Genome sequence of the dissimilatory metal ion–reducing bacterium *Shewanella oneidensis*

John F. Heidelberg[1,2], Ian T. Paulsen[1,3], Karen E. Nelson[1], Eric J. Gaidos[4,5], William C. Nelson[1], Timothy D. Read[1], Jonathan A. Eisen[1,3], Rekha Seshadri[1], Naomi Ward[1,2], Barbara Methe[1], Rebecca A. Clayton[1], Terry Meyer[6], Alexandre Tsapin[4], James Scott[7], Maureen Beanan[1], Lauren Brinkac[1], Sean Daugherty[1], Robert T. DeBoy[1], Robert J. Dodson[1], A. Scott Durkin[1], Daniel H. Haft[1], James F. Kolonay[1], Ramana Madupu[1], Jeremy D. Peterson[1], Lowell A. Umayam[1], Owen White[1], Alex M. Wolf[1], Jessica Vamathevan[1], Janice Weidman[1], Marjorie Impraim[1], Kathy Lee[1], Kristy Berry[1], Chris Lee[1], Jacob Mueller[1], Hoda Khouri[1], John Gill[1], Terry R. Utterback[1], Lisa A. McDonald[1], Tamara V. Feldblyum[1], Hamilton O. Smith[1,8], J. Craig Venter[1,9], Kenneth H. Nealson[4,10], and Claire M. Fraser[1,11]*

Shewanella oneidensis is an important model organism for bioremediation studies because of its diverse respiratory capabilities, conferred in part by multicomponent, branched electron transport systems. Here we report the sequencing of the *S. oneidensis* genome, which consists of a 4,969,803–base pair circular chromosome with 4,758 predicted protein-encoding open reading frames (CDS) and a 161,613–base pair plasmid with 173 CDSs. We identified the first *Shewanella* lambda-like phage, providing a potential tool for further genome engineering. Genome analysis revealed 39 *c*-type cytochromes, including 32 previously unidentified in *S. oneidensis*, and a novel periplasmic [Fe] hydrogenase, which are integral members of the electron transport system. This genome sequence represents a critical step in the elucidation of the pathways for reduction (and bioremediation) of pollutants such as uranium (U) and chromium (Cr), and offers a starting point for defining this organism's complex electron transport systems and metal ion–reducing capabilities.

Shewanella oneidensis MR-1 (formerly *Shewanella putrefaciens* strain MR-1; ref. 1) is a facultatively aerobic Gram-negative bacterium with remarkably diverse respiratory capacities. As in other species, *S. oneidensis* uses oxygen as the terminal electron acceptor during aerobic respiration; however, under anaerobic conditions, *S. oneidensis* undertakes respiration by reducing alternative terminal electron acceptors such as oxidized metals (including Mn(III) and (IV), Fe(III), Cr(VI), U(VI)), fumarate, nitrate, trimethylamine *N*-oxide, dimethyl sulfoxide, sulfite, thiosulfate, and elemental sulfur[2,3]. Such plasticity in alternative electron acceptors for anaerobic respiration has not been observed in any other organism.

The biological activities of metal ion–reducing bacteria have considerable implications with regard to environmental pollutants. *Shewanella oneidensis* and other *Shewanella* species can directly reduce both uranium and chromium from the dissolved liquid state (U(VI) and Cr(VI)) to insoluble oxides (U(IV) and Cr(III)). Such abilities could facilitate the removal of dilute metal pollutants in both contained-storage and natural sites. Additionally, *S. oneidensis* can produce large quantities of sulfide from either thiosulfate or S⁰. This allows the use of *S. oneidensis* to immobilize toxic metals through the formation of insoluble metal sulfides. Finally, *S. oneidensis* has been proposed for bioremediation of anoxic polluted sites because the redox potentials of iron and manganese ions as terminal electron acceptors are high enough to drive the oxidation of organic pollutants. This suggests that *S. oneidensis* might prove useful for direct bioremediation of both metal and organic pollutants under anaerobic conditions.
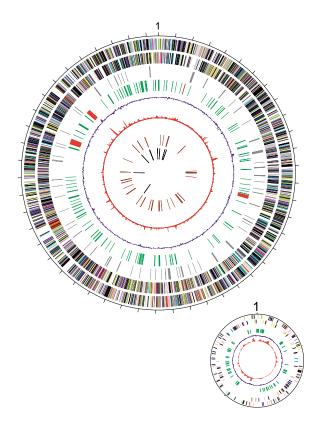
However, it is important to consider that the activities of such metal ion–reducing bacteria can also have negative effects. Many organic pollutants are strongly bound to iron and manganese oxides, and when such oxides are reduced, the pollutants are solubilized[3]. It is therefore essential to understand such activities by these organisms before undertaking environmental bioremediation. The rationale for selecting *S. oneidensis* for whole-genome sequencing was to advance biological understanding of this organism, thereby expediting efforts to use it for bioremediation of dissolved metal ions and organic toxins in water supplies.

Results and discussion

General genome features. The *S. oneidensis* MR-1 genome was sequenced by the whole-genome sequencing method[4–6]. The genome features are summarized in Table 1 and Figure 1. The genome is a circular chromosome of 4,969,803 base pairs (bp). There are a total of 4,758 predicted protein-encoding open reading frames (CDSs); 54.4%

[1]*The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850. [2]The Center for Marine Biotechnology, 701 East Pratt Street, Baltimore, MD 21202. [3]Johns Hopkins University, Charles and 34th Streets, Baltimore, MD 21218. [4]Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, CA 91109. [5]Current address: Department of Geology and Geophysics, University of Hawaii, Honolulu, HI 96822. [6]Department of Biochemistry and Molecular Biophysics, University of Arizona, Tucson, AZ 85721. [7]Carnegie Institution of Washington, Geophysical Laboratory, 5251 Broad Branch Rd. NW, Washington, DC 20015. [8]Current address: Celera Genomics, 45 West Gude Drive, Rockville, MD 20850. [9]Current address: The Center for the Advancement of Genomics, 1901 Research Boulevard, Rockville, MD 20850. [10]Current address: Department of Earth Sciences, University of Southern California, Los Angeles, CA 90089. [11]George Washington University Medical Center, 2300 Eye Street NW, Washington, DC 20037. *Corresponding author (cmfraser@tigr.org).*

Figure 1. Circular representation of the *Shewanella oneidensis* genome. The chromosome and megaplasmid are depicted. From the outside inward: the first and second circle show predicted coding regions on the plus and minus strand, by biological role: salmon, amino acid biosynthesis; light blue, biosynthesis of cofactors and prosthetic groups and carriers; light green, cell envelope; red, cellular processes; brown, central intermediary metabolism; yellow, DNA metabolism; green, energy metabolism; purple, fatty acid and phospholipid metabolism; pink, protein fate/synthesis; orange, purines, pyrimidines, nucleosides, and nucleotides; blue, regulatory functions; gray, transcription; teal, transport and binding proteins; black, hypothetical and conserved hypothetical proteins. The third circle shows genes involved in the electron transport chain. The fourth circle shows transposon-related (green) and phage-related (red) genes. The fifth circle shows the percentage of G+C in relation to the mean G+C of the replicon in a 2,000 bp window (chromosome) or 200 bp window (megaplasmid). The sixth circle shows the χ^2 values for trinucleotide composition in a 2,000 bp window. The seventh and eighth circles are tRNAs and rRNAs, respectively.

genes, suggesting lineage-specific duplications (Fig. 2). Most of the duplicated CDSs encode products involved in energy metabolism (46), transport and binding (22), protein fate (22), cell envelope (22), transposition (258), or unknown functions encoded by conserved hypothetical (73) or hypothetical proteins (135). The extensive duplication of genes involved in electron transport (29 of the duplicated genes involved in energy metabolism encode proteins that specifically carry out electron transport) suggests the importance of these genes for *S. oneidensis* biology, and particularly, for its ability to function as a respiratory generalist (see below).

Consistent with 16S rRNA-based phylogeny, the most similar proteomes to *S. oneidensis*, from organisms for which complete genome sequence information is available, are those of the γ-Proteobacteria (Fig. 2). *Vibrio cholerae* is the only organism with extensive regions of similar gene order (synteny); however, these syntenic regions are only on the *V. cholerae* chromosome I. This suggests that the second chromosome of *V. cholerae* was captured after the divergence of *V. cholerae* and *S. oneidensis*, that the second chromosome was lost in the *S. oneidensis* lineage, or that the second chromosome was rearranged in the *Shewanella* lineage. The main structural difference between the *S. oneidensis* chromosome and *V. cholerae* chromosome I is a significant cross-alignment ($P = 6.05 \times 10^{-165}$) reflecting inversions around the origin and terminus, as has been seen in other closely related bacteria[8] (Fig. 3).

Genome analysis revealed a 51,857 bp lambda-like phage genome (referred to as lambdaSo), both integrated in the *S. oneidensis* genome and present in nonintegrated form, suggesting that it is a functional phage. LambdaSo shares syntenic regions with *Pseudomonas aeruginosa* D3 (ref. 9) and enterobacteria HK022 (ref. 10) phage. However, beyond these syntenic regions there are no orthologous genes between lambdaSo and D3 or HK022. This novel *Shewanella* phage represents a potentially valuable tool for *Shewanella* genome engineering because it probably integrates into the genome in a lysogenic response.

Also integrated in the *S. oneidensis* genome are two phylogenetically distinct phage related to the *Escherichia coli* Mu (referred to as MuSo1, 34,551 bp, SO0641–SO0683; and MuSo2, 35,666 bp, SO2652–SO2704). The *S. oneidensis* Mu-like phages share syntenic regions with Mu and Mu-like phages from *Haemophilus influenzae*, *Neisseria meningitidis* serogroup A, and *N. meningitidis* serogroup B[11], interspersed with unique regions (Fig. 4). All the predicted proteins in the Mu-like regions are either phage proteins or unique.

Electron transport. The anaerobic respiratory versatility of *S. oneidensis* is probably a consequence of

were assigned a biological function based on a classification scheme adapted from Riley[7], 22.2% matched predicted coding sequences without a defined function from other organisms (conserved hypothetical CDSs), and 23.4% were unique to this bacterium. *Shewanella oneidensis* contains a 161,613 bp iteron-type plasmid that has a total of 173 CDSs; 61.3% were assigned to a biological function[7], 15.6% were conserved hypothetical CDSs, and 23.1% were unique to this organism. The plasmid contains 59 CDSs that are probable transposases.

The majority of the *S. oneidensis* MR-1 CDSs are most similar to *Vibrio cholerae* genes (1,265 CDSs; 32.33% of the *V. cholerae* genome), but 683 (13.85% of the *S. oneidensis* genome) of the *S. oneidensis* CDSs have highest similarity to other *S. oneidensis*

Table 1. General features of the *Shewanella oneidensis* genome

	Chromosome	Plasmid
Size (bp)	4,969,803	161,613
G+C percentage	46.0	43.7
Number of predicted CDSs	4,758	173
Average size of CDS (bp)	893	742
Percentage coding	85.5	79.5
Number of rRNA operons (16S-23S-5S)	9	0
Number of tRNAs[a]	102	0
Number of CDSs similar to known proteins	2,430	91
Number of CDSs similar to proteins of unknown function[b]	371	7
Number of conserved hypothetical proteins[c]	843	28
Number of hypothetical proteins[d]	1,114	47
Number of rho-independent terminators	833	13

[a]There is one selenocysteine tRNA likely to be important in translation of the selenium-containing, nitrate-inducible formate dehydrogenase (SO0101).
[b]Proteins of unknown function have substantial similarity (homology) to a named protein for which there is currently no known function.
[c]Conserved hypothetical proteins have sequence similarity to a translation of another CDS, but there is currently no evidence a protein is expressed.
[d]Hypothetical proteins have no substantial similarity to another protein.
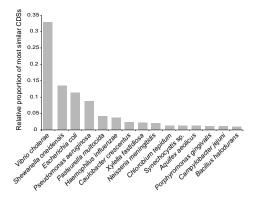
Figure 2. Comparison of the *Shewanella oneidensis* open reading frames with those of completely sequenced organisms. The sequences of all proteins from each completed genome were retrieved from National Center for Biotechnology Information and the TIGR Comprehensive Microbial Resource databases. All *S. oneidensis* CDSs were compared against the combined database of all genomes with FASTA3 (including *S. oneidensis*; however, the exact match CDS was removed for the results). The numbers of *S. oneidensis* CDSs with greatest significant similarity (E = 10^{-5}) are shown in proportion to the number of CDSs in the searched genome. Only the first 15 organisms, after adjustment for genome size, are presented.

its multicomponent branched electron transport system, composed of cytochromes, reductases, iron–sulfur proteins, and quinones[12]. Presumably, during anaerobic electron transport, multiple components of this system shunt electrons to insoluble metal oxides, which are reduced during contact with the bacterial surface[13]. *Shewanella oneidensis* possesses a large and diverse set of genes encoding proteins of the electron transport system. The elucidation of these components of the electron transport system is the first step in understanding the complexities of anaerobic electron transport and metal ion reduction.

Cytochromes. In *S. oneidensis*, 80% of membrane-bound *c*-type heme is localized to the outer membrane, suggesting a direct role for *c*-type cytochromes in metal reduction[14]. Genome analysis shows that *S. oneidensis* has more *c*-type cytochromes than any other organism sequenced to date (*S. oneidensis*, 39; *V. cholerae*, 12; *E. coli*, 7; *P. aeruginosa*, 32), including 14 *c*-type cytochromes with four or more heme-binding sites not described before in *S. oneidensis* (Supplementary Table 1 online).

The *S. oneidensis* genome contains eight genes encoding decaheme cytochrome *c* proteins that form two paralogous families (Supplementary Table 1 online). The first group consists of four periplasmic proteins with N-terminal leaders, some of which may be cleaved following translocation, and includes MtrA, which is essential for the reduction of Fe-citrate and MnO$_2$ (ref. 15). Members of the second family are similar only in the N-terminal region. Three members are likely outer-membrane lipoproteins with diacylglycerol attachment to N-terminal cysteines. These include OmcA and OmcB, which are known to be involved in the reduction of extracellular MnO$_2$ (ref. 16). Diversification of the *S. oneidensis* decaheme cytochrome *c* proteins has allowed for a specialization of labor of these proteins in both cellular location (i.e., the periplasmic side

Figure 3. Whole-genome proteome alignments between *Shewanella oneidensis* and *Vibrio cholerae*. The chromosomal locations of pairs of predicted proteins that have significant similarity (on the basis of FASTA3 comparisons, E = 10^{-5}) are indicated. Only the best match for each *S. oneidensis* CDS is shown. The filtering for top matches for each CDS removes noise due to the presence of many large multigene families.
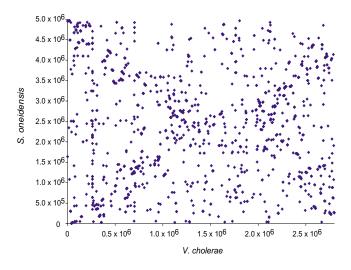
of the cytoplasmic membrane, the periplasmic space, and the outer membrane) and specific electron-transporting role.
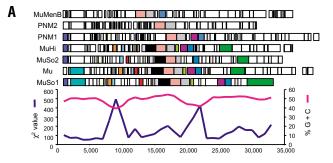
Genome analysis suggests a role for tetraheme cytochrome *c* proteins in the use of amino acids as electron acceptors. Two tetraheme *c* cytochromes are in gene clusters containing the cytochrome *c* subunit, a separate flavoprotein subunit (together forming a flavocytochrome *c*), and a Pal/histidase family protein (SO3299–SO3301; SO3056–SO3058). This suggests a novel type of anaerobic respiration involving deaminated amino acids as terminal electron acceptors. The Pal/histidase family protein could introduce a double bond by deamination of histidine, which could in turn be reduced by the flavocytochrome *c*, analogous to the reduction of fumarate. In *Wolinella*, a related flavocytochrome *c* (FccAB) is responsible for reduction of the unusual substrates acrylate and methacrylate[17].

The tetraheme cytochrome *c*, CymA, is localized to the periplasmic side of the cytoplasmic membrane and represents a branchpoint in the electron transport chain for reduction of a variety of substrates, including Fe-citrate, MnO$_2$, nitrate, and fumarate[18]. In *S. oneidensis, cymA* is not adjacent to other electron transfer protein genes, suggesting that it is regulated differently than the various pathways with which it interacts. This protein probably accepts electrons from quinol and shuttles them to periplasmic reductases and soluble cytochromes that, in turn, mediate electron transfer to the decaheme cytochromes in the outer membrane.

Hydrogenases. Hydrogenases are ubiquitous in prokaryotes and serve as intermediaries in numerous pathways involved in hydrogen production and/or consumption. Hydrogenases are classified into two metalloenzyme families based on the metal content of the active site: the Ni-containing hydrogenases (i.e., [NiFe] and [NiFeSe] hydrogenases) and the Fe-S cluster hydrogenase ([Fe] hydrogenase). *Shewanella oneidensis* has a classic [NiFe] hydrogenase operon (SO2097–SO2099) and a heterodimeric [Fe] hydrogenase (SO3920, SO3921), an important component in its anaerobic respiratory and potentially its metal-reducing capability. There have been no previous reports of the presence of the [Fe] hydrogenase enzyme family in a facultative aerobe (or any proteobacterial lineage other than δ-Proteobacteria). The *S. oneidensis* [Fe] hydrogenase is most similar to the periplasmic *Desulfovibrio* [Fe] hydrogenases, and includes a twin arginine translocation signal motif at the N terminus of the small subunit. Periplasmic [Fe] hydrogenases have been implicated in catalyzing hydrogen uptake for the purposes of electron donation to low-potential, multiheme *c*-type cytochromes, including those potentially involved in metal reduction[19,20].

Metabolism and transport. Fermentative end products rather than sugars have been suggested as the major energy source for *S. oneiden-*
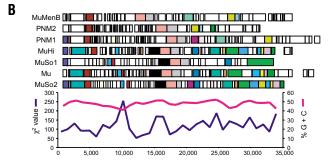
Figure 4. Comparison of the MuSo1 (A) and MuSo2 (B) phage with Mu, MuHi, PNM1, PNM2, and MuMenB phage. Top panel shows the locations of pairs of predicted proteins that have significant similarity (on the basis of FASTA3 comparisons; $E = 10^{-5}$) as indicated by colors. Bottom panel shows the χ^2 values for trinucleotide composition and G+C% in a 2,000 bp window.

sis[21]. Genome analysis predicts energy-generating pathways for many fermentative end products (e.g., acetate, fumarate, lactate, malate, pyruvate, and succinate) and transporters for lactate and other carboxylates. Genome analysis also suggests the presence of metabolic pathways for sugars and glycerol not previously described in *S. oneidensis*[1]. For example, *S. oneidensis* possesses a complete pentose phosphate pathway and all glycolytic pathway enzymes, except phosphofructokinase. Two proton-driven glucose/galactose symporters and a glucose phosphotransferase system (PTS) were identified. On the basis of these observations, glucose and galactose could both potentially be utilized via the pentose phosphate pathway, bypassing the missing phosphofructokinase step in glycolysis. However, previous biochemical characterization has suggested that while some species of *Shewanella* are capable of metabolizing several sugars, only galactose is utilized by *S. oneidensis* for growth[1]. There are two possible explanations for this apparent inconsistency between experimental and genomic observations: either glucose is a source of carbon and energy for *S. oneidensis*, but its utilization has not been experimentally detected, or *S. oneidensis* imports glucose or synthesizes glucose-6-phosphate via gluconeogenesis (with malate and serine as putative substrates) only for biosynthetic purposes such as cell wall synthesis or the synthesis of glycogen as a storage molecule.

Pathways for the utilization of gluconate and glycerol are present; the former is probably imported via proton symport. Bacterial glycerol uptake is typically via major intrinsic protein (MIP) family glycerol channels, and no glycerol channel was identified. This might explain previous observations that glycerol does not support growth[1]; however, a MIP aquaporin is present, and a variety of aquaporins are permeable to glycerol[22].

Glutamate appears central for metabolism, being a direct precursor or amino group donor for the biosynthesis of nine amino acids. Additionally, glutamate can be converted to succinate by the γ-aminobutyric acid pathway. Underscoring the importance of glutamate, seven probable proton or sodium ion–driven glutamate

transporters were identified: six homologs of GltP and a homolog of GltS. This appears to be an amplification of GltP in *Shewanella*, as no other sequenced organism possesses more than two GltP-like glutamate transporter paralogs.

All remaining amino acids can be synthesized *de novo*. Several peptide uptake systems and an extensive array of intracellular peptidases are also present in *S. oneidensis*, implying that peptides are an important nutrient for this organism. This may explain the presence in *S. oneidensis* of seven LysE family amino acid efflux proteins, which have been implicated in preventing accumulation of amino acids to bacteriostatic concentrations during growth of *Corynebacterium glutamicum* on peptides[23].

Under anaerobic conditions, many metal ions are used as alternative electron acceptors. Transporters for a range of transition metal ions including Fe^{2+}, Fe^{3+}, Cu^+, CrO_4^{2-}, Hg^{2+}, Co^{2+}, Mn^{2+}, and Ni^{2+} are present, as well as other divalent metal cation transporters whose specificity could not be determined. Despite the importance of metal ions as electron acceptors, *S. oneidensis* does not have an unusually high number of metal ion transporters, nor amplifications of any metal ion transporter families. Indeed, *S. oneidensis* has fewer transition metal ion transporters relative to genome size than other sequenced organisms such as *E. coli*, *V. cholerae*, and *P. aeruginosa*. This is consistent with theories that metal reduction occurs extracellularly through direct contact with the bacterial surface rather than by transport of metal ions into the cell[14].

There are nine MexB-like multidrug efflux proteins belonging to the resistance–nodulation–cell division (RND) protein family encoded in the genome, more than in any sequenced organism other than *P. aeruginosa*[24]. Whether these play a role in mediating intrinsic drug resistance or in secreting secondary metabolites or other hydrophobic compounds is unclear.

Regulatory function. *Shewanella oneidensis* is found in a wide variety of environments (including soil, sediments, the water column, and clinical isolates)[1]. An expansion in the proportion of regulatory genes (both two-component and transcriptional regulators) has been observed in environmental bacteria and suggested as a mechanism to allow them to adapt to changing and diverse conditions[4,24]. In *Shewanella oneidensis* we identified 88 two-component regulatory system proteins (23 histidine protein kinase genes (HPK), 57 response regulators (RR), and 8 HPK–RR hybrids) that could allow rapid detection and response to environmental changes. However, there is a relative overall paucity of regulatory genes in *S. oneidensis* compared with other environmental bacteria[4,5,24,25]. This is probably a consequence of its limited use of carbon sources rather than environmental sensing. For example, *S. oneidensis* lacks homologs to most members of the deoR family of DNA-binding transcriptional regulators of *E. coli*, which, among other functions, control the metabolism of complex carbon substrates[26].

Pathogenicity. *Shewanella oneidensis* has two type IV pilin gene clusters (MSHA and *tapABCD*) that may be important as host colonization factors, phage receptors, and mediators of DNA transfer. Additionally, *S. oneidensis* is known to attach to solid surfaces (i.e., iron and manganese oxides) and form biofilms[13,27,28], and MSHA plays a role in biofilm formation in *V. cholerae*[29,30].

Shewanella species are infrequent opportunistic human pathogens[1,31]. Little is known about the pathogenicity of *S. oneidensis*, but complete genome analysis revealed several potential virulence determinants. *Shewanella oneidensis* contains a putative pore-forming RTX toxin operon (SO4317–SO4319). This toxin could be involved in a process analogous to *Shewanella putrefaciens*–mediated cellulitis and other reported clinical manifestations[31]. *Shewanella oneidensis* also contains a putative virulence protein with fibronectin type III domain N-terminal repeats and a C-terminal transmembrane domain (SO0189) that could be important in colonization of

eukaryotic tissues. Fibronectin has been characterized as an extracellular matrix glycoprotein capable of multiple interactions with cell surfaces and other matrix components, potentially allowing *S. oneidensis* to bind host cells in a manner similar to *S. putrefaciens*[32].

Plasmid functions. Overall, 73.4% of the genes on the plasmid have no known function or are transposases; however, several genes with putative functions were identified, including the DNA damage bypass polymerase, umuCD (refs 33, 34), a putative restriction/modification system, and a multidrug efflux protein. Additionally, 20 genes on the plasmid (excluding transposases) appear to be recent duplications. One duplicated segment encoding three hypothetical proteins and two proteins composing a metal ion transport system is found in the same order on the chromosome. The abundance of transposase genes on the plasmid and chromosome makes it likely that homologous recombination could move genetic material between the two molecules.

Insertion sequences. *Shewanella oneidensis* has a large and diverse population of insertion sequence (IS) elements, which may have had a critical role in shaping this genome (Supplementary Table 2 online). The two most common ISs (ISSo1 and ISSo4, with 49 and 53 copies, respectively) are frequently found to be adjacent, suggesting a common IS preference. Some IS isotypes with high copy numbers are restricted to only one replicon (e.g., ISSo11 found only on the chromosome; ISSo8 and ISSo9 found only on the plasmid). However, most ISs are distributed on both the plasmid and chromosome, and there are transposase genes with identical sequence present on both molecules, suggesting intra-replicon transposition.

There is evidence of insertion of IS elements within previously intact genes (Supplementary Table 3 online). Some elements (especially ISSo1) are configured in a manner suggestive of a compound transposon; for instance, the region between IS elements SO4293–SO4294 and SO4300–SO4301 includes a putative chloramphenicol acetyltransferase (SO4299) and cAMP-binding protein (SO4298). The 4,211 bp duplication in the major chromosome (3,152,031–3,156,241 and 3,588,759–3,592,669 in reverse) resembles a composite transposon consisting of an intact ISSo4 element, truncated ISSo1, and two members of a family of conserved hypothetical paralogs.

Conclusions. *Shewanella oneidensis* MR-1 has great potential for the bioremediation of both metal ions and organic compounds. One of the major challenges in using microbes for bioremediation is the need to understand and predict their activities, and knowledge of the *S. oneidensis* gene complement provides a foundation for this research. In particular, the complete genomic sequence will allow the identification and characterization of the metal reductase(s) and of the regulatory networks that control them. It will also allow a more compete picture of the electron transport and metabolic capabilities of this bacterium. Finally, the discovery of the *Shewanella* lambda-like phage may provide an avenue for genetic manipulation of this group of microbes and allow the design of strains for specific bioremediation purposes.

Experimental protocol

Sequencing. *Shewanella oneidensis* MR-1 (American Type Culture Collection no. 700550) was grown from a single isolated colony. Cloning, sequencing, and assembly were as described for genomes sequenced by TIGR[4–6]. One small-insert (2–3 kb) plasmid library was constructed in pUC-derived vectors after random mechanical shearing (nebulization) of genomic DNA. One large-insert (~18 kb) shotgun library was constructed in λ-DASH II vectors (Stratagene, La Jolla, CA) after partial *Sau*IIIA digestion of genomic DNA. Sequencing of the small-insert libraries was achieved at a success rate of 72%, with an average read length of 528 bp. The plasmid and λ sequences were jointly assembled using TIGR Assembler[35]. The coverage criteria were that every position required at least double-clone coverage (or sequence from a PCR product amplified from genomic DNA), and either sequence from both strands or with two different sequencing

chemistries. The sequence was edited manually, and additional PCR[36] and sequencing reactions were carried out to close gaps, improve coverage, and resolve sequence ambiguities. All repeated DNA regions were verified with PCR amplification across the repeat and sequencing the product. The final genome is based on 71,777 sequences.

Genome analysis. The replicative origin was determined by similarity to the *Vibrio cholerae oriC*, co-localization of genes (*dnaA*, *dnaN*, *recF*, and *gyrA*) often found near the origin in prokaryotic genomes, and GC nucleotide skew (G–C/G+C) analysis[37]. On this basis, we designated base pair 1 in an intergenic region that is located in the putative origin of replication.

An initial set of open reading frames likely to encode proteins were predicted using Glimmer software[38]. All predicted proteins >30 amino acids were searched against a nonredundant protein database as described[4–6]. Frameshifts and point mutations were detected and corrected where appropriate. Remaining frameshifts and point mutations are considered to be authentic and were annotated as "authentic frameshift" or "authentic point mutation". Protein membrane-spanning domains were identified by TopPred[39]. The 5′ regions of each CDS were inspected to define initiation codons using homologies, and position of ribosomal binding sites and transcriptional terminators. Two sets of hidden Markov models were used to determine CDS membership in families and superfamilies: pfam version 5.5 (ref. 40) and TIGRFAMs 1.0 (ref. 41). Pfam version 5.5 hidden Markov models were also used with a constraint of minimum two hits to find repeated domains within proteins and mask them.

Domain-based paralogous families were then built by conducting all-versus-all searches on the remaining protein sequences using a modified version of an earlier described method[4–6].

Distribution of all 64 trinucleotides (3-mers) for each chromosome was determined, and the 3-mer distribution in 2,000 bp windows that overlapped by half their length (1,000 bp) across the genome was computed. For each window, we computed the χ^2 statistic on the difference between its 3-mer content and that of the whole chromosome. A large value for this statistic indicates that the 3-mer composition in this window is different from the rest of the chromosome. Probability values for this analysis are based on the assumption that the DNA composition is relatively uniform throughout the genome. Because this assumption may be incorrect, we prefer to interpret high χ^2 values merely as indicators of regions on the chromosome that appear unusual and demand further scrutiny.

The extent of potential lineage-specific gene duplications in this genome and the most similar whole proteome from organisms with competed genomes was determined by searching the *S. oneidensis* CDSs against all other competed genomes. The sequences of all proteins from each completed genome were retrieved from National Center for Biotechnology Information and the TIGR Comprehensive Microbial Resource databases. All *S. oneidensis* CDSs were searched with FASTA3 against all CDSs from the complete genomes, and matches with a FASTA E ≤ 10⁻⁵ were considered significant. Potential lineage-specific gene duplications was estimated by identification of CDSs that are more similar to other CDSs within the *S. oneidensis* MR-1 genome than to CDSs from other complete genomes.

URLs. The annotated genome and the gene family alignments are available online at http://www.tigr.org/tigr-scripts/CMR2/GenomePage3.spl?database=gsp. The sequences have been deposited in GenBank with accession numbers AE014299 (chromosome) and AE014300 (plasmid).

Note: Supplementary information is available on the Nature Biotechnology website.

1. Venkateswaran, K. *et al.* Polyphasic taxonomy of the genus *Shewanella* and description of *Shewanella oneidensis* sp. nov. *Int. J. Syst. Bacteriol.* **49 (**Pt. 2), 705–724 (1999).
2. Myers, C.R. & Nealson, K.H. Bacterial manganese reduction and growth with manganese oxide as the sole electron acceptor. *Science* **240**, 1319–1321 (1988).
3. Nealson, K.H. & Saffarini, D. Iron and manganese in anaerobic respiration: environmental significance, physiology, and regulation. *Annu. Rev. Microbiol.* **48**, 311–343 (1994).
4. Heidelberg, J.F. *et al.* DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae. Nature* **406**, 477–483 (2000).
5. Nierman, W. C.*et al.* Complete genome sequence of *Caulobacter crescentus. Proc. Natl. Acad. Sci. USA* **98**, 4136–4141 (2001).
6. Tettelin, H. *et al.* Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae. Science* **293**, 498–506 (2001).
7. Riley, M. Functions of the gene products of *Escherichia coli. Microbiol. Rev.* **57**, 862–952 (1993).
8. Eisen, J.A., Heidelberg, J.F., White, O. & Salzberg, S.L. Evidence for symmetric chromosomal inversions around the replication origin in bacteria. *Genome Biol.* **1**, 0111.1–0111.9 (2000).
9. Kropinski, A.M. Sequence of the genome of the temperate, serotype-converting, *Pseudomonas aeruginosa* bacteriophage D3. *J. Bacteriol.* **182**, 6066–6074 (2000).
10. Juhala, R.J. *et al.* Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. *J. Mol. Biol.* **299**, 27–51 (2000).
11. Masignani, V. *et al.* Mu-like prophage in serogroup B *Neisseria meningitidis* coding for surface-exposed antigens. *Infect. Immun.* **69**, 2580–2588 (2001).
12. Richardson, D.J. Bacterial respiration: a flexible process for a changing environment. *Microbiology* **146**, 551–571 (2000).
13. Lower, S.K., Hochella, M.F., Jr. & Beveridge, T.J. Bacterial recognition of mineral surfaces: nanoscale interactions between *Shewanella* and alpha-FeOOH. *Science* **292**, 1360–1363 (2001).
14. Myers, C.R. & Myers, J.M. Localization of cytochromes to the outer membrane of anaerobically grown *Shewanella putrefaciens* MR-1. *J. Bacteriol.* **174**, 3429–3438 (1992).
15. Beliaev, A.S. & Saffarini, D.A. *Shewanella putrefaciens* mtrB encodes an outer membrane protein required for Fe(III) and Mn(IV) reduction. *J. Bacteriol.* **180**, 6292–6297 (1998).
16. Myers, J.M. & Myers, C.R. Role for outer membrane cytochromes OmcA and OmcB of *Shewanella putrefaciens* MR-1 in reduction of manganese dioxide. *Appl. Environ. Microbiol.* **67**, 260–269 (2001).
17. Gross, R., Simon, J. & Kroger, A. Periplasmic methacrylate reductase activity in *Wolinella succinogenes. Arch. Microbiol.* **176**, 310–313 (2001).
18. Myers, C.R. & Myers, J.M. Cloning and sequence of cymA, a gene encoding a tetraheme cytochrome *c* required for reduction of iron(III), fumarate, and nitrate by *Shewanella putrefaciens* MR-1. *J. Bacteriol.* **179**, 1143–1152 (1997).
19. Brugna, M. *et al.* Kinetics and interaction studies between cytochrome *c*3 and Fe-only hydrogenase from *Desulfovibrio vulgaris* Hildenborough. *Proteins* **33**, 590–600 (1998).
20. Pohorelic, B.K. *et al.* Effects of deletion of genes encoding Fe-only hydrogenase of *Desulfovibrio vulgaris* Hildenborough on hydrogen and lactate metabolism. *J. Bacteriol.* **184**, 679–686 (2002).
21. Scott, J.H. & Nealson, K.H. A biochemical study of the intermediary carbon metabolism of *Shewanella putrefaciens. J. Bacteriol.* **176**, 3408–3411 (1994).
22. Meinild, A.K., Klaerke, D.A. & Zeuthen, T. Bidirectional water fluxes and specificity for small hydrophilic molecules in aquaporins 0-5. *J. Biol. Chem.* **273**, 32446–32451 (1998).
23. Bellmann, A. *et al.* Expression control and specificity of the basic amino acid exporter LysE of *Corynebacterium glutamicum. Microbiology* **147**, 1765–1774 (2001).
24. Stover, C.K. *et al.* Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. *Nature* **406**, 959–964 (2000).
25. Galibert, F. *et al.* The composite genome of the legume symbiont *Sinorhizobium meliloti. Science* **293**, 668–672 (2001).
26. Perez-Rueda, E. & Collado-Vides, J. The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. *Nucleic Acids Res.* **28**, 1838–1847 (2000).
27. Bagge, D., Hjelm, M., Johansen, C., Huber, I. & Gram, L. *Shewanella putrefaciens* adhesion and biofilm formation on food processing surfaces. *Appl. Environ. Microbiol.* **67**, 2319–2325 (2001).
28. Little, B. *et al.* The role of biomineralization in microbiologically influenced corrosion. *Biodegradation* **9**, 1–10 (1998).
29. Watnick, P.I., Fullner, K.J. & Kolter, R. A role for the mannose-sensitive hemagglutinin in biofilm formation by *Vibrio cholerae* El Tor. *J. Bacteriol.* **181**, 3606–3609 (1999).
30. Chiavelli, D.A., Marsh, J.W. & Taylor, R.K. The mannose-sensitive hemagglutinin of *Vibrio cholerae* promotes adherence to zooplankton. *Appl. Environ. Microbiol.* **67**, 3220–3225 (2001).
31. Chen, Y.S. *et al.* Skin and soft-tissue manifestations of *Shewanella putrefaciens* infection. *Clin. Infect. Dis.* **25**, 225–229 (1997).
32. Khashe, S. & Janda, J.M. Biochemical and pathogenic properties of *Shewanella alga* and *Shewanella putrefaciens. J. Clin. Microbiol.* **36**, 783–787 (1998).
33. Reuven, N.B., Arad, G., Maor-Shoshani, A. & Livneh, Z. The mutagenesis protein UmuC is a DNA polymerase activated by UmuD', RecA, and SSB and is specialized for translesion replication. *J. Biol. Chem.* **274**, 31763–31766 (1999).
34. Tang, M. *et al.* UmuD'(2)C is an error-prone DNA polymerase, *Escherichia coli* pol V. *Proc. Natl. Acad. Sci. USA* **96**, 8919–8924 (1999).
35. Sutton, G.G., White, O., Adams, M.D. & Kerlavage, A.R. TIGR assembler: a new tool for assembling large shotgun sequencing projects. *Genome Seq. Technol.* **1**, 9–19 (1995).
36. Tettelin, H., Radune, D., Kasif, S., Khouri, H. & Salzberg, S.L. Optimized multiplex PCR: efficiently closing a whole-genome shotgun sequencing project. *Genomics* **62**, 500–507 (1999).
37. Lobry, J.R. Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol. Biol. Evol.* **13**, 660–665 (1996).
38. Salzberg, S.L., Delcher, A.L., Kasif, S. & White, O. Microbial gene identification using interpolated Markov models. *Nucleic Acids Res.* **26**, 544–548 (1998).
39. Claros, M.G. & von Heijne, G. TopPred II: an improved software for membrane protein structure predictions. *Comput. Appl. Biosci.* **10**, 685–686 (1994).
40. Bateman, A. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **28**, 263–266 (2000).
41. Haft, D.H. *et al.* TIGRFAMs: a protein family resource for the functional identification of proteins. *Nucleic Acids Res.* **29**, 41–43 (2001).