

Supplementary Online Material

Evidence for homologous recombination in
intracellular chemosynthetic clam symbionts

Frank J. Stewart¹, Curtis R. Young², and Colleen M. Cavanaugh²

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

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

Supplementary table 1. GenBank accession numbers for vesicomylid symbiont sequences

Taxon ^a	Symbiont												
	16S	23S	<i>acrB</i>	<i>ccb3</i>	<i>COI</i>	<i>dnak</i>	<i>dsrB</i>	<i>ftsK</i>	<i>ftsY</i>	<i>groEL</i>	<i>pfk</i>	<i>sigpep</i>	<i>soxA</i>
<i>Calyptogena kilmeri</i>	EU403431	EU403440	FJ899937	FJ899960	FJ899949	EU403449	FJ899972	FJ899984	FJ899996	FJ900008	FJ900020	FJ900032	EU403458
<i>Calyptogena magnifica</i> ^b	NC_008610	NC_008610	YP_903767	YP_903313	YP_903309	YP_903600	YP_904056	YP_903317	YP_903977	YP_903785	YP_903364	YP_903841	YP_903997
<i>Calyptogena okutanii</i> ^c	NC_009465	NC_009465	YP_001219342	YP_001218903	YP_001218898	YP_001219195	YP_001219624	YP_001218909	YP_001219547	YP_001219359	YP_001218958	YP_001219414	YP_001219566
<i>Calyptogena ponderosa</i>	EU403436	EU403445	FJ899944	FJ899967	FJ899955	EU403454	FJ899979	FJ899991	FJ900003	FJ900015	FJ900027	FJ900039	EU403463
<i>Ectenagena extenta</i>	EU403433	EU403442	FJ899939	FJ899962	FJ899951	EU403451	FJ899974	FJ899986	FJ899998	FJ900010	FJ900022	FJ900034	EU403460
<i>Vesicomya cordata</i>	FJ899928	FJ899931	FJ899945	FJ899968	FJ899956	FJ899934	FJ899980	FJ899992	FJ900004	FJ900016	FJ900028	FJ900040	FJ899936
<i>Vesicomya gigas</i>	EU403432	EU403441	FJ899938	FJ899961	FJ899950	EU403450	FJ899973	FJ899985	FJ899997	FJ900009	FJ900021	FJ900033	EU403459
<i>Vesicomya</i> sp. mt-I	EU403439	EU403448	FJ899948	FJ899971	FJ899959	EU403457	FJ899983	FJ899995	FJ900007	FJ900019	FJ900031	FJ900043	EU403466
symA, <i>V.</i> sp. mt-II (JdF)	EU403438	EU403447	FJ899947	FJ899970	FJ899958	EU403456	FJ899982	FJ899994	FJ900006	FJ900018	FJ900030	FJ900042	EU403465
symB, <i>V.</i> sp. mt-II (JdF)	EU403434	EU403443	FJ899941	FJ899964	FJ899952	EU403452	FJ899976	FJ899988	FJ900000	FJ900012	FJ900024	FJ900036	EU403461
<i>Vesicomya</i> sp. mt-III	EU403437	EU403446	FJ899946	FJ899969	FJ899957	EU403455	FJ899981	FJ899993	FJ900005	FJ900017	FJ900029	FJ900041	EU403464
<i>Vesicomya</i> sp. BR	FJ899927	FJ899930	FJ899943	FJ899966	FJ899954	FJ899933	FJ899978	FJ899990	FJ900002	FJ900014	FJ900026	FJ900038	FJ899935
<i>Vesicomya</i> sp. GOM	FJ899926	FJ899929	FJ899940	FJ899963	NA	FJ899932	FJ899975	FJ899987	FJ899999	FJ900011	FJ900023	FJ900035	NA
<i>Vesicomya</i> sp. MAR	EU403435	EU403444	FJ899942	FJ899965	FJ899953	EU403453	FJ899977	FJ899989	FJ900001	FJ900013	FJ900025	FJ900037	EU403462

^asymbiont sequences are identified according to the host species from which they were obtained

^bsymbiont sequences were taken from the genome of the *C. magnifica* symbiont, Candidatus *Ruthia magnifica* (GenBank NC_008610); protein accession numbers are listed for protein coding loci

^csymbiont sequences were taken from the genome of the *C. okutanii* symbiont, Candidatus *Vesicomysocius okutanii* (GenBank NC_009465); protein accession numbers are listed for protein coding loci

NA = not available

Supplementary table 2. PCR and sequencing primers for vesicomid symbiont loci

gene	gene	position ^a	forward primer (5' -> 3') ^b	T _m (°C)	reverse primer (5' -> 3') ^b	T _m (°C)
16S	small subunit rRNA	1,081,270	27F - AGAGTTTGATCMTGGCTCAG	53.2	1492R - ACGGCTACCTTGTTACGACTT	56.1
			536F ^c - TGCAAGCGTTAATCGGAATTACTG	55.9	1029R ^c - CCATGCAGCACCTGTATTCGCATT	60.2
23S ^d	large subunit rRNA	1,077,544	2399F - AGAAGGCGATGAAGGACGTGATAATC	58.4		
			2769F - CCAAGGCTAAATACTCTCTAATGACCG	56.7		
			3505F - GACCGTCAGCTAAGGTCCCAA	58.9	3505R - TTGGGACCTTAGCTGACGGTC	58.9
			4181F - GAGAACTCGGGTGAAGGAAGTAG	56.6	4181R - CTAGTTCCTTACCCGAGTTCTC	56.6
			4761F - CATGGTGGGTAGTTTGACTGG	55.6	4761R - CCAGTCAAACACCCACCATG	55.6
					5413R - GTCAAGCCTCACGATCAATTAGTACAGG	58.8
<i>acrB</i> ^e	acriflavin resistance protein	594,313	<i>acrBF</i> - ATTGGTATTTTGGTGGATGACGC	60.1	<i>acrBR</i> - GGACCTGGTGGCATTTCACAACAACA	56.0
<i>ccb3</i>	cytochrome c oxidase, <i>ccb3</i> -type, subunit I	44,618	<i>ccb3F</i> - GGTGTTGTTAAATGGTTTGC	50.5	<i>ccb3R</i> - CATGACCAATCGTCCAATCTG	53.7
<i>COI</i> ^f	cytochrome c oxidase	40,373	<i>symCOIF</i> -CCATCACGGTCTGAAAAAGGC	58.6	<i>symCOIR</i> -GCGTACATAAAAATACTAACC	51.9
<i>dnaK</i> ^e	chaperone protein	381,478	<i>dnaKF</i> - TGGTGATGCCTGGGTTGAAGTTA	58.6	<i>dnaKR</i> - CAGCATGTGCTTCAGCATCT	55.9
<i>dsrB</i> ^e	sulfite reductase, dissimilatory, beta subunit	925,037	<i>dsrBF</i> - GGCTGAAGCACCAAGAATGCCTATC	60.1	<i>dsrBR</i> - GCTGGCCAACCAATACGATC	56.4
<i>ftsK</i> ^e	cell division protein	49,963	<i>ftsKF</i> - ATGCCGCACCTTGCTTGTGCTG	61.5	<i>ftsKR</i> - ATACGTGCTGCACGGTTATAGCC	59.3
<i>ftsY</i> ^e	signal recognition particle docking protein	825,783	<i>ftsYF</i> - ATCACTGCTGATATTGGCATT	52.3	<i>ftsYR</i> - CCTACGCCAATATAACGAATAGGT	54.3
<i>groEL</i> ^f	chaperonin	612,860	<i>groEL F2</i> - GAAGRYAARTTYGARAAYATGGG	52.1	<i>groEL R2</i> -TCWACRCVGCYTTYTYTCTTTC	57.2
<i>pfk</i> ^e	phosphofructokinase	110,906	<i>pfkF</i> - TGCTTTTTATGCACAATCAGG	51.8	<i>pfkR</i> - CAAGTAATCTGCAACTGCCAGT	57.6
<i>sigpep</i> ^e	signal peptidase I	675,432	<i>sigpepF</i> - CACAAAGTAATGCCCTTCTGG	53.9	<i>sigpepR</i> - GCATGGGATGTTTCTTCATTCTTGGC	58.8
<i>soxA</i> ^e	sulfur oxidation protein	858,411	<i>soxAF</i> - CCACAGTGGCTAGTTTGGGTTTGT	59.7	<i>soxAR</i> - TCAAACAACCATCATAACGACGATG	55.7

^a gene position (in basepairs) on the genome of the vesicomid symbiont *Candidatus Ruthia magnifica* (GenBank NC_008610; total chromosome length: ~1.16 Mbp)

^b PCR conditions: initial denaturation for 5min@94°C; 35 cycles of 45sec@94°C, 45sec@T_m-4°C, 2min@72°C; extension for 7min@72°C. PCR mixture: 1X Taq buffer (Invitrogen™), 0.5mM dNTP, 1.0 uM each primer, 2.0 mM MgCl₂, 0.03 U ul⁻¹ Taq DNA polymerase (Invitrogen™)

^c primers 536F and 1029R were designed based on conserved regions in the 16S sequences of the symbionts used in this study; they were used in conjunction with 27F and 1492R for direct sequencing

^d primers for 23S were designed based on conserved regions in the ribosomal RNA operon (16S, ITS, 23S regions) from the genomes of the vesicomid symbionts *Ruthia magnifica* and *Candidatus Vesicomiosocius okutanii* (NC_009465); primer names correspond to basepair positions relative to the start of the 16S rRNA gene; primers were used in varying combinations for PCR and direct sequencing

^e PCR and sequencing primers for symbiont protein coding loci were developed based on conserved regions in genes from the *R. magnifica* and *V. okutanii* genomes

^f degenerate symbiont *groEL* primers provided by A. Baik

Supplementary table 3. Substitution models for phylogenetic analyses of symbiont loci

locus ^a	model ^b
<i>COI</i>	GTR+SS
<i>cbb3</i>	GTR+SS
<i>ftsK</i>	GTR+SS
<i>pfK</i>	GTR+SS
<i>dnaK</i>	GTR+SS
<i>acrB</i>	GTR+SS
<i>groEL</i>	GTR+SS
<i>sigpep</i>	GTR+SS
<i>ftsY</i>	GTR+SS
<i>soxA</i>	HKY+G
<i>dsrB</i>	GTR+SS
23S	GTR+I
16S	GTR+I+ Γ

^a locus: loci ordered by chromosome position (fig. 1)

^b best-fit substitution model used for Bayesian phylogenetic analyses and in maximum likelihood tests of gene tree incongruence (table 3); models are those with the lowest Akaike Information Criterion (AIC) score, as determined using ModelTest;

GTR = general time reversible

HKY = Hasegawa-Kishino-Yano

SYM = symmetric (equal) base frequencies

SS = site-specific rate variation

I = proportion of invariant sites

Γ = gamma-distributed rate variation

Supplementary table 4. Nucleotide diversity and substitution patterns in vesicomylid symbionts (n=14)

locus ^a	G+C	π	K_a	K_s	K_a/K_s
<i>COI</i> ^b	0.40	0.076 (0.008)	0.015 (0.009)	0.395 (0.211)	0.040 (0.020)
<i>cbb3</i>	0.38	0.084 (0.007)	0.017 (0.008)	0.418 (0.194)	0.041 (0.014)
<i>ftsK</i>	0.40	0.111 (0.011)	0.057 (0.032)	0.413 (0.224)	0.163 (0.127)
<i>pfk</i>	0.39	0.100 (0.008)	0.036 (0.019)	0.432 (0.204)	0.094 (0.046)
<i>dnaK</i>	0.38	0.092 (0.007)	0.021 (0.009)	0.480 (0.239)	0.056 (0.062)
<i>acrB</i>	0.37	0.105 (0.008)	0.035 (0.018)	0.496 (0.236)	0.075 (0.044)
<i>groEL</i>	0.39	0.094 (0.007)	0.028 (0.013)	0.422 (0.211)	0.073 (0.035)
<i>sigpep</i>	0.31	0.116 (0.010)	0.068 (0.032)	0.392 (0.183)	0.182 (0.049)
<i>ftsY</i>	0.38	0.119 (0.010)	0.053 (0.022)	0.520 (0.266)	0.121 (0.066)
<i>soxA</i> ^b	0.39	0.109 (0.008)	0.062 (0.032)	0.376 (0.159)	0.166 (0.059)
<i>dsrB</i>	0.42	0.081 (0.006)	0.018 (0.009)	0.392 (0.176)	0.046 (0.018)
<i>23S</i>	0.48	0.024 (0.002)	NA	NA	NA
<i>16S</i>	0.50	0.022 (0.002)	NA	NA	NA

G+C = mean G+C proportion

π = nucleotide diversity, the mean number of nucleotide differences per site between two sequences

K_a = mean number of nonsynonymous substitutions per nonsynonymous site for any pair of sequences

K_s = mean number of synonymous substitutions per synonymous site for any pair of sequences

K_a/K_s = mean K_a/K_s for any pair of sequences

^a locus: loci ordered by chromosome position (fig. 1)

^b *COI* and *soxA* analyses did not include the symbiont of *Vesicomyla* sp. GOM; these loci could not be amplified from this symbiont

Parentheses = standard deviations

Analyses were run in DNAsp v.4.0; outgroups excluded

Supplementary figure 1. Bayesian phylogenies for vesicomylid 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 unrooted 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 rooted phylogenies for 13 symbiont loci. Posterior probabilities (PP) 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 ~5X greater in protein-coding loci than in 16S and 23S rRNA genes (note differences in scale bars). For each locus, the vesicomylid symbiont clade is separated from the 3 most-closely related free-living bacteria identified via BLAST. Trees are midpoint rooted and ordered based on the position of loci in the *Ruthia magnifica* genome (fig. 1). Here, *V. sp. JdF* = *V. sp. mt-II* of Goffredi et al. (2003) and Stewart et al. (2008).

Supplementary figure 2. Collection sites (red stars) of vesicomylid host species used in this study. Here, *Vesicomyla sp. JdF* is equivalent to *Vesicomyla sp. mt-II* of Goffredi et al (2003). Map courtesy of R.C. Vrijenhoek.

Supplementary figure 2

