Evidence for Homologous Recombination in Intracellular Chemosynthetic Clam Symbionts

Frank J. Stewart, Curtis R. Young,† and Colleen M. Cavanaugh†*

*Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA; and †Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA

Homologous recombination is a fundamental mechanism for the genetic diversification of free-living bacteria. However, recombination may be limited in endosymbiotic bacteria, as these taxa are locked into an intracellular niche and may rarely encounter sources of foreign DNA. This study tested the hypothesis that vertically transmitted endosymbionts of deep-sea clams (Bivalvia: Vesicomyidae) show little or no evidence of recombination. Phylogenetic analysis of 13 loci distributed across the genomes of 14 vesicomyid symbionts revealed multiple, well-supported inconsistencies among gene tree topologies, and maximum likelihood–based tests rejected a hypothesis of shared evolutionary history (linkage) among loci. Further, multiple statistical methods confirmed the presence of recombination by detecting intragenic breakpoints in two symbiont loci. Recombination may be confined to a subset of vesicomyid symbionts, as some clades showed high levels of genomic stability, whereas others showed clear patterns of homologous exchange. Notably, a mosaic genome is present in symB, a symbiont lineage shown to have been acquired laterally (i.e., nonvertically) by *Vesicomya* sp. JdF clams. The majority of loci analyzed here supported a tight sister clustering of symB with the symbiont of a host species from the Mid-Atlantic Ridge, whereas others placed symB in a clade with symA, the dominant phylotype of *V.* sp. JdF clams. This result raises the hypothesis that lateral symbiont transfer between hosts may facilitate recombination by bringing divergent symbiont lineages into contact. Together, the data show that homologous recombination contributes to the diversification of vesicomyid clam symbionts, despite the intracellular lifestyle of these bacteria.

Introduction

Vertically transmitted (parent to offspring) bacterial endosymbionts face severe evolutionary constraints. As these bacteria are effectively confined to an intracellular niche, opportunities for the horizontal exchange of genetic material may drastically decline or stop entirely (Silva et al. 2003; Papke and Ward 2004; Wernegreen 2005). This is hypothesized to be true for genetic exchange involving mobile elements (e.g., plasmids, phage, and transposable elements), as well as for homologous recombination, the process by which DNA in a donor genome replaces homologous DNA in a recipient genome (Bordenstein and Reznikoff 2005). Without these mechanisms, the generation of novel sequences in endosymbionts is driven only by mutations in existing alleles and by intragenomic recombination events (e.g., duplications; Narra and Ochman 2006). Further, as genetic deletions cannot be balanced by the influx of new material, endosymbiont genomes experience a characteristic size reduction, often containing only the core genes required for life in the host cytoplasm or those forming the functional basis of the symbiotic interaction (Moran and Mira 2001; Pérez-Brocal et al. 2006). Similarly, in the absence of recombination, mildly harmful mutations can be removed from a lineage only through rare back mutational events, and therefore tend to accumulate in symbiont genomes, contributing to the overall pattern of genomic degradation (Muller 1964; Moran 1996). These evolutionary consequences of endosymbiosis have been studied most thoroughly for mutualistic symbionts of insects (e.g., *Buchnera* sp., the symbiont of aphids), pathogenic bacteria, and organellar genomes (e.g., of the mitochondrion) that originated from ancient endosymbiotic events (Wernegreen 2004, 2005). However, vertically transmitted endosymbioses occur widely in nature, involving both pathogenic and beneficial symbiont–host interactions. Additional research will show the extent to which general features of symbiont evolution, including patterns of recombination, vary across the tree of life.

Endosymbioses between chemosynthetic proteobacteria and marine invertebrates constitute unique and relatively unexplored systems for testing models of symbiont evolution. In these mutualisms, which dominate the fauna at deep-sea hydrothermal vents and cold seeps, the invertebrate host facilitates access to substrates (sulfur or methane, oxygen, and carbon dioxide) needed for the chemoautotrophic metabolism of the internal symbiont. In exchange, symbiont carbon fixation supports most, if not all, of the host's nutrition (Stewart et al. 2005). These associations have evolved in at least six invertebrate phyla (Cavanaugh et al. 2006) but thus far have not lent themselves to the same comparative genomic and evolutionary analyses that are proving informative for insect endosymbioses. Fortunately, recent genome sequencing projects are helping to reverse this trend, providing a valuable comparative framework for exploring the impact of obligate intracellularity on bacterial evolution (Woyke et al. 2006; Kuwahara et al. 2007; Newton et al. 2007). Facilitated by such a framework, the present study uses multi-locus sequence data to test for recombination in chemosynthetic symbionts of deep-sea clams (Vesicomyidae). Specifically, we test the hypothesis that obligate chemosynthetic endosymbionts exhibit limited or no evidence of homologous recombination.

Clams of the Vesicomyidae and their intracellular symbionts are ideal systems for examining genome diversification as a symbiont nears total integration with its host. Occurring ubiquitously throughout the world's oceans, vesicomyids form an ecologically dominant component of hydrothermal vent and seep communities, obtaining almost all of their nutrition from sulfur-oxidizing bacteria living in specialized cells of the clam gills. Vesicomyid symbionts

Key words: genome evolution, recombination, lateral transmission, chemosynthesis, *Calyptogena*, hydrothermal vent.

E-mail: cavanaug@fas.harvard.edu.

Mol. Biol. Evol. 26(6):1391–1404. 2009
doi:10.1093/molbev/msp049
Advance Access publication March 16, 2009

Table 1
Vesicomyid Endosymbionts[a] Examined in this Study

Host Species	Location[b]	Habitat	Latitude, Longitude	Dive[c]	Depth (m)
Clade[a]					
gigas/kilmeri					
Vesicomya gigas	P, Mendocino Fault	Seep	40–20.06N, 127–-39.69W	T 473	1,694
Calyptogena kilmeri	P, Monterey Canyon	Seep	36–46.53N, 122–5.21W	V	970
Calyptogena okutanii[d]	P, Sagami Bay, Japan	Seep	34–57N, 139–12E	HP 305	1,157
Ectenagena extenta	P, Monterery Canyon	Seep	36–40.92N, 122–7.21W	V	1,464
Vesicomya sp. GOM	A, Gulf of Mexico	Seep	26–22.04N, 94–31.13W	J2 283	2,283
cordata/ponderosa					
Vesicomya cordata	A, Gulf of Mexico	Seep	27–40.88N, 91–32.10W	J 3276	720
Calyptogena ponderosa	A, Gulf of Mexico	Seep	27–40.88N, 91–32.10W	J 3276	720
Vesicomya sp. mt-I[e]	P, Monterey Bay	Seep	36–47N, 122–5W	V	~1,000
Vesicomya sp. JdF (mt-II) symA[f]	P, Juan de Fuca Ridge	Vent	47–57.79N, 129–05.48W	A 3459	2,190
Vesicomya sp. mt-III[d]	P, Monterey Bay	Seep	36–37.79N, 122–19.92W	T 217	2,200
MAR/BR					
Vesicomya sp. BR	A, Blake Ridge	Seep	32–29.69N, 76–11.08W	A 3710	2,155
Vesicomya sp. MAR	A, Mid-Atlantic Ridge	Vent	14–45.32N, 44–58.79W	A 3668	3,028
Vesicomya sp. JdF (mt-II) symB[f]	P, Juan de Fuca Ridge	Vent	47–57.79N, 129–05.48W	A 3459	2,190
Calyptogena magnifica[d]	P, East Pacific Rise	Vent	09–50.88N, 104–17.61W	A 3951	2,507

[a] Endosymbionts: named according to host taxa and subdivided into clades resolved by analysis of the concatenated symbiont loci (described in Results and fig. 2).
[b] Location: sites of host clam collection—also see Supplementary figure 2; A = Atlantic Ocean, P = Pacific Ocean.
[c] Dive: A = *DSV* Alvin, ATV = *ROV* Advanced Tethered Vehicle (Dive 95-48-149, High Rise Expedition), HP = *ROV* Hyper Dolphin, J = *DSV* Johnson Sea Link, J2 = *ROV* Jason II, T = *ROV* Tiburon, V = *ROV* Ventana (no associated dive numbers).
[d] *Calyptogena magnifica* and *C. okutanii* symbiont sequences obtained from GenBank NC008610 and NC009465; collection details are in Newton et al. (2007) and Kuwahara et al. (2007).
[e] *Vesicomya* sp. mt-I, II, and III are unnamed clam species that have been designated according to mitochondrial type (mt) by Goffredi et al. (2003); here, *Vesicomya* sp. JdF = *Vesicomya* sp. mt-II.
[f] *Vesicomya* sp. JdF clams host two distinct symbiont lineages, symA and symB (Stewart et al. 2008).

(gamma Proteobacteria) are presumed to be obligately symbiotic, as they have yet to be cultured or detected in a free-living form, have a relatively reduced genome size (Kuwahara et al. 2007; Newton et al. 2007), and are transmitted vertically between successive host generations via the egg (Endow and Ohta 1990; Cary and Giovannoni 1993; Peek et al. 1998; Hurtado et al. 2003). However, recent evidence suggests that some clam lineages may acquire divergent symbionts laterally via uptake from an environmental population or horizontal transfer from a co-occurring host (Stewart et al. 2008). Lateral symbiont acquisition may open up avenues for homologous recombination, potentially by bringing together distinct symbiont strains in the same host individual (e.g., Jiggins 2002). Precedence for such a pattern is evident in *Wolbachia*, the intracellular parasite of insects. *Wolbachia*, which undergoes maternal transmission as well as occasional instances of lateral symbiont transfer (e.g., Haine et al. 2005; Sintupachee et al. 2006), exhibits extensive evidence of recombination (Bordenstein and Reznikoff 2005; Baldo et al. 2006). In contrast, several mutualistic symbionts of insects (e.g., *Buchnera*, *Wigglesworthia*) appear to undergo nearly strict vertical transmission (e.g., Chen et al. 1999; Clark et al. 2000; Thao and Baumann 2004) and show minimal evidence of recombination (Tamas et al. 2002; van Ham et al. 2003; Wernegreen 2004, 2005; Degnan et al. 2005). Here, we examine where chemosynthetic clam symbionts fall along this continuum.

Using phylogenetic reconstructions and statistically based recombination detection methods, this study reveals evidence of homologous recombination across multiple clam symbiont loci, a pattern akin to that of *Wolbachia* and other obligate endosymbionts that show at least some level of lateral transmission. The detection of recombination in clam symbionts has important implications for chemosynthetic endosymbiont phylogeny and suggests a mechanism by which certain symbiont lineages may partially escape the debilitating effects of obligate intracellularity (e.g., Muller's rachet; Moran 1996). This study complements a growing body of literature challenging models of genomic stasis in obligate intracellular bacteria (Bordenstein and Reznikoff 2005; Sirand-Pugnet et al. 2007).

Materials and Methods

Vesicomyid Clam Symbionts

Symbiont strains ($n = 14$ total) used in the analysis are identified in table 1 according to host clam species. Symbiont-containing gill tissue was obtained from 11 species of vesicomyid clams inhabiting hydrothermal vent and cold seep environments in both the Atlantic and Pacific Oceans (table 1 and map in supplementary fig. 2, Supplementary Material online). Clams were typically dissected immediately following collection or frozen whole (−80 °C) until dissection and DNA extraction (below). Together, the 11 sampled clam species represent 12 symbiont lineages, as *Vesicomya* sp. JdF clams from the Juan de Fuca Ridge host two distinct symbiont phylotypes: symA and symB (Stewart et al. 2008; table 1). Here, *Vesicomya* sp. JdF clams correspond to *Vesicomya* sp. mt-II clams identified by host haplotype in Goffredi et al. (2003). Additionally, gene sequences from two other vesicomyid symbionts, Candidatus *Ruthia magnifica* and

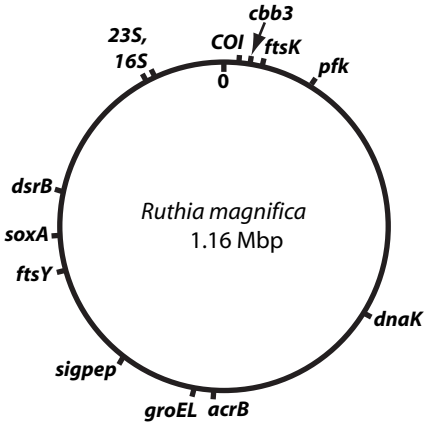

FIG. 1.—Relative positions of symbiont loci on the 1.16-Mb chromosome of Candidatus *Ruthia magnifica*, the symbiont of the deep-sea vesicomyid *Calyptogena magnifica*. Locus descriptions are in table 2.

Candidatus *Vesicomyosocius okutanii* (the symbionts of host clams *Calyptogena magnifica* and *C. okutanii*, respectively), were obtained from the published genomes of these bacteria (GenBank NC008610 and NC009465; Kuwahara et al. 2007; Newton et al. 2007).

Symbiont Loci, DNA Extraction, Polymerase Chain Reaction (PCR), and Sequencing

This study tested for recombination using sequences for portions of 13 single-copy loci of vesicomyid symbionts, including the bacterial *16S* and *23S* rRNA genes and 11 protein-coding genes (fig. 1; table 2). These loci provide broad coverage across most regions of the ~1.1-Mb symbiont chromosome (fig. 1), encoding diverse products involved in processes of translation (*16S*, *23S*), sulfur metabolism (*soxA*, *dsrB*), cell division (*ftsK*, *ftsY*), protein folding (*dnaK, groEL*) and processing (*sigpep*), energy transduction (*COI*, *cbb3*) and glycolysis (*pfk*), and membrane transport (*acrB*). GenBank accession numbers for sequences analyzed here are listed in supplementary table 1, Supplementary Material online. Sequence data for the *C. okutanii* and *C. magnifica* symbionts were taken from published genomes (Kuwahara et al. 2007; Newton et al. 2007). The *16S*, *23S*, *dnaK*, and *soxA* sequences for *Calyptogena ponderosa*, *Calyptogena kilmeri*, *Ectenagena extenta*, *Vesicomya gigas*, *Vesicomya* sp. MAR, *Vesicomya*. sp. mt-I, *Vesicomya* sp. mt-III, symA and symB symbionts (of *V*. sp. JdF clams) were taken from Stewart et al. (2008). All remaining sequences were obtained in this study, as follows:

Total DNA was extracted from the symbiont-containing gill tissue of clam specimens using the DNeasy tissue kit (Qiagen, Valencia, CA). The primer sets listed in supplementary table 2, Supplementary Material online, were used to PCR amplify and sequence portions of the symbiont loci (table 2). Primers for protein-coding loci were developed based on conserved regions in the genomes of *C. okutanii* and *C. magnifica* symbionts. Following visualization via gel electrophoresis, PCR products were purified using the QIAquick PCR purification kit (Qiagen) and direct sequenced in both the forward and reverse directions using standard BigDye techniques.

Base Composition and Genetic Divergence

Forward and reverse sequences were assembled into contigs and edited manually in Sequencher v.4.7. Alignments were generated automatically in ClustalW (Thompson et al. 1994) and then edited manually based on amino acid translation in MacClade 4.0 (Maddison and Maddison 2000). GC content and genetic divergence were calculated in the program DnaSP vs. 4.10.2 (Rozas et al. 2003). Nucleotide diversity (π) was calculated as the average number of nucleotide differences per site between two sequences (Nei 1987). Ratios of nonsynonymous substitutions per nonsynonymous site (K_a) to synonymous substitutions per synonymous site (K_s) were determined using the method of Nei and Gojobori (1986). The assumption of homogeneity of nucleotide frequency among taxa was tested using the method of Rzhetsky and Nei (1995), which accounts for error due to possible phylogenetic correlations among sequences.

Table 2
Vesicomyid Clam Symbiont Loci

Locus[a]	Gene/Region	Function	Length (bp)[b]
COI	Cytochrome *c* oxidase subunit I	Energy transduction	789
cbb3	*cbb3*-type cytochrome *c* oxidase subunit I	Energy transduction	801
ftsK	Cell division protein	Cell division	753
Pfk	Phosphofructokinase	Glycolysis	744
dnaK	Molecular chaperone	Protein folding	735
acrB	Acriflavin resistance protein	Multi-drug efflux	729
groEL	Molecular chaperone	Protein folding	885
Sigpep	Signal peptidase I	Protein processing	549
ftsY	Signal recognition particle docking protein	Signal recognition	582
soxA	Sulfur oxidation protein	Sulfur metabolism	594
dsrB	Sulfite reductase, dissimilatory-type beta subunit	Sulfur metabolism	864
23S	Large subunit rRNA gene	Ribosomal RNA	1,816
16S	Small subunit rRNA gene	Ribosomal RNA	1,253

[a] Locus: loci ordered by chromosome position (fig. 1).
[b] Length: basepairs sequenced for each locus, excluding gaps.

Phylogenetic Analyses

Bayesian inference was used to reconstruct phylogenies for single-gene and concatenated-gene data sets using MrBayes v3.1.2 (Huelsenbeck and Ronquist 2001). Single-gene phylogenies were run both unrooted and rooted. For rooted analyses, phylogenies included a clade of nonsymbiont taxa consisting of the three most-closely related bacterial sequences identified via Blast (using the *V. cordata* symbiont as the reference sequence). Nonsymbiont taxa necessarily varied among loci, as no single taxon that contained all of the vesicomyid symbiont genes analyzed here could be identified in public databases. Consequently, the concatenated-gene phylogeny, based on an alignment of all 13 loci (11,094 total characters) concatenated by their order on the *R. magnifica* genome (fig. 1), was run only as unrooted. Despite repeated attempts, the *COI* and *soxA* genes could not be amplified from the *Vesicomya* sp. GOM symbiont; these data were therefore coded as missing in the concatenated alignment.

Bayesian tree reconstruction was as follows: For single-gene analyses, the best-fit model of nucleotide substitution per locus was chosen based on the Akaike information criterion (AIC) using standard procedures in PAUP (Akaike 1974; Posada and Crandall 2001b; Swofford 1998; see the list of models in supplementary table 3, Supplementary Material online). For the concatenated analysis, the data were partitioned under two substitution models. Protein-coding loci were assigned a general time reversible model (GTR) with site-specific rate partitions for each codon position, whereas noncoding rRNA genes were assigned a GTR model with a proportion of invariable sites (I) and gamma distributed rate variation among sites (Γ). For all analyses, five Markov chain Monte Carlo (MCMC) runs were conducted with Metropolis coupling and 20 parallel chains. Each run was iterated for 5.0×10^6 generations, with parameters sampled every 1,000 generations. Of these 5,000 samples, 1,000 were discarded as burn-in. All repetitions of the analyses converged on very similar parameter estimates, and the five runs were combined for a total of 20,000 posterior samples. MCMC convergence was assessed using the CODA package in R Development Core Team (2007).

The significance of topological differences among loci was tested using methods implemented in the program CONSEL (Shimodaira and Hasegawa 2001). To avoid potential bias introduced by heterogenous base composition among taxa (see below), topological discordance was evaluated using reduced taxon data sets that excluded the symbionts of *V. gigas*, *C. kilmeri*, *V. okutanii*, *E. extenta*, and *V.* sp. GOM (i.e., the *gigas*/*kilmeri* clade, see fig. 2), as these taxa were shown to be enriched in AT composition relative to the other nine symbionts (see fig. 3 and results of the Rzhetsky and Nei test below). Briefly, using the heuristic search option in PAUP, an unrooted maximum likelihood (ML) tree was generated for a given symbiont gene using the same substitution model as specified in the single-gene Bayesian analysis (see supplementary table 3, Supplementary Material online, for best-fit models), tree bisection and reconnection branch-swapping, and 100 random addition replicates. The sitewise log likelihoods for this tree were compared with those generated when the same sequence data were constrained to the ML topologies for each additional symbiont gene. Using sitewise log likelihoods for both constrained and unconstrained topologies, CONSEL was implemented to calculate *P* values according to the Approximately Unbiased (AU) test using multiscale bootstrapping (Shimodaira 2002) and the Shimodaira–Hasegawa (SH) test (Shimodaira and Hasegawa 1999). *P* values correspond to the probability of obtaining a more extreme test statistic (the difference in log likelihoods between constrained and unconstrained topologies) by chance under the null hypothesis of no difference between topologies. These analyses were run reciprocally for each possible pair of symbiont loci. Topologies were judged significantly discordant at P values < 0.05, following sequential scaling for multiple tests using the Bonferroni correction method of Holm (1979).

Recombination Detection Program (RDP) and Genetic Algorithm for Recombination Detection (GARD) Tests for Intragenic Recombination

Five methods were applied to screen for intragenic recombination in single-gene alignments. RDP v.3.15 (Martin et al. 2005) was used to implement four nonparametric methods that detect regions of the aligned sequences delimited by putative recombination breakpoints: MaxChi (Maynard Smith 1992; Posada and Crandall 2001a), Chimaera (Posada and Crandall 2001a), RDP (Martin and Rybicki 2000), and GENECONV (Padidam et al. 1999). These methods, which have proved effective in detecting recombination in bacterial data sets (e.g., Baldo et al. 2006; Coscollá and González-Candelas 2007; Degnan and Moran 2008), are among the most powerful at detecting recombination when present yet produce relatively low rates of false positives (Posada and Crandall 2001a; Posada 2002). Sequences were considered circular, a Bonferroni correction for multiple comparisons was applied, and breakpoint positions were approximated using the polishing function of the RDP program. A range of window sizes (20, 40, and 60 nt) was used, as well as a setting that allows window size to vary during the analysis.

Additionally, a model-based approach, the GARD (Kosakovsky Pond et al. 2006), was used to search for putative breakpoints delimiting sequence regions having distinct phylogenies. Briefly, GARD compares a nonrecombinant model in which the sequence data are fitted to a single phylogeny to models in which breakpoints partition the sequence data into two or more regions having varying phylogenies. Support for recombination is reflected by changes in the goodness of fit between nonrecombinant and recombinant models, as assessed by the AIC. To further substantiate breakpoints detected by GARD, the SH test was applied as described above to verify whether adjacent sequence fragments yield statistically different tree topologies ($P < 0.05$, with sequential scaling for multiple tests using the Bonferroni correction method of Holm 1979). Analyses were implemented on the GARD cluster (http://www.datamonkey.org/GARD/) using the multiple breakpoint detection method. The nucleotide substitution

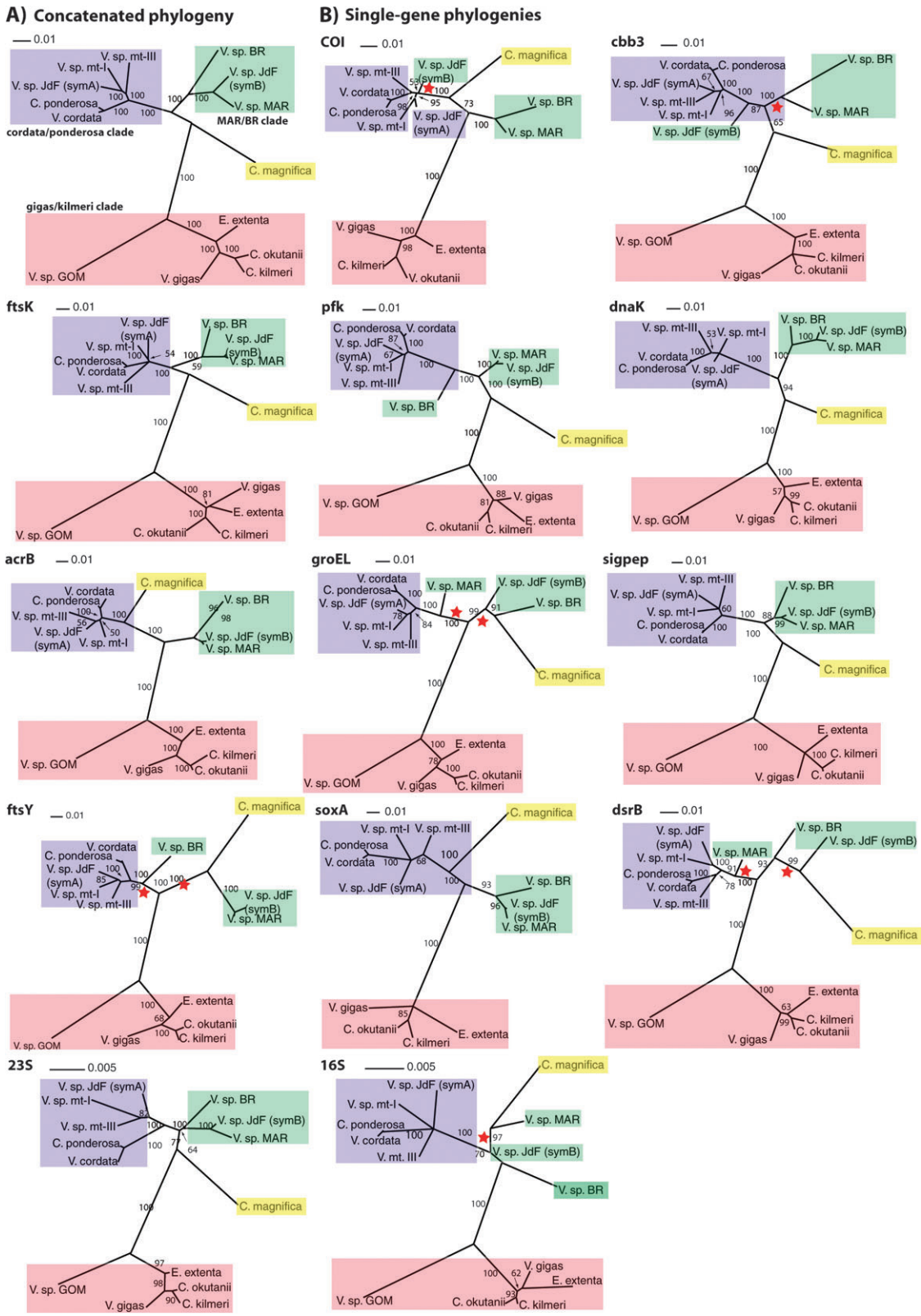

FIG. 2.—Unrooted Bayesian phylogenies for vesicomyid symbiont loci showing among-gene variation suggestive of recombination (see fig. 1 and table 2 for locus descriptions). Symbionts are identified according to host taxon. (*A*) A concatenated phylogeny (upper left) based on all symbiont loci (11,094 bp). Shaded boxes reflect clade designations: *cordata*/*ponderosa* (purple), MAR/BR (green), and *gigas*/*kilmeri* (red). (*B*) Single-gene phylogenies for 13 symbiont loci. Posterior probabilities of taxon bipartitions are displayed if >0.50. Symbiont clades are shaded as in the concatenated phylogeny, with red stars indicating groupings that deviate substantially (PP > 0.95) from the concatenated phylogeny. Substitution rates are ~5× greater in protein-coding loci than in *16S* and *23S* rRNA genes (note differences in scale bars). Trees are ordered based on the position of loci in the *Ruthia magnifica* genome (fig. 1). Here, *Vesicomya* sp. JdF = *V.* sp. mt-II of Goffredi et al. (2003) and Stewart et al. (2008).

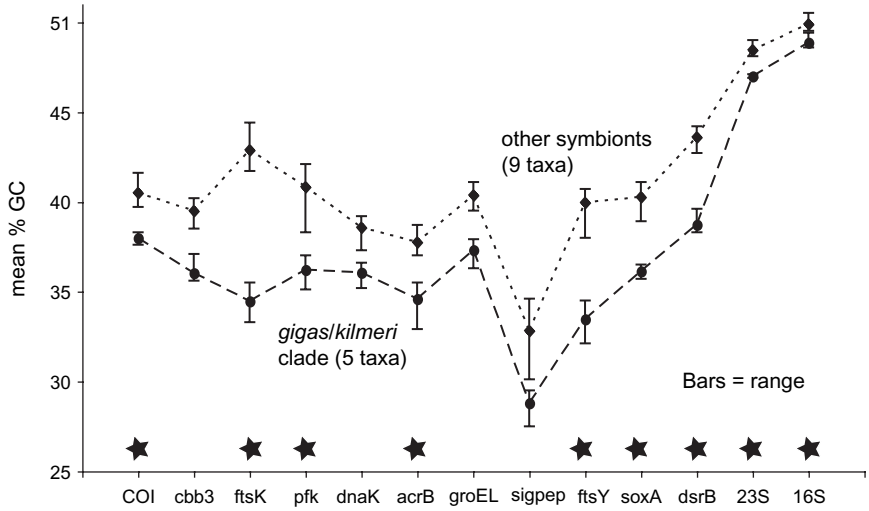

FIG. 3.—Discrepancies in base composition among vesicomyid symbionts. Base composition (% GC content) is significantly lower in the symbionts of the *gigas*/*kilmeri* clade (fig. 2; lower line) relative to the nine symbionts of other vesicomyid clam hosts (upper line). Black stars indicate loci for which significant heterogeneity in base composition ($P < 0.05$) was detected across all 14 taxa via the method of Rzhetsky and Nei (1995). For all loci, excluding *COI*, this method failed to detect heterogeneity when the *gigas*/*kilmeri* clade was removed from the analysis. The *COI* and *soxA* analyses did not include the *Vesicomya* sp GOM symbiont, as these loci could not be amplified from this symbiont. Loci are ordered by their position in the *Ruthia magnifica* genome (fig. 1).

matrix was optimized for each data set via Model Test (Posada and Crandall 2001b), as implemented via the HyPhy package (Kosakovsky Pond et al. 2005), and among-site rate variation was modeled by a discretized gamma distribution with four rate classes.

Results

Vesicomyid Symbiont Clades—Concatenated Loci Phylogeny

Unrooted Bayesian analysis of the concatenated sequences from 13 symbiont loci clustered the vesicomyid symbionts into 2 major clades (fig. 2, upper left): 1) the *gigas*/*kilmeri* clade (red) containing the symbionts of *V. gigas*, *C. kilmeri*, *V. okutanii*, *E. extenta*, and *V.* sp. GOM (posterior probabilities, PP = 1.00), and 2) the clade containing the other nine vesicomyid symbiont lineages (PP = 1.00). The nine taxa outside the *gigas*/*kilmeri* clade are further subdivided into two minor clades: 1) the *cordata*/*ponderosa* clade (purple) containing the symbionts of *V. cordata*, *C. ponderosa*, *V.* sp. mt-I, and III and the symA symbiont of *V.* sp. JdF (PP = 1.00), and 2) the MAR/BR clade (green) containing the symbionts of *Vesicomya* sp. MAR and *Vesicomya* sp. BR and the symB symbiont of *Vesicomya*. sp. JdF, and (PP = 1.00). Placement of the *C. magnifica* symbiont is poorly resolved, with this lineage appearing outside the other three clades.

Base Composition and Genetic Divergence Patterns

Base composition varied across loci, as well as across taxa for a given locus. Percent GC averaged 49% across *23S* and *16S* rRNA genes and 38% across 11 protein-coding loci (fig. 3; supplementary table 4, Supplementary Material online). Across-locus variation was anticipated, as genes are expected to vary in amino acid composition, selective constraint, proportion of neutral sites, and location on the chromosome, all of which influence base composition (Bentley and Parkhill 2004). Surprisingly, however, the null hypothesis of homogenous base composition across taxa was rejected for 9 of the 13 loci ($P < 0.05$; Rzhetsky and Nei 1995 test; fig. 3). Base composition was distinctly bimodal; across all loci, AT content was significantly enriched in symbionts of the *gigas*/*kilmeri* clade (figs. 2 and 3). When these five taxa were excluded from the analysis, homogeneity of base composition was rejected for only one locus (*COI*), implying that the discrepancy between this clade and the nine other symbiont taxa underlies the rejection of base composition homogeneity. These data divide the phylogeny into two primary lineages based on GC content.

Genetic diversity varied between noncoding and coding loci, with mean pairwise nucleotide divergence (π) at 2% for both *16S* and *23S* rRNA genes and ranging from 8% (in *COI*) to 12% (in *ftsY*) among the 11 protein-coding loci (supplementary table 4, Supplementary Material online; fig. 2). The synonymous substitution rate (K_s) for the coding loci ranged from 0.38 (in *soxA*) to 0.52 (in *ftsY*), with the ratio of K_a (nonsynonymous substitution rate) to K_s ranging from 0.04 to 0.16. Though sequence divergence can impact the power of recombination detection methods (e.g., Posada and Crandall 2001a; Tsaousis et al. 2005), the relative homogeneity of divergence values observed here suggests that variable substitution rates among loci are unlikely to bias our ability to detect recombination.

Topological Discordance among Symbiont Gene Trees

Clear topological differences occurred among symbiont gene trees (fig. 2). These differences were consistent across both unrooted and rooted Bayesian analyses, though support values were marginally lower for rooted trees (likely due to ambiguity over root placement). Here, we

Table 3
Pairwise Tests for Topological Congruence between Vesicomyid Symbiont Gene Trees

Locus[a]	*COI*	*cbb3*	*ftsK*	*pfk*	*dnaK*	*acrB*	*groEL*	*sigpep*	*ftsY*	*soxA*	*dsrB*	*23S*	*16S*
COI	–												
cbb3		–											
ftsK	A,S		–										
pfk	A,S			–									
dnaK	A,S				–								
acrB	A,S			A,S		–							
groEL	A,S	A	A	A	A	A,S	–						
sigpep	A,S						A	–					
ftsY	A,S	A,S	A		A	A,S	A,S	A	–				
soxA	A,S						A		A	–			
dsrB	A,S	S	A	A	A,S	A,S		A	A,S	A	–		
23S	A,S						A		A		A,S	–	
16S	A,S		A			A,S	A,S		A,S		S		–

All analyses were run unrooted using a reduced taxon set excluding the AT-enriched *gigas*/*kilmeri* clade to avoid potential bias due to heterogenous base composition (see text); analyses were run in CONSEL (Shimodaira and Hasegawa 2001).

[a] Locus: symbiont loci ordered by chromosome position (fig. 1) A or S = indicates rejection of topological congruence by the Approximately Unbiased (A) or the Shimodaira Hasegawa (S) test ($P < 0.05$, sequentially scaled for multiple tests using the Bonferroni method of Holm 1979).

present only unrooted trees, with rooted results presented as supplementary figure 1, Supplementary Material online. Topological inconsistencies primarily involved the placement of the *C. magnifica* symbiont and the three symbionts of the MAR/BR clade (fig. 2 and supplementary fig. 1, Supplementary Material online). To highlight this variation, the red stars in figure 2 identify nodes at which the placement of one of these four taxa deviated substantially (PP > 0.95) from the concatenated phylogeny (upper left in fig. 2). Aberrant placement of 1 of these 4 taxa occurred in 6 of the 13 loci: *16S*, *cbb3*, *COI, dsrB*, *ftsY*, and *groEL*. In contrast to the concatenated phylogeny, which may not accurately reflect the complex evolutionary history of these taxa, these aberrant loci (excluding *ftsY*) show a breakdown in the tight sister relationship between the symbiont of *V.* sp. MAR and the symB symbiont of *V.* sp. JdF. For example, the *dsrB* and *groEL* analyses moved the *V.* sp. MAR symbiont out of the MAR/BR clade to the base of the *cordata*/*ponderosa* clade (PP = 1.00). Further, the *COI* analysis suggested that the symB symbiont of *V.* sp. JdF is more closely related to *V.* sp. mt-I and III symbionts and to symA, placing symB at the base of the *cordata*/*ponderosa* clade (PP = 1.00). Overall, the *cordata*/*ponderosa* and *gigas*/*kilmeri* clades were much more stable; though minor topological inconsistencies involving members of these clades were observed, these differences were not well supported (PP < 0.95).

Tests for Symbiont Gene Tree Concordance

The topological inconsistencies shown in figure 2 were verified using AU and SH tests (table 3). To avoid any potential bias due to variance in base composition across taxa (see above), these comparisons were performed for a reduced taxon set excluding the AT-enriched *gigas*/*kilmeri* clade. Topological concordance was rejected by either the AU or SH test (or both) in 54% (42 of 78) of gene-by-gene comparisons ($P < 0.05$, sequentially scaled for multiple tests by a Bonferroni correction). Of these, 55% (23 of 42) were rejected by both methods and 98% (41 of 42) involved at least one of the six loci identified as deviating substantially (PP > 0.95 at nodes) from the concatenated topology for this group: *16S*, *cbb3*, *COI, dsrB*, *ftsY*, and *groEL* (red stars in fig. 2). Together, these results suggest distinct evolutionary histories among loci, a pattern suggestive of homologous recombination occurring intergenically between the analyzed genes.

RDP and GARD Tests for Intragenic Recombination

The potential for recombination in vesicomyid symbionts was verified by detecting putative breakpoints within symbiont genes (table 4). RDP methods detected recombination events in two genes, inferring breakpoints at bp 422 in *cbb3* and at bp 348 and 741 in *dsrB* (table 4). Similarly, GARD placed breakpoints at bp 399 in *cbb3* and bp *357* in *dsrB* based on AICc goodness of fit. GARD-estimated breakpoints were further substantiated by positive results of the SH test, which demonstrated significant incongruence between topologies before and after each breakpoint (table 4). Though RDP and GARD vary in their methods for optimizing breakpoint locations, the relative agreement in these methods provides strong support for intragenic recombination events involving *cbb3* and *dsrB*. Additionally, GARD detected breakpoints at bp 801 and 1,070 in the *23S* gene fragment. However, support for these breakpoints came only from ΔAICc; flanking topologies were not significantly discordant via the SH test, suggesting that other processes (e.g., substitution rate heterogeneity in the *23S* gene) may be contributing to phylogenetic variation before and after the breakpoint. No intragenic breakpoints were detected by RDP or GARD when the four taxa showing instability in the single-gene phylogenies (symbionts of the MAR/BR clade and *C. magnifica*) were removed from the analysis, indicating that the recombination signal is limited to these taxa.

Visual Evidence of Recombination

The strongest evidence for recombination in the data set comes from a direct visualization of the sequence alignment (fig. 4*a*). Figure 4 shows all parsimony informative sites ($n = 242$) in the concatenated sequences (11,094 total

Table 4
Recombination Breakpoints Detected within Symbiont Genes

Gene	Length (bp)[a]	RDP Results			GARD Results		
		n[b]	Position[c]	Support[d]	n[b]	Position[c]	Δ cAIC[e]
cbb3	801	1	422	M,C	1	**399**	36.7
dsrB	864	2	348,741	M,C	1	**357**	10.0
23S	1,816	0	ND	ND	2	801, 1,070	16.4

[a] Length: basepairs of analyzed sequence.

[b] *n*: number of breakpoints detected.

[c] Position: approximate breakpoint positions identified via RDP methods or GARD.

[d] Support: methods implemented in RDP v.3.15 that identified statistically significant recombination breakpoints ($P < 0.05$, sequentially scaled for multiple tests using the Bonferroni method of Holm 1979); M = MaxChi, C = Chimaera, R = RDP, and G = GENECONV.

[e] Δ cAIC: improvement in cAIC of the breakpoint-partitioned model over a no-recombination single phylogeny model. Data in bold highlights GARD-identified breakpoints for which flanking trees were significantly discordant via an SH test (P value <0.05, sequentially scaled using the Bonferroni method of Holm 1979).

bp) from four symbiont taxa: the three symbionts of the MAR/BR clade that show variability among gene tree topologies (*Vesicomya* sp. MAR, *V*. sp. BR, and symB symbionts) and symA, whose phylogenetic placement is relatively stable across loci (fig. 2). To highlight differences in phylogenetic patterns across genes (see colors in fig. 4), sites were color coded according to each of the three possible unrooted trees containing these four taxa (fig. 4). Further, RDP methods were applied to the full length concatenated sequences for these four taxa and the positions of any breakpoints detected by at least two RDP methods were mapped onto the alignment. The data indicate broad stretches of sites supporting each of the three distinct phylogenetic groupings, with transitions between stretches occurring at or near statistically supported breakpoints ($P < 0.05$). Although random homoplasy (e.g., via back mutation) would be expected to create occasional phylogenetic discordances among sites (reflected by intermittent changes in color), the spatial clustering of sites that support alternate phylogenies argues strongly against a single evolutionary history for all loci in the genome.

The mosaic sequences in figure 4 highlight a complex pattern of recombination in these four taxa. For example, informative sites downstream of the intragenic breakpoint in the *cbb3* gene fragment supported a clustering of symB and the *Vesicomya* sp. MAR symbiont (blue sites; fig. 4). This pattern was confirmed in a Bayesian phylogenetic analysis of the full-length-sequence data lying between the *cbb3* breakpoint and the next downstream breakpoint in the concatenated alignment (3,451 bp; PP = 1.00; fig. 4*C*), as well as in the majority of the gene trees shown in figure 2 (8 of 13). In contrast, informative sites upstream (left) of the *cbb3* breakpoint supported a grouping of symB with symA (red sites; fig. 4), the symbiont phylotype present in the majority of *Vesicomya* sp. JdF clams analyzed to date (Goffredi et al. 2003; Stewart et al. 2008). This pattern was consistent across both the 5′ part of the *cbb3* gene fragment as well as the *COI* gene fragment. Further, when the full sequence data corresponding to these regions (*COI* and 5′ of *cbb3*; 1,217 bp) were used to reconstruct the symbiont phylogeny, symB was shown to fall into the clade containing symA (PP = 1.00; fig. 4*B*). Finally, the majority of informative sites in both *groEL* and *dsrB* (yellow sites) suggest that symB is more closely related to the *V. sp. BR* symbiont than to either symA or the *V. sp. MAR* symbionts; this pattern is supported by the Bayesian phylogenies in figure 2. The occurrence of three distinct phylogenetic patterns involving these four taxa suggests a complex history of recombination, potentially involving multiple distinct recipient and donor genomes.

Discussion

Obligate endosymbiotic bacteria are confined to an intracellular niche and should therefore face limits to gene exchange with free-living or other symbiotic bacterial lineages (Bordenstein and Reznikoff 2005). Chemosynthetic endosymbionts of deep-sea vesicomyid clams have coevolved in close association with their hosts, live intracellularly, and undergo nearly strict vertical transmission between host generations (Peek et al. 1998; Stewart et al. 2008). Given this lifestyle, vesicomyid symbionts may experience a lack of recombination similar to that of other obligate endosymbionts. Here, this hypothesis was tested using sequence data from 13 symbiont loci.

Topological differences between symbiont gene trees showed that no single phylogeny accurately described the evolutionary history of all analyzed loci. Indeed, almost half of the loci (*16S*, *COI*, *cbb3*, *dsrB*, *ftsY*, and *groEL*) showed phylogenetic groupings with high node support that differed from the concatenated sequence phylogeny for these symbionts (fig. 2; table 3). Such topological variation may be attributed to several factors. First, deviations from assumptions of the evolutionary model may generate systematic errors in tree building (Jeffroy et al. 2006). Here, the evolutionary model was optimized for individual loci, and the six aberrant loci did not show substitution patterns consistently different from those of the other symbiont loci, suggesting that locus-specific differences in the mutation or substitution process (e.g., selection) did not bias the phylogenetic analyses (supplementary table 4, Supplementary Material online). Second, in recently diverged species, the incomplete sorting of ancestral polymorphism can cause topological incongruence, with distinct genes yet to converge on the same underlying "species" phylogeny (Castillo-Ramirez and Gonzalez 2008; Galtier and Daubin 2008). For incomplete lineage sorting to unlink phylogenies, the evolutionary histories of the loci under comparison must, at some point, have been decoupled (e.g., via recombination, or the duplication and subsequent extinction of gene copies). Incomplete lineage sorting may be contributing to topological incongruence in our data set, but, if so, suggests that symbiont genes have not evolved in parallel during their diversification. Third, incorrectly designating paralogs as orthologs can lead to disagreement among gene trees (Stewart and Cavanaugh 2007; Soria-Carrasco and Castresana 2008). Here, only genes occurring as single copies in the two sequenced vesicomyid symbiont genomes were analyzed (Kuwahara et al.

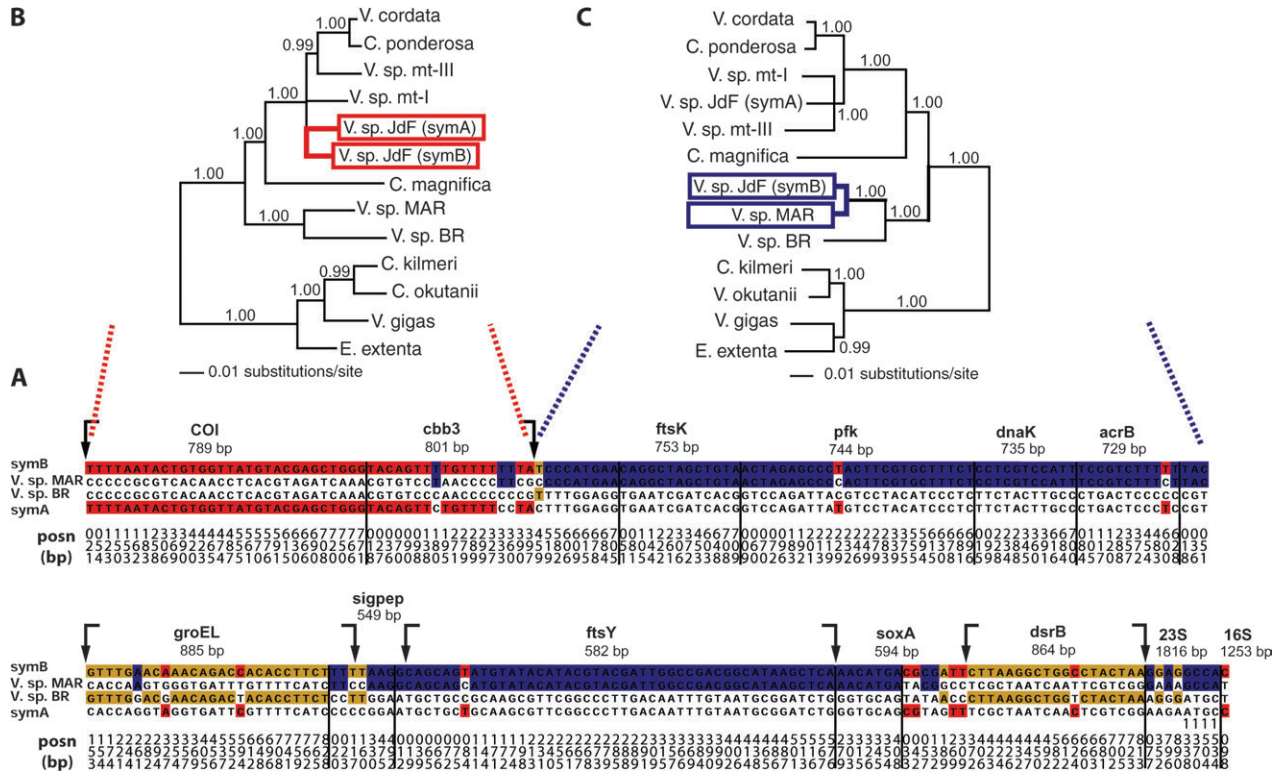

FIG. 4.—Patterns of homologous recombination in chemosynthetic vesicomyid symbionts. (*A*) Parsimony informative sites (242) from a concatenated alignment of 13 loci from 4 symbionts: symA and symB of *Vesicomya* sp. JdF and symbionts of *V*. sp. MAR and *V*. sp. BR. Phylogenetic variation among sites is highlighted by transitions in color patterns, with sites colored according to each of three possible four-taxon phylogenetic groupings. Loci are ordered by their position in the *Ruthia magnifica* genome (fig. 1), with the positions of each site listed below the alignment (relative to the start of each gene fragment). Arrows mark breakpoints delimiting potential recombinant regions identified by at least two methods in RDP analysis of the full length concatenated sequences for these four taxa (11,094 bp; $P < 0.05$). Breakpoint positions in *cbb3* and *dsrB* also correspond to the intragenic breakpoints identified in RDP analysis of the single-gene data sets (table 4). (*B*) Bayesian phylogeny of all symbiont taxa based on the first RDP-identified recombinant region (1,217 bp) containing the *COI* gene fragment (789 bp) and the 5′ region of the *cbb3* gene fragment (428 bp). (*C*) Bayesian phylogeny based on 3,451 bp from six symbiont gene fragments (3′ of *cbb3, ftsK, pfk, dnaK, acrB*, 5′ of *groEL*) between the RDP-identified breakpoints in *cbb3* and *groEL*. Colored boxes in (*B*) and (*C*) reflect phylogenetic groupings indicated by the alignment of parsimony informative sites (below). PP>0.50 are shown.
Note: the *V. sp*. GOM symbiont was excluded from (*B*) and (*C*), as the COI gene could not be amplified from this taxon.

2007; Newton et al. 2007). However, without additional whole genome sequencing, we can neither rule out the possibility that hidden paralogy is obscuring symbiont gene trees, nor definitively link all sequenced loci to the same symbiont chromosome. Finally, variation among gene trees may be caused by homologous recombination that reshuffles allelic variants among related lineages, effectively creating a mosaic genome in which genomic regions have varying evolutionary histories (Maynard Smith 1992; Doolittle and Papke 2006). Here, the phylogenetic data are consistent with a pattern of recombination occurring intergenically in chromosome regions between loci, decoupling the phylogenies of the genes being analyzed. The six loci with aberrant nodes encode proteins of diverse function and are distributed across the symbiont chromosome (fig. 1), suggesting that recombination has not been restricted to specific genomic regions.

The strongest evidence for recombination in vesicomyid symbionts comes from the detection of breakpoints directly within symbiont genes and from direct visualization of the sequence data. Our analyses, which examined ~1% of the estimated 1.1–1.2 Mb vesicomyid symbiont chromosome (Kuwahara et al. 2007; Newton et al. 2007), detected well-supported recombination breakpoints in two loci, *cbb3* and *dsrB* (tables 4 and 5). The *cbb3* breakpoint can be clearly visualized in the four-taxon alignment of figure 4, in which informative sites before and after the breakpoint clearly support alternate phylogenetic groupings. A clear transition between phylogenetic groupings is less clear for the *dsrB* gene. However, our analyses cannot effectively distinguish between one or multiple recombination events acting at these sites. Indeed, multiple recombination events, potentially involving several distinct taxa, might be expected to produce conflicting patterns of polymorphism similar to those observed in *dsrB* (fig. 4). Given the frequency at which intragenic recombination was detected in this study (ca. two breakpoints per 1% of symbiont genome analyzed, further analyses involving whole genome comparisons will likely reveal many other instances of genetic exchange, helping to determine whether the recombination events detected here are localized to specific hotspots (Smith 1994), or whether the recombination rate is uniform across the chromosome.

Variation in Recombination Patterns across Symbiont Lineages

The extent of recombination in vesicomyid symbionts likely varies across lineages. Vesicomyid clams occur ubiquitously throughout the world's oceans (Peek et al. 1997, 1998, 2000; Kojima et al. 2004; Krylova and Sahling 2006), but only a small subset of the estimated 50–100 species have been characterized molecularly with regard to symbiont phylotype. This study, examining 14 symbiont lineages, found the clearest signal of recombination in members of the MAR/BR clade (*V*. sp. MAR, *V*. sp. BR, symB symbionts); when these taxa and the *C*. *magnifica* symbiont were removed from the data set, RDP and GARD analyses failed to detect intragenic recombination. The mosaic sequences in figure 4, which suggest a pattern of symA–symB recombination, hint that analyses of additional loci may reveal signatures of recombination within the *cordata*/*ponderosa* clade (e.g., involving symA). However, the topological stability of this clade across the loci in figure 2 argues against the recent acquisition of foreign DNA by *cordata*/*ponderosa* symbionts from bacteria outside this clade. Similarly, the members of the *gigas*/*kilmeri* clade did not show strong evidence of recombination and varied little in taxonomic placement among loci, a pattern suggestive of the genomic stasis shown for other maternally inherited endosymbionts (e.g., *Buchnera*; Tamas et al. 2002; Silva et al. 2003).

Curiously, symbionts of the *gigas*/*kilmeri* clade were significantly elevated in AT composition relative to the other symbiont lineages (fig. 3). Genomic AT-enrichment is an underlying feature of obligately intracellular bacteria. In vertically transmitted insect endosymbionts with low effective population sizes (N_e), AT-enrichment may stem in part from an underlying GC to AT mutational bias, an enhanced rate of fixation via genetic drift, and the inability of symbionts to effectively purge mutations from the population via recombination (i.e., Muller's ratchet; Moran 1996; Wernegreen 2004, 2005; Wernegreen and Funk 2004; Klasson and Andersson 2007). Thus, reduced recombination rates may underlie the AT-enrichment of *gigas*/*kilmeri* symbionts; in the absence of recombination, these lineages would have few opportunities to remove mildly deleterious mutations from the population, leading to an enhanced rate of AT-enrichment (assuming an underlying GC to AT mutational bias). However, this hypothesis requires rigorous testing, as several other mechanisms could explain elevated AT-bias in this clade, including clade-specific variation in effective population size (N_e), selective codon usage, or DNA repair capacity (Clark et al. 1999; Moran 2002; Sharp et al. 2005). Regardless of the mechanistic basis of AT-enrichment, the recombination and bimodal base composition patterns observed here imply nonuniform evolutionary dynamics across the vesicomyid symbiont phylogeny.

Lateral Symbiont Transmission and Recombination

The recombination pattern in the symB symbiont of *Vesicomya* sp. JdF clams raises the hypothesis that lateral symbiont transmission creates opportunities for symbiont recombination. SymB symbionts occur in *Vesicomya* sp. JdF clams at vents on the Juan de Fuca Ridge (northeastern Pacific; Stewart et al. 2008). Analysis of four symbiont loci recently showed that symB did not cospeciate with its host and is divergent from symA, the symbiont phylotype found in the majority (>90%) of the *V*. sp. JdF clams examined (Peek et al. 1998; Goffredi et al. 2003; Stewart et al. 2008). Rather, symB clustered tightly with the symbiont of a distantly related host clam, *Vesicomya* sp. from the Mid-Atlantic Ridge (MAR; Stewart et al. 2008; see map in supplementary fig. 2, Supplementary Material online). That study indicated that despite a dominant strategy of vertical symbiont transmission in these clams, symB entered the *V*. sp. JdF host lineage laterally (i.e., nonvertically), potentially via horizontal transfer from a clam within the clade containing *V*. sp. MAR and related species from the North Pacific (Stewart et al. 2008). Though the physical mechanism of transfer is unknown, it is possible that lateral acquisition brought symB into direct contact with symA, potentially allowing for recombination between these lineages.

The mosaic nature of the symB genome provides some support for this hypothesis. Although a tight coupling of symB and the *V*. sp. MAR symbiont was shown at 8 of the 13 loci examined here, the evolutionary history of symB deviated from this pattern at other loci. Notably, analysis of the two gene fragments upstream of the intragenic breakpoint in *cbb3* (*COI* and 5′ region of *cbb3*) placed symB in a clade with symA, to the exclusion of the *V*. sp. MAR symbiont (figs. 2 and 4*b*). SymA and symB are not identical in the *COI* and 5′ *cbb3* regions (2.0% nucleotide divergence), suggesting that the recombination event(s) that produced this pattern did not occur between the extant symA and symB phylotypes. Rather, the data are consistent with the hypothesis that a symB ancestor recombined with a symA ancestor at some point in the history of this symbiosis, creating a mosaic genome in the extant symB lineage. However, the mosaic sequences observed here cannot be explained solely by symA-symB recombination. Indeed, at both the *dsrB* and *groEL* loci, symB is more closely related to the *V*. *sp*. *BR* symbiont, than to either the *cordata*/*ponderosa* clade or to the *V*. *sp*. *MAR* symbiont (figs. 2 and 4). These patterns suggest a complex history of genomic exchange, likely involving multiple different recipient and donor taxa (in addition to symA or symB). An unambiguous reconstruction of this history is unfortunately not possible based on the current data and is potentially complicated by ancient recombination that occurred prior to the divergence of the extant clades. However, for recombination to have occurred among these bacteria, divergent lineages (or fragments of their genomes) must first have been brought into contact. Such contact could potentially have occurred between free-living forms of symbiont ancestors, but presumably could also have been facilitated by the lateral acquisition and maintenance of divergent symbionts within the same host.

Homologous recombination has been shown in other endosymbionts that exhibit lateral movement between hosts. Notably, the insect endosymbiont *Wolbachia*, though predominantly vertically transmitted, undergoes occasional lateral transmission and exhibits extensive evidence of recombination (Jiggins et al. 2001; Jiggins 2002; Reuter and

Keller 2003; Baldo et al. 2006; Verne et al. 2007). In contrast, insect endosymbionts that show strict vertical transmission typically exhibit little or no recombination (Bordenstein and Reznikoff 2005; Wernegreen 2004, 2005). These include *Buchnera* spp. (of aphids), *Wigglesworthia* spp. (of tsetse flies), and *Blochmannia* spp. (of ants), all of which live intracellularly in obligate mutualisms with their hosts and show remarkably high levels of genomic stasis (Akman et al. 2002; Tamas et al. 2002; Gil et al. 2003; van Ham et al. 2003). Our data suggest that genomic evolution in some vesicomyid symbionts (e.g., members of the MAR/BR clade) more closely resembles that of *Wolbachia*, with homologous recombination contributing to symbiont diversification in lineages brought into contact by lateral transmission. In contrast, other symbiont clades (e.g., the *gigas kilmeri* clade) may undergo a pattern of strict vertical transmission without recombination (Hurtado et al. 2003).

Presumably, the likelihood of recombination in these lineages is highest when divergent symbionts become sequestered for a period of time within the same host individual, or even within the same host cell (i.e., heteroplasmy). For instance, Reuter and Keller (2003) found high levels of symbiont recombination in an ant population with high levels of multiple infection by different *Wolbachia* strains. Further, focusing on mussels (Mytilidae) exhibiting doubly uniparental inheritance (Zouros et al. 1994), Ladoukakis and Zouros (2001) provided strong evidence of homologous recombination between maternal and paternal mitochondrial lineages that occur heteroplasmically in the host cytoplasm. Though prior studies suggest that vesicomyid symbionts are clonal within a host (e.g., Kim et al. 1995), preliminary analyses using deep-coverage amplicon pyrosequencing suggest that individual *V.* sp. JdF clams hosting both symA and symB strains do occur at low frequency (<0.08) in the population (Stewart and Cavanaugh, forthcoming), suggesting a mechanism by which these strains may have recombined. However, mixed symbiont populations were not observed in the individuals analyzed for the current study (Stewart FJ, unpublished data). Quantifying within-clam symbiont diversity across diverse vesicomyid taxa should be a priority, as the extent to which multiple symbiont phylotypes co-occur within a single host may reflect both the rate of lateral symbiont transmission as well as the probability of symbiont recombination.

The molecular mechanism by which vesicomyid symbionts undergo homologous recombination is not clear. Recent studies revealed that the genomes of two vesicomyid symbionts lack enzymes critical to recombination and DNA repair in other bacteria (e.g., *recA*; Rocha et al. 2005; Kuwahara et al. 2007; Newton et al. 2007). However, recombination potential cannot be inferred solely based on gene content (Rocha et al. 2005). Indeed, despite having lost all but a minimal set of energy metabolism genes, the genomes of diverse animal mitochondria exhibit evidence of homologous recombination (Tsaousis et al. 2005). In such instances, recombination is presumably mediated by nuclear-encoded enzymes imported into the mitochondrion (Thyagarajan et al. 1997). As chemosynthetic endosymbionts may have experienced evolutionary forces similar to those affecting mitochondrial evolution, it is not implausible that certain cellular functions, perhaps including recombination, may be mediated by host-encoded proteins or symbiont proteins transferred to the host genome. Given evidence for symbiont-to-host gene exchange in other endosymbioses (e.g., *Wolbachia*; Hotopp et al. 2007), the hypothesis of host-mediated symbiont recombination merits testing.

Conclusions

This study provided strong evidence for homologous recombination in tightly host-integrated chemosynthetic endosymbionts. Also, by highlighting the recombination pattern of the symB symbiont lineage, this study, interpreted with the results of Stewart et al. (2008), suggested lateral symbiont transmission as a potential mechanism for facilitating recombination in deep-sea symbionts; additional genome-level sequencing of multiple symbiont lineages may help verify this hypothesis. Our results have implications for symbiont phylogenetics, as recombinant mosaic genomes (e.g., of symB) preclude accurate reconstruction of the evolutionary history of a symbiont lineage, or the host–symbiont association, based on a single symbiont locus. More broadly, these results indicate that vesicomyid symbionts are not devoid of opportunities for genetic exchange with other bacteria. This result complements a growing body of literature challenging the assumption of genomic stasis in obligate intracellular bacteria (Tamas et al. 2002; Bordenstein and Reznikoff 2005; Sirand-Pugnet et al. 2007). By introducing new allelic variants, reshuffling existing variation, and purging mutations from populations, recombination may play an important, yet understudied, role in the diversification and adaptation of endosymbionts. Further, recombination may attenuate the negative effects of obligate intracellularity (e.g., Muller's ratchet) on the diversification of clam symbionts, rendering the evolution of symbiont genomes far more dynamic than previously thought. Additional genomic analyses at multiple taxonomic scales (e.g., across vesicomyid host species, among individuals within a host species) will help clarify both the rate of recombination and the overall impact of this process on the diversification of marine chemosynthetic endosymbionts.

This and other studies of chemosynthetic symbiont genome evolution provide an important counterpoint to better-studied symbioses involving organelles, symbiotic bacteria of insects, and bacterial pathogens. Described in more than 200 host species worldwide, chemosynthetic symbionts vary widely in their level of specificity and integration with their host, ranging from surface-associated epibionts to endosymbionts that live obligately within host cells, are intrinsic to host metabolism, and pass vertically between host generations (Stewart et al. 2005). The latter are analogous to the eukaryotic chloroplast, potentially representing intermediates in the evolution toward a chemosynthetic "sulfoplast." Examining symbiont diversity and genome structure across this spectrum of diverse associations will help elucidate the phases through which bacteria in ancient associations with eukaryotes may have passed as they evolved into organelles.

Supplementary Material

Supplementary figures 1 and 2 and supplementary tables 1–4 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank David Posada, Scott Edwards, Peter Girguis, James McCarthy, and two anonymous reviewers for their generous and constructive input during the preparation of this manuscript. We also thank Erik Cordes, Chuck Fisher, Peter Girguis, Zoe McCuddin, Katie Scott, and Jonathan Robinson for collecting vesicomyid samples. Finally, we thank the generous crews of the research vessels and submersibles that made this work possible. This work was supported by National Science Foundation grants EF-0412205 and OCE-0453901 awarded to C.M.C. and by the Genetics and Genomics Training Program (GGT) at Harvard University.

Literature Cited

Akaike H. 1974. A new look at the statistical model identification. IEEE Trans Autom Contr. 19:716–723.

Akman L, Yamashita A, Watanabe H, Oshima K, Shiba T, Hattori M, Aksoy S. 2002. Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. Nat Genet. 32:402–407.

Baldo L, Bordenstein S, Wernegreen JJ, Werren JH. 2006. Widespread recombination throughout *Wolbachia* genomes. Mol Biol Evol. 23:437–449.

Bentley SD, Parkhill J. 2004. Comparative genomic structure of prokaryotes. Annu Rev Genet. 38:771–792.

Bordenstein SR, Reznikoff WS. 2005. Mobile DNA in obligate intracellular bacteria. Nature Rev Microbiol. 3:688–699.

Cary SC, Giovannoni SJ. 1993. Transovarial inheritance of endosymbiotic bacteria in clams inhabiting deep-sea hydrothermal vents and cold seeps. Proc Natl Acad Sci USA. 90: 5695–5699.

Castillo-Ramirez S, Gonzalez V. 2008. Factors affecting the concordance between orthologous gene trees and species tree in bacteria. BMC Evol Biol. 8:300.

Cavanaugh CM, McKiness ZP, Newton ILG, Stewart FJ. 2006. Marine chemosynthetic symbioses. In: Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E, editors. The Prokaryotes, 3rd ed. A handbook on the biology of bacteria: symbiotic associations, biotechnology, applied microbiology. Springer, New York.

Chen X, Li S, Aksoy S. 1999. Concordant evolution of a symbiont with its host insect species: molecular phylogeny of genus *Glossina* and its bacteriome-associated endosymbiont, *Wigglesworthia glossinidia*. J Mol Evol. 48:49–58.

Clark MA, Moran NA, Baumann P. 1999. Sequence evolution in bacterial endosymbionts having extreme base compositions. Mol Biol Evol. 16:1586–1598.

Clark MA, Moran NA, Baumann PB, Wernegreen JJ. 2000. Cospeciation between bacterial endosymbionts (*Buchnera*) and a recent radiation of aphids (*Uroleucon*) and pitfalls of testing for phylogenetic congruence. Evolution. 54: 517–525.

Coscollá M, González-Candelas F. 2007. Population structure and recombination in environmental isolates of *Legionella pneumophila*. Environ. Microbiol. 9:643–656.

Degnan PH, Lazarus AB, Wernegreen JJ. 2005. Genome sequence of *Blochmannia pennsylvanicus* indicates parallel evolutionary trends among bacterial mutualists of insects. Genome Res. 15:1023–1033.

Degnan PH, Moran NA. 2008. Evolutionary genetics of a defensive facultative symbiont of insects: exchange of toxin encoding bacteriophage. Mol Ecol. 17:916–929.

Doolittle WF, Papke RT. 2006. Genomics and the bacterial species problem. Genome Biol. 7:116.

Endow K, Ohta S. 1990. Occurrence of bacteria in the primary oocytes of vesicomyid clam C*alyptogena soyoae*. Mar Ecol Prog Ser. 64:309–311.

Galtier N, Daubin V. 2008. Dealing with incongruence in phylogenomic analyses. Phil Trans R Soc B. 363: 4023–4029.

Gil R, Silva FJ, Zientz E, et al. (13 co-authors). 2003. The genome sequence of *Blochmannia floridanus*: comparative analysis of reduced genomes. Proc Natl Acad Sci USA. 100:9388–9393.

Goffredi SK, Hurtado LA, Hallam S, Vrijenhoek RC. 2003. Evolutionary relationships of deep-sea vent and seep clams (Mollusca: Vesicomyidae) of the '*pacifica/lepta*' species complex. Mar Biol. 142:311–320.

Haine ER, Pickup NJ, Cook JM. 2005. Horizontal transmission of *Wolbachia* in a *Drosophila* community. Ecol Entomol. 30:464–472.

Holm S. 1979. A simple sequential rejective multiple test procedure. Scand. J Statistics. 6:65–70.

Hotopp JCD, Clark ME, Oliveira DCSG, et al. (20 co-authors). 2007. Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. Science. 317: 1753–1756.

Huelsenbeck JP, Ronquist F. 2001. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics. 17:754–755.

Hurtado LA, Mateos M, Lutz RA, Vrijenhoek RC. 2003. Coupling of bacterial endosymbiont and host mitochondrial genomes in the hydrothermal vent clam C*alyptogena magnifica*. Appl Environ Microbiol. 69:2058–2064.

Jeffroy O, Brinkmann H, Delsuc F, Philippe H. 2006. Phylogenomics: the beginning of incongruence? Trends Genet. 22:225–231.

Jiggins FM. 2002. The rate of recombination in *Wolbachia* bacteria. Mol Biol Evol. 19:1640–1643.

Jiggins FM, von der Schulenberg HG, Hurst GDD, Majerus MEN. 2001. Recombination confounds interpretations of *Wolbachia* evolution. Proc R Soc Lond B. 268:1423–1427.

Kim YW, Yasuda M, Yamagishi A, Oshima T, Ohta S. 1995. Characterization of the endosymbiont of a deep-sea bivalve, *Calyptogena soyae*. Appl Environ Microbiol. 61:823–827.

Klasson L, Andersson SGE. 2007. Strong asymmetric mutation bias in endosymbiont genomes coincide with loss of genes for replication restart pathways. Mol Biol Evol. 23:1037–1039.

Kojima S, Fujikura K, Okutani T. 2004. Multiple trans-Pacific migrations of deep-sea vent/seep-endemic bivalves in the family Vesicomyidae. Mol Phylogenet Evol. 32:396–406.

Kosakovsky Pond SLK, Frost SDW, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics. 21:676–679.

Kosakovsky Pond SLK, Posada D, Gravenor MB, Woelk CH, Frost SDW. 2006. Automated phylogenetic detection of recombination using a genetic algorithm. Mol Biol Evol. 23: 1891–1901.

Krylova EM, Sahling H. 2006. Recent bivalve molluscs of the genus *Calyptogena* (Vesicomyidae). J Mollus Stud. 72: 359–395.

Kuwahara H, Yoshida T, Takaki Y, et al. (10 co-authors). 2007. Reduced genome of the thioautotrophic intracellular symbiont

in a deep-sea clam, *Calyptogena okutanii*. Curr Biol. 17: 881–886.
Ladoukakis ED, Zouros E. 2001. Direct evidence for homologous recombination in Mussel (*Mytilus galloprovincialis*) mitochondrial DNA. Mol Biol Evol. 18:1168–1175.
Maddison DR, Maddison WP. 2000. MacClade 4: analysis of phylogeny and character evolution. Sunderland (MA): Sinauer Associates.
Martin D, Rybicki E. 2000. RDP: detection of recombination amongst aligned sequences. Bioinformatics. 16:562–563.
Martin D, Williamson C, Posada A. 2005. RDP2: recombination detection and analysis from sequence alignments. Bioinformatics. 21:260–262.
Maynard Smith J. 1992. Analyzing the mosaic structure of genes. J Mol Evol. 34:126–129.
Moran NA. 1996. Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. Proc Natl Acad Sci USA. 93:2873–2878.
Moran NA. 2002. Microbial minimalism: genome reduction in bacterial pathogens. Cell. 108:583–586.
Moran NA, Mira A. 2001. The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola*. Genome Biol. 2:research0054.1–0054.12
Muller HJ. 1964. The relation of recombination to mutational advance. Mutat Res. 106:2–9.
Narra HP, Ochman H. 2006. Of what use is sex to bacteria? Curr Biol. 16:R705–R710.
Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.
Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol. 3:418–426.
Newton ILG, Woyke T, Auchtung TA, et al. (16 co-authors). 2007. The *Calyptogena magnifica* chemoautotrophic symbiont genome. Science. 315:998–1000.
Padidam M, Sawyer S, Fauquet CM. 1999. Possible emergence of new geminiviruses by frequent recombination. Virology. 265:218–225.
Papke RT, Ward DM. 2004. The importance of physical isolation to microbial diversification. FEMS Microbiol Ecol. 48: 293–303.
Peek AS, Feldman RA, Lutz RA, Vrijenhoek RC. 1998. Cospeciation of chemoautotrophic bacteria and deep-sea clams. Proc Natl Acad Sci USA. 95:9962–9966.
Peek AS, Gustafson RG, Lutz RA, Vrijenhoek . 1997. Evolutionary relationships of deep-sea hydrothermal vent and cold-water seep clams (Bivalvia: Vesicomyidae): results from the mitochondrial cytochrome oxidase subunit I. Mar Biol. 130:151–161.
Peek AS, Gaut BS, Feldman RA, Barry JP, Kochevar RE, Lutz RA, Vrijenhoek RC. 2000. Neutral and nonneutral mitochondrial genetic variation in deep-sea clams from the family Vesicomyidae. J Mol Evol. 50:141–153.
Pérez-Brocal V, Gil R, Ramos S, Lamelas A, Postigo M, Michelena JM, Silva FJ, Moya A, Latorre A. 2006. A small microbial genome: the end of a long symbiotic relationship? Science. 314:312–313.
Posada D. 2002. Evaluation of methods for detecting recombination from DNA sequences: empirical data. Mol Biol Evol. 19:708–717.
Posada D, Crandall KA. 2001a. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proc Natl Acad Sci USA. 98:13757–13762.
Posada D, Crandall KA. 2001b. Selecting the best fit model of nucleotide substitution. Syst Biol. 50:580–601.
R Development Core Team. 2007. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing.
Reuter M, Keller L. 2003. High levels of multiple *Wolbachia* infection and recombination in the ant *Formica exsecta*. Mol Biol Evol. 20:748–753.
Rocha EPC, Cornet E, Michel B. 2005. Comparative and evolutionary analysis of the bacterial homologous recombination systems. PLoS Genet. 1:247–259.
Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics. 19:2496–2497.
Rzhetsky A, Nei M. 1995. Tests of applicability of several substitution models for DNA sequence data. Mol Biol Evol. 12:131–151.
Sharp PM, Bailes E, Grocock RJ, Peden JF, Sockett RE. 2005. Variation in the strength of selected codon usage bias among bacteria. Nucleic Acids Res. 33:1141–1153.
Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. Syst Biol. 51:492–508.
Shimodaira H, Hasegawa M. 1999. Multiple comparisons of log-likelihoods and applications to phylogenetic inference. Mol Biol Evol. 16:1114–1116.
Shimodaira H, Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics. 17:1246–1247.
Silva FJ, Latorre A, Moya A. 2003. Why are the genomes of endosymbiotic bacteria so stable? Trends Genet. 19:176–180.
Sintupachee S, Milne JR, Poonchaisri S, Baimai V, Kittayapong P. 2006. Closely related *Wolbachia* strains within the pumpkin arthropod community and the potential for horizontal transmission via the plant. Microb Ecol. 51:294–301.
Sirand-Pugnet P, Lartigue C, Marenda M, et al. (13 co-authors). 2007. Being pathogenic, plastic, and sexual while living with a nearly minimal bacterial genome. PLoS Genet. 3:744–758.
Smith GR. 1994. Hotspots of homologous recombination. Experientia. 50:234–1124.
Soria-Carrasco V, Castresana J. 2008. Estimation of phylogenetic inconsistencies in the three domains of life. Mol Biol Evol. 25:2319–2329.
Stewart FJ, Cavanaugh CM. 2007. Intragenomic variation and evolution of the internal transcribed spacer of the rRNA operon in bacteria. J Mol Evol. 65:44–67.
Stewart FJ, Cavanaugh CM. Forthcoming. Pyrosequencing analysis of endosymbiont population structure: co-occurrence of divergent symbiont lineages in a single vesicomyid host clam. Environ Microbiol. MS: EMI 1933.
Stewart FJ, Newton ILG, Cavanaugh CM. 2005. Chemosynthetic endosymbioses: adaptations to oxic-anoxic interfaces. Trends Microbiol. 13:439–448.
Stewart FJ, Young CR, Cavanaugh CM. 2008. Lateral symbiont acquisition in a maternally transmitted chemosynthetic clam endosymbiosis. Mol Biol Evol. 25:673–687.
Swofford DL. 1998. Paup*. Phylogenetic analysis using parsimony (*and other methods). Sunderland (MA): Sinauer.
Tamas I, Klasson L, Canback B, Naslund AK, Eriksson AS, Wernegreen JJ, Sandstrom JP, Moran NA, Andersson SGE. 2002. 50 million years of genomic stasis in endosymbiotic bacteria. Science. 296:2376–2379.
Thao ML, Baumann P. 2004. Evolutionary relationships of primary prokaryotic endosymbionts of whiteflies and their hosts. Appl Environ Microbiol. 70:3401–3406.
Thompson JD, Higgins DG, Gibson TJ. 1994. ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucl Acids Res. 22: 4673–4680.
Thyagarajan B, Padua RA, Campbell C. 1997. Mammalian mitochondria possess homologous DNA recombination activity. J Biol Chem. 271:27536–27543.

Tsaousis AD, Martin DP, Ladoukakis ED, Posada D, Zouros E. 2005. Widespread recombination in published animal mtDNA sequences. Mol Biol Evol. 22:925–933.

van Ham RCHJ, Kamerbeek J, Palacios C, et al. (16 co-authors). 2003. Reductive genome evolution in *Buchnera aphidicola*. Proc Natl Acad Sci USA. 100:581–586.

Verne S, Johnson M, Bouchon D, Grandjean F. 2007. Evidence for recombination between feminizing *Wolbachia* in the isopod genus *Armadillidium*. Gene. 397:58–66.

Wernegreen JJ. 2004. Endosymbiosis: lessons in conflict resolution. PLoS Biol. 2:307–311.

Wernegreen JJ. 2005. For better or for worse: genomic consequences of intracellular mutualism and parasitism. Curr Opin Genet Dev. 15:572–583.

Wernegreen JJ, Funk DJ. 2004. Mutation exposed: a neutral explanation for extreme base composition of an endosymbiont genome. J Mol Evol. 59:849–858.

Woyke T, Teeling H, Ivanova NN, et al. (18 co-authors). 2006. Symbiosis insights through metagenomic analysis of a microbial consortium. Nature. 443:950–955.

Zouros E, Ball AO, Saavedra C, Freeman KR. 1994. An unusual type of mitochondrial DNA inheritance in the blue mussel *Mytilus*. Proc Natl Acad Sci USA. 91:7463–7467.

Barbara Holland, Associate Editor

Accepted March 9, 2009