Published in final edited form as:

Mol Ecol. 2014 March ; 23(6): 1418–1432.

Genetic differentiation among isolates of *Teredinibacter turnerae*, a widely occurring intracellular endosymbiont of shipworms

Marvin A. Altamia^a, Nicole Wood^b, Jennifer Fung^b, Sandra Dedrick^b, Eric W. Linton^c, Gisela P. Concepcion^a, Margo G. Haygood^d, and Daniel L. Distel^{e,*}

^aMarine Science Institute, University of the Philippines-Diliman, Quezon City, 1101, Philippines

^bLaboratory for Marine Genomic Research, Ocean Genome Legacy, Inc., Ipswich, MA 01938

^cDepartment of Biology, Central Michigan University, Mount Pleasant, MI 48859

^dInstitute of Environmental Health, Oregon Health & Science University, Portland, OR 97239

^eOcean Genome Legacy Center of New England Biolabs, Marine Science Center, Northeastern University, Nahant, MA 01908

Abstract

Teredinibacter turnerae is a cultivable intracellular endosymbiont of xylophagous (wood-feeding) bivalves of the Family Teredinidae (shipworms). Although *T. turnerae* has been isolated from many shipworm taxa collected in many locations, no systematic effort has been made to explore genetic diversity within this symbiont species across the taxonomic and geographic range of its hosts. The mode of symbiont transmission is unknown. Here we examine sequence diversity in fragments of six genes (16S rRNA, *gyrB*, *sseA*, *recA*, *rpoB*, and *celAB*) among 25 isolates of *T. turnerae* cultured from 13 shipworm species collected in 15 locations in the Atlantic, Pacific and Indian Oceans. While 16S rRNA sequences are nearly invariant among all examined isolates (maximum pairwise difference <0.26%) variation among examined protein coding loci is greater (mean pairwise difference 2.2–5.9%). Phylogenetic analyses based on each protein-coding locus differentiate the 25 isolates into two distinct and well-supported clades. With five exceptions, clade assignments for each isolate were supported by analysis of alleles of each of the five protein coding loci. These exceptions include (1) putative recombinant alleles of the *celAB* and *gyrB* loci in two isolates (PMS-535T.S.1b.3 and T8510), suggesting homologous recombination between members of the two clades, and (2) evidence for a putative lateral gene transfer event affecting a second locus (*recA*) in three isolates (T8412, T8503 and T8513). These results demonstrate that *T. turnerae* isolates do not represent a homogeneous global population. Instead they indicate the

*Corresponding author contact information: Daniel L. Distel, Ocean Genome Legacy, 240 County Road, Ipswich, MA, USA, 01938, Tel. 978-380-7425, distel@oglf.org.

Author contributions

Marvin A. Altamia, Nicole Wood, Jennifer Fung and Sandra Dedrick performed the research described in this manuscript under the supervision and direction of Gisela P. Concepcion, Margo G. Haygood and Daniel L. Distel. Daniel L. Distel conceived and designed the project. Daniel L. Distel, Marvin A. Altamia and Margo G. Haygood wrote the manuscript.

Data accessibility

DNA sequences used in this study have been submitted to GenBank (see Table S1). Other data including treefiles and alignments have been submitted to Dryad and may be viewed at Dryad DOI (doi:10.5061/dryad.r9693).

emergence of two lineages that, although distinct, likely experience some level of genetic exchange with each other and with other bacterial species.

Keywords

Symbiosis; Bivalvia: Teredinidae; Teredinibacter; shipworm; recombination; lateral gene transfer

Introduction

Teredinibacter turnerae is a gammaproteobacterium of the order Alteromonadales known to occur as an endosymbiont of wood-boring bivalves of the family Teredinidae (commonly known as shipworms). Shipworms are among the most important wood consumers in marine environments [1]. Although capable of filter feeding [2], these bivalves are the only marine animals known to sustain normal growth and reproduction with wood as their sole source of particulate food [3]. Shipworms are ecologically and economically important, acting both as the principal mineralizers of wood in marine environments and as costly nuisance species, causing extensive damage to wooden piers, vessels, fishing equipment, and other man-made structures [4]. Despite their economic importance, and the fact that they have been part of the scientific literature since their description by Sellius in 1753 [5], this family remains poorly studied and poorly understood. The taxonomy of the family Teredinidae was most recently reviewed by Turner in 1966 [1]. This work, which recognizes only 14 of 42 historically proposed genera and 66 species from among nearly 300 historical species names, forms the basis of the classification scheme applied here [6] but also underscores the chaotic state of the systematics of this family. Indeed, recent molecular phylogenetic analyses show major taxonomic divisions within the family to be nonmonophyletic [7]. The geographic distribution of Teredinidae is also poorly understood. Many species are distributed over vast oceanic areas “apparently limited only by temperature, salinity and [means of] transportation [8]. Current-driven dispersal likely contributes to the frequent transport of long-lived planktonic larvae over entire ocean basins [9]. Adults also may be frequently transported over great distances by drifting wood and by wooden vessels. Thus, range boundaries of many Teredinidae may be highly dynamic in time and space and are only poorly characterized.

The mode of wood digestion by shipworms is also largely unknown, but intracellular symbionts have been implicated in this process [10]. Most described wood-eating animals rely on cellulolytic enzymes produced by extracellular microbes in the gut and/or endogenous nuclear-encoded cellulases for digestion of wood [4, 11]. Shipworms lack such well-developed microbial communities in their gut [12], but harbor dense endosymbiotic microbial communities within a specialized tissue of the gills referred to as the “Gland of Deshayes” [13]. The presence of bacterial endosymbionts in shipworm gills was first demonstrated by Popham and Dixon in 1973 using transmission electron microscopy (TEM) [14, 15]. Various putative functions were attributed to these bacteria, including production and/or uptake of essential amino acids [16, 17]. However, the subsequent discovery and *in vitro* cultivation of a bacterium, called *Teredinibacter turnerae* [18], from gill homogenates from numerous shipworm species, suggested alternative functions for these intracellular

symbionts [10]. When grown in pure culture, *T. turnerae* secretes cellulase (endo-1, 4- β -D glucanase) [19-21] and other key enzymes that degrade wood lignocellulose. It also fixes nitrogen when grown micro-aerobically [10]. Based on this information, it was proposed that *T. turnerae* is the intracellular endosymbiont previously observed by Popham and Dixon and that it produces enzymes, including cellulases and nitrogenase, that may contribute to the host's ability to survive on an indigestible and nitrogen-deficient diet of wood [10, 20, 22].

The presence of *T. turnerae* in gill bacteriocytes of the shipworm *L. pedicellatus* was subsequently confirmed using culture independent 16S rRNA methods including fluorescence *in situ* hybridization (FISH) with 16S rRNA directed oligonucleotide probes [23] but these studies also revealed that *T. turnerae* is not alone in shipworm gills. Instead it may constitute one member of a community that may also include several additional as yet uncultivated but closely related symbiont phylotypes [24, 25]. For example, in the shipworm *Lyrodus pedicellatus*, five symbiont phylotypes were detected, one of which corresponded to *T. turnerae* (on average accounting for $11.5 \pm 8.2\%$ of total 16S rRNA copies), and all of which were shown to form a closely related (99.8-91.2% identical in 16S rRNA sequence) monophyletic group that excluded other known free-living bacteria based on analyses of 16S rRNA genes [12, 24, 25]. These symbionts have been shown to fix nitrogen that is subsequently utilized by the host [26], and have been proposed to produce cellulolytic enzymes that contribute to the host's ability to digest wood [10]. In addition, *T. turnerae* has been shown to produce antimicrobial compounds that may help it to defend its intracellular niche and to structure the microbial communities associated with the host [27]. The mode of transmission of *T. turnerae* remains unknown although evidence for vertical transmission has been reported for another closely related but as yet uncultivated shipworm gill endosymbiont species [28].

T. turnerae differs from other shipworm endosymbionts, and other bivalve gill endosymbionts, in that it has been described based on axenic culture and that it has been isolated from many shipworm specimens, species and genera collected in many locations around the world [10, 18]. Previous phylogenetic analyses of six *T. turnerae* isolates from diverse host species, based on reverse transcribed 16S rRNA sequences, detected no sequence variation among isolates [23]. These results raised the question of whether *T. turnerae* represents a single, largely undifferentiated, symbiont population with worldwide distribution, or whether as yet undetected genetic structure exists among isolates. To explore this question, we took advantage of, and added to, a unique existing collection of *T. turnerae* isolates obtained from a broad cross section of the taxonomic and geographic range of its host species between 1979 and 1986 by John B. Waterbury (WHOI) and Ruth D. Turner (Harvard University). We examined sequences of multiple gene loci including 16S rRNA and five protein-coding loci in twenty-five *T. turnerae* isolates. We reasoned that the extent and patterns of sequence variability at these loci might provide clues with regard to specific relationships among symbionts, incidence of genetic exchange among symbionts, and patterns of host-symbiont sorting. While the opportunistic sampling performed here is not suitable for rigorous analysis of host and symbiont biogeography or cophylogeny, nor to resolve the as yet unknown mode(s) of symbiont transmission, it is sufficient to place

constraints on, and therefore to provide context for, our understanding of this little studied and poorly understood symbiosis.

Materials and Methods

Isolation of *T. turnerae*

Isolates T7901, T7902, T7903, T8401, T8402, T8412, T8415, T8503, T8508, T8509, T8510, T8513, T8601 and T8602, were obtained from a collection of isolates cultured by J. Waterbury (WHOI) between 1979 and 1986 and subsequently deposited to the Ocean Genome Resource (www.oglf.org). These isolates were cultivated as in [10, 18]. Briefly, gills were removed from freshly dissected shipworm specimens, rinsed in sterile 0.22µm filtered seawater (FSW) and gently homogenized by hand in a glass Dounce homogenizer in 10 volumes of fresh FSW. One ml of homogenate was then used to inoculate soft agar cultures in shipworm basal medium (SBM) containing 0.2% agar and 0.2% powdered cellulose (Sigmacell cellulose Type 101) as the sole potential sources of organic carbon (SBM+S). Ten serial dilution steps were performed on each homogenate with ten-fold dilution (v/v) at each step. Individual colonies were obtained from the highest dilution in which growth was observed by streaking onto SBM agar plates with 0.2% Sigmacell and 5 mM ammonium chloride added as sole sources of organic carbon and combined nitrogen respectively (SBM+S+N plates). Plates were incubated at 22-30°C until colonies were observed (4-7 days). Individual colonies were picked and re-streaked on fresh plates at least twice, then used to inoculate liquid cultures. Liquid cultures for DNA extraction were grown from single colonies in 25ml of SBM+S+N in 250 ml Erlenmeyer flasks in a shaker incubator at 22-30°C for 5-10 days, shaken at 100 rpm. Stocks for long-term storage were prepared by adding 0.5 ml of 40% glycerol to 0.5 ml of 5 day liquid culture and freezing at -80°C. All other isolates were prepared by the same method except that the soft agar culture and serial dilution steps were omitted. Instead, homogenates were directly streaked on SBM+S+N plates.

Host and isolate selection

The twenty-five *T. turnerae* isolates examined in this investigation were isolated from thirteen host species, selected to represent a broad cross-section of the phylogenetic diversity and geographic range of the bivalve family Teredinidae. These included representatives of two of three extant subfamilies (Teredininae and Bankiinae) and seven of fourteen genera [1, 29]. Collection sites included locations in the Atlantic, Pacific and Indian oceans and the coasts of four continents (North and South America, Asia and Australia). Where possible, isolates were obtained from multiple representatives of single host species collected both in closely situated and distantly separated locations. Isolates T0609 and T0611 and isolates PMS-509L.S.1a.4 and PMS-509L.S.1a.6 were obtained from single specimens of the species *L. pedicellatus* and *Teredo clappi*, respectively. Isolates T0611 and T0609 were obtained from a specimen of *L. pedicellatus* from the same continuously maintained captive breeding colony as was T7902 twenty-seven years earlier. See Table 1 for a complete list of isolates, hosts and locations.

Selection of gene loci

Regions of six genes were selected for sequence analysis. These included a 1,125 bp region of the 16S rRNA gene (aligning with *E. coli* positions 227-1351), a highly conserved structural RNA widely used as a phylogenetic marker for bacteria, regions of four protein coding genes; *rpoB* (4,083 bp), *gyrB* (2,427 bp), *recA* (1,029 bp), and *sseA* (846 bp), that are typically present in a single copy and that may be considered as components of core metabolism, and *celAB* (1,011 bp) a single copy bifunctional glycosyl hydrolase that may be considered as a component of specialized or peripheral metabolism. Based on the available genome sequence for one isolate (T7901) the selected loci are widely distributed around the single circular chromosome, reflecting no regional bias in distribution (Figure S1).

Nucleic acid purification, PCR amplification, and sequencing

Bacterial genomic DNA was extracted from 10 ml aliquots taken from 5-day old 25 ml liquid cultures (SBM+S+N). Cells (and some undigested residual cellulose powder) were pelleted by centrifugation ($10,000 \times g$, 30 minutes, 4°C). Genomic DNA was extracted from the pellet using the DNeasy Blood & Tissue Kit (Qiagen) according to the manufacturer's suggested protocol. Gene fragments were amplified by polymerase chain reaction (PCR) using the mixture: 20-50 ng template DNA, 12.5 µl 2X Phusion High-Fidelity PCR Master Mix (Finnzymes), 0.5 µM forward primer, 0.5 µM reverse primer, 7.5 µl sterile deionized water, for a total reaction volume of 25 µl. Thermal cycling parameters were as follows: denaturation; (95°C, 4 min, one cycle); amplification; 95°C, 30 sec, primer-dependent annealing temperature (see Table 2), 30 sec, 72°C, 60 sec, 30 cycles, extension; 72°C, 300 sec, one cycle. The PCR primer sequences are shown in Table 2. The size and homogeneity of amplified products were evaluated by electrophoresis (1.2% agarose). Amplified products were purified using QIAquick PCR Purification Kit (Qiagen). Amplicons were either directly sequenced or cloned into pCR-Blunt II-TOPO (Invitrogen, Carlsbad, CA) vector and transformed into *Escherichia coli* cells (C3019, NEB). Transformed cells were screened for inserts and plasmids were purified using the Qiagen Miniprep kit according to manufacturer's instructions. Clones were selected and sequenced using both vector and gene specific primers (Table 2). For direct sequencing, PCR amplicons were purified using the QIAquick kit according to the manufacturer's suggested protocol and sequenced using gene specific primers (Table 2). All samples were bi-directionally sequenced on an ABI 3130xl Genetic Analyzer using ABI Prism BIG DYE Terminator v3.1.

Phylogenetic Analysis

Nucleotide sequences for each gene fragment were assembled and edited using Geneious (v. 6.0.5) (Biomatters, NZ, available from <http://www.geneious.com>). Alleles of each locus were aligned using MAFFT [30] implemented in Geneious (v. 6.0.5) under default parameters (alignment strategy selected automatically; scoring matrix 200PAM/k=2; gap open penalty 1.53; and offset value 0.123; and trimmed to equal length (16S rRNA, 1125 bp; *gyrB*, 1,053 bp; *recA*, 489 bp; *rpoB*, 1,020 bp; *celAB*, 804 bp; *sseA*, 591 bp). GenBank accession numbers are listed in Table S1. Sequence divergence and base composition of the aligned datasets were calculated using DnaSP version 5.10 [31]. Best-fit models of evolution

and parameter value estimates for single gene and concatenated datasets were identified by Akaike information criterion (AIC) through jModelTest 2 [32] (GTR+I+ Γ for 16S rRNA, *gyrB* and concatenated dataset; SYM+ Γ for *recA* and *rpoB*; SYM+I for *celAB*; HKY85+I for *sseA*.) The partition homogeneity test ($p = 0.01$) [33] was used to determine compatibility of the data prior to concatenation. Phylogenetic analyses were performed by employing both Bayesian inference (MrBayes) [34, 35] and maximum likelihood (PhyML) [36] methods implemented in Geneious version 6.0.5. Bayesian-inference analyses were executed using 5 million generations, subsampling every 2000 generations with a burn-in of 20%. The sump command in MrBayes was used to confirm convergence on similar optimal log likelihood scores and tree topologies in each analysis. Maximum likelihood trees were constructed using a combination of Nearest Neighbor Interchanges (NNIs) and Subtree Pruning and Regrafting (SPR) algorithms. A total of 100 bootstrap replicates were used to construct maximum likelihood trees. Trees based on amino acid sequences were inferred using the maximum likelihood algorithm implemented in Molecular Evolutionary Genetics Analysis (MEGA) software version 5.05 [37] employing Jones-Taylor-Thorton (JTT) amino acid substitution model with 5 discrete gamma categories for a total of 1000 bootstrap replicates. Congruence of host and symbiont phylogenies was analyzed using the randomization test implemented in TreeMap v.2 [38] comparing tree topologies inferred for symbionts as described above and hosts as described in [7]. To examine the relationships of the *T. turnerae* isolates to other gammaproteobacteria, additional phylogenetic analyses were performed as described above using 16S rRNA, *gyrB*, *recA*, and *rpoB* with inclusion of homologous reference sequences from related gamma proteobacterial species. BLAST searches of Genbank non-redundant nucleotide database failed to identify compelling homologues of *celAB* and *sseA* in bacterial species other than *T. turnerae* and so these were omitted from this analysis. *E. coli* sequences were specified as outgroups for *gyrB*, *recA*, *rpoB* and 16S rRNA in Bayesian inference analyses.

Detection of recombination

Individual gene datasets were screened for evidence of recombination using GENECONV [39], BootScan [40], MaxChi [41], Chimaera [42], and 3Seq [43] methods implemented in Recombination Detection Program (RDP) version 4.16 [44] with a P -value of 0.01, using a Bonferroni correction [45] to compare test results. For MaxChi, Chimaera, and RDP, a window size of 60 nucleotide positions was used to scan the sequences for recombination; while for Bootscan a step size of 20 was selected and the Felsenstein 1984 model was employed with 1000 bootstrap replicates. G-ENECONV analysis was performed with G-scale (mismatch penalty) set to 1.0.

Analysis of selection

Evidence for positive or negative selection was assessed by analysis of nonsynonymous and synonymous substitutions within the 5 protein-coding loci using SLAC, FEL, REL and FUBAR [46] methods implemented on the Datamonkey server (<http://www.datamonkey.org>) [47, 48]. Recombinant alleles identified by methods within RDP were removed from the dataset. The best-fit nucleotide substitution model was reestimated using HyPhy (*gyrB*, GTR 011020; *rpoB*, TN93; *celAB*, 010011; *recA* and *sseA*, HKY85) [49]. Positively and negatively selected sites were determined using a cutoff P -value of 0.05

for SLAC and FEL tests; Bayes Factor of 100 for REL; and a posterior probability of 0.95 for FUBAR. Results from all the four tests were then compared using the integrative selection tool within Datamonkey to determine consensus among the different methods.

Results

Sequence variation among examined loci

Partial gene sequences (489 to 1,125 nucleotides) were determined for five protein-coding gene loci (*rpoB*, *recA*, *gyrB*, *sseA* and *celAB*) from the genomes of 25 *T. turnerae* isolates and one structural RNA (16S rRNA) from 21 of the aforementioned isolates. All isolates examined were nearly identical with respect to 16S rRNA sequence. Of 1,125 nucleotide positions in the aligned 16S rRNA sequence set, only three were observed to be variable. The phylogenetic position of *T. turnerae* isolates with respect to other gammaproteobacteria is shown in Figure S2. The greatest pairwise difference observed among isolates in examined 16S rRNA sequences was 3/1,125 or 0.26% (data not shown). Because 16S rRNA sequences were nearly invariant, this locus was excluded from further analyses. The protein-coding loci showed considerably greater variation with mean pairwise differences ranging from 2.2-5.9% (Table 3). With the exception of three *recA* alleles in three strains (T8412, T8503 and T8513; see Recombination and Lateral Gene Transfer), no length variation was observed among PCR amplicons derived from the examined alleles.

Phylogenetic differentiation among isolates

Bayesian inference and maximum likelihood analyses based on partial nucleotide sequences of each protein-coding locus differentiate the 25 isolates into two distinct and well supported clades (posterior probabilities = 1.0, bootstrap proportions = 96%), hereafter designated clades I and II (Table 1; Figure 1). Tree topologies inferred by maximum likelihood analyses of amino acid sequences were consistent with those inferred from nucleotide sequences (Figure S3), with the exception of *recA*, for which only a single amino acid change in a single allele was observed. Mean interclade variation in nucleotide sequence is significantly greater (independent *t*-test, *P* = 0.01) than intraclade variation for all loci (Table 3). With five exceptions (*recA* in T8412, T8503 and T8513, *celAB* in PMS-535T.S.1b.3 and *gyrB* in T8510), clade assignments for each isolate were supported by analyses of all alleles of the five protein coding loci and by the concatenated sequence set (Table 1, Figure 1).

Phylogenetic relationship of *T. turnerae* to other gamma proteobacteria

Phylogenetic analyses based on 16S rRNA, *gyrB*, *rpoB* and *recA* genes (Figure S2) indicate that all examined *T. turnerae* isolates form a monophyletic taxon most closely related to gamma proteobacteria of the order Alteromonadales.

Host specificity

All isolates examined, with the exception of T0609 and T0611, which were isolated from the same host specimen, represent unique genotypes containing at least one allele or combination of alleles not observed in any other isolate. In at least five host species (*Dicyathifer manni*, *Lyrodus pedicellatus*, *Bankia gouldi*, *Bankia rochi*, *Teredo navalis*, and *Teredo clappi*) distinct genotypes of *T. turnerae* were obtained from specimens collected in

different locations. Two distinct genotypes of *T. turnerae* were obtained from a single specimen of *L. pedicellatus* (PMS-509L.S.1a.4 and PMS-509L.S.1a.6). Although genotypes differed, all specimens of each host species examined displayed symbiont clade specificity, i.e. each harbored *T. turnerae* isolates of the same clade. Symbiont clade specificity was observed among specimens of host species collected contemporaneously as well as those separated in time by up to 27 years.

Geographic distribution of collection sites

Figure 2 depicts the geographic distribution of collection sites for host specimens and associated *T. turnerae* isolates and clades.

Distribution of *T. turnerae* clades among host taxa

Clade I isolates were found in host species representing the genera *Dicyathifer*, *Neoterredo*, *Teredora*, and *Nausitora*. Clade II isolates were found in host species representing the genera *Teredo* and *Lyrodus*. Isolates representing clade I and clade II were found in the genus *Bankia*, in the species *B. gouldi* (clade I), *B. australis* and *B. rochi* (clade II) (see Figure 3).

Recombination between *T. turnerae* clades I and II

MaxChi, Chimaera, and 3Seq provided significant support ($P = 0.01$) for the conclusion that the *celAB* allele observed in isolate PMS-535T.S.1b.3 is the result of a recombination event between major parent T8602 (clade I) and minor parent T7902 (clade II) with breakpoints starting at position 372 of the aligned sequence set (Table 4). Similarly, results of analyses using GENECONV, MaxChi, BootScan, and 3Seq gave significant support ($P = 0.01$) for the conclusion that the *gyrB* allele observed in isolate T8510 results from a recombination event between major parent PMS-683H.S.1a.9 (clade II) and minor parent T8509 (clade I) with breakpoints starting at position 597 and ending at position 772 (Table 4). No difference in sequence length and no interruptions in reading frame were observed between the putative recombinant alleles and the consensus sequences of clades I and II. PCR amplification of the putative recombinants was reproducible in independent experiments using DNA extracts derived from the progeny of separate single colonies and so the putative recombinants are unlikely to be the result of PCR artifacts.

Evidence of lateral gene transfer

Evidence of lateral gene transfer was observed in the *recA* loci of isolates T8412, T8503, and T8513; all isolated from host species collected in the western Atlantic (Table 1). For each of these isolates the PCR product obtained was ~1600 nucleotides in length, ~740 nucleotides longer than the clade I and clade II consensus sequences. The sequence of these amplicons revealed fragments of two tandemly repeated, contiguous, non-identical *recA* loci (S4a), hereafter referred to as *recA*₁ and *recA*₂. In each isolate, sequence determined for the upstream locus, *recA*₁, encodes 163 amino acid residues corresponding to, and displaying 99.8% nucleotide sequence identity with, the 5' end of the *T. turnerae* clade II consensus sequence. This was followed by a short in frame sequence region encoding 35 amino acids without detectable nucleotide or amino acid homology to any sequence in the GenBank

nonredundant database. This in turn was followed by a region of 22 in-frame codons including 3 stop codons and a second sequence encoding 288 amino acid residues corresponding to the 5' end of a second putative *recA* coding sequence, *recA*₂, ending at the reverse primer. BLAST analysis of the observed fragment of the *recA*₂ locus against the Genbank non-redundant protein database showed this sequence to be most closely related (88.3% amino acid and 70.9 % nucleotide identity) to the *recA* gene of *Aeromonas simiae* (Genbank accession AEE38933.1). This sequence displayed 85.9% amino acid identity with the consensus sequences of the *T. turnerae* *recA* clades I and II. The two *recA* coding sequences also differ in %GC-content. The *recA*₁ sequences displayed 53%GC-content; closely matching the clade I and II consensus sequences (53%) as well as the %GC-content of the complete genome of *T. turnerae* isolate T7901 (50.8%) [50]. The *recA*₂ sequences displayed 41%GC-content. To exclude the possibility that this observation was the result of a PCR artifact, we independently amplified and sequenced fragments spanning the junction of the two *recA* loci using PCR primers specific to each of the detected *recA* alleles starting with genomic DNA from each of the three isolates, T8412, T8503, and T8513.

Analyses of selection

Analyses of synonymous and nonsynonymous substitutions, using four methods to determine selection, indicate an overall pattern of purifying selection for examined coding sequences (Table S2). No codon position received significant support for positive selection by more than one of the four methods applied.

Analysis of cophylogeny of hosts and symbionts

Randomization testing using TreeMap v.2 demonstrated that, given the observed data, neither the null hypothesis, that the symbiont phylogeny is no more congruent with the host phylogeny than would be a random symbiont tree with the same taxon set, nor the reciprocal null hypothesis, that the host phylogeny is no more congruent with the symbiont phylogeny than would be a random host tree with the same taxon set, could be rejected ($P = 0.05$).

Discussion

Here we report evidence that 25 isolates of the shipworm symbiont species *T. turnerae*, isolated from specimens representing a broad cross section of the phylogenetic and geographic range of their host species, form two distinct clades; a result that is well supported by independent analyses of each of five protein-coding loci examined and by the concatenated data set. The observed pattern of diversification for all loci demonstrates comparatively large sequence variation between clades as compared to low variability within clades (Table 3). Despite low within-clade variability, 24 of 25 examined isolates of this widespread shipworm symbiont species represent unique genotypes with respect to the examined loci.

The pattern of distribution of the two identified clades of *T. turnerae* among host species suggests a degree of specificity in sorting of symbiont clades among host species. Each host species examined was observed to be associated with a single symbiont clade. However, the reciprocal relationship was not observed. Instead, each symbiont clade is associated with

several host species. Also, although each host species was associated with a single *T. turnerae* clade, distinct genotypes of *T. turnerae* were found within different specimens of the same host species, and in one case within a single specimen of one host species. More comprehensive sampling of isolates from individual host species and specimens will be required to evaluate the extent of variability that may occur among *T. turnerae* isolates within and among shipworm species.

The specific association of each host species with just one *T. turnerae* clade was observed regardless of when and where the host specimens were collected. For example, all specimens of *Dicathifer manni* were found to be associated with *T. turnerae* isolates belonging to clade I, regardless of whether the host specimens were collected in Townsville, Australia in 1986 or in Bohol, Philippines in 2009 (Table 1). Similarly, clade II isolates were obtained from *Bankia rochi* whether collected in Karachi, Pakistan in 1984 or Sao Paulo, Brazil in 1985. Clade II isolates were obtained from specimens of *L. pedicellatus* collected in Cataban Island, Philippines in 2009 and in Long Beach, CA in 1979. Also, clade II isolates were obtained from descendants of the Long Beach specimens of *L. pedicellatus* even after maintenance in laboratory culture in natural seawater for twenty-seven years. These results suggest that the specific associations between individual host species and their respective *T. turnerae* clades are stable over long periods of time and great distances of geographic separation.

Consistent with these observations, host specimens and species harboring clade I and host specimens and species harboring clade II of *T. turnerae* were determined to be globally distributed (Figure 2) and to have vast and overlapping ranges. In half of the broadly distributed geographic locales examined, host specimens harboring clade I and host specimens harboring clade II isolates were found in close proximity, sometimes in the same piece of wood. These results suggest that geographic location may not be a critical determinant governing the occurrence of *T. turnerae* clades among host specimens and species. It should be noted, however, that although the sampling performed here was phylogenetically and geographically broad, it was not comprehensive. More comprehensive sampling will be required to establish the geographic dependence, or independence, of host-symbiont sorting in the *T. turnerae*-Teredinidae symbioses.

Similarly, the distribution or sorting of *T. turnerae* clades among host specimens examined here does not conform to host species boundaries as defined by anatomical [6] and/or molecular criteria [7]. On the contrary, symbiont clades are shared across broadly diverse host species (Table 1; Figure 3). Therefore, it seems unlikely that the observed divergence of *T. turnerae* into two distinct clades is explained by host and symbiont cospeciation. Cophylogeny at higher taxonomic levels is another possible explanation. When mapped to a previously published phylogenetic tree inferred for the family Teredinidae, clade I isolates appear to be primarily associated with more deeply branching taxa while clade II strains appeared in more recently emergent species [7] (Figure 3). However, this observation should be interpreted with caution. Not all of the host species sampled in this study were included in the previous phylogenetic analysis (and vice versa), sequences used in that study were obtained from different specimens than those sampled in this study, sampling across host taxa was not comprehensive, and a different set of genes was used to infer host and

symbiont phylogeny. Moreover, randomization tests show that the observed congruence of host and symbiont phylogenies is not significantly greater than might be expected to occur by chance among random phylogenies with the same number of taxa. Thus, analysis of additional host taxa and symbiont isolates will be necessary to determine to what extent the observed differentiation among symbiont clades may reflect coevolution of hosts and symbionts. We note however that the host taxa are well differentiated by analysis of highly conserved ribosomal genes, indicating comparatively ancient divergence, while within clades the symbiont strains are not well differentiated by more rapidly evolving protein coding genes, indicating more recent diversification. Therefore, while examination of additional genes may improve resolution among isolates within clades, this in itself should not be expected to improve congruence between host and symbiont trees.

Given the strong support observed for differentiation of *T. turnerae* isolates into two distinct clades, it is reasonable to ask whether this differentiation is the result of genetic drift, selection, or both. Analysis of the ratio of nonsynonymous and synonymous mutations can be used to infer positive or negative selection when this value either exceeds or falls short of the random expectation. Such analyses suggest that the examined gene loci are largely under negative selection (as might be expected for core functional genes) and indicate that the majority of the phylogenetic signal distinguishing clades in nucleotide sequence analyses is due to synonymous substitutions. Nonetheless, amino acid differences are observed in all loci, and with the exception of *recA*, which contains just a single amino acid substitution, trees inferred on the basis of amino acid sequences support the same clade boundaries as do nucleotide sequence trees (Figure S3), indicating that non-synonymous substitution also contributes to phylogenetic signal. By the criteria applied here ($P \leq 0.05$ or REL Bayes Factor ≥ 100 , posterior probability ≥ 0.95), no codon position received significant consensus support for positive selection from more than a single analytical method (Table S2). Based on these considerations it seems likely that drift rather than selection is responsible for the observed sequence differences among the examined loci. We note however that the examined genes, with the exception of *celAB*, are considered to be components of the core genome and so may not be strongly influenced by factors in the external environment. This does not preclude the possibility that other genes may be under positive selection.

While the majority of the 125 alleles examined in this study can be assigned unambiguously to either *T. turnerae* clade I or to clade II, exceptions were observed involving three loci (*celAB*, *gyrB* and *recA*) and five isolates (T8412, T8503, T8510, T8513, and PMS.535T.S. 1b.3). These exceptions shed light on the extent to which members of the two *T. turnerae* clades are isolated genetically. Putative recombinant alleles of the *celAB* and *gyrB* loci, apparently derived from recombination between clade I and clade II parent alleles, were observed in two *T. turnerae* isolates. Several analytical methods provide significant support ($P \leq 0.01$; Table 4) for these recombination events, suggesting that genetic exchange has occurred between members of the two clades since their divergence. This observation is not unprecedented. Evidence for symbiont exchange among host species [51] and homologous recombination [52] among symbiont isolates has been reported for gill endosymbionts of vesicomid clams, indicating that acquisition of the intracellular symbiotic habit may not preclude genetic exchange, even among symbionts of host species presumed to harbor

nearly monospecific symbiont populations and to acquire their symbionts primarily via vertical transfer. While the mode of transmission of *T. turnerae* in the shipworm symbiosis is unknown, these results suggest that some strains of the two *T. turnerae* clades have, at some point in time, been in sufficiently close physical proximity to allow genetic exchange. This might indicate that *T. turnerae* is environmentally transferred and that opportunity for genetic exchange has occurred in the external environment. Alternatively, concurrent multiple infections of individual hosts by members of both clades may occur with sufficient frequency to provide opportunities for genetic exchange between clades within the host tissues.

Similarly, features of the *recA* loci of three *T. turnerae* isolates (T8412, T8503 and T8513) suggest that genetic exchange between these and other bacterial species is possible. In these isolates portions of two distinct *recA* loci were observed, one of which, *recA*₁, appears to be a native locus that has been truncated by insertion of a sequence containing a second, apparently exogenous, *recA* locus (*recA*₂).

Several types of evidence support the exogenous origin of *recA*₂. First, with few reported exceptions, e.g. [53, 54], *recA* is present as a single copy gene in bacterial genomes examined to date [55], including the closed and finished genome sequence of *T. turnerae* isolate (T7901)[50]. Consistent with this, a single *recA* sequence was detected by PCR in genomes of all other *T. turnerae* isolates examined here. The truncated sequence of the *recA*₁ allele in each of the three isolates closely matches the *T. turnerae* clade II consensus sequence. All other loci examined in these isolates were also consistent with *T. turnerae* clade II. In contrast, in each of these isolates the sequence of the *recA*₂ allele is significantly divergent from the *recA* consensus sequences of both *T. turnerae* clades I and II, and shows equal or greater identity to *recA* sequences of distantly related gammaproteobacteria (Figure S4). The *recA*₂ sequences also differ significantly in %GC-content (41%) as compared to *T. turnerae* *recA*₁ sequences (53%) and to the complete genome of *T. turnerae* isolate T7901 (50.8%)[50]. Taken together, these observations suggest that the *recA*₂ locus in these three isolates entered these genomes as the result of a horizontal gene transfer event that disrupted the endogenous *recA* locus. It is most parsimonious to conclude that this event occurred after the divergence of clades I and II. Again this suggests that *T. turnerae*, regardless of its status as intracellular endosymbiont, is not entirely isolated from genetic exchange with other bacterial species. It remains unclear whether this putative genetic exchange was facilitated by exposure to other bacterial species in the external environment or to exchange with other symbiont species that coexist with *T. turnerae* within the host tissue.

In summary, sequence analysis of *T. turnerae* isolates examined here indicates the emergence of two distinct lineages within this ubiquitous shipworm symbiont species, suggesting diversification in the presence of a persistent barrier to genetic exchange between but not among members of the two clades. While each host species examined appears to associate with just one of the two *T. turnerae* clades, each *T. turnerae* clade is associated with several host species, each having wide and overlapping geographic distributions. To this extent, the observed pattern of diversification of *T. turnerae* does not appear to be consistent with, and therefore explained by, host geography or host speciation. However, additional sampling will be required to determine the extent to which geography,

environment or host and symbiont cophylogeny contribute to the assortment of *T. turnerae* clades among teredinid hosts. The results presented here also indicate that, in spite of the apparent genetic isolation of the two clades, both recombination between clades and acquisition of foreign DNA may contribute to the diversification of *T. turnerae* genomes. Finally, based on the results observed here, it is interesting to speculate that divergence among subpopulations of established endosymbiont types, rather than, or in addition to, de novo acquisition of new bacterial types from the environment, may have contributed to the observed coexistence of multiple endosymbiont types in shipworms.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported by NSF awards IOS092540 and IOS1258090 and the National Institutes of Health award U01TW008163. Philippine specimens were collected under a permit from the Philippine Department of Agriculture - Bureau of Fisheries and Aquatic Resources, and Prior Informed Consent certificates from the municipality of Talibon in the province of Bohol, where collection sites were located. We thank Romell Seronay, Joe Arbasto, Abigail Fusaro, Jemelyn Lapitan, Silberio Silver, Gary Rosenberg, Jovie Basalo, Isabelo Porfirio Aliño, and Meljune Chicote for collection assistance, and John Waterbury, Diana Franks and Frederica Valois of the Woods Hole Oceanographic Institution, Abigail J. Fusaro of Wayne State University and Roberta O'Connor of Tufts University for isolating and identifying bacterial isolates. We also acknowledge the efforts of Ruth D. Turner (Harvard University; deceased 2000), whose expertise in the biology and systematics of Terebinidae made this work possible.

References

1. Turner, RD. A survey and illustrated catalogue of the Terebinidae (Mollusca: Bivalvia). Harvard University; The Museum of Comparative Zoology; Cambridge, MA: 1966. p. 265
2. Mann R, Gallager SM. Physiological and biochemical energetics of larvae of *Teredo navalis* and *Bankia gouldi* (Bartsch). *Journal of Experimental Marine Biology and Ecology*. 1985; 85:211–228.
3. Gallager SM, Turner RD, Berg CJ. Physiological aspects of wood consumption, growth, and reproduction in the shipworm *Lyrodus pedicellatus* Quatrefages. *Journal of Experimental Marine Biology and Ecology*. 1981; 52:63–77.
4. Distel, DL. The biology of marine wood boring bivalves and their bacterial endosymbionts.. In: Goodell, B.; Nicholas, DD.; Schultz, TP., editors. *Wood Deterioration and Preservation*. American Chemical Society Press; Washington: 2003. p. 253-271.
5. Sellius G. *Historia naturalis Terebinis seu Xylophagi marini, tubullo chonchoidis speciatim Belgeci*. J. Christoph. 1753
6. Moore, RC., editor. *Treatise on invertebrate paleontology*. Vol. 1. The Geological Society of America, Inc. and the University of Kansas; 1969. p. 489
7. Distel DL, Amin M, Burgoyne A, Linton E, Mamangkey G, Morrill W, Nove J, Wood N, Yang J. Molecular phylogeny of Pholadoidea Lamarck, 1809 supports a single origin for xylophagy (wood feeding) and xylophagous bacterial endosymbiosis in Bivalvia. *Molecular Phylogenetics and Evolution*. 2011; 61(2):245–54. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=21684342<http://www.sciencedirect.com/science/article/pii/S1055790311002612>. [PubMed: 21684342]
8. Edmondson CH. *Dispersal of Shipworms Among Central Pacific Islands, with Descriptions of New Species*: Bernice P. Bishop Museum. 1946
9. Scheltema RS. Dispersal of phytoplanktotrophic shipworm larvae (Bivalvia: Terebinidae) over long distances by ocean currents. *International Journal on Life in Oceans and Coastal Waters*. 1971; 11(1):5–11.

10. Waterbury JB, Calloway CB, Turner RD. A cellulolytic-nitrogen fixing bacterium cultured from the gland of Deshayes in shipworms (Bivalvia: Teredinidae). *Science*. 1983; 221:1401–1403. <http://www.ncbi.nlm.nih.gov/pubmed/17759016>. [PubMed: 17759016]
11. Tanimura A, Liu W, Yamada K, Kishida T, Toyohara H. Animal cellulases with a focus on aquatic invertebrates. *Fisheries Science*. 2013; 79(1):1–13. <http://dx.doi.org/10.1007/s12562-012-0559-4>.
12. Betcher MA, Fung JM, Han AW, O'Connor R, Seronay R, Concepcion GP, Distel DL, Haygood MG. Microbial Distribution and Abundance in the Digestive System of Five Shipworm Species (Bivalvia: Teredinidae). *PLoS ONE*. 2012; 7(9):e45309. <http://dx.doi.org/10.1371/journal.pone.0045309>. [PubMed: 23028923]
13. Deshayes GP. Histoire naturelle des mollusques. Exploration Scientifique l'Algerie. *Zoologie*. 1. 1848:35–76.
14. Popham JD, Dickson MR. Bacterial associations in the teredo *Bankia australis* (Lamellibranchia, Mollusca). *Marine Biology*. 1973; 19:338–340.
15. Popham JD. Further observations of the gland of Deshayes in the teredo *Bankia australis*. *The Veliger*. 1974; 18(1):55–59.
16. Trytek RE, Allen WV. Synthesis of amino acids by bacterial symbionts in the gills of the shipworm *Bankia setacea* (Tryon). *Comparative Biochemistry and Physiology*. 1980; 67A:419–427.
17. Stewart MG, Dean RC. Uptake and utilization of amino acids by the shipworm *Bankia gouldi*. *Comparative Biochemistry and Physiology*. 1980; 66B:443–450.
18. Distel DL, Morrill W, MacLaren-Toussaint N, Franks D, Waterbury J. *Teredinibacter turnerae* gen. nov., sp. nov., a dinitrogen-fixing, cellulolytic, endosymbiotic gamma-proteobacterium isolated from the gills of wood-boring molluscs (Bivalvia: Teredinidae). *International Journal of Systematic and Evolutionary Microbiology*. 2002; 52(Pt 6):2261–9. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12508896. [PubMed: 12508896]
19. Imam SH, Greene RV, Hockridge ME. Zymographic analyses of carboxymethylcellulases secreted by the bacterium from wood-boring marine shipworms. *Biotechnology Techniques*. 1993; 7(8): 579–584.
20. Greene RV, Griffin HL, Freer SN. Purification and characterization of an extracellular endoglucanase from the marine shipworm bacterium. *Archives of Biochemistry and Biophysics*. 1988; 267(1):334–341. <http://www.sciencedirect.com/science/article/pii/0003986188900392>. [PubMed: 3196031]
21. Griffin HL, Freer SN, Greene RV. Extracellular endoglucanase activity by a novel bacterium isolated from marine shipworm. *Biochemical and Biophysical Research Communications*. 1987; 144(1):143–151. [PubMed: 3579899]
22. Greene RV. Challenges from the sea: marine shipworms and their symbiotic bacteria. *Society for Industrial Microbiology News*. 1994; 44(2):51–59.
23. Distel DL, DeLong EF, Waterbury JB. Phylogenetic characterization and in situ localization of the bacterial symbiont of shipworms (Teredinidae: Bivalvia) by using 16S rRNA sequence analysis and oligodeoxynucleotide probe hybridization. *Applied and Environmental Microbiology*. 1991; 57(8):2376–82. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=1722662. [PubMed: 1722662]
24. Distel DL, Beaudoin DJ, Morrill W. Coexistence of multiple proteobacterial endosymbionts in the gills of the wood-boring Bivalve *Lyrodus pedicellatus* (Bivalvia: Teredinidae). *Applied and Environmental Microbiology*. 2002; 68(12):6292–9. <http://www.ncbi.nlm.nih.gov/pubmed/12450854>. [PubMed: 12450854]
25. Luyten YA, Thompson JR, Morrill W, Polz MF, Distel DL. Extensive variation in intracellular symbiont community composition among members of a single population of the wood-boring bivalve *Lyrodus pedicellatus* (Bivalvia: Teredinidae). *Applied and Environmental Microbiology*. 2006; 72(1):412–7. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16391072. [PubMed: 16391072]
26. Lechene CP, Luyten Y, McMahon G, Distel DL. Quantitative imaging of nitrogen fixation by individual bacteria within animal cells. *Science*. 2007; 317(5844):1563–6. <http://>

[www.ncbi.nlm.nih.gov/entrez/query.fcgi?](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17872448)

[cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17872448](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17872448). [PubMed: 17872448]

27. Elshahawi, SI.; Trindade-Silva, AE.; Hanora, A.; Han, AW.; Flores, MS.; Vizzoni, V.; Schrago, CG.; Soares, CA.; Concepcion, GP.; Distel, DL.; Schmidt, EW.; Haygood, MG. Boronated tartrolon antibiotic produced by symbiotic cellulose-degrading bacteria in shipworm gills.. Proceedings of the National Academy of Sciences of the United States of America. 2013. <http://www.ncbi.nlm.nih.gov/pubmed/23288898>
28. Sipe AR, Wilbur AE, Cary SC. Bacterial symbiont transmission in the wood-boring shipworm *Bankia setacea* (Bivalvia: Teredinidae). Appl Environ Microbiol. 2000; 66(4):1685–91. <http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?db=m&form=6&dopt=r&uid=0010742261><http://aem.asm.org/cgi/content/full/66/4/1685> <http://aem.asm.org/cgi/content/abstract/66/4/1685>. [PubMed: 10742261]
29. Newell, ND. Classification of Bivalvia.. In: Moore, RC., editor. Treatise on Invertebrate Paleontology. The Geological Society of America and the University of Kansas; 1969. p. 205-224.
30. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research. 2002; 30(14):3059–66. <http://www.ncbi.nlm.nih.gov/pubmed/12136088>. [PubMed: 12136088]
31. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009; 25(11):1451–2. <http://www.ncbi.nlm.nih.gov/pubmed/19346325>. [PubMed: 19346325]
32. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. Nature methods. 2012; 9(8):772. <http://www.ncbi.nlm.nih.gov/pubmed/22847109>. [PubMed: 22847109]
33. Cunningham CW. Can tree incongruence tests predict when data should be combined? Molecular Biology and Evolution. 1997; 14(7):733–40. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9214746. [PubMed: 9214746]
34. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 2003; 19(12):1572–4. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12912839. [PubMed: 12912839]
35. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 2001; 17(8):754–5. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11524383. [PubMed: 11524383]
36. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology. 2003; 52(5):696–704. <http://www.ncbi.nlm.nih.gov/pubmed/14530136>. [PubMed: 14530136]
37. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution. 2011; 28(10):2731–9. <http://www.ncbi.nlm.nih.gov/pubmed/21546353>. [PubMed: 21546353]
38. Charleston MA, Robertson DL. Preferential Host Switching by Primate Lentiviruses Can Account for Phylogenetic Similarity with the Primate Phylogeny. Systematic Biology. 2002; 51(3):528–535. <http://sysbio.oxfordjournals.org/content/51/3/528.abstract>. [PubMed: 12079649]
39. Padidam M, Sawyer S, Fauquet CM. Possible emergence of new geminiviruses by frequent recombination. Virology. 1999; 265(2):218–25. <http://www.ncbi.nlm.nih.gov/pubmed/10600594>. [PubMed: 10600594]
40. Martin DP, Posada D, Crandall KA, Williamson C. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. AIDS Research and Human Retroviruses. 2005; 21(1):98–102. <http://www.ncbi.nlm.nih.gov/pubmed/15665649>. [PubMed: 15665649]
41. Smith JM. Analyzing the mosaic structure of genes. Journal of Molecular Evolution. 1992; 34(2): 126–9. <http://www.ncbi.nlm.nih.gov/pubmed/1556748>. [PubMed: 1556748]
42. Posada D, Crandall KA. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proceedings of the National Academy of Sciences of the United States of

- America. 2001; 98(24):13757–62. <http://www.ncbi.nlm.nih.gov/pubmed/11717435>. [PubMed: 11717435]
43. Boni MF, Posada D, Feldman MW. An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics*. 2007; 176(2):1035–47. <http://www.ncbi.nlm.nih.gov/pubmed/17409078>. [PubMed: 17409078]
 44. Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefevre P. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics*. 2010; 26(19):2462–3. <http://www.ncbi.nlm.nih.gov/pubmed/20798170>. [PubMed: 20798170]
 45. Holm S. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*. 1979; 6:65–70.
 46. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K. FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Molecular Biology and Evolution*. 2013; 30(5):1196–205. <http://www.ncbi.nlm.nih.gov/pubmed/23420840>. [PubMed: 23420840]
 47. Delpont W, Poon AF, Frost SD, Kosakovsky Pond SL. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics*. 2010; 26(19):2455–7. <http://www.ncbi.nlm.nih.gov/pubmed/20671151>. [PubMed: 20671151]
 48. Pond SL, Frost SD. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics*. 2005; 21(10):2531–3. <http://www.ncbi.nlm.nih.gov/pubmed/15713735>. [PubMed: 15713735]
 49. Pond SL, Frost SD, Muse SV. HyPhy: hypothesis testing using phylogenies. *Bioinformatics*. 2005; 21(5):676–9. <http://www.ncbi.nlm.nih.gov/pubmed/15509596>. [PubMed: 15509596]
 50. Yang JC, Madupu R, Durkin AS, Ekborg NA, Pedamallu CS, Hostetler JB, Radune D, Toms BS, Henrissat B, Coutinho PM, Schwarz S, Field L, Trindade-Silva AE, Soares CA, Elshahawi S, Hanora A, Schmidt EW, Haygood MG, Posfai J, Benner J, Madinger C, Nove J, Anton B, Chaudhary K, Foster J, Holman A, Kumar S, Lessard PA, Luyten YA, Slatko B, Wood N, Wu B, Teplitski M, Mougous JD, Ward N, Eisen JA, Badger JH, Distel DL. The complete genome of *Teredinibacter turnerae* T7901: an intracellular endosymbiont of marine wood-boring bivalves (shipworms). *PLoS ONE*. 2009; 4(7):e6085. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19568419. [PubMed: 19568419]
 51. Stewart FJ, Cavanaugh CM. Pyrosequencing analysis of endosymbiont population structure: co-occurrence of divergent symbiont lineages in a single vesicomid host clam. *Environmental Microbiology*. 2009; 11(8):2136–47. <http://www.ncbi.nlm.nih.gov/pubmed/19397674>. [PubMed: 19397674]
 52. Stewart FJ, Young CR, Cavanaugh CM. Evidence for homologous recombination in intracellular chemosynthetic clam symbionts. *Molecular Biology and Evolution*. 2009; 26(6):1391–404. <http://www.ncbi.nlm.nih.gov/pubmed/19289597>. [PubMed: 19289597]
 53. Nahrstedt H, Schroder C, Meinhardt F. Evidence for two recA genes mediating DNA repair in *Bacillus megaterium*. *Microbiology*. 2005; 151(Pt 3):775–87. <http://www.ncbi.nlm.nih.gov/pubmed/15758224>. [PubMed: 15758224]
 54. Norioka N, Hsu MY, Inouye S, Inouye M. Two recA genes in *Myxococcus xanthus*. *Journal of Bacteriology*. 1995; 177(14):4179–82. <http://www.ncbi.nlm.nih.gov/pubmed/7608099>. [PubMed: 7608099]
 55. Lin Z, Kong H, Nei M, Ma H. Origins and evolution of the recA/RAD51 gene family: evidence for ancient gene duplication and endosymbiotic gene transfer. *Proceedings of the National Academy of Sciences of the United States of America*. 2006; 103(27):10328–33. <http://www.ncbi.nlm.nih.gov/pubmed/16798872>. [PubMed: 16798872]

Figure 1. Phylogenetic trees inferred for *T. turnerae* isolates based on partial sequences of five protein-coding loci and the concatenated set

The displayed trees were inferred using Bayesian inference and are identical in topology with respect to all well supported nodes (posterior probability 0.90, bootstrap proportions 70%) to trees inferred from the same data by Maximum Likelihood (data not shown). Numerical values at the nodes represent Bayesian posterior probability estimates 0.90 (left) and maximum likelihood bootstrap proportions 70% (right) determined for the corresponding node in a topologically identical Maximum Likelihood consensus tree. Trees are midpoint rooted. Circles, clade I; diamonds, clade II; squares, putative recombinant alleles.

Figure 2. Map depicting geographic distribution of collection sites for host specimens and associated *T. turnerae* isolates and clades

Dashed lines indicate approximate northernmost (top) and southernmost (bottom) boundaries of occurrence of the Family Teredinidae. Circles, clade I; diamonds, clade II.

Figure 3. Occurrence of *T. turnerae* clades I and II mapped to host phylogeny

Occurrences of symbiont clade I and clade II isolates are shown here mapped to a phylogram depicting the proposed evolutionary relationships among their respective hosts. The displayed phylogram is a subtree excerpted from a larger tree, originally published in [10], that includes 38 additional taxa. The displayed trees were inferred using Bayesian analysis of concatenated partial sequences of small and large subunit nuclear rRNA genes. Numerical values that appear at selected nodes indicate posterior probabilities (0.90) determined from Bayesian analyses (left) and bootstrap proportions (70%) determined for a topologically identical consensus tree inferred by Maximum Likelihood. Details of these analyses are found in [10]. Circles, clade I; diamonds, clade II.

Table 1

Isolates, hosts, locations, loci and clades

Symbiont		Host			Clade					
Isolate	Genus	Species	Location	Latitude ²	Longitude ²	gyrB	sseA	celAB	recA	rpoB
T8602	Dicyathifer	manni	Townsville, QLD, Aus.	S 19.27631 ³	E 147.05784 ³	I	I	I	I	I
PMS-539Y.S.1a.1	Dicyathifer	manni	Bohol, Isl, Philippines	N 10.102222	E 124.334722	I	I	I	I	I
T8402	Teredora	malleolus	Floating wood	N 38.30667	W 69.59333	I	I	I	I	I
T8508	Neoterredo	reynei	Ubatuba, Brazil	S 23.49127 ³	W 45.16385 ³	I	I	I	I	I
T8509	Nausitora	fuscula	Ubatuba, Brazil	S 23.49127 ³	W 45.16385 ³	I	I	I	I	I
T7901	Bankia	gouldi	Beaufort, NC, USA	N 34.71737 ³	W 76.67198 ³	I	I	I	I	I
T8415	Bankia	gouldi	Fort Pierce, FL, USA	N 27.48063 ³	W 80.30967 ³	I	I	I	I	I
T0609 [*]	Lyrodus	pedicellatus	Long Beach, CA, USA	N 33.76138 ³	W 118.17281 ³	II	II	II	II	II
T0611 [*]	Lyrodus	pedicellatus	Long Beach, CA, USA	N 33.76138 ³	W 118.17281 ³	II	II	II	II	II
T7902	Lyrodus	pedicellatus	Long Beach, CA, USA	N 33.76138 ³	W 118.17281 ³	II	II	II	II	II
PMS-554M.S.1a.4	Lyrodus	pedicellatus	Cataban Isl., Philippines	N 10.233333	E 124.383333	II	II	II	II	II
PMS-574K.S.1a.1	Lyrodus	pedicellatus	Cataban Isl., Philippines	N 10.233333	E 124.383333	II	II	II	II	II
T8503	Lyrodus	floridanus	Key Biscayne, FL, USA	N 25.68976 ³	W 80.16380 ³	II	II	II	II/LGT ³	II
T8412	Lyrodus	bipartitus	Jim Isl., Fort Pierce, FL	N 27.476944	W 80.311944	II	II	II	II/LGT ³	II
T8513	Teredo	navalis	Sao Paulo, Brazil	S 23.81992 ³	W 45.40517 ³	II	II	II	II/LGT ³	II
T7903	Teredo	navalis	Woods Hole, MA, USA	N 41.52340 ³	W 70.67228 ³	II	II	II	II	II
PMS.683H.S.1a.9	Teredo	clappi	Cataban Isl., Philippines	N 10.233333	E 124.383333	II	II	II	II	II
PMS.535T.S.1b.3	Teredo	clappi	Cataban Isl., Philippines	N 10.233333	E 124.383333	II	II	I / II	II	II
PMS.509L.S.1a.4 ^{**}	Teredo	clappi	Bohol Isl., Philippines	N 10.307778	E 124.399167	II	II	II	II	II
PMS.509L.S.1a.6 ^{**}	Teredo	clappi	Bohol Isl., Philippines	N 10.307778	E 124.399167	II	II	II	II	II
PMS.517Y.S.1a.10	Teredo	mindanensis	Cataban Isl., Philippines	N 10.233333	E 124.383333	II	II	II	II	II
PMS.691W.S.1a.3	Teredo	mindanensis	Cataban Isl., Philippines	N 10.233333	E 124.383333	II	II	II	II	II

Symbiont Isolate ¹	Host				Clade					
	Genus	Species	Location	Latitude ²	Longitude ²	gyrB	sseA	celAB	recA	rpoB
T8401	Bankia	rochi	Karachi, Pakistan	N 24.84107 ³	E 66.91069 ³	II	II	II	II	II
T8510	Bankia	rochi	Sao Sebastiao, Brazil	S 23.81992 ³	W 45.40517 ³	I/ II	II	II	II	II
T8601	Bankia	australis	Brisbane, QLD, Aus.	S 27.78624 ³	E 153.46184 ³	II	II	II	II	II

¹ All isolates were obtained from separate host specimens except where otherwise indicated.

² Latitude and longitude are estimated with ~10m accuracy except where otherwise indicated

³ LGT = putative lateral gene transfer

* Estimated at ~10km accuracy

** T0609 and T0611 were isolated from the same specimen of *L. pedicellatus*

*** PMS.509L.S.1a.4 and PMS.509L.S.1a.6 were isolated from the same specimen of *T. clappi*.

Table 2

Amplification and sequencing primers

Locus	Product	Forward primer sequence (5' to 3') ^a	Reverse primer sequence (5' to 3') ^a	Annealing temp ^a	Expected amplicon size (bp)
<i>rrsA</i> , <i>B</i> , and <i>C</i>	Small subunit (16S) ribosomal RNA	27F - AGAGTTTGATCMTGGCTCAG 530F ^b - GTGCCAGCMGCCGCGG 1055F ^b - ATGGCTGTCGTCAGCT	1492R - TACGGYTACCTTGTTACGACTT 800R ^b - TACCAGGGTATCTAATCC 1101R ^b - AGGGTTGCGCTCGTT	56°C	1495
<i>rpoB</i>	DNA-directed RNA polymerase β-subunit	rpoB-F1 - ATGACAGACAACGGTACWTTTCGTWRTBAAYG rpoB-F2 ^b - GAAGANGGCCGTCGTATTACTGC	rpoB-R1 - GTGAGTCGGGTGTACGTCWCGYACYTC rpoB-R2 ^b - GGCGCATCATGCGGTARATYTCDAACCAR	61°C	1277
<i>gyrB</i>	DNA gyrase β-subunit	gyrB-F - CATGCTGCGGTAATTCGA	gyrB-R - TCTACGTCCGCATCCGTCAT	60°C	1211
<i>recA</i>	DNA recombinase A	recA-F1 - GCCAGTTTGTAAAGGCAACBGTTBATGCGYM recA-F2 ^b - GGTCGTTTCAGGCCGTCARTHGAYGTYT recA _{1-1'} -F ^{b,c} - CGTATGGGTGACAAAAGAGCTGGAAG	recA-R1 - CGCCGTTGTAGCTATACCARGCRCCV recA-R2 ^b - CGCAGCGCTTGGCACATYARTCGMGCY recA ₂ -R ^{b,c} - GAGCTGCAACAGAGTCAATAATTACA	63°C 60°C	831 1150
<i>sseA</i>	Thiosulfate/3-mercaptopyruvate sulfurtransferase	sseA-F - ATGGCTAATTCACCCCTGGTG	sseA-R - TTAGGTACTGGTGCGGATGGG	60°C	849
<i>celAB</i>	Bifunctional glycoside hydrolase	celAB-F - CCACCTCTGCAGCTTTCGCGG	celAB-R - GCCGTTACCGATTGTGGTGCC	61°C	941
	Cloned PCR amplicons	M13-Forward(-20) - GTAAAAACGACGGCCAG	M13-Reverse - CAGGAAAACAGCTATGAC		

^a PCR conditions: initial denaturation for 3 min at 98°C; 35 cycles of 30 sec at 98°C, 30 sec at annealing temperature, 1 min at 72°C; final extension for 7min at 72°C; hold at 4°C.

^b Internal sequencing primer

^c Internal primer pair specific to the *recA*₁ and *recA*₂ alleles of isolates T8412, T8503, and T8513. Complementary sites of *recA* primers are shown in Figure S2

Table 3

Mean interclade and intraclade distances (substitutions/nucleotide position calculated from ungapped nucleotide alignments¹).

Gene	All pairwise distances ^{1, 4}	Intraclade ²		Interclade ³
		Clade I	Clade II	Clade I to Clade II
<i>gyrB</i>	0.059	0.020	0.014	0.118
<i>rpoB</i>	0.047	0.003	0.006	0.103
<i>recA</i>	0.022	0.014	0.007	0.041
<i>celAB</i>	0.027	0.003	0.007	0.054
<i>sseA</i>	0.038	0.011	0.010	0.078
All loci (concatenated)	0.043	0.010	0.009	0.086

Brackets indicate significant differences between mean values based on all loci concatenated (independent *t*-test, $P < 0.01$)

¹ Excluding putative recombinant and laterally transferred alleles

² Mean of pairwise distances within clade

³ Mean of pairwise distances between clades

⁴ Mean of all pairwise distances

Table 4

Evidence for proposed recombination events

Locus	P value						0.01 scaled using a Bonferroni correction ^a		Position of breakpoints ^b	Recombinant allele (major and minor parents)
	RDP	GENECONV	BootScan	MaxChi	Chinaera	3Seq				
<i>celAB</i>	-	-	-	6.884×10^{-4}	5.935×10^{-3}	1.122×10^{-4}	372 ^c	PMS-535T.S.1b.3 (T8602 and T7902) ^c		
<i>gyrB</i>	-	2.427×10^{-4}	3.861×10^{-8}	7.166×10^{-3}	-	1.161×10^{-7}	597, 777 ^c	T8510 (PMS-683H.S.1a.9 and T8509) ^c		

Only P values 0.01 are displayed

^b In the aligned dataset

^c Inferred by 3Seq