## SUPPLEMENTAL MATERIALS AND METHODS

### *Sample evaluation via 16S rRNA barcoding*

Extracted DNA was amplified using primers targeting the V1 to V3 regions of the 16S rRNA gene (V1-9F, 5′-CCTATCCCCTGTGTGCCTTGGCAGTC-TCAG-AC-GAGTTTGATCMTGGCTCAG-3′, where the underline indicates the gene specific section, and V3-541R, 5′-CCATCTCATCCCTGCGTGTCTCCGAC-TCAG-*X*-AC-WTTACCGCGGCTGC-TGG-3′, where an *X* barcode was uniquely designed for each sample, followed by a common linker AC). PCR reactions were carried out under the following conditions: initial denaturation at 94 °C for 5 min, followed by 10 cycles of denaturation at 94 °C for 30 s, annealing at 60 °C to 55 °C with a touch-down program for 45 s, and elongation at 72 °C for 90 s. This was followed by an additional 20 cycles of denaturation at 94 °C for 30 s, annealing at 55 °C for 45 s, and elongation at 72 °C for 90 s. The amplified products were purified using resin columns (Qiagen), and 1 μg of PCR product for each sample was mixed and subjected to DNA pyrosequencing using the Roche/454 GS FLX Titanium platform, according to the manufacturer's instructions (performed by Chunlab Inc., Seoul, Korea).

## SUPPLEMENTAL RESULTS AND DISCUSSION

### *Results from pentanucleotide partitioning*

The *Osedax* symbiont Rs1 contigs partitioned into three distinct bins; the target symbiont (which comprised 71% of the contigs >4kb), the host genome (17%), and what is presumed to be the mitochondrial fraction from the host (12%; Figure S2). Binning confirmed that the nucleic acid sample for *Osedax* symbiont Rs2 had been contaminated by several other bacteria (Figure S1), which resulted in 3 distinct bacteria-related bins, including the target symbiont (30% of contigs), *Rhodothermus*-like (22%) and *Cytophaga*-like (48%; Figure S2). These non-target organisms belonged to the CFB group, which have been observed to be a common contaminant in the mucous sheath surrounding the host worms (Goffredi et al, 2007; Verna et al, 2010). We, therefore, suspect that these bacteria were present, transiently, on the outside of the root tissue prior to homogenization. Nonetheless, successful binning and subsequent validation by the singular occurrence of single-copy core genes (i.e., *dnaK, gyrA,*

*pgk, rpoC, secA*; Table S2) in the symbiont Rs2 reads provided confidence that this genome is robust.

***General metabolic capabilities***

A full complement of genes central to carbon metabolism, including glycolysis, gluconeogenesis, the pentose phosphate pathway, TCA cycle, and electron transport were identified in both *Osedax* symbionts (Tables S4-S5). They also notably possess genes encoding enzymes involved in pathways that shunt the TCA cycle, including the glyoxylate and methylcitrate cycles, as well as long chain fatty acid ligases. The symbionts lack carbon fixation genes, nitrogenases, and hydrogenases, the latter two of which are capabilities employed by some close relatives. Both symbionts also contained the genes necessary to synthesize NADH dehydrogenases (*nqr* genes), $N^+/H^+$ antiporters (*nhaA-D*), V-type ATP synthase subunits (*atpA-H*), and numerous cytochromes and oxidoreductases for aerobic energy metabolism. Genes involved in DNA and RNA metabolism, including purine and pyrimidine biosynthesis, replication and repair, ribosomal protein synthesis, and protein translation and modification were present in both genomes.

The symbionts appear capable of de novo synthesis of all B vitamins, except thiamine (B1) and biotin (B7) for which they are lacking one gene, like their closest relative *Neptuniibacter caesariensis*. They also possess the full complement of genes necessary to synthesize all essential amino acids, except asparagine and tyrosine. Both symbionts have the genomic capacity for ammonia and nitrate assimilation, as well as dentirification. Nitrate can be reduced to nitrite and further, in the Rs1 symbiont, to nitrogen gas. Although a nitrate uptake mechanism was not identified, genes for nitrate-, nitrite-, nitric oxide-, and nitrous oxide-reductases were present. Genes necessary for the ABC-uptake of urea were identified, however, ureases were not. The symbionts possess sulfate permeases, as well as most of the genes necessary to perform assimilatory sulfate reduction, with the exception of adenylylsulfate kinase. There were virtually no genes involved in secondary metabolism present in the genomes (Tables S4-S5). In both symbionts, mobile elements and transposases were identified, along with phage-related regions. The general dsitribution of gene groupings, based on SEED subsystems and COGs, was similar between the symbionts and *N.*

*caesariensis* (Tables S4-S5), with a few exceptions. The reduced number of genes involved in DNA and protein metabolism identified in symbiont Rs2 likely reflects the incomplete nature of this genome, despite similar values for other categories, such as RNA metabolism, stress response, and respiration, to name a few. Nevertheless, the *Osedax* symbiont Rs2 had a full complement of fourteen $Na^+$-translocating NADH-ubiquinone oxidoreductase genes (*nqrA-F*) and all eight ATP synthase subunits (*atpA-H*).

In addition to specific differences between the symbionts noted in the main text, they also showed differences in genomic aspects of respiration and energy production. For example, the Rs2 symbiont uniquely possessed genes encoding anaerobic DMSO reductases (for subunits ABC and regulation, *dorX*), NADH-ubiquinone oxidoreductases (*nuoA-N*), and general secretion pathway proteins (C-L), found in many pathogen and symbionts, yet the proteins remain poorly characterized. With regard to differences in carbon usage, glycolate utilization (*glcCDEF*) genes were identified in the Rs1 symbiont, whereas genes encoding maltose (*malFG*) and α-glucoside transport (*aglAEFGK*) were identified in the Rs2 symbiont, suggesting subtle differences in their carbohydrate usage profiles (Tables 3-4). Genes required for cell division, cell wall construction (LPS/murein/peptidoglycan), and capsule formation were observed for both symbionts, with some distinguishing characteristics between them, including the presence of genes involved in alternative capsular polysaccharide biosynthesis (*wcbQR*) and export (*kpsCDEMST*) only in the Rs1 symbiont (Table 3).

***Potential Virulence-associated genes.***

Filamentous extracellular organelles (ex. pili or fimbriae) are common among bacteria, and mediate adhesion, motility, biofilm formation, and pathogenesis. Both *Osedax* symbionts contain *pil* genes responsible for Type IV pili formation, including pilin subunits, a prepilin peptidase, assembly and retraction ATPases, an inner membrane protein, and a secretin. Similar to other bacteria, these genes were arranged along two operons of *pilBCD*/prepilin peptidase (4.8 kb in length) and *pilQMNOP*/ multimodular transpeptidase-transglycosylase *mrcA* (8.4 kb in length), with several copies of the *pilT* twitching motility gene interspersed throughout the genomes. The *Osedax* symbiont type-II *pilQMNOP* cluster shares sequence conservation among divergent species and is similar, in gene order, to the *pil* cluster in

*Pseudomonas aeruginosa* (Wall and Kaiser 1999; Craig and Li, 2008). Genes encoding a mannose-sensitive hemagglutinin (MSHA)-like Type-IVa pilus were identified in both symbionts (*mshBCDEGIJ* and *mshOP* in symbiont Rs1), which function in surface attachment and colonization in *Vibrio cholerae* and other pathogenic bacteria (Chiavelli et al, 2001; Dalisay et al, 2006). Additionally, the Rs1 symbiont uniquely possesses a Type-IVb tight adherence (*tad*) pilus-encoding gene cluster, including the ATP hydrolase and *flp* (fimbrial low-molecular-weight protein) pilus assembly proteins *tadABCG*. Tad genes are often present on widespread colonization islands, and are essential for biofilm formation and colonization in many pathogens, including *Hemophilus, Pseudomonas,* and *Yersinia* (Tomich et al, 2007), as well as the mutual human gut symbiont *Bifidobacterium breve* (O'Connell Motherway et al, 2011). Although found in many bacteria, close Oceanospirilalles relatives did not possess this gene cluster.

Type VI secretion systems, despite broad physiological significance, similarly appear to be a key virulence factor for certain pathogens (Pukatzki et al, 2007; Schell et al, 2007; Shalom et al, 2007; Bingle et al, 2008). Like Type IV secretion systems, Type VI systems are large multi-protein complexes encoded for by *imp* genes (or *vas*, virulence-associated secretion; Bingle et al, 2008; Schwarz et al, 2010). In the *Osedax* Rs1 symbiont the *imp* cluster is made of 10 contiguous genes (18.5 kbp in length), including *impBCGHIJK*, a lipoprotein, *clpB*, and a sigma-54-dependent transcriptional regulator, in a configuration similar to that observed for *Vibrio cholerae* and *Yersinia pestis,* with the exception of a different location for the *impD* gene in the symbionts. The nearby *clpB* and σ54-dependent transcriptional regulator genes have also been shown to be essential for virulence. *clpB* and other similar ATPase proteins have been specifically linked to host cell interactions, possibly serving as an energy source for the Type VI apparatus (Badger et al, 2000; Pukatzki et al, 2009). Disruption of the Imp-associated σ54-dependent transcriptional activator in *V. cholerae* resulted in reduced expression of effectors secreted by the Type VI system and attenuated virulence (Pukatzki et al, 2007). Both symbionts also have the valine-glycine repeat protein G gene (*vrgG*), additionally found to be necessary, once secreted by the Type VI secretion system, for contact-dependent cytotoxicity in *V. cholerae* (Pukatzki et al, 2007). These capabilities, related to typical secretion systems (Type-IV and Type-VI), could similarly mediate

adherence and virulence interactions between the *Osdeax* symbionts and their eukaryotic hosts.

The *Osedax* symbionts also possess genes that may further facilitate entry into the host. Both symbionts possess genes with homology to the internalin A (*inlA*) of *Listeria monocytogenes*, which functions in adhesion and invasion and is required for movement into host cells (Gaillard et al, 1991; Lecuit et al, 1997). One specific *inlA* homolog in the Rs1 symbiont was notably related to those found in both *Clostridia perfringens* and in *T. teridinibacter*, the Oceanospirillales symbiont of wood-feeding bivalves. Similarly, large genes (~up to 1350 aa) were identified in both symbionts with homology to RTX-like toxins (50% similarity in amino acid content) found widely among gram-negative bacteria (Coote 1992; Lally et al, 1999). These toxins, also known as T1SS secreted agglutinins, have been observed in both symbiotic and pathogenic *Vibrio* species (*V. fischeri* and *V. cholerae*, respectively) and, at least for *V. cholerae*, are known to aid in the disruption of membrane integrity and eventual invasion by the bacteria (Fullner et al, 2002; Ruby et al, 2005). Additional genes that encode large surface exoproteins were identified in both symbionts, and in many cases were homologous to hemolysin genes observed in other bacteria known to be involved in adhesion and intracellular heme utilization. Several hemolysin homologs identified in the Rs1 symbiont were most closely related to those found in the pathogens *Acinetobacter baumanni, Clostridia,* and *Pseudomonas* species. A well-known symbiont of entomopathogenic nematodes also possesses many similar adhesins, toxins, and hemolysins, suggesting their role in mediating both adverse and beneficial interactions with eukaryotes (Ffrench-Constant et al, 2000; Duchaud et al, 2003). Likewise, of the related Oceanospirillales species, only *Hahella chejuensis*, which is thought to be a pathogen of marine eukaryotes, possesses similar hemolysin and RTX-like toxins (Jeong et al, 2005). *Hahella chejuensi* is the bacterium for which the symbionts share most of their gene repertoire (~78%) perhaps providing insight into the dual nature of microbes that interact intimately with higher organisms. Genomic evidence will ultimately help understand the extent to which these bacteria, with larger genomes, are able to switch between free-living and symbiotic (both mutual and pathogens) lifestages. It is likely that an important contribution to our understanding of the common strategies between symbionts and pathogens will be the examination of invertebrate hosts, for

which there are many similarities to mammals, and thus could shed light on bacteria-animal interactions at numerous ecological and evolutionary levels.

**SUPPLEMENTAL REFERENCES**


Badger JL, Young BM, Darwin AJ, Miller VL. (2000). *Yersinia enterocolitica* ClpB affects levels of invasin and motility. J Bacteriol 182:5563–5571.

Bingle LE, Bailey CM, Pallen MJ. (2008). Type VI secretion: a beginner's guide. Curr Opin Microbiol 11:3–8.

Chiavelli DA, Marsh JW, Taylor RK. (2001). The mannose-sensitive hemagglutinin of *Vibrio cholerae* promotes adherence to zooplankton. Appl Environ Microbiol 67:3220–3225.

Coote, JG. (1992). Structural and functional relationships among the RTX toxin determinants in Gram-negative bacteria. FEMS Microbiol Rev 88:137–162.

Craig L, Li J. (2008). Type IV pili: paradoxes in form and function. Curr Opin Struct Biol 18:267–277.

Dalisay DS, Webb JS, Scheffel A, Svenson C, James S, Holmstrom C, Egan S and Kjelleberg S. (2006). A mannose-sensitive haemagglutinin (MSHA)-like pilus promotes attachment of *Pseudoalteromonas tunicata* cells to the surface of the green alga *Ulva australis*. Microbiol 152:2875–2883.

Duchaud E, Rusniok C, Frangeul L, Buchrieser C, Givaudan A, Taourit S, Bocs S, Boursaux-Eude C, Chandler M, Charles JF, Dassa E, Derose R, Derzelle S, Freyssinet G, Gaudriault S, Médigue C, Lanois A, Powell K, Siguier P, Vincent R, Wingate V, Zouine M, Glaser P, Boemare N, Danchin A, Kunst F. (2003). The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens*. Nat Biotechnol 21:1307–1313.

Ffrench-Constant RH, Waterfield N, Burland V, Perna NT, Daborn PJ, Bowen D, Blattner FR. (2000). A genomic sample sequence of the entomopathogenic bacterium *Photorhabdus luminescens* W14: Potential implications for virulence. Appl Environ Microbiol 66:3310–3329.

Folmer O, Black M, Hoeh W, Lutz R, Vriejenhoek R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates.

Mol Mar Biol Biotechnol 3:294–299.

Fullner KJ, Boucher JC, Hanes MA, Haines GK III, Meehan BM, Walchle C, Sansonetti PJ, Mekalanos JJ. (2002). The contribution of accessory toxins of *Vibrio cholerae* O1 El tor to the proinflammatory response in a murine pulmonary cholera model. J Exp Med 195:1455–1462.

Gaillard J-L, Berche P, Frehel C, Gouin E, Cossart P. (1991). Entry of *L. monocytogenes* into cells is mediated by internalin, a repeat protein reminiscent of surface antigens from gram-positive cocci. Cell 65:1127–1141.

Giribet G, Carranza S, Baguñà J, Riutort M, Ribera C. (1996). First molecular evidence for the existence of a Tardigrada + Arthropoda clade. Molec Biol Evol 13:76–84.

Goffredi SK, Johnson SB, Vrijenhoek RC. (2007). Genetic diversity and potential function of microbial symbionts associated with newly discovered species of *Osedax* polychaete worms. Appl Environ Microbiol 73:2314–2323.

Jeong H, Yim JH, Lee C, Choi SH, Park YK, Yoon SH, Hur CG, Kang HY, Kim D, Lee HH, Park KH, Park SH, Park HS, Lee HK, Oh TK, Kim JF. (2005). Genomic blueprint of *Hahella chejuensis*, a marine microbe producing an algicidal agent. Nucleic Acids Res 33:7066–7073.

Lally ET, Hill RB, Kieba IR, Korostoff J. (1999). The interaction between RTX toxins and target cells. Trends in Micro 7:356–361.

Lecuit M, Ohayon H, Braun L, Mengaud J, Cossart P. (1997). Internalin of *Listeria monocytogenes* with an intact leucine-rich repeat region is sufficient to promote internalization. Infect Immun 65:5309–5319.

Meyer B, Imhoff JF, Kuever J. (2007) Molecular analysis of the distribution and phylogeny of the *soxB* gene among sulfur-oxidizing bacteria – evolution of the Sox sulfur oxidation enzyme system. Environ Microbiol 9:2957–2977

O'Connell Motherway M, Zomer A, Leahy SC, Reunanen J, Bottacini F, Claesson MJ, O'Brien F, Flynn K, Casey PG, Munoz JA, Kearney B, Houston AM, O'Mahony C, Higgins DG, Shanahan F, Palva A, de Vos WM, Fitzgerald GF, Ventura M, O'Toole PW, van Sinderen D. (2011). Functional genome analysis of *Bifidobacterium breve* UCC2003 reveals type IVb tight adherence (Tad) pili as an essential and conserved host-colonization

factor. Proc Natl Acad Sci 108:11217–11222.

Pukatzki S, Ma AT, Revel AT, Sturtevant D, Mekalanos JJ. (2007). Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. Proc Natl Acad Sci 104:15508–15513.

Pukatzki S, McAuley SB, Miyata, ST. (2009). The type VI secretion system: translocation of effectors and effector-domains. Curr Opin Microbiol 12:11–17.

Rouse GW, Goffredi SK, Johnson SB, Vrijenhoek RC. (2011). Not whale-fall specialists, *Osedax* worms also consume fishbones. Biol Lett 7:736–739.

Ruby EG, Urbanowski M, Campbell J, Dunn A, Faini M, Gunsalus R, Lostroh P, Lupp C, McCann J, Millikan D, Schaefer A, Stabb E, Stevens A, Visick K, Whistler C, Greenberg EP. (2005). Complete genome sequence of *Vibrio fischeri*: A symbiotic bacterium with pathogenic congeners. Proc Natl Acad Sci 102:3004–3009.

Schell MA, Ulrich RL, Ribot WJ, Brueggemann EE, Hines HB, Chen D, Lipscomb L, Kim HS, Mrázek J. (2007). Type VI secretion is a major virulence determinant in Burkholderia mallei. Mol Microbiol 64:1466–*1485.*

Schwarz S, Hood RD, Mougous JD. (2010). What is type VI secretion doing in all those bugs? Trends Microbiol 18:531–537.

Shalom G, Shaw JG, Thomas MS. (2007). In vivo expression technology identifies a type VI secretion system locus in *Burkholderia pseudomallei* that is induced upon invasion of macrophages. Microbiol 153:2689–2699.

Tomich M, Planet PJ, Figurski DH. (2007). The tad locus: postcards from the widespread colonization island. Nat Rev Microbiol 5:363–375.

Verna C, Ramette A, Wiklund H, Dahlgren TG, Glover AG, Gaill F, Dubilier N. (2010). High symbiont diversity in the bone-eating worm *Osedax mucofloris* from shallow whale-falls in the North Atlantic. Environ Microbiol 12:2355–2370.

Wall D, Kaiser D. (1999). Type IV pili and cell motility. Molec Microbiol 32:1–10.

**SUPPLEMENTAL TITLES AND LEGENDS TO FIGURES**

**Figure S1.** (A) Phylogenetic distribution of sequence reads based on 16SrRNA community analysis, via barcoding, of the DNA samples prior to genome sequencing, classified according to family. (B) Taxonomic distribution of annotated reads following metagenomic sequencing, based on BLASTP best hits (E-value < $10^{-12}$), classified according to phyla.

**Figure S2.** Principal component analysis of tetranucleotide partitioning of sequencing reads from symbiont Rs1 (A) and Rs2 (B). To assign metagenomic contigs and scaffolds to their likely taxonomic origin, we used a combinatory binning approach based on GC-content and Markov model-based statistical evaluations of pentamer frequencies for contigs > 4Kb. The symbiont Rs1 contigs partitioned into three distinct bins; the target symbiont (71%), the host genome (17%), and what is presumed to be the mitochondrial fraction from the host (12%). Binning confirmed that the nucleic acid sample for symbiont Rs2 was contaminated by several other bacteria, which resulted in 3 distinctly-spaced bacteria-related bins, including the target symbiont (which comprised 30% of the contigs >4kb), and two *Cytophaga*-related bins (22-48%).

**Figure S3.** Selective restriction digest of the 16S rRNA genes of the *Osedax* symbionts Rs1 and Rs2, shown in Figure 1. Restriction enzymes *Eco*RI (lanes 1,4,7), *Hae*III (lanes 2,5,8), and *Sfan*I (lanes 3,6,9) were used to distinguish symbiont ribotypes. For the DNA ladder at left, the upper dark band is 500 bp, while the lower band corresponded to 100 bp. Symbiont 'P2' is closely related to, yet distinct from, the Rs1 symbiont (as noted in Goffredi at al, 2007).

**Figure S4.** Targeted PCR amplification of the *soxB* gene from *Osedax* symbionts Rs1 and Rs2. Expected products are ~1014 bp for lanes 2-4 (primers 432F-1446R) and ~ 750 bp for lanes 6-8 (primers 693F-1446R).

**Figure S5.** Gene sequences for methyl-accepting chemotaxis genes, designated *mcp3* and *mcp4* for *Osedax* individuals housing symbiont Rs1 with identical 16S rRNA genes, collected from two different species (*Osedax rubiplumus* (Or) and *O. frankpressi* (Ofp), during 13 different

dives. Dive numbers begin with the remotely operated vehicle name; T= *Tiburon*, DR = *Doc Ricketts*. Genes amplified from the *Osedax* symbiont Rs1 corresponding to the genome sample is labeled and was collected during dive DR236. Sequences are shorter than 200-bp length, and thus not submitted to GenBank.

**Table S1 | Specimens, and collection sites, used in this study**

| Site latitude/longitude | Dive[1] | Date | Time (months) | # of specimens |
|---|---|---|---|---|
| 2890m 36.313 N/-122.434 W | T486 | Oct 2002 | 8 | 2 |
| | T610 | Aug 2003 | 18 | 3 |
| | T1069 | Jan 2007 | 59 | 2 |
| | DR010 | Mar 2009 | 85 | 3 |
| | DR098 | Nov 2009 | 93 | 4 |
| | DR204 | Oct 2010 | 104 | 1 |
| | DR234 | Jun 2011 | 112 | 2 |
| 1820m 36.708 N/-122.105 W | T1048 | Oct 2006 | 7 | 3 |
| | T1071 | Jan 2007 | 10 | 3 |
| | DR012 | Mar 2009 | 36 | 4 |
| | DR236[2] | Jun 2011 | 63 | 2 |
| 840m 32.767 N/-117.483 W | DR471 | May 2013 | unknown | 10 |

[1]Dive numbers begin with the remotely operated vehicle name; T= *Tiburon*, DR = *Doc Ricketts* (both owned and operated by the Monterey Bay Aquarium Research Institute).
[2]Samples used for genomic analysis were collected during dive DR236. Only the root tips, free of host ovisac, were used for symbiont separation and genome analysis.

**Table S2 | Estimate of completeness of the *Osedax* symbiont Rs1 and Rs2 genomes based on the presence and absence of 111 universal bacterial single-copy core genes (identified by Dupont et al 2012).**

| Core Gene | Rs1 | Rs2 | Core Gene | Rs1 | Rs2 |
|---|---|---|---|---|---|
| *alaS* | 1 | 1 | Ribosomal_L16 | 1 | 0 |
| *aspS* | 1 | 1 | Ribosomal_L17 | 1 | 0 |
| *cgtA* | 1 | 0 | Ribosomal_L18 | 1 | 0 |
| *cysS* | 2 | 2 | Ribosomal_L19 | 1 | 1 |
| dephospho-CoA kinase | 1 | 0 | Ribosomal_L20 | 1 | 1 |
| *dnaA* | 1 | 1 | Ribosomal_L21 | 1 | 0 |
| *dnaG* | 1 | 1 | Ribosomal_L22 | 1 | 1 |
| *dnaK* | 1 | 1 | Ribosomal_L23 | 1 | 1 |
| *dnaN* | 1 | 1 | Ribosomal_L24 | 1 | 0 |
| *dnaX* | 1 | 1 | Ribosomal_L27 | 1 | 0 |
| *dnlJ* | 1 | 1 | Ribosomal_L28 | 1 | 1 |
| *engA* | 1 | 1 | Ribosomal_L29 | 1 | 0 |
| *era* | 1 | 0 | Ribosomal_S2 | 1 | 1 |
| *ffh* | 1 | 1 | Ribosomal_S3 | 1 | 1 |
| *fmt* | 1 | 1 | Ribosomal_S4 | 1 | 0 |
| *frr* | 1 | 1 | Ribosomal_S5 | 1 | 0 |
| *ftsY* | 1 | 1 | Ribosomal_S6 | 1 | 1 |
| *glyQ* | 1 | 1 | Ribosomal_S7 | 1 | 1 |
| *glyS* | 1 | 1 | Ribosomal_S8 | 1 | 0 |
| *grpE* | 1 | 1 | Ribosomal_S9 | 0 | 1 |
| guanyl_kin | 1 | 1 | Ribosomal_S10 | 1 | 1 |
| *gyrA* | 1 | 1 | Ribosomal_S11 | 1 | 0 |
| *gyrB* | 1 | 1 | Ribosomal_S12 | 1 | 0 |
| *hisS* | 1 | 1 | Ribosomal_S13 | 1 | 0 |
| *IF-2* | 1 | 1 | Ribosomal_S15 | 1 | 1 |
| *ileS* | 1 | 1 | Ribosomal_S16 | 1 | 1 |
| *infC* | 1 | 1 | Ribosomal_S17 | 0 | 0 |
| *lepA* | 1 | 0 | Ribosomal_S18 | 1 | 1 |
| *leuS* | 1 | 1 | Ribosomal_S19 | 0 | 1 |
| Methyltransf_5 | 1 | 1 | Ribosomal_S20 | 1 | 1 |
| *nhaD* | 0 | 0 | RNaseIII | 1 | 0 |
| *nusA* | 1 | 1 | *rpmF* | 1 | 1 |
| *nusG* | 1 | 0 | *rpmH* | 1 | 1 |
| *pgk* | 1 | 0 | *rpmI* | 1 | 1 |
| *pheS* | 1 | 1 | *rpoA* | 1 | 0 |
| *pheT*_bact | 1 | 0 | *rpoB* | 1 | 0 |
| *pheT*_arch | 0 | 0 | *rpoC* | 1 | 1 |
| *prfA* | 1 | 1 | *rpoC1* | 0 | 0 |
| *proS*_famI | 0 | 0 | *secA* | 1 | 0 |
| *proS*_famII | 1 | 1 | *secE* | 1 | 0 |
| *pyrG* | 1 | 1 | *secG* | 1 | 0 |
| *rbfA* | 1 | 1 | *secY* | 1 | 0 |
| *recA* | 1 | 1 | *serS* | 1 | 1 |
| Ribosomal_L1 | 1 | 0 | *smpB* | 1 | 1 |
| Ribosomal_L2 | 1 | 1 | *thrS* | 1 | 1 |
| Ribosomal_L3 | 1 | 1 | *tig* | 1 | 1 |
| Ribosomal_L4 | 1 | 1 | *tilS* | 1 | 1 |
| Ribosomal_L5 | 1 | 0 | *trmU* | 1 | 1 |
| Ribosomal_L6 | 2 | 0 | tRNA-synt_1d | 1 | 1 |
| Ribosomal_L9 | 1 | 1 | *tsf* | 1 | 1 |
| Ribosomal_L10 | 1 | 0 | *tyrS* | 1 | 1 |
| Ribosomal_L11 | 1 | 0 | *uvrB* | 1 | 1 |
| Ribosomal_L12 | 1 | 0 | *valS* | 1 | 0 |
| Ribosomal_L13 | 1 | 1 | *ybeY* | 1 | 1 |
| Ribosomal_L14 | 1 | 0 | *ychF* | 1 | 1 |
| Ribosomal_L15 | 1 | 0 | **Total** | **104** | **73** |

**Table S3 | PCR primers, used in this study, for chemotaxis and motility genes in the *Osedax* symbiont genome**

| Specific gene | | Forward primer (5'-3') | | Reverse primer (5'-3') | $T_A$[5] |
|---|---|---|---|---|---|
| Oseadax symbiont 16S rRNA gene[1] | 435F | CAGCWGTGAGGAAAGGTT | 1213R | TGTGTAGCCCAACTCG | 58 |
| Cytochrome c oxidase (COI)[2] | LCO1490 | GGTCAACAAATCATAAAGATATTGG | HCO2198 | TAAACTTCAGGGTGACCAAAAAATCA | 50 |
| Eukaryote 18S rRNA gene[3] | 3F | GTTCGATTCCGGAGAGGGA | 5R | CTTGGCAAATGCTTTCGC | 54 |
| methyl-accepting chemotaxis protein 2 | IFex | AGA TCC TAT TAG TGA AAA GAT T | IIRc | CTT CGA TTG CCG CAT TC | 48 |
| methyl-accepting chemotaxis protein 3 | IFa | CAA ACC AAT TTA CTG GCT TTA A | IIR | CGA TTG TTG TTC GAC AGC | 48 |
| methyl-accepting chemotaxis protein 4 | IF | CAC GCA AAA GAC CTT GGT A | IIR | TGG ATG GGA AAT GAT ATT ACC | 48 |
| signal transduction histidine kinase (*cheA*) | IF | CCG AGA TAG ATA AAA CGG T | IIR | TGA AAA ACC AGC TTT GAA AAT T | 48 |
| chemotaxis response regulator (*cheY*) | IF | TCC GGC GTC ATC GCT ATC T | IIIR | ATT TGA TTC CGA GGT AAC CAT | 48 |
| chemotaxis-specific methylesterase (*cheB*) | IFm | CTA AAG CTA TGC GTG CTT GGA | IIIR | ATC GCC AGA CTC TGC ATG TTT | 48 |
| chemotaxis methyltransferase (*cheR*) | IF | CCA ACG AAA CAT TAT GGT TTC | IIR | AAA TAT ATT AAG ACG TTC CGA | 48 |
| flagellar motor switch protein (*fliM*) | IF | TCA AGA CGA AAT CGA TGC TCT G | IIIR | ATT ATC TAC CAG CTT AAA CAC CAA C | 48 |
| flagellum-specific ATP synthase (*fliI*) | IF | AAA GGT TAT CCG CCT TCG G | IR | ATA TCG ATC GCG GGA TAA TG | 48 |
| Sulfide oxidase complex (*soxB*)[4] | 432F | GAY GGN GGN GAY ACN TGG | 1446R | CAT GTC NCC NCC RTG YTG | 47 |
| Sulfide oxidase complex (*soxB*)[4] | 693F | ATC GGN CAR GCN TTY CCN TA | 1446R | CAT GTC NCC NCC RTG YTG | 47 |
| filamentous hemagglutinin (region1) | CF | AAC GCT AAC GGC CAA GTG T | ER | GCA CTG GCA TCC AAG GTG | 48 |
| filamentous hemagglutinin (region2) | marIF | TTT ATT AAT GAT GCC GTC AGC TG | marIIR | CAA TCC ATC TAA AGA GCC ACT AAA | 48 |

[1] taken from Rouse et al, 2011
[2] taken from Folmer et al, 1994
[3] taken from Giribet et al, 1996
[4] taken from Meyer et al, 2007
[5] Thermal cycling conditions included 25-30 cycles of 60 s each of denaturation at 94°C, annealing at the temperature indicated ($T_A$ in °C), elongation at 72°C and a final extension at 72°C for 6 min.

**Table S4 | Comparative analysis by relevant functional gene repertoires, ranked by % in the *Osedax* symbiont Rs1, according to SEED subsystem abundance.**

| Functional Category[1] | *Osedax* symbiont Rs1[3] | | *Osedax* symbiont Rs2 | | *Neptuniibacter caesariensis* | | *Neptunomonas japonica* | |
|---|---|---|---|---|---|---|---|---|
| | Gene # | % | Gene # | % | Gene # | % | Gene # | % |
| Amino Acids and Derivatives | 429 | 16.0 | 482 | 19.0 | 400 | 15.8 | 504 | 17.1 |
| Cofactors and Vitamins | 298 | 11.1 | 211 | 8.3 | 285 | 11.2 | 271 | 9.2 |
| Membrane Transport | 262 | 9.8 | 315 | 12.4 | 210 | 8.3 | 240 | 8.1 |
| Carbohydrates[2] | 244 | 9.1 | 296 | 11.7 | 167 | 6.6 | 260 | 8.8 |
| Protein Metabolism | 215 | 8.0 | 117 | 4.6 | 184 | 7.3 | 221 | 7.5 |
| RNA Metabolism | 189 | 7.0 | 180 | 7.1 | 182 | 7.2 | 186 | 6.3 |
| Cell Wall and Capsule | 145 | 5.4 | 114 | 4.5 | 143 | 5.6 | 122 | 4.1 |
| DNA Metabolism | 133 | 5.0 | 74 | 2.9 | 124 | 4.9 | 130 | 4.4 |
| Stress Response | 118 | 4.4 | 119 | 4.7 | 154 | 6.1 | 182 | 6.2 |
| Motility and Chemotaxis | 95 | 3.5 | 77 | 3.0 | 113 | 4.5 | 110 | 3.7 |
| Nucleosides/Nucleotides | 93 | 3.5 | 85 | 3.4 | 69 | 2.7 | 77 | 2.6 |
| Fatty Acids/Lipids | 87 | 3.2 | 104 | 4.1 | 92 | 3.6 | 112 | 3.8 |
| Respiration | 82 | 3.1 | 112 | 4.4 | 121 | 4.8 | 102 | 3.5 |
| Regulation and Cell Signaling | 62 | 2.3 | 50 | 2.0 | 59 | 2.3 | 81 | 2.7 |
| Virulence and Defense | 53 | 2.0 | 32 | 1.3 | 55 | 2.2 | 61 | 2.1 |
| Miscellaneous[4] | 47 | 1.8 | 43 | 1.7 | 25 | 1.0 | 47 | 1.6 |
| Nitrogen Metabolism | 30 | 1.1 | 32 | 1.3 | 30 | 1.2 | 61 | 2.1 |
| Phosphorous Metabolism | 28 | 1.0 | 17 | 0.7 | 34 | 1.3 | 37 | 1.3 |
| Cell Division | 23 | 0.9 | 21 | 0.8 | 22 | 0.9 | 25 | 0.8 |
| Potassium metabolism | 17 | 0.6 | 15 | 0.6 | 17 | 0.7 | 22 | 0.7 |
| Iron acquisition | 16 | 0.6 | 6 | 0.2 | 11 | 0.4 | 14 | 0.5 |
| Sulfur Metabolism | 8 | 0.3 | 15 | 0.6 | 34 | 1.3 | 29 | 1.0 |
| Metabolism of Aromatics | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 34 | 1.2 |
| Secondary Metabolism | 6 | 0.2 | 4 | 0.2 | 4 | 0.2 | 7 | 0.2 |
| Phage | 1 | 0.0 | 15 | 0.6 | 1 | 0.0 | 15 | 0.5 |
| TOTAL[5] | 2681 | | 2536 | | 2540 | | 2950 | |

[1] Based on SEED subsystems

[2] This category includes Glycolysis/Gluconeogenesis, the TCA cycle and Pentose Phosphate Pathway

[3] Gene copy numbers based on the Nextera library construction are expected to reflect a relative dedication of the genome to each gene category.

[4] This category includes genes involved in iron-sulfur cluster assembly and phosphoglycerate mutase family proteins, as well as others not included in SEED subsystems.

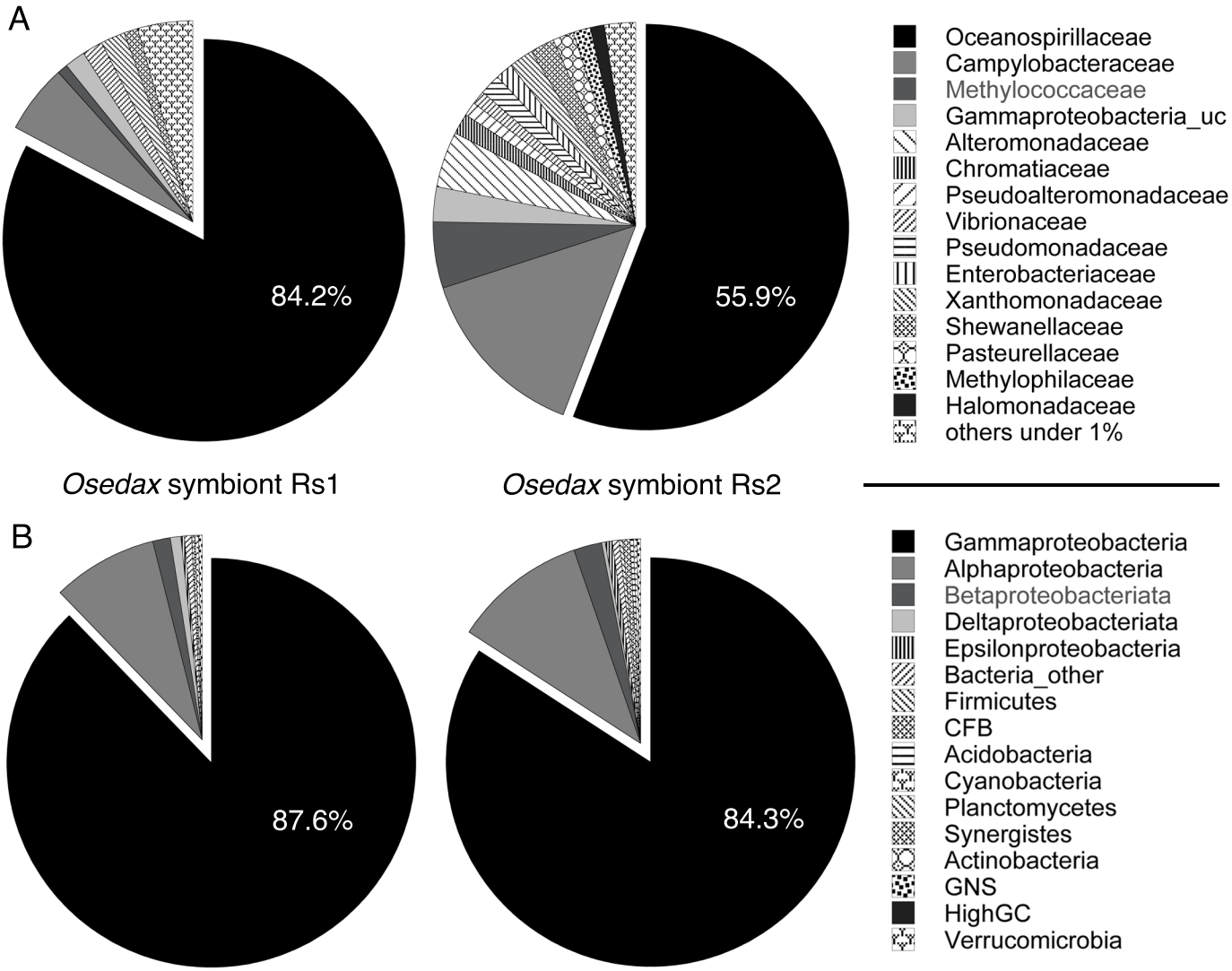
[5] Totals do not include conserved hypothetical

**Table S5 | Comparative analysis, by functional COG category, of gene repertoires in the *Osedax* symbionts Rs1 and Rs2**

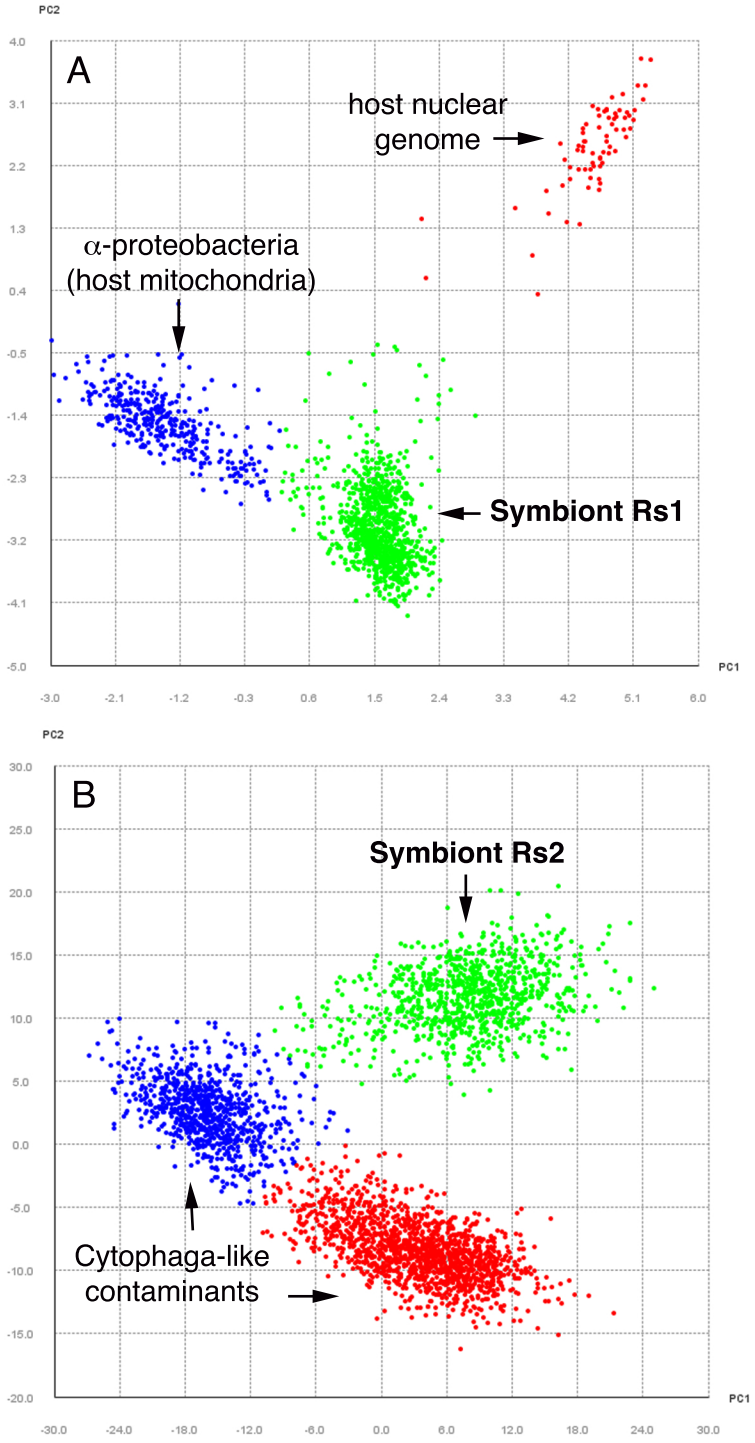
| Functional Category[1] | Symbiont Rs1 | | Symbiont Rs2 | |
|---|---|---|---|---|
| | gene # | % | gene # | % |
| C, energy production and conversion | 192 | 5.3 | 223 | 5.3 |
| D, cell division and chromosome partitioning | 52 | 1.4 | 37 | 0.9 |
| E, amino acid transport and metabolism | 410 | 11.4 | 527 | 12.5 |
| F, nucleotide transport and metabolism | 82 | 2.3 | 82 | 1.9 |
| G, carbohydrate transport and metabolism | 122 | 3.4 | 235 | 5.6 |
| H, coenzyme metabolism | 170 | 4.7 | 172 | 4.1 |
| I, lipid metabolism | 105 | 2.9 | 124 | 2.9 |
| J, translation, ribosomal structure and biogenesis | 202 | 5.6 | 176 | 4.2 |
| K, transcription | 274 | 7.6 | 444 | 10.5 |
| L, DNA replication, recombination and repair | 177 | 4.9 | 134 | 3.2 |
| M, cell envelope biogenesis, outer membrane | 196 | 5.4 | 198 | 4.7 |
| N, cell motility and secretion | 188 | 3.3 | 140 | 3.3 |
| O, posttranslational modification, protein turnover | 134 | 3.7 | 148 | 3.5 |
| P, inorganic ion transport and metabolism | 172 | 4.8 | 164 | 3.9 |
| R, general function predicted only | 402 | 11.2 | 506 | 12.0 |
| S, function unknown | 329 | 9.1 | 307 | 7.3 |
| T, signal transduction mechanisms | 226 | 6.3 | 341 | 8.1 |
| Intracellular trafficking, secretion | 206 | 2.9 | 100 | 2.4 |
| Secondary metabolite biosynthesis, transport | 78 | 2.2 | 102 | 2.4 |
| Other[2] | 58 | 1.6 | 50 | 1.2 |

[1] COG categories (Clusters of orthologous genes categories), based on the IMG database v4.1
[2] Includes Chromatin dynamics and Defense mechanisms

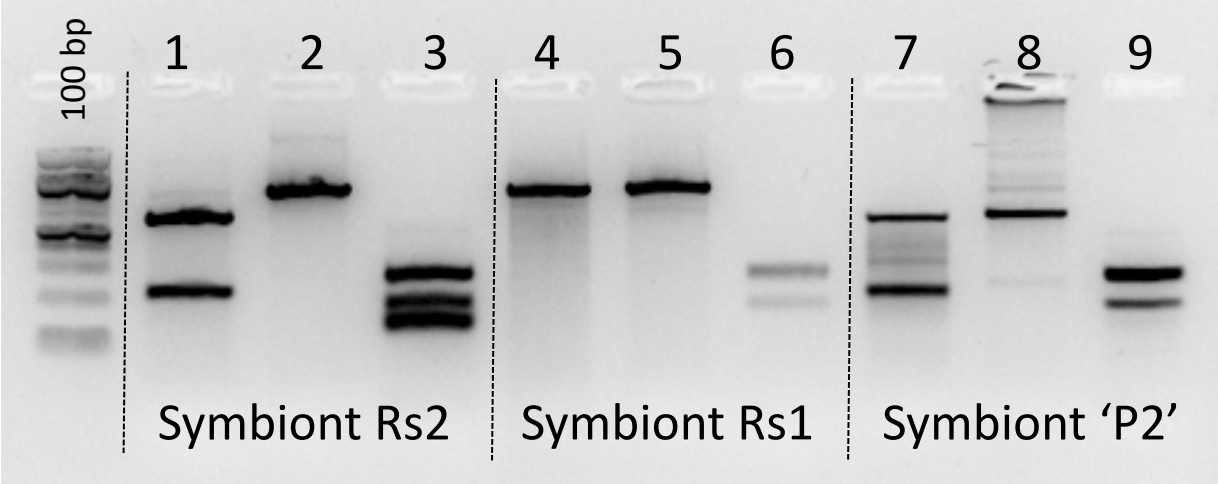
# Goffredi et al_Figure S1



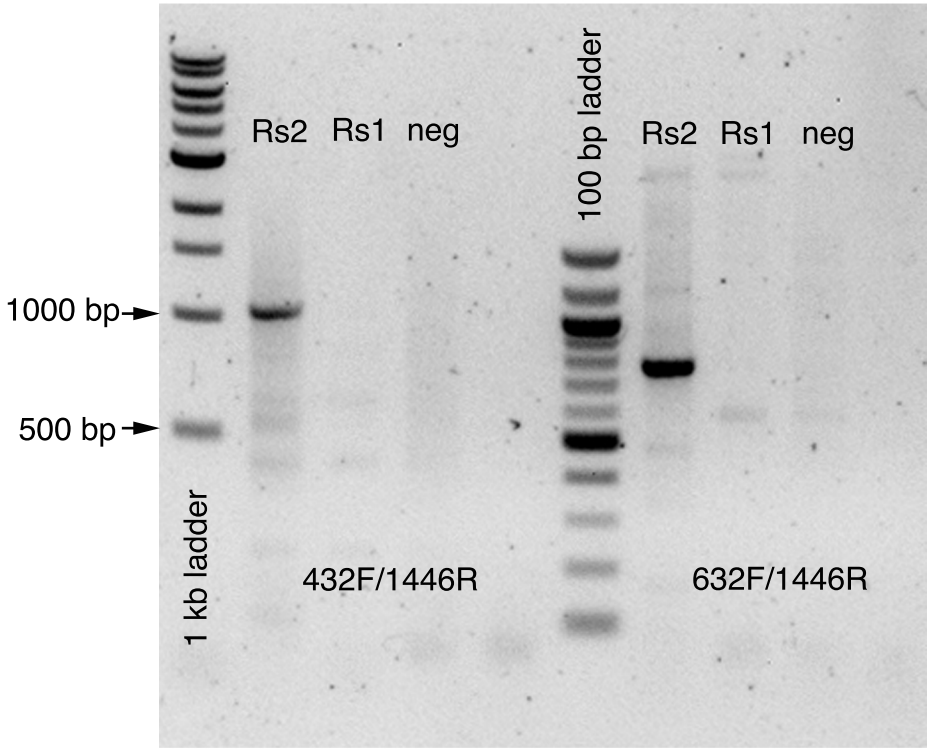
# Goffredi et al_Figure S2



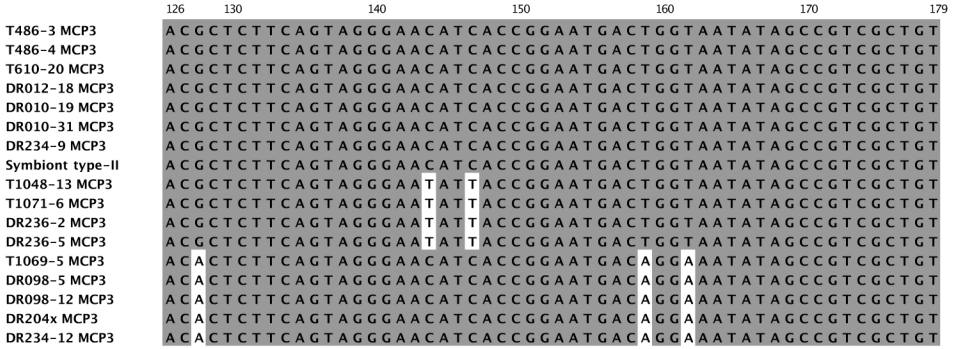
Goffredi et al_Figure S3



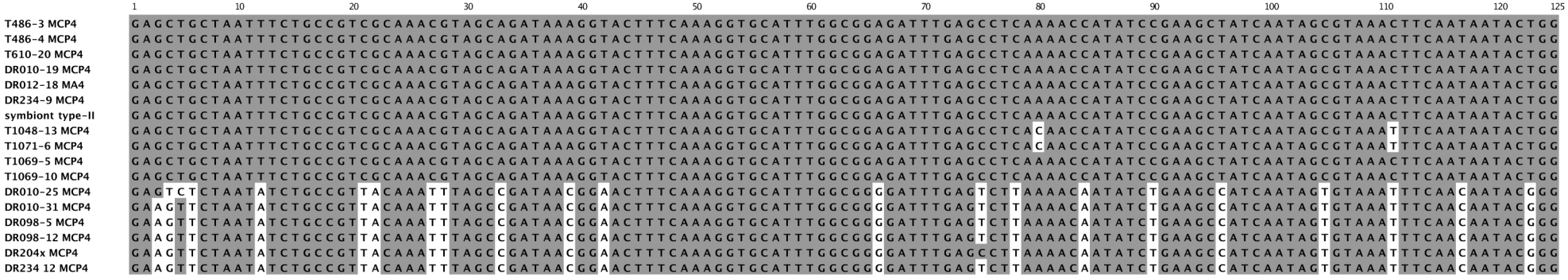
Goffredi et al_Figure S4

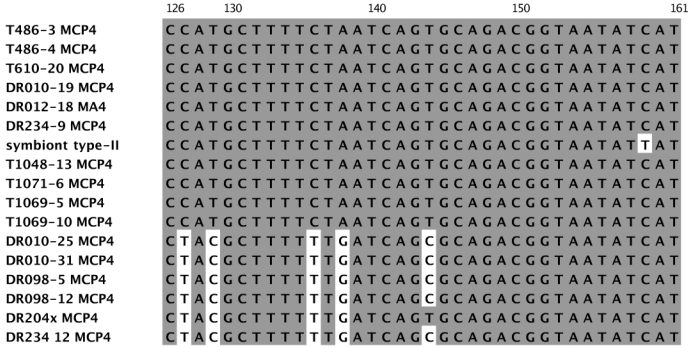


# Goffredi et al_FigureS5


methyl-accepting chemotaxis protein ‘3’




methyl-accepting chemotaxis protein ‘4’