Proceedings of the Royal Society B: DOI: 10.1098/rspb.2018.1281

Supporting Information

Polyclonal symbiont populations in hydrothermal vent tubeworms and the environment

Julia Polzin[1], Philip Arevalo[2], Thomas Nussbaumer[3,4], Martin F. Polz[2], Monika Bright[1]

[1]Department of Limnology and Bio-Oceanography, University of Vienna, Althanstrasse 14, A-1090 Vienna, Austria

[2] Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Parsons Laboratory, 15 Vassar Street, Cambridge MA 02139, USA

[3]Institute of Environmental Medicine (IEM), Helmholtz Center Munich, Ingolstädter Landstrasse 1, 85764 Neuherberg, Germany

[4] Institute of Network Biology (INET), Helmholtz Center Munich, Ingolstädter Landstrasse 1, 85764 Neuherberg, Germany

Supporting Material and Methods

Multi-locus gene sequencing

All primers were optimized by temperature gradient PCR. Cycling conditions included an initial 3 min denaturation at 95°C, followed by 35 cycles of 30 s denaturation at 95°C, 30 s specific annealing and specific extension time at 72°C. The final extension was for 10 min at 72°C. Reactions were carried out in a 50 µL containing dNTPs 0.2 mM, Taq polymerase 0.02 U/µL primers 0.5 µM, $MgCl_2$ 2 mM (Fermentas), BSA 0.01 mg/mL (Invitrogen). Annealing temperature and extension time were adjusted according to the gene amplified and the fragment length (the universal primers: 61°C 1 min ileS_gen, 53°C 30 s recA_gen_mod, and the specific primers: 42 °C 45 s atpA, 51 °C 30 s uvrD, 54 °C 1 min recA, 54 °C 1 min gyrB, 48 °C 2 min ileS (Table S3)). Amplified products were either sequenced directly or ligated and cloned with the TOPO TA Cloning Vector pCR2.1 and TOP10 chemically competent cells (Invitrogen) according to the manufacturer's instructions. After overnight growth, plates were kept at 4°C for 2 h to encourage blue staining of beta-glucosidase positive colonies. Inserts in plasmids from white colonies were amplified by PCR using universal M13 primers and subsequently Sanger sequenced (Microsynth).

Two universal primer pairs of the recA and ileS locus were used to amplify the gene fragments from one trophosome sample by PCR. The products were cloned into the pCR 2.1-TOPO vector using the TOPO TA Cloninig Kit (Invitrogen) according to the manufacturer's instructions.–The purpose for the cloning of the amplicons using universal primers was to compare the specificity of universal primers with the specificity of specific primers. We confirmed the dominance of a single genotype by using universal recA and ileS primers on ~ 300 clones for each of the two loci from one trophosome sample, to overcome the problem of choosing too specific targeted primers. The majority of clones from both genes (92% recA, 78% ileS) were identical to the reference metagenome sequences [24]. The remaining 8% of the *recA* clones had only one SNP each with 62.5 % of them representing non-synonymous

changes. The remaining 22% of ileS variants exhibited one SNP each (and 1% had 2 SNPs each) with 67.5 % being non-synonymous changes. These results indicate the presence of low abundance variants in the symbiont population.

Quantitative PCR analysis

An external standard for the multilocus gene approach approach was prepared by amplification of each of the housekeeping genes from trophosome DNA extracts. New primers were optimized with a gradient temperature PCR. Subsequently, each amplification was performed in a PCR reaction as described above using the specific annealing temperatures (56 °C atpA, 51 °C uvrD, 54 °C recA, 46 °C gyrB, 56 °C ileS) for 30 s. Amplification products were checked by agarose gel electrophoresis and stain with SYBR Gold, quantified using a Nanodrop spectrophotometer and the gene copy number was calculated from the concentration of the amplified DNA and the size of the fragment. Ten-fold serial dilutions in 10 mM Tris with 10^9 to 10^2 gene copies were prepared in triplicates.

The housekeeping gene abundance in the trophosome was determined for single-copy presence in samples 100-fold diluted in 10 mM Tris. All reactions were carried out in a 10 µL reaction volume. One µL of diluted or undiluted sample was added to the mix in each of the wells, using the LightCycler 480 SYBR Green Mastermix 480 (Roche), 0.5 µM of primers (Table S3), 0.2 mg/mL Bovine Serum Albumin (BSA) and ultra-pure sterile water (Sigma) in 96-well q-PCR plates with optical tape (Bio-Rad). Amplification was performed under the following conditions: 10 min initial denaturation at 95 °C; 50 cycles at 95 °C for 5 s, specific annealing temperature (50 °C 16S rRNA specific, 56° C 16S rRNA universal, 56 °C atpA, 51 °C uvrD, 54 °C recA, 46 °C gyrB, 56 °C ileS) for 5 s, 72 °C for 15 s, 80 °C for 3 s. To check for specificity of the q-PCR reaction, the melting curve analysis was done in a range of 65-95 °C with a read every 0.5 °C. Subsequently, PCR efficiencies and correlations for standard curves were calculated for triplicate reactions. Each of the gene fragments was detected using

a specific standard and gene abundances were compared to each other and to the 16S rRNA gene, which is considered as single copy gene from the published Endoriftia draft genomes [23, 24].

Supporting Figures

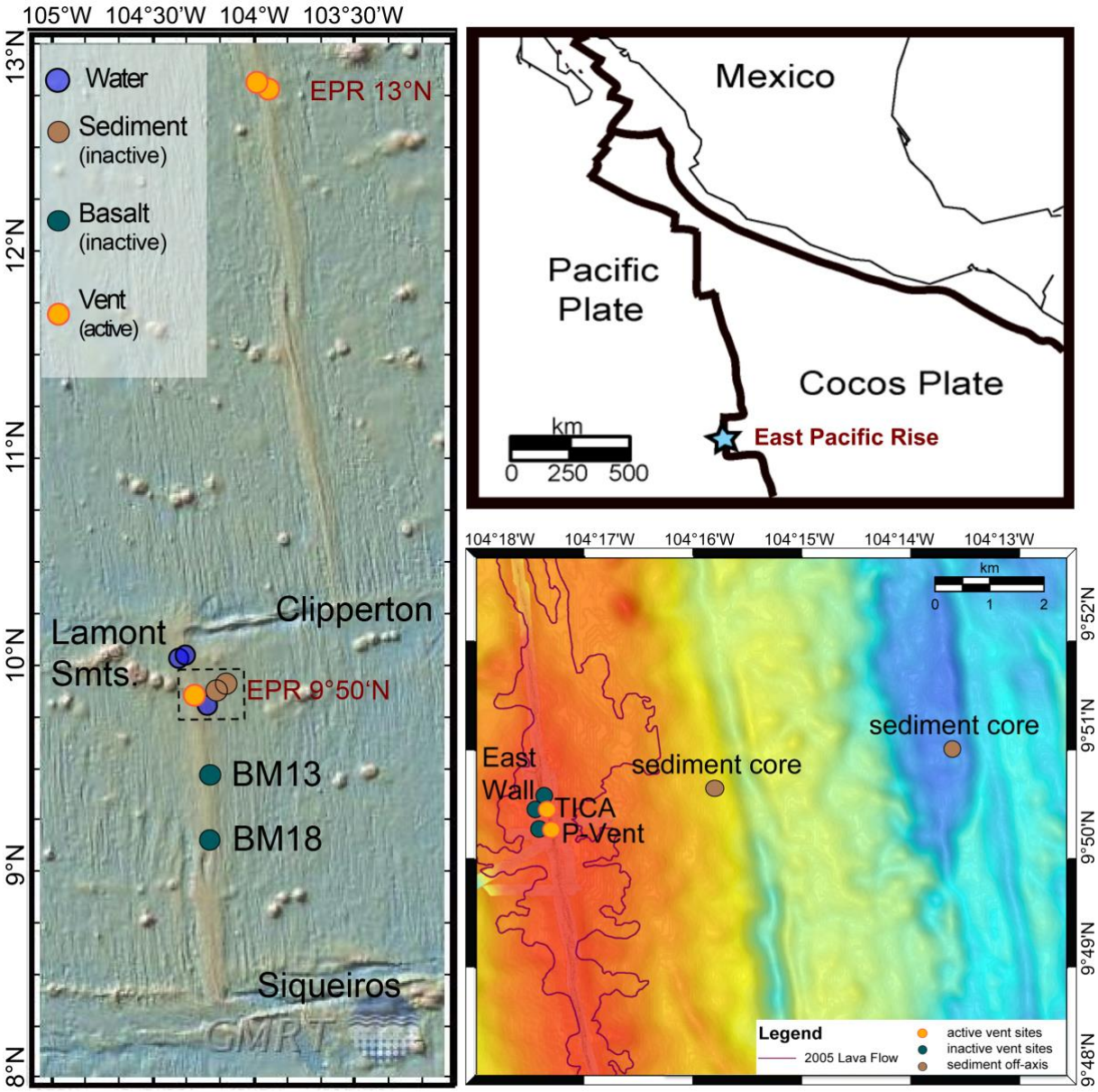

Figure S1: Sampling locations of the free-living 16S rRNA phylotype of *Cand.* Endoriftia persephone symbiont population at the axial summit trough of the 9°50 N region. Endoriftia is present at the active vent sites Tica and P-Vent and at inactive basalt close to these vent sites, in the water column up to 162 m, at inactive basalt (BM 13, BM 18), in the sediment off-axis the axial summit trough and also on biofilm of the inactive vent site East Wall.

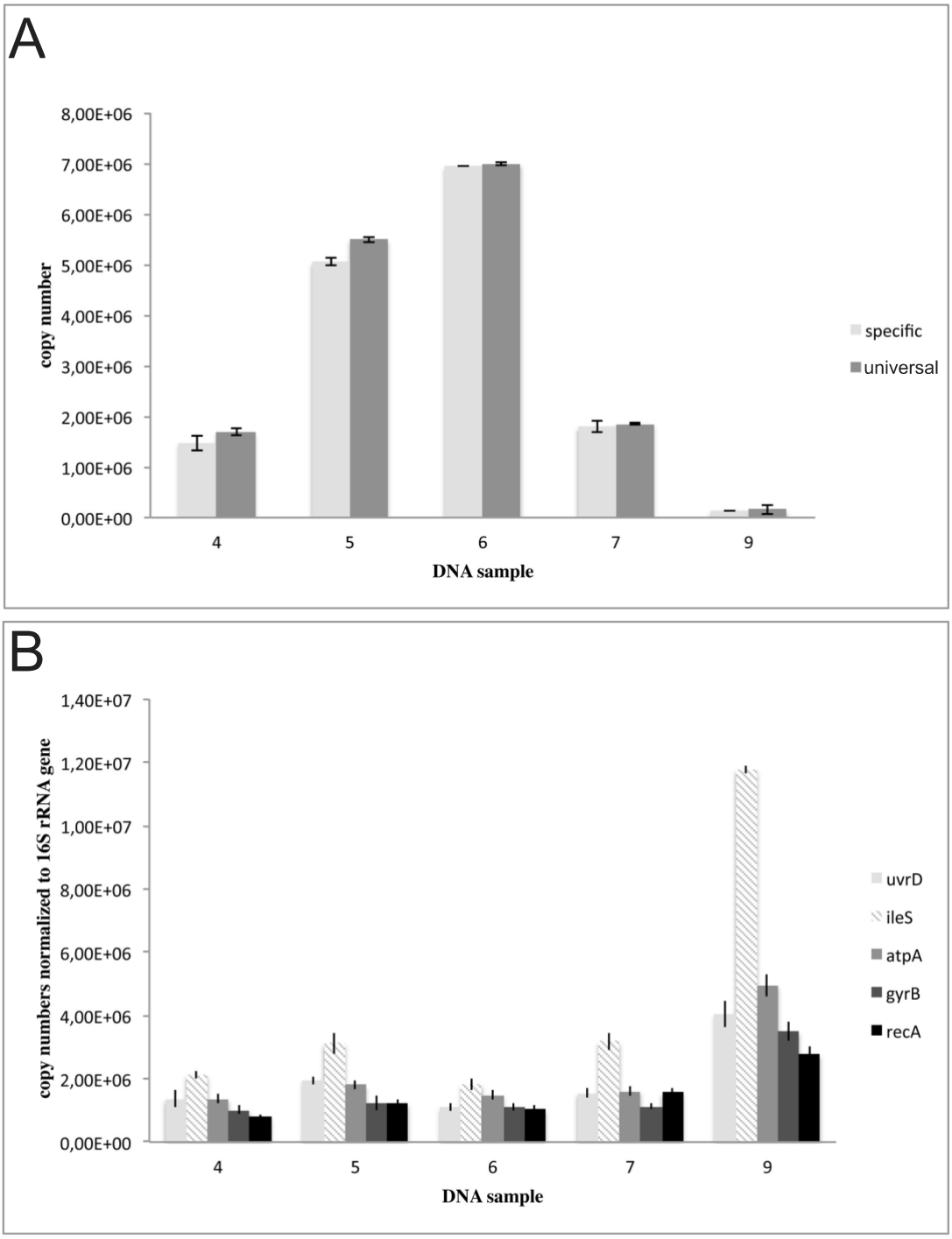

Figure S2. A) 16S rRNA universal bacterial and Endoriftia specific 16S rRNA gene copy numbers of trophosome samples. B) The 16S rRNA gene quantification of the trophosome revealed a 1:1 ratio for universal bacterial and Endoriftia specific 16S rRNA primers and quantitative comparison of genes in the six trophosome samples normalized to the 16S rRNA gene, as an endogenous reference gene. Relative abundance values were obtained for the target genes as well as for the control for each sample and the relative abundance of the target gene was divided by the value derived from the 16S control. The ileS gene is significantly higher in most samples analyzed and therefore is might not be a single copy gene.

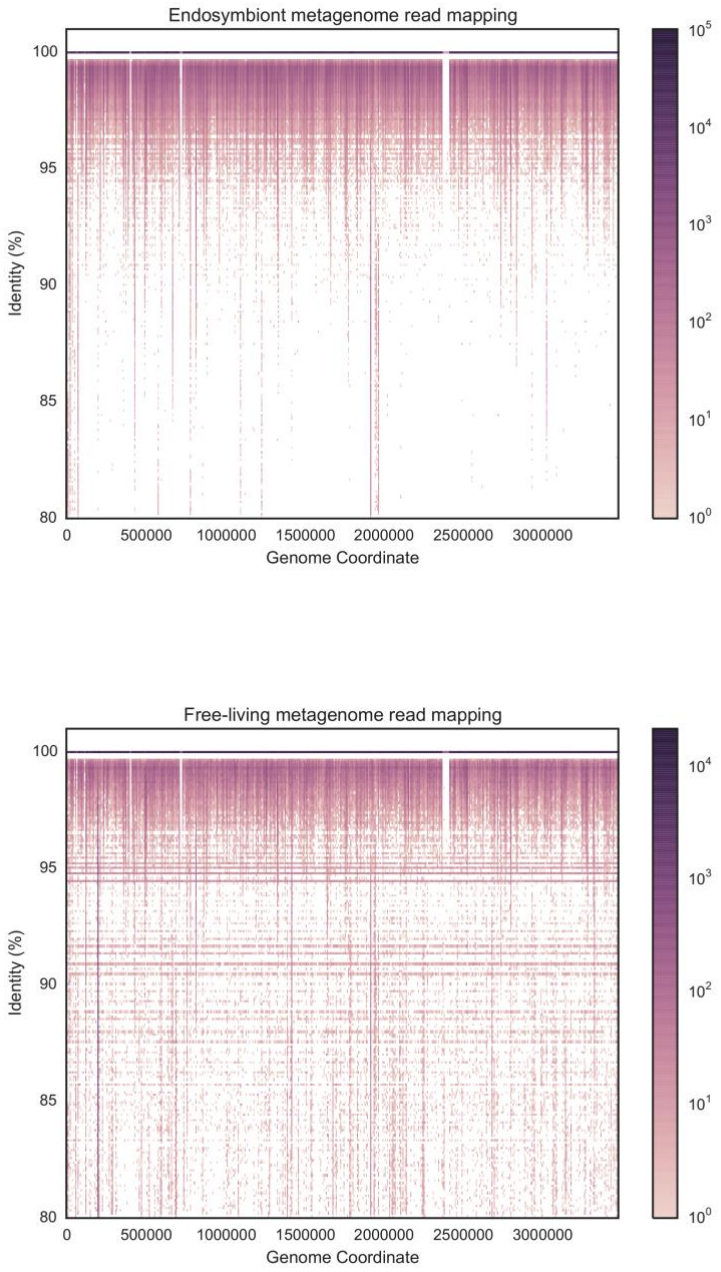

Figure S3. Read mapping of the host-associated and free-living endosymbiont metagenome to *Cand.* Endoriftia persephone reference metagenome *Riftia 1* [24]. X-axis represents the genome coordinate generated by an arbitrary ordering of reference contigs. Y-axis represents percent identity of reads mapped to a particular genome coordinate. Color represents the absolute number of reads mapped to a particular coordinate with darker colors signifying more reads.

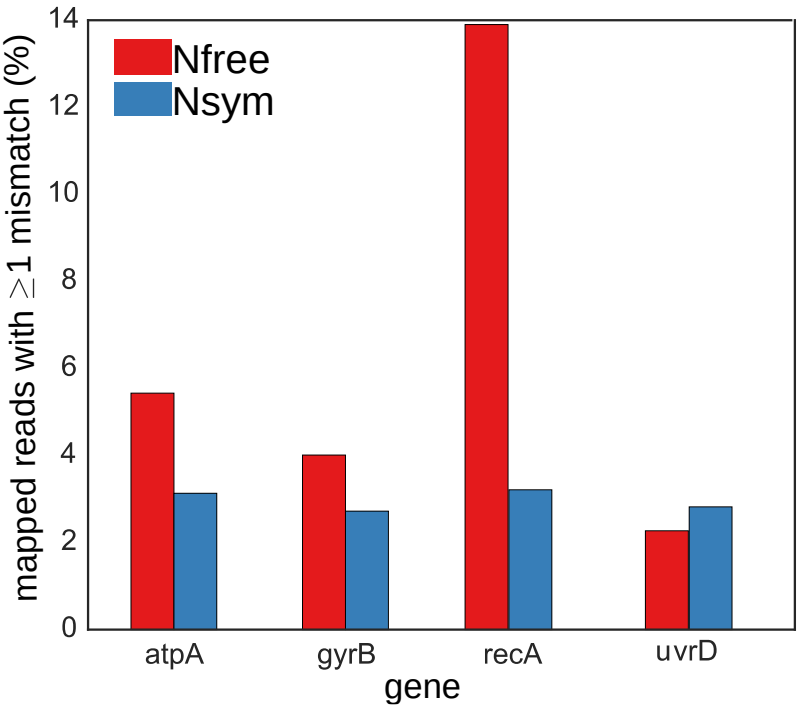

Figure S4: Metagenomic reads of the free-living and host-associated Endoriftia population with one or more mismatches mapped to the four multi-locus genes. The high percent of reads with one or more mismatches mapping to the recA locus detected in the free-living population likely stems from inclusion of several alleles that likely originate from bacteria other than Endoriftia.

Supporting Tables

Table S1. Sanger multi-locus (ML) gene sequencing approach for samples of *Riftia* adult and juvenile host-associated Endoriftia with allelic profile, sequence types (ST) and integration with Illumina metagenomes of host-associated (N_{sym}) and free-living (N_{free}) Endoriftia.

no	year	*Riftia*	sample #	length cm	trophosome[a]	location[b]	used for[c]	fixation[d]	atpA[e]	uvrD[e]	recA[e]	gyrB[e]	ST[e]	SNPs[f]	Metagenome Nsym reads support[g]	Metagenome Nfree read support[g]
1	2011	adult	1540	60	M	Tica	ML	cryo	1	1	1	3	3	1	2417/332	1345/336
2	2011	adult	1541	40	M	Tica	ML	cryo	1	1	1	4	4	2	2870/5,381	1103/9, 381
3	2011	adult	1542	54	M	Tica	ML	cryo	1	1	1	1	1			
4	2011	adult	1555	27	M	P-Vent	ML	cryo	1	1	1	1	1			
5	2011	adult	1556	21	M	P-Vent	ML	cryo	1	1	1	1	1			
6	2011	adult	1557	31.5	M	P-Vent	ML	cryo	1	1	1	5	5	1	2156/483	1125/none
7	2010	juvenile	1459	1	total	P-Vent	ML	cryo	1	1	1	1	1			
8	2010	juvenile	1460	0.3	total	P-Vent	ML	cryo	1	1	1	1	1			
9	2010	juvenile	1461	1	total	P-Vent	ML	cryo	1	1	1	2	2	2	4316/17,23	1903/none,1
10	2010	juvenile	1535_1	3	T	Tica	ML	ethanol	1	1	1	1	1			
11	2010	juvenile	1536_1	3	T	Tica	ML	ethanol	1	1	1	1	1			
12	2010	juvenile	1532	0.5	total	Janine	ML	cryo	1	1	1	1	1			
13	2010	juvenile	1533	1	total	Genesis	ML	cryo	1	1	1	1	1			
14	2010	adult	1534	10	M	Janine	ML	ethanol	1	1	1	1	1			
15	2010	adult	1533_5	33	M	Genesis	ML	cryo	1	1	1	1	1			
16	2010	adult	1532_1	23	T	Janine	ML	cryo	1	1	1	6	6	6	1845/11,27,3,12,9,7	1017/none,15,1, none/none/none
	2011	adult	1551	58	TMB	Tica	MG	cryo	1	1	1	1	1			

[a] specimen length and developmental stage distinguishes adults and juveniles, and trophosome part used for analyses is listed as: total = entire trophosome, T = top part, M = middle part, B = bottom part

[b] vent site (East Pacific Rise 9°50'N with the vent sites Tica and P-Vent, and 13°N with the vent sites Janine and Genesis)

[c] methods applied: ML = multi-locus sequencing, MG = metagenome

[d] fixation: cryo = frozen in liquid nitrogen, ethanol = 100% ethanol
[e] allelic profile for the four loci atpA, uvrD, recA and gyrB. Sequences were compared to the symbiont reference metagenome of Gardebrecht et al. [24] and a "1" indicates a 100% match to the reference. ST: resulting sequence type (ST) by concatenating the four loci for each sample
[f] SNPs: single nucleotide polymorphism compared to the symbiont reference metagenome [24]
[g] first number refers to the reads matching the reference (ST 1), second number refers to the reads matching the variant

Table S2. Multi-locus (ML) gene sequencing approach, quantification of free-living Endoriftia populations. For integration with Illumina metagenomes of host-associated (N_{sym}) and free-living (N_{free}) Endoriftia see Table S6.

no	year	sample	ID	lat.	long.	location[a]	vent activity	sym. presence	used for[b]	fixation[c]	DNA (ng/µl)[d]	qPCR copy (%)[e]	atpA clones[f]	uvrD clones[f]
37	2011	basalt	43	9°50.405	104°17.505	Tica, underneath tw	yes	yes	/	ethanol	9.5			
38	2011	basalt	44	9°50.405	104°17.505	Tica, underneath tw	yes	yes	ML/ MG	ethanol	10.2		17/9	25/20
39	2011	water	28	9°51.186	104°17.664	2495 m, 18 m alt.	no	yes	ML	cryo	10.6		3/2	27/17
40	2011	water	34	9°50.880	104°17.622	2490 m, 24 m alt.	no	yes	/	cryo	18.4			
41	2011	water	37	9°50.856	104°17.616	2420 m, 90 m alt.	no	yes	ML	cryo	10.0			15/14
42	2011	water	25	10°2.304	104°19.974	2424 m, 103 m alt.	no	yes	ML	cryo	8.6			16/12
43	2011	water	22	10°2.304	104°19.974	2443 m, 122 m alt.	no	yes	ML	cryo	12.1			11/8
44	2011	basalt	46	9°50.405	104°17.505	Tica vent, underneath tw	yes	yes	ML/ quantification	cryo	12.7	11.73	20/14	26/22
45	2011	basalt	49	9°50.405	104°17.505	Tica vent, underneath tw	yes	yes	quantification	cryo	11.2	6.62		
46	2011	basalt	52	9°50.409	104°17.499	Tica, next to tw	no	yes	quantification	cryo	15.9	1.76		
47	2011	basalt	55	9°50.409	104°17.499	Tica, next to tw	no	yes	quantification	cryo	11.3	1.59		
48	2011	basalt	60	9°50.409	104°17.499	Tica, next to tw	no	yes	ML	cryo	11			20/11
49	2011	water	4	9°53.191	104°15.111	2628 m, 20 m alt.	no	no	N.A.	cryo	10.9			
50	2011	water	7	9°53.190	104°15.014	2651 m, 20 m alt.	no	no	N.A.	cryo	9.8			
51	2011	water	10	9°53.181	104°14.451	2672 m, 20 m alt.	no	no	N.A.	cryo	12.0			
52	2011	water	13	9°53.180	104°14.321	2715 m, 20 m alt.	no	no	N.A.	cryo	14.7			
53	2011	water	16	10°2.561	104°20.710	2389 m, 167 m alt.	no	no	N.A.	cryo	12.9			
54	2011	water	19	10°2.748	104°20.082	2395 m, 162 m alt.	no	yes	/	cryo	11.1			
55	2011	water	31	9°51.144	104°17.664	2388 m 125 m alt.	no	yes	/	cryo	11.6			
56	2011	basalt	67	9°7.991	104°11.989	BM18, away from tw	no	yes	ML/ quantification	cryo	21.3	0.63	20/20	29/29

57	2011	basalt	70	9°27.002	104°14.278	BM13, away from tw	no	yes	ML/ quantification	cryo	27.4	1.79	15/14	20/20
58	2011	basalt	74	9°50.293	104°17.479	P-Vent, underneath tw	yes	yes	ML/ quantification	cryo	17,1	1,03	21/21	23/23
59	2011	basalt	76	9°50.293	104°17.479	P-Vent, next to tw	no	yes	ML/ quantification	cryo	20,9	1,74	14/14	
60	2009	basalt	1	9°50.056	104°17.445	Sketchy, away from tw	no	no	N.A.	ethanol	10,9			
61	2009	basalt	4	9°50.268	104°17.474	P-Vent, underneath tw	yes	no	N.A.	ethanol	11,3			
62	2009	basalt	7	9°50.268	104°17.474	P-Vent, underneath tw	yes	yes	/	ethanol	11,5			
63	2009	sediment	25	9°50.644	104°18.220	core 7, 3 km off-axis	no	yes	ML	ethanol	12,7		1/1	29/29
64	2009	sediment	28	9°51.006	104°13.677	core 8, 7 km off-axis	no	yes	ML	ethanol	15,4			23/23
65	2009	basalt	37	9°50.405	104°17.505	Tica, next to tw	no	no	N.A.	ethanol	11,9			
66	2009	basalt	40	9°50.540	104°17.506	East-Wall, away from tw	no	yes	/	ethanol	12,8			
67	2009	sediment	55	9°51.006	104°13.679	core 9, 3 km off-axis	no	yes	ML	ethanol	17,4			26/26
68	2009	sediment	58	9°50.649	104°15.828	core 10, 7 km off-axis	no	yes	ML	ethanol	17			25/25
69	2010	basalt	10	9°50.404	104°17.495	Tica active, underneath tw	yes	yes	ML/ quantification	ethanol	26,8	0,96	7/7	
70	2010	biofilm	104	9°50.408	104°17.494	Tica, active, TASC	yes	yes	/	ethanol	16,8			
71	2010	basalt	20	9°50.305	104°17.484	Bio 9, away from tw	no	no	N.A.	ethanol	27,6			
72	2009	basalt	38	9°50.405	104°17.494	Tica, underneath tw	yes	no	N.A.	ethanol	19,8			

[a] vent site (East Pacific Rise 9°50'N with the vent sites Tica and P-Vent), inactive basalt sites (bottom pressure recorder (BM)13, BM18, Sketchy, Bio9), and inactive sediment sites from cores 3-7 km off-axis, as well as water column samples with altitude (alt.) = distance to sea floor. For samples collected "close to tubeworms (tw)" we choose sites, where no tubeworms were located within 1 m diameter around the collection site; next to tw: basalt in up to 1 m distance; away from tw: basalt in more than 1 m distance. TASCs: tubeworm artificial settlement cubes [21]

[b] methods applied: ML = multi-locus sequening, MG = metagenome sequencing, symbiont quantification. N.A.: not applied – no symbiont presence

[c] fixation: cryo = frozen in liquid nitrogen, ethanol = 100% ethanol

[d] total bacterial DNA was extracted amplified with universal 16S rRNA primes and re-amplified with specific Endoriftia 16S rRNA primers

[e] copy number abundance was measured as described in the methods section

[f] number of clones analyzed for the atpA and uvrD locus (first number); clones, which are identical to the ST1 (second number)

Table S3. Primers for the multi-locus approach and fluorescence *in situ* hybridization probes.

PCR primer for Endoriftia

Gene	Primer sequences (5' - 3')	fragment	Reference
atpA	forw.: ATC TGA CTC TAT TTG GTC AGC rev.: TTA TAT CTC TTT AGC CAG GGC	461 bp	this study
uvrD	forw.: ATT ACC CAG CCG CTC TAC AGT C rev.: AGA GGG GAG AGA TTG GCG TAG G	209 bp	this study
recA specific	forw.: CAG ATC GAG AAG CAG TTT GG rev.: GTA CCC GCA GTT CAA TGT CC	928 bp	this study
recA universal	forw.: GC5 TTY ATY GAT GC GAR CA rev.: CCC AT5 TC5 CCT TCD ATY TC	207 bp	this study with courtesy of Eva Sintes
gyrB specific	forw.: GTG TCT CAG ATA ACG GGA GAG G rev.: TGA TTC TCG ATT CAT TGT CG	2072 bp	this study
ileS specific	forw.: TGG GAA GAG AAG AAG CTC TAC G rev.: CTT CGC CCA CGA CAT TCT C	2707 bp	this study
ileS universal	ileSBCUP1: GCC CGG CTG GGA YWS NCA YGG ileSBKDN1:TGG AGC CGG AGT CGA WCC ANM MNT	1337 bp	[34]
16S specific [a]	RifTO44: GGC CTA GAT TGA CGC TGC GG TA RifTO445: TCC TCA GGC TTT TCT TCC	401 bp	[21, 29]
16S universal	27F: AGA GTT TGA TCM TGG CTC AG 1492R: TAC GGY TAC CTT GTT ACG ACT T	1465 bp	[53]

qPCR primer for Endoriftia

Gene	Primer sequence (5' - 3')	fragment	Reference
atpA	forw.: AGA TCG AGC AGG AGA CCA ACC rev.: TCT TTA GCC AGG GCT TCA AC	142 bp	this study
uvrD	forw.: ATT ACC CAG CCG CTC TAC AGT C rev.: AGA GGG GAG AGA TTG GCG TAG G	209 bp	this study (see PCR primers)
recA specific	forw.: TGG CAC TTG TGC CTT CAT AG rev.: TCG ATC ACC ACC ACG TCT AC	171 bp	this study
recA universal	forw.: GC5 TTY ATY GAT GC GAR CA rev.: CCC AT5 TC5 CCT TCD ATY TC	207 bp	this study, with courtesy of Eva Sintes
gyrB specific	forw.: GCG CTA CAG TGG AAT GAG TC rev.: GAA CTG CTG TCA AGC CTT CC	206 bp	this study
gyrB universal	forw.: GCG GAA GCG GCC NGS NAT GTA rev.: CCG TCC ACG TCG GCR TCN GYC CAT	725 bp	[34]
ileS specific	forw.: CCG CGA CTA TTT GAG CTA CC rev.: TTC TCG TCG ACG GTA AAA CC	153 bp	this study
16S specific [a]	RifTO44: GGC CTA GAT TGA CGC TGC GG TA RifTO445: TCC TCA GGC TTT TCT TCC	401 bp	[21, 29]
16S univeral	1369F: CGG TGA ATA CGT TCY CGG 1492R: TAC GGY TAC CTT GTT ACG ACT T	124 bp	[53, 54]

Oligonucleotide probes for fluorescence *in situ* hybridization for Endoriftia

probe[b]	sequence	specificity	% FA[d]	reference
EUB338 I	5'-GCT GCC TCC CGT AGG AGT -3'	most bacteria (EUB I+II+III)	35 %	[55]
EUB338 II	5'-GCA GCC ACC CGT AGG TGT -3'	most bacteria (EUB I+II+III)	35 %	[56]
EUB338 III	5'-GCT GCC ACC CGT AGG TGT -3'	most bacteria (EUB I+II+III)	35 %	[56]
RifTO445[e]	5'-TCC TCA GGC TTT TCT TCC-3'	*Rif/Tev/Oas* symbiont	35 %	[21]
NON-338 EUB[c]	5'-ACT CCT ACG GGA GGC AGC-3'	negative control	10 %	[57]
RP1752[c]	5'-CGA CCT CTA AGC CGT CAA-3'	*Riftia* host	40 %	[58]

Primers for PCR and qPCR for the multi-locus gene sequencing and Endoriftia symbiont quantification. Primers are marked as "specific" and "universal" for amplification of five housekeeping genes and 16S rRNA distributed over the Endoriftia genome.

[a] symbiont specific 16S rRNA gene targeted primers (RifTO44, RifTO445) for PCR [21, 29]

[b] probes were labeled on the 5' end either with Cy3 or FITC to highlight simultaneously symbiont and all bacteria

[c] NON-338 EUB and a host-specific RP1752 probe were used as negative control

[d] all simultaneous hybridizations of symbiont specific probe and universal bacterial probe mix, as well as the negative controls, were carried out at 35% formamide stringency (given as % formamide (FA) in the hybridization buffer) and counter stained with the general DNA stain DAPI

[e] Endoriftia oligonucleotide probe RifTO445 [21] is specific for the 16S rRNA of the *Riftia pachyptila*, *Tevnia jerichonana* and *Oasisia alvinae* symbionts as all three vestimentiferans share an identical (*Tevnia*) or nearly identical (*Oasisia*) 16S rRNA symbiont phylotype (Rif/Tev/Oas) [16, 24, 26]

Table S4. General features of the Illumina HiSeq 2000 sequences of host-associated and free-living Endoriftia metagenomes from the East Pacific Rise 9°50'N Tica vent site.

	Free-living Endoriftia	***Riftia*-associated Endoriftia**
Sequenced bases	6.63×10^{10}	8.02×10^{9}
Sequenced reads	2.20×10^{8}	2.64×10^{7}
Bases after quality control and merging	3.15×10^{10}	3.10×10^{9}
Mapped bases to the reference	8.40×10^{8}	2.88×10^{9}
Mapped reads to the reference	2.78×10^{6}	9.54×10^{6}
% of mapped bases to the reference[a]	2.66	93.06
Coverage	241 x	828 x

[a] reference metagenome of Gardebrecht et al. [24]

Table S5. Latitude and longitude of sample collection sites and depth at the East Pacific Rise (EPR).

Site (EPR)	Tica	P-Vent N	Sketchy	East Wall	Bottom pressure recorder (BM) 13	Bottom pressure recorder (BM) 18	Bio 9	Genesis	Janine
latitude	9°50.404 N	9°50.293 N	9°50.056 N	9°50.540 N	9°27.002 N	9°7.9911 N	9°50.305 N	12°48.670 N	12°48.585 N
longitude	104°17.495 W	104°17.479 W	104°17.445 W	104°17.506 W	104°14.278 W	104°11.989 W	104°17.484 W	103°56.450 W	103°56.413 W
depth	2507 m	2505 m	2506 m	2350 m	2567 m	2537 m	2508 m	2639-2647 m	2624 m
vent activity	yes	yes	no	no	no	no	no	yes	yes

Table S6. Integration of Sanger multi-locus (ML) gene sequencing variants of free-living Endoriftia populations with Illumina metagenomes of host-associated (N_{sym}) and free-living (N_{free}) Endoriftia.

atpA clone	no	year	sample	lat.	long.	location	vent activity	SNPs	Metagenome N_{sym} read support	Metagenome N_{free} read support
1*	38	2011	basalt	9°50.405	104°17.505	Tica, underneath tw	yes	1	**2422/91**	**661/3**
2								1	2417/6	663/none
3								1	**2433/17**	**661/10**
4								2	2405/2,2	652/2,none
5								2	**2402/64,34**	**657/4,6**
6								1	**2419/75**	**661/28**
7								1	2424/7	663/none
8**								2	**2398/15,11**	**662/1,4**
1	39	2011	water	9°51.186	104°17.664	2495 m, 18 m alt.	no	3	**2368/6,15,11**	**659/2,1,4**
1	44	2011	basalt	9°50.405	104°17.505	Tica vent, underneath tw	yes	1	**2418/87**	**658/6**
2								1	**2423/7**	**663/1**
3								1	**2436/36**	**663/4**
4								2	2395/8,67	657/1,none
5								1	2424/4	664/none
6**								2	**2398/15,11**	**662/1,4**
1*	57	2011	basalt	9°27.002	104°14.278	BM13, away from tw	no	1	**2422/93**	**661/14**

uvrD clone	no	year	sample	lat.	long.	location	vent activity	SNPs	Metagenome N_{sym} read support	Metagenome N_{free} read support
1	38	2011	basalt	9°50.405	104°17.505	Tica, underneath tw	yes	2	487/8	172/none
2								1	**496/9**	**171/3**
3								1	478/none	167/none
4								1	**497/18**	**171/4**
5								1	480/none	168/none
1*[a]	39	2011	water	9°51.186	104°17.664	2495 m, 18 m alt.	no	2	491/18,none	163/5,none
2**								1	**505/6**	**174/1**
3								1	480/none	167/none
4**								1	**505/6**	**174/1**
5								1	480/1	167/none
6								3	480/15,4,none	161/5,none,none*
7								1	**513/11**	**174/7**
8								1	**506/32**	**174/15**
9*								1	**505/18**	**173/5**
1	41	2011	water	9°50.856	104°17.616	2420 m, 90 m alt.	no	1	503/12	176/none
1	42	2011	water	10°2.304	104°19.974	2424 m, 103 m alt.	no	1	478/none	168/1
2								1	**513/18**	**175/9**
3***								1	**493/2**	**170/3**
4*								1	**505/18**	**173/5**

uvrD clone	no	year	sample	lat.	long.	location[b]	vent activity	SNPs [c]	Metagenome N_{sym} read support[d]	Metagenome N_{free} read support[d]
1	43	2011	water	10°2.304	104°19.974	2443 m, 122 m alt.	no	1	**497/16**	**174/8**
2								2	487/16,4	163/8,none
3								2	489/7,5	174/none,none
1	44	2011	basalt	9°50.405	104°17.505	Tica vent, underneath tw	yes	1	**495/19**	**170/3**
2****								1	**504/11**	**174/6**
3								1	**506/16**	**174/8**
4								2	476/20,none	167/6,none
1****	48	2011	basalt	9°50.409	104°17.499	Tica, next to tw	no	1	**504/11**	**174/6**
2								1	501/3	175/none
3								1	**506/4**	**174/1**
4								1	501/3	175/none
5								1	500/7	175/none
6								1	478/2	167/none
7***								1	**476/2**	**161/3**
8								1	**480/1**	**168/1**
9								1	**491/9**	**170/1**

[a] * indicates the appearance of the same clone in different samples

[b] vent site (East Pacific Rise 9°50'N, with the vent site Tica), inactive basalt site (bottom pressure recorder (BM)13), and water column samples with altitude (alt.) = distance to sea floor

[c] single nucleotide polymorphisms (SNPs) indicate the number of SNPs detected compared to the reference metagenome of Gardebrecht et al. [24]

[d] first number in the column refers to the reads matching the reference and the second, third number refers to the total reads matching the variant clone for N_{sym} and N_{free} respectively. Bold: highlight the variants that are supported by N_{sym} and N_{free} metagenome reads