Title

Shewanella baltica populations in the Baltic Sea

Journal

Environmental Microbiology, 16(6)

ISSN

1462-2912

Authors

Deng, Jie
Brettar, Ingrid
Luo, Chengwei
et al.

Publication Date

2014-06-01

DOI

10.1111/1462-2920.12344

Peer reviewed

environmental microbiology

Environmental Microbiology (2014) **16**(6), 1854–1866 doi:10.1111/1462-2920.12344

Stability, genotypic and phenotypic diversity of *Shewanella baltica* in the redox transition zone of the Baltic Sea

Jie Deng,[1] Ingrid Brettar,[2] Chengwei Luo,[3,4] Jennifer Auchtung,[1] Konstantinos T. Konstantinidis,[4,5] Jorge L. M. Rodrigues,[6] Manfred Höfle[2] and James M. Tiedje[1]*

[1]*Center for Microbial Ecology, Michigan State University, East Lansing, MI, USA.*
[2]*Helmholtz Centre for Infection Research, Braunschweig, Germany.*
[3]*Center for Bioinformatics and Computational Genomics, Schools of* [4]*Biology and*
[5]*Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA, USA.*
[6]*Department of Biology, University of Texas, Arlington, TX, USA.*

Summary

Studying how bacterial strains diverge over space and time and how divergence leads to ecotype formation is important for understanding structure and dynamics of environmental communities. Here we assess the ecological speciation and temporal dynamics of a collection of *Shewanella baltica* strains from the redox transition zone of the central Baltic Sea, sampled at three time points over a course of 12 years, with a subcollection containing 46 strains subjected to detailed genetic and physiological characterization. Nine clades were consistently recovered by three different genotyping approaches: *gyrB* gene sequencing, multilocus sequence typing (MLST) and whole genome clustering of data from comparative genomic hybridization, and indicated specialization according to nutrient availability, particle association and temporal distribution. Genomic analysis suggested higher intra- than inter-clade recombination that might result from niche partitioning. Substantial heterogeneity in carbon utilization and respiratory capabilities suggested rapid diversification within the same 'named' species and physical habitat and showed consistency with genetic relatedness. At least two major ecotypes, represented by MLST clades A and E, were proposed based on genetic, ecological and physiological distinctiveness. This study suggests that genetic analysis in conjunction with phenotypic evaluation can provide better understanding of the ecological framework and evolutionary trajectories of microbial species.

Received 18 March, 2013; accepted: 25 November, 2013. *For correspondence. E-mail tiedjej@msu.edu; Tel. (+1) 517 353 9021; Fax (+1) 517 353 2917.

Introduction

The genus *Shewanella* belongs to the *Gammaproteobacteria* and is well known for its diverse metabolic capabilities (Nealson and Scott, 2006). *Shewanella* species are often found in energy-rich, redox-fluctuating environments, and are thought to play an important role in coupling the turnover of organic matter with anaerobic respiration of different electron acceptors (Venkateswaran *et al.*, 1999; Fredrickson *et al.*, 2008). In this work, we provide a systematic study of a collection of isolates from one species, namely *Shewanella baltica*, sampled at three time points over a course of 12 years, along a redox gradient from the stratified basin of the Gotland Deep in the central Baltic Sea. Previous studies have focused on comparative genomics of the sequenced members of this genus and inferred sexual evolutionary paths within a few clades (Caro-Quintero *et al.*, 2010; 2011). Our work broadened the comparative analysis with comprehensive genotypic and phenotypic analysis on a larger set of 46 strains, providing further implications on specialization of *S. baltica* into diverse environmental niches and their evolution over time.

The Baltic Sea is the world's largest brackish water environment and was impacted with pronounced eutrophication because of phosphorus (P) and nitrogen (N) overloads after World War II and at an accelerated pace since the 1960s (Elmgren, 1989). Baltic nutrient input was estimated to be increased by at least fourfold and eightfold for N and P, respectively, in surface waters from the 1960s to the 1980s, and levelled off by the beginning of the 21st century (Elmgren, 2001). At the Gotland Deep sampling station, eutrophication by N and P (as indicated by the concentrations in the winter surface water) had reached its maximum at the end of the 1970s,

and were on a comparable level from 1980 to 1998, meaning a comparable level of N and P eutrophication background for our samplings in 1986, 1987 and 1998 (Struck *et al*., 2004).

The Gotland Deep is the largest basin in the Baltic Sea and is considered representative of the central Baltic Sea environment (Brettar and Rheinheimer, 1992; Ziemke *et al*., 1997). Hydrological characteristics of the Gotland Deep include a halocline at 60–80 m depth, which inhibits vertical mixing of water, and stagnant deep anaerobic water with hydrogen sulfide accumulated from the sediment. Oxygen, nitrate and hydrogen sulfide meet directly near the oxic–anoxic interface, forming the energy rich layer where important biogeologic processes, including denitrification occur (Brettar and Rheinheimer, 1991). Denitrification was considered as the major factor counteracting eutrophication in the Baltic Sea (Rönner, 1985; Rönner and Sörensson, 1985). For most times of the year, denitrification in the water column of the central Baltic has been shown to be an autotrophic process fuelled by oxidation of reduced sulfur compounds (Brettar and Rheinheimer, 1991). Heterotrophic nitrate reduction can be expected in periods after enrichment of the deep water with organic carbon because of sedimentation of plankton blooms (Rönner and Sörensson, 1985; Brettar and Rheinheimer, 1992).

In order to study the microbial flora relevant for heterotrophic and autotrophic denitrification, techniques targeting these nitrate reducers were used to isolate bacteria from the central Baltic, i.e. from different depths of the Gotland Deep in 1986 and 1987, and from station T in 1986 (Brettar and Rheinheimer, 1991). As a result, a group of 134 *S. baltica* strains were obtained. Based upon dilution analysis of the isolates from 1986, *S. baltica* represented 77% of the total population of culturable nitrate reducers of the Gotland Deep (Brettar *et al*., 2001; 2002). Consistent results were observed among the 1987 isolates (Ziemke *et al*., 1997). *Shewanella baltica* was specifically enriched in the low oxic and oxic–anoxic transition zone from 80 m to 140 m. We selected 36 of these *S. baltica* strains for this study for further characterization based on their profiles from randomly amplified polymorphic DNA (RAPD) genotyping (Ziemke *et al*., 1997).

Stratification in the Gotland Deep was disrupted by a major salt water inflow from the North Sea in 1993 that led to deep water renewal (Matthäus and Lass, 1995; Håkansson *et al*., 1993), after which the typical stratification was re-established (Matthäus *et al*., 1996). In 1998, 10 additional *S. baltica* strains were recovered from the same sampling site in the Gotland Deep (Ziemke *et al*., 1998), although at the time of isolation, the *S. baltica* likely represented a less abundant population compared with the 1980s (Brettar *et al*., 2012). Together with those isolated from the 1980s, these strains provide a valuable resource for investigating bacterial specialization to the environmental gradients in the Baltic Sea, and tracking dynamics of the *S. baltica* populations, especially after the turnover event.

In this study, a series of comprehensive approaches were used to characterize the genotypic and phenotypic differentiation among these *S. baltica* strains. Phylogenetic relatedness was determined through the multilocus sequence typing (MLST) with a novel strategy in gene selection, which was compared with the traditionally recommended genotyping for *Shewanella* using the housekeeping gene, *gyrB*, encoding the B subunit of the DNA gyrase (Venkateswaran *et al*., 1999). Phylogenetic relatedness inferred from MLST was further confirmed by the whole genome clustering of data from comparative genomic hybridization (CGH). Phenotypic diversity was also examined. We tested respiratory capabilities of these *S. baltica* strains to use several organic and inorganic compounds relevant to potential alternative electron acceptors available in the environment. Utilization of 95 carbon compounds was also screened to examine their differentiation in metabolic capabilities. Through comparative genetic and metabolic profiling of these strains with their temporal and spatial distribution, we sought to identify environmental signatures of the *S. baltica* species, and to further understand the outcome of selection on the intraspecific populations.

Results

Genotypic analysis of the S. baltica *strains*

Analysis of gyrB genes revealed 10 well supported lineages of S. baltica. *gyrB* sequences were analysed to preliminarily determine the genetic relationships among the 46 *S. baltica* isolates, and were also used as a control for validation of the MLST genotyping approach, which employed a novel gene selection strategy. As a result, using *Shewanella oneidensis* MR-1 as an outgroup, all the *S. baltica* strains were clustered together (Supporting Information Fig. S1), which further confirmed that the 1986, 1987 and 1998 strains belong to the same species. The sequence analysis of the 1122 bp *gyrB* amplicon revealed a phylogeny of 10 well-supported lineages (bootstrap value > 90%).

MLST analysis revealed phylogenetic relationships among strains similar to those observed with gyrB. Although the *gyrB* and RAPD analysis (Ziemke *et al*., 1997) provided initial insights into the phylogenetic relationships among the strains, we sought a more refined characterization of the evolutionary history among the strains using more genome-wide information. Seven genes from the *Shewanella* conserved gene core were chosen for further phylogenetic characterization of these isolates (locus

Table 1. Marker genes used for MLST analysis.

Locus tag	Annotation	Length before concatenation/bp
SO_0578	Periplasmic metalloprotease	696
SO_0625	Periplasmic protein with Sel1-like repeats	537
SO_1771	D-glycerate transporter, GlyT	594
SO_2183	Periplasmic ErfK/YbiS/YcfS/YnhG family protein	321
SO_2615	Aminodeoxychorismate lyase, PabC	474
SO_2706	Succinylarginine dihydrolase, AstB	702
SO_4702	Glutathione reductase, Gor	903

tags as in *S. oneidensis* MR-1: SO_0578, SO_0625, SO_1771, SO_2183, SO_2615, SO_2706, SO_4702) (Table 1). Twenty-six unique sequence types were revealed from the concatenated sequences of these loci. Eleven well-supported clades were identified, among which nine were consistent with those observed in the *gyrB* phylogram (Fig. 1). The SO_4702, SO_0578 and SO_1771 genes account for more than half of all polymorphic sites, whereas others provide comparatively less discriminatory power. Synonymous (Ks) and non-synonymous (Ka) substitution rates were calculated based on sequence information at each locus, followed by Tajima's test of neutrality (Table 2). All the Ks/Ka ratios (except SO_2615 and SO_4702) are much larger than 1, indicating the effect of purifying selection against amino acid changes. Population recombination (ρ) and mutation (θ) rates were also estimated for each marker gene (Table 2). In particular, as the major MLST clades (A, D, J, E and H) included only identical or almost identical sequences, the calculated recombination to mutation ratios, ranging from 0.230 to 0.804, reflect low levels of inter-clade recombination. Because of low intra-clade heterogeneity, we found estimation of gene flow within clades not be feasible at this time.

Hierarchical clustering analysis on CGH profiles revealed clades that were identical to those identified through MLST. In order to obtain a more holistic picture of the genomic content of the *S. baltica* species, we used CGH to study the gene content in this collection of strains (Caro-Quintero *et al*., 2010). Hierarchical clustering was performed based upon the presence/absence profiles of 5635 genes generated from the CGH datasets (Supporting Information Fig. S2), representing the phylogenetic relationship among these strains on the level of whole genome composition. Ten clades were revealed from this method and were all identical to those identified from the MLST phylogeny, indicating that strains from the same clades share more gene contents than those from different clades. In addition, pairwise mean squared error between distance matrices from the *gyrB*, MLST and the CGH datasets showed higher similarity between the MLST and CGH datasets (Table 3), suggesting the advantage of the MLST analysis over the single gene-based genotyping method.

Habitat association of the S. baltica *strains.* We adopted AdaptML to investigate ecological differentiation by resource partitioning among the *S. baltica* strains (Hunt *et al*., 2008). AdaptML detects ecological associations of groups of strains if measured characteristics in one or more sampling dimensions are differentially apportioned among samples, and hypothesizes such distinct distributions as 'projected habitats'. Here we used both isolation procedures and sizes of inocula as the input for the AdaptML analysis. Clearly, the isolation procedures reflected both favoured redox state and nutrient sources while sizes of inocula related to the *in situ* state in the water column, i.e. small inocula (0.001–0.01 ml), may indicate higher concentration of free-living cells, whereas larger inocula (up to 10 ml) more likely suggest preferential particle attachment. Although depth of isolation could also be ecologically relevant, it was not considered here because of the possibility of minor horizontal and vertical circulation (Stigebrandt, 1987; Neretin *et al*., 2003). As a result, five projected habitats defined by distinct distributions of strains obtained from different isolation media and inoculum sizes were hypothesized, each associated with one to three MLST clades (Fig. 1). Interestingly, MLST clades A, B and D were projected to the same habitat H0, characterized by recovery most frequently from high-carbon (a, b, c) or thiosulfate-supplemented medium (e) using moderate size of inocula (0.1–1 ml), indicating some level of niche overlap that might have contributed to the convergence of MLST-A and D (Caro-Quintero *et al*., 2010). Both MLST clades E and F were associated with H1, characterized by isolation almost exclusively using high-carbon medium (a, b, c) and small inocula, suggesting a different lifestyle of these two clades. All strains isolated in 1998 as well as OS183 and OS225, both obtained from 1986 yet clustered with the 1998 strains, were associated with H2 with most strains obtained using medium-sized inocula from no-carbon and low-carbon medium (d, f, g). Habitats H3 and H4 covered relatively fewer strains. H3 included MLST-G and H strains that were obtained from several isolation procedures. H4 involved only three strains from MLST-C, two of which were obtained using large (10 ml) inocula. Impact of isolation medium on selection of ecologically distinct strains has also been reported in a previous study, which found putative ecotypes within the *Bacillus subtilis–Bacillus licheniformis* clade had significant differences in their tendencies to be recovered from various media (Kim *et al*., 2012). Yet resource partitioning by microbial communities most likely occurs at finer scales, and there must be more ecological diversity within the *S. baltica* that was not

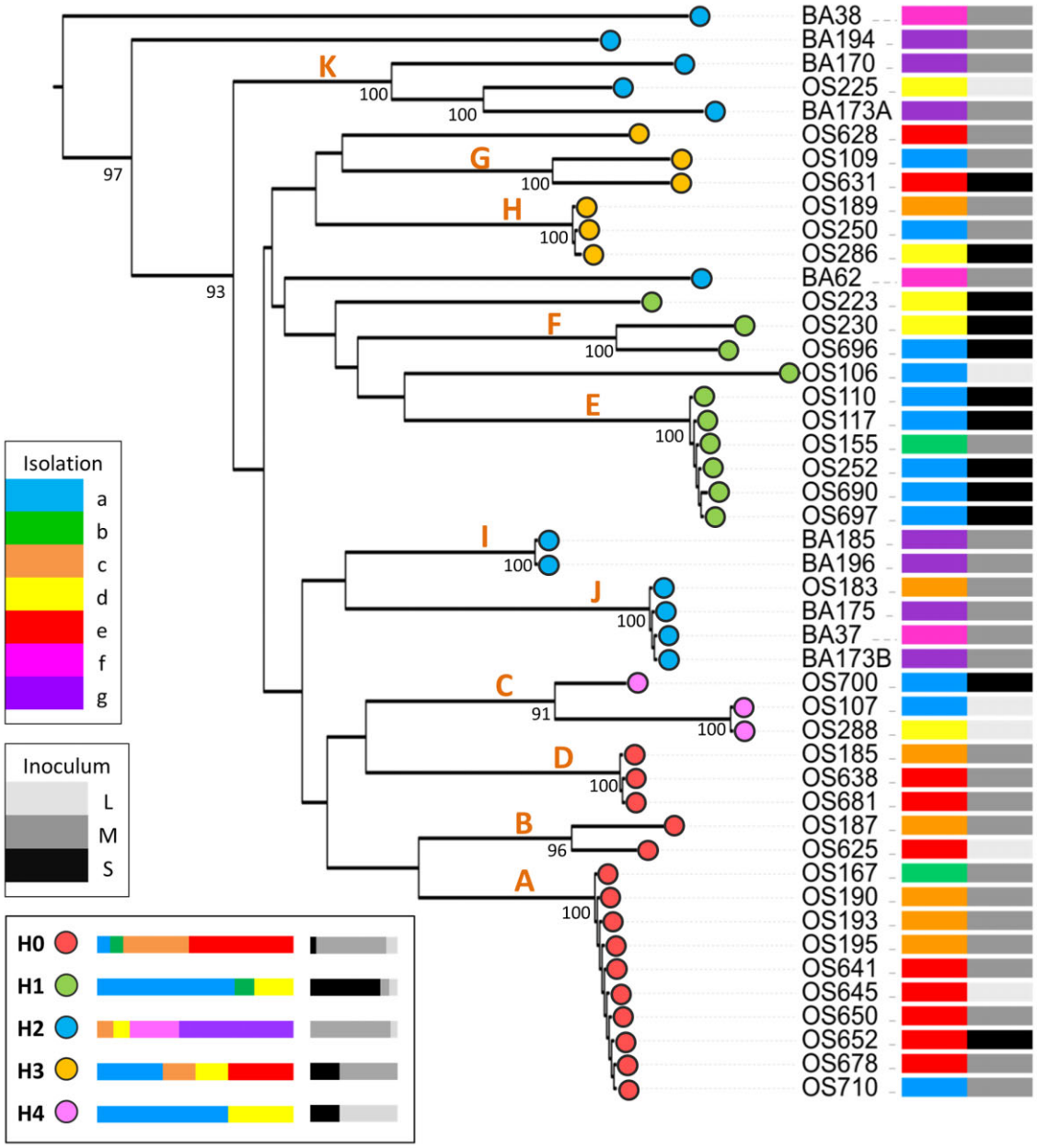

Fig. 1. Phylogeny and habitat association of the *S. baltica* populations. The tree was constructed with the Neighbor Joining algorithm using concatenated sequences from seven MLST genes. Strains starting with 'OS' were isolated from 1986 or 1987, whereas strains starting with 'BA' were from 1998. The two colour panels following each strain indicate the isolation procedure and the size of inoculum, respectively. The strain *S. oneidensis* MR-1 was used as outgroup but is not shown. Bootstrap values above 50% are shown on branches based on 1000 bootstrap replicates. The AdaptML modelled habitat association by identifying different trends in distributions of isolation procedures and sizes of inoculum, represented by 'projected habitats' as indicated in coloured circles. Characteristic distribution of isolation procedures and sizes of inoculum for each habitat is illustrated in the bottom-left panel. The isolation procedures are as follows: (a) Nutrient Broth medium supplemented with nitrate under anaerobic condition, (b) ZoBell agar plate under aerobic condition, (c) ZoBell agar plate under anaerobic condition, (d) 1:10 diluted Nutrient Broth medium supplemented with nitrate under anaerobic condition, (e) liquid nitrate plus thiosulfate medium under anaerobic condition without added organic carbon source, (f) ZoBell (1/5 dilute) agar plate under aerobic condition, (g) nitrate plus thiosulfate agar medium under anaerobic condition without added organic carbon source. In addition, 'S' indicates small inoculum of less than 0.1 ml; 'M' indicates moderate inoculum size between 0.1 ml and 1 ml; 'L' indicates inoculum of 10 ml.

Table 2. Neutrality test and estimation of recombination (ρ) and mutation rates (θ).

	SO_0578	SO_0625	SO_1771	SO_2183	SO_2615	SO_2706	SO_4702
Polymorphic sites	97	46	71	23	43	57	131
Ka	0.00667	0.00261	0.00188	0.00129	0.00529	0.00276	0.00528
Ks	0.13958	0.12357	0.08492	0.08012	0.05083	0.08665	0.13912
Ks/Ka	20.9	47.3	45.2	62.1	9.6	31.4	26.3
Watterson θ	22.1	10.5	16.2	5.2	9.8	13.0	29.8
Estimated ρ	35.5	10.5	24	4.5	4.5	7.5	21.5
θ per site	0.0317	0.0195	0.0272	0.0163	0.0206	0.0185	0.0330
ρ per site	0.0510	0.0196	0.0404	0.0140	0.0095	0.0107	0.0238
Recombination to mutation ratio	0.804	0.502	0.743	0.430	0.230	0.289	0.361

detected by the habitat association analysis. For instance, the strong signals from isolation medium may be due to specialization with regard to both nutrient sources and the redox gradient of the Baltic Sea; the effects from both isolation medium and inoculum size may also associate with macro-ecological conditions such as seasonal algal blooms. In addition, the populations might have adapted to environments not sampled in this study, such as colonization in animal intestines or phage predation. Despite these uncertainties, the observed partitioning according to isolation medium and inoculum size may have important implications for population biology of the *S. baltica* species.

Evaluation of intra- vs. inter-clade recombination using whole genome sequences. Acquisition of adaptive traits into distinct niches may lead to arising of new ecotypes, within which selective sweeps play important roles in maintaining the intra-population genetic homogeneity (Cohan, 2002a). Although it remains unknown what are the gene flow barriers among sympatric populations, reduced gene flow has been observed shortly after ecological divergence (Shapiro *et al*., 2012). Therefore, an important question in evaluating the effects of ecological differentiation on population structure is whether gene flow is depressed among genotypes that arose because of ecological distinctiveness. Because our MLST data lack the resolution to reveal intra-clade heterogeneity, we evaluated the available *S. baltica* genome sequences to assess the levels of intra- vs. inter-clade gene flow. Concatenation of 2894 core genes were used to construct the genome phylogeny of the *S. baltica* genomes, from which intra- and inter-clade recombination rates were estimated (Supporting Information Fig. S3). A total of 3083 genes that were commonly shared between at least two clades were examined for inter-clade recombination, out of which 0–1.14% [0–35 genes with $P < 0.05$ by embedded quartet decomposition analysis (EQDA) (Luo *et al*., 2011)] were suggested to have recombined (Supporting Information Fig. S4). Owing to the limited number of genomes within any clade for testing intra-clade recombination, we estimated the recombination rate using three strains OS195, OS678 and OS625 from clades A and B to approximate the level of intra-clade recombination. Although clades A and B were confidently separated in the MLST phylogeny, they were relatively closer compared with other clades, and were indistinguishable in the RAPD dendrogram. Recombination rates among these three strains were estimated by measuring concordance between gene phylogeny and genome phylogeny. Out of the 991 genes that qualified for the analysis, 3.13% (31 genes) showed significant signal of recombination ($P < 0.05$), which was over 2.7 times higher than the above inter-clade recombination rates (Dixon's Q test, $P = 0.003$). It should be noted that, because of higher similarity among genomes of OS195, OS678 and OS625, a smaller fraction of genes showed enough phylogenetic signal over background noise to be appropriate for assessing recombination. Hence, this rate (3.13%) most likely represents an underestimate of intra-clade recombination rate. However, high inter-clade recombination has also been observed between OS195 and OS185 (clades A and D), based on which clades A and D were hypothesized to have been undergoing convergent evolution (Caro-Quintero *et al*., 2010). This indicates that when environmental conditions favour, high inter-clade recombination can also occur. In summary, based on the only possible approximation of intra-clade recombination using the three relatively close genomes OS195, OS678 and OS625, we observed depressed inter-clade recombination, supporting genetic barriers among *S. baltica* genotypes that might have contributed to stability of the current population structure.

Phenotypic diversity among the S. baltica *strains*

Biolog profiling of carbon utilizing capabilities. Use of 95 different carbon sources by the *S. baltica* strains was

Table 3. Correlation among three distance matrices for phylogenetic inferences.

	MLST	gyrB	CGH
MLST	–	–	–
gyrB	0.669	–	–
CGH	0.757	0.647	–

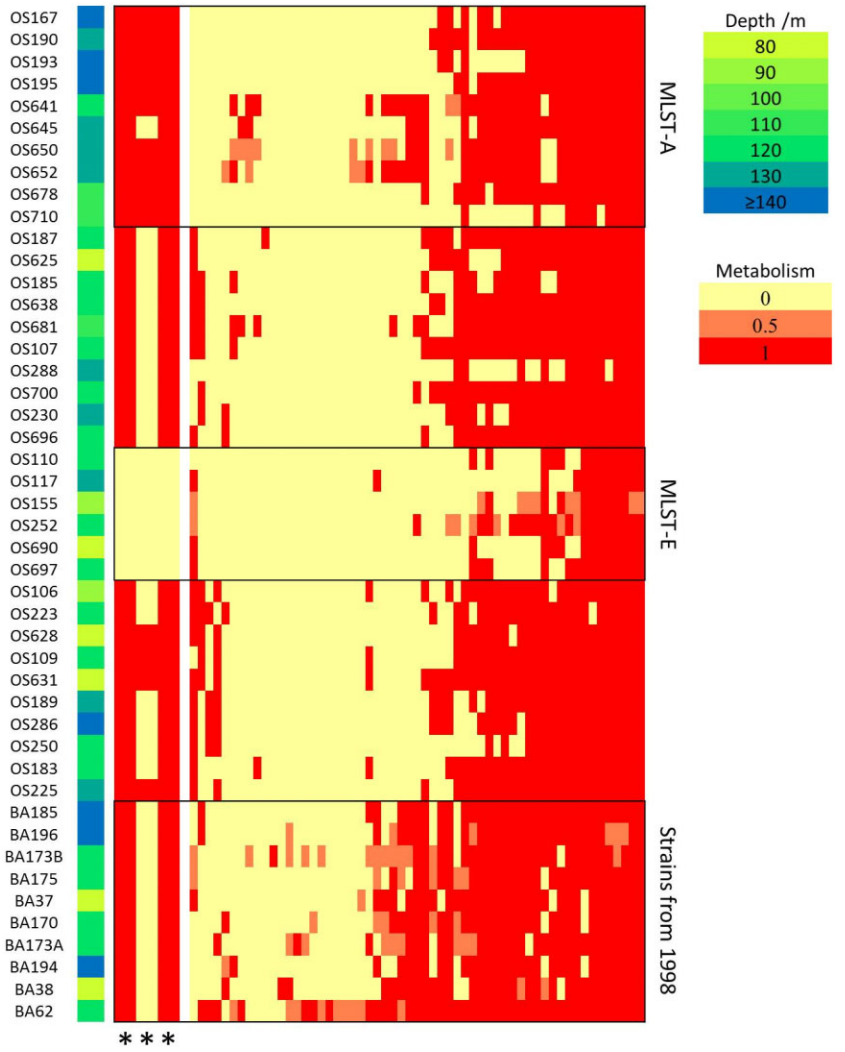

Fig. 2. Respiratory capabilities and carbon source utilization patterns among the 46 *S. baltica* strains. The first column on the left indicates the isolation depth of each strain. The three columns marked by an asterisk indicate the capabilities of respiring thiosulfate, DMSO and TMAO respectively. The rest of the columns correspond to 57 carbon sources with varied utilization patterns among the 46 strains. Also highlighted were MLST clades A and E, as well as strains isolated from 1998.

screened with the Biolog-GN system. Metabolic activities were measured and assays were scored as either growth or no growth (Fig. 2). The *S. baltica* strains were able to use an average of 36 compounds under aerobic conditions. Among the carbon sources tested, 19 were not used by any *S. baltica* strain, while the following 11 substrates were used by all strains: cis-aconitic acid, sucrose, D-gluconic acid, L-glutamic acid, dextrin, maltose, α-D-glucose, L-serine, N-acetyl-D-glucosamine, lactic acid, and inosine. Of the compounds tested, these carbon sources comprise part of the metabolic core of the *S. baltica* species. Interestingly, compared with *S. oneidensis* MR-1, whose carbon source utilization capability is limited primarily to three carbon compounds (Serres and Riley, 2006), these *S. baltica* strains were capable of using five or six carbon compounds, suggesting distinct carbon metabolism pathways present in the *S. baltica* strains. In order to mitigate the influence from phylogenetic relatedness in evaluating the relationship between Biolog profiles and environmental variables, the phylogenetic independent residuals from the Biolog data (Felsenstein, 1985) were taken and used to further examine the effects from the time of isolation, the culturing methods, as well as water chemistry measurements including the concentrations of NO_3^-, O_2 and H_2S. Significant correlation was observed only between these residuals with the time of isolation (analysis of variance $P < 0.01$), whereas effects of other environmental variables are either not significant on the phenotypic diversity, or are through influencing the phylogenetic relatedness among these isolates.

In order to discern relationships between the genetic content and phenotypic capability, we compared the phylogenetic independent residuals from the Biolog dataset and genetic distances from CGH, and found a strong and positive correlation ($P < 0.001$) (Supporting Information Fig. S5). This correlation can also be visualized through comparison of the hierarchical clustering structures of the CGH and the Biolog profiles (Supporting Information Fig. S6), where six clades were at least

partially shared between the two dendrograms. For instance, strains in MLST clade E were also grouped together in the Biolog dendrogram and consistently used fewer carbon sources compared with other clades. In particular, four carbon acids including succinate acid, aspartic acid, and α-ketobutyric acid failed to promote growth of any strains in this clade, implying the lack of relevant catalytic enzymes in their pathways. In summary, our results suggest that closely related strains share relatively similar carbon utilizing capabilities, and particularly, individual clades may have characteristic metabolic profiles.

Diversity of respiratory capabilities among the S. baltica *strains.* Although these *S. baltica* strains were initially identified as denitrifiers in the Baltic Sea (Brettar *et al*., 2001; 2002) by N_2O production (Brettar and Höfle, 1993), they are not respiratory denitrifiers because we detected no N_2 gas from any of these strains. But, because all strains were able to grow with nitrate as the sole electron acceptor and decrease the concentration of nitrate, we expect that they were respiring via dissimilatory reduction of nitrate to ammonia. *Shewanella oneidensis* MR-1 has been shown to respire nitrate via this pathway (Cruz-Garcia *et al*., 2006) and genes for this process are encoded in the *S. baltica* genomes (Caro-Quintero *et al*., 2010) whereas genes for denitrification were not found. The N_2O production shown for all 134 strains isolated in 1986 and 1987 may be attributed to the process of nitrate ammonification rather than denitrification (Bleakley and Tiedje, 1982; Brettar and Höfle, 1993; Baggs, 2008). Out of the 46 strains tested, 13 (all from 1986 and 1987) were able to grow with dimethyl sulfoxide (DMSO) as the only electron acceptor (Supporting Information Table S1). Thiosulfate and trimethylamine-N-oxide (TMAO) can be used as electron acceptors by most of these isolates, except for the strains in clade E, implying the possibilities of mutation or gene loss in their recent common ancestor.

Discussion

Identification of distinct intraspecific S. baltica *ecotypes*

This collection of *S. baltica* strains exemplified an assemblage of coexisting populations that inhabit different niches in the redox transition zone of the Baltic Sea. Distinct clusters of strains were consistently recovered from three different phylotyping approaches, and were also congruent with phenotypic profiles regarding both carbon source utilization and respiratory capabilities, suggesting genetic and phenotypic homogeneity within these clades. In particular, the MLST clades, especially regarding the large clades, are comparable with the former RAPD clades (Ziemke *et al*., 1997) based on the entire collection of 134 strains. For instance, the MLST clades E and D are congruent with the RAPD clades F and C respectively; the MLST clades A and B that share recent common ancestor also correspond well to the RAPD clade A. Therefore, what we found with regard to the selected 36 strains may exemplify the patterns underlying the larger strain collection. On the other hand, although the 1998 strains were not profiled using RAPD, the MLST cladistics suggested genetic continuity of these strains with those obtained in 1986 and 1987. Two clades, MLST-J and K, contain strains from both 1998 and 1986. Particularly, OS225 and BA173A from MLST-K, as well as OS183 and strains from 1998 in MLST-J, share high sequence similarities (98.7% and 100% respectively), suggesting that the later strains are likely descendants of those from 1986. In summary, consistency between the MLST and RAPD analyses suggests stability of several major *S. baltica* clades, and further supports the genetic homogeneity within them.

In addition, we provided evidence that the MLST clades were associated with distinct ecological niches, especially regarding the large clades MLST-A and E. Strains in each of MLST-A and E share nearly identical MLST and *gyrB* sequences, as well as great proportions of their genes based on CGH. Strains in MLST-E followed a characteristic probability distribution of isolation procedures and inocula sizes as indicated by AdaptML analysis, and were mostly recovered from high-carbon medium at the highest dilution steps, suggesting that they favour high-nutrient conditions and free-living lifestyle. Accordingly, MLST-E strains metabolized significantly fewer carbon sources compared with other clades, and were exclusively incapable of using four carbon compounds including succinate acid, aspartic acid and α-ketobutyric acid. In addition, strains in this clade were never recovered from thiosulfate-containing medium that was favoured by MLST-A, and exclusively lacked the capabilities of respiring thiosulfate, DMSO and TMAO. Therefore, we propose MLST-E to represent an individual ecotype that is physiologically less versatile and less adapted to anoxic water conditions. On the contrary, the MLST-A strains were frequently obtained from thiosulfate-containing medium using almost exclusively medium sizes of inocula, suggesting a different lifestyle, possibly more particle attached compared with MLST-E. They were capable to respire all three tested electron acceptors, and used more carbon sources, especially regarding those isolated in 1987. In addition, the RAPD-A clade that corresponds to MLST-A, although was only obtained near the oxic–anoxic transition zone (110–140 m) in 1986, was recovered from all depths (10–235 m) in 1987 and covered a much larger number of strains (31 strains in 1987 vs. 18 strains in 1986)

(Ziemke *et al*., 1997). This genotype was apparently rapidly adapting and expanding the size of its habitat, and therefore, is proposed to represent a physiologically versatile ecotype.

Although comparative analyses on sequenced genomes captured a snapshot of convergent evolution of MLST clades A and D, we suggest not merging them into one ecotype as there is significant heterogeneity between the two clades both genetically and phenotypically. The RAPD-C that correspond to MLST-D, although less frequently recovered compared with RAPD-A, was most often obtained from the suboxic zone near the denitrification layer, and was considered of potential importance in N cycling of the Baltic Sea (Ziemke *et al*., 1997). Therefore, MLST-D may represent a different ecotype, yet its ecological roles are awaiting further exploration. Among other MLST clades that contained only a few strains, we are less confident about whether each of them justifiably represents an individual ecotype, or if they contain multiple ecotypes. With data currently available, it is difficult to infer their characteristic physiology or ecological roles. Altogether, the above proposed ecotypes represented by MLST-A, E and D, as well as other clades that were also distinct yet statistically less compelling, together contribute to the intraspecific diversity of the *S. baltica* species.

Past studies on closely related bacteria strains repeatedly revealed distinct sequence clusters that often relate to ecologically heterogeneous populations. For example, Chisholm's group have identified distinct *Prochlorococcus* clades based on internal transcribed spacer (ITS) regions and several other gene markers and revealed niche partitioning among clades by light intensity, temperature and nutrient conditions (Rocap *et al*., 2002; Johnson *et al*., 2006); Polz and collaborators have found *Vibrio splendidus* populations from a single site north of Boston apportioned themselves according to seasons and particle sizes (Thompson *et al*., 2005; Hunt *et al*., 2008); work of Ward and colleagues on *Synechococcus* populations in hot spring mats in Yellowstone suggested clades identified by 16S rRNA and ITS sequences were specialized with regard to temperature and depth in the mat (Ward *et al*., 2006; 2007); Cohan's group studied populations of *Bacillus simplex* and *B. subtilis–B. licheniformis* clade from 'Evolution Canyons' of Israel and near Death Valley, and showed that specific sequence clusters were specialized to locations with varied levels of solar insolation (Koeppel *et al*., 2008; Connor *et al*., 2010). The ecotype concept has been proposed for such clusters, within which strains are genetically homogeneous and ecologically occupy the same habitat(s) (Cohan, 2001; 2002b). Therefore, it has been argued that ecotypes, rather than the species generally recognized in bacterial systematics, should represent the fundamental unit of bacterial diversity (Gevers *et al*., 2005; Cohan and Perry, 2007; Cohan and Koeppel, 2008). The *S. baltica* ecotypes recognized in this study, especially regarding the MLST clades A and E, clearly represent genetically and ecologically homogeneous populations despite their sympatric distribution. From the evolutionary perspective, identifying ecotypes and studying ecotype differentiation can reveal ecological adaptations that occurred over short time periods, while in the long run, providing further understanding of the mechanisms that develop and maintain the vast bacterial biodiversity.

Temporal dynamics of metabolic patterns among the S. baltica *populations*

The Biolog analysis characterized substantial metabolic differences among the *S. baltica* strains. In particular, temporal distribution was suggested to have a significant impact on their metabolic patterns. Strains isolated from 1998 grew on 44 carbon compounds on average, significantly more versatile than those from the 1980s which grew on only 34 compounds on average. Specifically, capability of using certain compounds including D-fructose, bromo-succinic acid, α-ketobutyric acid, glucose-1-phosphate, acetic acid, α-hydroxybutyric acid, glucose-6-phospate and L-alanine was found in a significantly larger percentage of strains isolated in 1998 compared with those isolated in the 1980s ($P < 0.001$), possibly corresponding to a shift in selective pressure in the Baltic Sea environment. We further looked into the metabolic patterns of OS183 and OS225, both isolated from 1986 but grouped with strains from 1998 (clades J and K). Although both strains share high MLST sequence similarities (100%, > 98%, respectively) with their closely related 1998 neighbours, they grew on fewer carbon sources compared with their 1998 phylogenetic neighbours and their Biolog patterns are more similar to strains from the 1980s (Supporting Information Fig. S6). In both clades, utilization of glucose-6-phosphate, alaninamide, glucose-1-phosphate, D-fructose and α-hydroxybutyric acid was only present in the 1998 strains, likely suggesting similar selective pressures on these 1998 strains.

Fast metabolic changes were also observed among strains over the shorter time stretch from 1986 to 1987. Especially in MLST clade A, within which all members share nearly identical sequences, strains isolated from 1986 and 1987 were mapped to two distinct Biolog clades, A1' and A2'. Biolog-A1' contains strains only from 1986 and used relatively fewer carbon sources. All Biolog-A2' strains were from 1987 and differ from the Biolog-A1' strains mainly at using an additional eight carbon sources (D-psicose, mono-methyl succinate, alaninamide, α-ketovaleric acid, glucose-1-phosphate, D-fructose, bromo-succinic acid and propionic acid), and shared

metabolic patterns more similar to those of the 1998 strains. Altogether, our findings revealed substantial phenotypic heterogeneity within the named species *S. baltica* and at the same sampling station. Fast phenotypic diversification within short evolutionary stretches has been observed in studies regarding both natural and experimental microcosms (Travisano and Lenski, 1996; Riley *et al*., 2001; Coleman and Chisholm, 2007; Walk *et al*., 2007; Koeppel *et al*., 2013), and may arise from processes including mutation, deletion, horizontal acquisition of genes, genome shuffling, etc. Adaptive changes can be rapidly fixed through selective pressure; non-adaptive changes may also spread by hitchhiking with beneficial mutations. Koeppel and colleagues (2013) further proposed that new ecotypes can arise as frequently as adaptive changes within ecotypes, suggesting that speciation may occur faster than expected. Taken together, our results indicate that phenotypic changes can occur faster than phylogeny can reflect, and further highlights the necessity of thorough phenotypic characterization while studying specialization of micro-organisms.

DMSO reduction may be related to specialization to anoxic deep water environment

Because the water chemistry profile exhibited a gradient of H_2S along the water column, we speculated that, other than nitrate, respiration of certain sulfur compounds might as well be involved in specialization of these *S. baltica* strains. Therefore, utilization of sulfur compounds as electron acceptors including DMSO and thiosulfate was tested. DMSO is naturally formed from the degradation of phytoplankton in marine environments and can be used as an alternative electron acceptor for many bacterial species where oxygen is deficient. Capabilities of reducing DMSO by *S. baltica* strains were consistent with gene presence/absence profiles from microarray experiments, with the exception of the strain OS645, which, though not capable of reducing DMSO, possesses all genes in the type I DMSO reductase operon (*dmsEFABGH*). The reason for loss in expression in OS645 is not clear. Furthermore, the expression level of this operon in strain OS195 is upregulated under anaerobic conditions of both nitrate and thiosulfate respiration (unpublished data). Hence, we predict that the ability of reducing DMSO may be related to levels of available electron acceptors in the water environment and is therefore potentially important for adaption to the anaerobic water environment.

Capability to reduce thiosulfate and TMAO was exclusively lacked in clade E

As a major product of anoxic sulfide oxidation, thiosulfate is a key intermediate in the dissimilatory sulfur cycle of aquatic environments (Sievert *et al*., 2007). Because H_2S, often produced during the reduction of thiosulfate, is only detectable at and below the oxic–anoxic interface, we postulate that the ability to reduce thiosulfate to H_2S may be one of the factors causing specialization of *S. baltica* strains. TMAO is also an important alternative electron acceptor in marine waters, and is the precursor to a variety of reduced nitrogenous biogases that mediate the marine N cycle. Reduction of TMAO results in the production of trimethylamine, which has been implicated as a major contributor to spoilage of seafood and fish (Barrett, 1985; Ward *et al*., 2006). The lack of capability to reduce both thiosulfate and TMAO exclusively in MLST-E may indicate selection on this clade against growth in anoxic water. Indeed, other than the nitrate supplemented nutrient broth medium, the RAPD-F clade that corresponds to MLST-E was only recovered from aerobic incubated ZoBell agar plates.

In addition, inability to reduce TMAO by OS155 appeared to result from a single nucleotide deletion at the 1243th position of *torA* (Sbal_3213), the gene encoding a periplasmic protein connected to the quinone pool via the cytochrome TorC, based on comparison of genomic sequences. No matter whether or not the incapability of reducing TMAO by the other strains in MLST-E is attributed to a similar mechanism, we note that a simple change in structural integrity can result in non-activity of the entire operon. Moreover, transcriptomic analysis showed that the TMAO reductase operon (Shew185_3211-Shew185_3214) was upregulated in OS195 under both anaerobic nitrate and thiosulfate respiring conditions, further indicating its potential role in advancing adaptation into anoxic niche conditions (unpublished data).

Conclusion

Through genotyping analyses using MLST, *gyrB* and CGH, we delineated the *S. baltica* strains into clades, some of which likely represent ecological units. We showed that the phylogenetic diversity observed for these *S. baltica* strains coexisting in the redox gradient is due, in part, to nutrient and particle associated conditions. For some of the clades, including the major clades MLST-A and E, some inferences can be made about the nature of their ecological niches, based on which we propose them as distinct ecotypes. In other cases, especially regarding the clades with fewer strains, there are not yet obvious ecological or physiological patterns to infer what niches they may occupy. This study will allow more informed selection of strains for further experimental characterization, as representatives from the MLST clades may also exemplify the larger RAPD genotypes. Furthermore, genetic differences revealed from this study can be used to design specific markers to probe the *in situ* abundance

and distribution of *S. baltica* populations over the redox gradient. The presence of multiple *S. baltica* ecotypes at a given site should allow for survival over a broader range of conditions and may contribute to their ecological success in the Baltic Sea environment, including the disruptive water turnover of 1993. Last but not the least, ecotypes revealed from this study, as well as those from many others, though not yet named into distinct species, hopefully will be granted some taxonomic status in the future that will recognize their genetic and ecologic distinctiveness.

Experimental procedures

Bacterial strains and growth conditions

Isolation procedures of *S. baltica* strains were previously described (Ziemke *et al*., 1997). For small inocula, small tips with narrow entrances were used, hence reducing the possibility to inoculate particles from the water samples. For larger inocula, bigger tips were used, which increases the likeliness to transfer particles into the medium. For 10 ml inocula, large glasses or plastic pipettes with wide entrances were used.

Luria-Bertani medium (Acumedia, Lansing, MI, USA) and modified M1 medium (Cruz-Garcia *et al*., 2006) were used for bacterial growth. Anaerobic M1 medium was made under O_2-free argon gas. Denitrification was tested in anaerobic M1 medium with 2 mM sodium nitrate as the sole electron acceptor. Durham tubes were inserted into the serum tubes for gas collection. Nitrate levels were tested after 3 days of incubation at 22°C using nitrate test strips (Merckoquant, limit of detection 160 μM, VWR International, Radnor, PA, USA). Thiosulfate reduction was tested in Triple Sugar Iron agar (TSI agar, Acumedia, Lansing, MI, USA). Strains were inoculated into the bottom of test tubes containing TSI agar and incubated at 22°C. Strains were classified as incapable of reducing thiosulfate into hydrogen sulfide (H_2S) if the agar did not turn black within a week. DMSO and TMAO reduction was tested in anaerobic M1 medium supplemented with 10 mM of each compound as electron acceptors, respectively, after which OD_{600} was monitored for 3 days. Strains were characterized as not capable of reducing DMSO or TMAO if the OD_{600} did not exceed 0.02 at the end of incubation; *S. oneidensis* MR-1 was used as the positive control. Testing of denitrification, thiosulfate, DMSO and TMAO reduction was repeated three times for all strains.

gyrB and MLST analysis

Seven genes from the conserved gene core of nine *Shewanella* genomes were selected based on their calculated reliability for reflecting genome phylogeny (Konstantinidis *et al*., 2006). Primers for polymerase chain reactions (PCR) were designed for these genes based on genomic sequences of four sequenced *S. baltica* strains (OS155, OS185, OS195 and OS223). GenBank accession numbers for deposited MLST sequences are KC785129–KC785450. *gyrB* genes and MLST genes were amplified by PCR (Supporting Information Table S1), their products were purified using ExoSap-IT (Affymetrix, Cleveland, OH, USA) and sequenced by the Michigan State University Research Technology Support Facility. Sequences were aligned with MUSCLE (Edgar, 2004). The Neighbor Joining tree was built using concatenated sequences of the seven genes by MEGA version 4 (MEGA, Tempe, AZ, USA) (Tamura *et al*., 2007).

The software DnaSP v5.10 (Librado and Rozas, 2009) was used to calculate the ratios of synonymous and non-synonymous substitution rates (Ks and Ka, respectively) and to perform Tajima's D test for neutrality. Population mutation and recombination rates were estimated using a likelihood based method implemented in the LDhat version 2.1 (McVean *et al*., 2002) using similar parameters as described in Konstantinidis and DeLong (2008). Distance matrices from *gyrB* and MLST analysis were generated using maximum composite likelihood model in MEGA 4.0.

CGH clustering and comparison with gyrB and MLST phylogeny

Microarray construction and analysis was performed as described in (Caro-Quintero *et al*., 2010). A Jaccard index was used for constructing distance matrix based on the gene presence/absence information (Levandowsky and Winter, 1971). The dendrogram was obtained using the average linkage agglomeration method in R version 2.9.2 (R core team, Vienna, Austria) (Team, 2005). Distance matrices of the *gyrB* gene and concatenated sequences from MLST analysis were generated by maximum composite likelihood model in Mega 4.0. All matrices were standardized before calculation of the mean squared error.

Habitat assignment

Habitat association was determined through AdaptML (Hunt *et al*., 2008), with the following parameter setting: Initial habitat number of 10, collapse threshold at 0.10, converge threshold at 0.00001, number of randomized topologies of 1000 and method for inferring mu or average habitat transition with 'avg'. Habitat association was mapped to the MLST phylogeny with the interactive tree of life (iTOL) programme (Letunic and Bork, 2011).

Estimation of whole-genome recombination rates

The construction of the genome phylogeny was carried out by Mr. Bayes (Altekar *et al*., 2004) based on 2894 concatenated core genes as defined by reciprocal best matches as detailed in (Luo *et al*., 2011). Five clades were observed (Fig. 1) from which intra- and inter-clade horizontal gene transfer rates were calculated. Specifically, we followed EQDA as introduced in Zhaxybayeva and colleagues (2006) and implemented in Luo and colleagues (2011). OS185 and OS223 were discarded from EQDA because they did not meet the minimal requirement (at least two genomes present in a clade). The cut-offs for synonymous mutation rate (Ks) and sites were set to be 0.02 and 50, respectively, to robustly capture recent recombination events (Luo *et al*., 2011).

Biolog analysis

Metabolic profiles of *S. baltica* strains were determined with BIOLOG-GN plates (BIOLOG Inc., Hayward, CA, USA),

essentially as described by Höfle and colleagues (2000). Distance matrices were calculated for the CGH profile and the Biolog dataset based on Jaccard index (Levandowsky and Winter, 1971). MLST distance matrix was exported based on the MLST phylogeny constructed as described above. After regression of the Biolog distances against MLST distances, the phylogenetic independent residuals were taken and used to estimate the correlation with genomic distances based on CGH, and the effects from environmental factors.

Acknowledgements

We thank the *Shewanella* Federation for supporting work on *Shewanella* genomics, and the Joint Genome Institute for sequencing selected strains. We also thank Amanda Geaslin for assistance with laboratory work. This work was partially supported by the US Department of Energy under Contract No. DE-FG02-07ER64389. The project is a component of the *Shewanella* Federation and contributes to the overall goal of applying the tools of genomics to better understand the ecophysiology and speciation of a microbial group with versatile metal-reducing and other electron-accepting capacities.

References

Altekar, G., Dwarkadas, S., Huelsenbeck, J.P., and Ronquist, F. (2004) Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics* **20:** 407–415.

Baggs, E.M. (2008) A review of stable isotope techniques for N_2O source partitioning in soils: recent progress, remaining challenges and future considerations. *Rapid Commun Mass Spectrom* **22:** 1664–1672.

Barrett, E.L. (1985) Bacterial reduction of trimethylamine oxide. *Annu Rev Microbiol* **39:** 131-149.

Bleakley, B.H., and Tiedje, J.M. (1982) Nitrous oxide production by organisms other than nitrifiers or denitrifiers. *Appl Environ Microbiol* **44:** 1342–1348.

Brettar, I., and Höfle, M.G. (1993) Nitrous oxide producing heterotrophic bacteria in the water column of the Baltic Sea: abundance and molecular identification. *Mar Ecol Prog Ser* **94:** 253–265.

Brettar, I., and Rheinheimer, G. (1991) Denitrification in the central Baltic: evidence for H_2S-oxidation as motor of denitrification at the oxic-anoxic interface. *Mar Ecol Prog Ser* **77:** 157–169.

Brettar, I., and Rheinheimer, G. (1992) Influence of carbon availability on denitrification in the central Baltic Sea. *Limnol Oceanogr* **37:** 1146–1163.

Brettar, I., Moore, E.R.B., and Höfle, M.G. (2001) Phylogeny and abundance of novel denitrifying bacteria isolated from the water column of the central Baltic Sea. *Microb Ecol* **42:** 295–305.

Brettar, I., Christen, R., and Höfle, M.G. (2002) *Shewanella denitrificans* sp. nov., a vigorously denitrifying bacterium isolated from the oxic-anoxic interface of the Gotland Deep in the central Baltic Sea. *Int J Syst Evol Microbiol* **52:** 2211–2217.

Brettar, I., Christen, R., and Höfle, M.G. (2012) Analysis of bacterial core communities in the central Baltic by comparative RNA-DNA-based fingerprinting provides links to structure-function relationships. *ISME J* **6:** 195–212.

Caro-Quintero, A., Deng, J., Auchtung, J., Brettar, I., Höfle, M.G., Klappenbach, J., and Konstantinidis, K.T. (2010) Unprecedented levels of horizontal gene transfer among spatially co-occurring *Shewanella* bacteria from the Baltic Sea. *ISME J* **5:** 131–140.

Caro-Quintero, A., Auchtung, J., Deng, J., Brettar, I., Höfle, M., Tiedje, J.M., and Konstantinidis, K.T. (2011) Genome sequencing of five *Shewanella baltica* strains recovered from the oxic-anoxic interface of the Baltic Sea. *J Bacteriol* **194:** 1236.

Cohan, F. (2002a) Sexual isolation and speciation in bacteria. *Genetica* **116:** 359–370.

Cohan, F.M. (2001) Bacterial species and speciation. *Syst Biol* **50:** 513–524.

Cohan, F.M. (2002b) What are bacterial species? *Annu Rev Microbiol* **56:** 457–487.

Cohan, F.M., and Koeppel, A.F. (2008) The origins of ecological diversity in prokaryotes. *Curr Biol* **18:** R1024–R1034.

Cohan, F.M., and Perry, E.B. (2007) A systematics for discovering the fundamental units of bacterial diversity. *Curr Biol* **17:** R373–R386.

Coleman, M.L., and Chisholm, S.W. (2007) Code and context: prochlorococcus as a model for cross-scale biology. *Trends Microbiol* **15:** 398–407.

Connor, N., Sikorski, J., Rooney, A.P., Kopac, S., Koeppel, A.F., Burger, A., *et al*. (2010) Ecology of speciation in the genus bacillus. *Appl Environ Microbiol* **76:** 1349–1358.

Cruz-Garcia, C., Murray, A.E., Klappenbach, J.A., Stewart, V., and Tiedje, J.M. (2006) Respiratory nitrate ammonification by *Shewanella oneidensis* MR-1. *J Bacteriol* **189:** 656–662.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32:** 1792–1797.

Elmgren, R. (1989) Man's impact on the ecosystem of the Baltic Sea: energy flows today and at the turn of the century. *Ambio* **18:** 326–332.

Elmgren, R. (2001) Understanding human impact on the Baltic ecosystem: changing views in recent decades. *Ambio* **30:** 222–231.

Felsenstein, J. (1985) Phylogenies and the comparative method. *Am Nat* **125:** 1–15.

Fredrickson, J.K., Romine, M.F., Beliaev, A.S., Auchtung, J.M., Driscoll, M.E., Gardner, T.S., *et al*. (2008) Towards environmental systems biology of Shewanella. *Nat Rev Microbiol* **6:** 592–603.

Gevers, D., Cohan, F.M., Lawrence, J.G., Spratt, B.G., Coenye, T., Feil, E.J., *et al*. (2005) Re-evaluating prokaryotic species. *Nat Rev Microbiol* **3:** 733–739.

Håkansson, B., Broman, B., and Dahlin, H. (1993) *The Flow of Water and Salt in the Sound During the Baltic Major Inflow Event in January 1993*. ICES Statutory Meeting Dublin, Paper CM1993/C:57.

Höfle, M.G., Ziemke, F., and Brettar, I. (2000) Niche differentiation of *Shewanella putrefaciens* populations from the

Baltic as revealed by molecular and metabolic fingerprinting. In *Microbial Biosystems: New Frontiers, Proceedings of ISME8*. Bell, C.R., Brylinsky, M., and Johnson-Green, P. (eds). Halifax, Canada, pp. 135–143.

Hunt, D.E., David, L.A., Gevers, D., Preheim, S.P., Alm, E.J., and Polz, M.F. (2008) Resource partitioning and sympatric differentiation among closely related bacterioplankton. *Science* **320:** 1081–1085.

Johnson, Z.I., Zinser, E.R., Coe, A., McNulty, N.P., Woodward, E.M.S., and Chisholm, S.W. (2006) Niche partitioning among prochlorococcus ecotypes along ocean-scale environmental gradients. *Science* **311:** 1737–1740.

Kim, J.-S., Makama, M., Petito, J., Park, N.-H., Cohan, F.M., and Dungan, R.S. (2012) Diversity of Bacteria and Archaea in hypersaline sediment from Death Valley National Park, California. *MicrobiologyOpen* **1:** 135–148.

Koeppel, A., Perry, E.B., Sikorski, J., Krizanc, D., Warner, A., Ward, D.M., *et al*. (2008) Identifying the fundamental units of bacterial diversity: a paradigm shift to incorporate ecology into bacterial systematics. *Proc Natl Acad Sci USA* **105:** 2504–2509.

Koeppel, A.F., Wertheim, J.O., Barone, L., Gentile, N., Krizanc, D., and Cohan, F.M. (2013) Speedy speciation in a bacterial microcosm: new species can arise as frequently as adaptations within a species. *ISME J* **7:** 1080–1091.

Konstantinidis, K.T., and DeLong, E.F. (2008) Genomic patterns of recombination, clonal divergence and environment in marine microbial populations. *ISME J* **2:** 1052–1065.

Konstantinidis, K.T., Ramette, A., and Tiedje, J.M. (2006) Toward a more robust assessment of intraspecies diversity, using fewer genetic markers. *Appl Environ Microbiol* **72:** 7286–7293.

Letunic, I., and Bork, P. (2011) Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res* **39:** W475–W478.

Levandowsky, M., and Winter, D. (1971) Distance between sets. *Nature* **234:** 34–35.

Librado, P., and Rozas, J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25:** 1451–1452.

Luo, C., Walk, S.T., Gordon, D.M., Feldgarden, M., Tiedje, J.M., and Konstantinidis, K.T. (2011) Genome sequencing of environmental *Escherichia coli* expands understanding of the ecology and speciation of the model bacterial species. *Proc Natl Acad Sci USA* **108:** 7200–7205.

McVean, G., Awadalla, P., and Fearnhead, P. (2002) A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* **160:** 1231–1241.

Matthäus, W., and Lass, H.-U. (1995) The recent salt water inflow into the Baltic Sea. *J Phys Oceanogr* **25:** 280–286.

Matthäus, W., Nehring, D., Ulrich Lass, H., Nausch, G., Nagel, K., and Siegel, H. (1996) The Baltic Sea in 1996 – continuation of stagnation and decreasing phosphate concentrations. *Dtsch Hydrogr Z* **48:** 161–174.

Nealson, K.H., and Scott, J. (2006) Ecophysiology of the genus *Shewanella*. *Prokaryotes* **6:** 1133–1151.

Neretin, L., Pohl, C., Jost, G., Leipe, T., and Pollehne, F. (2003) Manganese cycling in the Gotland Deep, Baltic Sea. *Mar Chem* **82:** 125–143.

Riley, M.S., Cooper, V.S., Lenski, R.E., Forney, L.J., and Marsh, T.L. (2001) Rapid phenotypic change and diversification of a soil bacterium during 1000 generations of experimental evolution. *Microbiology* **147:** 995–1006.

Rocap, G., Distel, D.L., Waterbury, J.B., and Chisholm, S.W. (2002) Resolution of prochlorococcus and Synechococcus ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. *Appl Environ Microbiol* **68:** 1180–1191.

Rönner, U. (1985) Nitrogen transformations in the Baltic proper-denitrification counteracts eutrophication. *Ambio* **14:** 134–138.

Rönner, U., and Sörensson, F. (1985) Denitrification rates in the low-oxygen waters of the stratified Baltic proper. *Appl Environ Microbiol* **50:** 801–806.

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

Shapiro, B.J., Friedman, J., Cordero, O.X., Preheim, S.P., Timberlake, S.C., Szabó, G., *et al*. (2012) Population genomics of early events in the ecological differentiation of bacteria. *Science* **336:** 48–51.

Sievert, S.M., Kiene, R.P., and Schulz-Vogt, H.N. (2007) The sulfur cycle. *Oceanography* **20:** 117–123.

Stigebrandt, A. (1987) A model for the vertical circulation of the Baltic deep water. *J Phys Oceanogr* **17:** 1772–1786.

Struck, U., Pollehne, F., Bauerfeind, E., and Bodungen, B.V. (2004) Sources of nitrogen for the vertical particle flux in the Gotland Sea (Baltic Proper) – results from sediment trap studies. *J Mar Syst* **45:** 91–101.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24:** 1596–1599.

Team, R.D.C. (2005) *R: a language and environment for statistical computing, reference index version 2.9.2*. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3–900051-07-0. URL http://www.R-project.org/.

Thompson, J.R., Pacocha, S., Pharino, C., Klepac-Ceraj, V., Hunt, D.E., Benoit, J., *et al*. (2005) Genotypic diversity within a natural coastal bacterioplankton population. *Science* **307:** 1311–1313.

Travisano, M., and Lenski, R.E. (1996) Long-term experimental evolution in *Escherichia coli*. IV. Targets of selection and the specificity of adaptation. *Genetics* **143:** 15–26.

Venkateswaran, K., Moser, D.P., Dollhopf, M.E., Lies, D.P., Saffarini, D.A., MacGregor, B.J., *et al*. (1999) Polyphasic taxonomy of the genus *Shewanella* and description of *Shewanella oneidensis* sp. nov. *Int J Syst Bacteriol* **49:** 705–724.

Walk, S.T., Alm, E.W., Calhoun, L.M., Mladonicky, J.M., and Whittam, T.S. (2007) Genetic diversity and population structure of *Escherichia coli* isolated from freshwater beaches. *Environ Microbiol* **9:** 2274–2288.

Ward, D.M., Bateson, M.M., Ferris, M.J., Kühl, M., Wieland, A., Koeppel, A., and Cohan, F.M. (2006) Cyanobacterial ecotypes in the microbial mat community of Mushroom Spring (Yellowstone National Park, Wyoming) as species-

like units linking microbial community composition, structure and function. *Philos Trans R Soc Lond B Biol Sci* **361:** 1997–2008.

Ward, D.M., Cohan, F.M., Bhaya, D., Heidelberg, J.F., Kuhl, M., and Grossman, A. (2007) Genomics, environmental genomics and the issue of microbial species. *Heredity* **100:** 207–219.

Zhaxybayeva, O., Gogarten, J.P., Charlebois, R.L., Doolittle, W.F., and Papke, R.T. (2006) Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events. *Genome Res* **16:** 1099–1108.

Ziemke, F., Brettar, I., and Höfle, M.G. (1997) Stability and diversity of the genetic structure of a *Shewanella putrefaciens* population in the water column of the central Baltic. *Aquat Microb Ecol* **13:** 63–74.

Ziemke, F., Höfle, M.G., Lalucat, J., and Rossellö-Mora, R. (1998) Reclassification of *Shewanella putrefaciens* Owen's genomic group II as *Shewanella baltica* sp. nov. *Int J Syst Bacteriol* **48:** 179–186.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Neighbour Joining tree of 46 *S. baltica* strains produced from *gyrB* sequences. The strain *S. oneidensis* MR-1 was used as out-group but is not shown. Bootstrap values above 90% are shown on branches based on 1000 bootstrap replicates.

Fig. S2. Hierarchical clustering of the CGH profile. Gene presence and absence information extracted from the CGH profile was used for constructing the distance matrix of strains using the Jaccard index. Hierarchical clustering was performed using the average linkage agglomeration method in R software version 2.9.2.

Fig. S3. Phylogeny based on 3079 commonly shared core genes among the nine *S. baltica* strains. Genome names are colour-coded for the clade they belong to.

Fig. S4. Recombination network on recently diverged commonly shared genes between at least two clades (core genome; with Ks < 0.02 and synonymous sites > 50 cut-off). The corresponding MLST clades were labelled next to the strain or the branch. We could not examine clades with only one member due to the limit of embedded quartet decomposition analysis (e.g. OS223 and OS185). The number of core genes exchanged between genomes range 0~35.

Fig. S5. Regression of Biolog dissimilarity and genomic distance. Pairwise genomic distances were estimated based on gene presence/absence matrix from the CGH data using the Jaccard index. Biolog dissimilarities were phylogenetic independent contrasts from Jaccard-based Biolog distances controlled by phylogenetic relatedness based on MLST.

Fig. S6. Hierarchical clustering of the Biolog profile. Pairwise Biolog dissimilarities were calculated using the Jaccard index. Hierarchical clustering was performed using the average linkage agglomeration method. Coloured letters identify the clades consistent with those from the MLST phylogeny.

Table S1. Primers used for MLST analysis.