Polyclonal symbiont populations in hydrothermal vent tubeworms and the environment

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

| Citation   | Polzin, Julia et al. “Polyclonal symbiont populations in hydrothermal vent tubeworms and the environment.” Proceedings of the Royal Society B: Biological Sciences 286, 1896 (February 2019): 20181281. © 2019 The Author(s) |
|------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| As Published | http://dx.doi.org/10.1098/rspb.2018.1281                                                                                                                                                     |
| Publisher  | Royal Society                                                                                                                                                                                  |
| Version    | Final published version                                                                                                                                                                       |
| Citable link | https://hdl.handle.net/1721.1/126276                                                                                                                                                           |
| Terms of Use | Creative Commons Attribution 4.0 International license                                                                                                                                          |
| Detailed Terms | https://creativecommons.org/licenses/by/4.0/                                                                                                                                                  |
Horizontally transmitted symbioses usually house multiple and variable symbiont genotypes that are acquired from a much more diverse environmental pool via partner choice mechanisms. However, in the deep-sea hydrothermal vent tubeworm *Riftia pachyptila* (Vestimentifera, Siboglinidae), it has been suggested that the *Candidatus Endoriftia persephone* symbiont is monoclonal. Here, we show with high-coverage metagenomics that adult *R. pachyptila* house a polyclonal symbiont population consisting of one dominant and several low-frequency variants. This dominance of one genotype is confirmed by multilocus gene sequencing of amplified housekeeping genes in a broad range of host individuals where three out of four loci (*atpA*, *uvrD* and *recA*) revealed no genomic differences, while one locus (*gyrB*) was more diverse in adults than in juveniles. We also analysed a metagenome of free-living Endoriftia and found that the free-living population showed greater sequence variability than the host-associated population. Most juveniles and adults shared a specific dominant genotype, while other genotypes can dominate in few individuals. We suggest that although generally permissive, partner choice is selective enough to restrict uptake of some genotypes present in the environment.

1. Introduction

Nearly all ecosystems harbour a large diversity of mutualistic interactions in which microbes live in physical contact with a eukaryotic host and the partners cooperate to increase each other's fitness compared with solitary life [1,2]. However, the persistence of these symbiotic mutualisms represents a challenge for evolutionary theory [1,3–7]. The evolution of cheaters among cooperating symbionts may lead to cooperators being outcompeted in the host-associated symbiont population, reducing the benefits to the host and thus decreasing its fitness [1,2,8,9]. In addition, a further detriment for the host population ensues if virulent variants arise that can successfully compete with cooperators [10]. Therefore, mechanisms are required to create a positive assortment of symbiont genotypes, so that cooperating partners preferentially interact with each other [11].

For horizontally transmitted symbionts, theory predicts that partner choice mechanisms through signalling between partners or screening of the partner play a crucial role in maintaining the associations [3,12,13]. Moreover, the structure and abundance of symbiont populations in the environment as well as dispersal and contact of both partners were identified as important factors influencing which symbiont strains are transmitted [14,15]. In particular, horizontal transmission creates the opportunity for uptake of symbiont variants that are optimized for local conditions [15]. For example, in well-studied
horizontally transmitted symbioses such as *Aliivibrio fischeri*—squad and rhizobia—legumes, symbionts within and between individual hosts are polyclonal [16,17], and environmental exceeds host-associated diversity. This suggests that partner choice is selective enough to restrict genotypes from entering the host, but is also permissive enough to allow a specific variety of genotypes into the host [16,18,19].

The hydrothermal vent tubeworms *Riftia pachyptila* (short *Riftia*, Vestimentifera, Siboglinidae) contain the sulfur-oxidizing bacterial symbiont *Candidateatus Endoriftia* persephone (short Endoriftia) in the symbiont-housing organ. These symbionts are transmitted horizontally upon settlement of the larvae through uptake from the environment [20], and viable bacteria can be released upon host death suggesting connectivity with the free-living symbiont population [21]. Indeed, several vestimentiferan host species share the same symbiont 16S rRNA phylotype [15,22–27], which was also detected free-living among tubeworm clumps in diffuse vent flow areas, on cold basalt and in the water column at ambient free-living among tubeworm clumps in diffuse vent flow.

16S rRNA activity with the free-living symbiont population [21]. Indeed, bacteria can be released upon host death suggesting connectivity with the free-living symbiont population [21]. Indeed, several vestimentiferan host species share the same symbiont 16S rRNA phylotype [15,22–27], which was also detected free-living among tubeworm clumps in diffuse vent flow areas, on cold basalt and in the water column at ambient free-living among tubeworm clumps in diffuse vent flow.

Indeed, several vestimentiferan host species share the same symbiont 16S rRNA phylotype [15,22–27], which was also detected free-living among tubeworm clumps in diffuse vent flow areas, on cold basalt and in the water column at ambient free-living among tubeworm clumps in diffuse vent flow.

The hydrothermal vent tubeworms *Riftia pachyptila* (short *Riftia*, Vestimentifera, Siboglinidae) contain the sulfur-oxidizing bacterial symbiont *Candidateatus Endoriftia* persephone (short Endoriftia) in the symbiont-housing organ. These symbionts are transmitted horizontally upon settlement of the larvae through uptake from the environment [20], and viable bacteria can be released upon host death suggesting connectivity with the free-living symbiont population [21]. Indeed, several vestimentiferan host species share the same symbiont 16S rRNA phylotype [15,22–27], which was also detected free-living among tubeworm clumps in diffuse vent flow areas, on cold basalt and in the water column at ambient free-living among tubeworm clumps in diffuse vent flow.

Indeed, several vestimentiferan host species share the same symbiont 16S rRNA phylotype [15,22–27], which was also detected free-living among tubeworm clumps in diffuse vent flow areas, on cold basalt and in the water column at ambient free-living among tubeworm clumps in diffuse vent flow.

16S rRNA activity with the free-living symbiont population [21]. Indeed, bacteria can be released upon host death suggesting connectivity with the free-living symbiont population [21]. Indeed, several vestimentiferan host species share the same symbiont 16S rRNA phylotype [15,22–27], which was also detected free-living among tubeworm clumps in diffuse vent flow areas, on cold basalt and in the water column at ambient free-living among tubeworm clumps in diffuse vent flow.

16S rRNA activity with the free-living symbiont population [21]. Indeed, bacteria can be released upon host death suggesting connectivity with the free-living symbiont population [21]. Indeed, several vestimentiferan host species share the same symbiont 16S rRNA phylotype [15,22–27], which was also detected free-living among tubeworm clumps in diffuse vent flow areas, on cold basalt and in the water column at ambient free-living among tubeworm clumps in diffuse vent flow.

16S rRNA activity with the free-living symbiont population [21]. Indeed, bacteria can be released upon host death suggesting connectivity with the free-living symbiont population [21]. Indeed, several vestimentiferan host species share the same symbiont 16S rRNA phylotype [15,22–27], which was also detected free-living among tubeworm clumps in diffuse vent flow areas, on cold basalt and in the water column at ambient free-living among tubeworm clumps in diffuse vent flow.
The clonal diversity of the host-associated symbiont population was examined with the eBURST tool, using the four loci atpA, recA, uvrD, and gyrB [32]. Genetic diversity was calculated based on the average nucleotide diversity and based on the number of segregating sites [35]. To test the neutral equilibrium model of selection, Tajima’s D [36] was calculated for each locus ($p < 0.01$), by using the number of segregating sites and the pairwise sequence differences in R with the software packages ‘ape’ v.3.4 and ‘pegas’ v.0.9.

(c) Metagenomic comparison
For the host-associated symbiont population from a single Riftia specimen from the Tica location at the EPR 9°50’N, a Percoll density gradient centrifugation was used to obtain an enriched symbiont fraction with depleted host content from the trophosome [23]. Trophosome was homogenized in imidazole-buffered saline (0.49 M NaCl, 0.03 M MgSO4, 0.011 M CaCl2, 0.003 M KCl and 0.05 M imidazole, pH 7.1) and centrifuged by Percoll centrifugation. Subsamples of the gradient were analysed by fluorescence in situ hybridization (FISH), and host-depleted fractions were washed in TE buffer and pooled for DNA extraction (see above). For the free-living population, we used a symbiont-enriched biofilm sample of basalt underneath the Riftia clump containing the host specimen used for metagenomics that was stored in glyceroll/TE buffer and flash frozen in liquid nitrogen. Enriched symbiont fractions of host-associated and free-living samples were quantified by FISH [37], and the absence of host tissue was confirmed using a host-specific FISH probe RP1752 (electronic supplementary material, table S3). DNA was extracted as described above.

A metagenomic library was made from the free-living Endoriftia and the symbiont-enriched host-associated Endoriftia population using the Illumina Nextera XT DNA Library Preparation Kit. The free-living and host-associated samples were barcoded and sequenced in a paired-end 150 bp multiplex run on the Illumina HiSeq 2000 at the Yale Center for Genome Analysis (YCGA). Reads were demultiplexed by the YCGA and imported into CLC GENOMICS WORKBENCH 8.0.2 (http://www.clcbio.com) for further processing and analysis. Adapters and low-quality regions were trimmed from the demultiplexed sequences, and overlapping paired reads were merged using default parameter in CLC (mismatch: 2, gap: 3). The processed reads were then mapped in CLC GENOMICS WORKBENCH to the reference genome sequence (Genbank assembly record: GCF_000224455.1, [23]) using the following parameters: mismatch cost = 2, insertion cost = 3, deletion cost = 3, length fraction = 0.9, similarity fraction = 0.8, auto-detected paired distances, random mapping of non-specific matches. Variants at segregating sites in the read mapping were called using the low-frequency variant caller in CLC GENOMICS WORKBENCH to the reference genome sequence (Genbank assembly record: GCF_000224455.1, [23]) using the following parameters: required significance = 0.1%; ignore positions with coverage > $1 \times 10^5$; ignore broken pairs; minimum coverage = 10; minimum count = 2; minimum frequency = 1.0%; base quality filter of radius 5, minimum central quality = 30, minimum neighbourhood quality = 30; relative read direction filter with significance = 0.1%. Variant overlap between the two samples was calculated using custom python scripts.

To integrate the Sanger multilocus gene sequencing approach with the Illumina metagenomics approach, reads from the host-associated and free-living symbiont metagenomes were mapped to the Sanger reads of Nsym and Nfree with segregating sites and variants, by considering default parameter with the help of Bowtie2 [38].

(d) Statistical analysis
To test whether the sampling location or year, trophosome subsample location or host size has an influence on the ST, we used the ImerTest package in R (v.2.0–33; (p-value cut-off of 0.01)) with the ANOVA method testing for linear mixed-effects models on the combined dataset to see any effect of interactions between fixed factors on the ST and with each factor separately (location, year, type of Riftia or trophosome compared to the ST). To obtain an overview of the data, tidyverse was used to include the data as factors and numbers.

(e) Quantitative polymerase chain reaction analysis
To quantify the abundance of Endoriftia in each seawater, biofilm, sediment and in the trophosome, we confirmed the presence of the 16S rRNA Endoriftia phylotype by sequencing the 16S rRNA gene and used quantitative PCR (qPCR) assays to estimate total bacterial and Endoriftia-specific 16S rRNA gene copy numbers (electronic supplementary material, table S1). The qPCR analysis was performed on a LightCycler 480 Real-Time PCR system (Roche) with the LightCycler RELEASE software 1.5.0. For both trophosome and free-living Endoriftia samples, we used universal 16S rRNA primers and compared the gene abundance with symbiont-specific 16S rRNA gene abundances (electronic supplementary material, table S3). To prepare the 16S rRNA standard, a gene fragment amplified with 16S rRNA universal primers was cloned into TOPO TA Cloning Vector pCR2.1 (Invitrogen). After plasmid isolation with the QIAprep Spin Mini-prep Kit (Qiagen), the plasmid was linearized with the HindIII high-fidelity restriction enzyme (BioLabs) and checked by agarose gel electrophoresis and SYBR Gold (Thermo Fisher) staining. For more details, see the electronic supplementary material.

We confirmed that the 16S rRNA gene was present in a single copy using qPCR of five samples with universal and specific primers (electronic supplementary material, figure S2 and table S3). The multiloci atpA, uvrD, recA, gyrB were also confirmed to be a single copy by qPCR comparisons to the 16S rRNA gene abundance in six samples (electronic supplementary material, figure S2 and table S3). The ileS locus revealed a distinct higher copy number abundance in all the samples, pointing to a non-single copy presence (electronic supplementary material, figure S2). Therefore, this locus was excluded from further analysis.

3. Results
To investigate the variability within the symbiont population in Riftia from the four vent sites Tica and P-Vent at 9°50’ N EPR and Janine and Genesis at 13° N EPR, we analysed a total of 16 specimens, including nine adults and seven juveniles using a multilocus gene framework by comparing the allelic profiles of the four housekeeping loci atpA, uvrD, recA, and gyrB. Because we used direct PCR amplification and sequencing for the Nsym population, it was possible to identify dominant symbiont genotypes as a single ST in each host-associated population. The multilocus gene sequencing revealed no variation of the loci (atpA, uvrD and recA) across all adult and juvenile Riftia specimens from 2010 and 2011 (electronic supplementary material, figure S1 and table S1). The fourth locus (gyrB) revealed nucleotide diversity per site of 1.19%.

Our eBURST analysis revealed one dominant sequence type 1 (ST 1), which is identical among a broad range of host individuals, including six juveniles and five adults, and to the metagenome of a single, previously studied Riftia trophosome specimen [23]. One (ST 2) out of seven juveniles and four (ST 3–6) out of nine adults differed in the gyrB locus from ST 1. Analysing the variable loci gyrB, we did not detect any significant effect of the host.
**Table 1.** *Candidatus Endoriftia persephone* symbiont abundances in biofilm close to host tubeworms and in sediment away from host tubeworms determined by quantitative PCR. (‘Vent activity’ indicates whether the samples were collected at an active or an inactive hydrothermal vent basalt biofilms. The total amount of total DNA recovered varied among different samples. The *Endoriftia* symbiont 16S rRNA copy number density was calculated for 1 ng recovered DNA. Each sample was analysed in triplicate and ± values give the standard deviation. Copy numbers were transformed to cell numbers by the factor 4.1 copies cell\(^{-1}\) as an average copy number variation in bacteria [39].

| site                  | vent activity | DNA sample | total DNA (ng) | distance to host (m) | total bacterial 16S rRNA copies per total DNA | symbiont 16S rRNA copies per total DNA | symbiont/total bacteria (%) cells | symbiont/total bacteria (%) copies |
|-----------------------|---------------|------------|----------------|----------------------|-----------------------------------------------|--------------------------------------|---------------------------------|----------------------------------|
| Tica, basalt under tw | active        | 44         | 635            | 0                    | \(4.3 \times 10^3 \pm 6.14 \times 10^1\)            | \(5.08 \times 10^2 \pm 1.65 \times 10^2\) | 11.73                           | 48.09                            |
| Tica, basalt under tw | active        | 45         | 560            | 0                    | \(1.85 \times 10^5 \pm 1.25 \times 10^4\)            | \(1.22 \times 10^3 \pm 0.96 \times 10^3\) | 6.62                            | 27.14                            |
| Tica, basalt under tw | active        | 69         | 1340           | 0                    | \(2.74 \times 10^5 \pm 1.83 \times 10^4\)            | \(2.65 \times 10^3 \pm 1.08 \times 10^2\) | 0.96                            | 3.94                             |
| P-Vent, basalt under tw | active  | 58         | 855            | 0                    | \(2.84 \times 10^5 \pm 4.15 \times 10^4\)            | \(2.93 \times 10^3 \pm 3.22 \times 10^2\) | 1.03                            | 4.22                             |
| Tica, basalt close to tw | inactive     | 46         | 795            | 1                    | \(5.96 \times 10^5 \pm 5.35 \times 10^4\)            | \(1.01 \times 10^3 \pm 5.79 \times 10^3\) | 1.76                            | 7.22                             |
| Tica, basalt close to tw | inactive     | 47         | 565            | 1                    | \(5.79 \times 10^5 \pm 5.66 \times 10^4\)            | \(9.20 \times 10^2 \pm 5.04 \times 10^3\) | 1.59                            | 6.52                             |
| P-Vent, basalt close to tw | inactive     | 59         | 1045           | 1                    | \(9.19 \times 10^5 \pm 2.34 \times 10^4\)            | \(1.58 \times 10^3 \pm 1.34 \times 10^3\) | 1.74                            | 7.13                             |
| inactive basalt, BM 13 | inactive     | 57         | 1370           | distance to the host \(> 50\) m | \(2.14 \times 10^4 \pm 4.38 \times 10^3\)            | \(3.83 \times 10^2 \pm 0.66 \times 10^3\) | 1.79                            | 7.34                             |
| inactive basalt, BM 18 | inactive     | 56         | 1065           | distance to the host \(> 50\) m | \(3.99 \times 10^4 \pm 2.68 \times 10^2\)            | \(2.53 \times 10^2 \pm 6.57 \times 10^3\) | 0.63                            | 2.58                             |
size (adults 10–60 cm length, juveniles 0.3–1 cm length), trophosome subsample location (median or top part of trophosome in adults and total trophosome in juveniles), vent location (Tica and P-Vent, 9°50’ N EPR; Janine and Genesis 13° N EPR) or year (2010, 2011) on the ST (linear model with fixed factors; p > 0.76), or a combination of these factors (linear mixed-effects model; p > 0.609).

We further tested whether additional genotypes can coexist within a single host specimen, reasoning that previous studies might have been unable to detect low abundance genotypes owing to low sequencing coverage (14 × [22] and 25 × [23]). We performed Illumina-based deep metagenomic sequencing with coverage of 828 × for one Riftia-associated symbiont population (Nsym) collected from Tica at 9°50’ N EPR in 2011 (electronic supplementary material, table S4). Our deep sequencing analysis shows that in this adult Riftia specimen, the majority of reads (76.2%) were identical to ST 1 (this study) and to the Riftia 1 reference metagenome [23]. The analysis of single and multiple nucleotide variants and indels revealed 7796 variants evenly spread across the reference metagenome, with a median frequency of 99.87% for these variants. Our dataset combines the low overall nucleotide heterogeneity of 0.17% in one specimen and suggests that the dominant phylotype is still the most abundant member of the endosymbiont community by a large margin (electronic supplementary material, figure S3).

To examine the genotypic variability in the free-living symbiont population (Nfree), we cloned and sequenced two multilocus genes (uvrD and atpA) that were also used in the Nsym gene analysis. We determined uvrD gene sequences from 14 environmental samples (11–29 clones sample−1) and atpA gene sequences from nine samples (1–21 clones sample−1) with seven samples overlapping for the two genes (electronic supplementary material, table S2). Moreover, we quantified population densities from the vent tubeworm habitat and also from surrounding environments with deep-sea conditions (inactive basalt, off-axis sediments and water) using qPCR with specific 16S rRNA gene primers. Over 70% of the samples (26 of 36) showed the presence of the free-living 16S rRNA Endoriftia phylotype (electronic supplementary material, table S2), and all samples were dominated by ST 1 (88% uvrD and 86% atpA clones). Highest symbiont densities were found underneath tubeworm clumps at Tica and P-Vent (0.52–2.93 × 10^3 16S rRNA copies per ng total DNA) and lower densities on inactive basalt close to and far away from these aggregations (0.16–1.01 × 10^3 16S rRNA copies per ng total DNA) (table 1; electronic supplementary material, table S2). These results indicate that the ST 1 genotype dominated the free-living Endoriftia population in all investigated environments. Apart from ST 1, other allelic variants were exclusively found in the free-living microbial population with a nucleotide heterogeneity of 5.16% (atpA) and 20.09% (uvrD), which indicated that the primers were not completely specific for Endoriftia. The data also showed considerable genotypic variation among the free-living populations in the number of alleles per locus with 14 for atpA and 30 for uvrD and segregating sites ranging between 20 for atpA and 37 for uvrD (table 2), respectively.

Table 2. Summary of nucleotide variation across genes within the host-associated (Nsym) and free-living (Nfree) Candidatus Endoriftia persephone population. (Nucleotide variation across genes of the host-associated and free-living Endoriftia population with n, sample size; seg. sites, segregating sites; bp, number of base pairs; η (s) %, percentage of segregating sites that are singleton alleles; π total, nucleotide diversity per site for all changes; π non, nucleotide diversity per site for non-synonymous changes; π sym, nucleotide diversity per site for synonymous changes; dN/dS, ratio of non-synonymous and synonymous changes. The free-living population was analysed with 118 clones for the atpA locus and 315 clones for the uvrD locus.)

| locus  | n | alleles | seg. sites | bp  | η (s) % | π total | π non | π sym | dN/dS | Tajima’s D |
|-------|---|---------|------------|-----|---------|---------|-------|-------|-------|------------|
| Nsym  |   |         |            |     |         |         |       |       |       |            |
| gyrB  | 16| 5       | 13         | 1846| 100     | 1.19    | 0.75  | 0.68  | 0.51  | -2.245277 |
| Nfree |   |         |            |     |         |         |       |       |       |            |
| atpA  | 9 | 14      | 20         | 446 | 85      | 5.16    | 4.71  | 1.34  | 0.71  | -2.731669 |
| uvrD  | 14| 30      | 37         | 209 | 86      | 20.09   | 18.10 | 5.57  | 0.73  | -2.686529 |

*aSome allelic variants of the free-living Endoriftia population were judged to not be part of the symbiont population due to greater nucleotide heterogeneity of 5.16% (atpA) and 20.09% (uvrD) compared with the host-associated population, with no variation in these loci.

Comparison of the multilocus genes with the same loci in the metagenomes confirmed that the dominant genotype in both deeply sequenced populations is identical to ST 1. However, the comparison also revealed that the multilocus gene sequence genes showed between approximately 2–14% and approximately 3–3.5% variation in the Nfree and Nsym metagenomes, respectively (figure 2; electronic supplementary material, figure S4). Among these additional variants, ST 3, 4 and 6 were also detectable in both metagenomes, while ST 5 from a single adult and ST 2 from a juvenile were only supported by the Nsym metagenome reads. Hence, not all variants were found in the Nfree metagenome, suggesting that the free-living symbiont population may have remained under-sampled or some variants may have been absent in this sample.

The Nfree population of Endoriftia shows higher heterogeneity than Nsym and also the recurrence of sequence
types from Nsym across different samples of Nfree. For the _uvrD_ locus, 15 of 30 alleles were supported by both Nsym and Nfree metagenomics reads, while nine alleles were only supported by Nsym reads (electronic supplementary material, table S6). While ST 1 dominated all habitats, some of the variants also occurred in more than one habitat. Interestingly, one allele that occurred at 18 m above the vents was also found at 103 m (electronic supplementary material, table S6). Another allele was present at the active vent site Tica next to tubeworm clumps, but also present in the pelagial at 9°30’ N of the East Pacific Rise. The _f_-values are the median frequency of each class of variant (e.g. variants shared between the two categories had a median frequency of 100%). The _n_-values are the number of variants in each class. (b) Variants with minimum frequencies in the free-living and host-associated symbiont population. SNP, single-nucleotide polymorphism.

Our study revealed that the polyclonal symbiont population in _Riftia_ is dominated by one genotype and several low-frequency variants based on our high-coverage metagenome sequencing analysis. Such variation cannot result from within host mutations alone, because the same symbiont variants were found across different samples in addition to the relatively short lifespan of the host [42]. Rather, the variant richness more likely results from infection of multiple strains during horizontal transmission in larvae [21]

The dominance of one strain in juvenile and adult hosts may result from frequency-dependent infection owing to dominance of this strain in the local environment or from preferential growth during the development of the host. The dominant genotype matches with the previously reported metagenomic analysis of a specimen collected at 13° N in 2008 [23] and is represented by the dominant sequence type (ST 1) in a broad range of juvenile and adult _Riftia_. This points to one dominating variant across the four studied vent sites on the EPR 9° and 13° N in two consecutive years (2010 and 2011). A previous study revealed that at the EPR Endoriftia clusters by host species, considering the core genome, but considering the accessory genome, it

4. Discussion

Symbioses with horizontal transmission face the challenge of maintaining long-term stability of the association, because each host generation relies on the availability of free-living cooperating symbionts in the environment for uptake into the host through partner choice [8,9,14,15]. Uptake is strongly influenced by the structure and abundance of symbiont populations in the local environment and their encounter with the settled larvae [14]. In this study, we show that instead of a clonal host-associated population, as previously suggested [23], there is a considerable variation of Endoriftia genotypes housed in the trophosome of a single _Riftia_ specimen. Accordingly, partner choice cannot be restricted to one specific genotype that is selected from the environment. We propose that partner choice is more permissive and allows several different Endoriftia strains to enter the host from a large and abundant pool of free-living variants in the _Riftia_ habitat. Multiple genotypes of very closely related symbionts have also been recently detected in the trophosome of the hydrothermal vent tubeworm species _Ridgea piscesae_ [40] applying whole genome shotgun sequencing and in the metagenomes of _Escarpia_ sp. and _Lamellibrachia_ sp. [41].

Figure 1. (a) Venn diagram showing the number of private and shared single-nucleotide variants of the free-living (Nfree) and host-associated (Nsym) Endoriftia metagenomes from the Tica vent site at 9°30’ N of the East Pacific Rise. The _f_-values are the median frequency of each class of variant (e.g. variants shared between the two categories had a median frequency of 100%). The _n_-values are the number of variants in each class. (b) Variants with minimum frequencies in the free-living and host-associated symbiont population. SNP, single-nucleotide polymorphism.
clusters by vent site, where Endoriftia in *Riftia* and *Tevnia* at EPR 9° N appeared to be more similar to each other than the Endoriftia population in *Riftia* collected from EPR 9° versus 13° N [40]. Our multilocus gene sequencing approach to discriminate closely related bacterial strains based on well-conserved housekeeping loci paints a more differentiated picture by showing that there is more variation within the same tubeworm species than previously anticipated.

Although 70% of the 17 sampled *Riftia* are dominated by one Endoriftia genotype (ST 1), other STs occur and can even dominate. These results further suggest that symbionts are locally adapted to a host-associated lifestyle and that symbiont reproduction primarily occurs within hosts and not in the environment [43,44].

In *Riftia*, other symbiont genotypes than ST 1 were dominant in several adults (ST 3, 4, 5 and 6) and one juvenile (ST 2) across vents and sampling years. The two tubeworms housing ST 3 and ST 4 were collected together with two other tubeworms that exhibited ST 1 from a single clump. Whether spatial and temporal heterogeneity of environmental

---

**Figure 2.** Metagenomic reads of the free-living (Nfree) and host-associated (Nsym) Endoriftia population mapped to the multilocus reference genes (atpA, gyrB, ileS, recA and uvrD). The most abundant STs correspond to the dominant alleles detected by metagenomics, while other low-frequency variants are present.
parameters known to be variable in the Riftia habitat [45–49] can lead to competitive replacement and selection of genotypes better adapted to different abiotic conditions, remains to be studied in a larger tubeworm collection with concurrent measurements of environmental parameters. Such selection may be enhanced by physiological adaptation of the host including enzymes and transporters, like a carbonic anhydrase [50], that are specific for a symbiotic lifestyle and may also control which symbiont genotype raises to dominance, potentially allowing the host to select the optimal partner for specific environmental conditions. Additionally, host genotype may influence the type of symbiont selected, but elucidating such specific genetic interactions will require host in addition to symbiont genotyping.

Quantitative PCR analyses from biofilms at active vent sites revealed high abundances of the 16S rRNA symbiont phylotype. ST 1 dominating in most tubeworms also dominated in the environmental biofilms of the Riftia habitat. The metagenome analysis of the environmental sample taken underneath a tubeworm clump showed that the dominant free-living genotype is identical to ST 1. Furthermore, the multilocus gene analyses of the free-living symbiont populations also revealed the dominance of sequences identical to ST 1 in two genes (atpA and uvrD). The prevalence of free-living and host-associated ST 1 (and other identified STs) in the Riftia habitat suggests that vigorous diffuse sulfide emissions [51] fuel chemosynthetic activity in the host [52]. The symbiont population outside of the host in the biofilms could be dormant, rely on heterotrophy or autotrophy but regardless of metabolic state, it guarantees the availability of source populations for transmission. Especially if the symbiont population is metabolically active, the unstable chemical environment may also introduce temporal variation in the symbiont pool available for host infection. Whether dominating symbiont genotypes in the trophosome are also favoured in the free-living environment after they escape from dead hosts [21] remains to be studied.

We could detect the 16S rRNA Endoriftia phylotype in a broad range of pelagic and benthic deep-sea environments, confirming and extending a previous study [28,29]. Owing to the presumably low concentration, we could not quantify free-living Endoriftia in non-vent environments from the pelagial, from far away sediments and from most inactive basalts far away from tubeworms. We were able, however, to sequence individual clones of atpA and uvrD genes and reported that the vast majority of retrieved sequences matches ST 1 (atpA clones: two samples 100% ST 1; two samples ST 1 dominance 67% and 93%; uvrD clones: seven samples 100% ST 1, three samples dominance 73, 75 and 93%). Presumably owing to its low concentration, we could not quantify Endoriftia in non-vent environments. However, our data suggest that ST 1 is abundant in various non-vent benthic and pelagic environments along the axial summit trough housing vents. Chemosynthesis should not be possible in those habitats in the absence of sulfide. Instead, functional analyses of metagenomes suggest that Endoriftia might be able to switch to heterotrophy [22]. Whether this metabolic versatility contributes to the distribution of Endoriftia and specifically of ST 1 in non-vent environments under deep-sea conditions is not known. Escape of symbionts from hosts upon cessation of vent flux into environments with deep-sea conditions and dispersal

Figure 3. Conceptual model of the connectivity between free-living and symbiotic Endoriftia populations. Aposymbiotic host larvae are initially infected by symbionts from the environmental pool [20] through a permissive partner choice mechanism. The development of the host results in preferential growth on one symbiont strain, most often the founder ST 1. Upon vent cessation the host dies, the symbionts are released into the environment and disperse to new active vent sites where infection of host may favour again the founder ST, resulting in the over-representation of the most beneficial partner for the initiation of the symbiosis.
may contribute to the broad distribution of ST 1 in such non-vent environments.

5. Conclusion
Our study suggests that genotypically diverse symbionts, available in high abundance for host infection, are acquired from the environment after settlement (figure 3). What functionally differentiates the Endoritida genotypes remains unknown. However, it is most likely that selection for the dominant genotype happens within the host rather than in the environment because environmental populations of bacteria are rarely dominated by a single clonal genotype [53]. This means that dominance in hosts causes dominance in environmental samples by the release of symbionts upon host death (figure 3). Regardless, our data support permissive partner choice similar to other horizontally transmitted symbionts because diverse populations persist in the host and different genotypes can dominate. Frequency-dependent uptake and growth or rather unspecific uptake and preferential growth of a dominant symbiont might be as crucial for the long-term persistence of this association, as the release of symbionts upon host death followed by proliferation or at least persistence in the extant environment.

Data accessibility. Sequencing data will be available at NCBI under the following accession number: project SRP117045; SRR6014239 host-associated symbiont population and SRR6014238 free-living symbiont population.

Authors contributions. M.B. and J.P. designed the research. J.P. and M.B. performed the research. J.P., P.A. and T.N. performed analyses. J.P. and M.B. wrote the manuscript and M.F.P. revised the manuscript.

Competing interests. We have no competing interests.

Funding. The European Marie Curie Network Symbiomics 264774 FP7-PEOPLE-2010-ITN granted to J.P. and M.B., and the Austrian Science Fund FWF 31543-B29 granted to M.B. funded this work. Additional funding was provided by a grant from the US National Science Foundation Biological Oceanography to M.F.P.

Acknowledgements. We thank the captains and the crew of L’Atalante with the submersible Nautilus and R/V Atlantis with the RON Jason and the chief scientists F. Lallier and S. Nooner. Special thanks to Christian Baranyi for technical support in the laboratory, Salvador Espada-H Fitzjosa for scientific exchange, Lukas Schuster for helping in with statistics, Eva Sintes for help with the qPCR and providing primers, and Michael Cutler for metagenomic library preparation.

References

1. Douglas AE. 2010 The symbiotic habit, vol. 51, pp. 197 – 198. Princeton, NJ: Princeton University Press.

2. Bornstein J. 2015 Mutualism. Oxford, NJ: Oxford University Press.

3. Sachs JL, Mueller UG, Wilcox TP, Bull JJ. 2004 The evolution of cooperation. Q. Rev. Biol. 79, 135 – 160. (doi:10.1086/383541)

4. Bhagavan R, Bergmüller R. 2008 Distinguishing four fundamental approaches to the evolution of helping. J. Evol. Biol. 21, 405 – 420. (doi:10.1111/j.1420-9101.2007.01482.x)

5. West SA, Griffin AS, Gardner A. 2007 Evolutionary explanations for cooperation. Curr. Biol. 17, R661 – R672. (doi:10.1016/j.cub.2007.06.004)

6. Gardner A, Foster KR. 2008 The evolution and ecology of cooperation — history and concepts. In Ecology of social evolution (eds J Korb, J Heinz), pp. 1 – 36. Berlin, Germany: Springer.

7. Cordero OX, Ventouras LA, DeLong EF, Polz MF. 2012 Bacterial public good dynamics in the ocean. Proc. Natl Acad. Sci. USA 109, 20 059 – 20 064. (doi:10.1073/pnas.1213441109)

8. Trivers RL. 1971 The evolution of reciprocal altruism. Q. Rev. Biol. 46, 35 – 57. (doi:10.1086/406755)

9. Hamilton WD. 1975 Innate social aptitudes of man: an approach from evolutionary genetics. In Biosocial anthropology (ed. R Fox), pp. 133 – 155. New York, NY: Wiley.

10. Frank SA. 1996 Host-symbiont conflict over the mixing of symbiotic lineages. Proc. R. Soc. B 263, 339 – 344. (doi:10.1098/rspb.1996.0052)

11. Fletcher JA, Doebeli M. 2009 A simple and general explanation for the evolution of altruism. Proc. R. Soc. B 276, 13 – 19. (doi:10.1098/rspb.2008.0829)

12. Bull JJ, Rice WR. 1991 Distinguishing mechanisms for the evolution of cooperation. J. Theor. Biol. 149, 63 – 74.

13. Archetti M. 2010 Contract theory for the evolution of cooperation: the right incentives attract the right partner. J. Theor. Biol. 269, 201 – 207. (doi:10.1016/j.jtbi.2010.10.025)

14. Genkai-Kato M, Yamamura N. 1999 Evolution of mutualistic symbiosis without vertical transmission. Theor. Pop. Biol. 55, 309 – 323.

15. Vrijenhoek RC. 2010 Genetics and evolution of deep-sea chemosynthetic bacteria and their invertebrate hosts. Top. Geobiol. 33, 15 – 49.

16. Wollenberg MS, Ruby EG. 2009 Population structure of Vibrio fischeri within the light organs of Euprymna scolopes squid from two Oahu (Hawaii) populations. Appl. Environ. Microbiol. 75, 193 – 202. (doi:10.1128/AEM.01792-08)

17. Denison RF. 2000 Legume sanctions and the evolution of symbiotic cooperation by rhizobia. Am. Nat. 156, 567 – 576. (doi:10.1086/316994)

18. Ehinger M, Mohr TJ, Starcevich JB, Sachs JL, Porter SS, Simms EL. 2014 Specialization-generalization within the light organs of Riftia pachyptila homogeneity among the endosymbionts of Riftia pachyptila and Tenia jerseyhonnae revealed by proteogenomics. ISME J. 6, 766 – 776. (doi:10.1038/ismej.2011.137)

19. Jones ML. 1981 Rifia pachyptila, new genus, new species, the vestimentiferan worm from the Galapagos Rift geothermal vents (Pogonophora). Proc. Biol. Soc. Wash. 93, 1295 – 1313.

20. Feldman RA, Black MB, Cary CS, Lutz RA, Vrijenhoek RC. 1997 Molecular phylogenetics of bacterial endosymbionts and their vestimentiferan hosts. Mol. Mar. Biol. Biotechnol. 6, 268 – 277.

21. Di Meo CA, Wilbur AE, Holben WE, Feldman RA, Vrijenhoek RC, Cary SC. 2000 Genetic variation among endosymbionts of widely distributed vestimentiferan tubeworms. Appl. Environ. Microbiol. 66, 651 – 658. (doi:10.1128/AEM.66.2.651-658.2000)

22. Perez M, Juniper SK. 2016 Insights into symbiont population structure among three vestimentiferan tubeworm host species at Eastern Pacific spreading centers. Appl. Environ. Microbiol. 82, 5197 – 5205. (doi:10.1128/AEM.00953-16)

23. Harmer TL, Rotjan RD, Nussbaumer AD, Bright M, Ng AW, DeCainne EG, Cavanaugh CM. 2008 Free-living tube worm endosymbionts found at deep-sea
vents. *Appl. Environ. Microbiol.* **74**, 3895 – 3898. (doi:10.1128/AEM.02470-07)

29. Gullmann LK, Beaulieu SE, Shank TM, Ding K, Seyfried WE, Sievert SM. 2015 Bacterial diversity and successional patterns during biofilm formation on freshly exposed basalt surfaces at diffuse-flow deep-sea vents. *Front. Microbiol.* **10**, 901.

30. Bright M, Bulgheresi S. 2010 A complex journey: transmission of microbial symbionts. *Nat. Rev. Microbiol.* **8**, 218 – 230. (doi:10.1038/nrmicro2262)

31. Bright M, Keckeis H, Fisher CR. 2000 An autoradiographic examination of carbon fixation, transfer and utilization in the *Riftia pachyptila* symbiosis. *Mar. Biol.* **136**, 621 – 632. (doi:10.1007/s002270050722)

32. Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG. 2004 Identification and phylogenetic sorting of bacterial lineages with universally conserved genes and proteins. *Cold Spring Harb. Protoc.* **7**, pdb.prot5253.

33. Le Bris N, Govenar B, Le Gall C, Fisher CR. 2006 Is the trophosome of *Ridgeia pachyptila* monoclonal? *Symbiosis* **74**, 55 – 65. (doi:10.1007/s13199-017-0490-7)

34. Russell SL, Corbett-Detig RB, Cavanaugh CM. 2017 Intrahost genetic diversity of bacterial symbionts exhibits evidence of mixed infections and recombiant haplotypes. *Mol. Biol. Evol.* **34**, 2747 – 2761. (doi:10.1093/molbev/msx188)

35. Daims H. 2009 Use of fluorescence in situ hybridization and the daime image analysis program for the cultivation-independent quantification of microorganisms in environmental and medical samples. *Cold Spring Harb. Protoc.* **7**, pdb.prot5253.

36. Tajima F. 1989 The effect of changes in population-size on DNA polymorphism. *Genetics* **123**, 597 – 601.

37. Le Bris N, Rodier P, Sarradin P-M, Le Gall C. 2006 Is temperature a good proxy for sulfide in hydrothermal vent habitats? *Cah. Biol. Mar.* **47**, 465 – 470.

38. Russell SL, Cavanaugh CM. 2017 Fast gapped-read alignment with Bowtie 2. *Nat. Methods.* **14**, 101 – 112. (doi:10.1038/nmeth.3923)

39. Klappenbach JA, Saxman PR, Cole JR, Schmidt TM. 2001 rrndb: the ribosomal RNA operon copy number database. *Nucleic Acids Res.* **29**, 181 – 184. (doi:10.1093/nar/29.1.181)

40. Perez M, Juniper SK. 2018 Is the trophosome of *Ridgeia pachyptila* monoclonal? *Symbiosis* **74**, 55 – 65. (doi:10.1007/s13199-017-0490-7)

41. Reveillaud J, Anderson R, Reves-Sohn S, Cavanaugh CM, Huber JA. 2018 Metagenomic investigation of vestimentiferan tubeworm endosymbionts from Mid-Cayman Rise reveals new insights into metabolism and diversity. *Microbiome* **6**, 19. (doi:10.1186/s40168-018-0411-x)

42. Langmead B, Salzberg SL. 2012 Fast gapped-read alignment with Bowtie 2. *Nat. Methods.* **9**, 57 – 60. (doi:10.1038/nmeth.f.308)

43. Russell SL, Cavanaugh CM. 2017 Intrahost genetic diversity of bacterial symbionts exhibits evidence of mixed infections and recombiant haplotypes. *Mol. Biol. Evol.* **34**, 2747 – 2761. (doi:10.1093/molbev/msx188)

44. Le Bris N, Govenar B, Le Gall C, Fisher CR. 2006 Is the trophosome of *Ridgeia pachyptila* monoclonal? *Symbiosis* **74**, 55 – 65. (doi:10.1007/s13199-017-0490-7)

45. Le Bris N, Rodier P, Sarradin P-M, Le Gall C. 2006 Is temperature a good proxy for sulfide in hydrothermal vent habitats? *Cah. Biol. Mar.* **47**, 465 – 470.

46. Sarradin PM, Caprais J-C, Briand P, Gaill F, Shillito B, Desbruyeres D. 1998 Chemical and thermal description of the environment of the genesis hydrothermal vent community (13°N, EPR). *Cah. Biol. Mar.* **39**, 159 – 167.

47. Contreira-Pereira L, Yu¨cel M, Omanovic D, Brulport J-P, Le Bris N. 2013 Compact autonomous voltammetric sensor for sulfide monitoring in deep sea vent habitats. *Deep Sea Res. Pt I: Oceanogr. Res. Pop.* **80**, 47 – 57. (doi:10.1016/j.dsr.2013.05.014)

48. Johnson KS. 1988 Chemical and biological interactions in the Rose Garden hydrothermal field, Galapagos spreading center. *Deep Sea Res.* **35**, 1723 – 1744. (doi:10.1016/0198-2466(88)90046-5)

49. Von Damm KL, Buttermore LG, Oosting SE, Bray AM, Formari DJ, Lilley MD, Shanks III WC. 1997 Direct observation of the evolution of a seafloor black smoker from vapor to brine. *Earth Planet. Sci. Lett.* **149**, 101 – 112. (doi:10.1016/S0012-821X(97)00059-9)

50. Bright M, Lallier FH. 2010 The biology of vestimentiferan tubeworms. *Oceanogr. Mar. Biol. Annu. Rev.* **48**, 213 – 266.

51. Etter RJ, Mullineaux L. 2000 Deep-sea communities. In *Marine community ecology* (eds MD Bertness, S Gaines, M Hay), pp. 367 – 393. Sunderland, MA: Sinauer Associates Inc.

52. Fisher C. 1996 IV: Unusual energy sources. *Ecophysiology of primary production at deep-sea vents and seeps. Biogeosyst.* **11**, 313 – 336.

53. Cordova OX, Polz MF. 2014 Explaining microbial genomic diversity in light of evolutionary ecology. *Nat. Rev. Microbiol.* **12**, 263 – 273. (doi:10.1038/nrmicro3218)