

ORIGINAL ARTICLE

Unprecedented levels of horizontal gene transfer among spatially co-occurring *Shewanella* bacteria from the Baltic Sea

Alejandro Caro-Quintero¹, Jie Deng², Jennifer Auchtung², Ingrid Brettar³, Manfred G Höfle³, Joel Klappenbach² and Konstantinos T Konstantinidis^{1,4}

¹School of Biology, Georgia Institute of Technology, Atlanta, GA, USA; ²Center for Microbial Ecology, Michigan State University, East Lansing, MI, USA; ³Department of Vaccinology and Applied Microbiology, Helmholtz Centre for Infection Research, Braunschweig, Germany and ⁴School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA, USA

High-throughput sequencing studies during the last decade have uncovered that bacterial genomes are very diverse and dynamic, resulting primarily from the frequent and promiscuous horizontal gene exchange that characterizes the bacterial domain of life. However, a robust understanding of the rates of genetic exchange for most bacterial species under natural conditions and the influence of the ecological settings on the rates remain elusive, severely limiting our view of the microbial world. Here, we analyzed the complete genomic sequences and expressed transcriptomes of several *Shewanella baltica* isolates recovered from different depths in the Baltic Sea and found that isolates from more similar depths had exchanged a larger fraction of their core and auxiliary genome, up to 20% of the total, compared with isolates from more different depths. The exchanged genes seem to be ecologically important and contribute to the successful adaptation of the isolates to the unique physicochemical conditions of the depth. Importantly, the latter genes were exchanged in very recent past, presumably as an effect of isolate's seasonal migration across the water column, and reflected sexual speciation within the same depth. Therefore, our findings reveal that genetic exchange in response to environmental settings may be surprisingly rapid, which has important broader impacts for understanding bacterial speciation and evolution and for modeling bacterial responses to human-induced environmental impacts.

The ISME Journal (2011) 5, 131–140; doi:10.1038/ismej.2010.93; published online 1 July 2010

Subject Category: microbial ecology and functional diversity of natural habitats

Keywords: comparative genomics; genome evolution; speciation; horizontal gene transfer; recombination; *Shewanella baltica*

Introduction

High-throughput sequencing during the last decade has revealed that bacterial genomes are much more diverse and dynamic than anticipated earlier (Lawrence and Ochman, 2002; Welch *et al.*, 2002; Konstantinidis and Tiedje, 2005). For instance, gene content variation among strains of the same bacterial species may comprise 30–35% of the genes in the genome (Konstantinidis and Tiedje, 2005; Tettelin *et al.*, 2005). This gene diversity and genome fluidity frequently underlies the emergence of new pathogens and the natural attenuation of important environmental pollutants, and hence, has important

health and economical consequences (Handelsman *et al.*, 2007). Horizontal gene transfer (HGT) accounts for a substantial fraction, if not the majority, of the bacterial genomic fluidity and diversity (Beiko *et al.*, 2005; Zhaxybayeva *et al.*, 2006; Lang and Beatty, 2007). However, a robust understanding of the rates of genetic exchange for most bacterial species under natural conditions and the influence of the ecological settings on the rates remain elusive (Gevers *et al.*, 2005; Konstantinidis *et al.*, 2006a; Handelsman *et al.*, 2007). An improved understanding of these issues has important broader impacts such as for reliable diagnosis of infectious disease agents, bioremediation efforts and for robust modeling of bacterial evolution and speciation.

Stratified aquatic systems are characterized by sharp physical, chemical and nutrient gradients and thus, offer unique opportunities for studying the role of the environment in shaping population (and genome) structure and dynamics. One such system, which is among the most stable systems on the

Correspondence: KT Konstantinidis, School of Civil and Environmental Engineering and School of Biology, Georgia Institute of Technology, 311 Ferst Drive, ES&T, Room 3224, Atlanta, GA 30332-0512, USA.

E-mail: kostas@ce.gatech.edu

Received 5 March 2010; revised 11 May 2010; accepted 14 May 2010; published online 1 July 2010

planet (for example, water retention time in the order of 20–30 years (Neumann, 2006)) and has been characterized extensively because of its long history of pollutant contamination, is the Baltic Sea (Backer *et al.*, 2010). *Shewanella baltica* dominates the pool of heterotrophic nitrate-reducing bacteria isolated from the oxic–anoxic interface of the Baltic Sea (Brettar *et al.*, 2001). For instance, *S. baltica* organisms (strains) accounted for 32–80% of total cultivable denitrifying bacteria under different growth conditions during our isolation efforts in 1986 (Ziemke *et al.*, 1997). These findings further corroborate the important role of *Shewanella* bacteria in cycling of organic and inorganic materials at redox interfaces (Myers and Nealson, 1988; Fredrickson *et al.*, 2008).

To identify the genetic elements that enable *S. baltica* to adapt to redox gradients and provide novel insights into the mechanisms and rates of genomic adaptation, we performed whole-genome sequence and DNA–DNA microarray comparative analyses of a large collection of isolates from the Baltic Sea ($n = 116$, Supplementary Figure S1). Our analyses revealed that *S. baltica* genomic adaptation to environmental settings, mediated by HGT, may be much more rapid and extensive compared with what was seen previously in other marine bacteria.

Materials and methods

Organisms used in this study

The *S. baltica* strains used in this study were isolated on denitrifying media (NHNO₃, THNO₃) or anaerobic ZoBell agar. More details on sampling, isolation conditions and genome fingerprinting patterns of each strain are provided in Ziemke *et al.* (1997). The complete genome sequences of the four *S. baltica* strains used in the study were obtained from GeneBank (Benson *et al.*, 2009). The strains and their GeneBank accession numbers were OS195 (NC_009997, NC_009998, NC_009999, NC_010000), OS185 (NC_009665, NC_009661), OS155 (NC_009052, NC_009035, NC_009036, NC_009037, NC_009038) and OS223 (NC_011663, NC_011664, NC_011665, NC_011668).

Identification of orthologs

Orthologs among the four *S. baltica* genomes were identified using a reciprocal best-match blastn approach, essentially as described earlier (Konstantinidis *et al.*, 2009). In brief, the sequences of the predicted genes in the genome of strain OS195 were searched, using the blastn algorithm (Altschul *et al.*, 1997), against the genomic sequence of each of the remaining three strains. The best match for each query gene, when better than at least 70% overall nucleotide identity (recalculated to an identity along the entire sequence) and an alignable region covering >70% of the length of the query gene

sequence, was extracted using a custom PERL script and searched against the complete gene complement of OS195 to identify reciprocal best matches. Such reciprocal best matches conserved genes were denoted as orthologs. Orthologs conserved in all four genomes were denoted as core orthologous genes. Genes that found no match better than the earlier standards against any of the remaining three genomes were denoted as OS195 specific (strain specific). Genes conserved in some but not all of the strains were denoted as variable (Supplementary Table S1, which includes all OS195 genes).

Recombination analysis

Recombination fragments were detected using a custom-made approach, essentially as described earlier (Konstantinidis and DeLong, 2008; Caro-Quintero *et al.*, 2009). Briefly, the genomic sequence of OS195 was cut *in silico* in 500 bp-long consecutive sequence fragments. The fragments were subsequently searched against the other *S. baltica* genomes for best matches, using blastn as described above for orthologs. A fragment was flagged as (potentially) recombined in another strain when its best blastn match in the latter strain showed >99.5% nucleotide identity, whereas its identity in the remaining strains was <98%, which corresponded to the average sequence divergence between the *S. baltica* strains (that is ~96.7%). Such fragments and their adjacent fragments were subsequently visually inspected to determine the presence of recent homologous recombination as shown graphically in Figure 2b. The recombined fragments identified this way were further validated by the genetic algorithm for recombination detection (GARD) (Kosakovsky Pond *et al.*, 2006a). Briefly, all core genes in all genomes were concatenated to provide a whole-genome core gene alignment. The alignment was scanned in 1 or 2 Kbp-long windows by GARD (longer windows are too computationally demanding for GARD) in a pairwise manner (that is two genomes at a time) and the sequence windows that provided δ AIC values higher than ~10 were flagged as containing recombined segments, as suggested earlier (Kosakovsky Pond *et al.*, 2006a). The recombined fragments identified by GARD were contrasted with those identified by visual inspection of the nucleotide identity patterns (blastn approach). Sequence fragments or genes that showed high nucleotide identity (>98%) between all four genomes encoded typically for highly conserved housekeeping genes such as the rRNA operon genes. Such fragments were excluded from the recombination analysis because it could not be established whether the identity patterns observed were due to recombination or high sequence conservation. Fewer than 100 fragments were excluded from the analysis for the latter reason (from >3000, in total; see Supplementary Table S1). The number of synonymous substitutions per synonymous site

(Ks) for every gene was calculated based on the nucleotide codon-based alignment using the codeml module of the PAML package (Yang, 2007).

DNA microarray construction and analysis

A detailed description of the DNA–DNA and gene expression microarray experiments is provided in the Supplementary Material.

Results

Unprecedented levels of genetic exchange among spatially co-occurring *S. baltica* strains

To unravel the genetic diversity within our *S. baltica* strain collection, four strains that represented the most abundant lineages recovered among the 116 isolates comprising our collection (Supplementary Figure S1) were fully sequenced. These strains were OS155, OS185, OS223 and OS195 and were recovered from three different depths of the Baltic Sea, 90, 120, 120 and 140 m, respectively. These depths were characterized by different redox potentials and nutrient availability at the time of isolation. In particular, the 140-m depth represented an anoxic environment, with absence of oxygen, very low nitrate concentrations (below $0.5 \mu\text{mol l}^{-1}$) and high abundance of alternative electron acceptors such as sulfur compounds (for example, S° , thiosulfate) and metal oxides. The 120 and 90 m depths were highly comparable in terms of low oxygen and high nitrate concentrations; at 120 m reduced sulfur compounds

were also (occasionally) present, originating presumably from deeper anoxic waters (Figure 1a).

The four *S. baltica* genomes showed very similar evolutionary relatedness among each other, for example, they had identical 16S rRNA gene sequences. To provide for a higher resolution, the genome-aggregate average nucleotide identity (ANI) (Konstantinidis and Tiedje, 2005) of all core genes ($n = 2500$) with no detectable signal of recombination according to PhiTest analysis (Bruen *et al.*, 2006) was used. ANI analysis revealed that these four genomes were not only very closely related but also show comparable evolutionary relatedness among each other, with their ANI values being $\sim 96.7\%$ for each pair of genomes compared (Figure 1b). These values are higher than the 95% ANI that corresponds to the 70% DNA–DNA hybridization standard frequently used for species demarcation (Goris *et al.*, 2007); hence, these genomes belong justifiably to the same species, *S. baltica*.

Despite the comparable evolutionary relatedness among all strains, strains from more similar depths shared, in general, substantially more genes compared with strains from more different depths. For instance, OS195 shared 580 (non-core) genes with OS185 and 350 with OS223, but none of these three strains shared >150 genes with OS155 (Supplementary Figure S2). Remarkably, most (that is ~ 350) of the 580 genes shared exclusively between OS195 and OS185 and an additional $\sim 10\%$ of their core genes showed 99.5–100% nucleotide identity between OS185 and OS195, contrasting sharply with

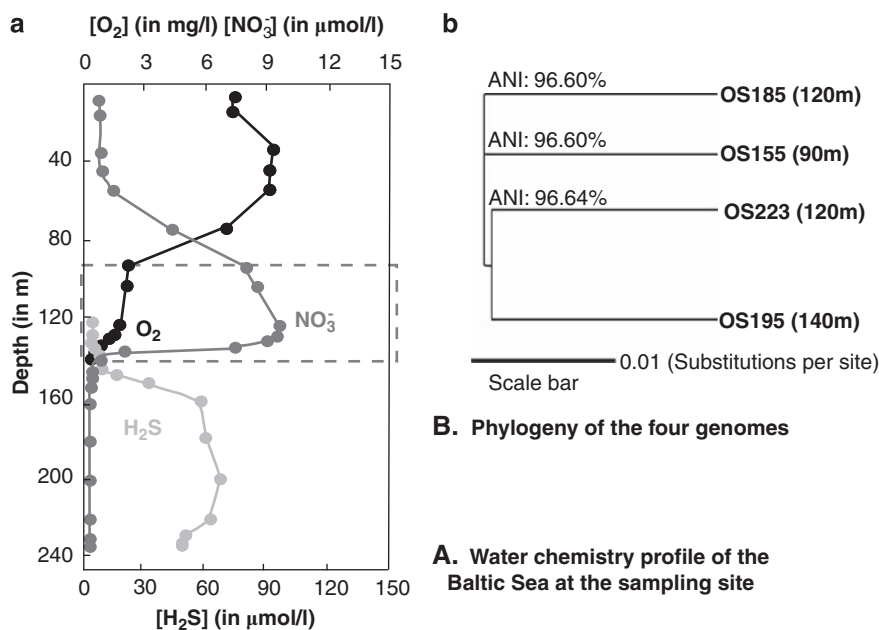


Figure 1 The *S. baltica* genomes. (a) The water chemistry profile at the site of isolation of the four genomes. Note that the appearance of H_2S at around 140 m depth is, at least in part, because of the reduction of sulfur compounds, including sulfur disproportionation. The whole-genome phylogeny of the genomes based on maximum likelihood analysis of the concatenated sequences of all core genes ($n = \sim 2500$) that showed no evidence of recombination, performed as described earlier (Konstantinidis *et al.*, 2006b), is shown in (b). ANI values among the genomes based on the non-recombined core genes are also provided. (a) Adapted from Brettar *et al.* (2001).

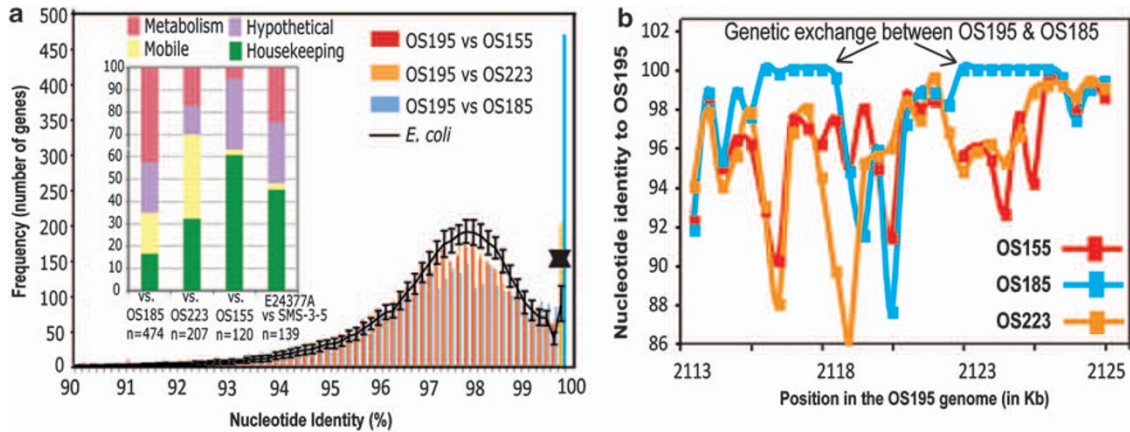


Figure 2 Nucleotide identity distribution of orthologous genes in the *S. baltica* genomes. (a) All genes in the OS195 genome were compared with their orthologs in strain OS185, OS223 and OS155. For each pairwise comparison (see figure key), the number of orthologs is plotted against their nucleotide identity. The solid line represents the average of 125 comparison of between *E. coli* genomes with similar ANI (~97%) and number of orthologs genes (~3500) with the *S. baltica* genomes. Error bars represent 1 s.d. from the mean and the 'X' represents the value of the most outlier *E. coli* genome pair (discussed in the text). The inset in (a) shows the functional annotation of the 100% nucleotide identity genes identified for each pairwise comparison (for details, see text). An graphical representation of the type of recent genetic exchange events assessed by our analysis is provided in (b). Note that the sequences of OS155 and OS223 show consistently lower, and close to the genome average, nucleotide identities to their recombined counterparts in OS195 and OS185.

~97% identity for the rest of the genes in the genome and <3% of high identity core genes (that is 99.5–100%) among the remaining pairs of genomes, respectively (Supplementary Table S1). This pattern became more obvious when the frequency of genes was plotted against their nucleotide identity for each pair of genomes compared (nucleotide identity histograms, see Figure 2a). Notably, a similar analysis of all pairs of genomes available in GenBank with similar ANI (96.5–97.5%) and genome size (3500–4500 genes) to the *S. baltica* genome pairs revealed that the gene nucleotide identity distribution in the OS185 vs OS195 case was unparalleled and significantly different from any other distribution based on the z-test (P -value <0.001). For instance, among the 125 pairwise comparisons of all available *Escherichia coli* genomes, only *E. coli* strains E24377A and SMS-3-5 had about 150 genes with higher than 99.5% nucleotide identity (still, four times fewer genes compared with the OS185 and OS195 case; see Figure 2a). We also observed about 200 genes with >99.5% nucleotide identity between OS195 and OS223, while comparing OS155 against OS195, OS185 or OS223 did not reveal more high identity genes than the average of all genome pairs from Genbank (that is $n < 100$).

As all *S. baltica* genomes show comparable evolutionary divergence among each other (Figure 1b), the high identity genes shared between OS185 and OS195 cannot be attributed simply to higher evolutionary relatedness between these two genomes. These findings cannot be explained by preferential deletion of the corresponding genes in OS155 or OS223 either, because the pool of high identity genes included several core genes that showed nucleotide identities in the 95–98% range against their OS155 or OS223 orthologs (see also

below). Instead, these findings are, most likely, attributed to recent extensive horizontal exchange between OS195 and OS185 or their immediate ancestors.

Unconstrained homologous recombination mediates the genetic exchange events

To further validate the earlier findings and provide insights into the mechanisms mediating the genetic exchange among the *S. baltica* strains, we examined the functional role of all 100% nucleotide identity genes shared between OS195 and OS185. The genes were assigned to one of the following four categories: (1) genes related to metabolism and regulation, (2) mobile elements (integrases, transposases and genes contained within prophages, integrons and plasmids), (3) hypothetical and (4) housekeeping genes (genes related to central cell functions such as replication and translation), which tend to be more conserved than the genome average at the sequence level (Konstantinidis and DeLong, 2008). The analysis showed that most of the genes were neither housekeeping nor mobile; instead, most of them encoded for metabolic, transport and regulatory functions related mainly to secondary metabolism. This functional gene distribution contrasted strikingly with that of the OS195 vs OS155 pair or the *E. coli* genome pairs, which were enriched in housekeeping and hypothetical genes (Figure 2a, inset). Thus, the majority of the exchanged genes do not seem to be the product of a single, specialized vector of HGT such as a bacteriophage or a plasmid.

Further examination of the nucleotide identity patterns of the recently exchanged core genes showed that these genes have been brought into the genome through a homologous recombination mechanism. For instance, the nucleotide identity of the exchanged core genes between OS195 and

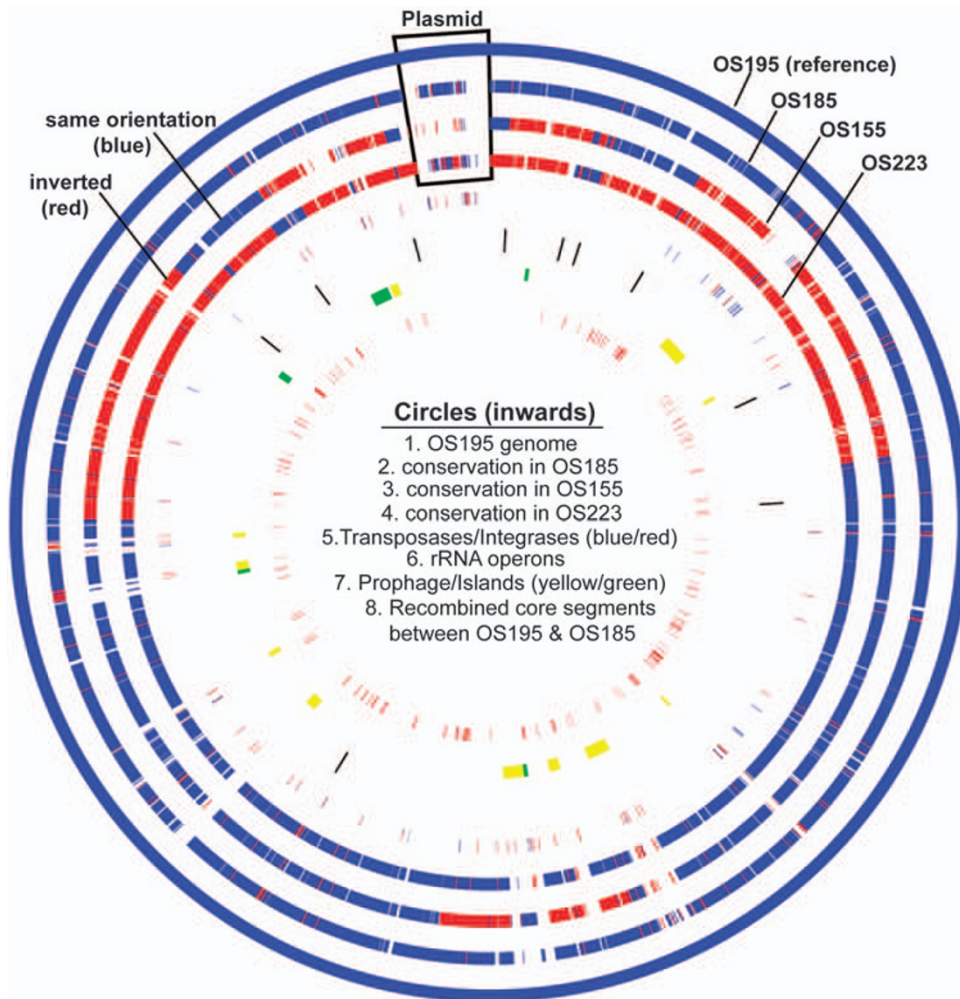


Figure 3 Preferential genome wide and extensive genetic exchange between the *S. baltica* genomes. Circles represent (inwards): the genome of OS195 (1); the conservation of the OS195 genome in OS185 (2), OS155 (3) and OS223 (4), with red denoting segments of the genome that have been inverted in the latter genomes relative to the OS195; the positions of transposase (blue) and integrase (red) genes in the genome of the OS195 (5); the position of the rRNA operons (6); all genomic islands shared between OS195 and OS185, colored either yellow if they corresponded to prophage genomes and prophage remnants or green if they encoded probable ecologically important genes (7); the position of the recombined segments between OS195 and OS185 that contained only core genes (8). Note that the latter segments do not show any spatial bias in the genome, are not typically associated with the mobile genes in the genome and represent a substantial fraction of the core genome. Genomic rearrangements are frequently associated with mobile and rRNA genes (note the spatial correspondence of the boundaries of the rearrangements with the position of the latter genes).

OS185 against their orthologs in OS155 or OS223 was consistently <100%, and typically in the 95–98% range (for a graphical representation, see Figure 2b; all data are available in Supplementary Table S1). In addition, the majority of the recombined core segments between OS185 and OS195 were randomly distributed in the genome (Figure 3, innermost circle), did not show any strong biases in terms of the function of the genes they contained when compared against the rest of the genome (Supplementary Figure S3) and were 0.5 to ~10 Kbp long (average ~1.5 Kbp; Figure 4). Genes identified as recombined based on such simple sequence comparisons were further validated by GARD, an advanced algorithm for homologous recombination detection (Kosakovsky Pond *et al.*, 2006b). In general, there was a high agreement between the

two methods (>80%) in identifying recently recombined fragments (Supplementary Figure S4). About 10-fold more recombined core genes were observed between strains OS195 and OS185 ($n=308$) than between OS195 and OS223 ($n=48$) or OS195 and OS155 ($n=28$), which is consistent with higher genetic flow between OS195 and OS185 compared with the other genome pairs. The majority of the non-core genes shared between OS195 and OS185 showed similar patterns to those described above for core genes, suggesting that they were also brought in the genome through a similar mechanism as the core genes. These patterns are best explained by invoking an unconstrained mechanism for genetic exchange among the *S. baltica* genomes such as transformation or conjugation and homologous recombination as the process through which the

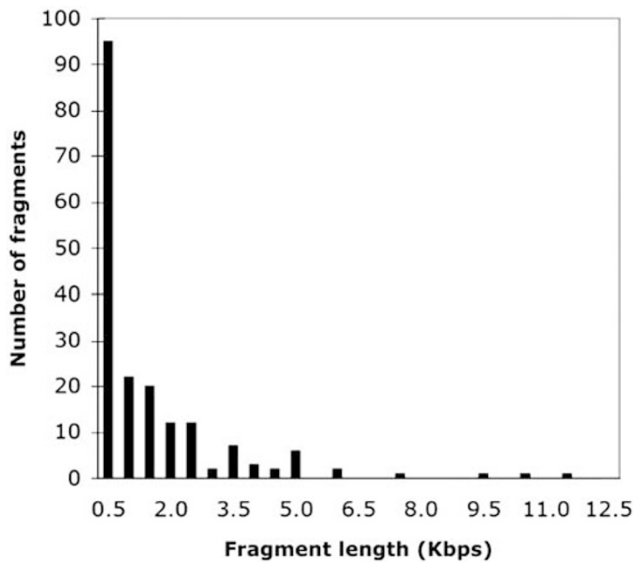


Figure 4 Length distribution of the recently recombined fragments between OS185 and OS195. All genetic exchange events between OS195 and OS185 similar to the two events shown in Figure 2b were identified based on visual inspection of the whole-genome alignments (as shown in Figure 2b and described in the Materials and methods section). The graph shows the length distribution of these recombined fragments.

exchanged DNA was incorporated into the genome. Although the exact mechanism for genetic exchange remains to be elucidated, the genome of *S. baltica* encodes several genes with strong amino acid similarity to known conjugative DNA transfer genes and a complete *recA*-dependent homologous recombination protein complex (Supplementary Table S1).

Assessing historical, as opposed to recent (Figure 2b), recombination among the *S. baltica* genomes was severely impeded by the very high nucleotide relatedness of the genomes, multiple (old) recombination events on the same segment of the genome, and the process of amelioration of the newly introduced DNA sequence into the recipient cell (Lawrence and Ochman, 1997). Accordingly, we report here on easily detectable, recent recombination events only.

Clonal or sexual speciation?

Even though precise dating of the genetic exchange events cannot be made because of lack of understanding of important population parameters such as the *in situ* generation time (Fraser *et al.*, 2007), a relative dating was attempted based on the predicted number of generations (*g*). We quantified *g* by dividing the average *K_s* value (synonymous substitutions per synonymous site) of all core genes with no obvious signal of recent recombination by the mutation rate of bacterial genomes (5.4×10^{-10} substitution/site/generation; Drake *et al.*, 1998), as suggested earlier (Wilson *et al.*, 2009; Jarvik *et al.*, 2010). (Synonymous substitutions are thought to be neutral and thus, reflect the intrinsic mutational

rate). The distribution of the *K_s* values of the core genes approximated the normal distribution and was very similar among all pairs of *S. baltica* genomes (six pairs in total; see Supplementary Figure S5A for OS195 vs OS185, Supplementary Figure S6 for all pairs). The average *K_s* was ~ 0.0898 , providing for a divergence time since the last common ancestor of all genomes that corresponded to 1.66×10^8 generations ($\pm 1.03 \times 10^7$ generations), with 95% confidence. By the same token, and using the average *K_s* of all recently recombined core genes between OS195 and OS185 (*K_s* = 0.0015), that is, the substitutions accumulated since the onset of recombination, we estimated that the recent recombination events identified here took place within the latest $\sim 2.77 \times 10^6$ generations. Thus, recombination between OS195 and OS185 occurred within the latest $\sim 2\%$ of the total divergence time since the last common ancestor of the *S. baltica* strains (Supplementary Figure S5B). We also used the codon usage bias of each gene, essentially as described earlier (Retchless and Lawrence, 2007), to normalize the *K_s* values (and derived divergence time estimates) for the different mutational rates of the genes because of the varied selection pressures acting on each gene. The normalized *K_s* values provided for similar results to those obtained with non-normalized *K_s* values (data not shown).

Using a simple strategy based on the *K_s* values, we also attempted to quantify the relative importance of recombination to mutation. For the time that recombination had been taking place between OS195 and OS185, we assumed that the synonymous substitutions brought in the genome by mutation equal the total length of all core genes (3.5 Mb) multiplied by the number of substitutions observed during this time (that is the *K_s* of recombined genes, which equaled 0.0015). During the same time, recombination purged a total number of synonymous substitutions that equaled the average number of substitutions between two genomes before the onset of recombination (that is, *K_s* of non-recombined genes—*K_s* of recombined genes; or $0.0898 - 0.0015 = 0.0883$) multiplied by the total length of the recombined core genes (0.20 Mb for OS195 vs OS185). Accordingly, the recombination (ρ) to mutation (θ) ratio was $\sim 3.4:1$ for OS195 and OS185, indicating sexual speciation (Fraser *et al.*, 2007). In contrast, and using the same methods and standards, the recombination to mutation ratio for the OS195 vs OS155 and OS195 vs OS223 pairs was 1:5 and 3:5, suggesting clonal divergence for these genome pairs.

Are the exchanged genes neutral or ecologically important?

DNA–DNA microarray experiments using a *S. baltica* pangenome oligoarray revealed that all OS195-like ($n = 10$) and OS185-like ($n = 3$) strains in our

collection examined had consistently greater hybridization signal for probes that corresponded to recombined vs non-recombined core genes (Supplementary Figure S7B). In addition, half of these strains, including OS195 and OS185, were isolated from the Gotland Deep sampling station in 1986 and the remaining half in 1987, whereas the *S. baltica* population was estimated to be about 1000 cells per ml of seawater in both sampling years based on most probable number estimates used with several liquid media (Ziemke *et al.*, 1997). Therefore, the genetic exchange patterns revealed by the sequenced genomes apply to a large collection of strains and were persistent over a time (1986–1987) in the natural *S. baltica* population.

Our data collectively reveal that the OS195 and OS185 lineages have exchanged recently >20% of their genome (core plus variable genes). The factors that have fostered the recent and extensive genetic exchange between OS195 and OS185 lineages are not fully understood but several lines of evidence seem to indicate that at least some of the exchanged genes are ecologically important as opposed to neutral. For instance, the strains of the OS185 lineage and particularly those of the OS195 lineage were isolated from depths (Supplementary Figure S7A) that were characterized by oxygen depletion and presence of alternative electron acceptors such as nitrate, manganese oxides and sulfur compounds (Figure 1a). To take advantage of the available electron acceptors, the strains possessed in their genomic islands several complete operons that encoded for anaerobic respiratory complexes and associated transport and cytochrome proteins (Supplementary Figures S7C and S8; Supplementary Table S2). In fact, the genes shared only by OS195 and OS185 represented either prophage related (that is ephemeral) or genes related, almost exclusively, to anaerobic metabolism and transport (Figure 3, 7th circle). It also appeared that the isolated OS195 strain, which apparently had migrated (sink?) in deeper waters after the recombination event(s) between the OS195 and OS185 lineages, had presumably adapted further to the more anoxic environment of the deeper waters. For instance, its genome encoded additional genomic islands for anaerobic lifestyle, such as a dimethyl sulfoxide reductase containing island (Supplementary Figure S7C) and OS195-like strains were more abundant and consistently recovered from this depth in both sampling years (Supplementary Figure S7A).

Although the substrates of the anaerobic genes shared between OS185 and OS195 remain speculative, laboratory microarray analysis revealed that some of these genes were expressed in OS185 and OS195 strains in response to anaerobic growth with nitrate or thiosulfate, indicating that they may be functional. The level of induction of the anaerobic metabolism genes examined typically varied between OS185 and OS195. For instance, the *nrf*

operon, which was shared exclusively between OS185 and OS195 (Supplementary Figure S8) and encodes for genes putatively involved in the dissimilatory nitrate reduction to ammonia (Hussain *et al.*, 1994), was significantly induced by thio-sulfate in both strains but by nitrate only in OS195 (Supplementary Figure S9). These variations in the level of induction may be due to the artificial batch conditions used in the laboratory compared with the *in situ* conditions, the experimental noise of the microarray measurements, and/or the varied degrees of ecological/genomic adaptations, which may have altered metabolic and regulatory networks between the two strains.

Consistent with their ecological role, bioinformatics sequence (Supplementary Table S1) and DNA–DNA microarray (Supplementary Figure S7C) comparisons suggested that most of the anaerobic metabolism genes shared between OS195 and OS185 were absent from strains of the OS155 lineage, which originated from (more) oxic waters (90–120 vs 120–140 m for strains of the OS195 lineage). In addition, competition growth experiments suggested that OS155 was outcompeted by OS195 under anaerobic conditions; for example, OS195 growth rate and final optical density in anaerobic medium (ZoBell agar or with thiosulfate as electron acceptor) were twice as high as those for OS155. Some of the potentially ecologically important genes shared between OS195 and OS185 (but not OS155), but not all (for example, thiosulfate/nitrate respiration; see Supplementary Figure S7C), were also present in OS223 (isolated from 120 m depth), whereas the number of genetic exchange events between OS195 and OS223 was higher compared with OS195 and OS155 (48 vs 28, respectively) but not as high as between OS195 and OS185 (308 events). These findings might indicate that although OS223 was isolated from the same depth as OS185, it might have occupied a slightly different ecological niche in the water column relative to OS185 or OS195, for example, being associated with sinking particles as opposed to being planktonic (or *vice versa*) or being transient or allochthonous at the 120–140 m depth (see also Discussion below). In agreement with the latter hypothesis, only one other OS223-like strain was recovered in our 1986 or 1987 isolation efforts.

Regardless of what the exact ecological niche of the strains or the environmental stimuli that the genes respond to may be, our findings collectively indicate that more anaerobic metabolism genes had been exchanged between strains from more similar (deeper) waters and these genes were apparently important for the successful adaptation of the strains in the deeper, more anoxic, waters. They also reveal that genomic adaptation of the *S. baltica* strains to their immediate environmental conditions, mediated by HGT, may be very fast and lead to sexual divergence (speciation).

Discussion

To the best of our knowledge, such rapid, extensive and genome-wide adaptation in immediate response to environmental settings, mediated by directed (as opposed to promiscuous) genetic exchange, as the one seen in the OS195 and OS185 or OS223 genomes, has never been observed earlier (Figure 2a). Thus, our findings advance understanding of the speed and mode of bacterial adaptation and underscore the important relationships between ecological setting, biotic interactions and genetic mechanisms that together shape and sustain microbial population structure. Extensive genetic exchange between co-occurring strains has been previously implied by metagenomic studies of natural populations (Tyson *et al.*, 2004; Konstantinidis and DeLong, 2008), but the fragmented nature of these data sets did not allow robust estimations of the magnitude of the genetic exchange at the whole-genome level or assessment of its ecological consequences (Eppley *et al.*, 2007; Konstantinidis and DeLong, 2008). Recent studies of isolated strains have also reported elevated levels of genetic exchange between pathogenic bacteria such as between distinct *Campylobacter* species (Sheppard *et al.*, 2008) or within *Vibrio cholerae* (Chun *et al.*, 2009). However, the genes exchanged in these cases are typically limited to a few environmentally selected functions and show strong biases in terms of spatial location in the genome (Caro-Quintero *et al.*, 2009). Accordingly and in contrast with *S. baltica*, genetic exchange is unlikely to lead to sexual speciation and population cohesion in such cases.

The *S. baltica* genomes reveal that genetic exchange, mediated by homologous recombination, could constitute an important mechanism for population cohesion among spatially co-occurring prokaryotes, similar to the role of sexual reproduction in higher eukaryotes. Therefore, our results provide the experimental evidence in support of recent computer simulation studies that suggested that recombination-driven sexual speciation is possible in bacteria (Fraser *et al.*, 2007). Despite the extensive recombination observed, the *S. baltica* genomes show no evidence in support of the recently proposed fragmented speciation model for bacteria (Retchless and Lawrence, 2007). For instance, the predicted signature of this model, that is, ecological genomic islands are surrounded by increased levels of nucleotide divergence between ecologically distinct (for example, OS195 vs OS155) but not between ecologically coherent (for example OS195 vs OS185) populations, was not observed (Supplementary Figure S10). The signature was also not observed in comparisons between selected *S. baltica* strains and other closely related (that is, sharing 80–88% ANI to *S. baltica*) but ecologically distinct sequenced *Shewanella* genomes of *Shewanella* sp. MR-4 and MR-7 from the Black Sea, *Shewanella* sp.

ANA-3 and *Shewanella oneidensis* MR-1 from freshwater ecosystems in the United States (Konstantinidis *et al.*, 2009). These results may be due to the fact that the recombined fragments are too small (Figure 4) for recombination to be affected (reduced) by the presence of genomic islands (which would act as barriers to recombination because the sequence is not conserved) among ecologically distinct organisms. Alternatively, the genetic exchange between the incipient ecological distinct species may not be maintained for long enough evolutionary time as hypothesized earlier (Retchless and Lawrence, 2007) for recombination to create the signature of the model in the *S. baltica* case.

To what extent the patterns of genetic exchange observed between OS195 and OS185 (Figure 2) and their sister strains (Supplementary Figure S7) apply to other natural sub-populations of *S. baltica* in the Baltic Sea and what accounts for the reduced genetic flow between OS185 and OS223 (same isolation depth) compared with OS195 (different depth) remain currently unknown. To address these issues, *in situ* genomic studies (for example, metagenomics) and sampling of the natural populations over time will be required. However, the OS195 and OS185 example does raise the possibility that bacterial adaptation through genetic exchange may be much more rapid and extensive than previously anticipated and thus, it has broader implications for understanding bacterial evolution and adaptation. Our independent analyses have also ruled out the possibility that the results reported here for OS195 and OS185 are attributable to manmade mixing of the genomic DNA submitted to sequencing or the derived sequences. For instance, if the results were attributable to DNA mixing, we would not have observed a significantly greater hybridization signal with the recombined vs the non-recombined genes during DNA–DNA microarray experiments (Supplementary Figure S7). It also appeared that the genomes of OS155 and OS223 had numerous and extensive genomic rearrangements (transposition and inversions) compared with those of OS195 and OS185, whereas OS185 and OS195 genomes were syntenic in almost their entire length (Figure 3, outer cycles). Whether or not these rearrangements, which could act as barriers to recombination because the sequence is not syntenic, are responsible for the reduced genetic flow between OS223 or OS155 and OS195 relative to OS185 and OS195 is not clear, but does represent an intriguing hypothesis that warrants further investigations.

In summary, it seems as if the genome of *S. baltica* adapts through continuous internal genome-wide genetic exchange and rearrangement events (Figure 3), in a highly dynamic (electron donors as well as electron acceptors), nutrient-rich pelagic environment. This differs fundamentally from what was observed previously in other important marine bacteria such as the *Pelagibacter ubique* (Giovannoni *et al.*, 2005) and *Prochlorococcus marinus* (Coleman

et al., 2006), which have streamlined genomes, developed over eons in rather constant, nutrient-poor environments. The latter organisms represent the ultimate marine *k*-strategist, whereas *S. baltica* is very close to the ultimate *r*-strategist. The patterns observed in *S. baltica* may be broadly applicable to other bacteria that experience frequent environmental fluctuations in the marine environment and elsewhere. Therefore, our findings expand understanding of the rate and mode of bacterial adaptation and underscore the important relationships between ecological setting, biotic interactions and genetic mechanisms that together shape and sustain microbial population structure.

Acknowledgements

We thank Professors James Tiedje and Frank Loeffler for helpful suggestions regarding the manuscript and the Shewanella Federation for supporting work on Shewanella genomics. Contributions of the Joint Genome Institute for the genome sequences used in this study are also acknowledged. This work is supported by the US Department of Energy under Contract No. DE-FG02-07ER64389.

References

- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W *et al.* (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl Acids Res* **25**: 3389–3402.
- Backer H, Leppanen JM, Brusendorff AC, Forsius K, Stankiewicz M, Mehtonen J *et al.* (2010). HELCOM Baltic Sea action plan—a regional programme of measures for the marine environment based on the Ecosystem approach. *Mar Pollut Bull* **60**: 642–649.
- Beiko RG, Harlow TJ, Ragan MA. (2005). Highways of gene sharing in prokaryotes. *Proc Natl Acad Sci USA* **102**: 14332–14337.
- Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. (2009). GenBank. *Nucleic Acids Res* **37**: D26–D31.
- Brettar I, Moore ER, Hofle MG. (2001). Phylogeny and abundance of novel denitrifying bacteria isolated from the water column of the central Baltic sea. *Microb Ecol* **42**: 295–305.
- Bruen TC, Philippe H, Bryant D. (2006). A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**: 2665–2681.
- Caro-Quintero A, Rodriguez-Castano GP, Konstantinidis KT. (2009). Genomic insights into the convergence and pathogenicity factors of *Campylobacter jejuni* and *Campylobacter coli* species. *J Bacteriol* **191**: 5824–5831.
- Chun J, Grim CJ, Hasan NA, Lee JH, Choi SY, Haley BJ *et al.* (2009). Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae*. *Proc Natl Acad Sci USA* **106**: 15442–15447.
- Coleman ML, Sullivan MB, Martiny AC, Steglich C, Barry K, Delong EF *et al.* (2006). Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science* **311**: 1768–1770.
- Drake JW, Charlesworth B, Charlesworth D, Crow JF. (1998). Rates of spontaneous mutation. *Genetics* **148**: 1667–1686.
- Eppley JM, Tyson GW, Getz WM, Banfield JF. (2007). Genetic exchange across a species boundary in the archaeal genus *ferroplasma*. *Genetics* **177**: 407–416.
- Fraser C, Hanage WP, Spratt BG. (2007). Recombination and the nature of bacterial speciation. *Science* **315**: 476–480.
- Fredrickson JK, Romine MF, Beliaev AS, Auchtung JM, Driscoll ME, Gardner TS *et al.* (2008). Towards environmental systems biology of *Shewanella*. *Nat Rev Microbiol* **6**: 592–603.
- Gevers D, Cohan FM, Lawrence JG, Spratt BG, Coenye T, Feil EJ *et al.* (2005). Opinion: re-evaluating prokaryotic species. *Nat Rev Microbiol* **3**: 733–739.
- Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D *et al.* (2005). Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309**: 1242–1245.
- Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. (2007). DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* **57**: 81–91.
- Handelsman J, Tiedje J, Alvarez-Cohen L, Ashburner M, Cann I, Delong E *et al.* (2007). *The New Science of Metagenomics: Revealing the Secrets of Our Microbial Planet*. The National Academies Press: Washington, DC.
- Hussain H, Grove J, Griffiths L, Busby S, Cole J. (1994). A seven-gene operon essential for formate-dependent nitrite reduction to ammonia by enteric bacteria. *Mol Microbiol* **12**: 153–163.
- Jarvik T, Smillie C, Groisman EA, Ochman H. (2010). Short-term signatures of evolutionary change in the *Salmonella enterica* serovar typhimurium 14028 genome. *J Bacteriol* **192**: 560–567.
- Konstantinidis KT, DeLong EF. (2008). Genomic patterns of recombination, clonal divergence and environment in marine microbial populations. *ISME J* **2**: 1052–1065.
- Konstantinidis KT, Ramette A, Tiedje JM. (2006a). The bacterial species definition in the genomic era. *Philos Trans R Soc Lond B Biol Sci* **361**: 1929–1940.
- Konstantinidis KT, Ramette A, Tiedje JM. (2006b). Toward a more robust assessment of intraspecies diversity, using fewer genetic markers. *Appl Environ Microbiol* **72**: 7286–7293.
- Konstantinidis KT, Serres MH, Romine MF, Rodrigues JL, Auchtung J, McCue LA *et al.* (2009). Comparative systems biology across an evolutionary gradient within the *Shewanella* genus. *Proc Natl Acad Sci USA* **106**: 15909–15914.
- Konstantinidis KT, Tiedje JM. (2005). Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci USA* **102**: 2567–2572.
- Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. (2006a). Automated phylogenetic detection of recombination using a genetic algorithm. *Mol Biol Evol* **23**: 1891–1901.
- Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. (2006b). GARD: a genetic algorithm for recombination detection. *Bioinformatics* **22**: 3096–3098.
- Lang AS, Beatty JT. (2007). Importance of widespread gene transfer agent genes in alpha-proteobacteria. *Trends Microbiol* **15**: 54–62.

- Lawrence JG, Ochman H. (1997). Amelioration of bacterial genomes: rates of change and exchange. *J Mol Evol* **44**: 383–397.
- Lawrence JG, Ochman H. (2002). Reconciling the many faces of lateral gene transfer. *Trends Microbiol* **10**: 1–4.
- Myers CR, Nealson KH. (1988). Bacterial manganese reduction and growth with manganese oxide as the sole electron acceptor. *Science* **240**: 1319–1321.
- Neumann T. (2006). The fate of river-borne nitrogen in the Baltic Sea: an example for the River Oder. *Estuar Coast Shelf Sci* **73**: 1–7.
- Retchless AC, Lawrence JG. (2007). Temporal fragmentation of speciation in bacteria. *Science* **317**: 1093–1096.
- Sheppard SK, McCarthy ND, Falush D, Maiden MC. (2008). Convergence of *Campylobacter* species: implications for bacterial evolution. *Science* **320**: 237–239.
- Tettelin H, Maignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL *et al.* (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial ‘pan-genome’. *Proc Natl Acad Sci USA* **102**: 13950–13955.
- Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM *et al.* (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**: 37–43.
- Welch RA, Burland V, Plunkett III G, Redford P, Roesch P, Rasko D *et al.* (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci USA* **99**: 17020–17024.
- Wilson DJ, Gabriel E, Leatherbarrow AJ, Cheesbrough J, Gee S, Bolton E *et al.* (2009). Rapid evolution and the importance of recombination to the gastroenteric pathogen *Campylobacter jejuni*. *Mol Biol Evol* **26**: 385–397.
- Yang Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**: 1586–1591.
- Zhaxybayeva O, Gogarten JP, Charlebois RL, Doolittle WF, Papke RT. (2006). Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events. *Genome Res* **16**: 1099–1108.
- Ziemke F, Brettar I, Hofle MG. (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.

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