Title: Population genomics of a bioluminescent symbiosis sheds light on symbiont transmission and specificity

A.L. Gould^{1,2*} and P.V. Dunlap²

¹ *Ichthyology Department, California Academy of Sciences, San Francisco, CA 94118*

² *Department of Ecology & Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109*

**Corresponding author:*

Alison Gould
California Academy of Sciences
55 Music Concourse Drive
San Francisco, CA 94118
(415) 350-5708
algould@umich.edu
ORCID iD: 0000-0002-2358-303X

Abstract

All organisms depend on symbiotic associations with bacteria for their success, yet the processes by which specific symbioses are established and persist remain largely undescribed. To examine the ecological mechanisms involved in maintaining symbiont specificity over host generations, we examined the population genomics of a binary symbiosis involving the coral reef cardinalfish *Siphamia tubifer* and the luminous bacterium *Photobacterium mandapamensis*. Using restriction site-associated sequencing (RAD-Seq) methods we demonstrate that the facultative symbiont of *S. tubifer* exhibits genetic structure at spatial scales of tens of kilometers in Okinawa, Japan in the absence of physical dispersal barriers and in contrast to the host fish. These results suggest the host's behavioral ecology help structure symbiont populations at a reef site by symbiont enrichment, consequently fostering symbiont specificity. This approach also revealed several symbiont genes that were divergent between host populations including genes known to play a role in other host-bacteria associations.

Introduction

In open marine environments, where there are few physical barriers to dispersal, microbes are expected to exist as well-mixed populations, supporting the classic panmictic view in microbial biogeography that “everything is everywhere, but the environment selects”¹. However, this view has been challenged in recent years,

providing evidence of geographic structure in certain marine microbes²⁻³, although often over broad geographic scales and with coarse taxonomic resolution⁴⁻⁷. In contrast, there are few studies that have quantitatively examined the spatial scale of intra-specific patterns of distribution in marine bacteria with sufficient fine-scale genetic resolution to differentiate populations⁸⁻¹⁰. We define the population genomic structure of a symbiotically luminous marine bacterium over a relatively small geographic region and reveal the important role a host animal's behavioral ecology can play symbiont transmission and in structuring populations of its microbial symbiont in a highly connected ocean environment.

The evolution of a symbiotic association requires that the specific partnership between host and symbiont remains stable over host generations. Such specificity is maintained in vertically transmitted symbioses via the intimate transfer of symbiont from parent to offspring and commonly leads to co-speciation¹¹⁻¹⁴. In contrast, the evolution of horizontally transmitted symbioses requires both the selection of a particular symbiont from the environmental pool and the maintenance of that relationship over time. Therefore, a lower level of host-symbiont specificity is expected for horizontally acquired symbioses than that of their vertical counterparts. Nevertheless, many such associations are highly specific¹⁵⁻¹⁷, indicating that mechanisms such as environmental filtering and the ecology or physiology of the host ensure that a certain symbiont is acquired from the environment by each new generation.

Bioluminescent symbioses, especially those involving fish hosts, display a higher degree of specificity than what would be expected due to random symbiont acquisition from the marine environment¹⁸⁻¹⁹. How such specificity is achieved and maintained

remains poorly understood, but the ecology of the host can play an important role²⁰⁻²². Unfortunately, little is known of the ecology of most symbiotically luminous fishes, especially with respect to symbiont acquisition, a process that is challenging to study in nature due to the planktonic larval state of most fishes and their often deep or open water habitats²³. The symbiotically luminous cardinalfish, *Siphamia tubifer*, inhabits shallow coral reefs in the Indo-Pacific and has known life history attributes²⁴ and behavior²⁵ that make it an ideal luminous fish species with which to investigate the host's role in symbiont acquisition and in maintaining specificity of the symbiosis.

The bioluminescent symbiosis between *S. tubifer* and the luminous bacterium, *Photobacterium mandapamensis*, is a mutualistic association in which the symbiont is provisioned in an abdominal light organ by the fish in exchange for light production by the bacteria. The host, which remains quiescent among the long spines of sea urchins during the daytime²⁵⁻²⁶, uses the bacterially produced light to illuminate its ventral surface while foraging at night, possibly for countershading against the moonlight or for attracting prey. The host's light organ is colonized by the luminous bacteria during larval development, although the exact timing in the wild remains unknown. In culture, *S. tubifer* larvae are not receptive to colonization by *P. mandapamensis* until at least eight days after having been released into the plankton²⁷. Therefore, the direct transfer of symbiont from parent to offspring is not possible, and acquisition of the symbiont from the environment by aposymbiotic larvae occurs sometime prior to settlement, as the light organs of all juvenile *S. tubifer* collected from a reef have an established symbiont population. Several aspects of the initiation of the symbiosis remain undefined, including

the timing and location of symbiont acquisition in the wild as well as the number of bacterial cells that initially colonize a light organ.

To gain a better understanding of symbiont acquisition for the *S. tubifer-P. mandapamensis* association we used a population genomics approach to quantify the within- and among- population genetic variability of the luminous bacteria from *S. tubifer* light organs collected from coral reefs in the Okinawa Islands, Japan. Specifically, we applied double digest restriction site-associated sequencing (ddRAD-Seq) methods to identify and analyze geographic patterns across hundreds of variant sites throughout the bacterial symbiont's genome to test the hypothesis that the luminous symbionts of *S. tubifer* within a reef site are more genetically similar to each other than they are to the symbionts at other reefs. We also tested for evidence of population structure in the host fish and compared the patterns of genetic variability of the host to that of their symbiotic bacteria. We then analyzed the temporal stability of the light organ symbiont population at a single reef site over three consecutive years. This study provides insights into the process of symbiont acquisition for this highly specific, bioluminescent association, and we discuss the results in the context of the role of the host's ecology in shaping the specificity of this horizontally transmitted symbiosis.

Results

Genetic admixture of the host fish

We carried out an analysis of the RAD-Seq data for a total of 282 *S. tubifer* specimens sampled from eleven locations that were separated by two to tens of kilometers (local scale) around Okinawa Island and up to 140 km (regional scale) between Okinawa Island and Kume Island. The host fish exhibited no evidence of genetic structure at either scale sampled in 2013 and 2014 (Figure 1) and all pairwise F_{ST} values were extremely low, ranging between 0.001 and 0.005 (SI Tables 1-3). These results confirm previous results that *S. tubifer* comprise one panmictic population in the Okinawa Islands likely due to larval dispersal²⁸ during their month long pelagic stage²⁴. Northward flow of the Kuroshio Current from the Philippine Islands promotes connectivity of marine populations in the region²⁹⁻³⁰, and a high frequency of typhoons and heavy predation on adult fish²⁴, presumably all contribute to the observed genetic admixture in the region.

Figure 1. Genomic analysis of the light organ symbionts of *Siphamia tubifer* reveals structure in populations of symbiotic *Photobacterium mandapamensis* despite a lack of structure in the host fish. (a) Samples were collected from ten locations in 2013 and from four locations in 2014 in the Okinawa Islands, Japan. Principal coordinates analysis of genetic differentiation of (b) symbiotic *P. mandapamensis* across 534 loci analyzed from light organs sampled in 2014, (c) the corresponding *S. tubifer* hosts across 8,637 SNPs, (d) symbiotic *P. mandapamensis* across 465 loci analyzed from light organs sampled in 2013, and (e) the corresponding *S. tubifer* hosts across 8 637 SNPs. Points represent individuals along the first and second axes of genetic variation. The first 20 axes of variation for each analysis are depicted in SI Figure 5. Ellipses indicate standard deviation for symbiont populations sampled at each location.

Population structure of the bacterial symbiont

Using a ddRAD-Seq approach, we identified a total of 601 bacterial loci for the light organ symbionts of *S. tubifer* across all 282 individuals with the restriction enzymes *EcoRI* and *MseI*. These loci were distributed approximately evenly throughout the *P. mandapamensis* genome (Figure 2a,b) and thus, provide a representative snapshot of the symbiont genome. Ordination analyses of the filtered subset of these loci (n=534) from fish sampled in 2014 revealed a signature of genetic structure for populations of the symbiotic bacteria in contrast to the host fish (Figure 1), and was confirmed to be significant by PERMANOVA (Table 1). This structure was particularly evident when comparing symbiont populations between Okinawa Island and Kume Island, located approximately 140 km to the West (Figure 1b), which had significant differences in genetic distance to the light organ symbionts sampled from all three other locations (SI Table 4). However, loci analyzed (n=465) from light organs sampled in 2013 at the more local geographic scale of two to tens of kilometers suggest that the symbiotic bacteria are more genetically homogeneous around Okinawa Island (Table 1), although

significant differences between populations in Okinawa were also detected (SI Table 5). Furthermore, no correlation in genetic distances was evident between the bacteria and their host fish for any of the three datasets (2013: $r = -0.38$, $p = 1.00$; 2014: $r = -0.01$, $p = 0.53$; Sesoko: $r = -0.12$, $p = 0.89$).

Loci correlated with symbiont population divergence

We identified putative loci driving the observed pattern of genetic divergence between populations of the symbiotic bacteria in 2014 and identified 24 outlier haplotypes from 22 loci. These haplotypes were highly correlated, either positively or negatively, with the primary axis of genetic variance differentiating the Kume Island and Okinawa Island populations (Figure 2c-e, SI Figure 4). All 22 loci were located in protein coding regions of the genome, and most of the variant effects examined for each of the outlier haplotypes were non-synonymous, including two that resulted in the addition of a stop codon (Table 2). Interestingly, some of these variants were in genes of known importance for host association in species of *Aliivibrio* and *Vibrio*, genera closely related to *Photobacterium*, including *hfq* (RNA-binding protein), *barA* (histidine kinase), and *gcvP* (glycine dehydrogenase), and could therefore, also be of importance for the bioluminescent association between *S. tubifer* and *P. mandapamensis*. Although the bacterial genes involved in the symbiosis have yet to be identified and little is known regarding the genetics of this association, except that homologs of non-*lux* genes involved in the bioluminescent symbiosis of *Aliivibrio fischeri* with the sepiolid squid, *Euprymna scolopes*, are absent in the *P. mandapamensis* genome³¹.

Figure 2. Restriction site-associated sequencing (RAD-Seq) yields a high density of genomic markers for the analysis of population structure and divergence of a symbiotic bacterium. (a) Annotated genome map of *Photobacterium mandapamensis*, the light organ symbiont of *Siphamia tubifer*. Protein-coding genes are color-coded according to which prokaryotic cluster of orthologous group (COG) they belong. Non-coding regions are white. Outer tic marks represent the locations of the 601 total loci analyzed for genomic differentiation between symbiont populations. The inner plot highlights the relative proportion of COG functions in which these loci are located. The 22 loci in which outlier haplotypes were identified as potential drivers of population differentiation are indicated on the inner circle of the map. (b) Distribution of the frequency of nucleotide distances between each pair of the 601 loci located in a scaffold of the *P. mandapamensis* genome. (c) Constrained analysis of principal coordinates (CAP) analysis of the 534 loci examined in 2014 (d) Corresponding CAP scores of the haplotypes analyzed across all 534 loci (e) Heatmap indicating the relative abundance within each population of the identified outlier haplotypes associated with each locus indicated.

Temporal stability of symbiont populations

To determine the extent to which light organ symbiont populations were stable over time, we also analyzed the genetic variance of bacteria from fish light organs collected from one location (Sesoko, Figure 1a) in three consecutive years (2012, 2013, and 2014). In contrast to the genetic differentiation observed between sampling locations in 2014, populations of the symbiotic bacteria were not significantly different between years at this location (Figure 3, Table 1). This lack of variation suggests that the bacterial populations are somewhat stable at a reef over time. The host fish also exhibited little genetic differentiation between years (Figure 3), however a previous analysis of outlier loci from this dataset revealed a strong signature of temporal differentiation likely due to variable larval supply from different upstream sources²⁸.

Figure 3. Temporal genomic analysis of the light organ symbionts of *Siphamia tubifer* reveals population stability at the same reef site over time. Light organs were sampled from reefs at Sesoko Island in Okinawa in the summer of 2012, 2013, and 2014. Principal coordinates analysis of genetic differentiation (a) of symbiotic *P. mandapamensis* across 552 loci and (b) of the corresponding *S. tubifer* hosts across 8,637 SNPs. Points represent individuals along the first and second axes of genetic variation. Ellipses indicate standard deviation for symbiont populations sampled each year.

Within-population symbiont strain diversity

In addition to defining the genomic diversity of symbiotic *P. mandapamensis* populations between reefs, we also examined symbiont diversity within individual light organs. To estimate the number of genetically distinct bacterial strains comprising each light organ's symbiont population, we determined the number of bacterial genotypes across all 601 loci within each host. However, one limitation of using RAD-Seq to analyze entire symbiont populations, is that haplotypes cannot be concatenated across loci. We therefore estimated the within light organ diversity to be the maximum number

of haplotypes observed among all bacterial loci within a host. Using this approach, an average of six (± 1.6 SD) distinct bacterial genotypes were detected within a light organ (Figure 4), although this could be an underestimate of the total strain diversity present due to our inability to concatenate haplotypes. Individual hosts had anywhere between two and ten symbiont types present in their light organ, and no fish harbored a population comprised entirely of a single symbiont genotype (Figure 4). Furthermore, there was no correlation between fish length, a proxy for fish age²⁴, and the minimum or maximum number of symbiont genotypes in a light organ ($F_{1,290} = 1.477$, $R^2_{Adj} = 0.00164$, $P = 0.225$) (Figure 4), indicating the number of strains within a light organ does not increase over time. These results suggest that the population of bacteria within a light organ is likely founded by an average of six genetically distinct strains of *P. mandapamensis*, which then persist within the light organ throughout the host's life.

Figure 4. An average of six distinct *Photobacterium mandapamensis* strains are present within a light organ of *Siphamia tubifer* (a) Frequency distribution of the minimum possible number of symbiotic *P. mandapamensis* genotypes present within a light organ as determined by the maximum number of haplotypes observed across all 601 symbiont loci analyzed. (b) The minimum number of distinct *P. mandapamensis* strains present within a light organ relative to the standard length of the host fish.

Discussion

We obtained data for discrete, live populations of bacterial symbionts by extracting and sequencing genomic DNA directly from whole light organs of *S. tubifer*, a symbiotically luminous fish commonly found on coral reefs in the Okinawa Islands. This approach together with RAD-Seq methods allowed us to obtain deeply sampled, fine-scale genomic data capable of distinguishing populations of this bacterium from various locations sampled in the region, and in the absence of obvious barriers to dispersal. Although RAD-Seq methods have become increasingly popular in recent years³²⁻³³, they have not previously been used to examine the genomic structure of natural populations of bacteria nor have they been used to simultaneously examine populations of a host animal and its bacterial symbiont population. By applying these methods, we were able to identify potential genes of interest that are driving the patterns of genetic divergence between symbiont populations. Most notably, we discovered missense variations in *barA*, and *hfq* that were completely absent in the two of the four populations examined (Table 2, Figure 2e). The two-component *barA/uvrY* homologs *varS/varA* in *Vibrio cholerae* have been recognized as important for virulence³⁴⁻³⁵, quorum sensing³⁶, and dissemination from a host to an aquatic environment³⁷, and *varS* deletion strains of *V. cholerae* have colonization deficiencies in mice³⁵. Similarly, the RNA-binding protein Hfq is essential for the virulence of *V. cholerae*, and strains lacking *hfq* also fail to colonize mice hosts³⁸. Interestingly, *hfq* has also been identified as a negative regulator of bioluminescence in *V. harveyi*³⁹. Additional variants of interest were observed in *gcvP*, which plays a role in the virulence of *V. cholerae* in *Drosophila*

hosts⁴⁰, and in *hpt*, which is located just upstream of the quorum-sensing regulating gene *luxR* in *V. harveyi*⁴¹. While the potential roles of these genes in the *S. tubifer*-*P. mandapamensis* symbiosis are unknown, our approach uncovered them as interesting targets for future studies of the underlying mechanisms regulating this bioluminescent association.

Despite the absence of physical barriers to dispersal and in contrast to the host fish, we discovered within-species level genetic differences between the luminous symbionts of *S. tubifer* sampled from different reefs in the Okinawa Islands. Furthermore, we determined that symbiont populations are seemingly stable at a reef between years. Environmental factors known to influence the biogeography of marine microbes include both temperature and salinity². However, variation in these factors between sampling locations were unlikely to influence populations of symbiotic *P. mandapamensis* in this study due to the relatively small geographic area examined and shallow collection depths (<10 m). Therefore, we suggest the observed patterns of genetic structure between light organ symbionts from reefs in Okinawa and Kume Islands result from the behavioral ecology of the host.

The *S. tubifer* light organ is connected by a duct to the host's intestine, and luminous bacterial cells are continually released via the duct from the light organ into the intestines and then the seawater with fecal waste⁴². The release of bacterial cells from the light organ presumably promotes the growth and light production of the symbiont population in the host while also enriching the surrounding seawater with symbiont cells^{22,43-45}. Once settled as juveniles, *S. tubifer* exhibit fidelity to a home reef and typically return to the same urchin at that reef after foraging each night²⁵. Therefore,

the *P. mandapamensis* genotypes present in the light organs of resident fish at a reef are regularly released into the surrounding seawater and can become enriched in the local environment. Thus, the observed genetic structure of *P. mandapamensis* symbionts is likely a consequence of symbiont acquisition by larval fish occurring in the locally symbiont-enriched seawater near their settlement site. As a result, *S. tubifer* hosts residing at a particular reef, although not closely related, share more similar symbiont genotypes with each other than with fish from other reefs. This host-mediated symbiont population structure is consistent with the lack of temporal structure observed for symbiont populations sampled at the same reef over three years and has been observed for other marine symbioses⁴⁶, including the bioluminescent symbiosis between *E. scolopes* and *A. fischeri*⁴⁷.

In a highly connected ocean environment, where the classic view of microbial biogeography suggests that “everything is everywhere”¹, we illustrate that fine-scale patterns of genomic population structure in a facultative bacterial symbiont can arise, even in a region dominated by strong ocean currents and in the absence of physical barriers to gene flow. These results provide new insight on the timing and location of symbiont acquisition by *S. tubifer* larvae in the wild, a process that is difficult to study for most horizontally acquired marine symbioses. We hypothesize that the distinct life history and behavioral ecology of *S. tubifer*²⁴⁻²⁵ results in the local enrichment of *P. mandapamensis* in the surrounding seawater, which consequently ensures that the next generation of host fish can successfully initiate its symbiosis with its species-specific luminous bacterial symbiont. This host facilitation presumably helps to foster the high degree of specificity observed for the partnership¹⁹ over time, and incidentally, can

promote genetic divergence between symbiont populations. Future applications of RAD-Seq methods with other symbiotic associations have the potential to broaden our understanding of the influence of host animals on the biogeographic patterns of their microbial symbionts and the evolutionary processes that lead to host-symbiont integration and specificity.

Materials and Methods

Sampling, DNA extraction and library preparation

Siphamia tubifer specimens were collected over three years from locations in Okinawa, Japan (Table 3). Ten collection sites on Okinawa Island were sampled during June and July of 2013, and in June of 2014, samples were collected from an additional site on Kume Island as well as three sites on Okinawa Island that were previously sampled in 2013 (Figure 1a). One reef site (Sesoko, “S”) that was sampled in 2013 and 2014 was also sampled in July of 2012, providing a three-year dataset from that location. All collection sites were less than 10 m depth. Approximately 20 fish ranging in body lengths and associated with different sea urchins were collected at each sampling location. Fish were immediately euthanized upon collection by immersion in seawater containing a lethal dose (500 mg/L) of buffered tricaine methanesulfonate (MS-222) as approved by the University of Michigan Institutional Animal Care and Use Committee.

The whole, intact light organ, composed of fish tissue and containing the

luminous bacterial symbiont population, was aseptically removed from each fish specimen and individually preserved in RNAlater®. Genomic DNA was extracted from each light organ using a QIAGEN DNeasy Blood and Tissue Kit following the manufacturer's protocol, and double digest restriction site-associated sequencing (ddRAD-Seq) DNA libraries were built from the total DNA extracted from *S. tubifer* light organs, each individually barcoded with unique 10 bp DNA sequences for downstream identification. The protocol used to construct the ddRAD-Seq libraries followed a modified combination of previously described methods⁴⁸⁻⁴⁹, using the restriction enzymes *MseI* and *EcoRI*, which were tested for their efficacy to produce a significant number of loci for the analysis of both the host fish and its symbiotic bacteria on an initial set of 50 light organs. Up to 50 DNA libraries were pooled per lane and sequenced at the Center for Applied Genomics, Toronto, ON, Canada, on the Illumina HiSeq2000 platform (San Diego, CA), generating 100 bp single-end sequence reads.

Sequence processing and data analysis

Raw sequence reads were de-multiplexed and quality-filtered for a Phred score of 33 or higher with the *process_radtags* command in *Stacks* v1.35⁵⁰⁻⁵¹. After the DNA sequence barcodes were removed, the remaining 90-bp sequence reads were aligned against the ~4.5 Mb reference genome of *Photobacterium mandapamensis*³¹ using the *very_sensitive* command in *Bowtie2*⁵² v2.2.0 to separate the host fish sequences from those of its luminous symbiont. The unaligned sequences were additionally filtered against the genomes of both the marine bacterium *Vibrio campbelli* or *Escherichia coli*

K12. All remaining unaligned sequences were classified as belonging to *S. tubifer*. Aligned *P. mandapamensis* sequences in .SAM file format were then additionally quality filtered using *SAMtools*⁵³ v1.3, retaining only reads with a MAPQ score greater than 20. The quality-filtered, aligned *P. mandapamensis* sequences were then processed with the *ref_map* command in *Stacks*, requiring a minimum stack depth of three (-m 3). Individuals with fewer than 100,000 total sequence reads or a mean depth of coverage per stack less than 100x were discarded from the analysis as they were considered to be outliers with respect to evenness of coverage across the individual sequence libraries (SI Figure 1). The final dataset consisted of sequence reads from 282 light organs collected from 11 locations over three years in the Okinawa Islands (Table 3).

To minimize the effects of missing data, only bacterial loci present in at least 70% of all light organs were extracted from the entire catalog of loci produced by *Stacks* and used for the downstream analysis. Subsequently, loci not present in at least 50% of individuals in each population were removed from the dataset, resulting in an initial dataset of 601 RADtags (loci) across all individuals. Additional missing data filters were also tested but did not have a notable affect on the analysis (SI Figure 2), thus we present the results for the missing data filters described above. After this initial filtering step, individuals sampled in 2013, 2014, and from Sesoko (S) over three consecutive years (2012-14) were subsequently grouped into sub-datasets to analyze independently for genomic structure. To further minimize any potential effects of missing data within each dataset, both individuals and loci with more than 15% missing data were removed. The numbers of remaining loci analyzed were 465, 534, and 552 for the 2013, 2014, and Sesoko datasets, respectively. Within each locus, haplotypes with at least 6x

coverage were then analyzed for patterns of genomic structure across light organ symbiont populations within each dataset. Additional haplotype depth filters (5x, 8x, and 10x) were also examined, but they also did not have a notable effect on the analysis (SI Figure 3). To visualize patterns of genetic variation in symbiont populations, principal coordinates analyses (PCoA) were performed on the Bray-Curtis dissimilarity matrices calculated from the presence or absence of these haplotypes across all light organs in each dataset. Significant differences in genetic differentiation were confirmed by a permutational multivariate analysis of variance (PERMANOVA) on the genetic dissimilarity matrices using the *adonis* function in the “vegan”⁵⁴ package in R⁵⁵ using sampling location (or year) as a factor. Pairwise *adonis* tests were then performed between all pairs of locations within the 2013 and 2014 datasets to determine which symbiont populations were significantly different from one another.

Patterns of genomic structure of the host fish were also analyzed and compared at the same scales as their light organ symbionts. *Siphamia tubifer* sequence reads were assembled *de novo* in *Stacks* and genotypes were assigned to each individual fish. Specifically, the first single nucleotide polymorphism (SNP) within each of the 8,637 loci present in 70% of individuals in each population with a minor allele frequency of 5% or greater were analyzed for each corresponding host fish. We selected these data filters based on previous RAD-Seq studies of other marine organisms⁵⁶⁻⁵⁷ in order to maximize the number of loci examined while minimizing the effects of missing data. Notably, the nonparametric analyses applied to these data tend to show little sensitivity to the minor allele filter⁵⁸. We then performed principal coordinates analyses on the Euclidean genetic distances between individuals in each of the *S. tubifer* SNP datasets

(2013, 2014, and Sesoko) using the “adegenet”⁵⁹ package in R. Weir and Cockerham’s pairwise F_{ST} values⁶⁰ between populations were calculated with 1,000 permutations in GenoDive, and corresponding p-values were adjusted for multiple comparisons using the Bonferroni method. To test for significant differences in genetic differentiation between sampling locations (or year) additional PERMANOVAs were performed on the genetic dissimilarity matrices as described above for the symbiotic bacteria. Mantel tests with 1,000 permutations were also performed on the dissimilarity matrices calculated for each dataset (2013, 2014, and Sesoko) to test for correlations between the genetic distances of the host fish to that of its light organ symbionts with the “vegan” package⁵⁴ in R.

To identify loci driving the patterns of genetic divergence between light organ symbionts at Kume Island and Okinawa Island in 2014, a constrained analysis of principal coordinates (CAP) using the *capscale* function in the “vegan” package⁵⁴ in R was performed on the 2014 *P. mandapamensis* genetic dissimilarity matrix. Haplotypes with the highest (>99%) and lowest (<1%) scores along first two axes of variation were identified and of these, 24 outlier haplotypes were further selected as potential outliers driving the observed genetic divergence between symbiont populations based on their large differences in CAP scores relative to all other haplotypes (SI Figure 4, Figure 2c-e). For each of these outlier haplotypes, the associated locus and gene function were identified, and their variant type and functional effect were determined using the program *SnpEff*⁶¹.

Acknowledgements

We thank Dr. Saki Harii and the staff of Sesoko Station, University of the Ryukyus for logistical assistance as well as Katherine Dougan and Maggie Grundler, University of Michigan (UM), for technical and field assistance. We gratefully acknowledge members of Dr. Lacey Knowles' Lab (UM) for guidance and assistance with the RAD-Seq protocols and data analysis. We also thank Raquel Rivadeneira and the UM Genomic Diversity Lab for technical services and the Center for Applied Genomics for their sequencing services. Funding for this project was provided by National Science Foundation grant DEB-1405286.

Competing Interests

The authors declare no conflict of interest.

References

1. Baas-Becking LG. Geobiologie; of inleiding tot de milieukunde. Van Stockum WP, Zoon NV; 1934.
2. Martiny JB, Bohannan BJ, Brown JH, Colwell RK, Fuhrman JA, Green JL, Horner-Devine MC, Kane M, Krumins JA, Kuske CR, Morin PJ. Microbial biogeography: putting microorganisms on the map. Nat Rev Microbiol. 2006; 4: 102-112.
3. Ivars-Martínez, E. et al. Biogeography of the ubiquitous marine bacterium *Alteromonas macleodii* determined by multilocus sequence analysis. Mol Ecol. 2008; 17: 4092-4106.
4. Pommier T, Pinhassi J, Hagström, Å. Biogeographic analysis of ribosomal RNA clusters from marine bacterioplankton. Aquat Micro Ecol. 2005; 41: 79-89.
5. Pommier T, Canbäck B, Riemann L, Boström KH, Simu K, Lundberg P, Tunlid A, Hagström Å. Global patterns of diversity and community structure in marine bacterioplankton. Molc Ecol. 2007; 16: 867-80.
6. Brown MV, Lauro FM, DeMaere MZ, Muir L, Wilkins D, Thomas T, Riddle MJ, Fuhrman JA, Andrews-Pfannkoch C, Hoffman JM, McQuaid JB. Global biogeography of SAR11 marine bacteria. Molc Systems Biol. 2012; 8: 595.
7. Sul WJ, Oliver TA, Ducklow HW, Amaral-Zettler LA, Sogin ML. Marine bacteria exhibit a bipolar distribution. PNAS. 2013; 110: 2342-2347.
8. Acinas SG, Klepac-Ceraj V, Hunt DE, Pharino C, Ceraj I, Distel DL, Polz MF. Fine-scale phylogenetic architecture of a complex bacterial community. Nature. 2004; 430: 551-554.
9. Hunt, D. E., David, L. A., Gevers, D., Preheim, S. P., Alm, E. J. & Polz, M. F. Resource partitioning and sympatric differentiation among closely related bacterioplankton. Science. 2008; 320: 1081-1085.
10. DeChaine, E. G., Bates, A. E., Shank, T. M., & Cavanaugh, C. M. Off-axis symbiosis found: characterization and biogeography of bacterial symbionts of Bathymodiolus mussels from Lost City hydrothermal vents. Environ Microbiol. 2006; 8: 1902-1912.
11. Chen X, Li S, Aksoy S. Concordant evolution of a symbiont with its host insect species: Molecular phylogeny of genus *Glossina* and its bacteriome-associated endosymbiont, *Wigglesworthia glossinidia*. J Mol Evol. 1999; 48: 49-58.
12. Clark MA, NA Moran, P Baumann, Wernegreen JJ. Cospeciation between bacterial endosymbionts (*Buchnera*) and a recent radiation of aphids (*Uroleucon*) and pitfalls of testing for phylogenetic congruence. Evolution. 2000; 54: 517-525.

13. Moran NA, Dale C, Dunbar H, Smith WA, Ochman H. Intracellular symbionts of sharpshooters (Insecta: Hemiptera: Cicadellinae) from a distinct clade with a small genome. *Environ Microbiol.* 2003; 5: 116–126.
14. Hosokawa T, Kikuchi Y, Nikoh N, Shimada M, Fukatsu T. Strict host–symbiont cospeciation and reductive genome evolution in insect gut bacteria. *PLoS Biol.* 2006; 4: 337.
15. Cavanaugh CM (1994) Microbial Symbiosis: Patterns of Diversity in the Marine Environment. *Amer Zool.* 1994; 34: 79-89.
16. Nyholm SV, McFall-Ngai MJ. The winnowing: establishing the squid–*Vibrio* symbiosis. *Nat Rev Microbiol.* 2004; 2: 632-642.
17. Nishiguchi MK. Co-evolution of symbionts and hosts: The sepiolid-*Vibrio* model. In: Seckbach J (ed). *Symbiosis: Mechanisms and Model Systems*, Cole-Kluwer Academic Publishers: Dordrecht, The Netherlands, 2001, pp 757-774.
18. Dunlap PV, Ast JC, Kimura S, Fukui A, Yoshino T, Endo H. Phylogenetic analysis of host-symbiont specificity and codivergence in bioluminescent symbioses. *Cladistics.* 2007; 23: 507-532.
19. Kaeding AJ, Ast JC, Pearce MM, Urbanczyk H, Kimura S, Endo H, *et al.* Phylogenetic diversity and co-symbiosis in the bioluminescent symbioses of *Photobacterium mandapamensis*. *Appl and Environ Microbiol.* 2007; 73: 3173-3182.
20. Nyholm SV, Stabb EV, Ruby EG, McFall-Ngai MJ. Establishment of an animal-bacterial association: recruiting symbiotic vibrios from the environment. *PNAS.* 2000; 97: 10231-10235.
21. Visick KL, Ruby EG. *Vibrio fischeri* and its host: it takes two to tango. *Curr Opin Microbiol.* 2006; 9: 632-638.
22. Lee KH, Ruby EG. Effect of the squid host on the abundance and distribution of symbiotic *Vibrio fischeri* in nature. *Appl Environ Microbiol.* 1994; 60: 1565-1571.
23. Herring PJ, Morin JG. Bioluminescence in fishes. In: Herring PJ (ed). *Bioluminescence in action*, Academic Press Inc: London, UK, 1978, pp 272-329.
24. Gould AL, Dougan KE, Koenigbauer ST, Dunlap PV. Life history of the symbiotically luminous cardinalfish *Siphamia tubifer* (Perciformes: Apogonidae). *J Fish Biol.* 2016; 89: 1359-1377.
25. Gould AL, Harii S, Dunlap PV. Host preference, site fidelity, and homing behavior of the symbiotically luminous cardinalfish, *Siphamia tubifer* (Perciformes: Apogonidae). *Mar Biol.* 2014; 161: 2897-2907.

26. Tamura R. Experimental observations on the association between the cardinalfish (*Siphamia versicolor*) and the sea urchin (*Diadema setosum*). *Galaxea*. 1982; 1: 1–10.
27. Dunlap PV, Gould AL, Wittenrich ML, Nakamura M. Symbiosis initiation in the bacterially luminous sea urchin cardinalfish *Siphamia versicolor*. *J Fish Biol*. 2012; 81: 1340-1356.
28. Gould AL, Dunlap PV. Genomic analysis of a cardinalfish with larval homing potential reveals genetic admixture in the Okinawa Islands. *Mol Ecol*. 2017; 26: 3870–3882.
29. Rodriguez-Lanetty M, Hoegh-Guldberg O. The phylogeography and connectivity of the latitudinally widespread scleractinian coral *Plesiastrea versipora* in the Western Pacific. *Mol Ecol*. 2002; 11: 1177-1189.
30. Nishikawa A, Sakai K. Genetic connectivity of the scleractinian coral *Goniastrea aspera* around the Okinawa Islands. *Coral Reefs*. 2005; 24: 318-323.
31. Urbanczyk H, Ogura Y, Hendry TA, Gould AL, Kiwaki N, Atkinson JT, *et al*. Genome sequence of *Photobacterium mandapamensis* strain svers. 1.1, the bioluminescent symbiont of the cardinal fish *Siphamia versicolor*. *J Bacteriol*. 2011; 193: 3144-3145.
32. Rowe HC, Renaut S, Guggisberg A. RAD in the realm of next-generation sequencing technologies. *Mol Ecol*. 2011; 20: 3499-3502.
33. Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat Rev Genetics*. 2016; 17: 81-92.
34. Wong SM, Carroll PA, Rahme LG, Ausubel FM, Calderwood SB. Modulation of expression of the ToxR regulon in *Vibrio cholerae* by a member of the two-component family of response regulators. *Infection and Immunity*. 1998; 66: 5854-5861.
35. Jang J, Jung KT, Park J, Yoo CK, Rhie GE. The *Vibrio cholerae* VarS/VarA two-component system controls the expression of virulence proteins through ToxT regulation. *Microbiol*. 2011; 157: 1466-73.
36. Lenz DH, Miller MB, Zhu J, Kulkarni RV, Bassler BL. CsrA and three redundant small RNAs regulate quorum sensing in *Vibrio cholerae*. *Mol Microbiol*. 2005; 58: 1186-1202.
37. Kamp HD, Patimalla-Dipali B, Lazinski DW, Wallace-Gadsden F, Camilli A. Gene fitness landscapes of *Vibrio cholerae* at important stages of its life cycle. *PLoS Pathogens*. 2013; 9: e1003800.

38. Ding Y, Davis BM, Waldor MK. Hfq is essential for *Vibrio cholerae* virulence and downregulates σ E expression. *Mol Microbiol*. 2004; 53: 345-354.
39. Lenz DH, Mok KC, Lilley BN, Kulkarni RV, Wingreen NS, Bassler BL. The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in *Vibrio harveyi* and *Vibrio cholerae*. *Cell*. 2004; 118: 69–82.
40. Vanhove AS, Hang S, Vijayakumar V, Wong AC, Asara JM, Watnick PI. *Vibrio cholerae* ensures function of host proteins required for virulence through consumption of luminal methionine sulfoxide. *PLoS Pathogens*. 2017; 13: e1006428.
41. Showalter RE, Silverman MR Nucleotide sequence of a gene, *hpt*, for hypoxanthine phosphoribosyltransferase from *Vibrio harveyi*. *Nucleic Acids Research*. 1990; 18(15): 4621.
42. Dunlap PV, Nakamura M. Functional morphology of the luminescence system of *Siphonia versicolor* (Perciformes: Apogonidae), a bacterially luminous coral reef fish. *J Morph*. 2011; 272: 897-909.
43. Ruby EG, Greenberg EP, Hastings JW. Planktonic marine luminous bacteria: species distribution in the water column. *Appl Environ Microbiol*. 1980; 39: 302-306.
44. Haygood MG, Tebo BM, Nealson KH. Luminous bacteria of a monocentrid fish (*Monocentris japonicus*) and two anomalopid fishes (*Photoblepharon palpebratus* and *Kryptophanaron alfredi*): population sizes and growth within the light organs, and rates of release into the seawater. *Mar Biol*. 1984; 78: 249-254.
45. Nealson KH, Haygood MG, Tebo BM, Roman M, Miller E, McCosker JE. Contribution by symbiotically luminous fishes to the occurrence and bioluminescence of luminous bacteria in seawater. *Micro Ecol*. 1984; 10: 69-77.
46. Taylor MW, Schupp PJ, De Nys R, Kjelleberg S, Steinberg PD. Biogeography of bacteria associated with the marine sponge *Cymbastela concentrica*. *Environ Microbiol*. 2005; 7: 419-433.
47. Wollenberg MS, Ruby EG. Population structure of *Vibrio fischeri* within the light organs of *Euprymna scolopes* squid from two Oahu (Hawaii) populations. *Appl Environ Microbiol*. 2005; 7: 419-433.
48. Parchman TL, Gompert Z, Mudge J, Schilkey FD, Benkman CW, Buerkle C. Genome-wide association genetics of an adaptive trait in lodgepole pine. *Mol Ecol*. 2012; 21: 2991-3005.
49. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*. 2012; 7: e37135.

50. Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics*. 2011; 1: 171-182.
51. Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. Stacks: an analysis tool set for population genomics. *Mol Ecol*. 2013; 22: 3124-3140.
52. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature Methods*. 2012; 9: 357-359.
53. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, *et al*. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009; 25: 2078-2079.
54. Oksanen J, Kindt R, Legendre P, O'Hara B, Stevens MH, Oksanen MJ, *et al*. The vegan package. *Community Ecology Package*. 2007; 10: 631-637.
55. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2014.
56. Benestan L, Gosselin T, Perrier C, Sainte-Marie B, Rochette R, Bernatchez L. RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (*Homarus americanus*). *Mol Ecol*. 2015; 24: 3299-3315.
57. Xu T, Sun J, Lv J, Watanabe HK, Li T, Zou W, Rouse GW, Wang S, Qian PY, Bao Z, Qiu JW. Genome-wide discovery of single nucleotide polymorphisms (SNPs) and single nucleotide variants (SNVs) in deep-sea mussels: Potential use in population genomics and cross-species application. *Deep Sea Research Part II: Topical Studies in Oceanography*. 2017; 137: 318-326.
58. Linck E, Battey CJ. Minor allele frequency thresholds strongly affect population structure inference with genomic datasets. *Mol Ecol Resources*. 2019; 19: 639-647.
59. Jombart T, Ahmed I. *adegenet 1.3-1*: new tools for the analysis of genome-wide SNP data. *Bioinformatics*. 2011; 27: 3070-3071.
60. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Evolution*. 1984; 38: 1358-1370.
61. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, *et al*. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*. 2012; 6: 80-92.

Tables

Table 1. Permutational multivariate analyses of variance (PERMANOVA) of the genetic distances between *Siphamia tubifer* hosts and between their light organ symbionts. For each dataset the degrees of freedom (DF), sum of squares (SS), predicted F-value (F), R^2 (R), and P-value are shown. For the 2014 and 2013 datasets, sampling location was used as the grouping factor, whereas sampling year was used for the Sesoko dataset. Significant P-values are in bold

Dataset	Host					Symbiont				
	DF	SS	F	R	P	DF	SS	F	R	P
2014	3	15694	1.004	0.032	0.343	3	0.523	1.597	0.050	<0.001
2013	9	47367	1.008	0.054	0.131	9	1.099	1.249	0.064	<0.001
Sesoko	2	10718	1.027	0.038	0.033	2	0.225	1.092	0.043	0.093

Table 2. Outlier haplotypes for *Photobacterium mandapamensis* identified as potential drivers of genetic divergence between symbiotic populations at Kume and Okinawa Islands, Japan in 2014. Listed are the locus ID assigned by *Stacks* (Catchen et al. 2011, 2013) corresponding to each haplotype, its variant type and effect as determined with *SnpEff*, and the putative gene product and category (COG) of the protein coding region in which each locus is located

ID	Scaffold	Variant Type	Variant Effect	Gene	Gene product
92	2	missense	moderate	<i>hfq</i>	RNA chaperone Hfq
286	4	synonymous	LOW	PMSV_1336	unknown function
314	4	synonymous	LOW	PMSV_145	methyl-accepting chemotaxis protein
355a	4	missense	moderate	PMSV_1452	efflux RND transporter permease subunit
355b		consensus	-		
377	4	missense	moderate	<i>yeaG</i>	PrkA family serine protein kinase
379	4	missense	moderate	<i>yeaG</i>	PrkA family serine protein kinase
576	4	upstream	modifier	PMSV_554	unknown function
585	4	missense	moderate	PMSV_564	lipopolysaccharide assembly protein LapB
667a	4	consensus	-	PMSV_772	Hsp70 family protein
667b		stop gained	high		
705	4	missense	moderate	PMSV_851	inosine/guanosine kinase
712	4	missense	moderate	VV12383	DUF2786 domain-containing protein
772	5	missense	moderate	<i>hpt</i>	hypoxanthine phosphoribosyltransferase
811	5	missense	moderate	<i>barA</i>	two-component sensor histidine kinase BarA
891	6	missense	moderate	PMSV_3308	Trk system potassium uptake protein
950	7	missense	moderate	PMSV_3736	sigma-54-dependent Fis family transcriptional regulator
1162	8	missense	moderate	PMSV_2985	unknown function
1195	8	consensus	-	PMSV_3012	AsmA family protein
1263	8	missense	moderate	PMSV_2003	alpha/beta fold hydrolase
1265	8	stop gained	high	PMSV_2003	alpha/beta fold hydrolase
1269	8	synonymous	low	PMSV_2013	cobalamin ABC transporter substrate-binding protein

7

1354	8	missense	moderate	PMSV_2243	PTS fructose transporter subunit IIC
1407	8	missense	moderate	<i>gcvP</i>	aminomethyl-transferring glycine dehydrogenase

Table 3. Location, year, and number of *Siphamia tubifer* light organs sampled for this study in the Okinawa Islands, Japan

ID	Location	Latitude	Longitude	Year	N
S	Sesoko	26.635409	127.865832	2012	16
				2013	18
				2014	21
M	Motobu	26.655806	127.880286	2013	20
N	Nago	26.603673	127.932404	2013	21
Hd	Hedo	26.848756	128.252513	2013	17
It	Itoman	26.095109	127.658478	2013	14
				2014	27
O	Ou	26.127916	127.768981	2013	16
Y	Yonabaru	26.203032	127.771178	2013	16
Ik	Ikei	26.393535	127.988601	2013	15
				2014	22
Hk	Henoko	26.534554	128.046181	2013	17
A	Ada	26.741936	128.321107	2013	16
K	Kume	26.351617	126.820149	2014	26