structure in these associations. In many cases, the effects of host dependence and ecological partitioning are likely interlocked. For example, related to its mixotrophic feeding mode, *B. septemdierum* occupies niches with less variable fluid flows, which might promote affiliations with the same symbiont strains across geographically distinct, but ecologically similar habitats. Our findings further suggest that symbiont-related factors, such as increased recombination rates combined with the possibility of long-distance dispersal, might decrease population differentiation among vent locations. However, the relative importance of these processes for symbiont geographic structure and adaptation remains unclear. Future investigations of environmental versus intra-host strain abundances at different vent sites as well as extracellular presence of symbionts on host larvae will help to shed light on this question.

Acknowledgements

We thank the captain and crew of the R/V *Falkor* as well as the pilots of the ROV *Ropos* for their support of the sample collections during Schmidt Ocean Institute cruise FK160407. The technical staff at the Argonne National Laboratory, Novogene, Co. and Psomagen, Inc. are gratefully acknowledged for preparing and sequencing the Illumina 16S rRNA amplicon, metagenomic and RNA-seq libraries, respectively. We further thank the Kingdom of Tonga for providing access to their national waters and Abi Goodman for her help in verifying the host species identities of mussel samples used in this study. This work was funded by the Key Special Project for Introduced Talents Team of Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou) (GML2019ZD0409), the General Research Fund of Hong Kong SAR (16101219, 16101822), the Schmidt Ocean Institute, the U.S. National Science Foundation (grant numbers OCE-1536331, 1819530 and 1736932 to R.A.B. and EPSCoR Cooperative Agreement OIA-#1655221) and the National Institutes of Health (grant numbers 5K99GM135583-02 to S.L.R. and 5R35GM128932-03 to R.C.D.).

Competing Interests

The authors declare no competing interests.

Data Availability Statement

Raw metagenomic, 16S rRNA amplicon and RNA-seq reads are available in the National Center for Biotechnology Information under BioProject numbers PRJNA855930, PRJNA473258 and PRJNA849542. The *Bathymodiolus septemdierum* transcriptome has been deposited in the Transcriptome Shotgun Assembly database under accession number GJZH00000000. Host *COI*

sequences and *Ca.* Thiodubiliella endoseptemdiera MAGs have been submitted to GenBank under accession numbers listed in Supplementary Tables S1 and S3, respectively.

References

1. Russell SL. Transmission mode is associated with environment type and taxa across bacteria-eukaryote symbioses: a systematic review and meta-analysis. *FEMS Microbiol Lett* 2019; **366**: fnz013.
2. Fisher RM, Henry LM, Cornwallis CK, Kiers ET, West SA. The evolution of host-symbiont dependence. *Nat Commun* 2017; **8**: 15973.
3. LaJeunesse TC, Pettay DT, Sampayo EM, Phongsuwan N, Brown B, Obura DO, et al. Long-standing environmental conditions, geographic isolation and host-symbiont specificity influence the relative ecological dominance and genetic diversification of coral endosymbionts in the genus *Symbiodinium*. *J Biogeogr* 2010; **37**: 785–800.
4. McIlroy SE, Cunning R, Baker AC, Coffroth MA. Competition and succession among coral endosymbionts. *Ecol Evol* 2019; **9**: 12767–12778.
5. Russell SL, Pepper-Tunick E, Svedberg J, Byrne A, Ruelas Castillo J, Vollmers C, et al. Horizontal transmission and recombination maintain forever young bacterial symbiont genomes. *PLoS Genet* 2020; **16**: e1008935.
6. Nishiguchi MK, Ruby EG, McFall-Ngai MJ. Competitive dominance among strains of luminous bacteria provides an unusual form of evidence for parallel evolution in sepiolid squid-*Vibrio* symbioses. *Appl Environ Microbiol* 1998; **64**: 3209–3213.
7. Nishiguchi MK. Host-symbiont recognition in the environmentally transmitted sepiolid squid-*Vibrio* mutualism. *Microb Ecol* 2002; **44**: 10–18.

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

9. Breusing C, Genetti M, Russell SL, Corbett-Detig RB, Beinart RA. Horizontal transmission enables flexible associations with locally adapted symbiont strains in deep-sea hydrothermal vent symbioses. *Proc Natl Acad Sci USA* 2022; **119**: e2115608119.

10. Leydet KP, Hellberg ME. Discordant coral–symbiont structuring: factors shaping geographical variation of *Symbiodinium* communities in a facultative zooxanthellate coral genus, *Oculina*. *Coral Reefs* 2016; **35**: 583–595.

11. Nussbaumer AD, Fisher CR, Bright M. Horizontal endosymbiont transmission in hydrothermal vent tubeworms. *Nature* 2006; **441**: 345–348.

12. Page HM, Fiala-Medioni A, Fisher CR, Childress JJ. Experimental evidence for filter-feeding by the hydrothermal vent mussel, *Bathymodiolus thermophilus*. *Deep Sea Res Pt A* 1991; **38**: 1455–1461.

13. Le Pennec M, Donval A, Herry A. Nutritional strategies of the hydrothermal ecosystem bivalves. *Prog Oceanogr* 1990; **24**: 71–80.

14. Colaço A, Bettencourt R, Costa V, Lino S, Lopes H, Martins I, et al. LabHorta: a controlled aquarium system for monitoring physiological characteristics of the hydrothermal vent mussel *Bathymodiolus azoricus*. *ICES J Mar Sci* 2011; **68**: 349–356.

15. Fisher CR, Childress JJ, Arp AJ, Brooks JM, Distel D, Favuzzi JA, et al. Microhabitat variation in the hydrothermal vent mussel, *Bathymodiolus thermophilus*, at the Rose Garden vent on the Galapagos Rift. *Deep Sea Res Pt A* 1988; **35**: 1769–1791.

16. Roeselers G, Newton ILG. On the evolutionary ecology of symbioses between chemosynthetic bacteria and bivalves. *Appl Microbiol Biotechnol* 2012; **94**: 1–10.

17. Duperron S, Lorion J, Samadi S, Gros O, Gaill F. Symbioses between deep-sea mussels (Mytilidae: Bathymodiolinae) and chemosynthetic bacteria: diversity, function and evolution. *C R Biol* 2009; **332**: 298–310.

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

19. Thubaut J, Puillandre N, Faure B, Cruaud C, Samadi S. The contrasted evolutionary fates of deep-sea chemosynthetic mussels (Bivalvia, Bathymodiolinae). *Ecol Evol* 2013; **3(14)**: 4748–4766.

20. Podowski E, Ma S, Luther G, Wardrop D, Fisher C. Biotic and abiotic factors affecting distributions of megafauna in diffuse flow on andesite and basalt along the Eastern Lau Spreading Center, Tonga. *Mar Ecol Prog Ser* 2010; **418**: 25–45.

21. Podowski EL, Moore TS, Zelnio KA, Luther GW, Fisher CR. Distribution of diffuse flow megafauna in two sites on the Eastern Lau Spreading Center, Tonga. *Deep Sea Res Pt I* 2009; **56**: 2041–2056.

22. Yang Y, Sun J, Chen C, Zhou Y, Van Dover CL, Wang C, Qiu JW, Qian PY. Metagenomic and metatranscriptomic analyses reveal minor-yet-crucial roles of gut microbiome in deep-sea hydrothermal vent snail. *Anim Microbiome* 2022; **4(1)**: 3.

23. Laming SR, Hourdez S, Cambon-Bonavita MA, Pradillon F. Classical and computed tomographic anatomical analyses in a not-so-cryptic *Alviniconcha* species complex from hydrothermal vents in the SW Pacific. *Front Zool* 2020; **17**: 12.

24. Chen C, Watanabe HK, Gena K, Johnson SB. Anatomical shifts linked with unusual diets in deep-sea snails. *Ecology* 2022 (in press).

25. Olu-Le Roy K, Cosel R von, Hourdez S, Carney SL, Jollivet D. Amphi-Atlantic cold-seep *Bathymodiolus* species complexes across the equatorial belt. *Deep Sea Res Pt I* 2007; **54**: 1890–1911.

26. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014; **30**: 2114–2120.

27. Song L, Florea L. Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. *GigaSci* 2015; **4**: 48.

28. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 2011; **29**: 644–652.

29. Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* 2014; **15**: 356.

30. Bhatia G, Patterson N, Sankararaman S, Price AL. Estimating and interpreting FST: The impact of rare variants. *Genome Research* 2013; **23**: 1514–1521.

31. Wickham H. ggplot2: Elegant Graphics for Data Analysis. 2016. Springer New York.

32. R Core Team. R: A language and environment for statistical computing. 2021. R Foundation for Statistical Computing, Vienna, Austria.

33. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* 2017; **27**: 824–834.

34. Uritskiy GV, DiRuggiero J, Taylor J. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* 2018; **6**: 158.

35. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 2019; **36**: 1925–1927.

36. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 2018; **9**: 5114.

37. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2015; **25**: 1043–1055.

38. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, Lees JA, et al. Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol* 2020; **21**: 180.

39. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. *arXiv:12073907 [q-bio]* 2012.

40. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 2019; **35**: 526–528.

41. Romero Picazo D, Dagan T, Ansorge R, Petersen JM, Dubilier N, Kupczok A. Horizontally transmitted symbiont populations in deep-sea mussels are genetically isolated. *ISME J* 2019; **13**: 2954–2968.

42. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: Community Ecology Package. 2020. https://CRAN.R-project.org/package=vegan

43. Bjornstad ON. ncf: Spatial Covariance Functions. 2020. https://CRAN.R-project.org/package=ncf

44. Beghini F, McIver LJ, Blanco-Míguez A, Dubois L, Asnicar F, Maharjan S, et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *eLife* 2021; **10**: e65088.

45. Quince C, Nurk S, Raguideau S, James R, Soyer OS, Summers JK, et al. STRONG: metagenomics strain resolution on assembly graphs. *Genome Biol* 2021; **22**: 214.

46. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 2019; **7**: e7359.

47. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 2016; **32**: 2847–2849.

48. Yu G, Smith DK, Zhu H, Guan Y, Lam TT. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* 2017; **8**: 28–36.

49. Chen J, Zhang X, Yang L. GUniFrac: Generalized UniFrac distances, distance-based multivariate methods and feature-based univariate methods for microbiome data analysis. 2022. https://CRAN.R-project.org/package=GUniFrac

50. Quince C, Delmont TO, Raguideau S, Alneberg J, Darling AE, Collins G, et al. DESMAN: a new tool for de novo extraction of strains from metagenomes. *Genome Biol* 2017; **18**: 181.

51. Walters W, Hyde ER, Berg-Lyons D, Ackermann G, Humphrey G, Parada A, et al. Improved bacterial 16S rRNA gene (V4 and V4-5) and fungal internal transcribed spacer marker gene primers for microbial community surveys. *mSystems* 2016; **1**: e00009-15.

52. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 2010; **26**: 2460–2461.

53. Eren AM, Maignien L, Sul WJ, Murphy LG, Grim SL, Morrison HG, et al. Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods Ecol Evol* 2013; **4**: 1111–1119.

54. Breusing C, Castel J, Yang Y, Broquet T, Sun J, Jollivet D, et al. Global 16S rRNA diversity of provannid snail endosymbionts from Indo-Pacific deep-sea hydrothermal vents. *Environ Microbiol Rep* 2022; **14**: 299–307.

55. McMurdie PJ, Holmes S. phyloseq: an R Package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* 2013; **8**: e61217.

56. Breton S, Beaupré HD, Stewart DT, Hoeh WR, Blier PU. The unusual system of doubly uniparental inheritance of mtDNA: isn't one enough? *Trends Genet.* 2007; **23(9)**: 465–474.

57. Hahn C, Bachmann L, Chevreux B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads--a baiting and iterative mapping approach. *Nucleic Acids Res.* 2013; **41(13)**: e129.

58. Wilm A, Aw PP, Bertrand D, Yeo GH, Ong SH, Wong CH, et al LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res.* 2012; **40(22)**: 11189–11201.

59. Didelot X, Wilson DJ. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLoS Comput Biol.* 2015; **11(2)**: e1004041.

60. Osvatic JT, Wilkins LGE, Leibrecht L, Leray M, Zauner S, Polzin J, et al. Global biogeography of chemosynthetic symbionts reveals both localized and globally distributed symbiont groups. *Proc Natl Acad Sci USA* 2021; **118**: e2104378118.

61. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* 2010; **5(6)**: e11147.

62. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 2002; **30(14)**: 3059–3066.

63. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* 2009; **25(15)**: 1972–1973.

64. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014; **30(9)**: 1312–1313.

65. Russell SL, Cavanaugh CM. Intrahost genetic diversity of bacterial symbionts exhibits evidence of mixed infections and recombinant haplotypes. *Mol Biol Evol.* 2017; **34(11)**: 2747–2761.

66. Perez M, Breusing C, Angers B, Beinart RA, Won YJ, Young CR. Divergent paths in the evolutionary history of maternally transmitted clam symbionts. *Proc Biol Sci.* 2022; **289(1970)**: 20212137.

67. González-Torres P, Rodríguez-Mateos F, Antón J, Gabaldón T. Impact of homologous recombination on the evolution of prokaryotic core genomes. *mBio.* 2019; **10(1)**: e02494–18.

68. Ansorge R, Romano S, Sayavedra L, Porras MÁG, Kupczok A, Tegetmeyer HE, et al. Functional diversity enables multiple symbiont strains to coexist in deep-sea mussels. *Nat Microbiol* 2019; **4**: 2487–2497.

69. Ikuta T, Takaki Y, Nagai Y, Shimamura S, Tsuda M, Kawagucci S, et al. Heterogeneous composition of key metabolic gene clusters in a vent mussel symbiont population. *ISME J* 2016; **10**: 990–1001.

70. Perez M, Juniper SK. Insights into symbiont population structure among three vestimentiferan tubeworm host species at Eastern Pacific spreading centers. *Appl Environ Microbiol* 2016; **82**: 5197–5205.

71. Perez M, Angers B, Young CR, Juniper SK. Shining light on a deep-sea bacterial symbiont population structure with CRISPR. *Microbial Genomics* 2021; **7**: 000625.

72. Ho P-T, Park E, Hong SG, Kim E-H, Kim K, Jang S-J, et al. Geographical structure of endosymbiotic bacteria hosted by *Bathymodiolus* mussels at eastern Pacific hydrothermal vents. *BMC Evol Biol* 2017; **17**: 121.

73. Ücker M, Ansorge R, Sato Y, Sayavedra L, Breusing C, Dubilier N. Deep-sea mussels from a hybrid zone on the Mid-Atlantic Ridge host genetically indistinguishable symbionts. *ISME J* 2021; **15**: 3076–3083.

74. Wentrup C, Wendeberg A, Schimak M, Borowski C, Dubilier N. Forever competent: deep-sea bivalves are colonized by their chemosynthetic symbionts throughout their lifetime. *Environ Microbiol* 2014; **16**: 3699–3713.

75. Debray R, Herbert RA, Jaffe AL, Crits-Christoph A, Power ME, Koskella B. Priority effects in microbiome assembly. *Nat Rev Microbiol* 2022; 20: 109–121.

76. Fraser C, Hanage WP, Spratt BG. Recombination and the nature of bacterial speciation. *Science* 2007; **315**: 476–480.

77. 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. https://doi.org/10.1111/eva.13326

78. Dick GJ. The microbiomes of deep-sea hydrothermal vents: distributed globally, shaped locally. *Nat Rev Microbiol* 2019; **17**: 271–283.

79. Wirshing HH, Feldheim KA, Baker AC. Vectored dispersal of *Symbiodinium* by larvae of a Caribbean gorgonian octocoral. *Mol Ecol* 2013; **22**: 4413–4432.

80. Won Y-J, Hallam SJ, O'Mullan GD, Pan IL, Buck KR, Vrijenhoek RC. Environmental acquisition of thiotrophic endosymbionts by deep-sea mussels of the genus *Bathymodiolus*. *Appl Environ Microbiol* 2003; **69**: 6785–6792.

Tables

Table 1 F_{ST} values between host (above diagonal) and symbiont (below diagonal) populations from different vent sites.

	Tahi Moana	**ABE**	**Tu'i Malila**
Tahi Moana	*	-0.011652	0.002158
ABE	0.075300	*	-0.012374
Tu'i Malila	0.129200	0.023500	*

Table 2 Mantel tests between host, symbiont and geographic distance matrices. Significant tests are indicated in bold.

	Symbiont nucleotide counts		**Dominant symbiont haplotype**		**Symbiont strain abundance**		
	Host-Symbiont	*Symbiont-Geography*	*Host-Symbiont*	*Symbiont-Geography*	*Host-Symbiont*	*Symbiont-Geography*	*Host-Geography*
***r* statistic**	0.1084	**0.3444**	-0.0521	-0.0221	0.1018	**0.1925**	**0.0766**
***p*-value**	0.1683	**0.0002**	0.2659	0.2991	0.1387	**0.0013**	**0.0326**

Table 3 Rate and effect ratios of homologous recombination to mutation events in dominant chemosynthetic symbionts of *Bathymodiolus septemdierum*, *Alviniconcha* spp. and *Ifremeria nautilei* from the Lau Basin and Tonga Volcanic Arc. A 95% confidence interval is given for each ratio. R/θ = recombination to mutation rate ratio, r/m = relative effects of recombination to mutation, Rδ = key determinant for algorithm performance

Symbiont	**# MAGs**	**Host**	**Alignment length [bp]**	**R/θ [95% CI]**	**r/m [95% CI]**	**Rδ**
Ca. Thiodubiliella	30	*B. septemdierum*	726,156	0.6394 (0.6246–0.6543)	2.0184 (1.9415–2.0895)	88470.0722
Epsilon	8	*A. boucheti*	1,152,639	0.1026 (0.0976–0.1077)	0.2766 (0.2514–0.2977)	14109.1698
Gamma1	35	*A. kojimai*, *A. strummeri*	738,988	0.0415 (0.0338–0.0499)	0.5249 (0.4174–0.6928)	881.9823
Ifr-SOX*	58	*I. nautilei*	803,984	0.0542 (0.0446–0.0647)	0.2302 (0.1874–0.3082)	1589.2253

*Contigs for each MAG were reordered based on the Ifr-SOX reference assembly (GCA_021766955.1) to reduce the computational burden of whole genome alignments in PROGRESSIVEMAUVE

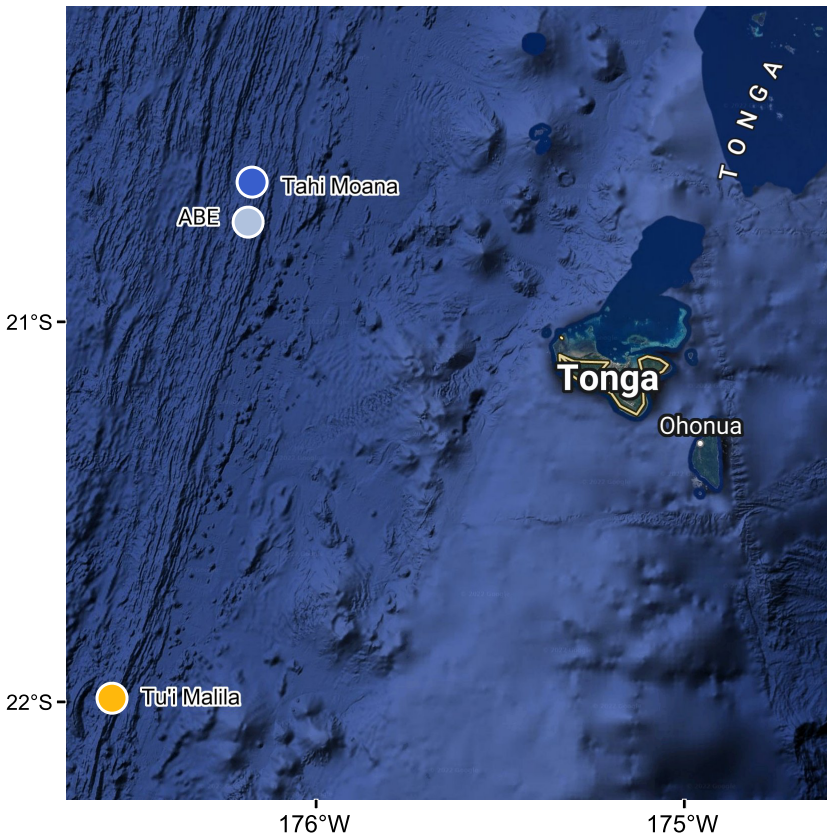
Figure Legends

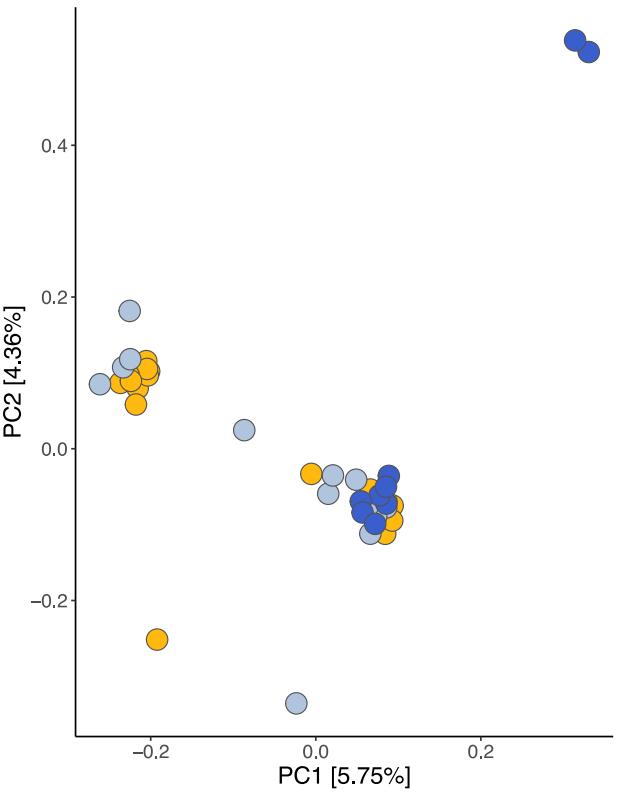
Fig. 1 Sampling locations of *Bathymodiolus septemdierum* in the Eastern Lau Spreading Center (A) and genetic structure of host (B) and symbiont (C) populations across vent sites. Vent locations are between 9 and 151 km apart. Host populations are genetically undifferentiated, while weak structuring by geography is observed in the symbiont. In contrast to patterns in co-occurring snail symbionts no evidence for isolation-by-distance is found. (D) Pairwise F_{ST} values (lower diagonal) and beta diversities (upper diagonal) among symbiont populations of single mussel hosts. In contrast to patterns observed on a geographic level, differentiation on a per-sample level is notably higher, indicating that symbiont genetic diversity is largely driven by variation among host individuals. Map data ©2022 Google.

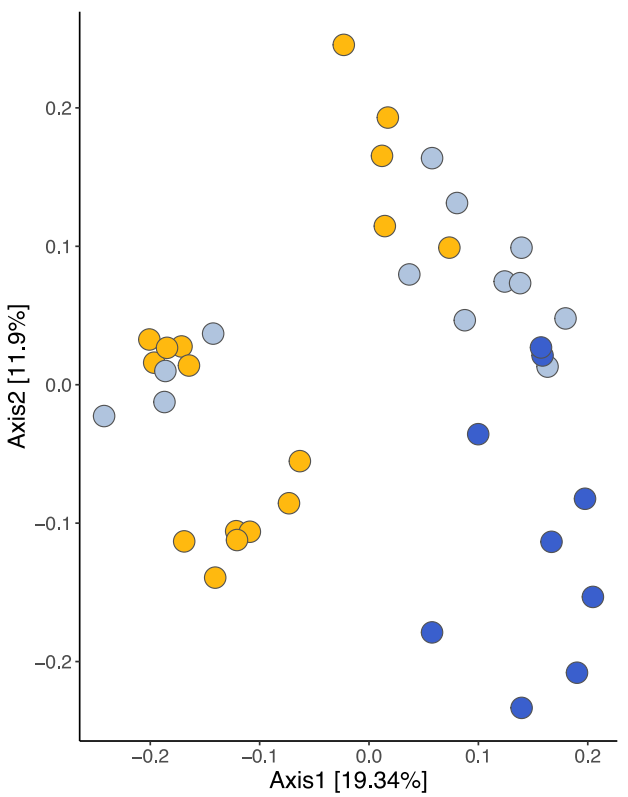
Fig. 2 Environmental associations (A) and intra-host abundances (B) of decomposed *Ca.* Thiodubiliella endoseptemdiera strains. Strain abundances are weakly associated with fluid geochemistry of the different vent sites, with strains 4, 5, 7 and 9 being relatively frequent in (some) host individuals from Tahi Moana, and strains 3 and 6 being abundant in host individuals from Tu'i Malila. The environmental associations among the distinct symbiont strains appear to be unrelated to phylogenetic proximity.

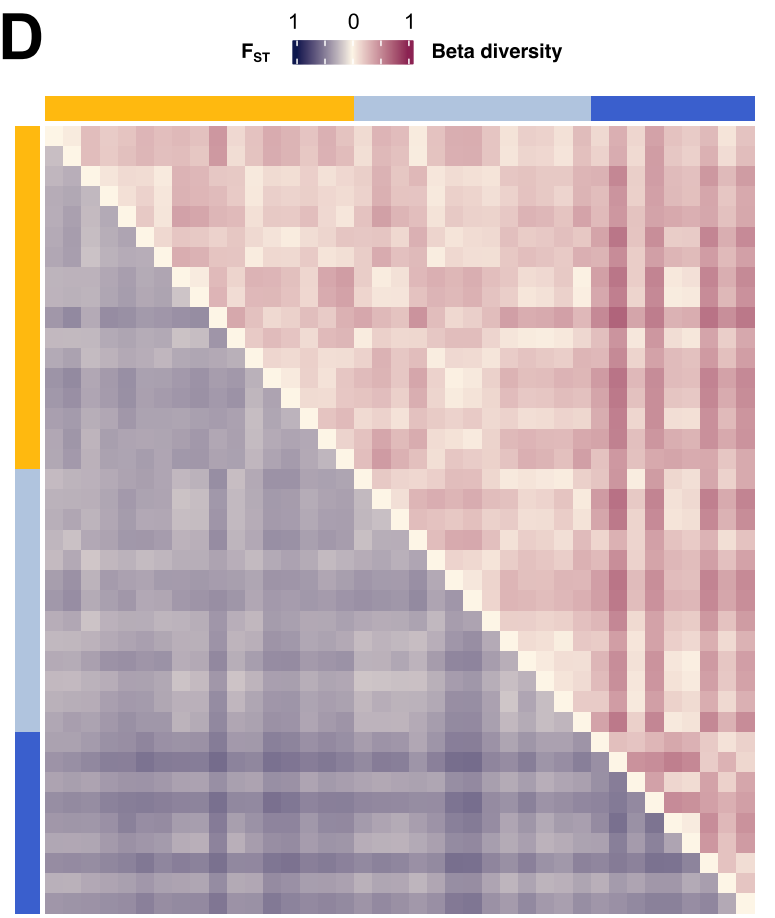
Fig. 3 Gene content differences among *Ca.* Thiodubiliella endoseptemdiera populations from vent locations in the Eastern Lau Spreading Center. Samples are ordered based on similarity. Symbiont populations appear to differ mostly in genes associated with unknown functions, DNA replication and transposition, with smaller variation being observed in other metabolic categories. In accordance with patterns of genetic differentiation, the Tahi Moana population is most distinct in terms of gene content.

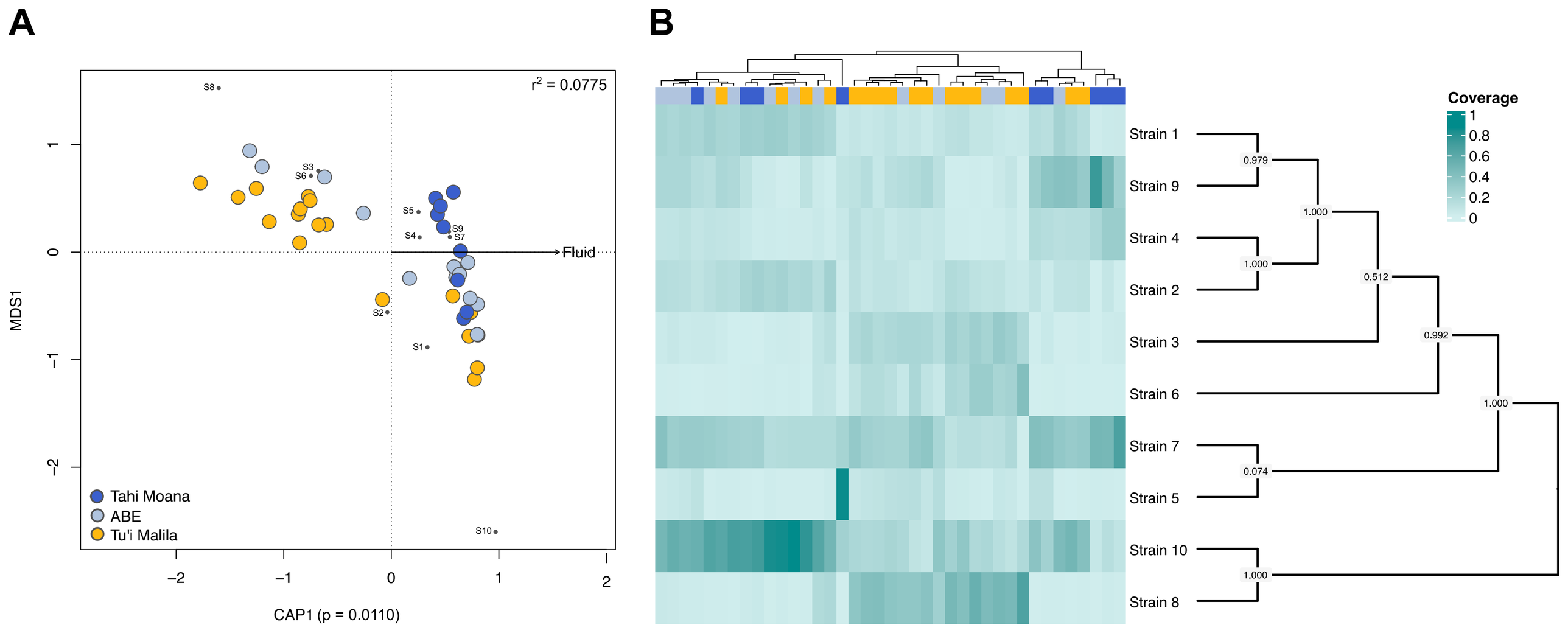
Fig. 4 Analysis of genomic recombination in *Ca.* Thiodubiliella endoseptemdiera. Left: recombination-corrected phylogeny, with tip colors indicating the isolation source of the symbiont genome. Right: recombination and mutation events across the core genome. Recombination events are shown as midnightblue bars, while mutation events are indicated as white to red bars, with white denoting non-homoplasic substitutions and increasing redness denoting increasing levels of homoplasy. Light purple bars indicate absence of substitutions. Recombination is pervasive across the core genome of *Ca.* Thiodubiliella endoseptemdiera, though rates are likely underestimated due to repeated recombination events at any given genomic position.

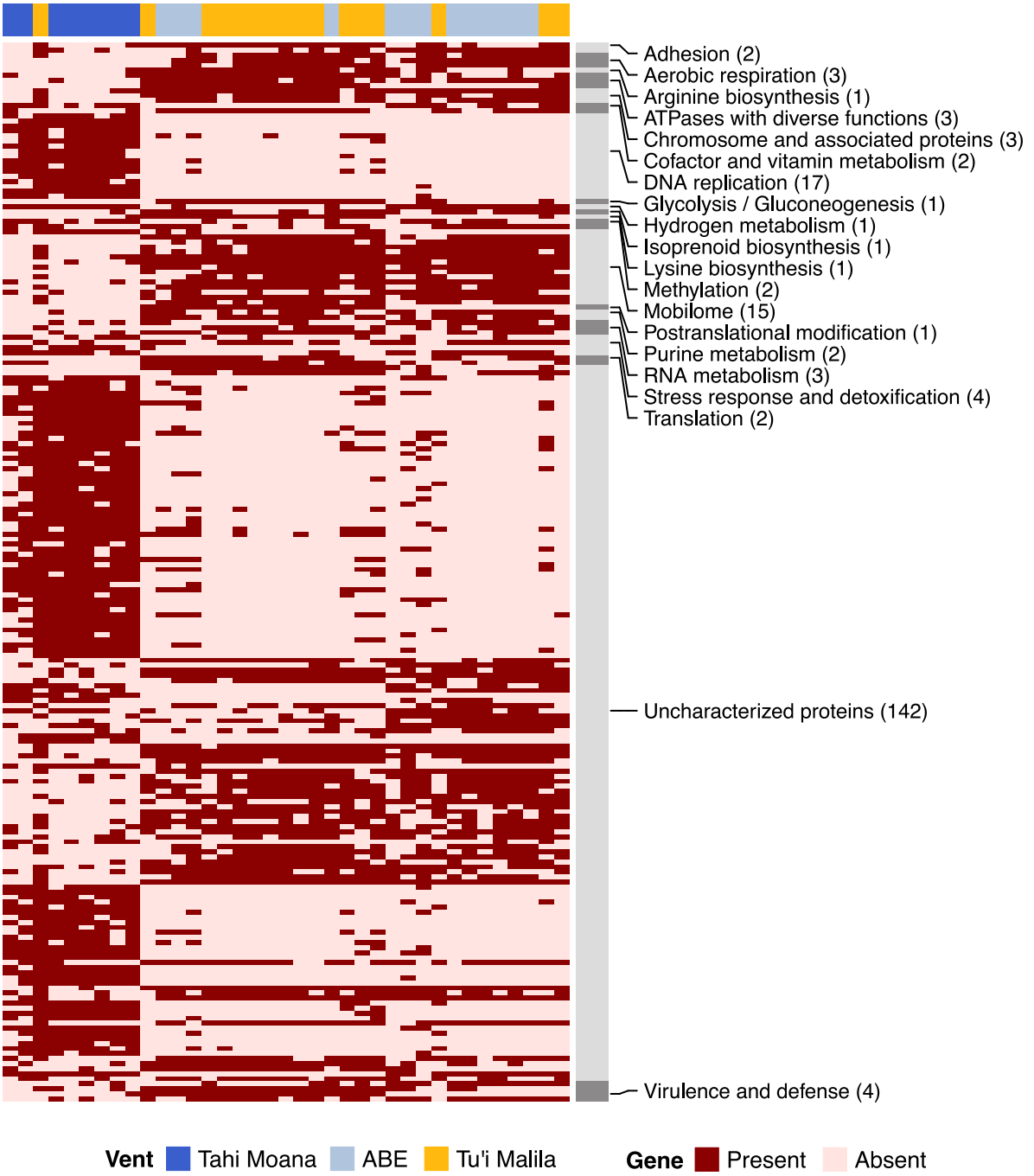

A
Tahi Moana
ABE
Tonga
Ohonua
Tu'i Malila
TONGA
21°S
22°S
176°W
175°W

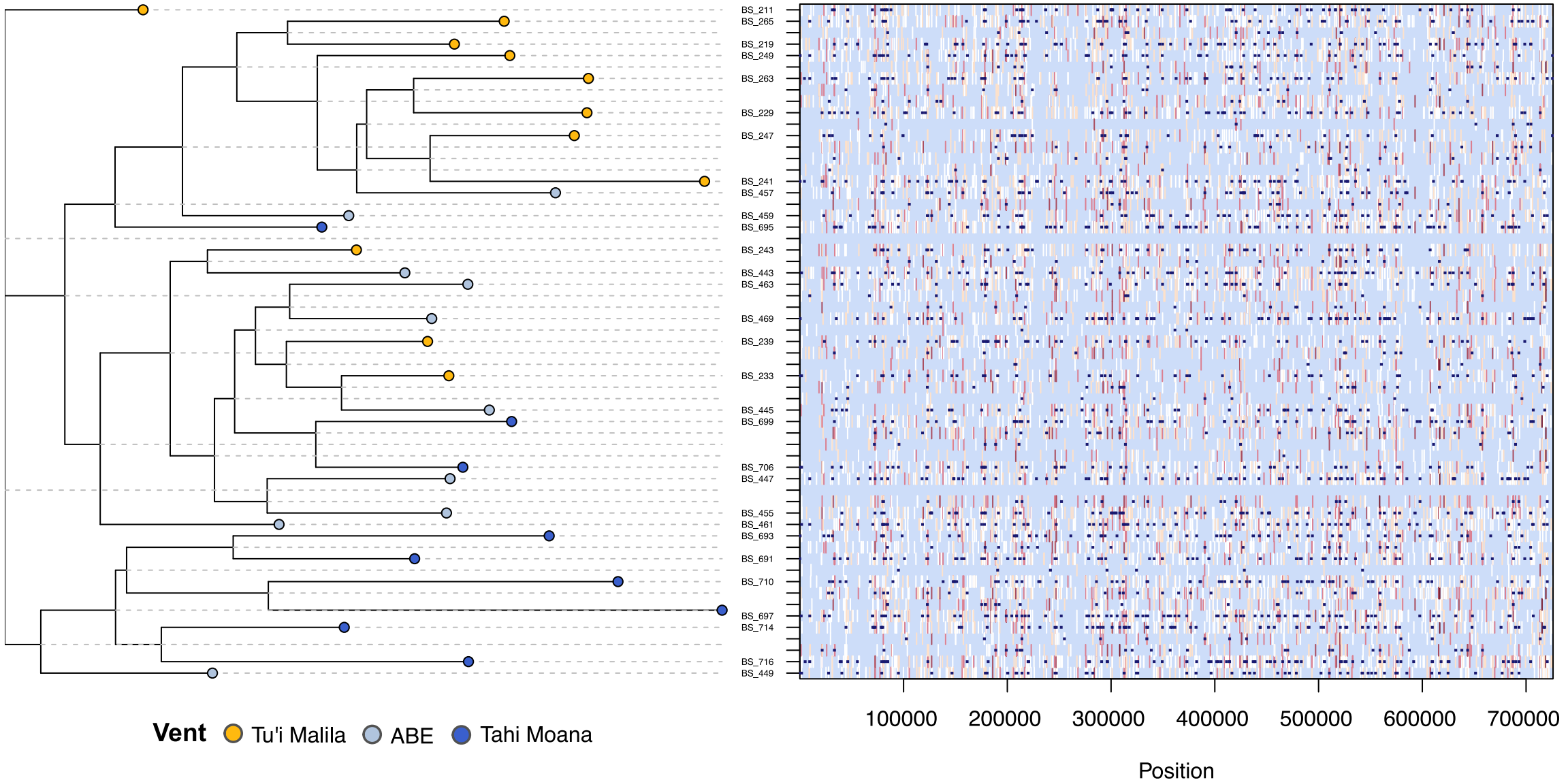

B B. septemdierum
PC2 [4.36%]
PC1 [5.75%]
0.4
0.2
0.0
−0.2
−0.2
0.0
0.2


C Ca. T. endoseptemdiera
Axis2 [11.9%]
Axis1 [19.34%]
0.2
0.1
0.0
−0.1
−0.2
−0.2
−0.1
0.0
0.1
0.2


D
1
0
1
FST
Beta diversity


A
r2 = 0.0775
S8
S3
S6
S5
S4
S9
S7
S2
S1
S10
Fluid
MDS1
CAP1 (p = 0.0110)
Tahi Moana
ABE
Tu'i Malila
B
Coverage
Strain 1
Strain 9
Strain 4
Strain 2
Strain 3
Strain 6
Strain 7
Strain 5
Strain 10
Strain 8
0.979
1.000
1.000
0.512
0.992
1.000
0.074
1.000


Adhesion (2)
Aerobic respiration (3)
Arginine biosynthesis (1)
ATPases with diverse functions (3)
Chromosome and associated proteins (3)
Cofactor and vitamin metabolism (2)
DNA replication (17)
Glycolysis / Gluconeogenesis (1)
Hydrogen metabolism (1)
Isoprenoid biosynthesis (1)
Lysine biosynthesis (1)
Methylation (2)
Mobilome (15)
Postranslational modification (1)
Purine metabolism (2)
RNA metabolism (3)
Stress response and detoxification (4)
Translation (2)
Uncharacterized proteins (142)
Virulence and defense (4)
Vent
Tahi Moana
ABE
Tu'i Malila
Gene
Present
Absent


BS_211
BS_265
BS_219
BS_249
BS_263
BS_229
BS_247
BS_241
BS_457
BS_459
BS_695
BS_243
BS_443
BS_463
BS_469
BS_239
BS_233
BS_445
BS_699
BS_706
BS_447
BS_455
BS_461
BS_693
BS_691
BS_710
BS_697
BS_714
BS_716
BS_449
100000
200000
300000
400000
500000
600000
700000
Position
Vent
Tu'i Malila
ABE
Tahi Moana