Genetic Variation among Endosymbionts of Widely Distributed Vestimentiferan Tubeworms

CAROL A. DI MEO,1 AMI E. WILBUR,2 WILLIAM E. HOLBEN,3 ROBERT A. FELDMAN,4 ROBERT C. VRIJENHOEK,5† AND S. CRAIG CARY1*

Graduate College of Marine Studies, University of Delaware, Lewes, Delaware 19958; Florida Marine Research Institute, St. Petersburg, Florida 33701; Division of Biological Sciences, The University of Montana, Missoula, Montana 59812-1002; Molecular Dynamics, Sunnyvale, California 94086-4520; and Center for Theoretical and Applied Genetics, Cook College, Rutgers University, New Brunswick, New Jersey 08903-0231

Received 29 July 1999/Accepted 30 November 1999

Vestimentiferan tubeworms thriving in sulfidic deep-sea hydrothermal vents and cold seeps are constrained by their nutritional reliance on chemosynthetic endosymbiotic bacteria. In a recent phylogenetic study using 16S ribosomal DNA, we found that endosymbionts from vent and seep habitats form two distinct clades with little variation within each clade. In the present study, we used two different approaches to assess the genetic variation among biogeographically distinct vestimentiferan symbionts. DNA sequences were obtained for the noncoding, internal transcribed spacer (ITS) regions of the rRNA operons of symbionts associated with six different genera of vestimentiferan tubeworms. ITS sequences from endosymbionts of host genera collected from different habitats and widely distributed vent sites were surprisingly conserved. Because the ITS region was not sufficient for distinguishing endosymbionts from different habitats or locations, we used a DNA fingerprinting technique, repetitive-extragenic-palindromic PCR (REP-PCR), to reveal differences in the distribution of repetitive sequences in the genomes of the bacterial endosymbionts. Most of the endosymbionts displayed unique REP-PCR patterns. A cladogram generated from these fingerprints reflected relationships that may be influenced by a variety of factors, including host genera, geographic location, and bottom type.

Invertebrates that are endemic to the highly sulfidic, reducing environments at deep-sea hydrothermal vents and cold seeps are commonly associated with chemosynthetic endosymbiotic bacteria (7, 18, 19). These bacteria oxidize the reduced sulfur compounds that are abundant in hydrothermal fluid (8, 36, 46). The resultant energy produced by the endosymbiont is coupled to the production of carbon sources that support the growth and maintenance of the invertebrate hosts (6, 8, 37, 48). The metabolic link between the invertebrate hosts and their endosymbionts has obvious implications for the dispersal and colonization strategies employed by the host organisms.

Vestimentiferan symbiont transmission in these associations by vertical transmission of the bacteria (12, 13, 16). Host specificity is maintained in the reservoir of repetitive sequences in the genomes of the bacterial endosymbionts.
between endosymbionts of geographically distributed vestimentiferan hosts but that these species- or strain-level differences have not been detected by approaches that have been employed to date.

The goal of our study was to resolve the question of genetic variation among widely distributed vestimentiferan endosymbionts by utilizing two molecular techniques that have historically proven useful for comparing closely related strains of bacteria. Our first approach entailed sequencing of the internal transcribed spacer (ITS) region, a noncoding region within the rRNA that often contains genetic variation sufficient for differentiating species of prokaryotes (2). Our second approach involved using a DNA fingerprinting technique that utilizes repetitive extragenic palindromes (REPs) as priming sites to differentiate species of prokaryotes (2). Our second approach employed to date.

**Materials and Methods**

**Sample collection.** Vestimentiferans were collected with the aid of submersibles from several locations, including two sites in the Western Pacific Ocean, four sites along the EPR, two sites in the Pacific Northwest, one site along the Galapagos Rift (GR), two sites along the western coast of California, and two sites in the Gulf of Mexico (GOM) (Table 1). Samples included in this study were chosen to represent the entire range of host vestimentiferan distribution. Habitat types included hydrothermal vent sites with a basaltic substrate, cold seep habitats with a sedimented bottom, and one sedimented site associated with a whale carcass (Table 1). All samples were carried to the surface in chilled seawater (0.5 to 2°C).

**DNA extraction.** The symbiont-containing trophosome tissue from each vestimentiferan host was aseptically removed, and a small portion of it was homogenized using the IsoQuick DNA purification kit (Orca, Bothell, Wash.) according to the manufacturer’s instructions and quantified spectrophotometrically. In addition to symbiont DNA, host DNA from three different genera was extracted from vestimentiferan tissue (non-symbiont-containing tissue) using the same extraction protocol to serve as controls. A number of the DNA samples used in this study were amplified from a mixed population of symbiont and host DNA for sequencing.

**16S rRNA gene characterization.** The methods used to PCR amplify, purify, and sequence the 16S rRNA symbiont genes were as described in the work of Feldman et al. (20). DNA sequence alignments were initially constructed using PileUp and then optimized (minimizing overall alignment differences) by eye in the SeqLab environment (Genetics Computing Group). Insertions and deletions were eliminated from the alignment before phylogenetic trees were determined.

**ITS region characterization.** PCRs amplified the ITS region of the symbiont ribosomal DNA (rDNA) using two ITS-specific primers, ITS16F-G1 (5′-GAAG TCTGAAACAGG-3′) (27) and ITS23R-L1 (5′-CAGGCCATCACCAGT-3′) (27). ITS16F-G1 is the 5′ end of the 16S rRNA gene, approximately 20 bases downstream from the spacer boundary in the 3′ end of the 16S rRNA gene (positions 1491 to 1505, Escherichia coli 16S rRNA gene). ITS23R-L1 is nested in the 5′ end of the 23S rRNA gene, approximately 20 bases downstream from the spacer boundary (positions 21 to 35, E. coli 23S rRNA gene). PCR mixtures (50-μl total volume) contained final concentrations of the following: 50 ng of symbiont DNA, 1× PCR buffer, 0.2 μM of each dNTP, 1.5 μM of MgCl2, 5% (vol/vol) acetamide, and 1.25 U of Taq polymerase (Promega). PCR conditions were as follows: 35 cycles, each consisting of denaturation at 92°C for 1 min, hybridization at 55°C for 2 min, and elongation at 72°C for 2 min added to each extension per cycle. A hot start (9) with denaturation at 95°C for 2 min was used at the beginning of the reaction to heighten reaction specificity. In addition, acetamide was added to the PCR to encourage more efficient amplification of GC-rich templates (38). All PCRs were performed on an MJ-Minicycler (MJ Research, Inc., Watertown, Mass.).

Because our attempts to directly sequence the smaller ITS PCR amplicon yielded poor results, amplified ITS regions were cloned to facilitate sequencing. Amplification products were pooled from three separate PCRs and cloned directly using the TA cloning kit with the PCK-II or the pCR-2.1-TOPO cloning vectors (Invitrogen, San Diego, Calif.) according to the manufacturer’s instructions. Plasmid preparations were made using a standard alkaline-lysis preparation (39). The sizes of the inserts were verified before sequencing by restriction fragment analysis. Plasmid DNA was further purified for sequencing using the Plasmid Miniprep Kit (Qiagen, Inc., Valencia, Calif.) according to the manufacturer’s instructions. Two clones from each amplified ITS region were cycle sequenced using the Perkin-Elmer (Foster City, Calif.) ABI BigDye dye termination cycle sequencing ready reaction kit with Ampli Taq DNA polymerase FS according to the manufacturer’s instructions. Sequencing was performed on an ABI PRISM 310 genetic analyzer.

**The ITS regions** were bidirectionally sequenced and confirmed prior to their alignment using the Sequence Navigator and AutoAssembler programs (Applied Biosystems, Inc., Foster City, Calif.). Only sequences that overlapped with 99% identity were included in the analysis. Final alignments were generated in the Genetic Data Environment, version 3.2. Phylogenetic relationships were determined from

**TABLE 1. Collection sites of vestimentiferan symbionts included in this study**

<table>
<thead>
<tr>
<th>Host species</th>
<th>Collection site</th>
<th>Latitude; longitude</th>
<th>Substrate, community type</th>
<th>Depth (m)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>R. pachyptila</td>
<td>18°S, EPR (18S)</td>
<td>18°36′S; 113°24′W</td>
<td>Basaltic, vent</td>
<td>2,637</td>
<td>20</td>
</tr>
<tr>
<td>R. pachyptila</td>
<td>9°N, EPR (9N)</td>
<td>9°49′N; 104°17′W</td>
<td>Basaltic, vent</td>
<td>2,516</td>
<td>20</td>
</tr>
<tr>
<td>R. pachyptila</td>
<td>Guaymas Basin vent (GY-V)</td>
<td>26°59′N; 111°24′W</td>
<td>Basaltic, vent</td>
<td>2,016</td>
<td>20</td>
</tr>
<tr>
<td>R. pachyptila</td>
<td>Guaymas Basin vent (GY-V)</td>
<td>26°59′N; 111°24′W</td>
<td>Basaltic, vent</td>
<td>2,451</td>
<td>20</td>
</tr>
<tr>
<td>Oasiaiasi avinae</td>
<td>21°N, EPR (21N)</td>
<td>20°47′N; 109°08′W</td>
<td>Basaltic, vent</td>
<td>2,630</td>
<td>This study</td>
</tr>
<tr>
<td>T. jerichonana</td>
<td>13°N, EPR (13N)</td>
<td>12°49′N; 103°56′W</td>
<td>Basaltic, vent</td>
<td>2,516</td>
<td>20</td>
</tr>
<tr>
<td>Ridgea pischae</td>
<td>JDF</td>
<td>45°58′N; 130°01′W</td>
<td>Basaltic, vent</td>
<td>1,550</td>
<td>This study</td>
</tr>
<tr>
<td>Lamellibrachia sp.</td>
<td>Green Canyon, GOM</td>
<td>27°41′N; 91°32′W</td>
<td>Sedimented, seep</td>
<td>700</td>
<td>5</td>
</tr>
<tr>
<td>Lamellibrachia sp.</td>
<td>Gorda Ridge (Gorda)</td>
<td>42°45′N; 126°42′W</td>
<td>Sedimented, seep</td>
<td>3,243</td>
<td>20</td>
</tr>
<tr>
<td>E. lamarina</td>
<td>WFE</td>
<td>26°07′N; 84°55′W</td>
<td>Sedimented, seep</td>
<td>2,016</td>
<td>This study</td>
</tr>
<tr>
<td>E. spicata</td>
<td>Santa Catalina Basin (whale)</td>
<td>33°12′N; 118°30′W</td>
<td>Sedimented, whale fall</td>
<td>1,240</td>
<td>20</td>
</tr>
<tr>
<td>E. spicata</td>
<td>Guaymas Basin vent (GY-V)</td>
<td>26°59′N; 111°24′W</td>
<td>Basaltic, vent</td>
<td>2,016</td>
<td>This study</td>
</tr>
<tr>
<td>E. spicata</td>
<td>Guaymas Basin vent (GY-S)</td>
<td>26°59′N; 111°24′W</td>
<td>Basaltic, vent</td>
<td>2,016</td>
<td>This study</td>
</tr>
<tr>
<td>L. columnna</td>
<td>Lau Basin, Hine Hina (Lau-Fiji)</td>
<td>22°32′S; 170°43′W</td>
<td>Sedimented, seep</td>
<td>1,859</td>
<td>42</td>
</tr>
<tr>
<td>Undescribed species</td>
<td>Nikko Scamount (Nikko)</td>
<td>30°05′N; 142°20′E</td>
<td>Basaltic, vent</td>
<td>433</td>
<td>This study</td>
</tr>
</tbody>
</table>

*See the work of Feldman et al. (20) for a map of the sample distribution.*
these aligned sequences using PHYLIp version 3.572 (22). A Kimura two-
parameter distance matrix was constructed, from which neighbor-joining trees with 100 bootstrap replications were generated. A second distance matrix, based on pairwise comparisons of the sample habitat types, was created by assigning “0” values to pairs of samples that were taken from similar habitat types (both vent or both seep) and “1” values to pairs of samples that were taken from different habitat types.

Mantel correlation tests were performed to compare ITS sequence distance with geographic and habitat distance (14, 33). Two- and three-way Mantel tests were performed using different combinations of the three distance matrices to assess whether ITS genetic distance significantly correlates with the geographic distances between endosymbiont collection sites or their habitat types (41). These Mantel tests were performed using the R package (32).

**REP elements.** REP sequences were amplified from each of the vestimentiferan symbionts with PCR using two universal REP primers, REP1 I (5'-IGCIGICATCIIGGCCTAC-3') (44). Each 25-μL REP-PCR mixture contained 50 ng of symbiont or host DNA and final concentrations of 1× Gitschier buffer [16.6 mM (NH4)2SO4, 67 mM Tris-HCl, 6.7 mM MgCl2, 6.7 μM EDTA, 30 mM β-mercaptoethanol (29)], 160 μg of DNA-grade bovine serum albumin per ml, 10% dimethyl sulfoxide, 1.25 mM (each) four deoxynucleoside triphosphates, 10 pmol of each REP primer, and 2 U of Taq polymerase (Gibco). One cycle of 95°C for 7 min was performed to denature the genomic DNA, followed by 30 cycles of 94°C for 1 min, 44°C for 1 min, and 65°C for 2 min, and a final extension at 65°C for 15 min. PCRs were performed with an MJ-Minicycler (MJ Research, Inc.). Reaction products (7.5 μl) were analyzed via agarose gel electrophoresis.

A digital image of each gel was captured and stored using the Alpha Imager 2000 Documentation and Analysis System (AlphaInnotech Corp., San Leandro, Calif.). Each gel image was analyzed using the software package GelCompar (version 4; Applied Maths, Kortrijk, Belgium). This software was used to normalize the molecular weight of each of the bands in each gel to all of the others based on the inclusion of the same set of molecular weight markers on each gel. Once normalized, all of the fingerprints in the database were compared to each other using an unweighted-pair-group-method-with-averages clustering algorithm and Jaccard coefficient.

**Nucleotide sequence accession numbers.** GenBank accession numbers for the new ITS sequences reported in this paper are as follows: AF076795, *Ridgella piscesae* endosymbiont, *Closcia gemmifera*; AF076796, Escarpia laminata endosymbiont, West Florida Escarpment (WFE); AF076797, *Lamellibrachia* sp. endosymbiont, GOM-10; AF076798, *E. spicata* endosymbiont, GY-S; AF076799, *R. pachyptila* endosymbiont, GY-V; AF076800, *Teunia jerichonana* endosymbiont, 9N; AF076801, *R. pachyptila* endosymbiont, 9N; AF076802, *T. jerichonana* endosymbiont, 13N; AF076803, *R. pachyptila* endosymbiont, 18S; AF076804, *Lamellibrachia* sp. endosymbiont, Gorda; AF076805, *R. pachyptila* endosymbiont, GR; AF076806, *R. pachyptila* endosymbiont, 18S; AF076807, O. alvinae endosymbiont, 21N; AF076808, *E. spicata* endosymbiont, whale; AF076809, undescribed species endosymbiont, Nikko Seamount; AF076810, L. columna endosymbiont, Lau-Fiji; AF076811, *Lamellibrachia* sp. endosymbiont, GOM-12; AF076812, *E. spicata* endosymbiont, GY-V; AF076813, *L. columna* endosymbiont, GY-V.

**RESULTS AND DISCUSSION**

**16S rDNA sequence variation among vestimentiferan endosymbionts.** In order to maintain a comparable data set, many of the endosymbiont DNA samples used in our ITS and REP-PCR analyses are the same DNA samples used in the previous 16S rDNA study of geographically distinct endosymbionts (20). Our study also includes additional samples to expand the representation of symbiont distribution. In concordance with our expanded data set, a new 16S rDNA phylogenetic tree was generated to include three of the endosymbionts that were added: GY-V *E. spicata*, GY-S *E. spicata*, and the undescribed species from the Nikko Seamount (Fig. 1A). In this new analysis of 16S rDNA sequences, the symbionts continue to cluster into two main groups: group I includes all of the symbionts from seep- or sediment-dwelling host organisms, while group II includes all of the symbionts from vent-dwelling host fauna. These relationships are in accordance with our previous findings (20). Interestingly, the difference between vent and seep
symbionts is maintained between symbionts from the same host, *E. spicata*, collected from vent (GY-V) and seep (GY-S) habitats at a single site in Guaymas Basin. Likewise, symbionts from two different vestimentiferan hosts are closely related when collected from the same habitat type (e.g., 9N *R. pachyptila* and *T. jerichonana* symbionts, both from vent habitats).

*ITS sequence variation among vestimentiferan symbionts.* ITS amplifications of vestimentiferan symbionts were generally 550 bp in length, although some individuals yielded a second, larger PCR product of roughly 800 bp in length (data not shown). The larger amplicon did not match any ITS sequences in the database, and we found that its intensity in the total PCR product was reduced by the use of higher annealing temperatures. For these reasons, we believe that the larger product represented an artifact of the PCR rather than rRNA operon heterogeneity. To separate these two distinct amplicons, we cloned the ITS PCR amplification products and used only the smaller amplicon from each symbiont for sequence analysis. As in many bacteria, all of the endosymbiont ITS regions contained putative sequences for the tRNA genes for alanine and isoleucine (Fig. 2). Five of the symbionts contained prominent insertion sequences (approximately base position 312 in ITS region) that totaled approximately 50 bp (Fig. 2).

ITS sequences of endosymbionts collected from several vestimentiferan host genera were compared phylogenetically using the corresponding ITS sequence of the symbiont of the lucinid bivalve *L. floridana* as an outgroup. A similarity matrix of ITS sequences was constructed based upon a Kimura two-parameter distance estimate. Genetic distances were calculated based on 500 bp of aligned sequences after the nucleotide insertions and deletions were removed. The sequences were surprisingly conserved overall, with percent sequence similarity ranging from 87.0 to 100% (data not shown). The symbionts from *R. pachyptila* hosts collected from Guaymas Basin, 9°N, and 18°S displayed 100% sequence similarity.

Symbiont relationships were depicted as a cladogram based upon ITS sequence diversity (Fig. 1B). Because of the high similarity among the ITS sequences, many of the bootstrap values at branch points in the neighbor-joining analysis were low or insignificant. Only bootstrap values of ≥50% were reported. The ITS cladogram showed a prominent break within the symbionts: group I included the symbionts of *E. spicata* from the whale carcass, *L. columna* from the Lau-Fiji basin, the undescribed species from a vent site at the Nikko Seamount, one of the two *Lamellibrachia* spp. collected from a seep site in the GOM (GOM-12), and *E. laminata* from the seep site at Guaymas Basin; group II included the rest of the symbionts (Fig. 1B). Apart from differing in sequence from the members of group II, the symbionts in group I contained prominent insertion sequences that totaled approximately 50 bp (Fig. 2).

ITS sequences of endosymbionts collected from several vestimentiferan host genera were compared phylogenetically using the corresponding ITS sequence of the symbiont of the lucinid bivalve *L. floridana* as an outgroup. A similarity matrix of ITS sequences was constructed based upon a Kimura two-parameter distance estimate. Genetic distances were calculated based on 500 bp of aligned sequences after the nucleotide insertions and deletions were removed. The sequences were surprisingly conserved overall, with percent sequence similarity ranging from 87.0 to 100% (data not shown). The symbionts from *R. pachyptila* hosts collected from Guaymas Basin, 9°N, and 18°S displayed 100% sequence similarity.

Symbiont relationships were depicted as a cladogram based upon ITS sequence diversity (Fig. 1B). Because of the high similarity among the ITS sequences, many of the bootstrap values at branch points in the neighbor-joining analysis were low or insignificant. Only bootstrap values of ≥50% were reported. The ITS cladogram showed a prominent break within the symbionts: group I included the symbionts of *E. spicata* from the whale carcass, *L. columna* from the Lau-Fiji basin, the undescribed species from a vent site at the Nikko Seamount, one of the two *Lamellibrachia* spp. collected from a seep site in the GOM (GOM-12), and *E. laminata* from the seep site at Guaymas Basin; group II included the rest of the symbionts (Fig. 1B). Apart from differing in sequence from the members of group II, the symbionts in group I contained prominent insertion sequences that totaled approximately 50 bp (Fig. 2). The sequences of these insertions were nearly identical in all of the symbionts in which they were found. Groups I and II in the 16S rDNA cladogram (Fig. 1A) correlate well with groups I and II in the ITS cladogram (Fig. 1B), with one exception: *E. laminata* symbiont, WFE, appears in group I according to its 16S rDNA sequence and in group II according to its ITS sequence (Fig. 1).

To determine if there was a significant correlation between ITS sequence diversity and either geography or habitat type, we performed two- and three-way Mantel tests to compare distance matrices generated from pairwise differences between ITS sequences, geographic distances, or habitat types. The
results of these Mantel tests are shown in Table 2. Despite the remarkable similarity among the ITS sequences, the genetic distances among ITS regions correlate significantly with geographic distances. Although none of the correlations were particularly strong (r ≤ −0.334 for all tests), P values from two different Mantel tests indicated that ITS sequence similarity and geography were significantly correlated (P = 0.0349, test 3; P = 0.0480, test 4). In contrast, P values for a two-way Mantel test of ITS sequence similarity versus habitat type alone (test 1) or a three-way test versus habitat type with geography factored out (test 5) were not statistically significant (P = 0.146 and 0.102, respectively). Thus, ITS sequence similarity was significantly correlated with broad-scale endosymbiont geography (P ≤ 0.05) but not significantly correlated with habitat type. In light of the fact that the ITS sequences for 9°N, 18°S, and Guaymas *R. pachyptila* were identical, our results suggest that the ITS region is variable enough to show broad-scale biogeographic variation but is not variable enough to distinguish among symbionts located along the same ridge axis. These results are in contrast to the results of Feldman et al. (20) and the new analysis of 16S rDNA relationships presented in this study, which suggest that symbiont sequence distances correlate with habitat type (basaltic, vent versus sedimented, seep substrates). Although our data suggest that the geographic distribution of the hosts may be a dominant influence in determining genetic variation of the endosymbions, it is clear that several factors affect endosymbiont diversity.

Historically, the ITS region has been useful for discriminating among strains of several bacterial species, including a hyperthermophilic archaeon (11), *Rhizobium* sp. (30), *Bifidobacterium* sp. (31), infectious *Pseudomonas* sp. (43), and *Trichodesmium* sp. (49). In the present study, we found that the ITS region was much more conserved than expected. Although the ITS region is usually very informative at the subspecies and/or strain level (11, 30, 31, 49), it has been unreliable in some cases. In studies of the *Mycobacterium tuberculosis* complex (23) and with several *Listeria* isolates (15), the ITS region failed to differentiate between species and/or strains.

**REP-PCR fingerprints of the vestimentiferan symbionts.** The REP-PCR technique is particularly well suited for investigations of bacterial symbiosis, in that it allows investigators to survey the symbiont genome in the presence of host DNA because the REP priming sites are exclusively bacterial (25). To confirm that REP sequences were not present in the genomes of the vestimentiferan hosts, we performed REP-PCR amplifications on DNA that was aseptically acquired from vestimentum (non-symbiont-containing) tissue from three different host genera (*Escarpia*, *Riftia*, and *Tevnia*). Most of these samples failed to amplify at all (data not shown). The exception was *E. laminata*, which produced a single band that did not correspond with any of the bands seen in the REP fingerprint of its endosymbiont.

Fingerprint patterns of the vestimentiferan endosymbiont DNA displayed remarkable variation (Fig. 3). Relationships based upon the number of shared bands among the fingerprint patterns were compared using GelCompar software and depicted as a cladogram with a corresponding, computer-generated gel image (Fig. 4). All of the symbionts included in the REP analysis displayed unique fingerprints, with the exception of 9°N *R. pachyptila*, and GR *R. pachyptila*, which exhibited indistinguishable patterns (Fig. 3 and 4). The REP analysis seems to divide the symbionts into four distinct groups, or clades (Fig. 4). Group I includes both *Lamellobacchia* symbionts from the GOM and the symbiont from *Ridgeia pachyptila* from the JDF. Group II includes the *R. pachyptila* symbiont and the symbiont from *T. jerichonana*. Group III includes all of the *Escarpia* symbionts, from both *E. spicata* and *E. laminata*. Group IV includes the endosymbionts from the undescribed species at Nikko Seamount and *L. columna* from the Lau-Fiji Basin, two distantly located sites in the western Pacific Ocean.

Upon closer examination, the REP analysis seems to reflect three different influences on endosymbiont differentiation: host genus, bottom or substrate type, and geography. Host genus seems to be the primary source of differentiation, as most of the clades break up into distinct groupings of congeners. The only clade where this breaks down is group II, which includes symbionts from *Riftia* and *Tevnia* host genera. This clade, however, supports previous studies using DNA-DNA hybridization and in situ hybridization that suggest that the endosymbions of *Riftia* and *Tevnia* are closely related (5, 17).

The secondary influences for symbiont genetic differentiation seem to be geographic location and substrate type. Within clades of congeners, it appears that symbionts differ based upon where they are collected and/or the habitat type (basaltic or sedimented). In group II, all of the *R. pachyptila* symbionts collected from the hosts at northern EPR sites (i.e. GY-V, 9N, and GR) are less closely related to the *R. pachyptila* symbiont collected from a site 18°S along the southern EPR. Within the *Escarpia* symbiont clade (group III), both geographic location and substrate type seem to be at work: symbionts collected from hosts at the whale fall in the Santa Catalina Basin (whale), Guaymas Basin seep (GY-S), and WFE, all soft-bot-
tom substrate types, clade together and group away from the symbionts collected from the Guaymas Basin vent (GY-V). Within the soft-bottom grouping of *Escarpia* symbionts, it appears that the symbionts from California and Guaymas (both on the western coast of North America) group away from the symbiont from Florida (on the eastern coast). The symbionts from the Nikko Seamount and Lau-Fiji Basin are clearly different from the rest of the symbionts, presumably because they are so geographically removed from the rest of the collection sites.

REP-PCR has been useful for genomic fingerprinting of various strains of bacteria, including *Actinobacillus* (1), *Rhizobium* (10, 30), *Legionella* (24), *Streptococcus* (44), *Bacillus* (26), and *Citrobacter diversus* (51) strains. In the present study, we found that the REP-PCR fingerprint analysis was sufficiently fine-scaled to reveal some interesting strain-level genetic variation among vestimentiferan symbionts. Presumably, this is because the REP analysis targets the whole genome, rather than a specific, relatively conservative region of the DNA. This physically maximizes the probability of finding strain-level variation. It should be noted that the symbiont of *O. alvinae* was not included in the REP analysis, because a reliable REP-PCR amplification and subsequent fingerprint pattern were not obtainable with this DNA. When the DNA was run on an agarose gel, it appeared sheared, and although it was sufficiently large enough for performance of the 550-bp ITS amplification, there was insufficient high-molecular-weight genomic DNA to support the REP analysis.

Implications for symbiont relatedness and acquisition. Previous studies have suggested that vestimentiferans acquire their symbionts through ingestion of free-living bacteria upon larval settlement (28, 42). If this mechanism alone is employed, one would expect that symbionts from two different vestimentiferan host species living at the same location would be identical or related more closely to each other than they are to other endosymbionts collected from distant vent sites. This, in fact, was not the case, as symbionts from 9°N *T. jerichonana* and 9°N *R. pachyptila* (both collected from the same rock) did not display identical ITS sequences or REP-PCR fingerprints. Instead, the symbionts grouped according to their respective host species as revealed through comparisons of shared bands in their REP-PCR fingerprints.

One explanation for this is that larvae could acquire their symbionts prior to attachment, either from a different substrate than where they eventually colonize or from the water column. This could enable two vestimentiferans that reside in the same site to harbor different symbionts. Alternatively, symbiont acquisition may be related to ecological succession. *T. jerichonana* is the first colonizer of new hydrothermal vent fields, in areas of the most intense diffuse flow (40). Presumably, this is because *T. jerichonana* is more tolerant (than *R. pachyptila*) of the high temperatures and elevated concentrations of H2S that are present in newly formed vent sites. It is only after *T. jerichonana* has become established and the levels of reduced chemicals have attenuated that *R. pachyptila* begins to colonize a new hydrothermal vent site (40). The *T. jerichonana* and *R. pachyptila* hosts may simply acquire a different subset of a free-living chemosynthetic community due to temporal and/or spatial consequences of ecological succession.

Specific host recognition mechanisms may exist that allow certain strains of endosymbiotic bacteria to survive in specific vestimentiferan hosts. Although these mechanisms have not been demonstrated in vestimentiferans, they could explain how two different hosts living on the same rock could harbor different symbionts. Alternatively, the larvae may acquire bacteria that have been released from congeners residing at the site of larval settlement, potentially through host decomposition. These mechanisms might yield REP-PCR fingerprint patterns that imply close relationships between endosymbionts of the same host genus. Because only one individual from each vestimentiferan genus was sampled from most of the locations, we cannot determine the extent of ITS sequence variation among endosymbionts at each site.

The extent of vestimentiferan symbiont relatedness and the host transmission mechanisms used to maintain these symbioses have remained ongoing questions. The results presented here demonstrate the existence of significant strain-level variation between endosymbionts of different vestimentiferan hosts collected from geographically separated areas. Further work to pinpoint the stage during vestimentiferan development when the symbiosis is first established will shed light on symbiont transmission and help to explain the genetic diversity among symbiont populations.

ACKNOWLEDGMENTS

We thank the crews and pilots of the R/V *Atlantis II* and the deep-submergence vehicle *Alvin*, the *Nautilus*, and the U.S. Navy Advanced Tethered Vehicle for their assistance in sample collection. Samples of
References


for simple extraction of DNA for PCR-based typing from forensic material.
carbon sources of mussels and tubeworms from Galapagos hydrothermal
the marine nitrogen-fixing cyanobacterium Trichodesmium sp. strain NIBB
1067, derived by 16S ribosomal RNA sequence analysis. Microbiology 140:
2159–2164.
relationships among isolates of Citrobacter diversus using DNA fingerprints
generated by repetitive sequence-based primers in the polymerase chain