Large scale comparative phenotypic and genomic analyses reveal ecological preferences of *Shewanella* species and identify metabolic pathways conserved at genus level

Jorge L. M. Rodrigues[1*], Margrethe H. Serres[2], and James M. Tiedje[3]

[1]Department of Biology, University of Texas, Arlington, TX 76019, [2]Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, MA 02543, and
[3]Center for Microbial Ecology, Michigan State University, 540 Plant and Soil Sciences Building, East Lansing, MI 48824.

Running Title: Genotypic and Phenotypic Comparisons in Shewanella species

Section: Evolutionary and Genomic Microbiology

* Corresponding author: University of Texas - Arlington
Department of Biology
501 S. Nedderman Drive, Arlington, TX, 76019.
Tel: (817) 272-1124
Fax: (817) 272-2855
E-mail: Jorge@uta.edu

ABSTRACT

The use of comparative genomics among different microbiological species has increased substantially as sequence technologies become more affordable. However, efforts to fully link a genotype to its phenotype remain limited to the development of one mutant at the time. In this study, we provide a high throughput alternative to this limiting step by coupling comparative genomics to phenotype arrays for five sequenced *Shewanella* strains. Positive phenotypes were obtained for 441 nutrients (C, N, P, and S sources), with N-based compounds being the most utilized for all strains. Many genes and pathways predicted by genome analyses were confirmed with the comparative phenotype assay, and three degradation pathways believed to be missing in *Shewanella* were confirmed. A number of previously unknown gene products were predicted to be part of pathways or to have a function, expanding the number of gene targets for future genetic analyses. Ecologically, the comparative high throughput phenotype analysis provided insights into niche specialization within the five different strains. For example, *Shewanella amazonensis* strain SB2B, isolated from the Amazon River delta, was capable of utilizing 60 C compounds, whereas *Shewanella* sp. strain W3-18-1, from the deep marine sediment, utilized only 25 of them. In spite of the large number of nutrient sources yielding positive results, our study indicated that except for the N-sources they were not sufficiently informative to predict growth phenotypes from increasing evolutionary distances. Our results indicate the importance of phenotypic evaluation for confirming genome predictions. This strategy will accelerate the functional discovery of genes and provide an ecological framework for microbial genome sequencing projects.

INTRODUCTION

The *Shewanella* genus is composed of facultative anaerobic bacteria known for their distinctive capability of utilizing a variety of electron acceptors such as NO_3^-, U, Cr, Tc, Pu, and nitroaromatic compounds (13). Members of this genus have also being regarded for their role as drivers of global biogeochemical cycles of C, N, and S in redox interfaces of marine environments (3, 28).

Being found in different environments such as salt and fresh waters, sediments, and subsurface formations, it is not surprising that the genus *Shewanella* developed its hallmark respiratory capability of utilizing many different electron acceptors. This diversity of respiratory phenotypes is a reflection of the genetic makeup carried by members of this genus. The sequenced genome of the *Shewanella oneidensis* strain MR-1 (MR-1) showed a large percentage of genes dedicated to the cell's electron transport system including cytochromes, reductases, iron-sulfur proteins, and quinones (12). As revealed by the genome sequencing of 22 additional *Shewanella* species and strains of the same species (10), the genetic diversity carried out by this genus is significant, with less than half of the genes being shared among ten of the sequenced *Shewanella* genomes (21).

Recently, several studies have used comparative genomics to systematize the genomic content into two groups: the core genome comprised of genes present in all strains and the accessory genome consisting of unique or strain-specific genes (21, 39). This approach has allowed for putative determination of the total number of genes and operons that might be involved in the ecological fitness of strains subjected to a specific environmental condition (18, 19, 24, 33). It is, however, less clear how this genomic diversity is translated into phenotypic traits and what their implications are to the ecological success of the species. Traditionally, a particular genotype has been linked to a phenotype through the development and characterization of mutants (23). Based on

the 862 genes (19.2%) that still remain to be characterized in the genome of the model microorganism *Escherichia coli* strain K12 (36), the above procedure is not only a labor intensive, but also a time-consuming activity.

High-throughput phenotype arrays can be used as an alternative approach to expedite the functional characterization of genes. The Biolog assay uses tetrazolium violet to monitor cell respiration, assuming that oxidation of the nutrient source will lead to respiration and, hence to purple dye formation (1). High-throughput phenotype arrays have been extensively used to characterize knockout mutants of single microorganisms (15, 43), but have yet to be tested for comparative analysis of phenotypes in light of genome sequence data (2).

In this study, we sought to gain access to the ecology of members of the *Shewanella* genus through a large-scale comparative analysis of phenotypes. Here, we took advantage of five fully sequenced *Shewanella* genomes and compared them to high-throughput phenotype arrays containing 561 nutrient sources. We established genotype-phenotype relationships, expanded the number of genes associated with specific phenotypes, and showed there is a limit in predicting phenotypes with increased phylogenetic distances.

MATERIAL AND METHODS

Strains used in this study. Microorganisms and their genome accession numbers (in parenthesis) used in this study were *Shewanella oneidensis* strain MR-1 (AE014299, AE14300), *Shewanella* sp. strain MR-4 (CP000446), *Shewanella* sp. strain MR-7 (CP000444, CP000445), *Shewanella* sp. strain W3-18-1 (CP000503), and *Shewanella amazonensis* strain SB2B (CP000507). Strain selection was based on the following: (1) representation of an evolutionary gradient with strains of the same species and different

species, (2) genomes that were curated manually. A description of habitat conditions at the time of sampling is presented in Table 1.

High-throughput phenotypic comparisons. Phenotype microarray assays were performed at Biolog Inc. (Hayward, CA) as previously described (1). Briefly, *Shewanella* strains were grown on R2A plates and incubated overnight at 22^{o}C. Colony swabs were used to suspend cells in IF-0 GN medium and diluted until 85% of transmittance of the cell suspension was achieved. A volume of 100 μl was added to each plate well. Plates were incubated in the OmniLog reader at 22^{o}C. Monitoring of color change was recorded for 48h for all plates. Only positive results for three replicate plates were reported. For the purpose of this work, a nutrient was defined as being used as the sole source for the element, e.g. when glycine was defined as a N source, another C source was provided in the well to be oxidized, providing electrons for the dye reduction.

Gene analysis and pathway reconstructions. Protein sequences encoded by the five *Shewanella* genomes were analyzed and compared. Function predictions for the gene products were obtained from the *Shewanella* Knowledgebase (17). Many of these predictions resulted from careful manual curation as previously described (17, 35, 37). Further, the gene products were assigned to metabolic pathways according to the MetaCyc schema (5) and the primary literature. In addition, we made use of the orthologous relationships previously determined for ten *Shewanella* strains, including the five strains used in our analysis (21). The sets of orthologs were identified based on their sequence similarity and genome neighborhoods and included genes present in one or more of the *Shewanella* genomes. The cut-off criteria for gene presence or absence was compiled from three different methods: (1) protein-protein pair-wise reciprocal BLAST-P, (2) Pair-wise alignments with PAM score of 100, and (3) reciprocal tBLAST-N (21). The ortholog table for the five *Shewanella* strains is available in Supplementary Material ST2.

Metabolic pathway predictions and the complete *Shewanella* ortholog dataset are also available through the *Shewanella* Knowledgebase (17).

Phenotypic and genome clustering analyses. Data analysis was performed using the kinetic and parametric modules of the software OmniLog v1.2. Tested compounds yielding negative or not reproducible results for all *Shewanella* strains were removed from the analysis. Results from three replicates were averaged and subtracted from control intensity value. Next, substrates with intensity values below the threshold of 10 were discarded as a conservative measure for respiration. Phenotypic profiles were converted into a matrix using the following parameters: 1 for values below 50, 2 for values between 50 and 100, and 3 for values above 100. Clustering was performed with the Cluster 3.0 software (6), using the single linkage and Euclidean distance.

Genomic clustering for all five strains was calculated based on gene orthologs identified in the genomes. When a gene ortholog was observed in the genome, it was scored as present (one in the matrix), whereas a gene not observed was scored as absent (zero in the matrix). Clustering analysis was carried out using a single linkage scheme and Euclidean distance. Pair-wise average nucleotide identity (%ANI) values between genomes were calculated according to Konstantinidis and Tiedje (20).

RESULTS

Large scale comparative phenotypic analysis. Five *Shewanella* strains were tested for metabolic diversity with 561 sources of carbon (C), nitrogen (N), sulfur (S), or phosphorus (P). Positive phenotypes for at least one strain were observed for a total of 441 sources (Supplemental Material SF1). Among the different substrates tested, 55 (92%) of the P, 293 (77%) of the N, 21 (58%) of S, and 67 (35%) of the C sources were utilized by one or more of the strains. The majority of the nutrient sources were used by

2-4 of the strains with only 15% and 16% being utilized by all strains or by one of the strains, respectively.

Carbon utilization patterns. Only 15 of the 190 carbon sources tested were utilized by all five strains (Fig. 1A). Growth on six of these substrates (L-lactate, adenosine, inosine, pyruvate, and N-acetyl-D-glucosamine) has been confirmed in other studies (7, 31, 42). The remaining compounds (2'-deoxyadenosine, methyl-pyruvate, uridine, Tween 80, Tween 40, Tween 20, gelatin, and three dipeptides) have not been shown to support growth of the five *Shewanella* strains and remain targets for further experimental studies. Hydrolyses of gelatin and Tween detergents have been previously reported for *S. loihica* strain PV-4 (11) and *S. affinis* strains KMM 3586 and KMM 3587 (14), respectively, and it is possible that these phenotypes are shared by the entire genus. The three di-peptides utilized by all strains had an N-terminal glycine (Gly-Glu, Gly-Pro, Gly-Asp). A fourth dipeptide with a C-terminal glycine (Ala-Gly) was only degraded by SB2B suggesting the need for a separate transporter or peptidase to utilize this dipeptide. The majority of the substrates degraded by all strains enter the central metabolism at the 2 or 3 carbon compound level (Fig. 2). This agrees with earlier predictions made in an analysis of the *S. oneidensis* MR-1 (MR-1) genome (37) where many of these pathways were outlined. It was also noted that MR-1 contained a smaller number of iso-enzymes for the degradation of 5 and 6 carbon compounds compared to *Escherichia coli*, an organism that is capable of using a plethora of 5 and 6 carbon compounds. Instead MR-1 was found to encode iso-enzymes for the utilization of 3 carbon compounds, i.e. 3 glyceraldehyde dehydrogenases. *Escherichia coli* also contains over 40 PTS systems for import of 4 – 6 carbon carbohydrates while MR-1 only has one such system. These trends hold up for the five *Shewanellas* included in this study. An additional mannose

specific PTS system was found on a mobile island in *Shewanella* sp. strain W3-18-1 (W3-18-1 from here on).

We observed a variation in the growth phenotype for sugar and polysaccharide utilization where each substrate was degraded by one to three strains. Several of the sugars tested included dimers or polymers of glucose subunits (i.e. α-, β-, and γ-cyclodextrin, dextrin, maltose, maltotriose, and sucrose). Since all of the glucose multimers were utilized by *Shewanella amazonensis* SB2B (SB2B), *Shewanella* sp. MR-4 (MR-4) and *Shewanella* sp. MR-7 (MR-7), and not by W3-18-1 or MR-1, it is likely that some of the same enzymes (pathways) and transporters are shared by different strains and involved in the degradation of the above compounds.

SB2B showed the most phenotypic versatility for carbon sources when compared to the other species. This strain was able to utilize 60 different substrates, while strain W3-18-1 used only 25. The limited substrate utilization by W3-18-1 includes its inability to utilize the many glucose-based multimers as well as several amino acids. In fact, none of the single amino acids tested were degraded by this strain. When the two most evolutionarily related strains isolated from the Black Sea water column were compared, namely strain MR-4 and MR-7, their carbon utilization pattern showed larger than expected differences. Both strains shared the ability to metabolize 31 carbon substrates with an additional 13 sources being solely utilized by strain MR-4 while strain MR-7 used another 6 substrates. MR-4 and MR-7 have an average nucleotide identity (ANI) of 97% suggesting that they belong to the same species.

Nitrogen utilization patterns. A total of 50 nitrogen sources were utilized by all five strains (a subset of sources is presented on Fig. 1B, Supplementary Material SF1). These substrates were all dipeptides with pronounced preference for amino acids with polar uncharged side chains: serine, threonine, asparagine, and glutamine. Amino acids

with electrically charged side chains (positive or negative) or hydrophobic side chains were of limited use for these strains, with exception of alanine and leucine. When analyzing for nitrogen substrates utilized by all *Shewanella* strains but one, the number of substrates increased to 126, with the inclusion of purine bases such as xanthine and adenine.

The Black Sea isolates, MR-4 and MR-7, were able to utilize 285 and 242 substrates, respectively; followed by strain MR-1 with a pattern of substrate utilization corresponding to 203 nitrogen sources. Inversely proportional to the carbon utilization profile, strain SB2B had the lowest number of positive Biolog phenotypes (81) for nitrogen sources. The low N-source diversity displayed by SB2B reflects its inability to utilize 194 (71%) of the dipeptides and tripeptides, and these made up 89% of the N-sources tested. Furthermore, the only amino acids or amines utilized by SB2B were L-tyrosine and N-acetyl glucosamine.

Phosphorus and sulfur utilization patterns. Adenosine 5'-monophosphate (AMP) was the sole phosphorus source utilized by all five strains. Of the 55 phosphorus compounds degraded by *Shewanella*, 39 were only utilized by MR-7 and MR-1. In fact, these two strains yielded a larger number of positive phenotypes for both the phosphate and sulfur sources tested. MR-7 was able to utilize 52 phosphate- and 15 sulfur- sources, and strain MR-1 used 48 and 19 substrates, respectively (Fig. 1C and 1D). The remaining strains degraded a very limited number of phosphate substrates besides AMP; W3-18-1 utilized TMP, AMP, GMP, CMP, UMP, pyrophosphate; while MR-4 utilized phosphogluconic acid and carbamylphosphate. Strain SB2B was able to utilize only TMP in addition to AMP (Fig. 1C). While W-3-18-1 tested positive for 3 of the sulfur compounds (L-cysteine sulfinic acid, glutathione, Glu-Met), MR-4 and SB2B did not utilize any sulfur compounds under the conditions tested (Fig. 1D).

Smaller scale comparisons of biolog phenotypes. We sought to explain some of the Biolog phenotypes from known phenotypes, genome contents, and orthologous relationships of the encoded genes (Supplementary Material ST1). The utilization of N-acetyl-D-glucosamine as a carbon source agrees with the published literature (42). Lactate and pyruvate degradation via acetyl-CoA to acetate or the TCA cycle has been shown with growth experiments (31). Also, the use of nucleotides and nucleosides, including adenosine and inosine, has previously been shown for MR-1 (7), and can be inferred for the other strains based on the presence of genes encoding for the degradation enzymes.

According to the high throughput phenotype assay, MR-1 was the only strain unable to utilize the C4 dicarboxylates L-malate, succinate, and fumarate. When these results were observed in light of the genome sequences, we found that the gene encoding the dicarboxylate carrier AbgT was absent from MR-1, but present in the other strains (MR4_3833/MR7_3926/W3181_3971/Sama_3559). Furthermore, an oxaloacetate decarboxylase was absent in MR-1 and present in the other strains (MR4_2984-7/MR7_3066-9/W3181_3133-6/Sama_1054-1). This enzyme decarboxylates oxaloacetate to pyruvate, and may be involved in the conversion of 4C dicarboxylic acids via pyruvate and the gluconeogenesis pathway to 5C and 6C essential metabolic intermediates. MR-1 also tested negative for L-arabinose utilization. An arabinose (and arabinoside) metabolism locus, including a TonB arabinose receptor, an ABC arabinose transporter, and enzymes degrading L-arabinose to xylose-5-phosphate, is found in the genomes of the strains MR-4 (MR4_1977-2001), MR-7 (MR7_1997-1974), and W3-18-1 (W3181_1944-1966), but not in MR-1, also observed by (34). Strain SB2B, also with an L-arabinose utilizing phenotype, did not contain the above locus and may use another hitherto unknown pathway to degrade arabinose. Two

additional sugars, D-mannose and D-fructose, were utilized by MR-7 and SB2B. A locus containing genes with similarity to the mannose utilization pathway is present in these organisms (MR7_3383-8/Sama_0565-60) with a second mannose transporter and utilization locus detected in SB2B (Sama_0303-4). While fructose can be degraded via the mannose degradation pathway, a transporter specific for fructose was not identified, agreeing with previous results (34).

In another example, N-acetyl-D-galactosamine was degraded by MR-4, MR-7, and SB2B. This phenotype agrees with the presence of the *aga* operon for uptake and degradation of N-acetyl-D-galactosamine in these strains (MR4_2530-6/MR7_2597-2603/Sama_1199-3). The Biolog assay also showed degradation of citric acid by the same three strains. Based on a recent paper describing citric acid utilization by *Corynebacterium glutamicum* (4), we identified genes for a citrate sensing two-component regulator and a three-component citrate transporter of the tricarboxylate transporter (TTT) family in MR-4 and MR-7 (MR4_3099-5/MR7_0873-7). Genes encoding these products were not found in the SB2B genome, suggesting that another citrate utilization path was taken in this organism.

Expanding testable predictions. The high throughput genotype-phenotype comparison among different strains allowed us to make targeted predictions for further experimental analysis. For example, degradation of Tween 80 has been shown to involve an outer membrane esterase in *Pseudomonas aeruginosa* (29). While no close homolog to the *P. aeruginosa* enzyme was found in the 22 *Shewanella* genome sequences, we did identify candidate genes present in the Tween degrading strains. Specifically, three candidate genes encoding for a surface expressed lipase (SO_2934/MR4_1469/MR7_1535/W3181_1613/Sama_2120), an outer membrane phospholipase (SO_0428/MR4_0432/MR7_3595/W3181_0530/Sama_0379) and a cold-

adapted lipase (SO_1994/MR4_2269/MR7_2341/W3181_1692/Sama_2029) were conserved among the strains. The latter gene encodes for a protein with sequence similiarity to a cold-active lipase isolated from a deep-sea sediment metagenome (16). We also identified two proton-dependent (oligo)peptide transporters (SO_0002/MR4_3938/MR7_4030/W3181_4066/Sama_2411 and SO_3195/MR4_1313/MR7_1380/W3181_1459/Sama_2266) that might be associated with the use of Gly-Glu, Gly-Pro, or Gly-Asp as carbon sources by *Shewanella*.

Phenotype assays yielded positive results for strains MR-1, MR-4, and SB2B tested with the heteropolysaccharide pectin. We searched the genomes for genes present in these three strains and absent in MR-7 and W3-18-1 genomes, and noted an outer membrane Ton-B dependent receptor (SO_1822/MR4_2467/Sama_1252). The involvement of TonB receptors in the uptake of sugars and their derivatives has been shown for MR-1 (42). We also identified a gene coding for a sugar-binding periplasmic protein that may be linked to the degradation of dextrins, maltose, and sucrose in MR-4 (MR4_0355), MR-7 (MR7_3671), and SB2B (Sama_3287).

Linking genotypic and phenotypic changes. To gain insight into the phenotypic variation and its relationship to the genome, we plotted the evolutionary distance (defined as the % average nucleotide identity (ANI)) among the five strains and the percent of identical phenotypes (Fig. 3**,** Supplementary Materials ST2 and ST3). The % rRNA similarity vs. % ANI regression line was selected to represent a conserved trait (correlation coefficient R^2 = 0.89). The % conservation for orthologs was also included, and it decreased more sharply with as the evolutionary distance increased (R^2 = 0.76). We also observed a correlation between N source utilization and evolutionary distance (R^2 = 0.86) (Fig 3). However, no such correlation was detected for the C source utilization (R^2 = 0.05). Based on the clustering of the P and S utilization phenotypes (Fig.

1C and 1D), it was evident that neither of these phenotypes would be conserved according to their evolutionary distance. Because the number of 294 phenotypes for N sources was significantly higher than the other compounds (67 C, 55 P, and 21 S), we selected sets of 60 N sources randomly to calculate the % similarity vs. % ANI in order to test for sample size bias. These datasets also showed a decrease in % similarity with increasing evolutionary distance suggesting that the N source phenotypic results are not dependent on the number of chemicals compared (data not shown).

In order to test whether the observed phenotypic diversity paralleled the gene content diversity, we performed clustering analyses with each one of the datasets (Fig. 4 A and B). Similar clustering branches were observed for closely related strains MR-4 and MR-7, but the same correspondence was not seen with greater evolutionary distances. The phenotypic diversity for closely related strains was much greater (longer branch lengths in Fig. 4B) than their gene content diversity (shorter lengths, Fig. 4A).

DISCUSSION

Historically, the establishment of a direct link between genotype and its phenotype has been performed through the study of mutants. However, the generation of thousands of mutants can be costly and laborious. Here, we explored a different avenue for developing such a link by making use of hundreds of Biolog phenotypes and by taking advantage of predicted genes and pathways through comparative genomics. Owing to the high phenotypic diversity within the *Shewanella* and availability of many sequenced genomes, we selected five isolates representing a gradient of evolutionary distances within the genus for this study.

The phenotype assay is not a measure of growth on the different substrates, but rather an indication of respiration when a nutrient source is provided. The presence of a transport system and catabolic pathway for a specific chemical compound would lead to

the production of NADH that then reduces a tetrazolium dye (1). Measuring respiration has an advantage over growth assays as the microbial cell metabolic response to a chemical compound is detected, even when growth support is not observed. Hence, in our study respiration is a reflection of the ability to utilize different substrates. Likewise, we used five genome sequences and their gene content to reflect the genetic potential of the *Shewanella* genus, without taking into account the many regulatory levels that determine whether a gene is expressed or whether the protein is synthesized and active. By analyzing the predicted functions of the encoded proteins as well as their presence or absence relative to a given phenotype, we were able to successfully confirm a series of genome predictions towards a specific metabolism and to identify a series of target genes for further genetic analysis.

Because the comparative phenotype analysis dealt with different numbers of C, N, P, and S sources, we normalized the utilization profile based on the percent of positive results over the total of number compounds tested. The emerging pattern of phenotypes suggested that *Shewanella* is capable of utilizing a variety of N compounds, (77% of the tested N sources) including several amino acids and dipeptides (Fig. 1). The particular ability to grow on several amino acids was previously observed for strain MR-1 (32), but the use of amino acids as a sole N source was not extensively tested in the *Shewanella* genus. This preference for N sources might explain the isolation of different species of this genus from a variety of environments such as dead fish (9), chicken breast (9), clinical samples (41), and marine sediments (26, 27, 40), where DNA and polypeptides are known to accumulate. Despite the large number of N sources tested, only 50, mainly dipeptides (90%), were utilized by all five strains. N-acetyl-D-glucosamine was used by these five strains as both N and as C sources (Fig. 2). While N-acetyl-D-glucosamine previously has been characterized as a carbon and energy source for the *Shewanellae* (42), its use as an N source has not been reported. We

identified 15 compounds that could be used both as N and C source by at least one of the strains. Interestingly, while SB2B was the only strain able to utilize L-threonine as a C source, the other four strains utilized it as an N source. Overall, there was a decrease in the number of N compounds shared among the strains as the ANI values increased. Fewer shared compounds might indicate a decrease in ecological niche overlap once speciation has taken place. To address this possibility, first we compared the N utilization profiles for the most closely related isolates, namely MR-4 and MR-7. These *Shewanellae*, with 97.05% ANI, are strains of the same species (21) and shared the largest number of N compounds (227). A comparison of these two strains to *S. oneidensis* showed the number of shared N sources decreasing to 187 for MR-4 vs. MR-1 (87.74% ANI) and 175 for MR-7 vs. MR-1 (87.70% ANI). Strains MR-4 and MR-7 came from the same environment, where niches have more similar conditions (Table 1). Accordingly, strain SB2B, which came from an environment differing significantly from the others, had the least number of overlaps in N source utilization with 81 positive phenotypes when compared to the other strains. None of these N sources were uniquely utilized by SB2B. However, upon inspection of the SB2B orthologs (Supplementary Material ST1) we identified a stretch of six genes (Sama_0018 to Sama_0023) absent from the other four strains that could be linked to nitrogen metabolism. Two of these genes, Sama_0022 and Sama_0023, encode a predicted urea carboxylase cleaving urea-1-carboxylate to CO_2 and ammonia, with the possibility of the later being assimilated as a nitrogen source. The other co-localized genes included two membrane proteins and a LamB family protein, but their specific role in nitrogen metabolism remains unknown.

When attempting to make in depth connections between the ecology and genomic information of microorganisms, environmental metadata is generally missing or not fully described. Although, not different in our study, the limited habitat description for

the *Shewanella* isolates in congruency with the large-scale comparative phenotype approach provided insights into the functional strategies devised by the five strains in their environment. *Shewanella amazonensis* strain SB2B was capable of utilizing a full range of C compounds, 60 out of the 67 tested, varying from well-known growth substrates for *Shewanella* such as pyruvate and lactate to previously unknown compounds such as the Tween series and laminarin. As SB2B was isolated from the coastal muddy waters of the Amazon River, a low salinity environment (Table 1), it is likely exposed to a full plethora of C compounds washed out from the Amazon forest (25). In contrast with the SB2B C utilization profile, strain W3-18-1 yielded positive results for only 25 C compounds. This might be a reflection of the W3-18-1 genetic adaptation to nutrient limitation imposed in the marine sediments of the Pacific Ocean (38).

Together, the five *Shewanella* strains selected for this study encode a total of 6,790 genes. Approximately 39% of the genes are present in all strains while the unique genes comprise 21 - 40% of the genomes. On average, 550 genes per strain were identified as unique, a number in accordance with the previously reported 468 unique genes when ten *Shewanella* genomes were compared (21). In our high-throughput genotype-phenotype analyses, we were able to link a few of these unique genes to specific phenotypes. In addition, pathways believed to be missing in these *Shewanella* strains (e.g. degradation of galactose, ribose, trehalose) were confirmed using the phenotype assay.

It is noteworthy that not all observed phenotypes were explained through the use of comparative genomics. For example MR-7 was the only strain unable to utilize acetate and thymidine as C sources. From the ortholog table, we identified only four proteins that were missing in MR-7 and present in the other four strains; two lipopolysacharides biosynthesis genes and a transcriptional regulator with an adjacent

multidrug transporter. It is not evident how any of these functions could explain the acetic acid and thymidine growth phenotypes. Likewise W3-18-1 is the only strain unable to degrade L-glutamate and L-glutamine, yet no proteins for transport or degradation of either compound could be found among the 137 genes absent from W3-18-1 and present in the other four strains. Glutamate transporters as well as degrading enzymes (i.e. glutamate decarboxylase, glutamate racemase, and mutase) were detected in all five genome sequences. We observed a glutamate racemase (W3181_0949) and a glutamine amidotransferase (W3181_0925) among proteins unique to W3-18-1, suggesting that this organism may have a different strategy for metabolizing glutamine and glutamate. Two possible reasons might explain the above differences. First, there is a large fraction of genes without function predictions (37). Presently 21% of the protein coding genes in the five strains remains of unknown function. Novel enzyme variants may also encode some of the "missing" activities as was uncovered in studies focused on revaluating the genomic content of *S. oneidensis* strain MR-1 (31, 42). Genome reconstruction allowed identification of novel enzymes for both degradation of L-lactate (lactate dehydrogenase, *LldEGF*) and N-acetyl-glucosamine (glucosamine-6-phosphate deaminase, *NagB* and N-acetylglucosamine kinase, *NagK*). Second, regulation of expression or activity at the gene or protein level affects whether a phenotype is exhibited, leading to weak phenotypic expression or a false negative in our analysis.

In order to validate our high-throughput phenotype approach, we compared our results to the recent work describing the reconstruction of *Shewanella* carbohydrate utilization pathways (34). Rodionov *et al*. determined the C degradation pathways for 19 *Shewanella* genomes using orthologous relationships, regulon predictions, and growth phenotypes. The increased number of genomes as well as the inclusion of growth phenotypes and regulatory networks allowed for thorough analysis of a subset of C-sources compared to our analyses. Growth phenotype data were obtained for eight of

the sugars (arabinose, cellobiose, fructose, glucose, mannose, maltose, N-acetylglucosamine, sucrose) tested in our high-throughput phenotype assay. A comparison between both studies indicated an agreement of 75% of the phenotypes (W3-18-1 was not included in their growth studies). Although the dataset used for comparison contains only a small number of sugars, we did not detect bias with regards to a particular substrate or strain, attesting the validity of our approach. We also found agreement in the utilization of amino acids by MR-1 between the Biolog phenotype and growth (30) for six of the seven amino acids tested in both assays.

The five strains selected for this study had varying degrees of relatedness as measured by the ANI. The two closest strains in the dataset, namely MR-4 and MR-7, shared the highest fraction of genes (93%) relative to the other genome pairs (e.g. from 80% between MR-4 and MR-1 to 72% between MR-4 and SB2B). However, this high sequence similarity between MR-4 and MR-7 was not reflected throughout the phenotypic results (Fig. 4). These strains were among the most similar for the carbon sources (88%) and nitrogen sources (77% similarity), but they were unable to both utilize any of the 21 S sources tested and could only both use one of the 55 P sources tested. Rather, MR-1 and MR-7 were able to utilize 87% and 95% of the P sources and 90% and 76% of the S sources, respectively. The remaining strains only utilized 14% or less of the P and S compounds. Considering that strains MR-1 and MR-7 have a more distant evolutionary history (87.70% ANI), their similar growth phenotypes for P and S sources might reflect the importance of whole cell networks and expression on regulating ecological fitness under certain environmental conditions. High-throughput phenotypic arrays were not designed to capture the interplay between whole cell gene expression and environmental conditions. While it is clear that the present study succeeds in deciphering novel phenotypic traits from a combination of high-throughput

phenotypes and comparative genomics, future work on linking the evolution of the phenotypes to environment fitness will need to include genome wide analyses.

The present study addressed the importance of bridging genomes and their counterpart phenotypes. The strategy devised here could accelerate the functional discovery of genes and provide an ecological framework for genome sequencing projects.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to the following members of the *Shewanella* Federation: Jim Fredrickson, Kostas Konstantinidis, Ken Nealson, and Margie Romine.

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract numbers DE-AC02-05CH11231 and DE-FG02-08ER64511.

REFERENCES

1. **Bochner, B.R.** 2008. Global phenotypic characterization of bacteria. FEMS Microbiol. Rev. **33:**191-205.
2. **Bochner, B.R.** 2009. New technologies to assess genotype-phenotype relationships. Nature Rev. Genetics **4:**309-314.
3. **Brettar, I., E.R. Moore, and M.G. Hofle.** 2001. Phylogeny and abundance of novel denitrifying bacteria isolated from the water column of the Central Baltic sea. Microb. Ecol. **42:**295-305
4. **Brocker, M., S. Schaffer, C. Mack, and M. Bott.** 2009. Citrate utilization by *Corynebacterium glutamicum* is controlled by the CitAB two-component system through positive regulation of the citrate transport genes *citH* and *tctCBA.* J. Bact. **191:**3869-3880.
5. **Caspi R., H. Foerster, C.A. Fulcher, P. Kaipa, M. Krummenacher, M. Latendresse, S. Paely, S.Y. Rhee, A.G. Shearer, C. Tissier, T.C. Walk, P. Zhang, and P.D. Karp.** 2008. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res. **36:**D623-631.
6. **de Hoon, I.J., S. Imoto, J. Nolan, S. Miyano.** 2004. Open source clustering software. Bioinformatics **12:**1453-1454.
7. **Driscoll, M.E., M.F. Romine, F.S. Jhun, M.H. Serres, L.A. McCue, A.S. Beliaev, J.K. Fredrickson, T.S. Gadner.** 2007. Identification of diverse carbon utilization pathways in *Shewanella oneidensis* MR-1 via expression profiling. Genome Inform. **18:**287-298.
8. **Eisenbeis, S., S. Lohmiller, M. Valdebenito, S. Leicht, and V. Braun.** 2008. NagA-dependent uptake of *N*-acetyl-glucosamine and *N*-acetyl-chitin

oligosaccharides across the outer membrane of *Caulobacter crescentus.* J. Bacteriol. **190:**5230-5238.

9. **Fonnesbech-Vogel, B., K. Jorgernsen, H. Christensen, J.E. Olsen, and L. Gram.** 1997. Differentiation of *Shewanella putrefasciens* and *Shewanella alga* on the basis of whole-cell protein profiles, ribotyping, phenotypic characterization and 16S rRNA gene sequence analysis. Appl. Environ. Microbiol. **73:**2189-2199.
10. **Fredrickson, J.K., M.F. Romine, A.S. Beliaev, J.M. Auchtung, M.E. Driscoll, T.S. Gardner, K.H. Nealson, A.L. Osterman, G. Pinchuk, J.L. Reed, D.A. Rodionov, J.L.M. Rodrigues, D.A. Saffarini, M.H. Serres, A.M. Spormann, I.B. Zhulin, and J.M. Tiedje.** 2008. Towards environmental systems biology of *Shewanella*. Nature Rev. Microbiol. **6:**592-603.
11. **Gao, H., A. Obraztova, N. Stewart, R. Popa, J.K. Fredrickson, J.M. Tiedje, K.H. Neason, and J. Zhou.** 2006. *Shewanella loihica* sp. nov., isolated from iron-rich microbial mats in the Pacific Ocean. Int. J. Syst. Evol. Microbiol. **56:**1911-1916.
12. **Heidelberg, J. F., I. T. Paulsen, K. E. Nelson, E. J. Gaidos, W. C. Nelson, T. D. Read, J. A. Eisen, R. Seshadri, N. Ward, B. Methe, R. A. Clayton, T. Meyer, A. Tsapin, J. Scott, M. Beanan, L. Brinkac, S. Daugherty, R. T. DeBoy, R. J. Dodson, A. S. Durkin, D. H. Haft, J. F. Kolonay, R. Madupu, J. D. Peterson, L. A. Umayam, O. White, A. M. Wolf, J. Vamathevan, J. Weidman, M. Impraim, K. Lee, K. Berry, C. Lee, J. Mueller, H. Khouri, J. Gill, T. R. Utterback, L. A. McDonald, T. V. Feldblyum, H. O. Smith, J. C. Venter, K. H. Nealson, and C. M. Fraser.** 2002. Genome sequence of the dissimilatory metal ion-reducing bacterium *Shewanella oneidensis*. Nat. Biotechnol. **20:**1118-1123.

13. **Hau, H.H. and J.A. Gralnick.** 2007. Ecology and biotechnology of the genus *Shewanella.* Ann. Rev. Microbiol. **61:**237-258.
14. **Ivanova, E.P., O.I. Nedashkovskaya, T. Sawabe, N.V. Zhukova, G.M. Frolova, D.V. Nicolau, V.V. Mikhailov, and J.P. Bowman.** 2004. *Shewanella affinis* sp. nov., isolated from marine invertebrates. Int. J. Syst. Evol. Microbiol. **54:**1089-1093.
15. **Johnson, D.A., S.G. Tetu, K. Phillippy, J. Chen, Q. Ren, I.T. Paulsen.** 2008. High-throughput phenotypic characterization of *Pseudomonas aeruginosa* membrane transport genes. PLOS Genetics **4:**e1000211.
16. **Jeon, J.H., J.T. Kim, Y.J. Kim, H.K. Kim, H.S. Lee, S.G. Kang, S.J. Kim, and J.H. Lee.** 2009. Cloning and characterization of a new cold-active lipase from a deep-sea sediment metagenome. Appl. Microbiol. Biotechnol. **81:**865-874.
17. **Kapinets, T.V., M.F. Romine, D.D. Schmoyer, G.H. Kora, M.H. Syed, M.R. Leuze, M.H. Serres, B.H. Park, N.F. Samatova, and E.C. Uberbacher EC.** 2010. *Shewanella* Knowledgebase: integration of the experimental data and computational predictions suggests a biological role for transcription of intergenic regions. Database doi:10.1093/database/baq012
18. **Kettler, G.C., A.C. Martiny, K. Huang, J. Zucker, M.L. Coleman, S. Rodrigue, F. Chen, S. Ferriera, J. Johnson, C. Steglich, G.M. Church, P. Richardson, and S.W. Chisholm.** 2007. Patterns and implications of gene gain and loss in the evolution of *Prochlorocuccus.* PLOS Genetics **3:**2515-2528.
19. **Keymer, D.P., M.C. Miller, G.K. Schoolnik, and A.B. Boehm.** 2007. Genomic and phenotypic diversity of coastal *Vibrio cholerae* strains is linked to environmental factors. Appl. Environ. Microbiol. **73:**3705-3714.
20. **Konstantinidis, K.T. and J.M. Tiedje.** 2005. Genomic insights that advance the species definition for prokaryotes. Proc. Natl. Acad. Sci. U.S.A. **102:**2567-2572.

21. **Konstantinidis, K., M.H. Serres, M. Romine, J. Rodrigues, J. Auchtung, L.-A. McCue, M. Lipton, A. Obraztsova, C. Giometti, K. Nealson, J. Fredrickson, and J. Tiedje.** 2009. Comparative systems biology across an evolutionary gradient within the *Shewanella* genus. Proc. Natl. Acad. Sci. U.S.A. **106:**15909-15914.

22. **Lohmiller, S., K. Hantke, S.I. Patzer, and V. Braun.** 2008. TonB-dependent maltose transport by *Caulobacter crescentus.* Microbiology **154:**1748-1754.

23. **Merlin, C., S. McAteer, and M. Masters.** 2002. Tools for characterization of *Escherichia coli* genes of unknown function. J. Bacteriol. **184:**4573-4581.

24. **Moran, M.A., R. Belas, M.A. Schell, J.M. Gonzales, F. Sun. S. Sun, B.J. Binder, J. Edmonds, W. Ye, B. Orcutt, E.C. Howard, C. Meile, W. Palefsky, A. Goesmann, Q. Ren, I. Paulsen, L. Ulrich, L.S. Thompson, E. Saunders, and A. Buchan.** 2007. Ecological genomics of marine roseobacters. Appl. Environ. Microbiol. **73:**4559-4569.

25. **Moreira-Turcq, P., P. Seyler, J.L. Guyot, and H. Etcheber.** 2003. Exportation of organic carbon from the Amazon river and its main tributaries. Hydrol. Process. **17:**1329-1344.

26. **Murray, A.E., D. Lies, G. Li, K. Nealson, J. Zhou, and J.M. Tiedje**. 2001. DNA/DNA hybridization to microarrays reveals gene-specific differences between closely related microbial genomes. Proc. Natl. Acad. Sci. U.S.A. **98:**9853-9859.

27. **Myers, C.R. and K.H. Nealson.** 1988. Bacterial manganese reduction and growth with manganese oxide as sole electron acceptor. Science **240:**1319-1321.

28. **Nealson, K.H., C.R. Myers, and B.B. Wimpee.** 1991. Isolation and identification of manganese-reducing bacteria and estimates of microbial Mn(IV)-reducing potential in the Black Sea. Deep Sea Res. A **38:**S907-S920.

29. **Ohkawa, I., S. Shiga, and M. Kageyama.** 1979. An esterase on the outer membrane of *Pseudomonas aeruginosa* for the hydrolysis of long chain acyl esters. J. Biochem. **86:**643-656.

30. **Pinchuk, G.E., E.A. Hill, O.V. Geydebrekht, J. De Ingeniis, X. Zhang, A. Osterman, J.H. Scott, S.B. Reed, M.F. Romine, A.E. Konopka, A.S. Beliaev, J.K. Fredrickson, and J.L. Reed.** 2010. Constraint-based model of *Shewanella oneidensis* MR-1 metabolism: a tool for data analysis and hypothesis generation. PLoS Comput. Biol. **6**(6):e1000822.

31. **Pinchuk, G.E., D.A. Rodionov, C. Yang, X. Li, A.L. Osterman, E. Dervyn, O.V. Geydebrekht, S.B. Reed, M.F. Romine, F.R. Collart, J.H. Scott, J. Fredrickson, and A.S. Beliaev.** 2009. Genomic reconstruction of *Shewanella oneidensis* MR-1 metabolism reveals a previously uncharacterized machinery for lactate utilization. Proc. Natl. Acad. Sci. U.S.A. **106:**2874-2879.

32. **Ringo, E., E. Stenberg, and A. R. Strom.** 1984. Amino acid and lactate catabolism in trimethylamine oxide respiration of *Alteromonas putrefaciens* NCMB 1735. Appl. Environ. Microbiol. **47:**1084-1089.

33. **Rocap, G., F.W. Larimer, J. Lamerdin, S. Malfatti, P. Chain, N.A. Ahigren, A. Arellano, M. Coleman, L. Hauser, W.R. Hess, Z.I. Johnson, M. Land, D. Lindell, A.F. Post, W. Regala, M. Shah, S.L. Shaw, C. Steglish, M.B. Sullivan, C.S. Ting, A. Tolonen, E.A. Webb, E.R. Zinser, and S.W. Chisholm.** 2003. Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. Nature **424:**1042-1047.

34. **Rodionov, D.A., C. Yang, X. Li, I.A. Rodionova, Y. Wang, A.Y. Obraztsova, O.P. Zagnitko, R. Overbeek, M.F. Romine, S. Reed, J.K. Fredrickson, K.H. Nealson, A.L. Osterman.** 2010. Genomic encyclopedia of sugar utilization pathways in the *Shewanella* genus. BMC Genomics. **11:**494-512.

35. **Romine, M.F., T.S. Carlson, A.D. Norbeck, L.A. McCue, and M.S. Lipton.** 2008. Identification of mobile elements and pseudogenes in the *Shewanella oneidensis* MR-1 genome. Appl. Environ. Microbiol. 74:3257-3265.

36. **Serres, M.H., S. Gupal, L.A. Nahum, P. Liang, T. Gaasterland, and M. Riley.** 2001. A functional update of the *Escherichia coli* K-12 genome. Genome Biol. 2 [Online.] http://genomebiology.com/2001/2/9/research/0035./

37. **Serres, M.H. and M. Riley.** 2006. Genomic analysis of carbon source metabolism of *Shewanella oneidensis* MR-1: predictions versus experiments. J. Bacteriol. **188:**4601-4609.

38. **Stapleton, R., Zakee L. Sabree, Anthony V. Palumbo, Craig Moyer, Allan Devol, Y. Roh, and J.Z. Zhou.** 2005. Metal reduction at cold temperatures by *Shewanella* isolates from various marine environments. Aquatic Microbiol. **38:**81–91.

39. **Tettelin, V. Masignani, M.J. Cieslewicz, C. Donati, D. Medini, N.L. Ward, S.V. Angiuoli, J. Crabtree, A.L. Jones, A.S. Durkin, R.T. DeBoy, T. M. Davidsen, M. Mora, M. Scarselli, I. Margarit y Ros, J.D. Peterson, C.R. Hauser, J.P. Sundaram, W.C. Nelson, R. Madupu, L.M. Brinkac, R.J. Dodson, M.J. Rosovitz, S.A. Sullivan, S.C. Daugherty, D.H. Haft, J. Selengut, M.L. Gwinn, L. Zhou, N. Zafar, H. Khouri, D. Radune, G. Dimitrov, K. Watkins, K.J. O'Connor, S Smith, T.R. Utterback, O. White, C.E. Rubens, G. Grandi, L.C. Madoff, D.L. Kasper, J.L. Telford, M.R. Wessels, R. Rappuoli, and C.M. Fraser.** 2005. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial "pan-genome". Proc. Nat. Acad. Sci. U.S.A. **105:**13950-13955.

40. **Wenkateswaran, K., M.E. Dollholpf, R. Aller, E. Stackebrandt, and K.H. Nealson.** 1998. *Shewanella amazonensis* sp. nov., a novel metal-reducing

facultative anaerobe from Amazonian shelf muds. Int. J. Syst. Bacteriol. **48:**965-972.

41. **Wenkateswaran, K., D.P. Moser, M.E. Dollholpf, D.P. Lies, D.A. Saffarini, B.J. MacGregor, D.B. Ringelberg, D.C. White, M. Nishijima, H. Sano, J. Burghardt, E. Stackebrandt, and K.H. Nealson.** 1999. Polyphasic taxonomy of the genus *Shewanella* and description of *Shewanella oneidensis* sp. nov. Int. J. Syst. Bacteriol. **49:**705-724.
42. **Yang, C., D.A. Rodionov, X. Li, O.N. Laikova, M.S. Gelfand, O.P. Zagnitko, M.F. Romine, A.Y. Obraztosova, K.H. Nealson, and A.L. Osterman.** 2006. Comparative genomics and experimental characterization of *N*-acetylglucosamine utilization pathway of *Shewanella oneidensis.* J. Biol. Chem. **281:**29872-29885.
43. **Zhou, L., X.-H. Lei, B.R. Bochner, and B.L. Wanner.** 2003. Phenotype micro-array analisys of *Escherichia coli* K-12 mutants with deletions of all two-component systems. J. Bacteriol. **185:**4956-4972.

FIGURE LEGENDS

Fig. 1. Two-dimensional map representation of the large scale comparative analysis of nutrient utilization for five *Shewanella* strains. Differences in color shade intensities represent differences in metabolic activity of transformed data as follows: black (no activity was detected), light color (below 50 units), dark color (between 50 and 100 units), intense color (above 100 units). The units are arbitrary and represent the calculated difference between the areas of a specific well and the control well.

Fig. 2. Carbon source utilization by five *Shewanella* strains. The substrates and their entry points into central metabolism via key metabolic intermediates are shown. Substrates utilized by all five strains are highlighted in bold. Abbreviations include: PP, pentose phosphate pathway; EDD, Entner-Doudoroff pathway; TCA, tricarboxylic acid pathway; G6P, glucose-6-phosphate; GA3P, glyceraldehyde-3-phosphate; PYR, pyruvate; AcCoA, acetyl-CoenzymeA.

Fig. 3. Phenotypic and genotypic changes over evolutionary time. Five *Shewanella* strains were compared in a pair-wise manner for their similarity at the 16S rRNA (green), number of gene orthologs (purple), nitrogen utilization phenotype (blue), and carbon utilization phenotype (red). The similarity of these traits, measured by % of shared ortholog, positive Biolog phenotypes, or rRNA residues, is plotted relative to the % average nucleotide identity (ANI) of the genome pair. A regression line and the curve fitness (R-value) are shown for each trait. When fewer than 10 symbols are shown, results overlap.

Fig. 4. Cladograms showing genomic and phenotypic comparisons among five *Shewanella* strains. Panels: (A) gene-content clustering based on orthologs identified

among the five genomes and (B) phenotypic transformed data clustering based on the presence/absence of respiratory activity for 441 nutrient sources. Scale bars indicate the percent similarity obtained from calculated matrices.

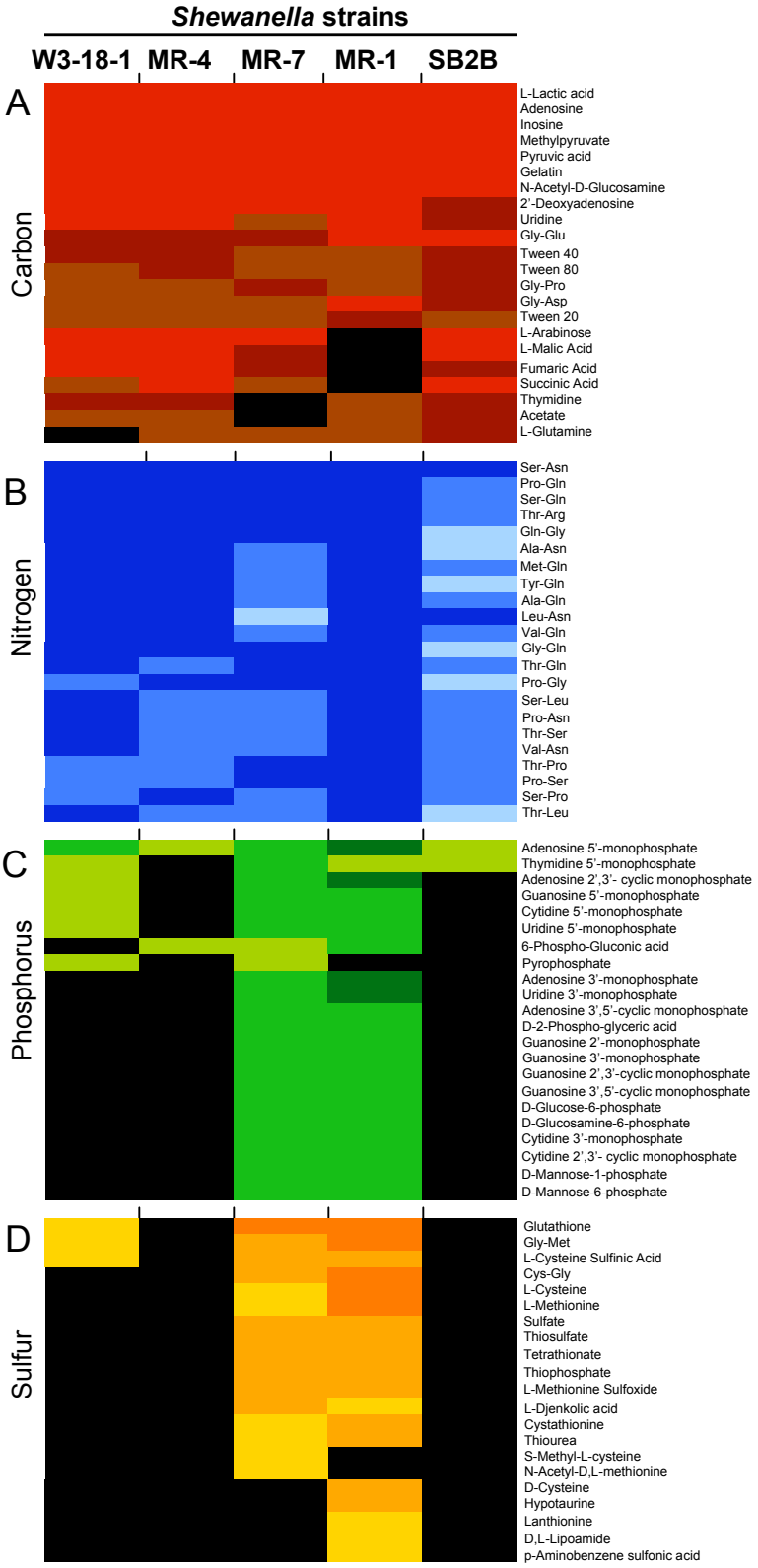
SUPPLEMENTARY MATERIAL

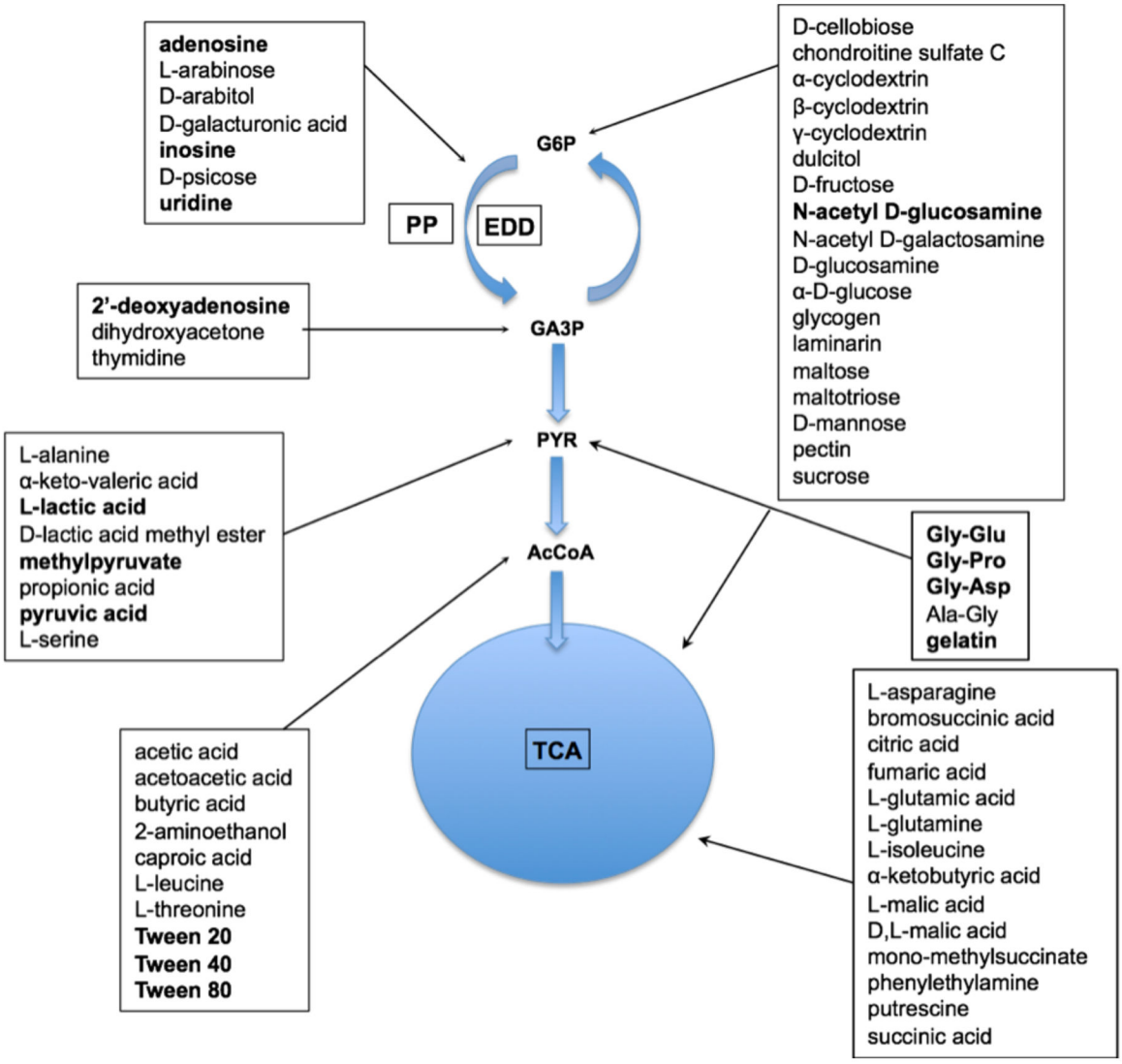
1. **Figure SF1.** Complete two-dimensional map of nutrient utilization profile for five sequenced *Shewanella* strains. Only nutrient sources with positive results for at least one strain are depicted: 67 carbon sources (red), 299 nitrogen sources (blue), 55 phosphorous sources (green), and 21 sulfur sources (yellow). Differences in color shade intensities represent differences in metabolic activity of transformed data as follows: black (no activity was detected), light color (below 50 units), dark color (between 50 and 100 units), intense color (above 100 units).

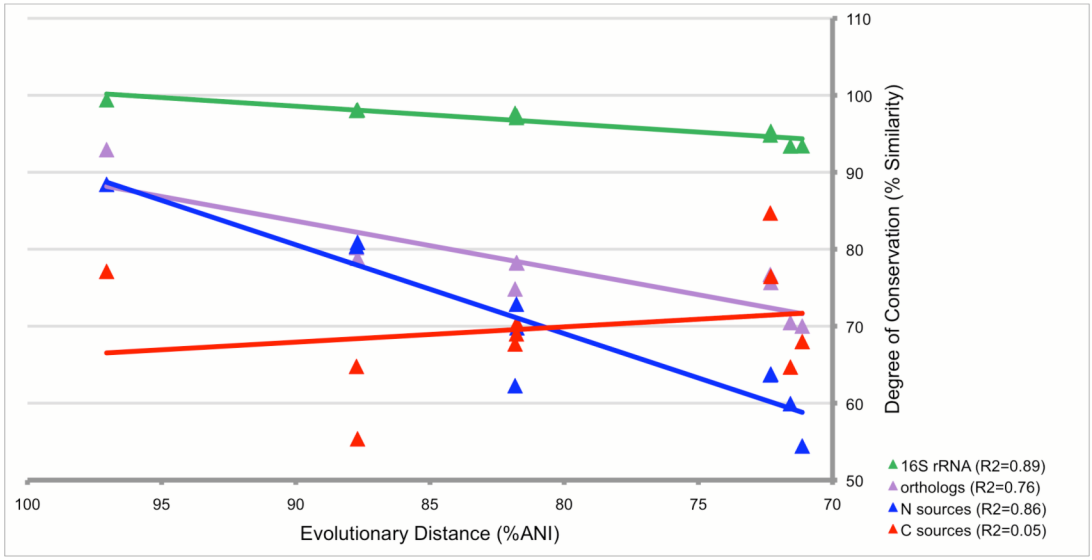
2. **Table ST1**. Table of *Shewanella* gene products and their orthologous relationships. The complete sets of proteins encoded by the five *Shewanella* strains compared are included in the table. The proteins are labeled with their locus tag, and their orthologous relationships are indicated in the table. Functional descriptions of the gene products are listed as well as the metabolic roles for gene products discussed in the paper.

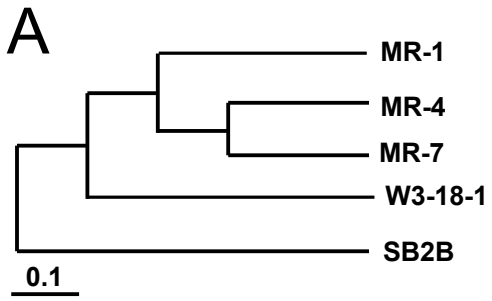
3. **Table ST2**. Average nucleotide identity values for the strains compared.

4. **Table ST3**. Calculations of trait similarity across evolutionary distance.


Shewanella strains
W3-18-1
MR-4
MR-7
MR-1
SB2B
A
Carbon
L-Lactic acid
Adenosine
Inosine
Methylpyruvate
Pyruvic acid
Gelatin
N-Acetyl-D-Glucosamine
2'-Deoxyadenosine
Uridine
Gly-Glu
Tween 40
Tween 80
Gly-Pro
Gly-Asp
Tween 20
L-Arabinose
L-Malic Acid
Fumaric Acid
Succinic Acid
Thymidine
Acetate
L-Glutamine
B
Nitrogen
Ser-Asn
Pro-Gln
Ser-Gln
Thr-Arg
Gln-Gly
Ala-Asn
Met-Gln
Tyr-Gln
Ala-Gln
Leu-Asn
Val-Gln
Gly-Gln
Thr-Gln
Pro-Gly
Ser-Leu
Pro-Asn
Thr-Ser
Val-Asn
Thr-Pro
Pro-Ser
Ser-Pro
Thr-Leu
C
Phosphorus
Adenosine 5'-monophosphate
Thymidine 5'-monophosphate
Adenosine 2',3'- cyclic monophosphate
Guanosine 5'-monophosphate
Cytidine 5'-monophosphate
Uridine 5'-monophosphate
6-Phospho-Gluconic acid
Pyrophosphate
Adenosine 3'-monophosphate
Uridine 3'-monophosphate
Adenosine 3',5'-cyclic monophosphate
D-2-Phospho-glyceric acid
Guanosine 2'-monophosphate
Guanosine 3'-monophosphate
Guanosine 2',3'-cyclic monophosphate
Guanosine 3',5'-cyclic monophosphate
D-Glucose-6-phosphate
D-Glucosamine-6-phosphate
Cytidine 3'-monophosphate
Cytidine 2',3'- cyclic monophosphate
D-Mannose-1-phosphate
D-Mannose-6-phosphate
D
Sulfur
Glutathione
Gly-Met
L-Cysteine Sulfinic Acid
Cys-Gly
L-Cysteine
L-Methionine
Sulfate
Thiosulfate
Tetrathionate
Thiophosphate
L-Methionine Sulfoxide
L-Djenkolic acid
Cystathionine
Thiourea
S-Methyl-L-cysteine
N-Acetyl-D,L-methionine
D-Cysteine
Hypotaurine
Lanthionine
D,L-Lipoamide
p-Aminobenzene sulfonic acid


adenosine
L-arabinose
D-arabitol
D-galacturonic acid
inosine
D-psicose
uridine
G6P
PP
EDD
D-cellobiose
chondroitine sulfate C
α-cyclodextrin
β-cyclodextrin
γ-cyclodextrin
dulcitol
D-fructose
N-acetyl D-glucosamine
N-acetyl D-galactosamine
D-glucosamine
α-D-glucose
glycogen
laminarin
maltose
maltotriose
D-mannose
pectin
sucrose
2'-deoxyadenosine
dihydroxyacetone
thymidine
GA3P
PYR
L-alanine
α-keto-valeric acid
L-lactic acid
D-lactic acid methyl ester
methylpyruvate
propionic acid
pyruvic acid
L-serine
Gly-Glu
Gly-Pro
Gly-Asp
Ala-Gly
gelatin
AcCoA
TCA
acetic acid
acetoacetic acid
butyric acid
2-aminoethanol
caproic acid
L-leucine
L-threonine
Tween 20
Tween 40
Tween 80
L-asparagine
bromosuccinic acid
citric acid
fumaric acid
L-glutamic acid
L-glutamine
L-isoleucine
α-ketobutyric acid
L-malic acid
D,L-malic acid
mono-methylsuccinate
phenylethylamine
putrescine
succinic acid


110
100
90
80
70
60
50
Degree of Conservation (% Similarity)
100
95
90
85
80
75
70
Evolutionary Distance (%ANI)
16S rRNA (R2=0.89)
orthologs (R2=0.76)
N sources (R2=0.86)
C sources (R2=0.05)


A
MR-1
MR-4
MR-7
W3-18-1
SB2B
0.1

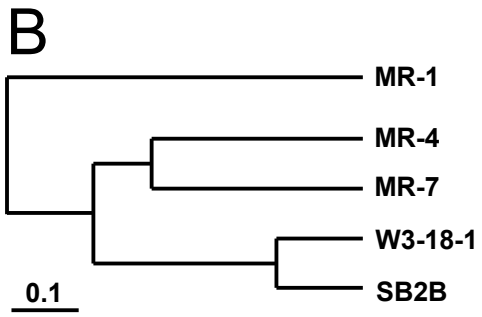

B
MR-1
MR-4
MR-7
W3-18-1
SB2B
0.1

TABLE 1. Description of habitats of isolation and genome information for the *Shewanella* strains used in this study.

Habitat of Isolation	Environmental characteristics	*Shewanella* species	Genome size (number of genes)[b]	Reference
Lake Oneida (USA) fresh sediment	Anaerobic redox condition, pH 7.5-8.2, 4°C (Winter) and 20°C (Summer), > 100 μM Mn^{+4}	*Shewanella oneidensis* MR-1	5,131,416 (4,745)	27
Amazon River delta (Brazil) sediment	Non-sulfidic, suboxic redox conditions, 26.4-31.5°C, 0.1-1 mM Fe^{+2}, 0.1-0.2 mM Cl^{-} (low salinity), depth of 1 m	*Shewanella amazonensis* SB2B	4,306,142 (3,785)	40
Pacific Ocean (USA) marine sediment	Washington coast, 3.4°C, 10 μM Fe^{+2}, 76 μM NH_4, depth of 997 m	*Shewanella* sp. W3-1-18	4,708,380 (4,217)	26, 38
Black Sea sediments	8°C, 5 μM O_2, 7 μM NO_3, depth of 60 m	*Shewanella* sp. MR-7	4,799,109 (4,186)	28[a]
Black Sea water	16°C, 280 μM O_2, depth of 5 m	*Shewanella* sp. MR-4	4,706,287 (4,084)	28[a]

[a] Values were calculated based on reference.
[b] Genome size (Mb) and number of predicted genes were calculated as a sum of chromosome and plasmid, when the later was present.