Title

Horizontal transmission enables flexible associations with locally adapted symbiont strains in deep-sea hydrothermal vent symbioses.

Journal

Proceedings of the National Academy of Sciences of USA, 119(14)

Authors

Breusing, Corinna
Genetti, Maximilian
Russell, Shelbi
et al.

Publication Date

2022-04-05

DOI

10.1073/pnas.2115608119

Peer reviewed

PNAS RESEARCH ARTICLE | EVOLUTION

Horizontal transmission enables flexible associations with locally adapted symbiont strains in deep-sea hydrothermal vent symbioses

Corinna Breusing[a,1], Maximilian Genetti[b], Shelbi L. Russell[c], Russell B. Corbett-Detig[b], and Roxanne A. Beinart[a,1]

Edited by Margaret McFall-Ngai, University of Hawaii at Manoa, Honolulu, HI; received August 24, 2021; accepted March 2, 2022

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, environmentally acquired 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 and influences 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 an ∼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 locally adapted strains of their specific symbiont phylotypes, which supports a long-standing but untested paradigm of the benefits of horizontal transmission. Symbiont strain flexibility in these snails likely enables host populations to exploit a range of habitat conditions, which might favor widespread genetic connectivity and ecological resilience unless physical dispersal barriers are present.

chemosynthetic symbiosis | hydrothermal vents | host–symbiont population genomics | habitat adaptation | symbiont transmission and specificity

Significance

In marine ecosystems, transmission of microbial symbionts between host generations occurs predominantly through the environment. Yet, it remains largely unknown how host genetics, symbiont competition, environmental conditions, and geography shape the composition of symbionts acquired by individual hosts. To address this question, we applied population genomic approaches to four species of deep-sea hydrothermal vent snails that live in association with chemosynthetic bacteria. Our analyses show that environment is more important to strain-level symbiont composition than host genetics and that symbiont strains show genetic variation indicative of adaptation to the distinct geochemical conditions at each vent site. This corroborates a long-standing hypothesis that hydrothermal vent invertebrates affiliate with locally adapted symbiont strains to cope with the variable conditions characterizing their habitats.

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 hosts' diets (1, 2). For symbiotic associations to persist over evolutionary time, hosts must successfully transmit their symbionts from one generation to the next, either through symbiont acquisition from the environment (horizontal transmission), direct inheritance of symbiont lineages (vertical transmission), often through the host germline, or a combination of both mechanisms (mixed transmission) (3, 4). The mode of transmission has significant implications for the composition and variation of symbionts within and between host individuals. Vertical transmission necessarily results in strong genetic coupling between host and symbiont lineages and an accompanied reduction in intrahost symbiont diversity (3). By contrast, horizontal transmission exposes aposymbiotic hosts to a potentially heterogeneous environmental pool of symbiont lineages (3). This can promote the formation of generalist partnerships, where multiple hosts and symbionts associate with each other, to more specialized associations between only one or a few potential partners (1, 5). This range is often referred to as host–symbiont specificity, which can vary in its taxonomic level for both partners (i.e., partner fidelity) depending on the symbiotic system. The degree of partner fidelity and its effect on symbiont composition can have dramatic impacts on holobiont functioning (6). For example, host–symbiont specificity, both at the species and genotype level, is crucial for light production in bioluminescent squid (7), while shifts in symbiont communities have been shown to affect phytoplankton growth rates (8) and the efficiency of nitrogen fixation in nodule-forming legumes (9–11).

Environmental transmission of obligate symbionts is particularly common in marine ecosystems (2). Apart from host–symbiont genetic interactions, isolation by distance through geographically limited dispersal, selection by environment, or interactive

Author contributions: C.B., S.L.R., R.B.C.-D., and R.A.B. designed research; C.B. and M.G. performed research; S.L.R. and R.B.C.-D. contributed new reagents/analytic tools; C.B., M.G., S.L.R., R.B.C.-D., and R.A.B. analyzed data; and C.B., M.G., S.L.R., R.B.C.-D., and R.A.B. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: corinnabreusing@gmail.com or rbeinart@uri.edu.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2115608119/-/DCSupplemental.

Published March 29, 2022.

effects of these processes can shape the composition, diversity, and structure of horizontally transmitted symbionts across host populations (12). However, despite their importance for holobiont biology, the relative contributions of these factors to symbiont composition within and among hosts remain understudied in most environmentally acquired marine symbioses. It has long been hypothesized that horizontal transmission enables host organisms to associate with locally adapted symbiont strains (i.e., subtypes within a bacterial species), conferring fitness advantages in spatially and temporally variable marine habitats (3, 13–15). Especially for long-dispersing aposymbiotic larvae that are likely to encounter new habitat conditions when they settle, association with a locally adapted symbiont strain may be advantageous compared to carrying a vertically transmitted symbiont that might be maladapted at a nonnative site. This hypothesis has been indirectly supported by evidence that marine animals with horizontally transmitted obligate microbial symbionts often host location-specific strains (15–20). However, the influence of local adaptation relative to neutral evolutionary processes on symbiont geographic structure and genomic traits has not been formally evaluated.

Chemosynthetic animal–microbe symbioses are globally significant phenomena that dominate hydrothermal vent and hydrocarbon seep ecosystems in the deep sea. In these associations, the bacterial partner uses chemical energy from the oxidation of reduced compounds, such as hydrogen, sulfide, or methane, to synthesize organic matter, which serves as primary nutrition for the host (21). Vent animals harboring chemosynthetic symbionts are typically very specific in their partnerships: In the predominant number of cases, host individuals associate with only one or two phylotypes of gammaproteobacterial or campylobacterial symbionts (i.e., species- or genera-level phylogenetic clades based on 3 and 5% 16S ribosomal RNA [rRNA] sequence divergence, respectively) (21), whereas symbionts can exhibit a comparatively broad host range. Chemosynthetic endosymbioses in deep-sea snails of the Indo-Pacific sister genera *Alviniconcha* and *Ifremeria* are examples of reciprocally relatively specific partnerships, where host species harbor only particular symbiont species or genera across their geographic distribution. Given the absence of host–symbiont phylogenetic concordance, the symbionts are assumed to be environmentally acquired in both genera (22) despite contrasts in reproductive mode. Although reproductive strategy does not always predict symbiont transmission mode (23), a vertical transmission component is more likely in *Ifremeria*, as this genus is known to brood its larvae (24), giving symbionts more developmental time to transfer.

In the Eastern Lau Spreading Center (ELSC), previous work suggested that specificity to functionally distinct symbiont phylotypes drives local- and regional-scale habitat partitioning among four co-occurring *Alviniconcha* and *Ifremeria* species (25–28). *Alviniconcha boucheti* from the ELSC contains a campylobacterial phylotype (Epsilon) and is usually found at northern vent sites with high concentrations of sulfide and hydrogen, while *Alviniconcha kojimai* and *Alviniconcha strummeri* associate with different gammaproteobacterial phylotypes (Gamma1 and Gamma1/GammaLau, respectively) and usually occupy midlatitude to southern vent sites, where the concentrations of these chemical reductants are lower (25, 27–29). *Ifremeria nautilei* establishes dual symbioses with thiotrophic and methanotrophic gammaproteobacterial endosymbionts (29) and is codistributed with *Alviniconcha* across their geographical range, although it typically segregates into habitat patches with reduced fluid flow relative to its sister genus (30). While it has been well-established that niche differentiation across hydrothermal vents is likely linked to symbiont phylotype specificity in *Alviniconcha* and *Ifremeria* host species, nothing is known about the fidelity of these associations at the population level, how strain-level specificity might influence host population structure and connectivity, and how regional adaptation is conferred functionally.

In this study, we applied population genomic methods to assess symbiont strain-level genetic variation and patterns of host–symbiont genetic subdivision in *Alviniconcha* and *Ifremeria* species from the ELSC and Tonga Volcanic Arc (Fig. 1*A* and Dataset S1). We first assembled transcriptomes and pangenomes for each host species and symbiont phylotype, respectively, and then determined sequence and functional variation between populations by mapping reads from multiple individuals against the reference assemblies. Using multivariate statistical analyses, we subsequently evaluated the impact of host genetics, environment, and geography on symbiont composition in populations of all four snail species and assessed the effect of local adaptation on symbiont geographic structure. To test if vertical transmission has linked symbiont and host genealogies or if horizontal transmission has randomized these relationships, we further assessed the topological concordance among symbiont and host mitochondrial phylogenies.

Results

Host Populations Are Not Structured across Geographic Locations. Host transcriptomes were coassembled from 11 to 25 individual gill samples per species and used as references for variant calling to assess host population genomic structure. Transcriptome assemblies consisted of 24,176 to 35,654 transcripts (totaling 20.68 to 28.68 Mb) and were ∼30.30 to 55.40% complete (Dataset S2). Mapping of 1,138,186 to 541,418,710 (mean 20,984,011) host reads against the transcriptome references (Dataset S3) and subsequent filtering of variant sites yielded 1,655 to 9,185 single-nucleotide polymorphisms (SNPs) per species for population genetic analyses. Irrespective of host taxon, F_{ST} and ordination analyses revealed low genetic differentiation among host populations sampled from different vent localities (Fig. 1*B* and Dataset S4; $F_{ST} = 0.0400$ to 0.1579). With the exception of four SNP sites in *A. kojimai*, no F_{ST} outliers that could be indicative of adaptive variation were detected in any host species at a q value $\leqq 0.05$ (Dataset S5). The effect of geography on host population genetic structuring was not significant in Mantel correlations (Dataset S6).

Snail Symbionts Belong to Six Bacterial Species That Are Frequently Horizontally Transmitted. Pangenomes of each symbiont phylotype were reconstructed from 4 to 59 high-quality symbiont metagenome assembled genomes for population genomic and phylogenomic analyses (Datasets S7 and S8). Based on average nucleotide identities and taxonomic assessments, symbiont phylotypes recovered in our dataset comprised six species from five genera of chemosynthetic bacteria: *Sulfurimonas* "Epsilon" (*A. boucheti* from the ELSC), a second *Sulfurimonas* species (*A. boucheti* from Niua South), unclassified Gammaproteobacteria "Gamma1" (*A. kojimai*, *A. strummeri*), Sedimenticolaceae "GammaLau" (*A. strummeri*), *Thiolapillus* "Ifr-SOX", and Methylomonadaceae "Ifr-MOX" (*I. nautilei*) (Dataset S7). Because of the low coverage of the GammaLau and Ifr-MOX species that resulted in poor variant recovery (*SI Appendix*, Fig. S1 and Dataset S7), these symbionts were excluded from further analyses. In addition, the *Sulfurimonas*

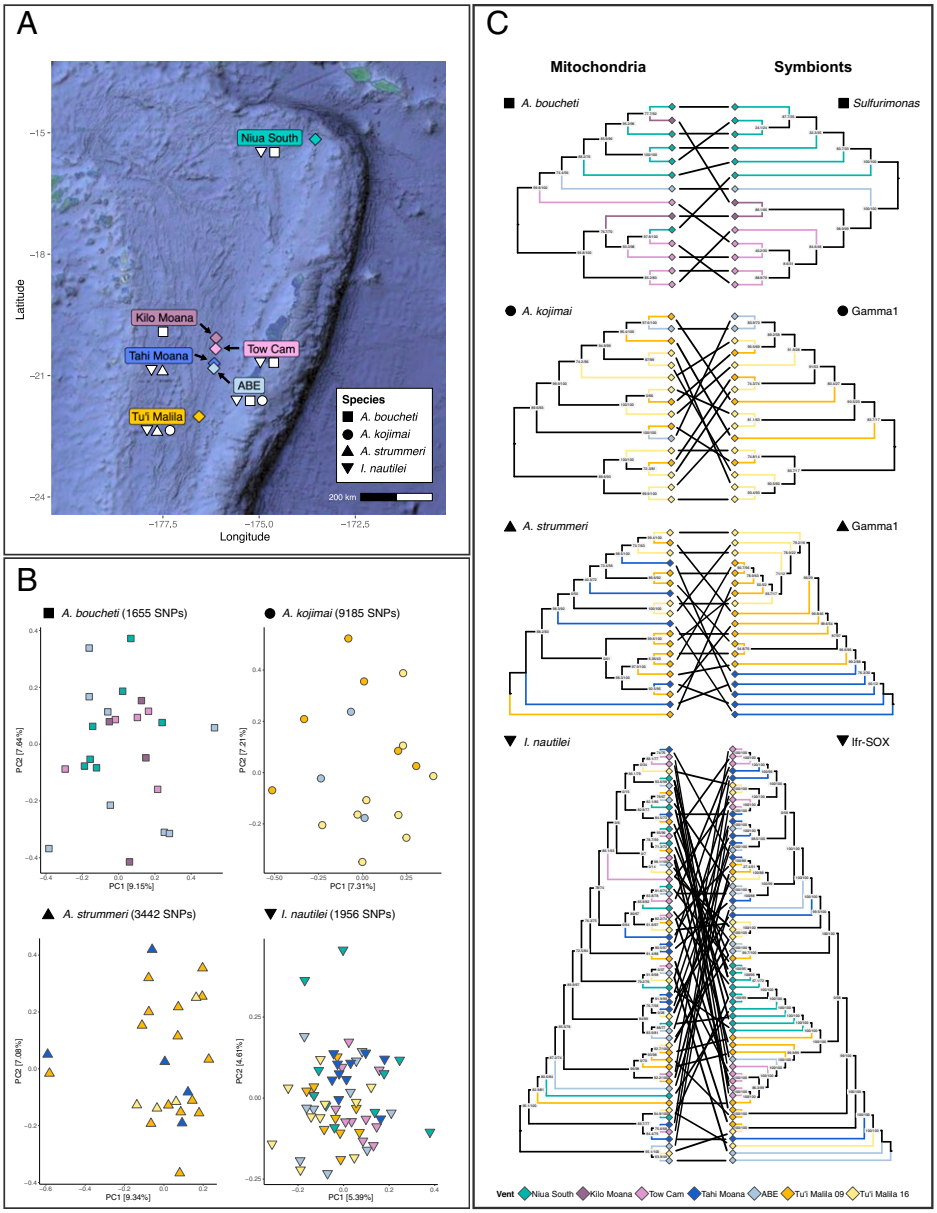

Fig. 1. (*A*) Sampling map for *Alviniconcha* and *Ifremeria* species at vent fields along the ELSC (Kilo Moana, Tow Cam, Tahi Moana, ABE, and Tu'i Malila) and the Tonga Volcanic Arc (Niua South). Vent field locations are shown as colored diamonds and the symbols below each site label indicate the species sampled from the respective vent field. Map data ©2021 Google. (*B*) Principal-component plots for *Alviniconcha* and *Ifremeria* host species based on genetic covariance matrices. Axes show percent of variation explained by the first and second principal components. Each symbol represents a single snail sample colored by vent field of origin. Symbols for all species cluster without identifiable pattern, indicating that host individuals across locations are genetically similar and not isolated through ecological or physical barriers. PC, principal component. (*C*) Host mitochondrial and symbiont cladograms. Node labels indicate support values from ultrafast bootstrapping and Shimodaira–Hasegawa approximate-likelihood ratio tests. Lines connecting branches on each tree indicate host–symbiont pairs. The tree topologies are clearly discordant, indicating that symbionts are frequently environmentally acquired.

species from Niua South was not included in population genomic analyses due to taxonomic divergence from the Epsilon symbiont associated with *A. boucheti* from the ELSC. The dominant symbiont species within the ELSC (Epsilon, Gamma1, and Ifr-SOX) were considered for all downstream analyses.

To assess if symbionts exhibit evidence of vertical transmission, we determined the congruence between tree topologies of the host mitochondrial and symbiont core genomes. For all host species, symbiont and mitochondrial phylogenies were clearly discordant (Fig. 1*C*). This indicates that symbionts are primarily environmentally acquired or that enough horizontal transmission has occurred to erode the signal of vertical transmission in both *Alviniconcha* and *Ifremeria* despite differences in reproductive strategy.

Symbiont Populations Are Segregated by Geography, with Little Effect of Host Genetics. Populations of the dominant ELSC symbiont species, Epsilon, Gamma1 and Ifr-SOX, were structured by geography based on 1,716 to 7,741 variants (Fig. 2, *SI Appendix*, Fig. S2, and Dataset S9). Genetic variation between symbiont populations typically increased with geographic distance between vent fields and could be broadly partitioned into four regions: Tu'i Malila, ABE + Tahi Moana, Tow Cam + Kilo Moana, and Niua South (Dataset S9; F_{ST} = 0.1012 to 0.8682). In general, differentiation within the ELSC was larger for the *Alviniconcha* symbionts than for the *Ifremeria* symbionts across the same geographic scale. Genotype–site associations were significant even when corrected for host genetic variation (Dataset S6; $P \leqq 0.0285$, $r = 0.2861$ to

0.8512), with the strongest geographic effect being observed in the Epsilon and Ifr-SOX symbionts of *A. boucheti* and *I. nautilei*, respectively.

The Gamma1 symbiont was associated with both *A. kojimai* and *A. strummeri* and we therefore investigated whether populations of this phylotype varied between host species. Ordination analyses and Mantel tests indicated a clear affiliation to host taxon that superseded the effect of geography by about a factor of 3, suggesting that different Gamma1 strains associate selectively with either *A. kojimai* or *A. strummeri* (Fig. 2, *SI Appendix*, Fig. S3, and Dataset S6; $P \leqq 0.0002$, $r = 0.6844$ to 0.7424). By contrast, effects of host genetics on symbiont genetic variation within species appeared to be absent.

Symbiont Genetic Variation Is Partly Explained by Local Adaptation. To determine if some of the recovered genetic variants between geographic and host-specific symbiont strains were subject to positive selection, we evaluated genotype–environment associations through redundancy analyses (RDAs). In all but one analysis performed for symbionts from single host species, 3.42 to 7.57% of genetic variants exhibited significant correlations with hydrothermal fluid composition, depth, and/or year ($P \leqq 0.05$), with constrained ordinations explaining between 8.88 and 13.66% of the variance in the respective RDA models (Fig. 2 and Dataset S10). The RDA model for the Gamma1 symbiont from *A. kojimai* was not significant. When combining strains from both *A. kojimai* and *A. strummeri* in the analysis, however, 4.18% of variants of the Gamma1 symbiont were significantly associated with host species, while 3.27% were correlated with other environmental factors (Fig. 2 and Dataset S10; $P = 0.001$, $r^2_{adj} = 38.54\%$). Likewise, the overall marginal effects of each explanatory variable that was included in the RDA models for the different host–symbiont pairs were significant, except for year in the case of the Gamma1 symbiont (Dataset S11).

Candidate adaptive loci in each symbiont species encompassed a wide range of functional categories, from metabolism of macronutrients, cofactors, and cell wall/membrane to energy production, maintenance of cellular homeostasis, motility, mobile elements, and response to environmental stressors and phages (Dataset S10). In concordance with the known environmental differences among vent sites in the ELSC and previous observations of habitat partitioning between *Alviniconcha* holobionts linked to vent geochemistry (25), a subset of adaptive variants in the Epsilon and Gamma1 symbionts was located in

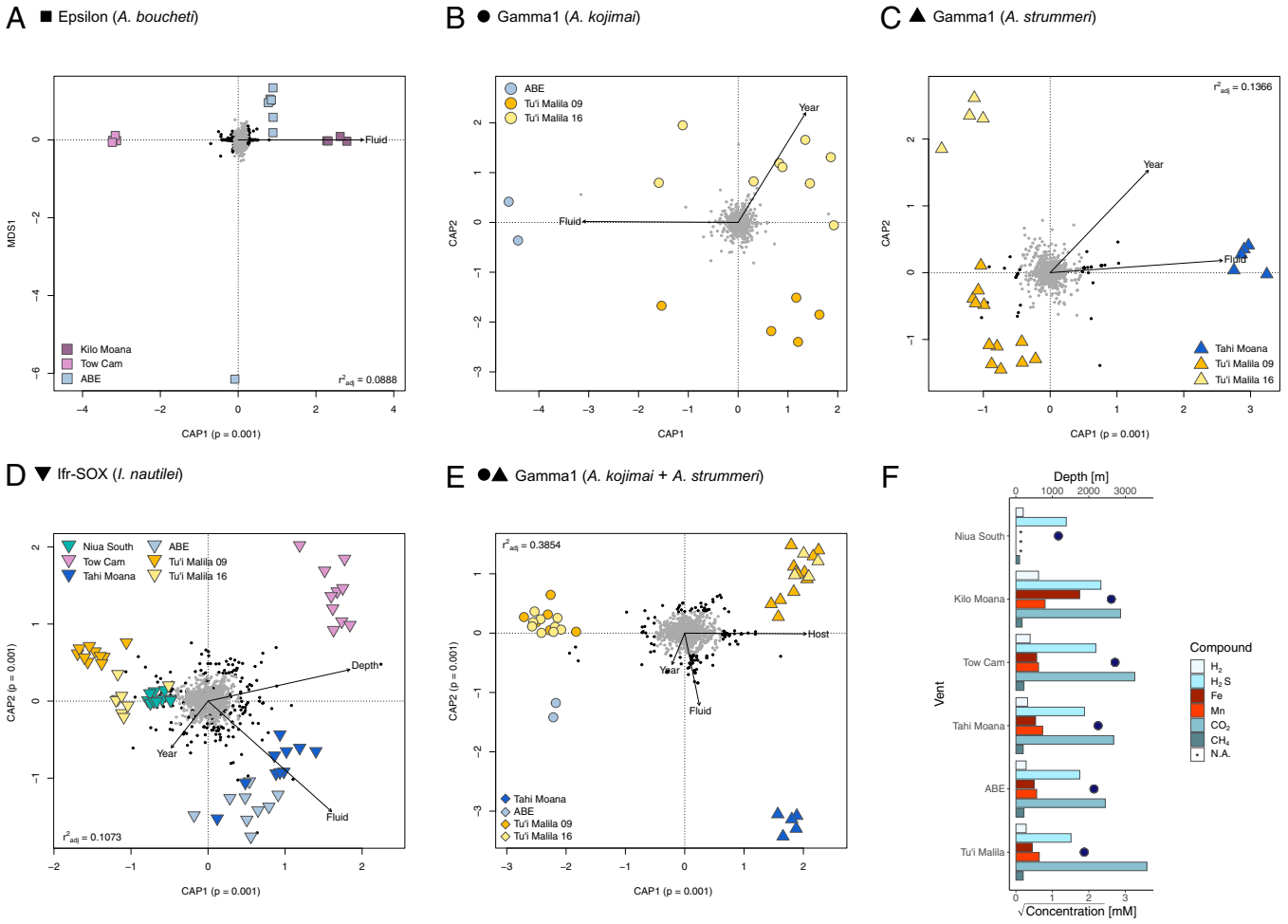

Fig. 2. Redundancy analysis plots of the first two ordination axes for *Alviniconcha* and *Ifremeria* symbionts. (*A*–*E*) Small dots in the center of the plots indicate genetic variants, with outlier loci shown in black. Large colored shapes represent intrahost symbiont populations sampled from different vent localities. Vectors show the environmental predictors. Redundancy analyses were conditioned by geography for all symbionts except for Gamma1. *P* values for constrained ordination axes are shown in the axis labels. Genotype–environment associations were not significant for Gamma1 associated with *A. kojimai* when analyzed separately by host species. r^2 values adjusted for the number of parameters indicate the proportion of variation explained and are expectedly low as only a fraction of variants is adaptive. CAP (constrained analysis of principal coordinates), constrained ordination axis. MDS (multidimensional scaling), unconstrained ordination axis. (*F*) Average depth (dark blue dots) and endmember concentrations of fluid compounds (bars) at vent sites of the Tonga Volcanic Arc and ELSC from north to south. Chemical concentrations were square root–transformed to improve representation of small values. N.A., not available.

genes related to sulfur and hydrogen metabolism (e.g., sulfide-quinone oxidoreductase, sulfate permeases, carbamoyltransferase HypF). Conversely, variants associated with these functions appeared to not be under positive selection in the Ifr-SOX symbiont. Notably, this symbiont encoded loci involved in conjugation that showed evidence for adaptive evolution and that were not detected as outliers in the *Alviniconcha* symbionts. Interestingly, all symbionts contained variants in genes related to vitamin B12 metabolism that were linked to local adaptation. In the Gamma1 symbiont, adaptive loci related to host species comprised several genes involved in immunity (e.g., production of cell-wall/membrane compounds, virulence, response to oxidative stress), nutrient provisioning (carbon fixation, cofactor and macronutrient metabolism), and cellular respiration. In general, candidate variants in all symbiont species were characterized by elevated F_{ST} values and were located in genes with increased or reduced pN/pS ratios (Dataset S12), which further supports their potential role in habitat adaptation and highlights the essentiality of the respective genes for symbiont functioning.

Symbiont Populations Differ in Gene Content. We used the symbiont pangenome references to further assess variation in gene content between symbiont populations from different vent localities or broader geographic regions based on groupings observed in the ordination and F_{ST} analyses (Fig. 3, *SI Appendix*, Fig. S4, and Dataset S13). For the Gamma1 symbiont, we additionally determined gene content variation between populations from different host species, given that this symbiont species occurs in both *A. strummeri* and *A. kojimai* (*SI Appendix*, Fig. S3 and Dataset S13). Differentially preserved genes between symbiont populations comprised about 1.33 to 3.84% of protein-coding regions.

Independent of symbiont species, the most common differences among geographic or host-specific populations concerned genes of unknown or miscellaneous function as well as a smaller number of genes related to mobilome (e.g., transposases, integrases), antiviral defense (e.g., restriction–modification systems, CRISPR-Cas systems), and DNA metabolism (e.g., DNA polymerases, helicases, repair systems) (Fig. 3, *SI Appendix*, Figs. S3 and S4, and Dataset S13). Populations of the *A. boucheti* Epsilon symbiont from Kilo Moana/Tow Cam and ABE further differed in several cell wall/membrane-related genes, a protein phosphatase, a 3-dehydro-bile acid delta(4, 6)-reductase, and an inosine-5′-monophosphate dehydrogenase (Fig. 3, *SI Appendix*, Fig. S4, and Dataset S13). Within the Gamma1 symbiont of *A. kojimai*, populations from Tu'i Malila contained multiple private genes that were related to cell signaling, iron–sulfur cluster biosynthesis (*sufS*), iron uptake, and other membrane transport processes, whereas populations from ABE encoded genes for succinate dehydrogenase assembly and hydrogenase maturation. Differences in hydrogenase-related genes were also evident in the Gamma1 symbiont of *A. strummeri*, where genes for uptake and sensory hydrogenases (*hoxKL*, *hydB*) as well as hydrogenase maturation and regulation (*hypB*, *hypDEF*, *hupR1*, *hupU*) were largely missing in populations from Tu'i Malila (Fig. 3, *SI Appendix*, Fig. S4, and Dataset S13). In addition, these populations lacked several transcription factors and RNA helicases that were present in lineages from Tahi Moana, while containing private genes involved in cell signaling, membrane transport, as well as carbon, vitamin B12, and secondary metabolism. Similar contrasts in metabolic functions were observed between host-specific strains of the Gamma1 symbiont. Most notably, host-specific populations differed in the presence of genes for a sulfite dehydrogenase complex (*soeABC*), which was present in strains from *A. kojimai* but not *A. strummeri* (*SI Appendix*, Fig. S3 and Dataset S13). The *I. nautilei* Ifr-SOX populations from Niua South and the ELSC showed contrasting conservation patterns in genes that spanned a variety of metabolic and cellular functions, including asparagine biosynthesis, macronutrient, cofactor, nitrogen, nucleotide, and RNA metabolism, transport of iron and other inorganic compounds, cell regulation, conjugation, and cellular respiration (Fig. 3, *SI Appendix*, Fig. S4, and Dataset S13).

Discussion

Strain-level variation within microbial symbionts is increasingly recognized as an important driver of host ecology and evolution (6), yet its patterns, determining factors, and functional implications remain poorly investigated in many nonmodel symbioses. In this study, we used metagenomic analyses to assess sequence and gene content differences in chemosynthetic symbionts associated with four co-occurring species of deep-sea hydrothermal vent snails and determined the impact of host genetic, geographical, and environmental factors on symbiont strain composition and variation.

Despite fidelity between hosts and symbionts at the host species level (25, 27), our results indicate that specificity between host genotypes and symbiont strains in *Alviniconcha* and *Ifremeria* is weak: Host populations were not partitioned across an ∼800-km gradient, whereas symbiont populations were clearly structured between vent locations or broader geographic regions. Correlations between symbiont and host genetic distances were either insignificant or low, suggesting that host genetics has a minor impact compared to environment or geography on strain composition within both *Alviniconcha* and *Ifremeria*.

These findings qualitatively agree with observations in deep-sea mussel hybrids that appear to associate with locally available symbiont strains (19), but contrast markedly with patterns in other horizontally transmitted partnerships, such as the squid–*Vibrio* symbiosis and some legume–rhizobia associations, where hosts exhibit strong strain specificity (6). Environmental uptake of locally adapted symbiont strains provides host organisms with the opportunity to optimally exploit novel habitats, but carries the risk of unsuccessful symbiont acquisition and infection by cheaters (3). Depending on the amount of partner reliance, chances of symbiont encounter, and fitness variation among symbiont strains, holobionts likely find different trade-offs between these opposing factors. Associations with chemosynthetic bacteria are obligate for hydrothermal vent animals, but are restricted to relatively ephemeral habitats that are characterized by large temporal and spatial fluctuations in environmental conditions and associated shifts in microbial communities (31–33). Strong nutritional dependency in vent symbioses combined with the uncertainty of habitat (and thus symbiont) encounter might promote specificity toward a mutualistic symbiont phylotype, while enabling enough flexibility toward different local strains of that phylotype to maximize recruitment success of host larvae at a new vent site. Although vent tubeworms, for example, are known to release symbionts into their environment upon death (34), it is not guaranteed that larvae self-recruit and take up the same symbionts as their population of origin, given the transient nature of hydrothermal vents. By contrast, juvenile bobtail squids affiliate with symbiotic bacteria that are expelled by adult hosts into the native environment (35), which might favor increased strain selectivity in this association. Similarly, some legume species

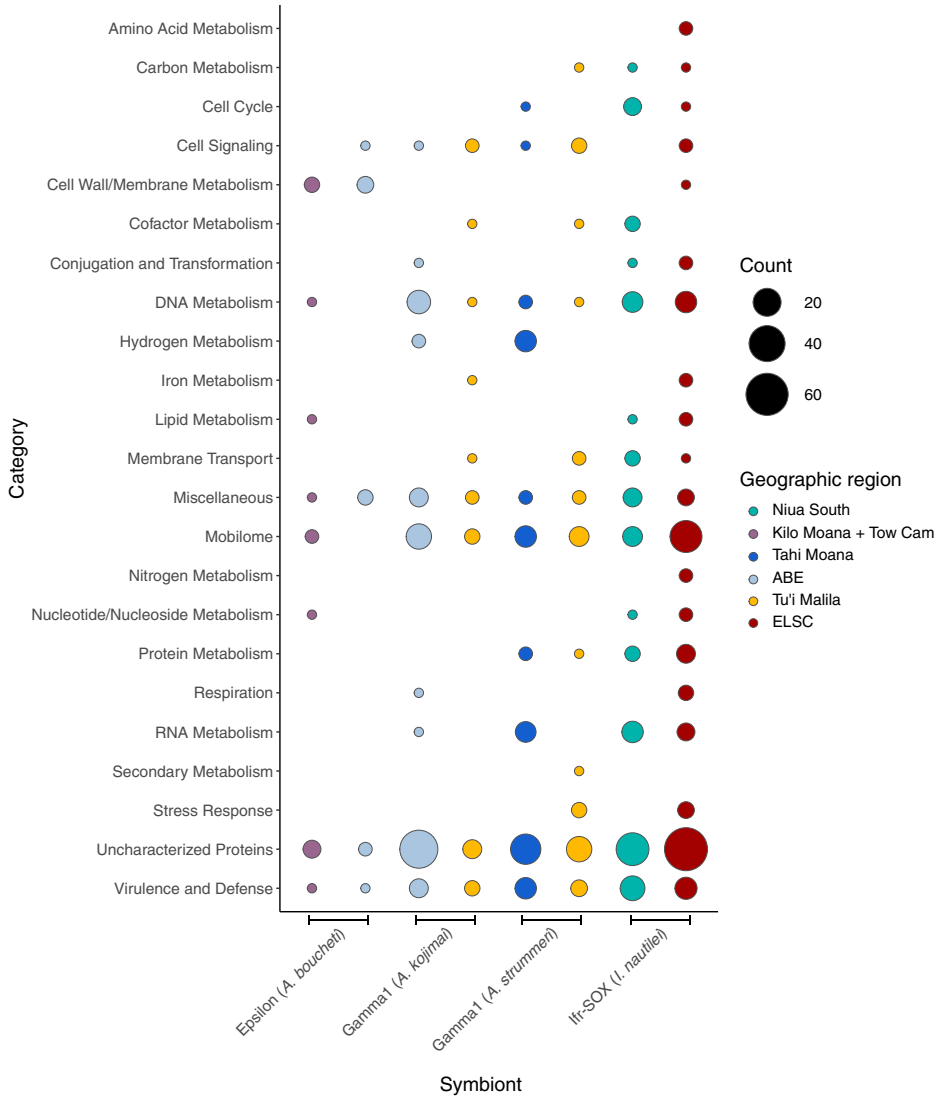

Fig. 3. Dotplot showing differences in gene content for each functional category between habitats and symbiont species. Dots are scaled by the number of genes in each category and colored by geographic region. ELSC includes all vent sites in the Lau Basin sampled for *Ifremeria* (Tow Cam, Tahi Moana, ABE, Tu'i Malila). Presence/absence of genes was determined in PanPhlAn based on gene coverage profiles.

appear to preferentially associate with certain rhizobial strains that are abundant in the native host range (36). Although the factors underlying strain specificity are not fully understood, it is possible that the relative fitness advantages provided by these locally available strains (36) coupled with the benefit of decreased cheater invasion (11) might have selected for strong partner fidelity in these symbioses.

A. kojimai and *A. strummeri* represent notable exceptions to the observed patterns of symbiont specificity, as they share one species of gammaproteobacterial symbionts (Gamma1), but appear to take up functionally distinct, site-specific strains even in habitats where these host species co-occur. Strains of both host species showed evidence for positive selection in genes encoding outer-membrane proteins and factors for the biosynthesis of cell-wall lipopolysaccharides, which likely play a role in host–symbiont recognition and specificity (37).

The phylogenetic divergence between *A. kojimai* and *A. strummeri* (~25 million years ago; MYA) is more recent than that between either species and *A. boucheti* (~38 MYA) or *I. nautilei* (~113 MYA) (27). Perhaps the closer evolutionary relationship between these two species favors associations with the same symbiont phylotype but distinct symbiont strains. A similar scenario has been suggested for *Bathymodiolus azoricus* and *Bathymodiolus puteoserpentis* on the Mid-Atlantic Ridge (~8.4 MYA) (38) as well as *Bathymodiolus thermophilus* and *Bathymodiolus antarcticus* on the East Pacific Rise (~2.5 to 5.3 MYA) (39), which do not show phylotype specificity in contrast to their more divergent relatives from Atlantic cold seeps (~15.4 MYA) (19, 38). However, compared with the bathymodiolin mussel system, the split between *A. kojimai* and *A. strummeri* is significantly older, which could indicate that the timing of phylotype specificity evolution varies among taxonomic groups, possibly as a result of contrasting ecological or evolutionary contexts. The fact that *A. kojimai* and *A. strummeri* nevertheless associate with distinct symbiont strains in sympatry is potentially a mechanism to avoid competition for niche space. Gamma1 strains of these two species notably differed in the presence of a molybdenum-containing sulfite dehydrogenase (SoeABC), which was preserved in strains of *A. kojimai* but not *A. strummeri*. SoeABC is the predominant enzyme for sulfite oxidation in many purple sulfur bacteria (40) and is further involved in taurine and dimethylsulfoniopropionate degradation in *Roseobacter* clade bacteria (41–43). Although the physiological role of SoeABC in *A. kojimai*'s Gamma1 strain

is unknown, it is possible that it contributes to partitioning of sulfur resources among co-occurring snail holobionts.

Our data support a model of horizontal symbiont transmission in both *Alviniconcha* and *Ifremeria*. While these results are expected for *Alviniconcha* that produce free-swimming planktotrophic larvae and do not invest in their young (44), they are rather surprising for *Ifremeria* that brood their offspring in a modified pouch in the female's foot (24) and would thus have more opportunities to vertically transmit their symbionts (independent of possible transovarial inheritance). Although topological discordance between host mitochondrial and symbiont phylogenomic trees suggests frequent horizontal transmission, it is possible that *Ifremeria* symbionts exhibit a mixed transmission mode and maternally acquired symbionts get replaced or complemented by more competitive strains in the habitat where the snail larvae settle. Our current dataset cannot distinguish this possibility from a strict horizontal transmission mode, though future studies assessing symbiont composition in different developmental stages of *Ifremeria* would be helpful to address this hypothesis.

While the majority of genetic variants in the symbionts did not deviate from neutral expectations, a subset of loci showed evidence for natural selection, implying a role of both genetic drift and local adaptation in shaping symbiont population structure. Candidate adaptive loci spanned a surprisingly broad range of metabolic functions, many of which were probably not causally linked to the investigated environmental predictors but other correlated variables that we could not account for in this study. For example, in all investigated symbiont species, we observed adaptive variation in genes that were related to antiviral defense (e.g., CRISPR-associated proteins, restriction–modification systems) or mobile elements. Although differences in depth or geochemistry might contribute to these patterns, it is more likely that they reflect exposure of symbiont strains to distinct viral assemblages that might covary with local habitat conditions, as has been suggested for vestimentiferan tubeworm symbionts (45). Similarly, all symbionts showed evidence for local adaptation associated with vitamin B12 metabolism. This could be due to habitat-specific differences in the availability of cobalt, which is an essential trace element for the formation of vitamin B12. Though cobalt concentrations are unknown for hydrothermal fluids in the Lau Basin, they are typically correlated with those of iron and manganese and might thus show similar variation between vent fields, given the solubility dependence of all of these metals on temperature and chlorinity (46).

Polymorphisms in other genes, by contrast, are likely directly explained by variation in the analyzed environmental factors. Geographic strains of *A. boucheti*'s Epsilon symbiont, for instance, contained several variants under positive selection that were located in genes involved in sulfur oxidation and respiration. Likewise, the Gamma1 strains of *A. kojimai* and *A. strummeri* showed adaptive differences in variants related to hydrogen and sulfur metabolism. Niche-specific differences in hydrogen metabolic genes were also observed in comparisons of gene content among symbiont strains. In both *A. kojimai* and *A. strummeri*, Gamma1 strains from Tu'i Malila lacked subunits of uptake and hydrogen-sensing hydrogenases as well as various genes for hydrogenase maturation and regulation. Given that H_2 concentrations at Tu'i Malila can drop to 35 μM in endmember fluids and are probably lower in diffuse-flow habitats (25), it is possible that hydrogen does not constitute a major energy source for Gamma1 strains from this locality. This hypothesis is in agreement with previous physiological experiments that revealed strikingly low hydrogen oxidation and associated carbon fixation rates in Gamma1 symbionts from Tu'i Malila (28). An alternative though mutually nonexclusive explanation for the loss of hydrogenase-related genes at least in the *A. strummeri* strain from Tu'i Malila could be avoidance of intrahost competition with co-occurring GammaLau strains, which sometimes codominate in *A. strummeri* individuals at this vent site (25). Such functional diversity is predicted to enable symbiont coexistence in a variety of hydrothermal vent symbioses, including bathymodiolin mussels (47, 48), alvinocaridid shrimp (49), and vestimentiferan tubeworms (50). In contrast to the *Alviniconcha* symbionts, the *Ifremeria* Ifr-SOX symbiont did not show evidence for adaptive variation related to hydrogen and sulfur metabolism. This finding is not surprising, given that *Ifremeria* holobionts typically inhabit patches with less variable sulfide concentrations than *Alviniconcha* (51) and do not show the same patterns of regional-scale niche partitioning. Interestingly, we observed a higher number of adaptive variants linked to conjugative transfer in this symbiont. Perhaps site-specific selection in the respective genes promotes increased exchange of genetic material in the Ifr-SOX symbiont, which could improve adaptation to diverse environmental conditions and might thus enable this symbiont to occupy a broader range of ecological niches.

Although geography and environment appeared to be the dominant drivers of symbiont strain composition, some differences were linked to year of collection. In particular, strain pools of the Ifr-SOX symbiont from Tu'i Malila showed significant temporal variation. Given that the Kilo Moana vent field expired between 2009 and 2016, it is likely that the ELSC experienced changes in hydrothermal activity and associated geochemical conditions during this time period, which could have induced shifts in biological communities. If individuals only acquire their symbionts during a short developmental window, the observed strain pools might reflect distinct host cohorts, or, alternatively, it is possible that long-lived host individuals replace their symbionts over time, resulting in compositional differences between years. Under both scenarios, however, potential symbiont subdivision within vent fields might have exacerbated temporal variation in intraspecific strain composition, given that different snail patches (41 to 87 m apart) were sampled between years.

Compared with their symbionts, host populations were markedly less structured across the same spatial scales. Although a proportion of genetic markers was differentiated across vent sites (0.41 to 12.70% with $F_{ST} > 0.15$), we did not find strong evidence for local adaptation in any host species, suggesting that these patterns likely reflect random variation among localities. The limited genetic subdivision between host populations agrees with predictions from biophysical models that indicate absence of physical dispersal barriers for vent larvae within the Lau Back-Arc Basin (52). While symbiont population structure, by contrast, appears to be at least partly driven by natural selection, it is possible that symbionts also experience stronger dispersal limitations than their hosts, as has been hypothesized in some coral–algae symbioses (18). The environmental distributions and life cycles of the free-living stages of *Alviniconcha* and *Ifremeria* symbionts are currently unknown, and a better understanding of these aspects will be necessary to evaluate the relative importance of dispersal barriers in symbiont biogeography.

Overall, our findings reveal that a lack of strain-level specificity in *Alviniconcha* and *Ifremeria* symbioses is common and possibly reflects an evolutionary strategy to cope with the transient and dynamic nature of hydrothermal vent habitats. However, a deviation from this pattern appears to occur in relatively closely related sympatric host taxa, which can evolve strong fidelity to local strains of the same symbiont species. In general,

strain flexibility might contribute to the widespread genetic connectivity observed among host populations, since the ability to partner with local strains should facilitate success across broad geographic and environmental scales and decrease ecological barriers to gene flow. In turn, this could favor resilience to natural but also anthropogenic environmental disturbances, a relevant consideration given the increasing human pressures on hydrothermal ecosystems worldwide (53). Our observations further support the fundamental hypothesis that horizontal transmission enables host organisms to associate with locally adapted symbiont strains (3, 12–14). Though the genomic basis of local adaptation can be detected in natural populations using population genomics methods, as we did here, evaluation of the phenotypic consequences of the observed strain-level genomic trait variation will be necessary to confirm local adaptation in these symbiont strains. Future work using organism-based manipulative experiments will be helpful to compare the fitness of hydrothermal vent animals hosting site-specific symbiont strains when exposed to native and foreign conditions (54). These assessments will be critical to understand the commonality of horizontal symbiont transmission in the marine environment (2), given the hypothesis that local adaptation is less common in marine than terrestrial systems due to higher levels of gene flow (55).

Materials and Methods

Sample Collection, Nucleic Acid Extraction, and Sequencing. Samples of *Alviniconcha* and *Ifremeria* were collected from six vent sites (1,164 to 2,722 m) of the Lau Basin and Tonga Volcanic Arc in 2009 and 2016 using remotely operated vehicles (Fig. 1 and Dataset S1). Upon recovery, animal samples were dissected, placed in RNALater Stabilization Solution (Thermo Fisher Scientific), and frozen at −80 °C until further analysis. DNA was extracted with the Quick-DNA 96 Plus Extraction Kit (Zymo Research) and further purified with the MO BIO PowerClean DNA Pro Clean-Up Kit (Qiagen).

Host Transcriptome Sequencing and Assembly. Illumina RNA-sequencing reads for host transcriptome assemblies were obtained from sequencing experiments performed in refs. 28 and 56. Adapter clipping, quality trimming, and removal of rRNA and symbiont reads were performed as in ref. 28. Cleaned reads were error-corrected with Rcorrector (57) and filtered for uncorrectable and overrepresented sequences with the TranscriptomeAssemblyTools package (https://github.com/harvardinformatics/TranscriptomeAssemblyTools). Host transcriptome coassemblies for each *Alviniconcha* and *Ifremeria* species were performed with Trinity (58) with the PasaFly algorithm using 12,487,748 to 559,864,144 total reads per species. For each *Alviniconcha* species, additional transcripts were reconstructed from 6,434 to 46,731 454 reads obtained from ref. 59. Assembled contigs for each species were clustered with Cd-Hit-Est (60) at a 95% identity threshold to reduce transcript redundancies. Open reading frames (ORFs) were predicted with TransDecoder (https://github.com/TransDecoder/TransDecoder) considering homologies to known proteins (UniRef90) and protein domains (Pfam) as ORF retention criteria. Transcripts that did not contain any ORF or had a non-eukaryotic origin based on taxonomy classifications with BlobTools (61) were removed from the assembly. Final transcriptome assemblies were evaluated for quality and completeness with Busco (62).

Symbiont Metagenomic Sequencing and Pangenome Generation. One hundred and ninety-two barcoded high-throughput DNA-sequencing libraries were prepared with a Tn5 transposase-based protocol after ref. 63 at the University of California, Santa Cruz and then sent for 150-bp paired-end sequencing on a NovaSeq 6000 instrument at the University of California, Davis. However, due to low read allocation, 66 of the libraries were excluded from further analysis. Raw reads were trimmed with Trimmomatic (64) and filtered for sequence contaminants through mapping against the human (GRCh38) and PhiX reference genomes. An average of 31,747,773 (2,015,718 to 882,674,692) filtered Illumina reads for each host individual were then assembled with metaSPAdes (65), using *k*-mers from 21 to 121 in 10-step increments, and independently binned with MetaBAT 2 (66), MaxBin 2 (67), and GraphBin (68). Binning results were evaluated with DasTool (69). Taxonomic identity of the top-scoring symbiont metagenome-assembled genomes (MAGs) was assessed with GTDB-Tk (70) and assembly quality was checked with CheckM (71) based on Gammaproteobacteria- or Campylobacteria-specific marker genes. High- to moderate-quality MAGs with >85% completeness and <10% contamination were chosen for further analysis (Dataset S6). Additional reassemblies of previously published draft symbiont genomes (29) were constructed from a combination of Illumina and Nanopore reads (*SI Appendix, Methods*). Average nucleotide identities between MAGs were calculated with fastANI (72) to determine species boundaries at the 95% threshold (73). All MAGs were annotated with Prokka (74) and combined into species-specific symbiont pangenomes with Panaroo (75), using default parameters for gene clustering and a gene-refinding step to locate genes that were missed by gene prediction software within a radius of 1,000 nt. Genes without functional annotation were characterized through comparisons against the UniRef90 database with BlastP (76) at an *e*-value threshold of 1e-10. As the *A. boucheti* symbiont from Niua South represented a distinct *Sulfurimonas* species from that of the ELSC, we constructed an additional genus-specific pangenome, using a sequence identity threshold of 0.9 and a protein family sequence identity threshold of 0.5.

Phylogenomic Analyses of Host Mitochondria and Symbionts. To evaluate the presence of vertical transmission in all host species, we reconstructed phylogenies for host mitochondrial and symbiont core genomes and determined the concordance between tree topologies. Host mitochondrial genomes were assembled with MITObim (77) and annotated with GeSeq (78), using the *A. boucheti* (NC_049893.1) and *I. nautilei* (NC_024642.1) reference mitogenomes as bait for mitochondrial reconstruction and functional characterization. Concatenated alignments of 15 mitochondrial (13 protein-coding and 2 rRNA genes) and 222 to 1,679 symbiont-specific core genes were produced with Mafft (79) in Geneious Prime (https://www.geneious.com) and Panaroo, respectively. Phylogenomic trees were calculated with IQ-Tree 2 (80) in 10 independent runs using a gene-wise best-fit partition model selected with ModelFinder based on a greedy or relaxed hierarchical clustering strategy. Branch support was assessed through the ultrafast bootstrap approach and the Shimodaira–Hasegawa–like approximate-likelihood ratio test with 5,000 replicates by resampling both partitions and then sites within resampled partitions. To reduce the impact of model violations, each bootstrap tree was optimized using a hill-climbing nearest-neighbor interchange search (option –bnni). The best maximum-likelihood trees were midpoint-rooted with Phangorn (81) and cladograms were plotted with GGtree (82) in R.

Population-Level Variant Identification in Hosts and Symbionts. Filtered metagenomic reads from each host individual were mapped against the corresponding host transcriptome and symbiont pangenome with Bowtie 2 (83) in very sensitive mode. Optical duplicates were removed with Picard's MarkDuplicates tool (https://github.com/broadinstitute/picard). To resolve common alignment errors and improve base call accuracy, we locally realigned reads around indels and recalibrated base quality scores with LoFreq (84).

Host population genomic variation was assessed in Angsd (85) by inferring genotype likelihoods based on Hardy–Weinberg equilibrium considering individual inbreeding coefficients. To increase accuracy of the analyses, variant sites with mapping qualities <30 (minMapQ = 30), base qualities <20 (minQ = 20), and minimum minor-allele frequencies <0.01 (minMaf = 0.01) were excluded. We further filtered sites based on strand bias (sb_pval = 0.05), heterozygote bias (hetbias_pval = 0.05), and probability of being variable (SNP_pval = 1e-6). In addition, we removed spurious and improperly paired reads, adjusted mapping qualities for excessive mismatches (C = 50), and computed per-base alignment qualities (BAQ = 1) to disregard variants close to indel regions. Putative paralogous variants were excluded by discarding reads with multiple mappings and by considering only sites that had a maximum depth of 40 to 80 (based on distributions of average coverage). Genetic distances between individuals were inferred by calculating pairwise covariance matrices.

Symbiont population genomic variation was determined with FreeBayes (86) using input parameters adjusted for the analysis of metagenomic data (-F 0.01 -C 1 -p 1 –pooled-continuous –haplotype-length 0 –report-monomorphic). Variant

calls were restricted to sites with a minimum base quality of 20, a minimum mapping quality of 30, and a minimum coverage of 10. To eliminate bias in variant identification and other downstream analyses due to uneven read depth between samples, we normalized all samples to the lowest amount of coverage found in a sample for a particular symbiont phylotype (>10× coverage). Variants were further filtered based on strand bias (SRP > 5 && SAP > 5 && EPP > 5), proximity to indels (5 bp), and maximum mean depth with BcfTools (87) and VcfTools (88). In addition, sites and individuals with more than 25% missing data were excluded from the analysis. Allele counts (=symbiont strain abundances) and consensus haplotypes (=dominant symbiont strains) were extracted with GATK's VariantsToTable tool (89).

Population Genomic Structure and Differentiation. We performed ordination analyses with the Ape and Stats packages in R (90, 91) to assess genetic variation between symbiont and host populations. Host genetic structure was determined through principal-component analyses based on genetic covariance matrices, while symbiont genetic structure was inferred through principal-coordinate analyses based on both consensus haplotype and allele count data transformed into Euclidean distances and Bray–Curtis dissimilarities, respectively. For each sample, absolute allele counts were normalized to relative counts prior to analysis. Negative eigenvalues were corrected using the method by Cailliez (92) and final ordination plots were produced with GGplot2 (93). F_{ST} values between host and symbiont populations were calculated in Angsd and SciKit-Allel (https://github.com/cggh/scikit-allel), respectively, following the procedure in ref. 94. For the host species, potential outlier loci under selective pressures were inferred with OutFlank (95) based on neutral F_{ST} distributions that were obtained from quasi-independent SNP subsets determined with Plink (96).

Assessment of Symbiont Variation Based on Environment and Host Genetics. We conducted traditional and partial Mantel tests with the NCF package in R (97) to assess relationships between host identity-by-state distances, symbiont genetic distances, and geography based on Spearman rank correlations. Geographic distances were determined by calculating the geodesics between vent sites with the GeoSphere package (98). We used redundancy analyses following the approach in ref. 99 to evaluate the influence of hydrothermal fluid composition and depth on the genetic structure of each symbiont species. We further included the effect of sampling year, which encompasses changes in hydrothermal circulation within the ELSC as indicated by the cessation in fluid flow at the Kilo Moana vent field between 2009 and 2016. Endmember concentrations for geochemical compounds were obtained from the literature (25, 100, 101) or data provided by A. Diehl, MARUM, Bremen, Germany, and J. Seewald, Woods Hole Oceanographic Institution, Falmouth, MA (Dataset S1). Due to multicollinearity among the chemical species, we used the first eigenvector from principal-coordinate analyses (corresponding to the eigenvalue with the largest explanatory power) as the composite value. For each symbiont species, we further assessed the strength of correlation with other predictors to exclude variables that were highly collinear. Unless geographic location was strongly linked to environmental predictors, we used latitude as a conditioning factor in the analyses to correct for isolation by distance, given that vent sites are separated along a latitudinal range. Based on these collinearity evaluations, we tested the effect of all factors on the *Ifremeria* SOX symbiont, the effect of fluid composition on the *A. boucheti* Epsilon symbiont, and the effect of fluid composition and year (uncorrected for geography) on the *A. kojimai* and *A. strummeri* Gamma1 symbionts. As the Gamma1 symbiont was present in two host species, we ran an additional model including host as explanatory variable. SNPs were considered candidates for local adaptation if their loadings on significant constrained RDA axes deviated more than 2.5 SDs from the mean of the distribution. For each symbiont species, we further calculated the ratio of nonsynonymous polymorphisms per nonsynonymous site to synonymous polymorphisms per synonymous site (pN/pS) to infer candidate genes evolving under natural selection. The number of observed nonsynonymous and synonymous polymorphisms was determined with SNPeff (102), while the number of expected mutations was assessed with a modified Python script from ref. 103.

Assessment of Symbiont Gene Content Variation. We used PanPhlAn (104) to investigate potential differences in gene composition between symbiont populations from different vent localities. Downsampled symbiont reads for each species were mapped against the corresponding symbiont pangenome using custom annotations for functional categorizations. As exact zeros are not an appropriate threshold for considering a gene absent due to the possibility of mapping artifacts, PanPhlAn calculates the expected strain abundance by assessing the plateau in the coverage curves. Genes that are called absent through this method therefore include alignment errors but also genes that might be present in strains with very low abundance. Samples were profiled for gene presence/absence based on the following sensitive parameter thresholds: –min_coverage 1 –left_max 1.70 –right_min 0.30. Only samples that passed the retention thresholds in the variant-calling analyses were included. Gene content variation between symbiont strains was assessed by identifying genes that were present/absent in about 90% of samples from one or the other geographic region. For populations with small sample sizes (<5), we reduced this threshold down to 80% to avoid being overly stringent in determining differential gene preservation between locations. Presence/absence heatmaps were produced with the ComplexHeatmap package in R (105).

Data Availability. The raw Nanopore and Illumina reads, host transcriptomes, and symbiont reference genomes reported in this article have been deposited in the National Center for Biotechnology Information under BioProjects PRJNA523619, PRJNA526236, and PRJNA741492. Scripts for bioinformatic analyses are available on GitHub, https://github.com/cbreusing/Provannid_host-symbiont_popgen. All other study data are included in the article and/or supporting information.

ACKNOWLEDGMENTS. We thank the captains, crews, and remotely operated vehicle (ROV) pilots of the research vessel (R/V) *Thomas G. Thompson* (ROV *Jason II*) and R/V *Falkor* (ROV *Ropos*) for supporting the sample collections that made this study possible. We thank the Kingdom of Tonga for allowing access to their national waters, Peter Girguis for his contribution of the 2009 samples to this project, Alexander Diehl and Jeff Seewald for providing geochemical data, Michelle Hauer and Erin Frates for their assistance with sample preparation, and the NSF Established Program to Stimulate Competitive Research Cooperative Agreement OIA-1655221 for providing access to Brown University's high-performance computing cluster, where the bioinformatic analyses were performed. We further thank the technical staff at the University of California, Davis Genome Center for sequencing our Illumina metagenomic libraries. This work was supported by the NSF (Grants OCE-1536331, 1819530, and 1736932 to R.A.B.) and NIH (Grants 5K99GM135583-02 to S.L.R. and 5R35GM128932-03 to R.B.C.-D.).

Author affiliations: [a]Graduate School of Oceanography, University of Rhode Island, Narragansett, RI 02882; [b]Jack Baskin School of Engineering, University of California, Santa Cruz, CA 95064; and [c]Department of Molecular, Cell, and Developmental Biology, University of California, Santa Cruz, CA 95064

1. R. M. Fisher, L. M. Henry, C. K. Cornwallis, E. T. Kiers, S. A. West, The evolution of host-symbiont dependence. *Nat. Commun.* **8**, 15973 (2017).
2. S. L. Russell, Transmission mode is associated with environment type and taxa across bacteria-eukaryote symbioses: A systematic review and meta-analysis. *FEMS Microbiol. Lett.* **366**, fnz013 (2019).
3. R. C. Vrijenhoek, "Genetics and evolution of deep-sea chemosynthetic bacteria and their invertebrate hosts" in *The Vent and Seep Biota: Aspects from Microbes to Ecosystems*, S. Kiel, Ed. (Springer, Dordrecht, The Netherlands, 2010), pp. 15–49.
4. M. Bright, S. Bulgheresi, A complex journey: Transmission of microbial symbionts. *Nat. Rev. Microbiol.* **8**, 218–230 (2010).
5. G. Chomicki, E. T. Kiers, S. S. Renner, The evolution of mutualistic dependence. *Annu. Rev. Ecol. Evol. Syst.* **51**, 409–432 (2020).
6. D. R. Ginete, H. Goodrich-Blair, From binary model systems to the human microbiome: Factors that drive strain specificity in host-symbiont associations. *Front. Ecol. Evol.* **9**, 614197 (2021).
7. S. C. Verma, T. Miyashiro, Quorum sensing in the squid-*Vibrio* symbiosis. *Int. J. Mol. Sci.* **14**, 16386–16401 (2013).
8. S. L. Jackrel, J. W. Yang, K. C. Schmidt, V. J. Denef, Host specificity of microbiome assembly and its fitness effects in phytoplankton. *ISME J.* **15**, 774–788 (2021).
9. Q. Wang, J. Liu, H. Zhu, Genetic and molecular mechanisms underlying symbiotic specificity in legume-*Rhizobium* interactions. *Front. Plant Sci.* **9**, 313 (2018).
10. B. B. Allito, N. Ewusi-Mensah, V. Logah, Legume-*Rhizobium* strain specificity enhances nutrition and nitrogen fixation in faba bean (*Vicia faba* L.). *Agronomy (Basel)* **10**, 826 (2020).
11. L. Walker, B. Lagunas, M. L. Gifford, Determinants of host range specificity in legume-rhizobia symbiosis. *Front. Microbiol.* **11**, 585749 (2020).
12. T. C. LaJeunesse *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.* **37**, 785–800 (2010).
13. D. C. Smith, A. E. Douglas, *The Biology of Symbiosis* (Edward Arnold, 1987).

14. A. Hilário *et al.*, New perspectives on the ecology and evolution of siboglinid tubeworms. *PLoS One* **6**, e16309 (2011).
15. D. J. Thornhill, E. J. Howells, D. C. Wham, T. D. Steury, S. R. Santos, Population genetics of reef coral endosymbionts (*Symbiodinium*, Dinophyceae). *Mol. Ecol.* **26**, 2640–2659 (2017).
16. R. C. Vrijenhoek, M. Duhaime, W. J. Jones, Subtype variation among bacterial endosymbionts of tubeworms (Annelida: Siboglinidae) from the Gulf of California. *Biol. Bull.* **212**, 180–184 (2007).
17. D. T. Pettay, T. C. Lajeunesse, Long-range dispersal and high-latitude environments influence the population structure of a "stress-tolerant" dinoflagellate endosymbiont. *PLoS One* **8**, e79208 (2013).
18. P.-T. Ho *et al.*, Geographical structure of endosymbiotic bacteria hosted by *Bathymodiolus* mussels at eastern Pacific hydrothermal vents. *BMC Evol. Biol.* **17**, 121 (2017).
19. S. W. Davies, K. N. Moreland, D. C. Wham, M. R. Kanke, M. V. Matz, *Cladocopium* community divergence in two *Acropora* coral hosts across multiple spatial scales. *Mol. Ecol.* **29**, 4559–4572 (2020).
20. M. Ücker *et al.*, Deep-sea mussels from a hybrid zone on the Mid-Atlantic Ridge host genetically indistinguishable symbionts. *ISME J.* **15**, 3076–3083 (2021).
21. N. Dubilier, C. Bergin, C. Lott, Symbiotic diversity in marine animals: The art of harnessing chemosynthesis. *Nat. Rev. Microbiol.* **6**, 725–740 (2008).
22. Y. Suzuki *et al.*, Host-symbiont relationships in hydrothermal vent gastropods of the genus *Alviniconcha* from the southwest Pacific. *Appl. Environ. Microbiol.* **72**, 1388–1393 (2006).
23. A. C. Hartmann, A. H. Baird, N. Knowlton, D. Huang, The paradox of environmental symbiont acquisition in obligate mutualisms. *Curr. Biol.* **27**, 3711–3716.e3 (2017).
24. K. C. Reynolds *et al.*, New molluscan larval form: Brooding and development in a hydrothermal vent gastropod, *Ifremeria nautilei* (Provannidae). *Biol. Bull.* **219**, 7–11 (2010).
25. R. A. Beinart *et al.*, Evidence for the role of endosymbionts in regional-scale habitat partitioning by hydrothermal vent symbioses. *Proc. Natl. Acad. Sci. U.S.A.* **109**, E3241–E3250 (2012).
26. R. A. Beinart, A. Gartman, J. G. Sanders, G. W. Luther, P. R. Girguis, The uptake and excretion of partially oxidized sulfur expands the repertoire of energy resources metabolized by hydrothermal vent symbioses. *Proc. Biol. Sci.* **282**, 20142811 (2015).
27. C. Breusing *et al.*, Allopatric and sympatric drivers of speciation in *Alviniconcha* hydrothermal vent snails. *Mol. Biol. Evol.* **37**, 3469–3484 (2020).
28. C. Breusing *et al.*, Physiological dynamics of chemosynthetic symbionts in hydrothermal vent snails. *ISME J.* **14**, 2568–2579 (2020).
29. R. A. Beinart, C. Luo, K. T. Konstantinidis, F. J. Stewart, P. R. Girguis, The bacterial symbionts of closely related hydrothermal vent snails with distinct geochemical habitats show broad similarity in chemoautotrophic gene content. *Front. Microbiol.* **10**, 1818 (2019).
30. E. Podowski, S. Ma, G. Luther, D. Wardrop, C. Fisher, 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.* **418**, 25–45 (2010).
31. M. Tivey, Generation of seafloor hydrothermal vent fluids and associated mineral deposits. *Oceanography* **20**, 50–65 (2007).
32. G. J. Dick *et al.*, The microbiology of deep-sea hydrothermal vent plumes: Ecological and biogeographic linkages to seafloor and water column habitats. *Front. Microbiol.* **4**, 124 (2013).
33. G. J. Dick, The microbiomes of deep-sea hydrothermal vents: Distributed globally, shaped locally. *Nat. Rev. Microbiol.* **17**, 271–283 (2019).
34. J. Klose *et al.*, Endosymbionts escape dead hydrothermal vent tubeworms to enrich the free-living population. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 11300–11305 (2015).
35. S. V. Nyholm, M. J. McFall-Ngai, The winnowing: Establishing the squid-*Vibrio* symbiosis. *Nat. Rev. Microbiol.* **2**, 632–642 (2004).
36. M. Andrews, M. E. Andrews, Specificity in legume-rhizobia symbioses. *Int. J. Mol. Sci.* **18**, 705 (2017).
37. J. Lin, S. Huang, Q. Zhang, Outer membrane proteins: Key players for bacterial adaptation in host niches. *Microbes Infect.* **4**, 325–331 (2002).
38. B. Faure, S. W. Schaeffer, C. R. Fisher, Species distribution and population connectivity of deep-sea mussels at hydrocarbon seeps in the Gulf of Mexico. *PLoS One* **10**, e0118460 (2015).
39. Y. Won, C. R. Young, R. A. Lutz, R. C. Vrijenhoek, Dispersal barriers and isolation among deep-sea mussel populations (Mytilidae: *Bathymodiolus*) from eastern Pacific hydrothermal vents. *Mol. Ecol.* **12**, 169–184 (2003).
40. C. Dahl, B. Franz, D. Hensen, A. Kesselheim, R. Zigann, Sulfite oxidation in the purple sulfur bacterium *Allochromatium vinosum*: Identification of SoeABC as a major player and relevance of SoxYZ in the process. *Microbiology (Reading)* **159**, 2626–2638 (2013).
41. J. M. Rinta-Kanto *et al.*, Analysis of sulfur-related transcription by *Roseobacter* communities using a taxon-specific functional gene microarray. *Environ. Microbiol.* **13**, 453–467 (2011).
42. S. Lehmann, A. W. Johnston, A. R. Curson, J. D. Todd, A. M. Cook, "SoeABC, a novel sulfite dehydrogenase in Roseobacters" in *EMBO Workshop on Microbial Sulfur Metabolism*, G. Muyzer, A. J. Stams, Eds. (Sieca Repro, 2012), vol. **1518**, p. 29.
43. S. Lenk *et al.*, *Roseobacter* clade bacteria are abundant in coastal sediments and encode a novel combination of sulfur oxidation genes. *ISME J.* **6**, 2178–2187 (2012).
44. A. Warén, P. Bouchet, New records, species, genera, and a new family of gastropods from hydrothermal vents and hydrocarbon seeps. *Zool. Scr.* **22**, 1–90 (1993).
45. M. Perez, S. K. Juniper, Insights into symbiont population structure among three vestimentiferan tubeworm host species at eastern Pacific spreading centers. *Appl. Environ. Microbiol.* **82**, 5197–5205 (2016).
46. E. Douville *et al.*, The rainbow vent fluids (36°14′N, MAR): The influence of ultramafic rocks and phase separation on trace metal content in Mid-Atlantic Ridge hydrothermal fluids. *Chem. Geol.* **184**, 37–48 (2002).
47. T. Ikuta *et al.*, Heterogeneous composition of key metabolic gene clusters in a vent mussel symbiont population. *ISME J.* **10**, 990–1001 (2016).
48. R. Ansorge *et al.*, Functional diversity enables multiple symbiont strains to coexist in deep-sea mussels. *Nat. Microbiol.* **4**, 2487–2497 (2019).
49. M.-A. Cambon-Bonavita, J. Aubé, V. Cueff-Gauchard, J. Reveillaud, Niche partitioning in the *Rimicaris exoculata* holobiont: The case of the first symbiotic Zetaproteobacteria. *Microbiome* **9**, 87 (2021).
50. J. Zimmermann *et al.*, Dual symbiosis with co-occurring sulfur-oxidizing symbionts in vestimentiferan tubeworms from a Mediterranean hydrothermal vent. *Environ. Microbiol.* **16**, 3638–3656 (2014).
51. G. W. Luther III *et al.*, Chemistry, temperature, and faunal distributions at diffuse-flow hydrothermal vents: Comparison of two geologically distinct ridge systems. *Oceanography* **25**, 234–245 (2012).
52. S. Mitarai, H. Watanabe, Y. Nakajima, A. F. Shchepetkin, J. C. McWilliams, Quantifying dispersal from hydrothermal vent fields in the western Pacific Ocean. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 2976–2981 (2016).
53. L. A. Levin, D. J. Amon, H. Lily, Challenges to the sustainability of deep-seabed mining. *Nat. Sustain.* **3**, 784–794 (2020).
54. F. Blanquart, O. Kaltz, S. L. Nuismer, S. Gandon, A practical guide to measuring local adaptation. *Ecol. Lett.* **16**, 1195–1205 (2013).
55. L. M. Schiebelhut, M. N. Dawson, Correlates of population genetic differentiation in marine and terrestrial environments. *J. Biogeogr.* **45**, 2427–2441 (2018).
56. S. L. Seston *et al.*, Metatranscriptional response of chemoautotrophic *Ifremeria nautilei* endosymbionts to differing sulfur regimes. *Front. Microbiol.* **7**, 1074 (2016).
57. L. Song, L. Florea, Rcorrector: Efficient and accurate error correction for Illumina RNA-seq reads. *Gigascience* **4**, 48 (2015).
58. M. G. Grabherr *et al.*, Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
59. J. G. Sanders, R. A. Beinart, F. J. Stewart, E. F. Delong, P. R. Girguis, Metatranscriptomics reveal differences in in situ energy and nitrogen metabolism among hydrothermal vent snail symbionts. *ISME J.* **7**, 1556–1567 (2013).
60. W. Li, A. Godzik, Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
61. D. R. Laetsch, M. L. Blaxter, BlobTools: Interrogation of genome assemblies. *F1000 Res.* **6**, 1287 (2017).
62. M. Seppey, M. Manni, E. M. Zdobnov, "BUSCO: Assessing genome assembly and annotation completeness" in *Gene Prediction*, M. Kollmar, Ed. (Springer, 2019), pp. 227–245.
63. S. Picelli *et al.*, Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014).
64. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
65. S. Nurk, D. Meleshko, A. Korobeynikov, P. A. Pevzner, metaSPAdes: A new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).
66. D. D. Kang *et al.*, MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).
67. Y.-W. Wu, B. A. Simmons, S. W. Singer, MaxBin 2.0: An automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605–607 (2016).
68. V. Mallawaarachchi, A. Wickramarachchi, Y. Lin, GraphBin: Refined binning of metagenomic contigs using assembly graphs. *Bioinformatics* **36**, 3307–3313 (2020).
69. C. M. K. Sieber *et al.*, Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat. Microbiol.* **3**, 836–843 (2018).
70. P. A. Chaumeil, A. J. Mussig, P. Hugenholtz, D. H. Parks, GTDB-Tk: A toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925–1927 (2019).
71. D. H. Parks, M. Imelfort, C. T. Skennerton, P. Hugenholtz, G. W. Tyson, CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
72. C. Jain, L. M. Rodriguez-R, A. M. Phillippy, K. T. Konstantinidis, S. Aluru, High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **9**, 5114 (2018).
73. K. T. Konstantinidis, J. M. Tiedje, Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 2567–2572 (2005).
74. T. Seemann, Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
75. G. Tonkin-Hill *et al.*, Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol.* **21**, 180 (2020).
76. C. Camacho *et al.*, BLAST+: Architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
77. C. Hahn, L. Bachmann, B. Chevreux, Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads–A baiting and iterative mapping approach. *Nucleic Acids Res.* **41**, e129 (2013).
78. M. Tillich *et al.*, GeSeq–Versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* **45**, W6–W11 (2017).
79. K. Katoh, K. Misawa, K. Kuma, T. Miyata, MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
80. B. Q. Minh *et al.*, IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
81. K. P. Schliep, phangorn: Phylogenetic analysis in R. *Bioinformatics* **27**, 592–593 (2011).
82. G. Yu, D. K. Smith, H. Zhu, Y. Guan, T. T. Y. Lam, ggtree: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).
83. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
84. A. Wilm *et al.*, LoFreq: A sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res.* **40**, 11189–11201 (2012).
85. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics* **15**, 356 (2014).
86. E. Garrison, G. Marth, Haplotype-based variant detection from short-read sequencing. arXiv [Preprint] (2012). https://arxiv.org/abs/1207.3907 (Accessed 11 August 2021).
87. H. Li *et al.;* 1000 Genome Project Data Processing Subgroup, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
88. P. Danecek *et al.;* 1000 Genomes Project Analysis Group, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
89. G. A. V. de Auwera, B. D. O'Connor, *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra* (O'Reilly, ed. 1, 2020).
90. E. Paradis, K. Schliep, ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
91. R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2020).
92. F. Cailliez, The analytical solution of the additive constant problem. *Psychometrika* **48**, 305–308 (1983).
93. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016).

94. G. Bhatia, N. Patterson, S. Sankararaman, A. L. Price, Estimating and interpreting F_{ST}: The impact of rare variants. *Genome Res.* **23**, 1514–1521 (2013).
95. M. C. Whitlock, K. E. Lotterhos, Reliable detection of loci responsible for local adaptation: Inference of a null model through trimming the distribution of F_{ST}. *Am. Nat.* **186**, S24–S36 (2015).
96. S. Purcell *et al.*, PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
97. O. N. Bjornstad, ncf: Spatial Covariance Functions (2020). https://CRAN.R-project.org/package=ncf. Accessed 28 January 2020.
98. R. J. Hijmans, geosphere: Spherical Trigonometry (2019). https://CRAN.R-project.org/package=geosphere. Accessed 26 May 2019.
99. B. R. Forester, J. R. Lasky, H. H. Wagner, D. L. Urban, Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. *Mol. Ecol.* **27**, 2215–2233 (2018).
100. M. J. Mottl *et al.*, Chemistry of hot springs along the Eastern Lau Spreading Center. *Geochim. Cosmochim. Acta* **75**, 1013–1038 (2011).
101. G. E. Flores *et al.*, Inter-field variability in the microbial communities of hydrothermal vent deposits from a back-arc basin. *Geobiology* **10**, 333–346 (2012).
102. P. Cingolani *et al.*, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w^{1118}; iso-2; iso-3. *Fly (Austin)* **6**, 80–92 (2012).
103. D. Romero Picazo *et al.*, Horizontally transmitted symbiont populations in deep-sea mussels are genetically isolated. *ISME J.* **13**, 2954–2968 (2019).
104. F. Beghini *et al.*, Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *eLife* **10**, e65088 (2021).
105. Z. Gu, R. Eils, M. Schlesner, Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).