

UC Davis

UC Davis Previously Published Works

Title

Genomic versatility and functional variation between two dominant heterotrophic symbionts of deep-sea *Osedax* worms

Permalink

<https://escholarship.org/uc/item/9f3942tx>

Journal

The ISME Journal: Multidisciplinary Journal of Microbial Ecology, 8(4)

ISSN

1751-7362

Authors

Goffredi, Shana K

Yi, Hana

Zhang, Qingpeng

et al.

Publication Date

2014-04-01

DOI

10.1038/ismej.2013.201

Peer reviewed

## ORIGINAL ARTICLE

# Genomic versatility and functional variation between two dominant heterotrophic symbionts of deep-sea *Osedax* worms

Shana K Goffredi<sup>1,5</sup>, Hana Yi<sup>2,5,6</sup>, Qingpeng Zhang<sup>3</sup>, Jane E Klann<sup>1</sup>, Isabelle A Struve<sup>1</sup>, Robert C Vrijenhoek<sup>4</sup> and C Titus Brown<sup>3</sup>

<sup>1</sup>Department of Biology, Occidental College, Los Angeles, CA, USA; <sup>2</sup>Institute of Molecular Biology and Genetics, Seoul National University, Seoul, Republic of Korea; <sup>3</sup>Computer Science and Engineering, Michigan State University, East Lansing, MI, USA and <sup>4</sup>Monterey Bay Aquarium Research Institute, Moss Landing, CA, USA

**An unusual symbiosis, first observed at ~3000 m depth in the Monterey Submarine Canyon, involves gutless marine polychaetes of the genus *Osedax* and intracellular endosymbionts belonging to the order Oceanospirillales. Ecologically, these worms and their microbial symbionts have a substantial role in the cycling of carbon from deep-sea whale fall carcasses. Microheterogeneity exists among the *Osedax* symbionts examined so far, and in the present study the genomes of the two dominant symbionts, Rs1 and Rs2, were sequenced. The genomes revealed heterotrophic versatility in carbon, phosphate and iron uptake, strategies for intracellular survival, evidence for an independent existence, and numerous potential virulence capabilities. The presence of specific permeases and peptidases (of glycine, proline and hydroxyproline), and numerous peptide transporters, suggests the use of degraded proteins, likely originating from collagenous bone matter, by the *Osedax* symbionts. <sup>13</sup>C tracer experiments confirmed the assimilation of glycine/proline, as well as monosaccharides, by *Osedax*. The Rs1 and Rs2 symbionts are genomically distinct in carbon and sulfur metabolism, respiration, and cell wall composition, among others. Differences between Rs1 and Rs2 and phylogenetic analysis of chemotaxis-related genes within individuals of symbiont Rs1 revealed the influence of the relative age of the whale fall environment and support possible local niche adaptation of ‘free-living’ lifestages. Future genomic examinations of other horizontally-propagated intracellular symbionts will likely enhance our understanding of the contribution of intraspecific symbiont diversity to the ecological diversification of the intact association, as well as the maintenance of host diversity.**

*The ISME Journal* (2014) 8, 908–924; doi:10.1038/ismej.2013.201; published online 14 November 2013

**Subject Category:** Microbe-microbe and microbe-host interactions

**Keywords:** *Osedax*; symbionts; deep sea; Oceanospirillales; whale fall

## Introduction

Nutritional symbioses involving bacteria and their invertebrate hosts have critical roles in sustaining high productivity in many marine environments. Some of these invertebrate taxa transmit their essential symbionts vertically through eggs, but surprisingly, many acquire their symbionts horizontally from the environment (Bright and Bulgheresi, 2010). Environmental symbionts must therefore face conflicting selective pressures for their host-

associated and free-living subpopulations. Coevolutionary pressures will constrain changes to bacterial genes that facilitate infection of aposymbiotic invertebrate larvae, engage in ‘hand-shaking’ with the host genome to avoid exclusion and coordinate metabolic interdependencies, as well as genes that enhance metabolism and reproduction in the host cytosol. The free-living subpopulation, on the other hand, might favor genes that slow metabolism and reproduction and enhance survival under nutrient-limited conditions (for example, in the deep sea). New details have emerged regarding the range of adaptations used by endosymbionts, parallels with pathogens, provisioning from symbiont to host and many ecological aspects, including transmission between host generations and the duality of negotiating between host-associated and free-living lifestages. Recent advances in molecular biology and genomics have enabled this insight for

Correspondence: SK Goffredi, Department of Biology Occidental College, 1600 Campus Road, Los Angeles, CA 90041, USA.

E-mail: sgoffredi@oxy.edu

<sup>5</sup>These are the co-first authors.

<sup>6</sup>Current address: Department of Public Health Science, Graduate School, Korea University, Seoul, Republic of Korea.

Received 21 April 2013; revised 21 September 2013; accepted 8 October 2013; published online 14 November 2013

chemosynthetic symbionts, in particular, hosted by the marine oligochaete *Olavius algarvensis* (Woyke *et al.*, 2006), vesicomid clam species (Kuwahara *et al.*, 2007; Newton *et al.*, 2007), the siboglinid tubeworm *Riftia pachyptila* (Robidart *et al.*, 2008) and the heterotrophic symbionts of wood-boring bivalves (Yang *et al.*, 2009).

An unusual symbiosis involving gutless marine polychaetes of the genus *Osedax* and heterotrophic gammaproteobacteria belonging to the order Oceanospirillales was first observed in 2002 living on a gray whale skeleton at ~3000 m depth in the Monterey Submarine Canyon, California (Goffredi *et al.*, 2004; Rouse *et al.*, 2004). Now, 17 morphologically diverse species are known worldwide from depths ranging from 25 to 3000 meters (Goffredi *et al.*, 2004; Rouse *et al.*, 2004; Glover *et al.*, 2005; Fujikura *et al.*, 2006; Braby *et al.*, 2007; Vrijenhoek *et al.*, 2009; Salathé and Vrijenhoek, 2012). Evidence for *Osedax* borings in fossilized bones suggests that the genus has historically colonized a large portion of the world's oceans (Higgs *et al.*, 2011; Kiel *et al.*, 2011; Higgs *et al.*, 2012). Much is now known about the role of these worms in degrading sunken bones and the recycling of deep-sea carbon (Goffredi *et al.*, 2005; Braby *et al.*, 2007), their phylogenetic relationships (Vrijenhoek *et al.*, 2009; Hilário *et al.*, 2011), and an unusual reproductive strategy that involves dwarf males and environmental sex determination (Rouse *et al.*, 2008; Vrijenhoek *et al.*, 2008; Rouse *et al.*, 2011; Miyamoto *et al.*, 2013).

*Osedax* are unique among members of the polychaete family Siboglinidae in that they possess a complex system of 'roots' with an ovisac covered by tissue which houses intracellular bacteria (Table 5; Goffredi *et al.*, 2004). Collagen from bone appears to be their nutritional foundation, based on  $\delta^{13}\text{C}$  signatures that are consistent between *Osedax* tissues and the collagen-specific fraction in whale bones (Goffredi *et al.*, 2005). The worms are hypothesized to use acidifying vacuolar  $\text{H}^+$ -ATPases in their root-like tissue to bore into and breakdown the hard, calcified matrix of bones (Tresguerres *et al.*, 2013). Based on unsuccessful attempts at PCR amplification of genes required for autotrophic  $\text{CO}_2$  fixation (that is, ribulose-1, 5-bisphosphate carboxylase/oxygenase and ATP citrate lyase), the symbionts are hypothesized to be uniquely heterotrophic, rather than chemosynthetic, as in all other examined siboglinids (Goffredi *et al.*, 2005). The symbionts are likely responsible for at least some of the nutritional needs of the host, as evidenced by concordant  $\delta^{13}\text{C}$  compound-specific isotopic measures between symbiont-derived fatty acids and host-produced wax esters (Goffredi *et al.*, 2005, 2007). The specific nature of the symbiont metabolism, and any exchange of compounds, however, has yet to be elucidated.

All species of *Osedax* examined to date host related bacteria from the Oceanospirillales order of

Gammaproteobacteria (Goffredi *et al.*, 2005, 2007; Verna *et al.*, 2010; Salathé and Vrijenhoek, 2012). In this study, we examine the dominant symbiont 'ribospecies' Rs1 and Rs2 associated with *Osedax rubiplumus* and *Osedax frankpressi* living at 1820–2891 m depths in Monterey Canyon (Goffredi *et al.*, 2005; Salathé and Vrijenhoek, 2012). A previous study noted that the frequencies of Rs1 and Rs2 symbionts hosted by these worms varied as the bones decompose and that the observed temporal shifts are independent of host species, suggesting a direct influence of the local environment on symbiont populations (Salathé and Vrijenhoek, 2012). A genomic comparison of the two dominant symbiont types in *Osedax* worms offers the opportunity to elucidate details regarding their specific metabolism and to assess whether spatial or temporal heterogeneity, even within the deep sea, can influence biological diversity.

## Materials and methods

### Sample collection

Females of *Osedax frankpressi* were collected for genome analysis from a whale fall at 1891 m depth in Monterey Canyon (36.708 N/–122.105 W; dive DR236, on 5 June 2011; Table 5, Supplementary Table S1). Additional samples of two host species (*O. frankpressi* and *O. rubiplumus*) from two locations (1891 m and 2891 m depth; the latter of which is at 36.313 N/–122.434 W) were collected over the course of 112 months (between 2002 and 2011; Supplementary Table S1), for PCR-based analysis of both symbiont identity and functional gene characteristics, described below. Finally, *O. roseus* worms for experiments with  $^{13}\text{C}$ -labeled substrate were collected from a whale fall at 840 m depth off southern California (32.767 N/–117.483 W; dive DR471, on 18 May 2013).

### DNA extraction and evaluation of purity

For genomic analysis, host root tissues were loosely homogenized in ice-cold  $1\times$  phosphate-buffered saline using a Kontes ground glass homogenizer. Symbionts were enriched via selective filtration from homogenate (3  $\mu\text{m}$  pre-filter and final 0.22  $\mu\text{m}$  filter), and collected on a hydrophilic Durapore membrane (Millipore, Billerica, MA, USA). Purity of captured cells was evaluated via fluorescence microscopy (DAPI and FISH, according to Goffredi *et al.* (2005)), before proceeding to nucleic acid extraction. DNA extraction directly from filters was performed using the Qiagen DNeasy kit (Valencia, CA, USA), following the manufacturer's protocol, with the exception of an initial overnight lysis in 400  $\mu\text{l}$  of ATL buffer and proteinase K. Symbiont identity for each sample was determined via molecular analysis, described below, and two samples containing the Rs1 and Rs2 symbionts were

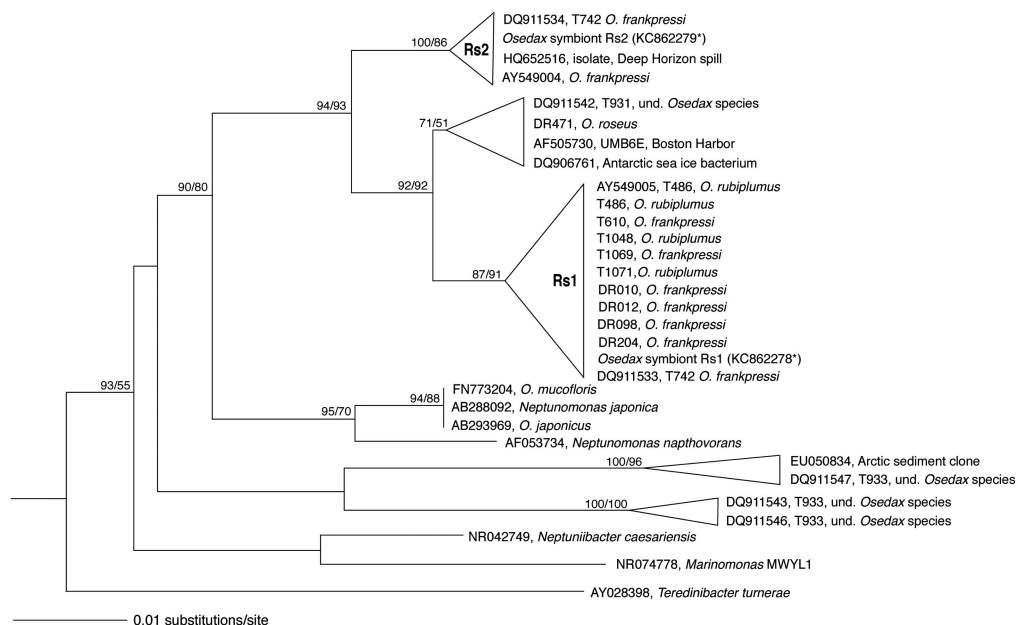
chosen for further genomic analysis. For simplification, we have maintained the ribospecies numbers (Rs1 and Rs2; Figure 1) associated with each symbiont based on the designation by Salathé and Vrijenhoek (2012), which correspond to the original 16S rRNA GenBank submissions AY549005 and AY549004, respectively (Goffredi *et al.*, 2005). The amount of extracted DNA, calculated using a Qubit fluorometer (Invitrogen) and Quant-iT PicoGreen (Invitrogen), was 150 and 50 ng for the Rs1 and Rs2 symbiont samples, respectively.

To ensure the relative symbiont purity in the nucleic acid extractions, we examined the microbial composition of the 16S ribosomal RNA gene (V1–V3 region) via pyrosequencing, according to Chun *et al.* (2010) (see Supplementary Methods for more details). Subsequent genomic analysis was performed on two samples that were determined to be relatively pure, one from each symbiont type, based on dominance of the target Oceanospirillales symbiont, which comprised 84% and 56% of the recovered 16S rRNA barcodes for symbionts Rs1 and Rs2, respectively (Supplementary Figure S1). The genomic DNA from both samples was subjected to whole-genome multiple displacement amplification (MDA) using the REPLI-g Mini kit (Qiagen) according to the manufacturer's recommendations. A UV irradiation method was employed to eliminate exogenous contaminant DNA found in MDA reagents (Woyke *et al.*, 2011) and 25 independent

amplification reactions were pooled for further analysis. Before genome sequencing, amplified DNA (~20 µg) was purified using the QIAamp DNA Mini kit (Qiagen) following the manufacturer's supplementary protocol.

#### Genome sequencing, assembly and partitioning

The *Osedax* symbiont Rs1 genome was sequenced using a combination of both Illumina and 454 FLX sequencing, from both a non-amplified and MDA-amplified nucleic acid extraction, described above. A 150-bp paired-end library was prepared from 8 ng of the non-amplified metagenomic DNA using the Nextera DNA sample preparation kit (Illumina, San Diego, CA, USA) and a total of 6 163 796 reads were obtained using an Illumina MiSeq instrument. Sequencing of the MDA-amplified Rs1 sample was performed from both an Illumina GAIIX 100 bp paired-end library (using the TruSeq DNA sample preparation kit) and from a 454 FLX Titanium 8-kb-span paired-end library (according to the Roche/454 protocol for titanium shotgun libraries), both from 5 µg of the MDA-amplified sample. A total of 203 574 paired reads were obtained from the 454 FLX run (average length = 356 bp), whereas 61 819 346 were obtained from the Illumina run. The *Osedax* symbiont Rs2 genome was sequenced from an Illumina GAIIX 100-bp paired-end shotgun library, prepared from 5 µg of the purified MDA



**Figure 1** Phylogenetic relationships among the *Osedax* symbionts, based on the 16S rRNA gene. Additional sequences were obtained from GenBank and aligned using the ARB automated alignment tool with subsequent manual refinements. Taxa in bold denote those discussed in detail in this paper. The bacterial 16S rRNA-encoding sequences in the metagenomic sequencing reads for *Osedax* symbiont Rs1 were extracted using rRNASelector program. Fragments that showed overlap with an rRNA (>60 bp; HMM profile with *E*-value lower than  $10^{-5}$ ) were identified as rRNA-coding regions. The 16S rRNA gene of the symbiont Rs2 sample was acquired via manual assembly of the 16S rRNA community sequence data, before metagenomic sequencing. For near full-length representatives and closest relatives, neighbor-joining analysis was conducted with Olsen distance correction. *Colwellia piezophila* (AB094412) was used as the outgroup. A more detailed phylogeny can be found in Salathé and Vrijenhoek (2012).

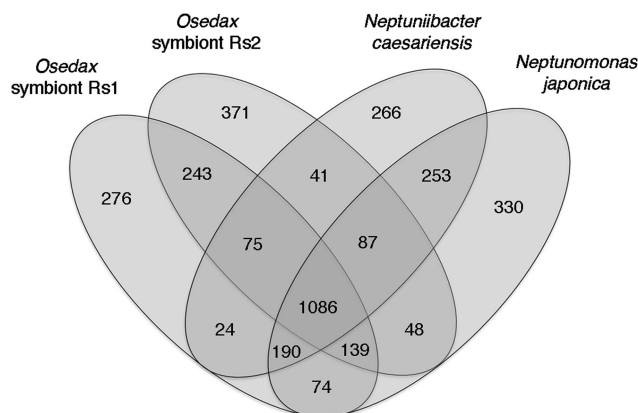
using the TruSeq DNA Sample Preparation kit (Illumina) and a total of 58 078 116 reads were obtained. The specific construction of shotgun libraries and Illumina/454 sequencing was performed at ChunLab, Inc. (Seoul, Korea).

Redundant high-coverage reads and sequencing errors from the MDA-amplified samples were removed via 'digital normalization' according to Brown *et al.* (2012). The resulting Illumina sequences for both genomes were assembled using CLC genomics wb5 (CLC Bio, Aarhus, Denmark), whereas the 454 reads of the Rs1 genome were assembled using the Newbler assembler 2.3 (Roche Diagnostics, Indianapolis, IN, USA). For the *Osedax* symbiont Rs1 genome, a hybrid assembly of both Illumina runs (non-amplified Nextera and MDA-amplified) and the 454 contigs was performed using CLC Genomics Workbench 5 and CodonCode Aligner (CodonCode Co, Centerville, MA, USA). The assembled contigs were subjected to *in silico* separation of the target genome from other possible metagenomic data sets using pentanucleotide content analysis as described previously (Woyke *et al.*, 2006; Kanehisa *et al.*, 2010; custom java code available at <http://sw.ezbiocloud.net>; Supplementary Figure S2 and Supplementary Results). In cases where a unique capability was attributed to the target symbiont genome, the partition to which it belonged was confirmed via recruitment and mapping of the contig against the bin, as well as the phylogenetic placement of surrounding contiguous genes.

Open-reading frames were predicted using Glimmer 3 (Delcher *et al.*, 2007) and the assembled contigs and scaffolds were uploaded into the RAST server (Aziz *et al.*, 2008). The predicted ORFs were annotated by searching against clusters of orthologous groups (COGs; Tatusov *et al.*, 1997) and SEED databases (Disz *et al.*, 2010). A comparative method was also employed, as described previously by Chun *et al.* (2009), in which target contigs with homology to a query ORF (>50% amino-acid sequence over 50% of the length of the gene) were identified using BLASTN. Nucleotide sequences of the query ORF, and selection of target homologous regions, were then aligned using a pairwise global alignment algorithm (Myers and Miller, 1988). Orthologs and paralogs were differentiated by reciprocal comparison. A comparison of ORFs present in both symbiont types, versus *Neptuniibacter caesariensis* and *Neptunomonas japonica*, was conducted using Venny (Figure 2; Oliveros, 2007).

#### Phylogenetic analysis—Average nucleotide identity of ORFs

A phylogenetic analysis, using highly conserved orthologous genes, compared the genomes of the *Osedax* symbionts Rs1 and Rs2 with 16 other Oceanospirillales genomes available in public databases. A set of 304 ORFs (107 966 amino acids) showing greater than 40% amino-acid sequence



**Figure 2** Common and unique genes in the *Osedax* symbionts Rs1 and Rs2, in comparison with the closest relatives *Neptuniibacter caesariensis* and *Neptunomonas japonica*, obtained by bidirectional best-hit analysis of all protein sequences in each genome. The Venn diagrams depict the overlap of those genes for which the function was assigned (that is, not hypothetical) and which were not duplicated. Unique genes may include some paralogs within each genome.

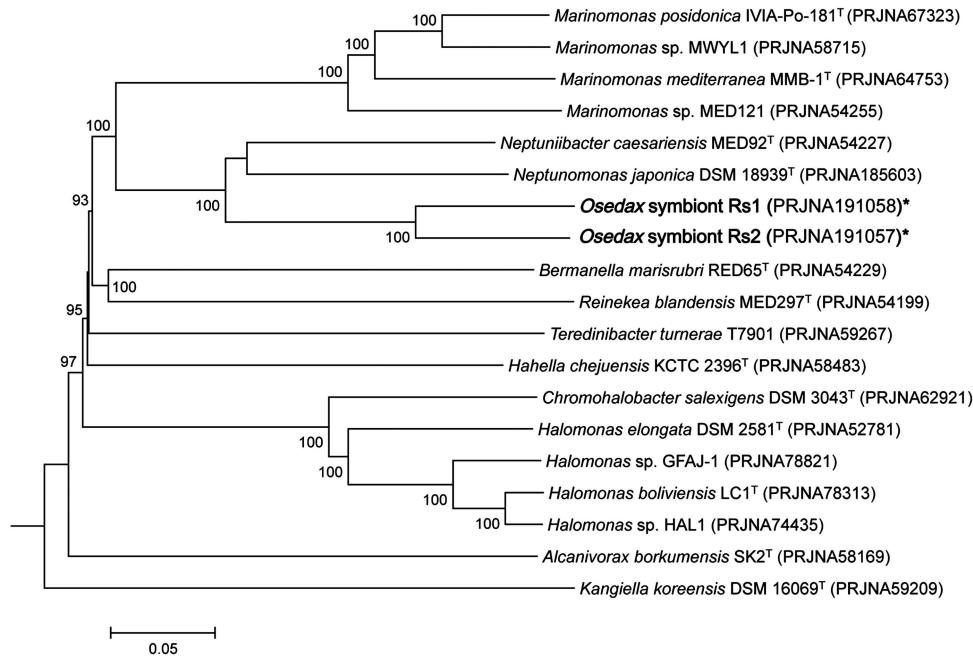
similarity to symbiont Rs1, over 70% of the length of the gene, was selected as highly conserved and aligned using CLUSTALX (Thompson *et al.*, 2002). Multiple alignments were concatenated and then used to construct a genome tree using the neighbor-joining method implemented in MEGA (Saitou and Nei, 1987; Kumar *et al.*, 2008). An evolutionary distance matrix for the neighbor-joining tree was generated according to the model of Jukes and Cantor (1969) and the resulting tree topology was evaluated by bootstrap analysis (Figure 3; Felsenstein, 1985). Evolutionary distance between the symbionts, and other relatives, was measured by the average nucleotide identity (ANI), computed by a BLAST algorithm, as previously described (Konstantinidis and Tiedje, 2005; Goris *et al.*, 2007).

#### Completeness estimation

Completeness of the genomes was estimated via the presence of 111 universal bacterial genes, identified by Dupont *et al.* (2012), using hmmer3 (Finn *et al.*, 2011), requiring 70% length match for each HMM, with the trusted cutoff as the minimum score (Supplementary Table S2). As many of the universal core genes are co-localized, including ribosomal and *sec* translocase genes, an additional estimate of completeness for symbiont Rs2 was assessed by comparison of ~169 lineage-specific uniformly spaced single-copy genes, identified in a minimum of 95% of gammaproteobacteria genomes in IMG release 4.0 (out of 418 genomes). The scripts for both estimates are available at [https://github.com/ Ecogenomics/CheckM](https://github.com/Ecogenomics/CheckM).

#### Diagnostic symbiont and host identification

Samples were extracted using the Qiagen DNeasy kit, according to manufacturer's instructions. Symbiont



**Figure 3** Genomic relatedness of the *Osedax* symbionts to 17 other bacteria within the order Oceanospirillales available in public databases (an ‘\*’ denotes genomes from this study). The comparison is based on sequence similarity among a set of 300 highly conserved orthologous genes, showing greater than 40% amino-acid sequence similarity to *Osedax* symbiont Rs1. Sequences were aligned using CLUSTALX and multiple alignments were concatenated and then used to construct a genome tree using the neighbor-joining method implemented in MEGA. An evolutionary distance matrix for the neighbor-joining tree was generated according to the model of Jukes and Cantor (1969) and the resulting tree topology was evaluated by bootstrap analysis (numbers at nodes indicate bootstrap support from 1000 re-sampled data sets). The scale bar represents 0.05 substitutions per site.

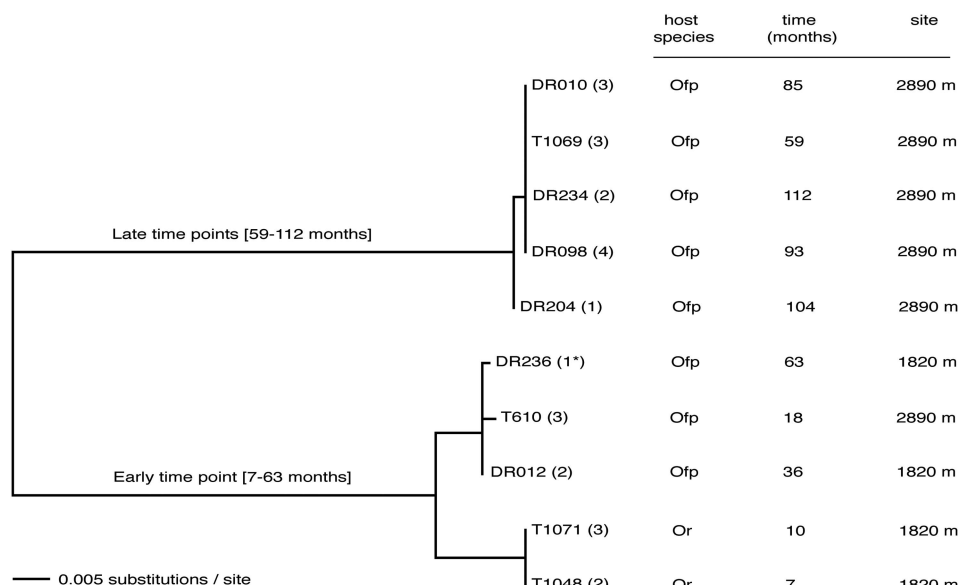
and host identities of these individual worms were confirmed by amplification and sequencing of the symbiont 16S rRNA and host 18S rRNA (or cytochrome *c* oxidase) genes. A low-cost alternative for determining the dominant symbiont strain within each worm involved the specific amplification of the symbiont 16S rRNA gene, using an Oceanospirillales-specific primer pair (Supplementary Table S3; Rouse *et al.*, 2011), followed by subsequent digestion with *EcoRI*, *HaeIII* and *AluII* (New England Biolabs, Ipswich, MA, USA), according to the manufacturer’s recommendation. The specific pattern indicative of each symbiont (Supplementary Figure S3) was confirmed via direct sequencing (Laragen, Culver City, CA, USA), according to Goffredi *et al.*, 2012. Sequences were assembled and edited using Sequencher v4.10.1 (GeneCodes Corp., Ann Arbor, MI, USA) and compiled in ARB, after initial alignment using the SILVA Aligner function (Ludwig *et al.*, 2004; Pruesse *et al.*, 2007). Additionally, bacterial 16S rRNA-encoding sequences in the subsequent metagenomic sequencing reads for *Osedax* symbiont Rs1 were extracted using rRNASelector program (Lee *et al.*, 2011). Fragments that showed overlap with an rRNA (>60 bp; HMM profile with E-value lower than  $10^{-5}$ ) were identified as rRNA-coding regions and extracted (Figure 1; GenBank acc. no. KC862278). The 16S rRNA gene of the symbiont Rs2 sample was acquired via manual assembly of the 16S rRNA community sequence data, before metagenomic sequencing (Figure 1, acc. no.

KC862279). Host identities were confirmed via amplification and sequencing of either COI or 18S rRNA gene sequences (see Supplementary Table S3) compared, via BLASTN, to *Osedax* host sequences available in GenBank.

The *soxB* gene (sulfate thioesterase/sulfate thiohydrolase) was amplified using previously published primers 432F/1446R (or 693F/1446R) and PCR conditions (Supplementary Table S3, Supplementary Figure S4), and sequenced using an Applied BioSystems (Foster City, CA, USA) 3500xL Genetic Analyzer. Additionally, a targeted gene amplification approach from worms specifically infected by symbiont Rs1 (both *O. rubiplumus* and *O. frankpressi*, was performed for two methyl-accepting chemotaxis genes (*mcp3*, *mcp4*), the histidine protein kinase *cheA* gene, the methyltransferase *cheR* gene, and the methyl-esterase gene *cheB*, using primers designed in this study and conditions shown in Supplementary Table S3. Sequences were assembled, edited and translated using Sequencher v4.10.1 (GeneCodes Corp.) and a concatenation, in the case of *mcp* and *che* genes, was compared via neighbor-joining analysis (Figure 4).

#### Nucleotide sequence accession numbers

The *Osedax* symbiont genomes were deposited in GenBank under accession number PRJNA191058 (symbiont Rs1) and PRJNA191057 (symbiont Rs2).



**Figure 4** Phylogenetic relationships among chemotaxis-related functional genes, based on amino-acid divergence, for *Osedax* symbiont Rs1 individuals with identical 16S rRNA genes, collected from two different species (*Osedax rubiplumus* (Or) and *O. frankpressi* (Ofp)), from two different sites, over 112 months. The midpoint-rooted neighbor-joining distance tree shown here was constructed from an alignment of five concatenated genes, including two methyl-accepting chemotaxis genes (*mcp3*, *mcp4*), the histidine protein kinase *cheA* gene, the methyltransferase *cheR* gene, and the methylesterase gene *cheB*. Two main clades cluster by early and late periods of time on the seafloor, with the early group showing a possible relationship to host species. Dive numbers begin with the remotely operated vehicle name; T = *Tiburion*, DR = *Doc Ricketts* (both owned and operated by the Monterey Bay Aquarium Research Institute). The number in parentheses indicates the number of worms analyzed from each dive. Genes amplified from the *Osedax* symbiont Rs1 corresponding to the genome sample was collected during dive DR236 (1\*).

GenBank accession numbers for *cheA*, *cheB*, *cheR* and *mcp2* gene sequences obtained in this study are KC814720–KC814787. The GenBank accession number for the *soxB* gene from symbiont Rs2 is KF444201. Sequences that were shorter than 200 bp length, yet still informative with regard to ecotype distinctions between symbiont populations, including *mcp3* and *mcp4*, can be found in Supplementary Figure S5. GenBank accession numbers for the 16S rRNA sequences corresponding to symbionts in this study are KC862278 (Rs1) and KC862279 (Rs2).

#### Carbon elemental analysis

On collection, worms for  $^{13}\text{C}$  tracer experiments were carefully removed from substrate and incubated, for 65 h at 4 °C on board ship in glass vials, in the presence of various substrates, of which ~20% was  $^{13}\text{C}$ -labeled (in 100 ml 0.2  $\mu\text{m}$ -filtered sea water; Table 5). Before analysis, specimens were divided into root and trunk tissues (symbiont- and non-symbiont containing, respectively) and were washed with 1  $\times$  phosphate-buffered solution (PBS) to remove residual-labeled salts. The uptake of  $^{13}\text{C}$  label was then analyzed via elemental analysis isotope ratio mass spectrometry. Inorganic carbon was dissolved by acidification with 85% phosphoric for 2 h at 50 °C. Samples (0.2–1.2 mg) were lyophilized and transferred to tin capsules (9  $\times$  5 mm).  $\delta^{13}\text{C}$  of bulk organic matter and weight percent total organic carbon were determined via continuous flow

(He; 100 ml min $^{-1}$ ) on a Costech Instruments Elemental Combustion System model 4010 (EA) by oxidation at 980 °C over chromium (III) oxide and silvered cobalt (II, III) oxide followed by reduction over elemental copper at 650 °C.  $\text{CO}_2$  was subsequently passed through a water trap and a 5-Å molecular sieve GC at 50 °C to separate  $\text{N}_2$  from  $\text{CO}_2$ .  $\text{CO}_2$  was diluted with helium in a ConFlo IV interface/open split before analysis.  $\delta^{13}\text{C}$  values were measured on a Thermo Scientific Delta V Plus irMS.  $\delta^{13}\text{C}$  and values were corrected for sample size dependency and then normalized to the VPDB scale with a two-point calibration and internal standards (Coplen *et al.*, 2006). Accuracy was  $\pm 0.09\%$  ( $n = 20$ ) and precision was  $\pm 0.02\%$  ( $n = 5$ ;  $1\sigma$ ), determined by analyzing independent standards as samples and precision was determined from international standards.

## Results and discussion

Here, we describe the genomes of the two dominant Oceanospirillales symbionts found in association with the deep-sea polychaete genus *Osedax*. Evidence from comparative analysis among free-living relatives and a related symbiont of marine bivalves, suggests metabolic versatility, strategies for intracellular survival, potential virulence capabilities and evidence for local niche adaptation between ribospecies of the free-living lifestyles of the *Osedax* symbionts (Rs1 and Rs2).

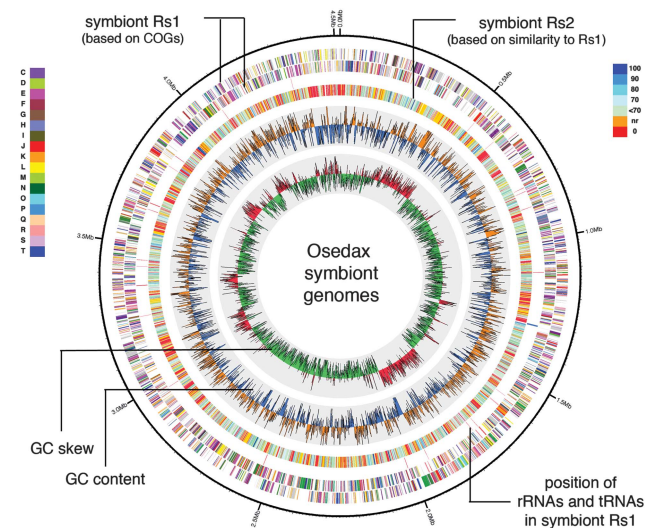
### General genomic features.

The *Osedax* symbiont Rs1 hybrid genome assembly resulted in 231 contigs (average length of 19.4 kb, range 1–180 kb), and 47 scaffolds ranging from 2 to 480 kb, with a total approximate length of 4 512 748 bp (Table 1; Figure 5). This genome was estimated to be ~94% complete with the singular occurrence of 104/111 core genes (Supplementary Table S2). For the *Osedax* symbiont Rs2 genome assembly, 411 contigs were acquired (average length of 12.8 kb, range 1–76 kb), with a total approximate length of 4 649 842 bp, and an estimate of 66% completeness (73/111 core genes; Table 1; Supplementary Table S2; Figure 5). As many of the genes that were missing from the symbiont Rs2 genome were ribosomal genes ( $n = 22$ ; Supplementary Table S2), which are typically co-localized on the genome, a further assessment of 169 lineage-specific uniformly-spaced core genes shared among gamma-proteobacteria revealed the symbiont Rs2 genome to be ~89% complete. Despite an initial contamination of this sample by *Cytophaga*-related sequences, the occurrence of single-copy core genes after binning (that is, *dnaK*, *gyrA*, *pgk*, *rpoC*, *secA*; Supplementary Table S2), suggests that the symbiont Rs2 genome contains information from the symbiont only. Both symbiont genomes are within the range of previous estimates in close relatives for genome size (3.9–5.2 Mb), as well as the number of coding genes (3500–4700; Table 1). Coding density for each of the symbiont genomes was ~80% and, after binning, 84% of the annotations in both genomes were most similar to those within the gammaproteobacteria (Supplementary Figure S1).

### Phylogenetic affiliation

A phylogenetic analysis, based on 304 highly conserved orthologous genes, was used to describe the genetic relatedness between the *Osedax* symbionts Rs1 and Rs2 to 17 other Oceanospirillales available in public databases (Figure 3). Based on

this lineage-specific analysis, the symbionts were most closely related to each other, and formed a monophyletic clade within the radiation of the order Oceanospirillales, consistent with single gene 16S rRNA analysis (Figure 3; Goffredi *et al.*, 2007; Salathé and Vrijenhoek, 2012). The symbionts shared an average genome nucleotide identity



**Figure 5** Circular representation of the genome of *Osedax* symbiont Rs1, and comparison with symbiont Rs2, generated using Circos 0.52 (Krzywinski *et al.*, 2009). The outer rings 1 and 2 represent the coding regions of symbiont Rs1, which are color coded according to the COG functional classes as designated in the inset at left and defined in Supplementary Table S5. Ring 3 shows the position of rRNAs and tRNAs in symbiont Rs1. Ring 4 represents the genome of symbiont Rs2, color coded according to homology with the genes of symbiont Rs1, depicted in the inset at right. Rings 5 and 6 are G+C content and G+C skew, respectively. The yellow and blue denote %G+C content (window size of 10 000 bp, step size 1000 bp), ranging from 30% to 53% with a step of 1%, which was higher and lower than the average G+C content of Rs1, respectively. The green and red denote G+C skew  $[(G - C)/(G + C)]$ ; (window size 10 000 bp, step size 1000 bp), ranging from -0.22 to 0.25 with a step of 0.01%, and values <0 and >0, respectively.

**Table 1** Comparison of the general genomic features between the *Osedax* symbionts, a closely related symbiont and two free-living members of the Oceanospirillales

Features	<i>Osedax</i> Rs1 symbiont	<i>Osedax</i> Rs2 symbiont	<i>Teredinibacter turnerae</i> symbiont	<i>Neptuniibacter caesariensis</i>	<i>Neptunomonas japonica</i> <sup>a</sup>
Size, total (Mb)	4.51	4.65	5.19	3.87	4.39
%G+C	43.4	44.3	50.9	46.6	43.7
No. of contigs	231	411	1	6	42
No. of predicted CDS <sup>b</sup>	4244	4421	4690	3641	4265
No. of function assigned	2888	3107	3067	2479	2558
No. of hypothetical proteins <sup>c</sup>	1356	1314	1615	1162	1707
No. of RNAs	68	11 <sup>d</sup>	49	46	1

<sup>a</sup>*Neptunomonas japonica* DSM 18939.

<sup>b</sup>Character statistics for both symbiont genomes were similar in RAST and IMG/M.

<sup>c</sup>Includes both conserved hypothetical proteins (inferred homology to hypothetical proteins encoded by other genomes) and those with no inferred homology to any previously identified proteins.

<sup>d</sup>This value reflects the incomplete nature of the symbiont Rs2 genome, in that it is missing many of the co-localized RNA subunit genes (Supplementary Table S2).



(ANI) value of 73%, which clearly suggests they belong to different species (Konstantinidis and Tiedje, 2005; Goris *et al.*, 2007). *Neptuniibacter caesariensis*, isolated from the Mediterranean Sea, and *Neptunomonas japonica*, isolated from sediments near a sperm whale carcass off Japan, were the closest sequenced relatives based on genomic data available (Arahal *et al.*, 2007; Miyazaki *et al.*, 2008), and both shared ~67% ANI value with the *Osedax* symbiont genomes. For this reason, many traits below are compared between the *Osedax* symbionts and these related bacteria.

#### Evidence for a free-living lifestage

The genomic characteristics of intracellular symbionts are greatly affected by the mode of transmission between host generations. In general, large genomes are thought to be more ecologically versatile (Konstantinidis and Tiedje, 2004), which would thus be important for an environmentally acquired symbiont that has both free-living and host-associated stages. The estimated genome sizes of the *Osedax* symbionts, larger than free-living relatives, and similar to the facultative intracellular endosymbiont *Teredinibacter turnerae* of the marine shipworm bivalve (Table 1; Yang *et al.*, 2009), indirectly suggests a possible retention of genes for both free-living and intracellular existence. Vestimentiferan tubeworms, siboglinid relatives of *Osedax*, are infected each generation with symbionts from the local environment in which the worm larvae settle (Nussbaumer *et al.*, 2006). It is not known definitively whether *Osedax* acquire symbionts in a similar manner, however, molecular and microscopic analysis did not detect symbionts in either eggs or early larval stages (Rouse *et al.*, 2008), supporting this assumption.

Both *Osedax* symbionts possess the full complement of genes necessary for flagella formation, including *fla*, *fle*, *flg*, *flh* and *fli* gene clusters, indicating the capability for movement. Similarly, the *Osedax* symbionts have a complex chemosensory system with >20 genes encoding for putative methyl-accepting chemotactic sensory genes (*mcp*s) involved in extracellular signal recognition and initiation of chemotaxis in bacteria (Wadhams and Armitage, 2004). The presence of genes responsible for motility-conferring flagella and sensing of the environment supports the assumption of a free-living stage for the *Osedax* symbionts. The large number of *mcp* genes identified in the symbionts may reflect complex associations with other organisms, as observed for many pathogens and some symbionts (~20–40 *mcp* genes is common; Dziejman *et al.*, 2002; Robidart *et al.*, 2008). However, the free-living relative *Neptunomonas japonica* also possesses a large number of *mcp* copies ( $n=17$ ), thus it is more likely that this reflects an increased need for detection and response to dissolved organic matter at abyssal

depths, as has been observed in the genomes of many deep-sea bacterial species (Lauro and Bartlett, 2008).

#### Metabolic flexibility

Lack of CO<sub>2</sub> fixation genes supports early evidence of the heterotrophic nature of the *Osedax* symbionts and confirms their unusual divergence from the typical chemosynthetic lifestyle of all other related siboglinids (Naganuma *et al.*, 2005; Thornhill *et al.*, 2008; Hilário *et al.*, 2011). Consistent with possible exposure to variable chemical regimes, and heterotrophy, the transport profile of the *Osedax* symbionts indicates flexibility in the types of compounds brought into the cell. The genome for both symbionts is devoted to ATP-binding cassette (ABC) and TRAP-like transport genes with predicted specificity for various carbohydrates, amino acids and peptides, suggesting a broad substrate spectrum for both carbon and nitrogen gain (Table 2, Supplementary Table S4). Membrane transport, as a category, comprised 10–12% of the total number of annotated genes in the *Osedax* symbiont genomes, compared with 8% in both *N. caesariensis* and *N. japonica* (Supplementary Table S4). Notably, ~30 genes involved in ABC transport of oligo- and dipeptides were observed in the symbiont genomes, compared with <10 in both *N. caesariensis* and *N. japonica*. The symbionts have the potential to metabolize monosaccharides, sugar alcohols, fatty acids, organic acids, and in the case of symbiont Rs2, disaccharides and amino sugars (Tables 2–4). This versatility in the *Osedax* symbionts likely reflects the diversity of compounds readily available in the whale fall environment, as well as those available intracellularly. Overall, the genomic potential for monosaccharide uptake by the *Osedax* symbionts, in particular, is in stark contrast to both *N. caesariensis* and *N. japonica*, which instead have greater numbers of genes dedicated to fermentation (Table 4; Arahal *et al.*, 2007; Miyazaki *et al.*, 2008). Apparent uptake of <sup>13</sup>C-labeled glucose supports the utilization of monosaccharides by *Osedax* (Table 5), however further analyses with larger sample sizes is necessary to confirm the importance of specific substrates in their overall metabolism.

Collagen is thought to be the source of nutrition for the *Osedax* symbiosis, based on consistent  $\delta^{13}\text{C}$  signatures between *Osedax* tissues and the collagen-specific fraction found in bone (Goffredi *et al.*, 2005). A potential for collagen breakdown by the *Osedax* symbiosis was established by the presence of collagenase in the root tissues of individual worms (Goffredi *et al.*, 2007), although at the time it was not possible to determine whether the symbiont or host was the source of the collagenase activity. In the present study, collagenase genes were not identified in the *Osedax* symbiont genomes and instead, genes with homology to glycine

**Table 2** The transport profile of the *Osedax* symbionts, in comparison with the closest relative *Neptunomonas japonica*

Transporter type—substrate	Symbiont Rs1	Symbiont Rs2	<i>Neptunomonas japonica</i>
<i>ABC-transporters amino acids</i>			
Arginine/ornithine	<i>artIMQ</i>	<i>aotPQ<sup>a</sup></i>	—
Branched chain	<i>livFGHKM<sup>b</sup></i>	<i>livFGHKM<sup>b</sup></i>	<i>livFGHKM</i>
Histidine	<i>hisJMPQ</i>	<i>hisJMPQ<sup>b</sup></i>	<i>hisMP</i>
Glutamate/aspartate	<i>gltIJKL</i>	<i>gltIJKL</i>	<i>gltIJK</i>
L-Proline/glycine betaine	<i>proXWV<sup>b</sup></i>	<i>proXWV<sup>b</sup></i>	<i>proXWV</i>
Amino acids (general)	16	21	7
<i>Derivatives/peptides</i>			
Spermidine/Putrescine	<i>potABCD<sup>b</sup></i>	<i>potABCD<sup>b</sup></i>	<i>potABCD</i>
Putrescine	<i>potFGHI<sup>b</sup></i>	—	<i>potFGHI</i>
Urea	<i>urtCDE<sup>a,b</sup></i>	<i>urtDE<sup>a</sup></i>	—
Oligopeptides	<i>oppABCDF<sup>b</sup></i>	<i>oppABCDF<sup>b</sup></i>	<i>oppBC<sup>c</sup></i>
Dipeptides	<i>dppBCDF<sup>b</sup></i>	<i>dppBCD<sup>b,c</sup></i>	<i>dppCDF</i>
<i>Carbohydrates</i>			
N-Acetyl-D-glucosamine	—	<i>ugpABE<sup>a,b</sup></i>	—
Glycerol-3-phosphate	<i>ugpABE<sup>a,b</sup></i>	<i>ugpABCE<sup>a,b</sup></i>	—
α-Glucoside transport	—	<i>aglEFGK<sup>b</sup></i>	—
Rhamnose	<i>ddpABC<sup>b,c</sup></i>	<i>ddpABCD<sup>b</sup></i>	—
Ribose	<i>rbsBC<sup>c</sup></i>	<i>rbsABCD<sup>b</sup></i>	—
Sugars (general)	6	13	—
<i>Vitamins/cofactors</i>			
Thiamin	<i>tbpA/thiPQ<sup>b</sup></i>	<i>tbpA/thiPQ<sup>b</sup></i>	<i>tbpA/thiPQ</i>
Vitamin B12	<i>btuCDF</i>	—	—
Hydroxymethylpyrimidine	<i>tauABC<sup>b</sup></i>	<i>tauABC<sup>b</sup></i>	—
Manganese	<i>sitABCD<sup>b</sup></i>	<i>sitABCD<sup>b</sup></i>	—
Molybdenum	<i>modABC<sup>b</sup></i>	<i>modABC<sup>b</sup></i>	<i>modABC</i>
Nickel	<i>nikKMOQ</i>	<i>nikKMOQ</i>	—
Nitrate	—	—	<i>ABC<sup>d</sup></i>
Phosphate	<i>pstABCS<sup>b</sup></i>	<i>pstABC<sup>b</sup></i>	<i>pstABCS</i>
Phosphonate	<i>phnCDE<sup>b</sup></i>	<i>phnCDE</i>	—
Tungstate	—	—	<i>ABC<sup>d</sup></i>
Zinc	<i>znuABC<sup>b</sup></i>	<i>znuABC<sup>b</sup></i>	<i>znuABC</i>
<i>ABC-transporters iron</i>			
Ferric iron	<i>afuABC<sup>b</sup></i>	<i>afuABC<sup>b</sup></i>	<i>afuABC</i>
Heme	<i>htsABC<sup>b,e</sup></i>	<i>htsABC<sup>b</sup></i>	—
Hemin	<i>hmuS/fepCD<sup>b</sup></i>	—	<i>hmuS/fepCD</i>
Iron-siderophore	—	<i>ceuABC<sup>b</sup></i>	—
<i>Other</i>			
Protease export	—	<i>PrtDE<sup>c</sup></i>	—
Multidrug system	4	5	8
General – no category	15	17	18
Single gene hits	5	10	4
Cell wall and capsule	<i>yrbCDEF<sup>b</sup></i>	<i>yrbCDEF<sup>b</sup></i>	<i>yrbCDEF</i>
Nucleosides	4	4	1
Capsular polysaccharide	<i>kpsCEMST</i>	—	<i>kpsCEMST</i>
Miscellaneous	<i>ynjBCD<sup>b</sup></i>	<i>ynjBCD<sup>b</sup></i>	—
<i>TRAP transporters</i>			
TAXI- dicarboxylates	TAXI	TAXI <sup>b</sup>	TAXI
DctP- dicarboxylates	<i>dctPQM</i>	<i>dctPQM<sup>b</sup></i>	<i>dctPQM</i>
<i>Other transporters</i>			
TTT- tricarboxylates	<i>tctABC<sup>b</sup></i>	<i>tctABC<sup>b</sup></i>	<i>tctABC</i>
<i>Permeases (non-ABC)</i>			
Acetate ( <i>actP</i> )	Yes	—	Yes
Drug/metabolite (DMT)	Yes	Yes	Yes
Drug resistance ( <i>bcr</i> )	—	—	Yes
Lipid A ( <i>msbA</i> )	Yes	Yes	Yes
L-Lysine	Yes	Yes	—
Major facilitator	Yes	Yes	Yes
Maltose ( <i>malFG</i> )	—	Yes	—
Muropeptide ( <i>ampG</i> )	Yes	Yes	Yes
Taurine ( <i>tauABC</i> )	—	—	Yes

Table 2 (Continued)

Transporter type—substrate	Symbiont Rs1	Symbiont Rs2	<i>Neptunomonas japonica</i>
Sulfate	Yes	Yes	Yes
Uracil-xanthine	Yes	—	Yes
Undefined/other	4	7	6

—, not identified or missing two out of three genes necessary for ABC cassette.

<sup>a</sup>Includes a substrate-binding protein-coding gene (no abbreviation given).

<sup>b</sup>Indicates genes found in a contiguous operon in the *Osedax* symbionts.

<sup>c</sup>Indicates that the ABC cassette is missing only one gene.

<sup>d</sup>Indicates a complete cassette, with no abbreviation given.

<sup>e</sup>Also annotated as a ferrichrome transporter in SEED.

oxidase, proline iminopeptidase, 4'-hydroxyproline dipeptidase were identified in both symbionts. In live animal tracer experiments, <sup>13</sup>C from labeled glycine (with proline amended), was taken up by the worms within 65 h of exposure (Table 5). It now seems likely, then, that collagen-derived amino acids (for example, glycine, proline and hydroxyproline) or peptides are liberated by the host, and used by the symbionts. Genes involved in proline catabolism were also identified in symbiont Rs2 by an indirect screen of bacterial mRNA transcripts from the root tissue of *Osedax frankpressi*, suggesting a role in the intact symbiosis (G. Rouse, SIO, personal communication). Additionally, in the present study, numerous copies of genes homologous to those that encode for ABC transport of both oligopeptides (OppABCDF) and dipeptides (DppBCDF; Monnet, 2003) were found in the *Osedax* symbionts, arranged in 5–6 kb operon-like structures. As a symbiont, peptidolytic activity would be important if, as mentioned, the host worm provides small peptides derived from the dissolution of the collagenous bone matrix.

There appears to be similar versatility with regard to phosphate and iron acquisition by both *Osedax* symbionts. Phosphate uptake genes identified in the *Osedax* symbionts included ABC-transporters for phosphate, phosphonate and glycerol-3-phosphate (glycerol-P). By comparison, only 3 of the 17 Oceanospirillales relatives (Figure 3) possessed all three mechanisms. Iron uptake, which is discussed in more detail later, is also potentially accomplished via three separate mechanisms in each symbiont, including ABC transport of ferric iron and heme by both symbionts, as well as heme uptake by the Rs1 symbiont and siderophore-mediated iron uptake in symbiont Rs2 (Table 3). All of these genes were identified along operons (3–8 kb), on large contigs (11–90 kb), and included the presence of the *Fur* gene, which mediates the transcriptional regulation, via its protein product, of iron-related genes (Litwin and Calderwood, 1993). Similarly, only 2 of 17 close relatives examined possessed all three distinct mechanisms for iron uptake. This flexibility in the *Osedax* symbionts for the solubilization and acquisition of iron may reflect adaptation to both the deep-sea water column and host cytosol.

#### Genomic implications for intracellular life

Heterotrophic intracellular symbionts must adapt their metabolism to the available nutrients and physical conditions within the cell or in vacuoles, as is the case for the *Osedax* symbionts. A dependence on cytosolic carbon and other nutrients requires special physiological adaptations that may not be observed in closely related free-living relatives. In general, the mechanisms of nutrient uptake by intracellular bacteria are not well understood and for the *Osedax* symbionts is a situation much more analogous to that of intracellular pathogens, rather than either chemo- or photoautotrophic endosymbionts.

As mentioned, the *Osedax* symbiont genomes reveal the potential for numerous methods of iron mobilization and transport, some of which are closely linked to virulence in other bacteria. For example, genes that encode large surface exoproteins, homologous to hemolysins, were identified in both *Osedax* symbionts. The acquisition of iron from host cells, most of which is bound, is accomplished by many intracellular bacteria by the production of hemolysins (Kuhn *et al.*, 1988; Portnoy *et al.*, 1988). The direct use of heme and/or hemin has also been observed in many pathogens and for *Yersinia*, *Neisseria* and *Vibrio* species, is the sole strategy for iron acquisition (Perry and Brubaker, 1979; Mickelsen and Sparling, 1981; Stoebner and Payne, 1988). Not only do both *Osedax* symbionts possess ABC-type heme transporters, but the Rs1 symbiont specifically carries the *hmu* gene cluster, which is essential for utilization of free hemin and heme-protein complexes by *Yersinia pestis* (Hornung *et al.*, 1996). Likewise, siderophores, for which coding genes were identified in the *Osedax* symbiont Rs2 genome, can also effectively scavenge iron from host iron-binding compounds. Interestingly, the siderophore transport genes observed in the Rs2 symbiont were similar (45–55% by amino-acid content) to those encoding the citrate-containing siderophores petrobactin, anguibactin and vibriobactin found in many pathogenic bacteria (Griffiths *et al.*, 1984; Actis *et al.*, 1986; Koppisch *et al.*, 2005). Iron scavenging has been shown to be critically important to successful virulence and may confer the same advantage for accessing iron in intracellular beneficial endosymbionts as it does

**Table 3** Select metabolic capabilities attributed to the *Osedax* symbionts Rs1 and Rs2, or both

Unique to <i>Osedax</i> symbiont Rs1	Common to both <i>Osedax</i> symbionts	Unique to <i>Osedax</i> symbiont Rs2
	<i>Metabolic versatility</i>	
<b>Glycolate dehydrogenase/utilization</b> ( <i>glcCDEF</i> )	Glyoxylate Bypass ( <i>aceBC</i> )	<b>N-Acetylglucosamine ABC transport/ regulation</b>
Glutamate decarboxylase	Methylcitrate cycle ( <i>prpBCD</i> )	Ribose ABC transport ( <i>rbsABCD</i> ) <sup>a</sup>
<b>Glycerate kinase</b>	Oligopeptide ABC transport ( <i>oppABCDF</i> )	$\alpha$ -glucosidase/ $\alpha$ -glucoside transport ( <i>aglAEFGK</i> )
	<b>Dipeptide ABC transport (<i>dppBCDF</i>)</b>	Maltose/maltodextrin ABC transport ( <i>malFG</i> )
	Glycerol-3-phosphate ABC transport/kinase	
	<b>Rhamnose ABC transport</b>	
	<b>4'-Hydroxyproline dipeptidase</b>	
	Proline iminopeptidase	
	<b>Creatinase</b>	
	Spermidine-Putrescine ABC transport ( <i>potABCD</i> )	
	Tripartite tricarboxylate transport (TTT)	
	TRAP dicarboxylate transport - substrate 6	
	<i>Iron/sulfur/nitrogen/phosphate/cofactors</i>	
Vitamin B12 ABC transport ( <i>btuCDF</i> )	Ferric iron ABC transport	<b>Iron-Siderophore (vibrobactin-like)</b>
Hemin ABC transport ( <i>hmuS</i> )	<b>Heme ABC transport<sup>b</sup></b>	<b>Sulfide oxidation (<i>soxABDXYZ</i>) + Mo cofactor</b>
<i>tonB</i> -dependent hemin receptor	<b>Nitrous oxide reductase proteins</b> ( <i>nosDFLYZ + R</i> )	
<b>Nitrate reductase (<i>napD</i>)</b>	Glycerol-3-phosphate ABC transport	
	Phosphate ABC transport	
	<b>Phosphonate ABC transport<sup>c</sup></b>	
	<b>Nickel ABC transport (<i>nikKLMQ</i>)<sup>d</sup></b>	
	<i>Virulence/outer membrane structures</i>	
<b>Type-IVB Flp pilus (<i>cpaBF/tadABC</i>)</b>	<b>Internalins (<i>inIA</i>)</b>	
	Hemolysins (other large exoproteins)	
	<b>RTX toxins (T1SS-secreted agglutinins)</b>	
	Mannose-sensitive hemagglutinin (MSHA)-like pilus	
	Flagella formation	
	Methyl-accepting chemotaxis proteins	
	Type IV pilus assembly ( <i>pil</i> genes)	
	Type VI pilus assembly ( <i>imp</i> genes)	
	<i>Cell wall/capsule</i>	
<b>Capsular polysaccharide biosynthesis</b> ( <i>wcbQR</i> ) <sup>e</sup>	Capsular polysaccharides ( <i>cpsBCD</i> )	
<b>Capsular polysaccharide export</b> ( <i>kpsCDEMST</i> ) <sup>e</sup>	<b>O-antigen flippase <i>wzx</i></b>	
O-antigen ligase		

Capabilities in bold were not observed in more than three closely related genera ( $n = 10$  examined; Figure 3), and thus appear less common to the Oceanospirillales family.

<sup>a</sup>This gene cluster is listed as also being involved in rhamnose, or other sugar transport.

<sup>b</sup>A single gene, *htsA*, was observed in more than four close relatives.

<sup>c</sup>A single gene, *phnE*, was observed in five close relatives.

<sup>d</sup>The COG function of this gene cluster is annotated as a cobalt transporter.

<sup>e</sup>Gene clusters for capsule formation were identified in all four *Halomonas* species, as well as in 1–2 other close relatives.

in pathogens. The Candidatus Poribacteria symbiont in sponges has the genomic potential for both ferric iron and heme uptake (Siegl *et al.*, 2010), Candidatus *Endoriftia persephone*, the symbiont in tubeworms, can uptake both ferrous and ferric iron (Robidart *et al.*, 2008), whereas *Teredinibacter turnerae*, the symbiont of wood-feeding bivalves, appears to possess only a single hemin-based mechanism for iron uptake (Yang *et al.*, 2009). Until additional data is available for other beneficial

symbionts, it remains unknown whether multiple modes of iron acquisition are an adaptation broadly employed by symbionts.

In addition, genes that bypass part of the TCA cycle, including isocitrate lyase (*aceA*), malate synthase (*aceB*) and methylisocitrate lyase (*prpB*) were identified in the *Osedax* symbionts. Although observed widely in bacteria, the glyoxylate and methylcitrate cycles have been attributed to enhancement of cytosolic residence in intracellular

**Table 4** Comparative analysis of genes involved in carbohydrate metabolism, ranked according to percentage of the subsystem in the *Osedax* symbiont Rs1

Functional category <sup>a</sup>	Osedax symbiont Rs1 <sup>b</sup>		Osedax symbiont Rs2		Neptuniibacter caesariensis		Neptunomonas japonica	
	Gene no.	%	Gene no.	%	Gene no.	%	Gene no.	%
Central metabolism <sup>c</sup>	116	49.2	129	43.6	103	61.7	126	48.6
Fermentation	32	13.6	41	13.9	40	24.0	62	23.9
C1 compounds	31	13.1	40	13.5	5	3.0	51	19.7
Monosaccharides	24	10.2	26	8.8	1	0.6	1	0.4
Organic acids	18	7.6	30	10.1	12	7.2	10	3.9
Sugar alcohols	11	4.7	18	6.1	4	2.4	5	1.9
Amino sugars	2	0.8	9	3.0	0	0.0	0	0.0
Other <sup>d</sup>	2	0.8	3	1.0	2	1.2	4	1.5
Total	236		296		167		260	

<sup>a</sup>Based on SEED subsystems within carbohydrate metabolism.<sup>b</sup>Gene copy numbers based on the Nextera library construction are expected to reflect a relative dedication of the genome to each gene category.<sup>c</sup>This category includes glycolysis/gluconeogenesis, the tricarboxylic acid cycle and pentose phosphate pathway.<sup>d</sup>This category includes di- and oligosaccharides and carbon storage.**Table 5**  $\delta^{13}\text{C}$  values for *Osedax* trunk and root tissue after exposure of whole animals to  $^{13}\text{C}$ -labeled substrates for ~ 65 h at 4 °C on board ship

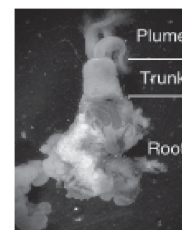
Substrate <sup>a</sup>	$\delta^{13}\text{C}$ (%)	
	Trunk	Root <sup>b</sup>
Unlabeled	- 14.1 <sup>b</sup>	- 14.1 <sup>c</sup>
$^{13}\text{C}$ -Glycine (100 mM, 25% label) + Proline (26 mM)	723.0	830.0
$^{13}\text{C}$ -Glucose (47 mM, 18% label)	460.6	778.4
$^{13}\text{C}$ -Ribose (17 mM, 25% label)	14.6	43.4

<sup>a</sup>Substrate concentrations were different between treatments.<sup>b</sup>The worms collected for this experiment harbored symbionts most closely related to Rs1, shown in Figure 1 (dive DR471).<sup>c</sup>Combined root and trunk ( $n = 4$ ). A previous study revealed trunk and root tissue to have similar  $\delta^{13}\text{C}$  values within 1.0% (Goffredi et al., 2005).

pathogens including *Mycobacterium* and *Salmonella*, presumably by allowing for rapid and efficient growth on host-produced odd-chain-length fatty acids (McKinney et al., 2000; Fang et al., 2005; Muñoz-Elías and McKinney 2005; Gould et al., 2006; Eisenreich et al., 2010). The *Osedax* symbionts possess many genes devoted to the specific degradation of branched-chain amino acids, which also results in the generation of propionyl-CoA, rather than acetyl-CoA (Muñoz-Elías et al., 2006), thus increasing the necessity for the methylcitrate cycle. The observation of genes for these two TCA shunts in both *Osedax* symbionts, as well as in the marine bivalve symbiont, *Ruthia magnifica* (Newton et al., 2007), suggests that they, too, may be important for intracellular life. Again, however, a more thorough comparison between heterotrophic symbionts and pathogens is critical to our understanding of the relative importance of these pathways for symbiosis.

Finally, although the *Osedax* symbionts presumably form beneficial associations, they possess genes known to be involved in attachment and colonization of eukaryotic cells. 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 (Supplementary Results; Craig and Li, 2008). Similarly, they possess *imp* genes that encode Type VI pili formation, as well as the *clpB* and *vrgG* genes observed to be necessary for host cell interactions and virulence in many pathogens (Table 3; Supplementary Results; Schlieker et al., 2005; Schell et al., 2007; Schwarz et al., 2010). Neither of the closest free-living relatives, *Neptuniibacter caesariensis* and *Neptunomonas japonica*, possess Type VI secretion systems. Genes encoding a mannose-sensitive hemagglutinin (MSHA)-like Type-IVa pilus were also identified in both symbionts and the *Osedax* Rs1 symbiont uniquely possesses a Type-IVb tight adherence (*tad*) pilus-encoding gene cluster. These pili function in surface attachment, biofilm formation and colonization and are considered a key virulence factor in *Vibrio cholerae* and other pathogenic bacteria (Chiavelli et al., 2001; Dalisay et al., 2006; Tomich et al., 2007; O'Connell Motherway et al., 2011). Both symbionts possess genes with homology



to the internalin A (*inlA*) of *Listeria monocytogenes* and *Clostridia perfringens*, which functions in adhesion and movement into host cells (Gaillard *et al.*, 1991; Lecuit *et al.*, 1997). Similarly, large genes were identified in both symbionts with homology to those that encode RTX-like toxins, which aid in membrane disruption and host cell invasion by pathogenic and symbiotic bacteria (Fullner *et al.*, 2002; Jeong *et al.*, 2005; Ruby *et al.*, 2005). Both *T. teredinibacter* and *Photorhabdus luminescens*, the symbiont of entomopathogenic nematodes, possess many similar secretion systems, adhesins and toxins, suggesting a role in mediating both adverse and beneficial interactions with eukaryotes (Ffrench-Constant *et al.*, 2000; Duchaud *et al.*, 2003; Yang *et al.*, 2009). The *Osedax* symbionts are not only suspected to infect nascent *Osedax* larvae, at a time when their immune system and developmental program might be more conducive to infection, but also to repeatedly infect adults, through their root tips, over time (Verna *et al.*, 2010). The latter scenario requires the symbionts to possess sufficient colonization mechanisms for adhesion and entry into the worms. The observed Type IV and VI secretion systems and putative virulence factors, such as internalins and toxins, may aid in this initial and repeated encounter and infection by the *Osedax* symbionts.

The intracellular symbionts in *Osedax* likely provide an essential function that enables the unusual ecology of these worms, yet we did not identify obvious genomic features that uniquely identify them as symbionts. The symbionts appear capable of *de novo* synthesis of most B vitamins, including the very long vitamin B12 synthesis pathway, as well as a capability for vitamin B12 transport (*btuCDF* genes identified in the *Osedax* symbiont Rs1 and the *btuC* gene in the Rs2 symbiont). It, thus, makes it tempting to speculate that vitamin B12 could be made and transported to the host, which is presumably incapable of its own B12 synthesis, as are all animals (Warren *et al.*, 2002). Similarly, the symbionts are capable of *de novo* synthesis of all amino acids, except asparagine and tyrosine. The intriguing presence of a large number of genes annotated as 'lysine exporter protein LysE/YggA', involved in the export of several essential amino acids, and an apparent absence of these genes in the genomes of fully sequenced relatives, suggests that they may be specifically relevant for the *Osedax* symbiosis. Presently, however, the specific role of the symbionts remains to be definitively elucidated and requires further examination using culture-dependent and -independent techniques.

#### *Environmental niche adaptation by symbiont populations*

A whale fall instantaneously creates a novel habitat that progressively changes as opportunistic

organisms exploit the massive influx of nutrients to the seafloor. These spatially heterogeneous and enriched habitats persist for varying timespans of 5 to 50 years, depending on a number of conditions (for example, depth and oxygen concentrations; Smith and Baco, 2003; Goffredi *et al.*, 2004; Braby *et al.*, 2007; Lundsten *et al.*, 2010). Many organisms, both animals and free-living microbes, tend to be stratified in space and time around the decomposing carcasses (Smith *et al.*, 2002; Goffredi *et al.*, 2004; Braby *et al.*, 2007; Goffredi *et al.*, 2008; Vrijenhoek *et al.*, 2009; Goffredi and Orphan, 2010; Lundsten *et al.*, 2010). Temporal variation also exists in relative abundances of the dominant symbiont types that infect *Osedax* (Goffredi *et al.*, 2007; Verna *et al.*, 2010). Disproportionately greater numbers of ribospecies Rs2 occur at later points in time (Salathé and Vrijenhoek, 2012). This shift occurs independently of shifts in host species composition, resulting perhaps from local adaptations to temporal changes in environmental gradients encountered by the pre-infectious 'free-living' bacteria (Verna *et al.*, 2010; Salathé and Vrijenhoek, 2012). Intra-symbiont temporal variation in this study was also indicated by genes involved in the chemotaxis pathway sequenced from *Osedax* which housed only symbiont Rs1 with identical 16S rRNA sequences, collected from 26 different worms from two sites, over a 112-month period. These genes, including *mcp3*, *mcp4*, *cheA*, *cheB* and *cheR*, differed among specimens by as much as 20% amino-acid content and clustered according to the duration of time the whale carcass was estimated to be on the seafloor (7–63 months and 59–112 months, respectively; Figure 4). It is currently not clear whether this pattern is solely derived from temporal variation of the environmental symbiont community or if there is also a certain amount of species specificity of symbiont strains involved.

Genomic differences in the metabolic and biochemical potential of Rs1 and Rs2 are similarly consistent with the hypothesis that the environment controls the symbiont populations at a given time, and that hosts primarily acquire the locally-adapted symbiont. For example, Rs2 uniquely carries genes for sulfide oxidation (*soxABDXYZ* and *molybdopterin C*), some in a contiguous 4.2-kb operon. *sox* gene sequences were similar (47–70% amino-acid content) to genes from *Neptuniibacter caesariensis*, and many other sulfide-oxidizing free-living species. Targeted PCR amplification of the *soxB* from the genomic DNA of symbiont Rs2, and not Rs1, confirmed the likely absence of these genes from the latter symbiont (Supplementary Figure S4). Advanced stages of whale fall decomposition generate significant increases in sulfide concentrations (Smith and Baco, 2003; Goffredi *et al.*, 2008) and the ability to detoxify this harmful substance should benefit Rs2 during these later stages. Rs1 and Rs2 also differ in their abilities to use carbohydrates (Tables 3 and 4). Specifically, Rs2 possesses a full complement of genes required for the uptake and metabolism of

monosaccharides (ribose), disaccharides (maltose) and amino sugars (*N*-acetylglucosamine). The fact that symbiont Rs1 may be deficient in these capabilities was indicated by the much lower tissue enrichment values when animals were exposed to <sup>13</sup>C-labeled ribose (Table 5). These preliminary experiments should be repeated with larger sample sizes and consistent substrate concentrations to further confirm the distinct capabilities, or lack thereof, of each *Osedax* symbiont. The clade encompassing Rs2 also contains free-living representatives, including several found at the Deep Horizon oil spill (Figure 1; for example, HQ652516, Kessler *et al.*, 2011), similarly suggesting a broader nutritional repertoire than that of Rs1 and the others. During later stages of whale fall decomposition Rs2 may be exposed to a diverse milieu of carbon substrates, including sugars in the water column or on surfaces. Examinations of the specific changes in the availability of carbohydrates, in relation to whale fall longevity, as well as the actual occurrence of the symbionts in the environment outside the symbiosis and the biogeochemical parameters that determine their presence, are needed to test this hypothesis.

### Conclusion

Despite a rise in publications on beneficial bacteria, detailed genomic information is available for relatively few intracellular endosymbionts that are newly acquired from the environment in each host generation. This endeavor is particularly relevant given our growing appreciation of the profound ways by which intimate interactions with bacteria can shape animal biology. In a recent synthesis of the top 100 fundamental questions in ecology, Sutherland *et al.* (2013) suggested that understanding the contribution of intraspecific diversity to the dynamics of mutualistic interactions, as well as the role of symbionts in maintaining host diversity, which must invariably relate to symbiont capabilities, are among the top research goals with regard to microbial ecology. Comparative genomics between closely related bacteria can reveal unique patterns of adaptation and ecological diversification. In this study, an assessment of the genomic capacity of the *Osedax* endosymbionts revealed metabolic versatility in carbon, phosphate and iron uptake strategies, mechanisms for intracellular survival, and numerous potential virulence/colonization capabilities. In the present case, and perhaps for other horizontally propagated symbioses, 'opportunistic acquisition' of locally adapted, environmental stages of the bacterial population may influence host diversity. Further genomic examinations of intracellular symbionts with free-living phases, and the elucidation of environmental controls on their temporal and spatial distributions, will establish a foundation for addressing hypotheses about the complex host-symbiont associations in marine and terrestrial environments.

### Conflict of Interest

The authors declare no conflict of interest.

### Acknowledgements

We thank the captain and crew of the RV *Western Flyer*, the pilots of the ROVs *Tiburón* and *Doc Ricketts*, Y-J Won for introducing us to the Chunlab and for shipboard assistance in June 2011, J Harvey and L Lundsten for shipboard support, S Johnson for support at sea and in the laboratory, as well as G Rouse (Scripps Institution of Oceanography) and VJ Orphan (California Institute of Technology) for continued professional support. We also thank the Moore Foundation for an initial attempt at sequencing the *Osedax* symbiont Rs2 genome, which, despite being aborted due to excessive host contamination, yielded early insights that informed subsequent analyses; E Allen (Scripps Institution of Oceanography) for bioinformatics assistance on the early genome attempts; K Dawson (California Institute of Technology) for <sup>13</sup>C analysis of *Osedax* tissues; C Skennerton (Australian Centre for Ecogenomics, University of Queensland) for the estimates of completeness for both genomes. Funding for undergraduate students was provided by a HHMI grant to Occidental College and the Undergraduate Research Center (Academic Student Projects) at Occidental College. This project was funded primarily by a National Science Foundation grant to S Goffredi (Occidental College, IOS-0923775).

### References

- Actis LA, Fish W, Crosa JH, Kellerman K, Ellenberger SR, Hauser FM *et al.* (1986). Characterization of anguibactin, a novel siderophore from *Vibrio anguillarum* 775(pJM1). *J Bacteriol* **167**: 57–65.
- Arahall DR, Lekunberri I, González JM, Pascual J, Pujalte MJ, Pedrós-Alió C *et al.* (2007). *Neptuniibacter caesariensis* gen. nov., sp. nov., a novel marine genome-sequenced gammaproteobacterium. *Int J Syst Evol Microbiol* **57**: 1000–1006.
- Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA *et al.* (2008). The RAST Server: Rapid annotations using subsystems technology. *BMC Genomics* **9**: 75.
- Braby CE, Rouse GW, Johnson SB, Jones WJ, Vrijenhoek RC. (2007). Bathymetric and temporal variation among *Osedax* boneworms and associated megafauna on whale-falls in Monterey Bay, California. *Deep Sea Res Part I Oceanogr Res* **54**: 1773–1791.
- Bright M, Bulgheresi S. (2010). A complex journey: transmission of microbial symbionts. *Nat Rev Microbiol* **8**: 218–230.
- Brown CT, Howe A, Zhang Q, Pyrkosz AB, Brom TH. (2012). A reference-free algorithm for computational normalization of shotgun sequencing data. *Genomics* arXiv:12034802 (q-bioGN).
- Chiavelli DA, Marsh JW, Taylor RK. (2001). The mannose-sensitive hemagglutinin of *Vibrio cholerae* promotes adherence to zooplankton. *Appl Environ Microbiol* **67**: 3220–3225.
- Chun J, Grim CJ, Hasan NR, Lee J, Choi SY, Haley BJ *et al.* (2009). Comparative genomics reveals mechanism for short-term and long-term clonal transitions in

- pandemic *Vibrio cholerae*. *Proc Natl Acad Sci USA* **106**: 15442–15447.
- Chun J, Kim KY, Lee JH, Choi Y. (2010). The analysis of oral microbial communities of wild-type and toll-like receptor 2-deficient mice using a 454 GS FLX Titanium pyrosequencer. *BMC Microbiol* **10**: 101.
- Coplen TB, Brand WA, Gehre M, Gröning M, Meijer HA, Toman B *et al*. (2006). New guidelines for  $\delta^{13}\text{C}$  measurements. *Anal Chem* **78**: 2439–2441.
- Craig L, Li J. (2008). Type IV pili: paradoxes in form and function. *Curr Opin Struct Biol* **18**: 267–277.
- Dalisy DS, Webb JS, Scheffel A, Svenson C, James S, Holmstrom C *et al*. (2006). A mannose-sensitive haemagglutinin (MSHA)-like pilus promotes attachment of *Pseudoalteromonas tunicata* cells to the surface of the green alga *Ulva australis*. *Microbiology* **152**: 2875–2883.
- Delcher AL, Bratke KA, Powers EC, Salzberg SL. (2007). Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* **23**: 673–679.
- Disz T, Akhter S, Cuevas D, Olson R, Overbeek R. (2010). Accessing the SEED genome databases via Web services API: tools for programmers. *BMC Bioinformatics* **11**: 319.
- Duchaud E, Rusniok C, Frangeul L, Buchrieser C, Givaudan A, Taourit S *et al*. (2003). The genome sequence of the entomopathogenic bacterium *Photobacterium luminescens*. *Nat Biotechnol* **21**: 1307–1313.
- Dupont CL, Rusch DB, Yooseph S, Lombardo MJ, Richter RA, Valas R *et al*. (2012). Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *ISME J* **6**: 1186–1199.
- Dziejman M, Balon E, Boyd D, Fraser CM, Heidelberg JF, Mekalanos JJ. (2002). Comparative genomic analysis of *Vibrio cholerae*: genes that correlate with cholera endemic and pandemic disease. *Proc Natl Acad Sci USA* **99**: 1556–1561.
- Eisenreich W, Dandekar T, Heesemann J, Goebel W. (2010). Carbon metabolism of intracellular bacterial pathogens and possible links to virulence. *Nat Rev Microbiol* **8**: 401–412.
- Fang FC, Libby SJ, Castor ME, Fung AM. (2005). Isocitrate lyase (AceA) is required for *Salmonella* persistence but not for acute lethal infection in mice. *Infect Immun* **73**: 2547–2549.
- Felsenstein J. (1985). Confidence-limits on phylogenies—An approach using the bootstrap. *Evolution* **39**: 783–791.
- Ffrench-Constant RH, Waterfield N, Burland V, Perna NT, Daborn PJ, Bowen D *et al*. (2000). A genomic sample sequence of the entomopathogenic bacterium *Photobacterium luminescens* W14: Potential implications for virulence. *Appl Environ Microbiol* **66**: 3310–3329.
- Finn RD, Clements J, Eddy SR. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res Web Server Issue* **39**: W29–W37.
- Fullner KJ, Boucher JC, Hanes MA, Haines GK III, Meehan BM, Walchle C *et al*. (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.
- Fujikura K, Fujiwara Y, Kawato M. (2006). A new species of *Osedax* (Annelida: Siboglinidae) associated with whale carcasses off Kyushu, Japan. *Zool Sci* **23**: 733–740.
- 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.
- Glover AG, Källström B, Smith CR, Dahlgren TG. (2005). World-wide whale worms? A new species of *Osedax* from the shallow north Atlantic. *Proc Biol Sci* **272**: 2587–2592.
- Goffredi SK, Paull CK, Fulton-Bennett K, Hurtado LA, Vrijenhoek RC. (2004). Unusual benthic fauna associated with a whale-fall in Monterey Canyon, California. *Deep Sea Res Part I Oceanogr Res* **51**: 1295–1306.
- Goffredi SK, Orphan VJ, Rouse GW, Jahnke L, Embaye T, Turk K *et al*. (2005). Evolutionary innovation: a bone-eating marine symbiosis. *Environ Microbiol* **7**: 1369–1378.
- 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.
- Goffredi SK, Wilpiseski R, Lee R, Orphan VJ. (2008). Temporal evolution of methane cycling and phylogenetic diversity of archaea in sediments from a deep-sea whale-fall in Monterey Canyon, California. *ISME J* **2**: 204–220.
- Goffredi SK, Orphan VJ. (2010). Bacterial community shifts in taxa and diversity in response to localized organic loading in the deep sea. *Environ Microbiol* **12**: 344–363.
- Goffredi SK, Morella N, Pulcrano M. (2012). Affiliations between bacteria and marine fish leeches (Piscicolidae), with emphasis on a new deep-sea species from Monterey Canyon, CA. *Environ Microbiol* **14**: 2429–2444.
- Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* **57**: 81–91.
- Gould TA, van de Langemheen H, Muñoz-Elías EJ, McKinney JD, Sacchettini JC. (2006). Dual role of isocitrate lyase 1 in the glyoxylate and methylcitrate cycles in *Mycobacterium tuberculosis*. *Mol Microbiol* **61**: 940–947.
- Griffiths GL, Sigel SP, Payne SM, Neilands JB. (1984). Vibriobactin, a siderophore from *Vibrio cholerae*. *J Biol Chem* **259**: 383–385.
- Higgs ND, Glover AG, Dahlgren TG, Little CT. (2011). Bone-boring worms: characterizing the morphology, rate, and method of bioerosion by *Osedax mucofloris* (Annelida, Siboglinidae). *Biol Bull* **221**: 307–316.
- Higgs ND, Little CTS, Glover AG, Dahlgren TG, Smith CR. (2012). Evidence of *Osedax* worm borings in Pliocene (~3Ma) whale bone from the Mediterranean. *Hist Biol* **24**: 269–277.
- Hilário A, Capa M, Dahlgren TG, Halanych KM, Little CT, Thornhill DJ *et al*. (2011). New perspectives on the ecology and evolution of siboglinid tubeworms. *PLoS One* **6**: e16309.
- Hornung JM, Jones HA, Perry RD. (1996). The hmu locus of *Yersinia pestis* is essential for utilization of free haemin and haem-protein complexes as iron sources. *Mol Microbiol* **20**: 725–739.
- Jeong H, Yim JH, Lee C, Choi SH, Park YK, Yoon SH *et al*. (2005). Genomic blueprint of *Hahella chejuensis*, a marine microbe producing an algicidal agent. *Nucleic Acids Res* **33**: 7066–7073.



- Jukes TH, Cantor CR. (1969). Evolution of protein molecules. In: Munro HN (ed) *Mammalian protein metabolism*, vol 3. Academic Press: New York, pp 21–132.
- Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. (2010). KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* **38**: D355–D360.
- Kessler JD, Valentine DL, Redmond MC, Du MR, Chan EW, Mendes SD. (2011). A persistent oxygen anomaly reveals the fate of spilled methane in the deep Gulf of Mexico. *Science* **331**: 312–315.
- Kiel S, Kahl W-A, Goedert JL. (2011). *Osedax* borings in fossil marine bird bones. *Naturwissenschaften* **98**: 51–55.
- Konstantinidis KT, Tiedje JM. (2004). Trends between gene content and genome size in prokaryotic species with larger genomes. *Proc Natl Acad Sci USA* **101**: 3160–3165.
- Konstantinidis KT, Tiedje JM. (2005). Towards a genome-based taxonomy for prokaryotes. *J Bacteriol* **187**: 6258–6264.
- Koppisch AT, Browder CC, Moe AL, Shelley JT, Kinkel BA, Hersman LE et al. (2005). Petrobactin is the primary siderophore synthesized by *Bacillus anthracis* str Sterne under conditions of iron starvation. *BioMetals* **18**: 577–585.
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D et al. (2009). *Circos*: an information aesthetic for comparative genomics. *Genome Res* **19**: 1639–1645.
- Kuhn M, Kathariou S, Goebel W. (1988). Hemolysin supports survival but not entry of the intracellular bacterium *Listeria monocytogenes*. *Infect Immun* **56**: 79–82.
- Kumar S, Nei M, Dudley J, Tamura K. (2008). MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform* **9**: 299–306.
- Kuwahara H, Yoshida T, Takaki Y, Shimamura S, Nishi S, Harada M et al. (2007). Reduced genome of the thioautotrophic intracellular symbiont in a deep-sea clam, *Calyptogena okutanii*. *Curr Biol* **17**: 881–886.
- Lauro FM, Bartlett DH. (2008). Prokaryotic lifestyles in deep sea habitats. *Extremophiles* **12**: 15–25.
- 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.
- Lee JH, Yi H, Chun J. (2011). rRNASelector: a computer program for selecting ribosomal RNA encoding sequences from metagenomic and metatranscriptomic shotgun libraries. *J Microbiol* **49**: 689–691.
- Litwin CM, Calderwood SB. (1993). Role of iron in regulation of virulence genes. *Clin Microbiol Rev* **6**: 137–149.
- Ludwig W, Strunk R, Westram L, Richter H, Meier H. (2004). ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363–1371.
- Lundsten L, Fraiser K, Schlining KL, Johnson SB, Kuhn LA, Harvey JB et al. (2010). Time-series analysis of six whale-fall communities in Monterey Canyon, California, USA. *Deep Sea Res Part 1 Oceanogr Res* **57**: 1573–1584.
- McKinney JD, Bentrup KH, Muñoz-Elias EJ, Miczak A, Chen B, Chan WT et al. (2000). Persistence of *Mycobacterium tuberculosis* in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. *Nature* **406**: 735–738.
- Mickelsen PA, Sparling PF. (1981). Ability of *Neisseria gonorrhoeae*, *Neisseria meningitidis*, and commensal *Neisseria* species to obtain iron from transferrin and iron compounds. *Infect Immun* **33**: 555–564.
- Miyamoto N, Yamamoto T, Yusa Y, Fujiwara Y. (2013). Postembryonic development of the bone-eating worm *Osedax japonicus*. *Naturwissenschaften* **100**: 285–289.
- Miyazaki M, Nogi Y, Fujiwara Y, Kawato M, Kubokawa K, Horikoshi K. (2008). *Neptunomonas japonica* sp. nov., an *Osedax japonicus* symbiont-like bacterium isolated from sediment adjacent to sperm whale carcasses off Kagoshima, Japan. *Int J Syst Evol Microbiol* **58**: 866–871.
- Monnet V. (2003). Bacterial oligopeptide-binding proteins. *Cell Mol Life Sci* **60**: 2100–2114.
- Muñoz-Elias EJ, McKinney JD. (2005). *Mycobacterium tuberculosis* isocitrate lyases 1 and 2 are jointly required for *in vivo* growth and virulence. *Nature Med* **11**: 638–644.
- Muñoz-Elias EJ, Upton AM, Cherian J, McKinney JD. (2006). Role of the methylcitrate cycle in *Mycobacterium tuberculosis* metabolism, intracellular growth, and virulence. *Mol Microbiol* **60**: 1109–1122.
- Myers EW, Miller W. (1988). Optimal alignments in linear space. *Comput Appl Biosci* **4**: 11–17.
- Naganuma T, Elsaied HE, Hoshii D, Kimura H. (2005). Bacterial endosymbioses of gutless tube-dwelling worms in nonhydrothermal vent habitats. *Mar Biotechnol* **7**: 416–428.
- Newton ILG, Woyke T, Auchtung TA, Dilly GF, Dutton RJ, Fisher CR et al. (2007). The *Calyptogena magnifica* chemoautotrophic symbiont genome. *Science* **315**: 998–1000.
- Nussbaumer AD, Fisher CR, Bright M. (2006). Horizontal endosymbiont transmission in hydrothermal vent tubeworms. *Nature* **441**: 345–348.
- O’Connell Motherway M, Zomer A, Leahy SC, Reunanen J, Bottacini F, Claesson MJ et al. (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 USA* **108**: 11217–11222.
- Oliveros JC. (2007). VENNY An interactive tool for comparing lists with Venn Diagrams. <http://bioinfogpcnbcscs/ tools/venny/indexhtml>.
- Perry RD, Brubaker RR. (1979). Accumulation of iron by *Yersinia*. *J Bacteriol* **137**: 1290–1298.
- Portnoy DA, Jacks PS, Hinrichs DJ. (1988). Role of hemolysin for the intracellular growth of *Listeria monocytogenes*. *J Exp Med* **167**: 1459–1471.
- Preusse E, Quast C, Knittel K, Fuchs BM, Ludwig WG, Peplies J et al. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35**: 7188–7196.
- Robidart JC, Bench SR, Feldman RA, Novoradovsky A, Podell SB, Gaasterland T et al. (2008). Metabolic versatility of the *Riftia pachyptila* endosymbiont revealed through metagenomics. *Environ Microbiol* **10**: 727–737.
- Rouse GW, Goffredi SK, Vrijenhoek RC. (2004). *Osedax*: Bone-eating marine worms with dwarf males. *Science* **305**: 668–671.
- Rouse GW, Worsaae K, Johnson SB. (2008). Acquisition of dwarf male ‘harems’ by recently settled females of *Osedax roseus* n. sp. (Siboglinidae; Annelida). *Biol Bull* **214**: 67–82.

- 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 *et al.* (2005). Complete genome sequence of *Vibrio fischeri*: a symbiotic bacterium with pathogenic congeners. *Proc Natl Acad Sci USA* **102**: 3004–3009.
- Saitou N, Nei M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406–425.
- Salathé RM, Vrijenhoek RC. (2012). Temporal variation and lack of host specificity among bacterial endosymbionts of *Osedax* bone worms (Polychaeta: Siboglinidae). *BMC Evol Biol* **12**: 189.
- Schell MA, Ulrich RL, Ribot WJ, Brueggemann EE, Hines HB, Chen D *et al.* (2007). Type VI secretion is a major virulence determinant in *Burkholderia mallei*. *Mol Microbiol* **64**: 1466–1485.
- Schlieker C, Zentgraf H, Dersch P, Mogk A. (2005). ClpV, a unique Hsp100/Clp member of pathogenic proteobacteria. *Biol Chem* **386**: 1115–1127.
- Schwarz S, Hood RD, Mougous JD. (2010). What is type VI secretion doing in all those bugs? *Trends Microbiol* **18**: 531–537.
- Siegl A, Kamke J, Hochmuth T, Piel J, Richter M, Liang C. (2010). Single-cell genomics reveals the lifestyle of *Poribacteria*, a candidate phylum symbiotically associated with marine sponges. *ISME J* **5**: 61–70.
- Smith CR, Baco AR. (2003). The ecology of whale-falls at the deep-sea floor. *Oceanogr Mar Biol Annu Rev* **41**: 311–354.
- Smith C, Baco A, Glover A. (2002). Faunal succession on replicate deep-sea whale-falls: time scales and vent-seep affinities. *Cah de Biolog Mar* **43**: 293–297.
- Stoebner JA, Payne SM. (1988). Iron-regulated hemolysin production and utilization of heme and hemoglobin by *Vibrio cholerae*. *Infect Immun* **56**: 2891–2895.
- Sutherland WJ, Freckleton RP, Godfray HCJ, Beissinger SR, Benton T, Cameron DD *et al.* (2013). Identification of 100 fundamental ecological questions. *J Ecol* **101**: 58–67.
- Tatusov RL, Koonin EV, Lipman DJ. (1997). A genomic perspective on protein families. *Science* **278**: 631–637.
- Thompson JD, Gibson TJ, Higgins DG. (2002). Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics* **Chapter 2**: Unit 2.
- Thornhill DJ, Fielman KT, Santos SR, Halanych KM. (2008). Siboglinid-bacteria endosymbiosis: A model system for studying symbiotic mechanisms. *Commun Integr Biol* **1**: 163–166.
- Tomich M, Planet PJ, Figurski DH. (2007). The tad locus: postcards from the widespread colonization island. *Nat Rev Microbiol* **5**: 363–375.
- Tresguerres M, Katz S, Rouse GW. (2013). How to get into bones: proton pump and carbonic anhydrase in *Osedax* boneworms. *Proc Royal Soc B* **280**: 20130625.
- Verna C, Ramette A, Wiklund H, Dahlgren TG, Glover AG, Gaill F *et al.* (2010). High symbiont diversity in the bone-eating worm *Osedax mucofloris* from shallow whale-falls in the North Atlantic. *Environ Microbiol* **12**: 2355–2370.
- Vrijenhoek RC, Johnson SB, Rouse GW. (2008). Bone-eating *Osedax* females and their ‘harems’ of dwarf males are recruited from a common larval pool. *Molec Ecol* **17**: 4535–4544.
- Vrijenhoek RC, Johnson SB, Rouse GW. (2009). A remarkable diversity of bone-eating worms (*Osedax*; Siboglinidae; Annelida). *BMC Biol* **7**: 74.
- Wadhams GH, Armitage JP. (2004). Making sense of it all: bacterial chemotaxis. *Nat Rev Mol Cell Biol* **5**: 1024–1037.
- Warren MJ, Raux E, Schubert HL, Escalante-Semerena JC. (2002). The biosynthesis of adenosylcobalamin (Vitamin B12). *Nat Prod Rep* **19**: 390–412.
- Woyke T, Teeling H, Ivanova NN, Huntemann M, Richter M, Gloeckner FO *et al.* (2006). Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* **443**: 950–955.
- Woyke T, Sczyrba A, Lee J, Rinke C, Tighe D, Clingenpeel S *et al.* (2011). Decontamination of MDA reagents for single cell whole genome amplification. *PLoS One* **6**: e26161.
- Yang JC, Madupu R, Durkin AS, Ekborg NA, Pedamallu CS, Hostetler JB *et al.* (2009). The complete genome of *Teredinibacter turnerae* T7901: an intracellular endosymbiont of marine wood-boring bivalves (shipworms). *PLoS One* **4**: e6085.

Supplementary Information accompanies this paper on The ISME Journal website (<http://www.nature.com/ismej>)