Supplemental Materials and Methods

Collections

Eggs were procured by inducing adult *S. velum* to spawn with temperature cycling [1]. In brief, seawater for spawning was collected from the *S. velum* habitat at the time of collection and filtered through a 0.2 µm nitrocellulose filter, and two one liter volumes were heated to 4°C and 25°C. *Solemya velum* were then alternated between cold and warm water baths for 30 min each, starting with cold water. This spawning method produced sperm much less consistently than eggs, precluding embryological work. For qPCR, eggs were rinsed in 0.2 µm filtered seawater (FSW) and half were surface sterilized for 1 min in 1% sodium hypochlorite in FSW, while the remaining half were left untreated to control for the bleach treatment. Eggs were then separated into 1, 3, 4, or 50 per 1 ml tube, flash frozen in a dry ice-96-100% ethanol slurry, and stored at -80ºC until DNA extraction. Seawater used to spawn *S. velum* was sampled before and after clams were placed in the water by filtering 120 ml through 0.2 µm nitrocellulose filters (Millipore, Billerica, MA, USA), and the filters were stored at -80ºC. To test the specificity of *S. velum* symbiont-specific probes, trophosome tissue from the tube worm *Riftia pachyptila*, which contains chemosynthetic symbionts, was obtained from a formaldehyde-fixed specimen and prepared for light microscopy.

Sediment samples were also collected from locations offshore of Georgia, USA (Figure 2B and Table S1). No *S. velum* were found in multiple sediment grabs and cores taken from the continental shelf in the sampled areas, making this a potential negative control for marine sediment that should not contain the S. velum symbionts unless they are widespread, and potentially free-living. A few grams were collected from sediment samples taken from the continental shelf floor using a corer or a grab, placed in cryotubes, and frozen at -20ºC. For extraction, approximately 0.25 g was taken from each tube quickly, as soon as enough thawing had occurred, and put in a PowerSoil (MoBio) tube for DNA extraction.

Light microscopy

Gross morphology hematoxylin and eosin (HE) staining: Specimens were deparaffinized with xylene and hydrated through an ethanol series. Following two washes in distilled water, sections were stained with Mayer's hematoxylin (EMS, Pennsylvania, USA 26043) for 30 sec with agitation, rinsed, and stained in 1% eosin Y (Electron Microscopy Supplies, Hatfield, PA, USA) for 10-30 sec with agitation. Sections were then dehydrated through an ethanol series, cleared in xylene (VWR, Radnor, PA, USA), and mounted in Canada balsam (Sigma Aldrich, Missouri, USA).

Specific Marker Design (SMD) workflow:

We designed a novel workflow, named SMD, that works by blasting the genome of the taxon of interest, one gene at a time, against a database of sampled bacterial sequences, and sorts and filters them according to find genes that are well-sampled, but highly divergent from other bacterial sequences (outlined in Figure S1). If population-level genetic data is available for the taxon of interest, this can be used with the previous data to also find genes that have high conservation within the taxon of interest, but low conservation to other bacterial sequences.

The scripts used to perform SMD are available at from the GitHub page for this manuscript at https://github.com/shelbirussell/SMD_workflow. BLAST needs to be installed locally and the nucleotide database of choice (*e.g.,* the nonredundant nucleotide database was used here) needs to be downloaded. As input, a genome consisting of a fasta sequence and general feature file (gff) containing the annotation for the taxon of interest are required. Optionally, population-level variation can be input as a variant call format (vcf) file. First, the fasta sequence and gff annotation are read and parsed to extract the sequence for each gene in the genome. Next, each gene is blasted against the database [2],

recording all hits with e-values less than 1e-4. Blast results are then parsed and filtered to retain genes with a minimum of 100 hits with at most 80% average percent identity to sequences not from the taxon of interest. The remaining genes are then sorted and output by descending number of hits and increasing average percent identity. These values were selected to return confident alignments with reasonably good representation in the database, but were fairly permissive to ensure we had enough candidate genes to test. Using these settings, 509,162 hits were returned for the 2,741 tested genes in the *S. velum* genome. If given a vcf file, this candidate marker gene output can also be sorted by increasing number of variant sites within each gene. The output format is a tab-delimited file reporting the gene id, number of hits, average percent identity, and (optionally) the number of variant sites. From this, the user can select genes starting from the top of the file (*i.e.,* genes optimizing: maximum hits, minimum % identity, minimum population-level variation) for primer design and testing.

Design of S. velum symbiont specific markers with SMD: Candidate markers were designed using the *S. velum* symbiont genome [3] and population genetic variant data [4], returning 853 genes out of 2741 passing the filtering cutoffs. The top eight candidates were distinctly better than the remaining candidates, and so were selected for primer design with NCBI primer blast [5]. These genes were: JV46_03550, JV46_09960, JV46_12670, JV46_12710, JV46_14840, JV46_15080, JV46_18880, JV46_21670. The thermodynamics for each primer set were checked and optimized with the Integrated DNA Technologies (IDT) Oligoanalyzer (www.idtdna.com). One primer set for each locus was ordered from IDT. See Supplemental Table S2 and Table 3 for primer information.

Mock symbiont-containing sediment test sample for SMD: First, the minimum detectable dilution of symbiont-containing gill DNA was ascertained with PCR by creating and amplifying a dilution series from 1:100 to 1:100000 gill DNA in dH_2O, utilizing the 16S-ITS primer set from [6]. A 1:2000 dilution (final 0.04 ng/µl) was found to be the optimal minimum concentration reliably detected by conventional PCR. Next, a symbiont-free sediment sample was collected from the Point Judith, RI mudflat where *S. velum* occurs, but distant to an *S. velum* burrow. The sediment DNA contained plenty of amplifiable bacterial DNA, however, when repeatedly tested with symbiont 16S-ITS and *gyrB* primers, in addition to mitochondrial *atp6* primers, no template amplified (n = 3 each), giving us confidence that no symbiont DNA was present. Gill DNA was spiked 1:2000 into this symbiont-free sediment DNA extract (final concentration of 0.04 ng/µl), and mixed thoroughly by vortexing. This spiked sample was used for primer testing to assess the specificity of symbiont primers in the presence of other bacterial sequences.

PCR primer testing and validation: Each primer set was tested on DNA extracted from the symbiont-containing gills at the melting temperature provided by IDT minus 2°C to ensure that a single band amplified from this natural mixture of host and symbiont DNA. For all end point PCR assays, either AccuStart HiFi Taq (Quanta Biosciences, Beverly, MA, USA) or EconoTaq (Lucigen, Middleton, WI, USA) was used. Reactions using AccuStart were performed with an initial denaturation at 94ºC for 1 min, followed by 30 cycles of 94ºC for 20 sec, primer set annealing temperature for 30 sec, and 68ºC for 1.5 min Reactions using EconoTaq with an initial denaturation at 94ºC for 2 min, followed by 30 cycles of 94ºC for 30 sec, primer set annealing temperature for 30 sec, and 58ºC for 1 min, followed by a final extension at 72ºC for 10 min. For initial tests of specificity, 1 µl (approximately 80 ng) of *S. velum* gill DNA was used in a reaction volume of 50 µl containing 0.5 µM of each primer and 200 µM dNTPs. The PCR products were analyzed by gel electrophoresis. Six of eight primer sets exhibited single bands of the appropriate size. Next, primer specificity was assessed by amplifying 1 µl of the sediment sample spiked with 0.04 ng of *S. velum* gill DNA with the primer sets that exhibited single bands across an annealing temperature gradient from 55.5-70°C (12 wells). The PCR products from the highest annealing temperature that successfully amplified were Sanger sequenced and analyzed in

Geneious (Geneious 8.1 (http://www.geneious.com, [7]) to assess specificity. The primer sets that exhibited absolutely clear, unambiguous traces were identified, and the primer set to *rhlE* was selected for use in the subsequent qPCR assays (Table S2).

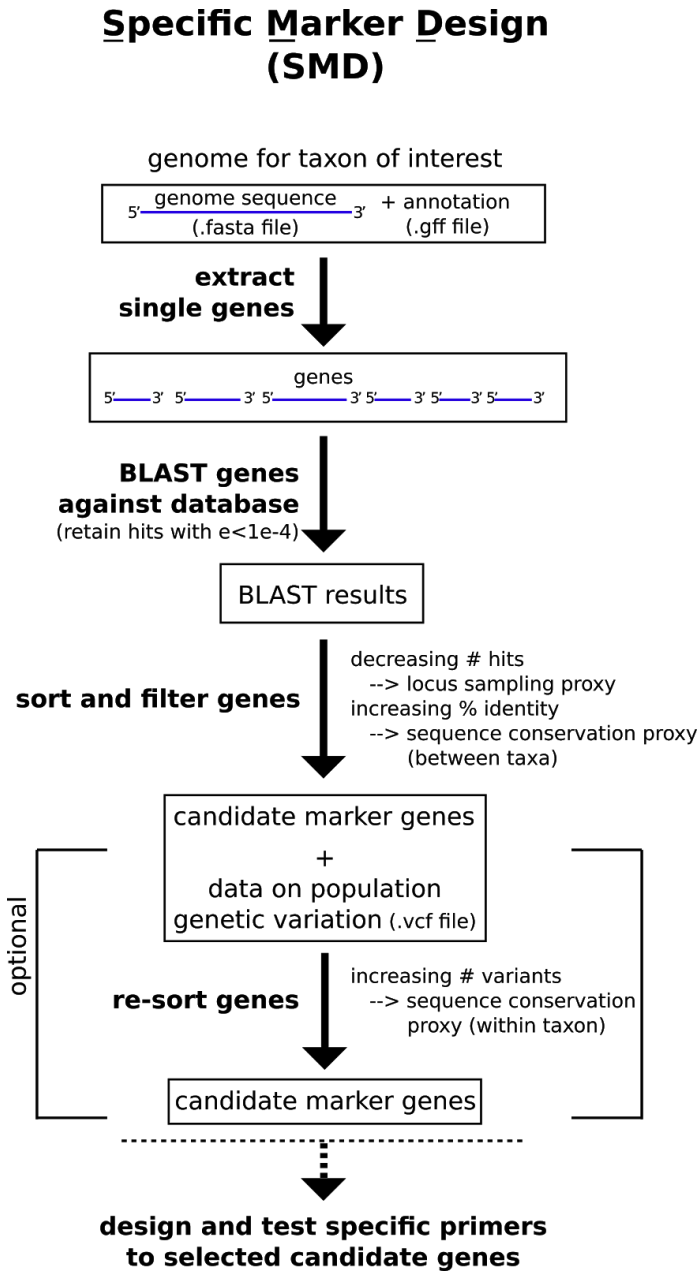

Figure S1. Workflow for the generation of specific marker candidates. First, genes are extracted from the genome supplied by the user for the taxon of interest. Each gene is then blasted against a database (the NCBI nr database, here), retaining all hits with e-values < 1e-4. The blast output is then filtered, removing genes with few hits or high percent identity to non-taxon of interest sequences. The remaining genes are sorted by the number of hits (proxy for extent of locus sampling) and percent sequence identity (proxy for sequence conservation between species), and then output for primer design. If population-level variation is known, output gene lists can be further sorted by the number of variant sites within the population (proxy for sequence conservation within species).

Symbiont detection via PCR

EconoTaq PCR protocol for S. velum symbiont gyrB: For PCR amplification of *gyrB*, EconoTaq (Lucigen, Middleton, WI) was used according to the manufacturer's protocol as follows. An initial denaturation at 94ºC for 2 min, followed by 30 cycles at 94ºC for 30 sec, 65ºC for 30 sec, and 58ºC for 1 min, followed by a final extension at 72ºC for 10 min 4 µl of DNA were used in a reaction volume of 50 µl.

Symbiont quantification via qPCR

Standard curves: Standard curves were made for the *S. velum* symbiont *rhlE* and mitochondrial *atp6* genes, and the bacterial 16S rRNA gene. Gill DNA was amplified with the *rhlE* primer set and a newly designed primer set to *S. velum* mitochondrial *atp6* (Table 3 and Supplemental Figure S2). The *E. coli* K12 16S sequence was amplified with mostly universal primers to the Bacterial V4 region (Table S2). These PCR products were cloned into the pJET1.2/blunt cloning vector (Thermo Fisher Scientific, Waltham, MA, USA) with OneShot Top10 competent cells (Thermo Fisher Scientific). Plasmids were purified with an EZNA plasmid extraction kit (Omega Bio-Tek, Norcross, GA, USA), linearized with PstI restriction enzyme (NEB, Ipswitch, MA, USA), re-purified with an EZNA Cycle Pure PCR cleanup kit (Omega), and quantified by Qubit (Thermo Fisher Scientific). Standard curves were made by serial dilution of linearized plasmids to make a 10-fold series from 1 to $1x10^8$ copies and 4 to $4x10^7$ copies of template per volume added to the qPCR reaction for *rhlE* and *atp6* quantification and 16S rRNA quantification, respectively.

qPCR reactions: qPCR assays were run on a Stratagene Mx3005P using PerfeCTa SYBR Green FastMix (Quanta Biosciences) with 0.3 µM of each primer (10 µM stock concentration) and 4 µl of template DNA in a 20 µl reaction volume. A 3-step cycling protocol with 40 cycles was used (see Quanta). Annealing temperatures were as follows for the three primer sets: 65ºC for *rhlE*, 62ºC for *atp6*, and 55ºC for 16SV4. qPCR data were analyzed with MxPro qPCR software (Stratagene, San Diego, CA).

In situ hybridization (ISH)

Probe testing/optimization: Hybridization stringency was optimized for the Svsym47 probe set by testing hybridization temperatures (35°C and 46°C) and formamide concentrations (20, 30, 40, 50, and 60%). Specific binding was achieved at the lower temperature of 35°C at an intermediate formamide concentration of 50%.

Control probes: Control probes consisted of two positive controls, EUB338 targeting bacterial 16S rRNA and EUK516 targeting eukaryotic 18S rRNAs, and two negative control probes corresponding to the reverse complement of the bacterial (NON338) and symbiont (NON47) 16S probes. See Table 1 for probe sequences. Hybridization was also specific at 35°C and 50% formamide for these probes.

Control binding: The EUB338 and Svsym47 probes were separately tested on tissue sections treated with RNase (2 mg/ml for 15 min) and resulted in no signal, verifying the probe signals are due to RNA-binding (assumed rRNA). *Riftia* trophosome tissue, containing gammaproteobacterial chemosynthetic symbionts, and *E. coli* OneShot competent cells attached to a filter were hybridized with EUB338 and Svsym47 probes, producing signal for EUB338, but not Svsym47, indicating that the designed probes are reasonably specific to the *S. velum* symbionts. Sets of alternating slides containing 5-10 sections were used for the test probe, positive control probes, and negative control probes.

Choice of probe labels: Amplified chromogenic and fluorescent approaches were used because autofluorescence was quite strong in *S. velum* tissues, even after quenching aldehyde groups with glycine (data not shown). Thus, monolabeled FISH for direct detection of FITC probes produced insufficient signal to be discernable from background autofluorescence. Furthermore, autofluorescence even made it difficult to detect HRP-amplified fluorescent signal in tissues exhibiting low signal intensity due to low probe target density (*e.g.,* gill in Figure S3A-C). Chromogenic detection avoided these background fluorescence issues, allowing more sensitive detection of low abundance probe targets (*e.g., S. velum* 16S rRNA in non-gill tissues (Figure 3D-G))[8].

In situ hybridization (CARD-FISH and CISH): For both catalyzed reporter deposition fluorescence *in situ* hybridization (CARD-FISH) and chromogenic *in situ* hybridization (CISH), tissue sections were deparaffinized and hydrated followed by permeabilization with 10 mg/ml lysozyme for 30 min at room temperature followed by 3 µg/ml proteinase K for 30 min at 37°C. Endogenous peroxidases were inactivated by treatment with 3% H_2O_2 in 1x PBS for 10 min at room temperature. Endogenous phosphatases were inactivated by treatment with Bloxall (Vector Labs, Burlingame, CA SP-6000) for 15 min at room temperature. Sections were pretreated with hybridization buffer (50% formamide, 0.9 M NaCl, 20 mM Tris HCl pH 8.0, 1% blocking reagent (Roche), 10% w/v dextran sulfate) at 46°C for 20 min, followed by hybridization at 35°C overnight (12-18 hours) with 1:100 probe (final concentration 0.5 ng/µl) in hybridization buffer. Following hybridization, sections were rinsed in wash buffer (5 mM EDTA pH 8.0, 18 mM NaCl, 20 mM Tris HCl pH 8.0, 0.01% w/v SDS) at 37°C for 25 min and in 1xPBS pH 8.0 with agitation for 15 min.

CARD-FISH signal detection: Amplification buffer (1xPBS pH 7.6, 0.1% blocking reagent (Roche), 2 M NaCl, 10% w/v dextran sulfate) was mixed 100:1 with 0.15% H_2O_2 in PBS and 1000:1 with fluorescent tyramides (Alexa 488, 555, or 647). Sections were incubated in the amplification mix at 40ºC for 20 min in a dark humidity chamber and then rinsed in 1xPBS and distilled water. A subset of sections were prepared for multiple labeling as follows: Sections were incubated in 0.1% H_2O_2 for 2 min to inactivate horseradish peroxidase from the previous hybridization, and then hybridization was repeated as described above with a different probe and fluorophore. Labeled sections were counterstained with DAPI (1 µg/ml; 2-(4-amidinophenyl)-1H -indole-6-carboxamidine) and mounted in Mowiol with DABCO (2.5% w/v; 1,4-diazabicyclo-[2,2,2]-octane; Sigma Aldrich).

CISH signal detection: Sections were blocked for a minimum of 20 min in 1xISH Blocking Solution (Vector Labs MB-1220) at 37ºC and then incubated in 1:1000 alkaline phosphatase anti-FITC in 1xISH Blocking Solution. Following antibody binding, slides were washed twice in 100 mM Tris pH 9.5 with agitation for 3 min each. Sections were incubated in NBT/BCIP (5-bromo-4-chloro-3-indolyl-phosphate with nitro blue tetrazolium) as substrate (Vector Labs SK-5400) in a dark humidity chamber overnight (>12 hrs). Slides were then washed in 100 mM Tris pH 9.5 for 5 min. Sections were counterstained with Nuclear Fast Red (EMS 26078-05) for 10 min, dehydrated, and mounted in Canada balsam (Sigma Aldrich).

Analysis: Images were imported into ImageJ using the Bio-Formats plugin for file conversion, scale bar annotation, 3D reconstruction of confocal stacks, and stitching together images of whole *S. velum* with the Grid/Collection stitching plugin. Oocyte measurements were made using the image tool in ImageJ. Images were annotated in the Inkscape vector graphics editor.

Supplemental Results

Morphology and anatomy of *S. velum* (continued)
The right and left halves of the mantle were fused along the posterior end of the shell, opening for the foot approximately halfway along the ventral-anterior axis.

Each ctenidium consisted of many gill filaments (gf) comprised primarily of bacteriocytes (b), intercalary cells (not labeled), and ciliated cells (c, Figure 1B,C).

Depending on the state of maturity and sex, the gonad visually comprised up to 50% of the remaining visceral mass. The aorta and digestive tract run along the dorsal side of the visceral mass and are intercalated in the gonad and foot tissue (data not shown; see [9]). The stomach is highly reduced in this species compared to protobranch bivalves without chemosynthetic gill symbionts but is still clearly distinguishable (st in Figure 1A).

As seen in Figure 1C with hematoxylin and eosin (HE) staining, oogonia (oo) and previtellogenic oocytes (pv) were basophilic whereas vitellogenic (vo) and mature oocytes (mo) were acidophilic. Oogonia (oo) were closely situated against the acinal wall and had distinctive germinal vesicles, which appeared as small, dense nuclei inside of a large transparent nuclear envelope. Previtellogenic oocytes (pv) extended out from the acinal wall and had a smooth to slightly granular cytoplasm. Vitellogenic oocytes (vo) contained numerous lipid-rich vesicles and protruded further from the acinal wall. Mature oocytes (mo) contained darkly pigmented vesicles along the perimeter of the vitelline membrane along with lighter colored lipid-rich vesicles in the cytoplasm. These cells were connected to the acinal wall by a stalk of cytoplasm at the base of the cell and contained large germinal vesicles (data not shown). Thus, the structure of the ovary, and its oocytes that maintain access to somatic cells via the acinal wall, offer access points for symbionts to enter eggs prior to spawning.

Specific primer design
The wealth of genetic data now available for unculturable as well as culturable bacterial taxa enables candidate markers for specific bacteria to be identified *in silico*. In principle, a potential marker will be conserved within the taxon of interest, yet variable outside of it. The only limitation to this approach is the completeness of taxonomic sampling in the queried dataset for each locus being tested. Thus, this attribute must be considered during the marker selection process. Additionally, genetic variation within the taxon of interest should also be considered to ensure that any primers designed amplify from all members of the taxon. Applying this approach. which we are calling Specific Marker Design (SMD) and is outlined in Supplemental Figure S1, we aligned the genome sequence and annotation for the *S. velum* symbiont to the NCBI non-redundant nucleotide database to identify genes that are well sampled from many bacterial taxa and have a low percent identity to non-symbiont sequences.

Using SMD, we successfully identified and designed three specific *S. velum* symbiont primer sets after testing an initial eight sets designed for eight different loci (Supplemental Table S2 and Figure S2), demonstrating the efficacy of this method. Eight loci were initially selected because, after sorting our candidate list by number of hits, average percent identity, and conservation within *S. velum* symbionts, this was the number of genes that were clearly better candidates than the remaining genes on the list. This testing effort proved more than adequate for designing specific primers to the *S. velum* symbionts. We excluded two of the eight primer sets immediately after the first amplifications with gill DNA (for loci JV46_03550 and JV46_12710), as they clearly amplified more than one region of the symbiont genome, based multiple gel electrophoresis bands. The remaining six primer sets were tested on a mock sediment sample intended to mimic the same template complexity of the *S. velum* habitat, which likely

contains close relatives of the symbionts, without the *S. velum* symbionts being present (see Materials and Methods for a full description of how this sample was identified).

Following the addition of 1:2000 *S. velum* gill/symbiont DNA to an aliquot of the symbiont-free sediment DNA, this spiked sample was tested to assess the specificity of the remaining six primer sets. Following amplification, the PCR products exhibited single bands in gel electrophoresis. Upon Sanger sequencing, three of the six sets produced unambiguous chromatograms, indicating that these primers amplified single sequence types even in complex sediment (JV46_09960, also referred to as *rhlE*, JV46_12670, and JV46_18880; see Supplemental Figure S2A,B,C and Table S2). For each of the remaining three primer sets/loci, only one of the two primers produced unambiguous chromatograms (*e.g.,* JV46_14840, JV46_15080, and JV46_21670; Supplemental Figure S2D,E,F and Table S2, and Supplemental Discussion). For downstream analyses, we decided to use the primers to *rhlE* (JV46_12670), as this gene is present once in the genome and the primers produced very robust and specific amplification.

Primer sets were also designed for two more classic bacterial and mitochondrial marker loci, specifically symbiont DNA gyrase subunit B (*gyrB,* aka. JV46_17100) and the host mitochondrially encoded ATP synthase 6 (*atp6*; Table 3 and Supplemental Table S2). These primer sets were tested on gill DNA, symbiont-free sediment DNA, and symbiont-spiked sediment DNA as described above. The symbiont *gyrB* primer set produced an unambiguous Sanger sequence chromatogram (Supplemental Figure S2G) from the forward primer, but not the reverse, so this primer set and sequencing with one primer was only used to confirm results found with the *rhlE* primer set. The host mitochondrial *atp6* primer set produced unambiguous chromatograms with both primers (Supplemental Figure S2H), indicating it is specific for quantification.

Owing to the greater specificity of the *rhlE* primer set and the smaller amplified product compared to the *gyrB* primers, the SvrhlE primer set was used for symbiont qPCR quantification on *S. velum* egg DNA and environmental samples.

qPCR quantification of bacteria and *S. velum* symbionts in environmental samples

Very few samples exclusively amplified either the symbiont sequence (6/87 samples) or mitochondrial (7/87 samples) sequence, whereas 33% (29/87) of samples amplified both. Abundant bacteria were detected in the environmental samples (on the order of expected values of 10^6/ml and 10^9/g [10-12]). The lack of symbiont sequences detected did not appear to be due to low DNA concentration or poor DNA quality, as bacterial abundance was loosely correlated with symbiont abundance (correlation coefficient of 0.11). These results support the presence of small amounts of symbiont DNA in the environment, which may have arisen via host tissue sources (*e.g.,* gametes, larvae, degrading gill tissue, etc.), temporary bacterial persistence in the environment, or perhaps even as a very small free-living/diving population.

None of the sediment samples from continental shelf sites off of Georgia (n = 6) amplified with the symbiont primers, despite having abundant bacterial 16S rRNA, average 2.92 +/- 3.94 x 10^8 copies, and in only one was the mitochondrial *atp6* sequence detected, with 357 copies/ml (Figure 2C,D).

The qPCR products from 10 *rhlE* reactions (8 from sediment and 2 from seawater) and 14 *atp6* reactions (12 from sediment and 2 from water) were Sanger sequenced to confirm that the genotypes detected from these samples are identical to the known symbiont genotypes from each subpopulation (from [4]). The two Shark River Island, NJ sediment samples sequenced both contained the T -> C transition at *rhlE*

position 288 reported in the population genetics data from [4]. One of the two sediment samples from Kingston Bay, MA contained the known G -> A transition at position 598 of *rhlE*. While these are the only two SNPs in the amplified region previously reported from the *S. velum* symbiont population [4], the two Great Bay, NJ samples contained a G -> A transition at position 228 and one contained an A -> T transversion at position 251. The *atp6* gene only had one high frequency variant site within the amplified region, a G -> A transition at position 279 of the gene, which has been found in individuals from all the subpopulations [4]. Consistent with this, 8/14 of the *atp6* sequences contained this substitution. Furthermore, the NC seawater sample contained peaks for both an A and G at this site, indicating that this sample contained both alleles. Interestingly, the PCR product from the one GA sediment sample that amplified with *atp6* primer did match the symbiont sequence and contained the A substitution.

Adult *S. velum in situ* hybridization

The eukaryotic 18S rRNA probe EUK516 was tested on whole-clam sections with CARD-FISH, as a positive control for tissue permeability, confirming that horseradish peroxidase-labeled probes could penetrate, bind, and be detected in the ovary tissue (Figure S3). Gills exhibited a very weak signal despite the fact that they must contain ribosomes for protein synthesis (Figure S3A-C), whereas oocytes, known to contain many ribosomal copies, exhibited a very intense signal (Figure S3D-F).

No probe signal was observed with either of the reverse complement negative control probes, NON338 (Figure 3M-R) and NONSvsym47 (Figure 5H-J). CARD-FISH signals were stronger than those for classical FISH, which uses probes labeled with a single fluorophore (data not shown), and was necessary to overcome background autofluorescence in the tissues (Figure S4A,B). The use of tyramides labeled with Alexa 647 helped to reduce noise because autofluorescence occurred primarily in the blue-green spectrum. Sections labeled chromogenically with chromogenic *in situ* hybridization (CISH) afforded greater sensitivity partially because they did not have the signal-to-noise issues seen in the fluorescently labeled samples.

Sections treated with RNase A failed to bind the Svsym47 probe, again verifying that the probe signal is due to binding symbiont rRNA (Figure S4F,G). The Svsym47 probe was shown to be specific to the *S. velum* symbionts, as it did not bind to the gammaproteobacteria *E. coli* (data not shown) or the symbionts of the deep sea tubeworm *Riftia pachyptila* (Figure S4H,I). Additionally, although no symbiont DNA amplified from male testes previously sampled [13], we examined two male specimens to verify this result. While symbionts were at high density in gill filament bacteriocyte cells of these males (Supplemental Figure S5A,D-F), no symbiont 16S signal was detected in the testes with either CARD-FISH (Supplemental Figure S5B,C) or CISH (Supplemental Figure S5G-I).

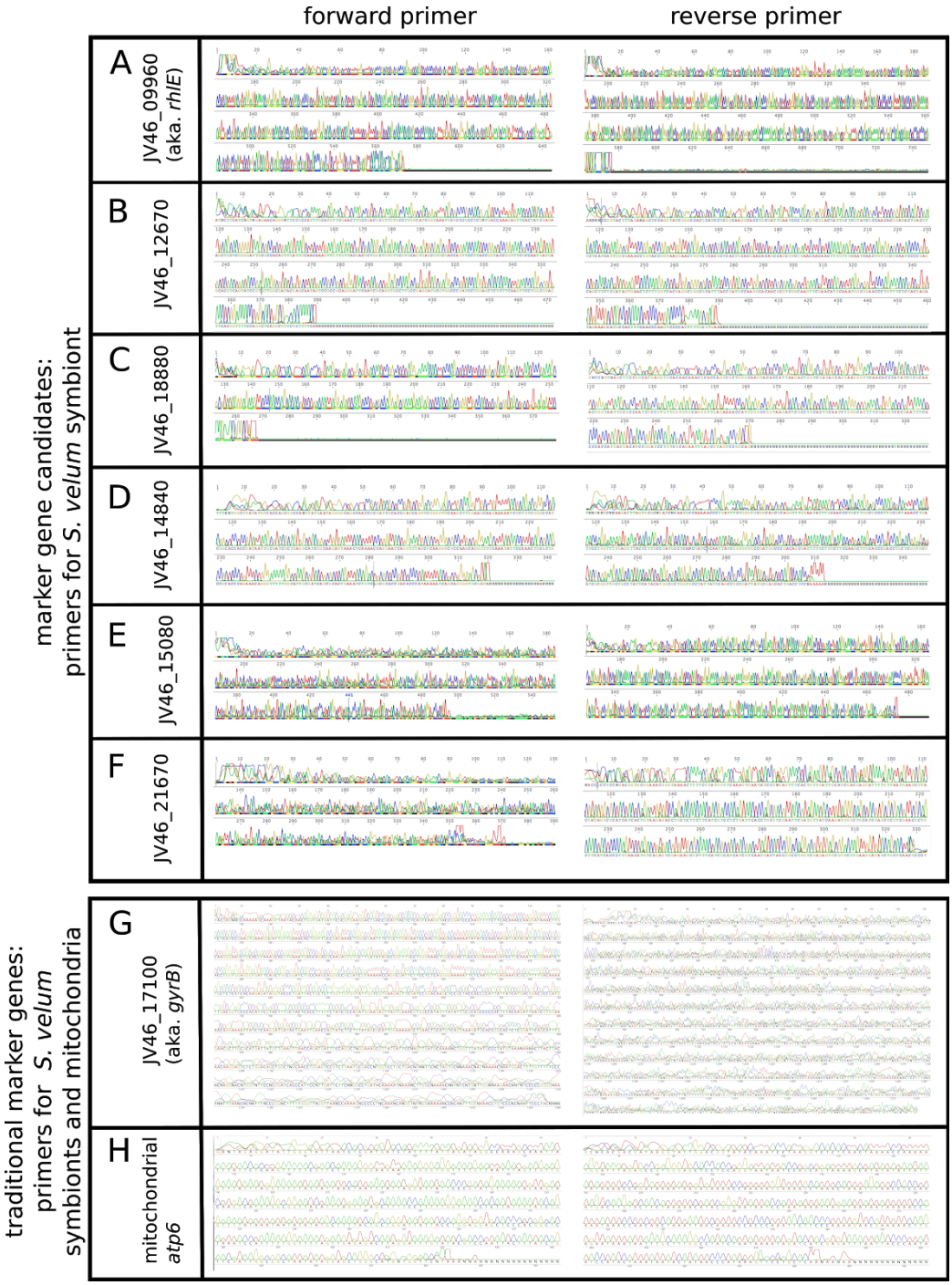

Figure S2. Raw chromatograms from PCR products amplified and sequenced using candidate primer sets designed for the *S. velum* (A-G) symbiont and (H) mitochondria. (A-G) Loci IDs from the reference *S. velum* genome (Dmytrenko *et al.* 2014) are indicated in the far left column (see also Table 3 and Supplemental Table S2). (A-F) Chromatograms for sequences amplified with primers designed for candidate genes identified through the method described in the Methods and Supplemental Figure S1. (G,H) Chromatograms for two additional marker sequences identified previously for symbiont and mitochondrial identification. Chromatograms were imaged in Geneious and exported as screenshots. Base pairs are called from the identity and sequential order of colored peaks along the chromatogram trace (left to right). Relative peak height indicates the intensity of the signal. A clear chromatogram is one for which each base pair has a single, well-defined peak of a single color, indicating that only a

single type of nucleotide is present at each position. Any "N"s after the sequence with no readout are just empty sequence, and those with some low intensity readout is noise after the polymerase finished reading through the amplicon.

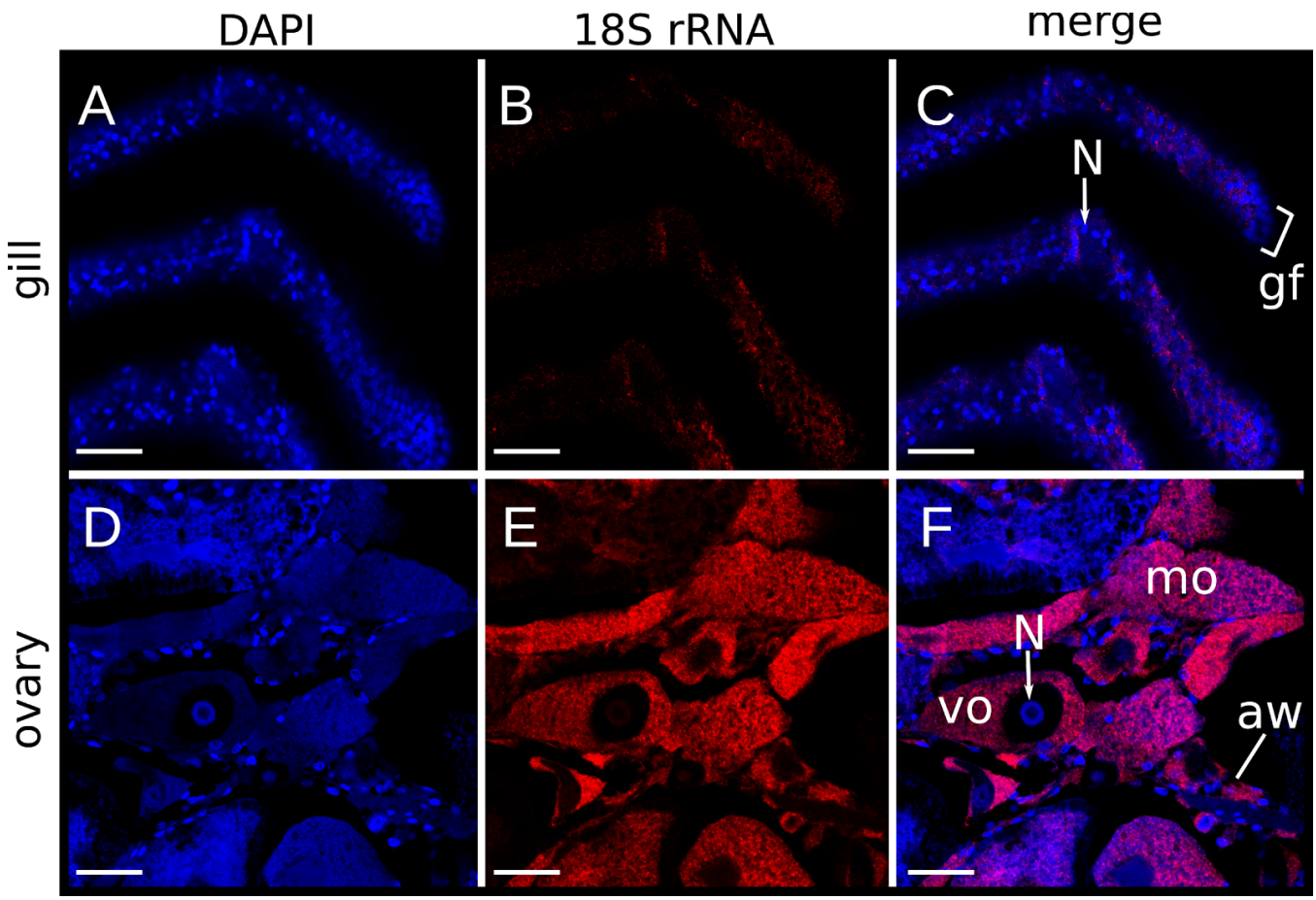

Figure S3. Positive controls for catalyzed reporter deposition fluorescence in situ hybridization (CARD-FISH) permeability via labeling host 18S rRNA in (A-C) *S. velum* gill and (D-F) ovary. Tissues were labeled with the EUK516 probe specific for Eukaryotic 18S rRNA. A-F Columns left through right: (A,D) DAPI DNA staining (blue), (B,E) 18S signal (red, Alexa 647), (C,F) merged images. Abbreviations: (aw) acinal wall, (gf) gill filament, (mo) mature oocyte, (N) nucleus, (vo) vitellogenic oocyte. Scale bars = 50 μm.

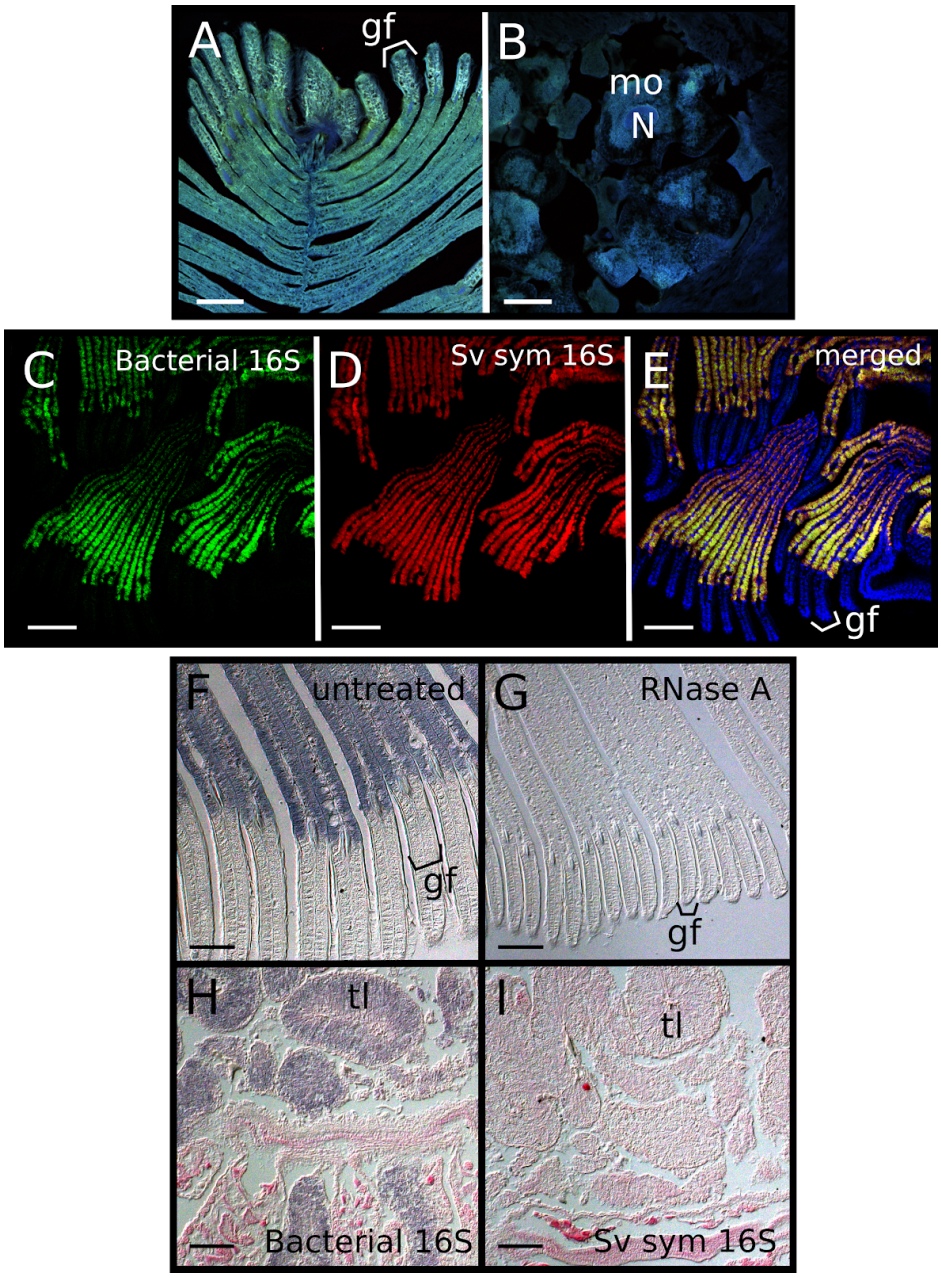

Figure S4. *In situ* hybridization (ISH) controls imaged with (A-E) confocal and (F-I) DIC microscopy. (A,B) Unlabeled (A) gill and (B) ovary tissue autofluorescence under high laser intensity. (C-E) CARD-FISH co-localization of universal bacterial and *S. velum* symbiont-specific probes in gill tissue. (C) EUB338 signal (green, Alexa 555). (B) Svsym47 16S (red, Alexa 647). (C) Merged with DNA stained with DAPI (blue). Overlapping probe signals result in yellow. (F,G) CISH negative controls for Svsym47 binding to rRNA. (F) Svsym47 probe signal (purple, NBT/BCIP) is present in untreated gill filament tissue, but not in (G) RNase A-treated tissue. (H,I) CISH test for symbiont specificity in *Riftia* tubeworm symbiont-containing trophosome tissue, showing that the EUB338 probe (purple in H) binds the chemosynthetic tubeworm symbionts but the symbiont probe Svsym47 does not (no purple in I). DNA counterstained with nuclear fast red in H,I. Abbreviations: (gf) gill filament, (mo) mature oocyte, (N) nucleus, (tl) trophosome lobe. Scale bars = 100 μm.

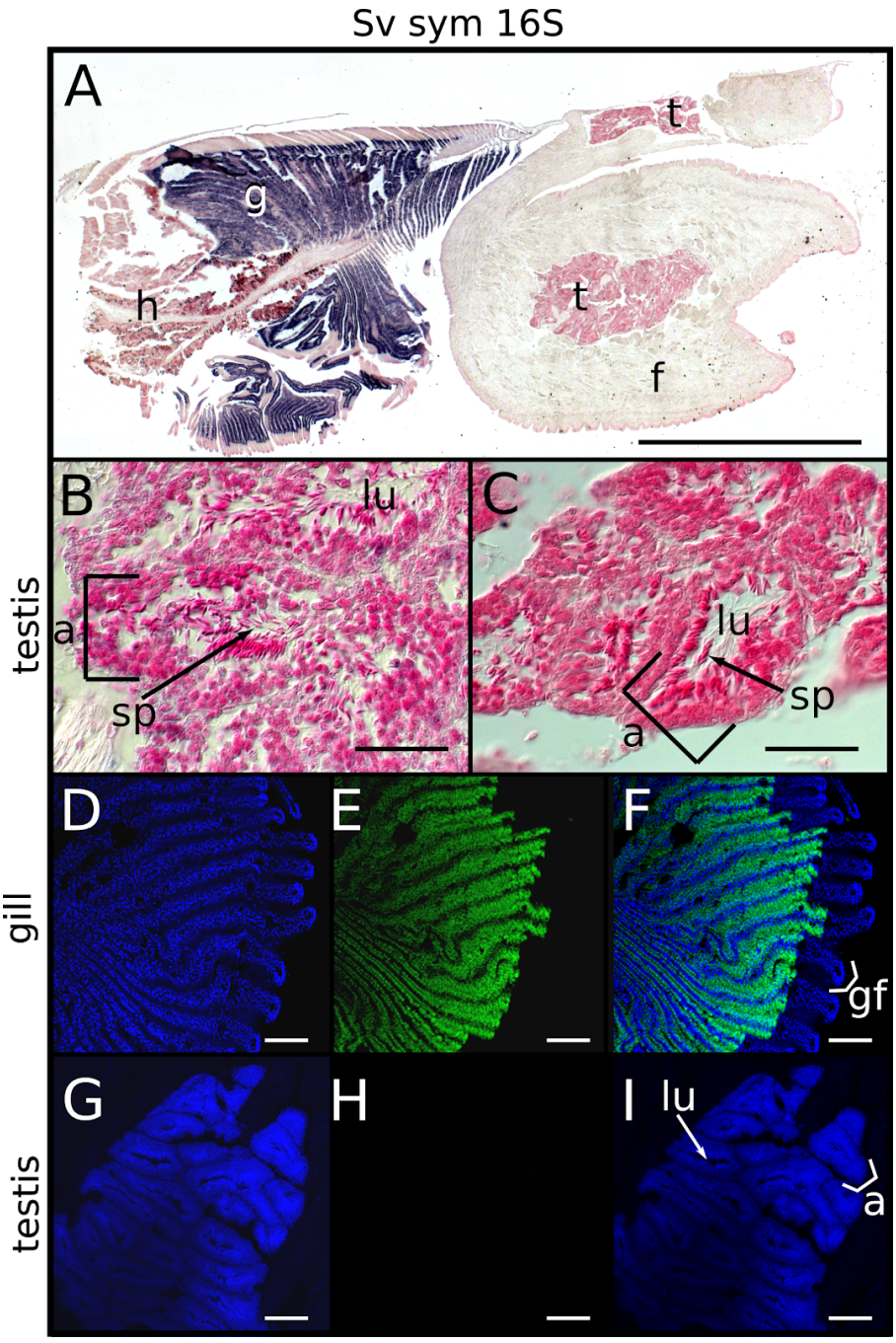

Figure S5. Symbiont 16S rRNA localized to gill tissue of male *S. velum*, but were not found in the testis. (A) Whole *S. velum* removed from shell and (B) testis hybridized with Svsym47 and imaged with DIC microscopy. (C) Testis hybridized with the negative control probe NONSvsym47. Probe signal in purple (NBT/BCIP) and DNA counterstained in pink (nuclear fast red). (D-F) Male *Solemya velum* gill and (G-I) testis labeled with eubacterial probe EUB338 and imaged with confocal microscopy. Left column (D,G): DAPI (blue). Middle column (E,H): EUB338 signal (green, Alexa 488). Right column (F,I): Merged images. Abbreviations: (a) acinus, (f) foot, (g) gill, (h) hypobranchial gland, (lu) lumen, (sp) spermatozoa, (t) testis. Scale bars = (A) 1 mm, (B,C) 50 µm. (D-I) 100 µm.

locality	sediment samples				water samples		
	samples (n)	average wet (g)	standard deviation wet (g)	total wet (g)	samples (n)	average filtered (ml)	total filtered (ml)
Duxbury, MA (42° 0'14.15"N, 70°40'43.79"W)	12	0.19	0.04	2.22	2	130	260
Point Judith, RI (41°22'55.91"N, 71°30'1.10"W)	22 (Power Soil prep)	0.27	0.17	6.01	2	130	260
	1 (Power Max prep)	9.00	na	9.00	8 (100 µm prefilter)	1000	8000
Watch Hill Cove, RI (41° 18' 36" N, 71° 51' 35.9994" W)	8	0.25	0.12	2.04	na	na	na
Shark River Island, NJ (40°11'9.60"N, 74° 1'48.00"W)	3	0.19	0	0.57	1	195	195
Great Bat, NJ (39° 31' 48" N, 74° 19' 47.9994" W)	6	0.18	0.04	1.06	na	na	na
Sinepuxent Bay, MD (38°14'58.56"N, 75° 9'8.06"W)	9	0.18	0.03	1.63	1	225	225
Piver's Island, NC (34°42'59.23"N, 76°40'26.33"W)	3	0.19	0	0.57	1	225	225
Skidaway Island, GA (0 to 98 km offshore) (31° 59' 31.14" N, 81° 0' 20.94" W) (31° 54' 33.3" N, 80° 54' 52.9194" W) (31° 42' 20.82" N, 80° 41' 27.82" W) (31° 28' 18.48" N, 80° 21' 21.78" W)	8	0.19	0	1.14	na	na	na
Total	72			24.24	15		9165

Table S1. Environmental samples collected to test for the presence of *S. velum* symbionts in the *S. velum* habitat, and in other habitats (*i.e.,* coastal sediment off Georgia, USA).

gene name in *S. velum* genome annotation	gene product	amplicon length (bp)	primer name	primer sequence (5'-3')	marker selection criteria				PCR products	
					hits (#)	% identity	Sv symbiont variant sites (#)	gene length (bp)	single band	specific Sanger sequence (F/R)
JV46_09960	DNA and RNA helicase *rhlE* (named primer set SvrhlE in main text)	606	09960-112F	GGACGGGATATCATGGGTGG	1103	71.3	4	1311	y	y/y
			09960-717R	CAGCCAGGAGAGGAGTTCAC						
JV46_12670	response regulator	421	12670-591F	CACGCAAAATGGGGACTTGG	1329	68.4	8	1362	y	y/y
			12670-1011R	GAAAAGAGACGCCTGAGCCT						
JV46_18880	type II protein secretion system E2	302	18880-196F	GCCGGACTACGCTCAATTCT	1030	70.6	7	771	y	y/y
			18880-497R	TCTGCCCGAACAAGTTCACA						
JV46_14840	Na-translocating oxaloacetate decarboxylase OadGAB subunit B2	341	14840-557F	TTTTCGGAGGTCAGTGCTCG	1057	69.3	7	1167	y	y/n
			14840-897R	ATCGCACCCCTGCTGATTTT						
JV46_15080	vacuolar-type H(+)-translocating pyrophosphatase	507	15080-818F	GCGATACCGACACCGTAGAG	1024	73.0	11	2007	y	n/y
			15080-1324R	CTGTGGCTCCATGATCGCTA						
JV46_21670	type IV pilus assembly major pilin protein PilB	375	21670-319F	TCACTTGAGGCGATTGAGCA	1085	69.7	6	1698	y	n/y
			21670-693R	GCCAGTTGTGATAGGGGGTC						
JV46_03550	transketolase	397	03550-211F	CGCCTCGAATGCATCAACAG	1093	69.8	5	1977	n	na
			03550-607R	CGTCTTCACCCACGACTCAA						
JV46_12710	2-oxoglutarate dehydrogenase E1 component	728	12710-304F	GACGAGGGCAACATTCTCCA	1030	68.7	9	2853	n	na
			12710-1031R	TGCTGGACGATGGCTTTCAT						

Table S2. Primer sets and PCR testing results for the top 8 candidate *S. velum* symbiont marker genes generated with our workflow, Specific Marker Design (SMD). The output selection criteria from SMD are reported under the columns marked "marker selection results". The specificity of each primer set that produced a single band when amplified using *S. velum* gill DNA as a template was tested on a symbiont-free environmental sediment sample spiked with a known amount of S. velum gill DNA. The PCR products of these reactions were Sanger sequenced (presented under "PCR results"). Markers and primer sets passing all tests for symbiont specificity are highlighted in green, those that produced single bands yet had contaminated chromatogram traces are in white, and those failing to produce single bands from gill DNA are in red.

See file Table-S3_qPCR_egg_data.xlsx, as the data spreadsheet is too large to display on a single page.
Table S3. qPCR data for eggs, environmental samples, and standard curves.

month	cell type	mean diameter	standard deviation	mumber of measurements	number of specimens
June	oogonia	29.0	4.1	16	3
	pre-vitellogenic oocytes	79.7	24.7	13	3
	vitellogenic oocytes	132.2	26.4	11	3
	mature oocytes	209.2	29.6	19	3
July	oogonia	24.6	6.5	22	3
	pre-vitellogenic oocytes	67.4	14.5	18	3
	vitellogenic oocytes	127.5	19.1	14	3
	mature oocytes	206.8	30.7	18	3
September	oogonia	18.4	4.5	35	4
	pre-vitellogenic oocytes	55.7	11.8	34	4
	vitellogenic oocytes	108.9	17.9	24	4
	mature oocytes	194.5	27.9	29	4
November	oogonia	19.3	4.8	25	3
	pre-vitellogenic oocytes	49.0	12.0	20	3
	vitellogenic oocytes	90.0	na	1	1
	mature oocytes	207.0	49.5	2	2

Table S4. Measurements of *S. velum* oocytes across developmental stages in the summer versus fall. The morphological and histological features described in the text were used to delimit oocyte stages. Values are aggregated from multiple oocytes measured from multiple individuals, as indicated.

Supplemental Discussion

The workflow implemented in this manuscript effectively identifies loci for specific primer design to use in detecting and quantifying bacteria in environmental samples. Only a few candidate loci need to be tested to generate specific primers for assays to detect the presence and abundance of the taxon of interest. This number will vary depending on the divergence from free-living relatives, with more divergent symbionts requiring fewer loci be tested compared to recently host-associated taxa. However, the relatively close relationship of the *S. velum* symbionts to free-living clone isolates (<=95% identity) indicates that this method will even be useful for taxa very recently diverged from their free-living ancestors. This method of identifying loci in a target genome for specific detection among a heterogeneous mixture of sequences will prove useful for many studies of environmental microbiology. In conjunction with recent evidence of horizontal transmission in the *S. velum* symbionts, these results provide additional supporting evidence for the presence of mixed modes of symbiont transmission. This substantiates the idea that vertical transmission through the individual is possible even when horizontal transmission events are evident in the population. Vertical transmission offers the benefit of ensuring that hosts have access to symbiont resources at critical stages of development [14,4]. However, in some associations vertical transmission is not always one hundred percent effective. This results in some offspring not inheriting symbionts, and requires the maintenance of horizontal transmission to supplement missing symbionts (as observed in *Wolbachia* bacteria of *Drosophila* [16]*, Aeromonas* bacteria of leeches, and facultative symbionts of aphids [17]. In addition, given the small number of symbionts associated with spawned eggs (ca. 50-100 symbiont genomes per egg in *S. velum*) relative to other vertically transmitted associations (*e.g.*, [18,19,20,21]), horizontal transmission may be beneficial in mitigating effects of transmission bottlenecks [20]. Furthermore, *S. velum* symbionts observed within

gill cells often contain multiple genomes per cell (pers. comm.). If this is also the case in the ovary/eggs, the number of symbiont cells (and unique genomes) associated with each egg may be less than the number of genomes quantified per egg, further exacerbating the effects of the transmission bottleneck. Thus, maintaining the capability for horizontal transmission in a vertically transmitted symbiosis could insure against the possible failure of vertical transmission, in addition to providing genetic diversity to the symbiont population.

References:

1. Helm MM, Bourne N, and Lovatelli A. 2004 *Hatchery culture of bivalves*. Food & Agriculture Organization of the United Nations.
2. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, and Madden TL. 2009 BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421.
3. Dmytrenko O, Russell SL, Loo WT, Fontanez KM, Liao L, Roeselers G, et al. 2014 The genome of the intracellular bacterium of the coastal bivalve, Solemya velum: a blueprint for thriving in and out of symbiosis. *BMC Genomics* 15, 924.
4. Russell SL, Corbett-Detig R, Cavanaugh CM. 2017 Mixed transmission modes and dynamic genome evolution in an obligate animal-bacterial symbiosis. *ISME Journal* **11**, 1359-1371.
5. Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T.L. (2012) Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* 13: 1.
6. Stewart FJ, Baik A, and Cavanaugh CM. 2009 Genetic subdivision of chemosynthetic endosymbionts of *Solemya velum* along the southern New England coast. *Applied Environmental Microbiology* 75, 6005–6007.
7. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. 2012 Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28. 1647–1649.
8. Nieto, M.A., Patel, K., and Wilkinson, D.G. (1996) *In situ* hybridization analysis of chick embryos in whole mount and tissue sections. In, *Methods in cell biology*. Elsevier, pp. 219–235.
9. Yonge CM. 1939 The protobranchiate Mollusca; a functional interpretation of their structure and evolution. *Philosophical Transactions of the Royal Society B* 230, 79–148.
10. Hobbie, J.E., Daley, R.J., and Jasper, S. (1977) Use of nuclepore filters for counting bacteria by fluorescence microscopy. *Applied and Environmental Microbiology* 33: 1225–1228.
11. Boulos L, Prevost M, Barbeau B, Coallier J, and Desjardins R. 1999 LIVE/DEAD® BacLightTM: application of a new rapid staining method for direct enumeration of viable and total bacteria in drinking water. *Journal of Microbiological Methods* 37, 77–86.
12. Taylor JP, Wilson B, Mills MS, Burns RG. (2002). Comparison of microbial numbers and enzymatic activities in surface soils and subsoils using various techniques. *Soil Biology and Biochemistry* 34: 387–401.
13. Krueger DM, Gustafson RG, and Cavanaugh CM. 1996 Vertical transmission of chemoautotrophic symbionts in the bivalve *Solemya velum* (Bivalvia: Protobranchia). *Biological Bulletin* 190, 195–202.

14. Zientz E, Beyaert I, Gross R, Feldhaar H. 2006 Relevance of the Endosymbiosis of Blochmannia floridanus and Carpenter Ants at Different Stages of the Life Cycle of the Host. *Applied and Environmental Microbiology* **72**, 6027–6033. (doi:10.1128/AEM.00933-06)
15. Stoll S, Feldhaar H, Fraunholz MJ, Gross R. 2010 Bacteriocyte dynamics during development of a holometabolous insect, the carpenter ant *Camponotus floridanus*. *BMC microbiology* 10, 1.
16. Hurst GD, Jiggins FM, Robinson SJ. 2001 What causes inefficient transmission of male-killing *Wolbachia* in *Drosophila*? *Heredity* 87, 220–226.
17. Peccoud J, Bonhomme J, Mahéo F, de la Huerta M, Cosson O, Simon J-C. 2014 Inheritance patterns of secondary symbionts during sexual reproduction of pea aphid biotypes: Sex and inheritance of aphid symbionts. *Insect Science* 21, 291–300. (doi:10.1111/1744-7917.12083)
18. Mira A, Moran NA. 2002 Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria. *Microbial Ecology* 44, 137–143. (doi:10.1007/s00248-002-0012-9)
19. Hosokawa T, Kikuchi Y, Fukatsu T. 2007 How many symbionts are provided by mothers, acquired by offspring, and needed for successful vertical transmission in an obligate insect-bacterium mutualism? *Molecular Ecology* 16, 5316–5325. (doi:10.1111/j.1365-294X.2007.03592.x)
20. Kaltenpoth M, Goettler W, Koehler S, Strohm E. 2010 Life cycle and population dynamics of a protective insect symbiont reveal severe bottlenecks during vertical transmission. *Evolutionary Ecology* 24, 463–477. (doi:10.1007/s10682-009-9319-z)
21. Serbus LR, White PM, Silva JP, Rabe A, Teixeira L, Albertson R, Sullivan W. 2015 The impact of host diet on *Wolbachia* titer in *Drosophila*. *PLoS Pathogens* 11, e1004777.