

Pyrosequencing analysis of endosymbiont population structure: co-occurrence of divergent symbiont lineages in a single vesicomid host clam

Frank J. Stewart¹ and Colleen M. Cavanaugh^{2*}

¹Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

²Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA.

Summary

Bacteria–eukaryote endosymbioses are perhaps the most pervasive co-evolutionary associations in nature. Here, intracellular chemosynthetic symbionts of deep-sea clams (*Vesicomidae*) were analysed by amplicon pyrosequencing to explore how symbiont transmission mode affects the genetic diversity of the within-host symbiont population. Vesicomid symbionts (*Gammaproteobacteria*) are presumed to be obligately intracellular, to undergo nearly strict vertical transmission between host generations, and to be clonal within a host. However, recent data show that vesicomid symbionts can be acquired laterally via horizontal transfer between hosts or uptake from the environment, potentially creating opportunities for multiple symbiont strains to occupy the same host. Here, genotype-specific PCR and direct sequencing of the bacterial internal transcribed spacer initially demonstrated the co-occurrence of two symbiont strains, symA and symB (93.5% nt identity), in 8 of 118 *Vesicomya* sp. clams from 3 of 7 hydrothermal vent sites on the Juan de Fuca Ridge. To confirm multiple strains within individual clams, amplicon pyrosequencing of two symbiont loci was used to obtain deep-coverage measurements (mean: ~1500× coverage per locus per clam) of symbiont population structure. Pyrosequencing confirmed symA–symB co-occurrence for two individuals, showing the presence of both genotypes in amplicon pools. However, in the majority of clams, the endosymbiont population was remarkably homogenous, with > 99.5% of sequences collapsing into a single symbiont

genotype in each clam. These results support the hypothesis that a predominantly vertical transmission strategy leads to the fixation of a single symbiont strain in most hosts. However, mixed symbiont populations do occur in vesicomids, potentially facilitating the exchange of genetic material between divergent symbiont lineages.

Introduction

Chemosynthetic endosymbioses represent ecologically and evolutionarily unique adaptations to life in sulfidic, marine habitats (e.g. hydrothermal vents, reducing sediments). In these mutualisms, autotrophic bacteria (typically *Gammaproteobacteria*) sheltered within the cells of an invertebrate host oxidize reduced chemicals to fuel carbon fixation, which in turn supports the nutrition of both the symbiont and host (Stewart *et al.*, 2005). The degree of symbiont–host co-adaptation can vary substantially among taxa. Some hosts (e.g. mussels of the Mytilidae) can supplement symbiont-based nutrition by heterotrophic feeding and can obtain their symbionts via lateral acquisition, defined here as uptake occurring either environmentally from a free-living population or horizontally from a co-occurring host individual. Other hosts, notably deep-sea clams of the *Vesicomidae*, exhibit a tighter symbiont–host association, depending almost entirely on their symbionts for sustenance and maintaining the specificity of the symbiont–host interaction by transmitting their symbionts vertically (via attachment to the egg) between host generations (Cary and Giovannoni, 1993). Given the recent availability of whole genome sequence data for two clam symbionts (Kuwahara *et al.*, 2007; Newton *et al.*, 2007), vesicomid symbioses are valuable systems for studying the impact of symbiont transmission strategy on the population structure, gene flow dynamics and genomic evolution of deep-sea endosymbiotic bacteria.

Lateral symbiont acquisition may strongly affect the diversity of symbiont populations within a host. In laterally transmitted endosymbioses, the acquisition of symbionts from a (potentially) diverse free-living population or from a co-occurring host may result in multiple symbiont genotypes (i.e. distinct lineages) coexisting within a single

Received 24 September, 2008; accepted 11 March 2009. *For correspondence. E-mail cavanaugh@fas.harvard.edu; Tel. (+1) 617 495 2177; Fax (+1) 617 496 6933.

host individual (i.e. a mixed population). Based on a combination of phylogenetic, cytological or symbiont uptake studies, lateral symbiont acquisition has been inferred for several deep-sea invertebrates, including mytilid mussels and vestimentiferan tubeworms (Won *et al.*, 2003; Kadar *et al.*, 2005; Nussbaumer *et al.*, 2006). Consistent with a lateral acquisition hypothesis, these taxa have also been shown to host mixed internal symbiont populations (DeChaine *et al.*, 2006; Vrijenhoek *et al.*, 2007; Duperron *et al.*, 2007; 2008; Harmer *et al.*, 2008).

In contrast, multiple genotypes are unlikely to be present in endosymbiont populations undergoing strict or nearly strict vertical inheritance, such as those of vesicomyid clams. In such associations, which are best exemplified by the maternally inherited nutritional symbioses of insects (e.g. *Buchnera*-aphid symbioses), diversification of the intracellular symbiont population is governed only by processes operating within a host (e.g. gene duplications, chromosomal re-arrangements; Narra and Ochman, 2006; Moran, 2007). Notably, gene exchange between the endosymbiont population and free-living bacterial populations (or those of distinct host lineages) is hypothesized to be low or non-existent, effectively reducing diversification via the acquisition of foreign genetic material (Bordenstein and Reznikoff, 2005). Furthermore, transmission from parent-to-offspring may force the endosymbiont population through repeated cycles of contraction and expansion (Funk *et al.*, 2001; Mira and Moran, 2002). Such bottlenecks facilitate random change in symbiont population diversity via genetic drift, often accelerating the rate at which a single symbiont genotype goes to fixation within a host (Wernegreen, 2005). As such, endosymbionts experiencing strict vertical transmission are hypothesized to be genetically homogenous (clonal) within a host individual.

This study tests the hypothesis that endosymbiont populations are clonal within deep-sea vesicomyid clams. Specifically, symbiont population structure is examined within *Vesicomya* sp. clams from the Juan de Fuca Ridge (referred to here as *Vesicomya* sp. JdF). Phylogenetic analysis of multiple loci previously showed that *Vesicomya* sp. JdF clams host one of two symbiont lineages: symA and symB (Stewart *et al.*, 2008). These symbionts both cluster unambiguously within the monophyletic vesicomyid symbiont clade, sharing 98.8% nucleotide identity across rRNA-coding genes (16S + 23S) and ~90% nucleotide identity across protein-coding genes (in Stewart *et al.*, 2008). However, symA is present in the majority of clams examined to date (>95%) and shows a general pattern of co-speciation with its host, suggesting that it is the ancestral symbiont lineage for this host species (Peek *et al.*, 1998; Goffredi *et al.*, 2003; Stewart *et al.*, 2008). In contrast, symB clusters with the symbiont of a diver-

gent vesicomyid species (*Vesicomya* sp. MAR, sharing ~99% nucleotide identity across two protein-coding loci) and shows no evidence of symbiont–host co-speciation. This pattern indicates that symB entered the *Vesicomya* sp. JdF clam lineage via a lateral (non-vertical) symbiont acquisition event (Stewart *et al.*, 2008). Subsequent to this event(s), symB presumably has been maintained in the host population via vertical transfer, the predominant symbiont transmission strategy for vesicomyid clams (Cary and Giovannoni, 1993; Peek *et al.*, 1998; Stewart *et al.*, 2008). It is hypothesized here that the lateral acquisition of symB involved displacement of the resident symbiont lineage (symA), perhaps over a period in which both genotypes occurred as a mixed population in the same host matriline. If this hypothesis is true, and depending on the rate at which a symbiont genotype is brought to fixation within a host, it is possible that both symA and symB may still co-occur in some *Vesicomya* sp. JdF individuals.

Prior work confirmed the presence of symA and symB symbionts among individuals of *Vesicomya* sp. JdF clams, but did not test for both symbionts in the same host individual. The dominant symbiont in 118 *Vesicomya* sp. JdF clams was characterized previously by direct sequencing of the symbiont 16S rRNA gene (Stewart *et al.*, 2008), which occurs as a single copy in vesicomyid symbionts (Kuwahara *et al.*, 2007; Newton *et al.*, 2007). This work detected the symA genotype in 116 of the 118 individuals; the remaining two clams (clams 2 and 64 in this study, Table 1) hosted symB. However, as the primers used in Stewart and colleagues (2008) did not discriminate between symA and symB symbionts, direct sequencing likely recovered only the dominant bacterial genotype in each clam. Thus, low-frequency minority genotypes, if present, may not have been detected by this analysis. Here, this study tests for the co-occurrence of symA and symB in the same host individual by obtaining a deep-coverage estimate of endosymbiont diversity within *Vesicomya* sp. JdF clams.

This work takes advantage of the sequencing depth made possible by amplicon pyrosequencing (Roche 454 Life Sciences technology). The 454 platform is a powerful alternative to traditional clone library-based assessments of microbial diversity. Using massively parallel pyrosequencing, a single 454 run can generate >200 Mb of sequence data, increasing by orders of magnitude the ability to detect rare but potentially functionally important genetic variants, while avoiding potential bias associated with clone library construction (Sogin *et al.*, 2006; Huber *et al.*, 2007; Dethlefsen *et al.*, 2008). This technology increasingly is being used to quantify the remarkable depth of genetic diversity in natural microbial populations, but has so far not been applied to studies of endosymbiont diversity. Here, pyrosequencing yielded the first

Table 1. Species identifications and collection sites of vesicomimid clams used in this study.

Host species	Clam No.	Year	Location	Latitude, longitude	Dive ^a
<i>Vesicomima</i> sp. JdF ^b	1–3 ^c	1991	Juan de Fuca (JdF) Ridge, North Endeavour	47–57.4°N, 129–05.9°W	A 2413
	4–35	1995	JdF	47–58.1°N, 129–05.2°W	ATV
	36–39	1999	JdF	47–57.78°N, 129–05.51°W	A 3454
	40–45	1999	JdF	47–57.66°N, 129–05.62°W	A 3456
	46–52	1999	JdF	47–58.11°N, 129–05.24°W	A 3457
	53–86 ^c	1999	JdF	47–57.79°N, 129–05.48°W	A 3459
	87–118	1999	JdF	47–57.73°N, 129–05.55°W	A 3462
<i>Calyptogena magnifica</i>	1	2003	East Pacific Rise (9°N)	09–50.88°N, 104–17.61°W	A 3951
	2	1990	Galapagos Rift	00–48.2°N 86–13.9°W	A 2224

a. Dive: A = DSV Alvin, ATV = ROV Advanced Tethered Vehicle (Dive 95-48-149, High Rise Expedition).

b. *Vesicomima* sp. JdF corresponds to *Vesicomima* sp. mt-II, an unnamed clam species designated according to mitochondrial haplotype by Goffredi and colleagues (2003).

c. Populations in which the symbiont genotype symB was first discovered (Stewart *et al.*, 2008).

deep-coverage characterization of endosymbiont population structure within a host. The results provide empirical support for theoretical models predicting the structure of endosymbiont populations in response to vertical transmission, as well as provide the first evidence of mixed symbiont populations in vesicomimids. The latter has important implications for models of symbiont gene flow, as the sequestration of distinct symbiont genotypes within the same host may facilitate the horizontal exchange of genetic material (i.e. recombination).

Results

Symbiont-specific PCR and direct sequencing

The co-occurrence of symA and symB symbiont DNA was demonstrated in gill extracts from eight *Vesicomima* sp. JdF clams (out of 118 total clams) collected from distinct Juan de Fuca (JdF) vent sites (specimens 1, 2, 3, 54, 59, 64, 111 and 112 in Table 2). These clams were collected over three DSV Alvin dives (in both 1991 and 1999) and processed in the lab on separate days

Table 2. Symbiont genotype diversity in *Vesicomima* sp. JdF clams assessed via pyrosequencing of two symbiont loci.

Specimen ^b	Direct sequence 16S genotype ^c	Symbiont-specific ITS screen ^d	Number of sequences ^a							
			16S				ITS			
			Total	A	B	Other	Total	A	B	Other
1	A	A/B	1837	1834	3	0	643	642	1	0
2	B	A/B	1223	0	1221	2	542	0	542	0
3	A	A/B	1094	1090	0	4	651	651	0	0
4	A	A	379	379	0	0	465	465	0	0
35	A	A	3234	3233	0	1	2582	2581	0	1
54	A	A/B	1571	1571	0	0	1166	1166	0	0
59	A	A/B	2529	2528	0	1	1740	1740	0	0
64 ^e	B	A/B	514	22	492	0	181	16	165	0
85	A	A	1120	1119	0	1	1190	1190	0	0
97	A	A	904	904	0	0	648	647	0	1
111	A	A/B	622	621	0	1	212	212	0	0
112	A	A/B	889	889	0	0	762	762	0	0
Cmag1	Rmag	na	4880	0	0	19	3372	0	0	0
Cmag2	Rmag	na	3810	0	0	2	2326	0	0	0

a. Number of sequences generated by 454 pyrosequencing of variable regions of the symbiont 16S rRNA gene (170 nt fragment) and the internal transcribed spacer (ITS; 200 nt fragment); primers listed in Table S1; A or B = sequences with > 99% nucleotide identity to symA or symB genotypes; other = sequences with < 99% nucleotide identity to symA or symB, or to the symbiont of *Calyptogena magnifica* (Cmag) clams; Rmag = *Ruthia magnifica*, the *C. magnifica* symbiont; BLAST results for 'other' sequences are in Table S2.

b. Specimen: *Vesicomima* sp. JdF clams selected for pyrosequencing (see text), as numbered in Table 1.

c. Direct sequence 16S genotype: symbiont genotype determined as either symA (A) or symB (B) in Stewart *et al.* (2008) by direct sequencing of the 16S rRNA gene using universal primers.

d. ITS screen: symbiont genotype detected via PCR and direct sequencing of the symbiont ITS using symA- and symB-specific primers (see Table S1); A/B indicates detection of both symA and symB genotypes.

e. Pyrosequencing results for additional gill replicates ($n = 4$) from specimen 64 are in Table 3.

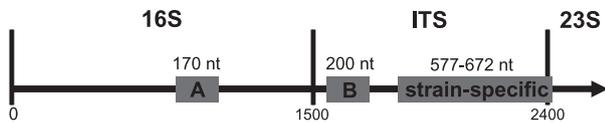


Fig. 1. Genetic markers targeted for vesicomid endosymbiont diversity analyses. The schematic shows a portion of the bacterial ribosomal rRNA operon. 'Strain-specific' marks the region of the internal transcribed spacer (ITS) analysed via strain-specific PCR and direct sequencing to test for the presence of *symA* and *symB* symbiont genotypes in *Vesicomya* sp. JdF clams. Fragments of the 16S rRNA gene (V5 region; box A) and the ITS (box B) were analysed via 454 pyrosequencing.

(Table 1). For each of these clams, PCR using strain-specific primers targeting the internal transcribed spacer (ITS) of the bacterial ribosomal RNA operon (Fig. 1 and Table S1) successfully yielded both *symA* and *symB* gel bands; all other individuals showed only *symA* bands. The identity of each PCR product (*symA* or *symB*) was then confirmed by direct sequencing. The eight clams showing evidence of a mixed symbiont population were selected for in-depth analysis by amplicon pyrosequencing.

Amplicon pyrosequencing read number

Massively parallel pyrosequencing can rapidly sequence 1000s to 10 000s of individual PCR amplicons, a method comparable in power to high-throughput direct sequencing of cloned PCR products but with substantially reduced monetary and labour costs per read generated (Huse *et al.*, 2008). If applied at high coverage levels, both pyrosequencing and clone library analysis can uncover sequence variants at low frequency in the symbiont population (Sogin *et al.*, 2006; Dethlefsen *et al.*, 2008). This provides a clear benefit over the direct sequencing of PCR product, which typically yields data for only the dominant sequence in the amplicon pool. Here, pyrosequencing of two loci, a 170 nt fragment spanning the V5 region of the 16S rRNA gene and a 200 nt fragment of the ITS (Fig. 1), generated a total of 41 185 full-length sequence reads across 14 clams (Table 2). These clams include the 8 *Vesicomya* sp. JdF clams that tested positive for a mixed symbiont population using symbiont-specific PCR and direct sequencing (described above), 4 *Vesicomya* sp. JdF clams (specimens #4, 35, 85, 97) with no evidence of a mixed population and 2 *Calyptogena magnifica* clams (control). Following the removal of sequences with ambiguous bases (N's), read number dropped by only 0.24%, yielding 41 086 sequences, with read number per clam (per locus) varying from 181 to 4880 (mean: 1467; Table 2). For all clams, read number per clam was higher for the 16S locus than for the ITS.

SymA–*symB* co-occurrence

Pyrosequencing confirmed the co-occurrence of *symA* and *symB* symbionts in two *Vesicomya* sp. JdF clams: clam 1 collected from the JdF in 1991, and clam 64 collected from the JdF in 1999 (Table 2). Consistent with direct sequencing results (Stewart *et al.*, 2008), 454 sequencing recovered *symA* as the dominant genotype of clam 1. However, *symB* was also detected in this individual, constituting approximately 1 in 600 reads (0.16%) generated for both the 16S and ITS loci (Table 2). This pattern was reversed for clam 64, in which *symB* dominated the sequence reads and *symA* occurred as the minority genotype. Relative to clam 1, the minority genotype of clam 64 (*symA*) occurred at a higher frequency, representing 4% of the 16S reads and 9% of the ITS reads (Table 2).

However, direct visual evidence (e.g. fluorescent *in situ* hybridization) is necessary to definitively show that *symA* and *symB* colocalize in the same host tissue or within the same host cells. The mixed populations reported here could theoretically be produced if *symA* and *symB* bacteria also occur in free-living forms that occasionally become associated with the gill surface but are not part of the intracellular symbiont community. Several lines of evidence argue against this hypothesis and suggest that *symA* and *symB* are indeed vesicomid symbionts. First, multilocus sequence analysis has definitively clustered these taxa within a monophyletic clade of intracellular vesicomid symbionts, to the exclusion of their free-living relatives (Peek *et al.*, 1998; Goffredi *et al.*, 2003; Stewart *et al.*, 2008; Stewart *et al.*, 2009). Second, recent genome sequencing of two related symbionts suggests that the vesicomid symbiont genome (~1.1 Mb) has experienced some level of size reduction (Kuwahara *et al.*, 2007; Newton *et al.*, 2007), a feature characteristic of other obligately endosymbiotic bacteria (Pérez-Brocá *et al.*, 2006; Gómez-Valero *et al.*, 2007). While larger than that of other endosymbionts, the reduced genome of vesicomid symbionts suggests adaptation to the intracellular environment and the potential for limited growth outside the host; however, this hypothesis requires further testing.

Finally, to confirm our results, we replicated the analysis using DNA extracted from four additional tissue sections (reps 2–5 in Table 3) sampled from both the left and right gills of clam 64 (adequate tissue was not available to repeat these analyses for clam 1). Positive results showing co-occurrence using symbiont-specific PCR and direct sequencing (of the ITS, as described above) were first obtained for all gill replicates (data not shown). Additionally, pyrosequencing of 16S and ITS amplicons generated from these tissue tissues (as described above) showed co-occurring *symA* and *symB* sequences (Table 3). The frequency of the dominant genotype

Table 3. Symbiont genotype diversity in gill sections from *Vesicomya* sp. JdF clam 64 assessed via 454 pyrosequencing.

Rep ^b	Number of sequences ^a							
	16S				ITS			
	Total	A	B	Other	Total	A	B	Other
1	514	22 (4.3%)	492 (95.7%)	0	181	16 (8.8%)	165 (91.2%)	0
2	1360	12 (0.9%)	1341 (98.6%)	7	457	11 (2.4%)	446 (97.6%)	0
3	1622	10 (0.6%)	1607 (99.1%)	5	901	6 (0.7%)	895 (99.3%)	0
4	1191	255 (21.4%)	934 (78.4%)	2	319	94 (29.5%)	225 (70.5%)	0
5	1361	46 (3.4%)	1311 (96.3%)	4	349	17 (4.9%)	331 (94.8%)	1

a. Number of sequences generated by 454 pyrosequencing of variable regions of the symbiont 16S rRNA gene (170 bp fragment) and the ITS (200 bp fragment); primers listed in Table S1; A or B = sequences with > 99% nucleotide identity to symA or symB genotypes; other = sequences with < 99% nucleotide identity to symA or symB; BLAST results for 'other' sequences are in Table S2.

b. rep = gill tissue replicate from clam 64 (see Table 1); rep 1 corresponds to the original tissue section analysed (in Table 2); reps 1–3 were taken from the right gill, reps 4–5 were taken from the left.

(symB) ranged from 71% to 99% of the sequences (Table 3), indicating that the minority genotype (symA) was a non-trivial fraction of the bacterial population in this clam. The frequencies of symA and symB sequences in these replicates were tightly correlated between the two loci (ITS and 16S; $r^2 = 0.99$; Table 3). However, the variance observed among replicates (Table 3) suggests the possibility that strain proportions may differ depending on location in the gill. These results indicate that DNA corresponding to both symA and symB genotypes is broadly distributed across the gills of clam 64, suggesting that both genotypes are indeed present throughout the symbiont population.

Endosymbiont homogeneity

For the majority of clams examined here, pyrosequencing yielded remarkably homogenous sequences for both the symbiont 16S and ITS loci. In contrast to the results for clams 1 and 64 (discussed above), pyrosequencing reads for the remaining 10 *Vesicomya* sp. JdF individuals revealed only a single symbiont genotype per clam. For each individual, these reads reflected the genotype detected previously by direct sequencing of the symbiont 16S rRNA gene (Stewart *et al.*, 2008), sharing > 99% nucleotide similarity to either symA or symB (Table 2). Indeed, parsimony networks showing the relatedness of read sequence variants from all 12 *Vesicomya* sp. JdF clams (excluding gill replicates 2–5 for clam 64, Table 3) show that the vast majority (> 99.5%) of all sequence reads were 100% identical to either symA or symB. Of those 0.5% of reads not identical to symA or symB, 56% occurred only once (i.e. were singletons; not shown in Fig. 2), and no single sequence variant occurred at a frequency higher than 0.07% of all reads for a given locus. At such low frequencies, these variants are unlikely to represent genuine components of population variation, as errors introduced during PCR or the 454

sequencing process occur at comparable rates (Huse *et al.*, 2007).

In some clams, a small fraction (0–0.4%) of the total reads constituted sequences with < 99% nucleotide identity to the dominant genotype (see 'other' in Table 2; not pictured in Fig. 2). These sequences did not BLAST to vesicomylid symbionts and likely reflected contamination by free-living bacteria associated with the clam gill (see Table S2 for BLAST results). Finally, analysis of the two *C. magnifica* control samples yielded 8690 (16S) and 5698 (ITS) sequences (Table 2), all of which identified a single symbiont phylotype corresponding to *Ruthia magnifica*, the previously characterized symbiont of this host species (Newton *et al.*, 2007). No symA or symB sequences were detected among the *C. magnifica* reads, suggesting a lack of cross-contamination during DNA extraction and PCR in preparation for pyrosequencing.

Discussion

Multiple symbiont genotypes in a single host

Prior studies have led to the prevailing hypothesis that individual vesicomylid clams host a single symbiont genotype within their gills. Using direct sequencing, Hurtado and colleagues (2003) examined variation in the symbiont ITS among individuals of the vesicomylid *C. magnifica*, finding 11 ITS genotypes across 80 clams but no evidence for multiple ITS genotypes within a single host. Similarly, studies focusing on a less variable marker (16S rRNA gene) have found no evidence of a mixed internal symbiont population (Kim *et al.*, 1995; Peek *et al.*, 1998; Yamamoto *et al.*, 2002; Goffredi *et al.*, 2003; Stewart *et al.*, 2008). However, these analyses utilized either direct sequencing or clone library analyses involving a small number of sequences (≤ 5), thereby making it difficult to detect potentially rare sequence variants in the endosymbiont population. Here, pyrosequencing analysis

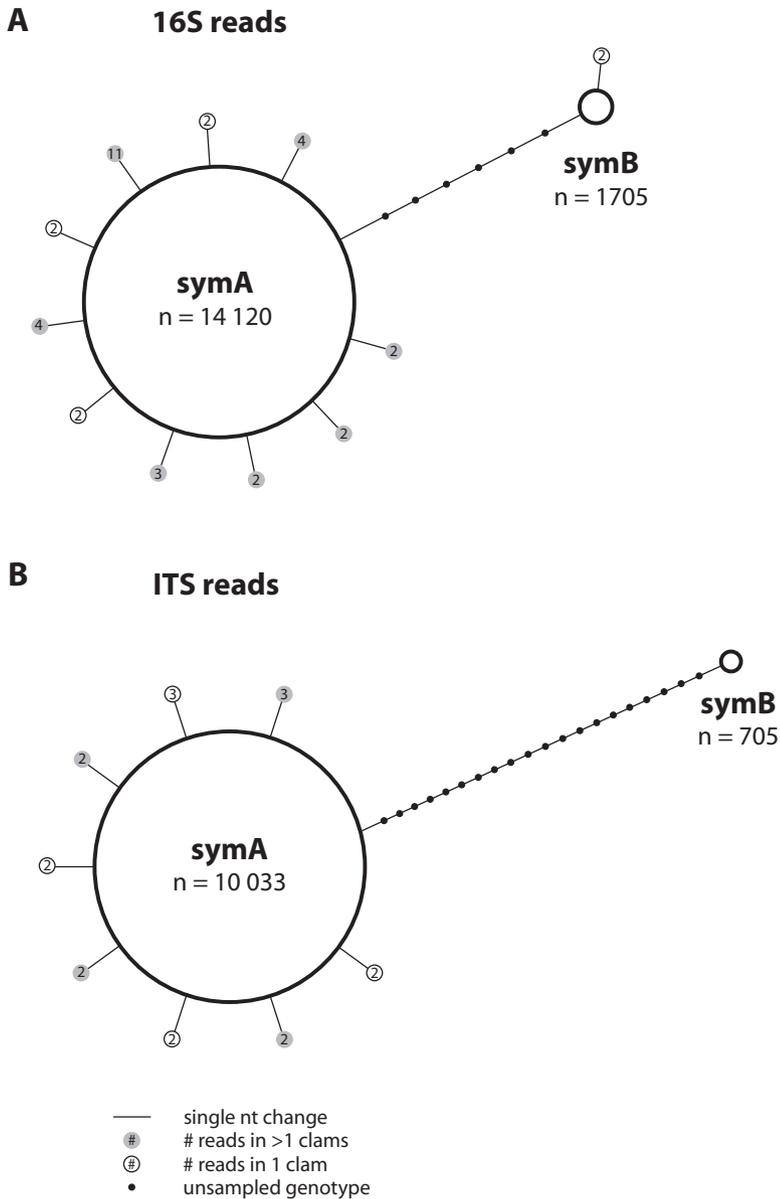


Fig. 2. Parsimony networks inferred from amplicon pyrosequencing of two symbiont loci across 12 *Vesicomya* sp. JdF clams. Networks include (A) 15 861 sequence reads spanning 170 nucleotides of the V5 region of the 16S rRNA gene (16S), and (B) 10 756 sequence reads spanning 200 nt of the internal transcribed spacer (ITS). Genotypes are shown as circles, with connecting lines representing one nucleotide difference and genotype frequencies (the number of reads with 100% identity to each genotype) inscribed within each circle. Small circles extending from symA and symB represent low frequency variants (< 0.5% of all reads) that occurred at least twice in the data sets; the rates at which these were observed are comparable to the rate of errors introduced by pyrosequencing or PCR (Huse *et al.*, 2007). Similarly, homopolymer variations and variants occurring only once in the data sets (i.e. singletons, 45 among all 16S reads, 24 among all ITS reads) were excluded from the analysis, as these could not be distinguished from read length and carry-forward errors introduced during pyrosequencing.

provided the first evidence that individual vesicomyid clams can host multiple symbiont genotypes.

Deep-coverage pyrosequencing made it possible to detect minority symbiont genotypes present at low frequency in the PCR amplicon pools. Here, the co-occurrence of symA and symB symbionts was detected in sequencing reads from two *Vesicomya* sp. JdF clams (out of 12): clam 1 and clam 64. These individuals were collected from separate sites on the JdF Ridge in two years (1991 and 1999), suggesting that the presence of mixed internal symbiont populations may not be a transient characteristic of *Vesicomya* sp. JdF clams. In both individuals, the minority symbiont genotype was present in low abundance. In clam 1, symB DNA represented approximately 1 in 600 reads at both the 16S and

ITS loci. At this frequency (0.16%), symB could easily have escaped detection in clone library analyses that sequence < 100 clones per sample. In contrast, the results for clam 64 suggested that the minority symbiont (symA in clam 64) may represent a sizable fraction of the internal symbiont population, here comprising 0.6–21.4% of the 16S sequences and 0.7–29.5% of the ITS sequences for this individual (replicates 1–5; Table 3). Interestingly, the results from this individual raise the possibility that the relative abundances of symA and symB differ throughout the gill. Replicate studies involving other clams hosting mixed populations will help determine whether this pattern is stochastic (perhaps reflecting either methodological variation or random variation throughout the tissue) or reflects a biological process controlling symbiont distribu-

tion (e.g. host developmental stage, selection acting on variation in symbiont function).

While amplicon pyrosequencing successfully revealed low-abundance sequence variants in this study, the method is not without limitations. First, while pyrosequencing is less expensive per sequence read generated than traditional clone-library analysis (Huse *et al.*, 2008), the method cannot be easily scaled to match the size of the application and therefore may not be cost-effective for researchers working with limited numbers of samples or requiring fewer reads (e.g. < 10 000). Second, as amplicon pyrosequencing is PCR-based, the method may be subject to DNA template-specific bias introduced during the amplification step (e.g. Polz and Cavanaugh, 1998). Therefore, pyrosequencing read frequencies may not accurately reflect the true abundance of microorganisms in mixed communities and should be interpreted with caution. Finally, despite the deep-coverage provided by pyrosequencing, this technique might not be adequate to detect rare variants in all cases. Indeed, six of the eight clams that tested positive for symA–symB co-occurrence via strain-specific PCR and direct sequencing (of the ITS) did not show evidence of co-occurrence by pyrosequencing (Table 2). This raises the possibility that strain-specific PCR, or other sequence-specific methods (e.g. Southern blots), may be equally sensitive to the detection of rare sequence variants; however, due to methodological variation, the absence of detection via PCR cannot unambiguously reflect the absence of a target organism from a sample. Pyrosequencing read number per sample may have to be increased beyond that of this study (mean: ~1500 reads per sample) to fully sample endosymbiont diversity within a host. Nonetheless, our results highlight the efficacy of amplicon pyrosequencing for obtaining high-resolution estimates of bacterial population structure, here providing evidence of mixed internal symbiont populations that may not have been obtained via traditional clone-based techniques.

The co-occurrence of symA and symB lineages suggests that lateral symbiont acquisition may create opportunities for mixed symbiont populations in vesicomids. Multilocus phylogenies show that symB is closely related to the symbiont of a non-*Vesicomya* sp. JdF clam (*Vesicomya* sp. MAR), suggesting that symB was acquired laterally (non-vertically) by *Vesicomya* sp. JdF hosts (Stewart *et al.*, 2008). Direct evidence of lateral transfer would come from the discovery of symB in other vesicomid host species co-occurring with *Vesicomya* sp. JdF; to date, however, such species have not been characterized and more comprehensive sampling of JdF vesicomid clams is warranted. As demonstrated with the *Wolbachia* endosymbionts of insects (e.g. Viljakainen *et al.*, 2008), lateral symbiont acquisition in vesicomids may lead to a mixed internal symbiont

population in which the invading symbiont lineage (e.g. symB) co-occurs with the resident symbiont lineage (e.g. symA) in the same host. Here, assuming that symA and symB are intracellular symbionts and are functionally and ecologically equivalent within a host (i.e. selectively neutral, an unfounded assumption at this point), genetic drift during subsequent maternal transfer events is hypothesized to randomly fix either symA or symB in descendant host lineages. In contrast, if symA and symB play independent roles within the host, it may be possible that mixtures of both genotypes will be maintained across successive maternal transfer events. Indeed, the vertical co-transmission of functionally divergent endosymbionts has been shown in other symbioses, including those of certain sap-sucking insects (e.g. sharpshooters; Takiya *et al.*, 2006; Moran *et al.*, 2008). Given the relatively high genetic similarity of symA and symB (~98.8% nucleotide identity across 16S and 23S rRNA genes; Stewart *et al.*, 2008) and the fact that both genotypes can be found individually in non-mixed populations within a host (Table 2), the assumption that these bacteria perform the same (or a very similar) functional role in the symbiosis is not unfounded. While additional sampling is needed to clarify how symA and symB are being sorted in JdF clams, the results presented here demonstrate that the specificity of the symbiont–host relationship is not absolute in vesicomids, and that a single host can accommodate two genetically divergent symbiont lineages.

The detection of two symbiont genotypes in *Vesicomya* sp. JdF clams is consistent with a growing number of studies showing multiple coexisting chemosynthetic symbionts within the same marine host. For example, strain-level variation has been observed within populations of sulfur-oxidizing symbionts in deep-sea mussels and vestimentiferan tubeworms (DeChaine *et al.*, 2006; Vrijenhoek *et al.*, 2007; Harmer *et al.*, 2008), both of which have been shown genetically or via ultrastructural analyses to acquire their symbionts laterally (Won *et al.*, 2003; Nussbaumer *et al.*, 2006). The co-occurrence of symbionts differing above the strain-level is also well documented (Dubilier *et al.*, 2008). Notably, several species of *Bathymodiolus* mussels simultaneously host both methanotrophic and thioautotrophic *Gammaproteobacteria* (Distel *et al.*, 1995; review in Stewart *et al.*, 2005), and gutless marine oligochaetes can host a consortium of species, including species of sulfate-reducing *Deltaproteobacteria* and sulfide-oxidizing *Gammaproteobacteria* (Dubilier *et al.*, 2001; Blazejak *et al.*, 2006; Ruehlend *et al.*, 2008). For many such associations, expanded molecular and experimental analyses should be applied to determine the relative roles that selective forces, neutral evolutionary processes (e.g. genetic drift) and lateral symbiont acquisition play in creating and maintain-

ing symbiont diversity within a host (Stewart *et al.*, 2008; review by Dubilier *et al.*, 2008).

Mixed endosymbiont populations in *Vesicomya* sp. JdF and other host taxa may create opportunities for genomic exchange (recombination) between symbiont strains. Vertically transmitted intracellular symbionts presumably experience few opportunities for recombination with divergent symbiont lineages or free-living bacteria (Bordenstein and Reznikoff, 2005). However, the likelihood of recombination should increase when divergent symbionts become sequestered for a period of time within the same host individual. Indeed, high levels of recombination have been reported for *Wolbachia* endosymbionts occurring as a mixed population within their ant hosts (Reuter and Keller, 2003). The co-occurrence of symA and symB symbionts in *Vesicomya* sp. JdF clams suggests the possibility that these strains may have exchanged genetic material. Indeed, multilocus sequencing and multiple statistical tests suggest that divergent vesicomid symbionts, including symA and symB, have undergone homologous recombination (Stewart *et al.*, 2009). Detection of recombination in these bacteria suggests that vesicomid symbionts, despite an obligate intracellular lifestyle, can evolve through the acquisition of foreign genetic material.

Clonality of vesicomid symbionts

This study empirically confirms a fundamental prediction of endosymbiont theory, that vertical transmission leads the internal symbiont population to clonality. In theory, vertical transmission forces endosymbionts through population bottlenecks at the start of each host generation, at which point only a subset of symbiont diversity is sampled to found the new population (Mira and Moran, 2002). Repeated passage through such bottlenecks enhances the role that genetic drift plays relative to selection in shifting genotype frequencies within the population (Frank, 1994; Moran, 1996; Wernegreen, 2002; 2005). New symbiont genotypes, resulting from mutation, recombination or perhaps lateral symbiont transfer into a host lineage, are therefore quickly lost or fixed in the population due to chance alone. This process should ultimately result in genetic homogeneity of the internal symbiont population and strict symbiont–host co-speciation, particularly in lineages, in which strict vertical transmission precludes the influx of novel genetic material via recombination or in which strong selection pressure for the maintenance of symbiont heterogeneity is lacking (Peek *et al.*, 1998; Frank *et al.*, 2002; Bordenstein and Reznikoff, 2005). Surprisingly, however, few studies have tested these predictions with empirical analyses of endosymbiont population diversity. Here, amplicon pyrosequencing at deep-coverage levels enabled testing of the clonality hypothesis.

The vesicomid endosymbiont population is remarkably homogenous within the majority of hosts. For each *Vesicomya* sp. JdF clam (excluding clams 1 and 64; see above), all of the sequence reads could be collapsed into a single symbiont genotype (> 99% cluster) for both the ITS and 16S loci. Indeed, $\geq 99.5\%$ of all reads were 100% identical to either symA or symB, with deviation from a perfect match due to rare, single nucleotide substitutions that could not be reliably distinguished from errors introduced during PCR or pyrosequencing (Fig. 2). This homogeneity was somewhat unexpected. Given the depth of the sequence coverage and the hypervariable nature of the ITS marker (Stewart and Cavanaugh, 2007), it would not have been surprising to have found ITS variation among the 12 *Vesicomya* sp. JdF clams, as these individuals were collected from 4 JdF vent sites over a collection period spanning 9 years (Tables 1 and 2). Furthermore, it would not be improbable to have detected low- to middle-frequency symbiont genotypes that arose due to mutations occurring during the lifespan of the host clam, similar to the somatic mutations that arise in mitochondrial genomes (Rand, 2001). Rather, the homogeneity observed here indicates that the internal vesicomid endosymbiont population is effectively clonal, with little to no influx of divergent phylotypes from the external environment. This result suggests that, despite recent evidence of lateral symbiont transfer among hosts, maternal transmission is the dominant strategy in vesicomids and, as such, likely exerts a strong force on symbiont genome evolution and population structure.

Conclusions

The powerful new technology of 454 pyrosequencing enabled a deep-coverage characterization of within-host endosymbiont population structure. Here, pyrosequencing provided the first evidence that individual vesicomid hosts can harbour multiple symbiont strains. However, the majority of clams hosted a clonal symbiont population consistent with a strict, or nearly strict, vertical transmission strategy. Given the paucity of molecular data for other vesicomid symbiont strains, it is unclear to what extent these patterns can be extrapolated to other taxa. The results presented here suggest that endosymbiont population structure is more complex than previously thought, underscoring the potential for new technology to challenge our assumptions of endosymbiont evolution.

These results suggest several additional studies. With regard to *Vesicomya* sp. JdF hosts, the spatial distribution of divergent strains within a clam gill should be examined. Although precluded in this study due to a lack of appropriately preserved tissue, *in situ* hybridizations with symbiont-specific probes would help to determine whether symA and symB symbionts co-occur together in

the same host cell, or even in the same host vacuole. Characterizing the symbiont population at an ultrastructural level might provide insight into whether divergent symbiont lineages perform different functions within the host, as well as into the potential for co-occurring bacteria to exchange genetic material (recombine). Furthermore, additional host clams should be tested for the presence of mixed internal symbiont populations. The *Vesicomysidae* comprises at least 50 species (Goffredi *et al.*, 2003; Kojima *et al.*, 2004; Krylova and Sahling, 2006), the vast majority of which have not been characterized with regard to symbiont population diversity.

Experimental procedures

Vesicomysid clam specimens

This study analyses *Vesicomys* sp. clams collected from the JdF Ridge in the north-eastern Pacific. Based on prior phylogenetic analyses of clam mitochondrial loci (Stewart *et al.*, 2008), these clams correspond to the clam lineage described as *Vesicomys* sp. mitochondrial type II by Goffredi and colleagues (2003), and as *Calyptogena pacifica* by Peek and colleagues (1998); as the taxonomy of this species is in flux, these clams are named *Vesicomys* sp. JdF throughout the text.

A total of 118 *Vesicomys* sp. JdF clams were collected from seven hydrothermal vent sites on the JdF Ridge using the submersibles *DSV Alvin* or *ROV Advanced Tethered Vehicle* (Table 1); these clams correspond to those characterized by direct sequencing (symbiont 16S rRNA gene) in Stewart and colleagues (2008). Additionally, *C. magnifica* clams ($n = 2$), collected from the East Pacific Rise (9°N) and the Galapagos Rift, were used as controls in this analysis. Clams were frozen whole (−80°C) until dissection in the laboratory (excluding *C. magnifica* specimens, which were dissected aboard ship). To avoid cross-contamination of tissues during dissection, sterile disposable dissecting trays (weigh boats) and razors were replaced between specimens, and forceps were washed in ethanol and flame-sterilized twice between specimens. Small sections (~20 mg) of symbiont-containing gill tissue were excised from each clam, rinsed with distilled water to help remove gill-associated free-living bacteria, and used for DNA extraction following the protocol of the DNeasy tissue kit (QIAGEN, Valencia, CA).

Symbiont-specific PCR and direct sequencing

PCR with strain-specific primers was used to test for the co-occurrence of both symA and symB in individual *Vesicomys* sp. JdF clams. PCR primers specific to either symA or symB were developed based on a variable fragment of the symbiont ITS region (Fig. 1; Table S1). To select ITS primers, a broad region (~4000 nt) of the bacterial ribosomal rRNA operon from nine vesicomysid symbiont taxa was direct sequenced using primer sets based on the genomes of two vesicomysid symbionts, *R. magnifica* and *Vesicomysosocius okutanii* (the symbionts of host clams *C. magnifica* and *C. okutanii* respectively; GenBank NC008610 and

NC009465; Kuwahara *et al.*, 2007; Newton *et al.*, 2007). The resulting sequences, spanning ~1200 nt of the 16S rRNA gene, the entire ITS region (variable lengths among taxa), and ~2000 nt of the 23S rRNA gene, were aligned automatically in ClustalX (Thompson *et al.*, 1997), and then edited manually in MacClade 4.0 (Maddison and Maddison, 2000). Based on this alignment, symA- and symB-specific primer sets were designed to amplify 577 nt (for symB) or 672 nt (for symA) of the ITS (Fig. 1; Table S1; GenBank FJ830599 and FJ830600). Forward primers were specific to either symA or symB, differing from other strains in the 9-taxon alignment by at least 3 nt. Both symA- and symB-specific primer sets utilized the same reverse primer (embedded in the 23S rRNA gene; Fig. 1).

Using symA- and symB-specific primer sets, the 118 *Vesicomys* sp. JdF clams described above were screened via PCR for the presence of symA and symB symbiont genotypes. PCR was performed in 50 µl reactions containing 1 U of a high fidelity polymerase (Herculase II Fusion, Stratagene), 1× Herculase PCR buffer, 2 mM Mg²⁺, 0.5 mM dNTP and 0.4 µM each primer. Reactions were run twice using both 1 and 3 µl DNA extract. Following an initial denaturation of 2 min at 92°C, reactions were run for 30 cycles of 25 s at 92°C, 25 s at 50°C, 90 s at 72°C, followed by a final extension of 3 min at 72°C. Resulting products were visualized by gel electrophoresis, purified with a QIAquick kit, and direct sequenced bidirectionally using the symA- and symB-specific PCR primer sets and traditional BigDye technology.

454 pyrosequencing

Massively parallel pyrosequencing (Roche 454 Life Sciences technology) of PCR amplicons was used to characterize intrahost symbiont diversity and to test for the co-occurrence of symA and symB genotypes within *Vesicomys* sp. JdF clams. Pyrosequencing was applied to a total of 14 clams: 8 *Vesicomys* sp. JdF clams that tested positive for a mixed symbiont population using the ITS PCR screen (described above), 4 *Vesicomys* sp. JdF clams with no evidence of a mixed population and 2 *C. magnifica* clams (control).

Using the 9-taxon alignment described above, PCR primers were designed to amplify two fragments of the rRNA operon having varying levels of phylogenetic resolution: a 170 nt fragment spanning the V5 region of the 16S rRNA gene, and a 200 nt fragment of the ITS (boxes A and B in Fig. 1; Table S1; GenBank FJ830601–FJ830604). Sequence variation at either marker can be used to distinguish among the nine taxa in the reference alignment. However, of the two loci, the fragment of the hypervariable ITS provides greater phylogenetic resolution, facilitating discrimination of symbiont genotypes that might differ at the population level (e.g. Stewart and Cavanaugh, 2007). Primers for both markers were designed based on conserved stretches flanking the variable target regions. Forward and reverse primers targeting the 16S V5 fragment blasted with 100% identity to a broad range of Gram-negative and Gram-positive bacteria, thereby ensuring amplification across divergent genotypes, including non-symbiotic, free-living bacteria potentially associated with the clam gills. The forward primer used to amplify the ITS region is embedded within the 23S rRNA gene and targets a similarly broad range of bacterial groups. In contrast, the reverse ITS primer lies within the

hypervariable ITS region itself, blasting with 100% identity to a more narrow set of taxa, including vesicomid clam and bathymodolid mussel symbionts and to uncultured environmental clones. Relative to the 16S V5 primer set, the ITS primer set is unlikely to amplify sequences from distantly related gill-associated bacteria.

A tagged sequencing strategy using 5'-nucleotide barcodes was implemented to facilitate the parallel processing of multiple samples (as in Binladen *et al.*, 2007; Huber *et al.*, 2007). Briefly, the 16S and ITS forward primers were modified by the 5'-addition of 14 unique 10-mer barcodes, with each barcode corresponding to one of the 14 clams analysed (Table S1). Barcodes enable the bioinformatic identification of sequences corresponding to a specific sample, thereby allowing PCR amplicons from multiple reactions to be combined in a single pyrosequencing run. To reduce the likelihood of misidentification, barcodes were designed to contain no homopolymers and to differ from other barcodes by at least two bases (<http://www.roche-applied-science.com>; Huber *et al.*, 2007). Finally, 19-mer sequences corresponding to either the 454 Life Sciences A Adapter (for forward primers) or B Adaptor were fused to each primer (Table S1); these sequences are used to immobilize single amplicons onto DNA capture beads during the emulsion PCR (emPCR) step of pyrosequencing.

Barcoded primers were used to PCR amplify 16S and ITS fragments from 14 clam samples. PCR was performed in 50 μ l reactions using a high fidelity polymerase as described above; reactions were run for only 17 cycles to minimize errors introduced during amplification. PCR products were visualized by gel electrophoresis, purified with a QIAquick kit, then quantified spectrophotometrically (NanoDrop 1000, Thermo Scientific), standardized to 100 ng μ l⁻¹ and pooled into a single sample. The pooled sample was used as template for emulsion PCR and pyrosequencing at the Environmental Genomics Facility (EnGenCore, U. South Carolina), with the sample constituting 1/8 of the template in a single LR70 run on a Genome Sequencer FLX instrument (Roche 454 Life Sciences).

Data analysis

A schematic of the 454 data analysis process (showing data from clam 64) is provided as Fig. S2. Briefly, sequence reads that passed the FLX internal quality filters were combined into a single SFF output file. Reads corresponding to distinct loci and samples were then binned according to barcode + primer combinations using the C+ program Barcode Sort, yielding a total of 28 output data sets (FASTA format), representing 14 clams and 2 loci. These data sets were edited by trimming off barcodes and primers from all reads and removing any reads with ambiguous bases (N's). Following editing, BLASTClust (NCBI) was used to cluster sequences into 100% identity groups and a Perl script was used to parse the output. For each data set, representatives from each 100% identity group were aligned automatically in ClustalX and then edited manually in MacClade 4.0. To assess whether homopolymer variation or single-base indels, the most common sources of error in 454 pyrosequencing (Huse *et al.*, 2007), could be reliably counted as genuine sources of variation, the program GSamplicon was used to evaluate read integrity using flowgram

data from the pyrosequencing runs. Based on these results, sequences differing by homopolymer variation or single-base indels were collapsed manually into consensus sequences in MacClade, yielding aligned reads differing only by substitution-level differences. These alignments were used to assess within-clam symbiont diversity and to construct genotype networks, using parsimony criteria implemented in the TCS program (Clement *et al.*, 2000).

DNA sequencing for symA and symB generated via ITS screening and pyrosequencing can be found under GenBank accession numbers FJ830599-FJ830604.

Acknowledgements

We thank Bob Vrijenhoek and two anonymous reviewers for their generous and constructive input during the preparation of this manuscript. We thank Katie Scott and Jonathan Robinson for collecting vesicomids, Li Tao for providing the program Barcode Sort, and Rob Young and Irene Newton for support in editing the pyrosequencing results. We also thank the expert crews of the research vessels and submersibles that made this work possible. This work was supported by National Science Foundation grants EF-0412205 and OCE-0453901 awarded to C. Cavanaugh, the Genetics and Genomics Training Program (GGT) at Harvard University, and by a seed grant from the Environmental Genomics Facility (EnGenCore; Joe Jones, Director) at the University of South Carolina.

References

- Binladen, J., Gilbert, M.T.P., Bollback, J.P., Panitz, F., Bendixen, C., Nielsen, R., and Willerslev, E. (2007) The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS ONE* **2**: e197.
- Blazejak, A., Kuever, J., Erséus, C., Amann, R., and Dubilier, N. (2006) Phylogeny of 16S rRNA, ribulose 1,5-bisphosphate carboxylase/oxygenase, and adenosine 5'-phosphosulfate reductase genes from gamma- and alphaproteobacterial symbionts in gutless marine worms (*Oligochaeta*) from Bermuda and the Bahamas. *Appl Environ Microbiol* **72**: 5527–5536.
- Bordenstein, S.R., and Reznikoff, W.S. (2005) Mobile DNA in obligate intracellular bacteria. *Nature Rev Microbiol* **3**: 688–699.
- Cary, S.C., and Giovannoni, S.J. (1993) Transovarial inheritance of endosymbiotic bacteria in clams inhabiting deep-sea hydrothermal vents and cold seeps. *Proc Natl Acad Sci USA* **90**: 5695–5699.
- Clement, M., Posada, D., and Crandall, K.A. (2000) TCS: a computer program to estimate gene genealogies. *Mol Ecol* **9**: 1657–1659.
- DeChaine, E.G., Bates, A.E., Shank, T.M., and Cavanaugh, C.M. (2006) Off-axis symbiosis found: characterization and biogeography of bacterial symbionts of *Bathymodiolus* mussels from Lost City hydrothermal vents. *Environ Microbiol* **8**: 1902–1912.
- Dethlefsen, L., Huse, S., Sogin, M.L., and Relman, D.A. (2008) The pervasive effects of an antibiotic on the human

- gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol* **6**: e280.
- Distel, D.L., Lee, H.K., and Cavanaugh, C.M. (1995) Intracellular coexistence of methano- and thioautotrophic bacteria in a hydrothermal vent mussel. *Proc Natl Acad Sci USA* **92**: 9598–9602.
- Dubilier, N., Mülders, C., Ferdelman, T., de Beer, D., Penthaler, A., Klein, M., *et al.* (2001) Endosymbiotic sulphate-reducing and sulphide-oxidizing bacteria in an oligochaete worm. *Nature* **411**: 298–302.
- Dubilier, N., Bergin, C., and Lott, C. (2008) Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. *Nature Rev Microbiol* **6**: 725–740.
- Duperron, S., Sibuet, M., MacGregor, B.J., Kuypers, M.M.M., Fisher, C.R., and Dubilier, N. (2007) Diversity, relative abundance and metabolic potential of bacterial endosymbionts in three *Bathymodiolus* mussel species from cold seeps in the Gulf of Mexico. *Environ Microbiol* **9**: 1423–1438.
- Duperron, S., Halary, S., Lorion, J., Sibuet, M., and Gaill, F. (2008) Unexpected co-occurrence of six bacterial symbionts in the gills of the cold seep mussel *Idas* sp. (*Bivalvia: Mytilidae*). *Environ Microbiol* **10**: 433–445.
- Frank, A.C., Amiri, H., and Andersson, S.G. (2002) Genome deterioration: loss of repeated sequences and accumulation of junk DNA. *Genetica* **115**: 1–12.
- Frank, S.A. (1994) Kin selection and virulence in the evolution of protocells and parasites. *Proc R Soc Lond B* **258**: 153–161.
- Funk, D.J., Wernegreen, J.J., and Moran, N.A. (2001) Intraspecific variation in symbiont genomes: bottlenecks and the aphid-*Buchnera* association. *Genetics* **157**: 477–489.
- Goffredi, S.K., Hurtado, L.A., Hallam, S., and Vrijenhoek, R.C. (2003) Evolutionary relationships of deep-sea vent and seep clams (*Mollusca: Vesicomidae*) of the 'pacifica/lepta' species complex. *Mar Biol* **142**: 311–320.
- Gómez-Valero, L., Silva, F.J., Simon, J.C., and Latorre, A. (2007) Genome reduction of the aphid endosymbiont *Buchnera aphidicola* in a recent evolutionary time scale. *Gene* **389**: 87–95.
- Harmer, T.L., Rotjan, R.D., Nussbaumer, A.D., Bright, M., Ng, A.W., DeChaine, E.G., and Cavanaugh, C.M. (2008) Free-living tubeworm endosymbionts found at deep-sea vents. *Appl Environ Microbiol* **74**: 3895–3898.
- Huber, J.A., Mark Welch, D., Morrison, H.G., Huse, S.M., Neal, P.R., Butterfield, D.A., and Sogin, M.L. (2007) Microbial population structures in the deep marine biosphere. *Science* **318**: 97–100.
- Hurtado, L.A., Mateos, M., Lutz, R.A., and Vrijenhoek, R.C. (2003) Coupling of bacterial endosymbiont and host mitochondrial genomes in the hydrothermal vent clam *Calyptogena magnifica*. *Appl Environ Microbiol* **69**: 2058–2064.
- Huse, S.M., Huber, J.A., Morrison, H.G., Sogin, M.L., and Mark Welch, D. (2007) Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol* **8**: R143.
- Huse, S.M., Dethlefsen, L., Huber, J.A., Mark Welch, D., Relman, D.A., and Sogin, M.L. (2008) Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PLoS Genet* **4**: e1000255.
- Kadar, E., Bettencourt, R., Costa, V., Santos, R.S., Lobo-Da-Cunha, A., and Dando, P. (2005) Experimentally induced endosymbiont loss and re-acquirement in the hydrothermal vent bivalve *Bathymodiolus azoricus*. *J Exp Mar Biol Ecol* **318**: 99–110.
- Kim, Y.W., Yasuda, M., Yamagishi, A., Oshima, T., and Ohta, S. (1995) Characterization of the endosymbiont of a deep-sea bivalve, *Calyptogena soyae*. *Appl Environ Microbiol* **61**: 823–827.
- Kojima, S., Fujikura, K., and Okutani, T. (2004) Multiple trans-Pacific migrations of deep-sea vent/seep-endemic bivalves in the family *Vesicomidae*. *Mol Phylogenet Evol* **32**: 396–406.
- Krylova, E.M., and Sahling, H. (2006) Recent bivalve molluscs of the genus *Calyptogena* (*Vesicomidae*). *J Mollus Stud* **72**: 359–395.
- Kuwahara, H., Yoshida, T., Takaki, Y., Shimamura, S., Nishi, S., Harada, M., *et al.* (2007) Reduced genome of the thioautotrophic intracellular symbiont in a deep-sea clam, *Calyptogena okutanii*. *Curr Biol* **17**: 881–886.
- Maddison, D.R., and Maddison, W.P. (2000) *Macclade 4: Analysis of Phylogeny and Character Evolution*. Sunderland, MA, USA: Sinauer Associates.
- Mira, A., and Moran, N.A. (2002) Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria. *Microb Ecol* **44**: 137–143.
- Moran, N.A. (1996) Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proc Natl Acad Sci USA* **93**: 2873–2878.
- Moran, N.A. (2007) Symbiosis as an adaptive process and source of phenotypic complexity. *Proc Natl Acad Sci USA* **104**: 8627–8633.
- Moran, N.A., McCutcheon, J.P., and Nakabachi, A. (2008) Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet* **42**: 165–190.
- Narra, H.P., and Ochman, H. (2006) Of what use is sex to bacteria? *Curr Biol* **16**: R705–R710.
- Newton, I.L.G., Woyke, T., Auchtung, T.A., Dilly, G.F., Dutton, R.J., Fisher, M.C., *et al.* (2007) The *Calyptogena magnifica* chemoautotrophic symbiont genome. *Science* **315**: 998–1000.
- Nussbaumer, A.D., Fisher, C.R., and Bright, M. (2006) Horizontal endosymbiont transmission in hydrothermal vent tubeworms. *Nature* **441**: 345–348.
- Peek, A.S., Feldman, R.A., Lutz, R.A., and Vrijenhoek, R.C. (1998) Cospeciation of chemoautotrophic bacteria and deep-sea clams. *Proc Natl Acad Sci USA* **95**: 9962–9966.
- Pérez-Brocá, V., Gil, R., Ramos, S., Lamelas, A., Postigo, M., Michelena, J.M., Silva, F.J., Moya, A., and Latorre, A. (2006) A small microbial genome: the end of a long symbiotic relationship? *Science* **314**: 312–313.
- Polz, M.F., and Cavanaugh, C.M. (1998) Bias in template-to-product ratios in multitemplate PCR. *Appl Environ Microbiol* **64**: 3724–3730.
- Rand, D.M. (2001) The units of selection on mitochondrial DNA. *Annu Rev Ecol Syst* **32**: 415–448.
- Reuter, M., and Keller, L. (2003) High levels of multiple *Wolbachia* infection and recombination in the ant *Formica exsecta*. *Mol Biol Evol* **20**: 748–753.
- Ruehlend, C., Blazejak, A., Lott, C., Loy, A., Erséus, C., and Dubilier, N. (2008) Multiple bacterial symbionts in two species of co-occurring gutless oligochaete worms from

- Mediterranean sea grass sediments. *Environ Microbiol* **10**: 3404–3416.
- Sogin, M.L., Morrison, H.G., Huber, J.A., Mark Welch, D., Huse, S.M., Neal, P.R., *et al.* (2006) Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *Proc Natl Acad Sci USA* **103**: 12115–12120.
- Stewart, F.J., and Cavanaugh, C.M. (2007) Intragenomic variation and evolution of the internal transcribed spacer of the rRNA operon in *Bacteria*. *J Mol Evol* **65**: 44–67.
- Stewart, F.J., Newton, I.L.G., and Cavanaugh, C.M. (2005) Chemosynthetic endosymbioses: adaptations to oxic–anoxic interfaces. *Trends Microbiol* **13**: 439–448.
- Stewart, F.J., Young, C.R., and Cavanaugh, C.M. (2008) Lateral symbiont acquisition in a maternally transmitted chemosynthetic clam endosymbiosis. *Mol Biol Evol* **25**: 673–687.
- Stewart, F.J., Young, C.R., and Cavanaugh, C.M. (2009) Evidence for homologous recombination in intracellular chemosynthetic clam symbionts. *Mol Biol Evol* (in press).
- Takiya, D.M., Tran, P.L., Dietrich, C.H., and Moran, N.A. (2006) Co-cladogenesis spanning three phyla: leafhoppers (*Insecta: Hemiptera: Cicadellidae*) and their dual bacterial symbionts. *Mol Ecol* **15**: 4175–4191.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl Acids Res* **25**: 4876–4882.
- Viljakainen, L., Reuter, M., and Pamilo, P. (2008) *Wolbachia* transmission dynamics in *Formica* wood ants. *BMC Evol Biol* **8**: 55.
- Vrijenhoek, R.C., Duhaime, M., and Jones, W.J. (2007) Subtype variation among bacterial endosymbionts of tube-worms (*Annelida: Siboglinidae*) from the Gulf of California. *Biol Bull* **212**: 180–184.
- Wernegreen, J.J. (2002) Genome evolution in bacterial endosymbionts of insects. *Nat Rev Genet* **3**: 850–861.
- Wernegreen, J.J. (2005) For better or for worse: genomic consequences of intracellular mutualism and parasitism. *Curr Opin Genet Dev* **15**: 572–583.
- Won, Y.J., Hallam, S.J., O'Mullan, G.D., Pan, I.L., Buck, K.R., and Vrijenhoek, R.C. (2003) Environmental acquisition of thiotrophic endosymbionts by deep-sea mussels of the genus *Bathymodiolus*. *Appl Environ Microbiol* **69**: 6785–6792.
- Yamamoto, H., Fujikura, K., Hiraishi, A., Kato, K., and Maki, Y. (2002) Phylogenetic characterization and biomass estimation of bacterial endosymbionts associated with invertebrates dwelling in chemosynthetic communities of hydrothermal vent and cold seep fields. *Mar Ecol Prog Ser* **245**: 61–67.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Schematic of the protocol used to analyse 454 pyrosequencing reads.

Table S1. Primers used for symbiont-specific PCR screening and 454 pyrosequencing.

Table S2. Top BLAST hits to non-vesicomid-symbiont bacterial DNA sequences recovered by pyrosequencing.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.