

## ORIGINAL ARTICLE

# Comparative genomics of two ecotypes of the marine planktonic copiotroph *Alteromonas macleodii* suggests alternative lifestyles associated with different kinds of particulate organic matter

Elena Ivars-Martinez<sup>1</sup>, Ana-Belen Martin-Cuadrado<sup>1</sup>, Giuseppe D'Auria<sup>1</sup>, Alex Mira<sup>1</sup>, Steve Ferriera<sup>2</sup>, Justin Johnson<sup>2</sup>, Robert Friedman<sup>2</sup> and Francisco Rodriguez-Valera<sup>1</sup>

<sup>1</sup>Evolutionary Genomics Group, Departamento Producción Vegetal y Microbiología, Universidad Miguel Hernández, San Juan de Alicante, Alicante, Spain and <sup>2</sup>J Craig Venter Institute, Rockville, MD, USA

***Alteromonas macleodii* is a common marine heterotrophic  $\gamma$ -proteobacterium. Isolates from this microbe cluster by molecular analysis into two major genotypic groups or ecotypes, one found in temperate latitudes in the upper water column and another that is for the most part found in the deep water column of the Mediterranean. Here, we describe the genome of one strain of the 'deep ecotype' (AltDE) isolated from 1000 m in the Eastern Mediterranean and compare this genome with that of the type strain ATCC 27126, a representative of the global 'surface' ecotype. The genomes are substantially different with DNA sequence similarity values that are borderline for microbes belonging to the same species, and a large differential gene content, mainly found in islands larger than 20 kbp, that also recruit poorly to the Global Ocean Sampling project (GOS). These genomic differences indicate that AltDE is probably better suited to microaerophilic conditions and for the degradation of recalcitrant compounds such as urea. These, together with other features, and the distribution of this genotypic group, indicate that this microbe colonizes relatively large particles that sink rapidly to meso and bathypelagic depths. The genome of ATCC 27126 on the other hand has more potential for regulation (two component systems) and degrades more sugars and amino acids, which is consistent with a more transient particle attachment, as would be expected for lineages specialized in colonizing smaller particulate organic matter with much slower sinking rates. The genomic data are also consistent with a picture of incipient speciation driven by niche specialization.**

*The ISME Journal* (2008) 2, 1194–1212; doi:10.1038/ismej.2008.74; published online 31 July 2008

**Subject Category:** integrated genomics and post-genomics approaches in microbial ecology

**Keywords:** biogeography; deep ecotype; marine bacteria; mediterranean microbiota; metagenomics; sympatric speciation; particulate organic matter (POM)

## Introduction

Prokaryotic species comprise strains (lineages) with very different gene content. The most extensively studied cases are those of pathogenic bacteria in which the presence of specific gene clusters or pathogenicity islands are often linked to the clinical

outcome of their infection (Tettelin *et al.*, 2005; Dempsey *et al.*, 2006; Hochhut *et al.*, 2006; Petrosino *et al.*, 2006; Willenbrock *et al.*, 2006). However, free-living bacteria have similar genomic strategies when exploiting different habitats or niches and comparative genomics is a powerful tool to understand the ecological specialization of the different lineages (Ting *et al.*, 2002; Kettler *et al.*, 2007). The diversity of species or ecotypes in marine waters is intriguing due to their extraordinary number under a seemingly homogenous environment (the so-called plankton paradox). This higher-than-expected diversity was attributed to an underestimate of the number of microenvironments in the sea

Correspondence: F Rodriguez-Valera, Evolutionary Genomics Group, Departamento Producción Vegetal y Microbiología, Universidad Miguel Hernández, San Juan de Alicante, Alicante 3550, Spain.

E-mail: frvalera@umh.es

Received 7 May 2008; revised 27 June 2008; accepted 4 July 2008; published online 31 July 2008

(Richerson *et al.*, 1970). We believe that genomic data from different ecotypes may shed light on the specific adaptations of lineages to these microenvironments and how these adaptations can drive speciation events.

*Alteromonas macleodii* was shown by molecular methods to dominate heterotrophic blooms in mesocosms (Pukall *et al.*, 1999; Schafer *et al.*, 2000). These data and the phenotypic properties characterize this species as a copiotroph that behaves as an *r*-strategist in the relatively nutrient depleted regions of the world's oceans. Concurrently, other authors demonstrated that sequences closely related to the only representative isolate at the time ATCC 27126 were also quite prevalent in open Mediterranean waters. The abundance seemed to increase in the particulate fraction and down to a depth of 400 m in the Western Mediterranean (Acinas *et al.*, 1999). Hybridization experiments confirmed these data (Glockner *et al.*, 1999; Eilers *et al.*, 2000; Garcia-Martinez *et al.*, 2002). *A. macleodii* is a common isolate by standard marine bacteria isolation protocols, and isolates belonging to this species from different studies have accumulated in individual researchers' collections (Sass *et al.*, 2001; Lopez-Lopez *et al.*, 2005). By collecting isolates from different parts of the world and, to a lesser degree, from different depths, it was shown that isolates from the deep Mediterranean were consistently variable at the molecular level. This discovery led to the proposal of the existence of a deep Mediterranean ecotype or deep ecotype (Lopez-Lopez *et al.*, 2005). The Mediterranean is different from other oceanic water masses in that it has a relatively warm water column (*ca* 13 °C) regardless of depth, which can reach 5000 m in the deepest basins. It has been proven by several independent methodologies (DeLong *et al.*, 2006; Zaballos *et al.*, 2006; Martin-Cuadrado *et al.*, 2007) that *A. macleodii* is present worldwide in photic zone depths and is present at all depths in the Mediterranean but not in other oceanic locations. We attribute this to low water temperature of the deep oceans that prevent the successful competition of an organism that is specialized in exploiting the concentrated nutrient havens of the particulate fraction (Grossart *et al.*, 2007). Other organisms that do not rely so heavily on the rapid growth would most likely outcompete *A. macleodii* in deeper, cold oceanic regions. Corroborating this, molecular evidence of significant *A. macleodii* contribution to the biomass was never found below 10 °C (Garcia-Martinez *et al.*, 2002).

It is only recently that the advances in genomics and metagenomics have been applied to clarify the specific adaptations of the deep ocean microbiota. In general, cultured deep ocean microbes seem to have higher representation of transposable and phage-related elements and larger intergenic spacers. This has been interpreted as reflecting an opportunistic way of life (*r*-strategists) (Lauro and Bartlett, 2008). Regarding types of genes, as would be expected,

deep ocean dwellers have small representation of light-related genes such as the photoreactivation genes. The absence of photolyases has been considered diagnostic of autochthonous deep ocean dwellers. Conversely, some genes are better represented in the deep communities, specifically those involved in membrane unsaturation, which are important for both cold and pressure adaptation (Lauro *et al.*, 2008). Sugar transporters are more common in shallow waters, whereas tripartite ATP-independent periplasmic (TRAP) transporters were more abundant in deep waters (Simonato *et al.*, 2006).

Here, we report the complete genome of *A. macleodii* 'deep ecotype' (AltDE) (Lopez-Lopez *et al.*, 2005) obtained from a depth of 1000 m in the South Adriatic Sea basin. This strain was sequenced as part of the 'Gordon and Betty Moore Foundation' Marine Microbial Initiative and has been fully assembled here. For comparative purposes, we have sequenced the genome of the first isolate of *A. macleodii* ATCC 27126 described by Baumann *et al.* (1972), obtained from the surface waters near Oahu (Hawaii). This genome has been pyrosequenced (at  $\times 12$  coverage) and has not been fully assembled. AltDE and its relatives have been found consistently at depths down to 3500 m in the Mediterranean. The cluster represented by ATCC 27126 was never found in deep water masses but is widespread on the surface of the world's oceans. Therefore, the specific adaptations detected in AltDE at the genomic level can be useful to understand the requirements imposed on the microbiota by the bathypelagic habitat, minus the low temperature normally associated with this habitat. Both organisms belong to the same species (for example, 16S rRNA similarity is close to 99%) but the sequences of housekeeping genes (MLSA) and some phenotypic properties (Lopez-Lopez *et al.*, 2005; Ivars-Martinez *et al.*, 2008) clearly indicate that both represent divergent ecotypes or strain clusters. Therefore, this is an excellent opportunity to study the genomic changes that accompany the initial stages of bacterial speciation within free-living bacteria. In modern evolutionary theory, genetic isolation of populations was considered the basic requirement for generating new species (Dobzhansky, 1970) and geographic barriers are the most extended cause for the initial splitting of populations. This allopatric speciation model was developed mainly with sexual eukaryotes in mind and needs to be reassessed for prokaryotic species (Franklin, 2007). In free-living bacteria, physical or geographic barriers are not well delineated and clear instances of genetic recombination have been found between isolates from very distant oceanic locations (Martin-Cuadrado *et al.*, 2008). In addition, microbial speciation models indicate that a decrease in recombination frequency is insufficient to generate species in nature (Hanage *et al.*, 2006) and therefore other mechanisms should be involved in separating

bacterial populations under sympatry. Experimental evolution studies with different strains indicate that differential resource usage may underlie population divergence, even under direct contact (Spencer *et al.*, 2007). Based on the genomic data from the two *A. macleodii* ecotypes, we intend to study whether bacterial divergence can also be driven by niche specialization and not just by a lack of genetic recombination.

## Materials and methods

### Sample collection

*A. macleodii* AltDE was isolated at 1000 m deep in the Adriatic Sea (41°36'N—17°22'E) from a sample collected during the R/V 'Urania' cruise to the Gulf of Manfredonia in May 2003. Biochemical parameters of the sampling site were: temperature: 12.5 °C; salinity: 38.6‰ and dissolved oxygen: 5.2 ml l<sup>-1</sup>. The strain is deposited in the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ) collection with reference DSMZ 17117. *A. macleodii* ATCC 27126 was isolated in 1972 in the coast of Oahu (Hawaii) from surface waters (Baumann *et al.*, 1972).

### Sample sequencing and assembly

*A. macleodii* deep ecotype strain was sequenced by the Gordon and Betty Moore Foundation as part of a large marine microbes sequencing effort. Two genomic libraries with insert sizes of 4 and 40 kbp were made as described in the methods in Goldberg *et al.* (2006). The prepared plasmid and fosmid clones were sequenced from both ends to provide paired-end reads at the J Craig Venter Institute Joint Technology Center on ABI 3730XL DNA sequencers (Applied Biosystems, Foster City, CA, USA). Successful reads for the organism were used as input for Celera assembler (Myers *et al.*, 2000). Whole-genome random shotgun sequencing (WGS) produced 111 contigs averaging 39 759 bp (which range from 1065 bp to 268 182 bp) for a total of approximately 4.4 Mbp of microbial DNA sequenced. WGS sequence produced by the assembler was then annotated automatically using Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP) at NCBI\*. Data are released to the Gordon and Betty Moore Foundation Marine Microbial Genome Sequencing Project website\*\*. Artemis Comparison Tool ACTv.6 (Carver *et al.*, 2005) was used to study the possible synteny of the 111 contigs with other genomes. Oligonucleotides at the extremes of the contigs of AltDE were designed and PCR was carried out to close the gaps. Strain ATCC 27126 was pyrosequenced (454 Life Sciences, Branford, CT, USA) and the final number of genomic fragments was 716 (with sizes ranging from 89 bp to 80 530 bp, with an average of 6435 bp). These sequences were not assembled in one single contig. The length of the nonoverlapping contigs was 4 607 844 bp. \*<http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html> (<http://www.jcvi.org/cms/research/projects/microgenome/overview/>). \*\*<http://www.moore.org/>.

### Bioinformatic analysis

**Gene prediction and annotation.** After assembling, protein-coding genes were predicted using GLIMMER (Delcher *et al.*, 1999) and SEEDs (Overbeek *et al.*, 2005), and were further manually curated. Spacers were subsequently searched against the nonredundant database using BLAST (Altschul *et al.*, 1990) to ensure that no open reading frame (ORF) had been missed. Identified ORFs were compared with the previous automatic annotation and also to known proteins in the nonredundant database using BLASTX. All hits with an *E*-value greater than 10<sup>-5</sup> were considered nonsignificant. tRNAs were identified using tRNAscan-SE (Lowe and Eddy, 1997).

**Functional classification.** All predicted ORFs were also searched for similarity using RPSBLAST to predict COG (clusters of orthologous groups) assignments (cutoff of *E*-value 10<sup>-7</sup>) (Tatusov *et al.*, 2001). The KEGG database was used to analyze the metabolic pathways (Kanehisa *et al.*, 2004).

**Genome analysis.** From the European Molecular Biology Open Software Suite (EMBOSS) package (Rice *et al.*, 2000), GEECEE was used to calculate GC content whereas CUSP and CODCMP were used for codon usage analysis of the genomic islands (GIs) shown in Table 3. The GC skew (Figure 2) and the tetranucleotide frequency (Supplementary Figure S7) were calculated using the Oligoweb interface <http://insilico.ehu.es/oligoweb/>. For comparative analyses, reciprocal BLASTN and TBLASTX searches between the genomes were carried out, leading to the identification of regions of similarity, insertions and rearrangements (*E*-value cutoff 10<sup>-3</sup>). Artemis (Rutherford *et al.*, 2000) and Artemis Comparison Tool ACTv.6 (Carver *et al.*, 2005) were used to allow interactive visualization of genomic fragment comparisons. For whole genome visualization and taxonomic comparison of each ORF in Supplementary Figure S1, the software tool GENOMEVIZ 1.1 was used (Ghai *et al.*, 2004). ANI (average nucleotide identity) was calculated as defined by Konstantinidis and Tiedje (2005). Reciprocal BLASTCLUST was used to calculate the orthologous proteins between *A. macleodii* ATCC 27216 and AltDE strains using a minimum cutoff of 50% similarity and 70% of the length of the query protein for Figure 2 (BLASTCLUST parameters: -S 50 -L 0.7 -b T). Bioedit software was used to manipulate the sequences (Hall, 1999). Sequences were aligned using MUSCLE 3.6 (Edgar, 2004) and ClustalW (Thompson *et al.*, 1994) and edited manually as necessary. Phylogenetic analysis of proteins was performed using the MEGA4 phylogenetic tool software package (Tamura *et al.*, 2007).

**Global Ocean Sampling (GOS) recruitment.** Global Ocean Sampling metagenome sequences were aligned against reference genome of AltDE using the MUMmer program version 3.19 (Kurtz *et al.*,

2004). Specifically, we used the 'promer' program with the maxmatch option and default parameters to calculate the alignments. 'Show-coords' option was used to obtain the coordinates and generate the percent identity plot depicted in Figure 2d. For Figure 3, a BLASTN comparison (cutoff of 50% identity and 70% of the length of the query sequence) between a database formed by the two genomes and the complete GOS database were performed (contigs of ATCC 27216 were artificially concatenated). This permitted us to determine the presence of the two genomes in each environmental marine database. After normalization of the results by database size, we observed an unexpected strong bias to sample GS000b, recruiting 70.94% of the total sequences for strain ATCC 27216. However, collection GS000d, other Sargasso Sea sample, which was recovered in the same time and conditions, recruits only 2.47% of the total. These inconsistencies lead us to withdraw the Sargasso Sea collections from the analysis of Figure 3.

#### Growth curves and biochemical assays

Some phenotypic features inferred from the genome sequenced were assessed by growth experiments. Growth at different temperatures (5 °C, 10 °C, 15 °C, 20 °C, 25 °C, 30 °C and 40 °C) was measured using a microplate optical density reader Fluostar Optima from BMG LABTECH GmbH (Offenburg, Germany) at 595 nm and marine broth as basal medium. To investigate heavy metals resistance, cultures were grown in the presence of zinc acetate, mercury chloride and lead chloride at different concentrations (0.001 mM, 0.01 mM, 0.1 mM, 1, 5 mM and 10 mM). Culture growth was followed by optical density at 595 nm. Urease tests were carried out as described previously by Christian (1946) and *Proteus vulgaris* was used as a positive control. Nitrate respiration was performed as described by Skerman (1967).

#### Accession numbers

The sequences have been deposited in GenBank under the genomic accession numbers CP001103 for *A. macleodii* AltDE and ABQB00000000 for ATCC 27126.

## Results

#### General features of the genomes

The genome of AltDE was assembled into a single replicon of 4 412 285 bp. *A. macleodii* strain ATCC 27126 was pyrosequenced and automatically assembled into 716 contigs, with an estimated genome size was of 4 607 844 bp.

Both organism's genome appear to be quite similar in structure and size. Among the 4102 ORFs found in AltDE, 2696 had orthologs in ATCC 27126 (see Materials and methods). The average nucleotide

identity between them was 81.24%, which is low for members of the same species, but is to be expected when examining organisms from widely divergent ecotypes (Lopez-Lopez *et al.*, 2005; Ivars-Martinez *et al.*, 2008). Showing phylogenetic and ecological consistency, most of each organism's genes have their closest homologs in other aquatic  $\gamma$ -proteobacteria (78.05%) (Supplementary Figure S1). *Pseudalteromonas atlantica* T6c is the bacterium, which shares more genes with AltDE (1753). This is a common marine bacterium originally isolated from lesions on crabs with shell disease but is found both in the water column and in biofilms attached to surfaces (Pernthaler *et al.*, 2001; Costa-Ramos and Rowley, 2004). Basic features of both *A. macleodii* genomes and of *P. atlantica* T6c are displayed in Table 1. The coding density and intergenic spacers are very similar for the three genomes and, in any case, not larger in AltDE. Larger intergenic spacers have been described as a typical feature of cultured bathytypes (Lauro and Bartlett, 2008). As expected, the number of genes shared and average similarity are much lower for *P. atlantica* T6c. We did not detect any significant difference in amino-acid usage (data not shown) and could not find the extra loop associated to some bathytypes in the 16S rRNA (Lauro *et al.*, 2007) in any of the strains. A BLASTX analysis provided reliable function predictions for 74.53% of AltDE ORFs, while 25.47% of the predicted ORFs showed only weak or no similarities. Similar values were found for ATCC 27126, with percentages of 67.47 and 32.53. Specifically, AltDE had 1045 ORFs with hypothetical function, 198 of which were conserved hypotheticals. In ATCC 27126, 1446 hypothetical proteins were found, from which 82 are conserved in other genomes. We have determined the pools of genes shared among AltDE, ATCC 27126 and *P. atlantica* T6c. There are more than 1200 genes that are unique for each of the *A. macleodii* ecotypes, while ca 2696 were found in both. As ATCC 27126 is not assembled, synteny cannot be completely assessed (Supplementary Figure S2a). However, in *P. atlantica* T6c, the genome appeared largely syntenic to AltDE and for the large ATCC 27126 contigs, orthologous genes conserved the relative positions in nearly all cases (Supplementary Figure S2b).

#### Differential gene content and metabolic reconstruction

The differential gene content has been analyzed to infer the biological meaning and its relationship to the two putative lifestyles, that is, surface vs deep dwellers. The different genes found reflect remarkably different biological properties for both strains, although most are not obviously associated to living in different strata of the water column. Upon initial screening, the AltDE and ATCC 27126 genes were functionally classified according to COG category and the frequencies were compared in both genomes. 68.52 and 68.85% genes in AltDE and ATCC 27126,

**Table 1** General features of genomes

	<i>A. macleodii</i> AltDE	<i>A. macleodii</i> ATCC27126 <sup>a</sup>	<i>P. atlantica</i> T6c
Size (bp)	4 412 285	4 607 844	5 187 010
GC content	44.90	44.62	44.6
Contigs	1	716	1
Total ORFs	4102	4444	4281
Function assigned	3057	3319	3408
Hypotetical protein	1045	1588	873
Genes share with AltDE	—	2696	1753
Exclusive genes	1242	1588	2368
rRNA operons	5	5	5
tRNAs	40	48	68
ANI <sup>b</sup>	—	81.24	68.11
Coding density	(86.63%)	(87.29%)	(86.14%)
Average spacers size	143	181	167
Average ORFs size	931	1043	855

Abbreviations: ANI, average nucleotide identity; *A. macleodii*, *Alteromonas macleodii*; AltDE, *A. macleodii* 'deep ecotype'; *P. atlantica*, *Pseudoalteromonas atlantica*.

<sup>a</sup>Only automatic annotation.

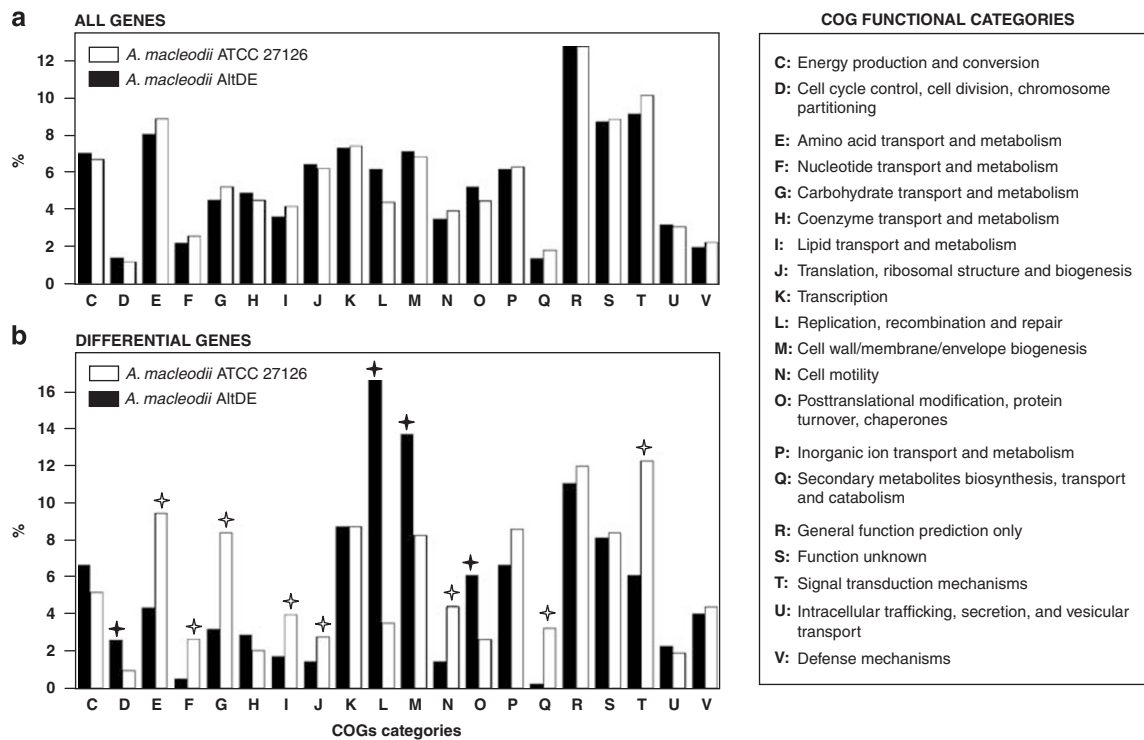
<sup>b</sup>ANI (Konstantinidis and Tiedje, 2005).

respectively were clearly affiliated into one COG category. Considering the total pools of genes in each strain, the differences seemed small (Figure 1a). However, analyzing only the genes that were unique to each strain, the differences appeared more conspicuous and indicated that they belong largely to certain categories (Figure 1b). The genes included in these categories were submitted to a more refined functional analysis (Table 2) and whenever possible, were experimentally confirmed by assessing the expected phenotype.

The most remarkable difference was the number of transposable elements with 65 belonging to 17 families in AltDE vs only three in two families in ATCC 27126. Only one unique insertion sequence (IS) element in ATCC 27126 was not found in AltDE, whereas 63 exist in AltDE without homologs in ATCC 27126. The presence of large numbers of IS elements is considered a typical characteristic associated with deep-sea bacteria (Vezi et al., 2005; DeLong et al., 2006; Martin-Cuadrado et al., 2007). Although the reason for this is unknown, it may reflect a different mode of gene propagation, likely related to the slower growth, lower productivity and lower effective population sizes of deep-sea microbial communities. Another interesting difference is the larger number of genes found in AltDE that could be involved in interaction with phages, particularly phage integrases and the clustered regularly interspaced short palindromic repeats (CRISPR) cluster (Sorek et al., 2008). Large numbers of phage-related genes have also been described as a typical feature of deep-sea cultured bacteria and is, again, poorly understood (Camparano et al., 2005; Vezi et al., 2005). Contrarily, DeLong et al. (2006) observed an enrichment of phage genes in the photic zone indicating a greater role for phage parasites in the more productive upper water column relative to deeper waters. Nonetheless, the ratio of phage genes found in the

deep Mediterranean at 3000 m (Martin-Cuadrado et al., 2007) was higher compared with the North Atlantic at 700 m and 4000 m. In our case, both characteristics might reflect a more strict association to particulate material and relatively high local population densities (see below).

*A. macleodii* deep ecotype contains more dioxygenases (Table 2) indicating more potential for degrading recalcitrant compounds. AltDE also contains more chaperons such as a Cochaperonin (GroE), a flagellin-specific chaperonine, a hydrogenase chaperone and a lipase chaperone. Chaperons could help the organism to grow at lower temperatures by assisting in protein folding and these type of genes are predicted to be highly expressed in deep-sea genomes (Xu and Ma, 2007). Another very conspicuous set of genes over-represented in AltDE were related to heavy metal resistance (see below, GI 9). In contrast, the numbers of sigma factors, transcriptional regulators and particularly, histidine kinases were much larger in the surface isolate ATCC 27126. This potentially signifies a higher adaptability, as should be expected in an environment subjected to wider environmental changes and gradients. However, most types of two component systems were shared by both strains, except the nitrogen sensor of GI12 (see below) and a turgor pressure sensor (KdpD) related to the osmoregulatory Kdp pathway that was only present in the ATCC 27126. The Kdp pathway is the first step in adaptation to changes in medium salinity by increasing potassium intracellular concentration. Its absence in AltDE might indicate further adaptation to a more homogeneous environment, in this case, salinity wise. Along the same lines, ATCC 27126 has many more unique TonB receptors, which could indicate that a wider range of substrates can be transported and utilized. Nevertheless, the number of ABC transporters was similar, yet many were unique for each isolate (Table 2). Owing to the



**Figure 1** Distribution of COG functional classes. Percentage of COGs predicted in AltDE and ATCC 27126 genomes. All genes (a) and genes found only in one of the genomes (b) are indicated. Asterisks indicate categories where a significant variation was found comparing the two genomes. AltDE, *Alteromonas macleodii* ‘deep ecotype’; COG, clusters of orthologous groups.

**Table 2** Gene categories found in different frequencies in the two *A. macleodii* strains

Genes categories	ATCC 27126	AltDE
Phage integrase	12 (9)	32 (23)
Transposase and IS elements	3 (1)	65 (63)
Chaperones	9 (1)	13 (5)
Sigma factors	10 (0)	8 (0)
Histidine kinases	53 (18)	40 (5)
Dioxygenases	8 (0)	11 (3)
Heavy metals resistance	24 (1)	36 (18)
TonB receptors	82 (31)	52 (7)
Toxin–antoxin	7 (1)	12 (5)
Acr systems	50 (11)	45 (11)
ABC transporters	79 (19)	82 (22)

Abbreviations: *A. macleodii*, *Alteromonas macleodii*; AltDE, *A. macleodii* ‘deep ecotype’.

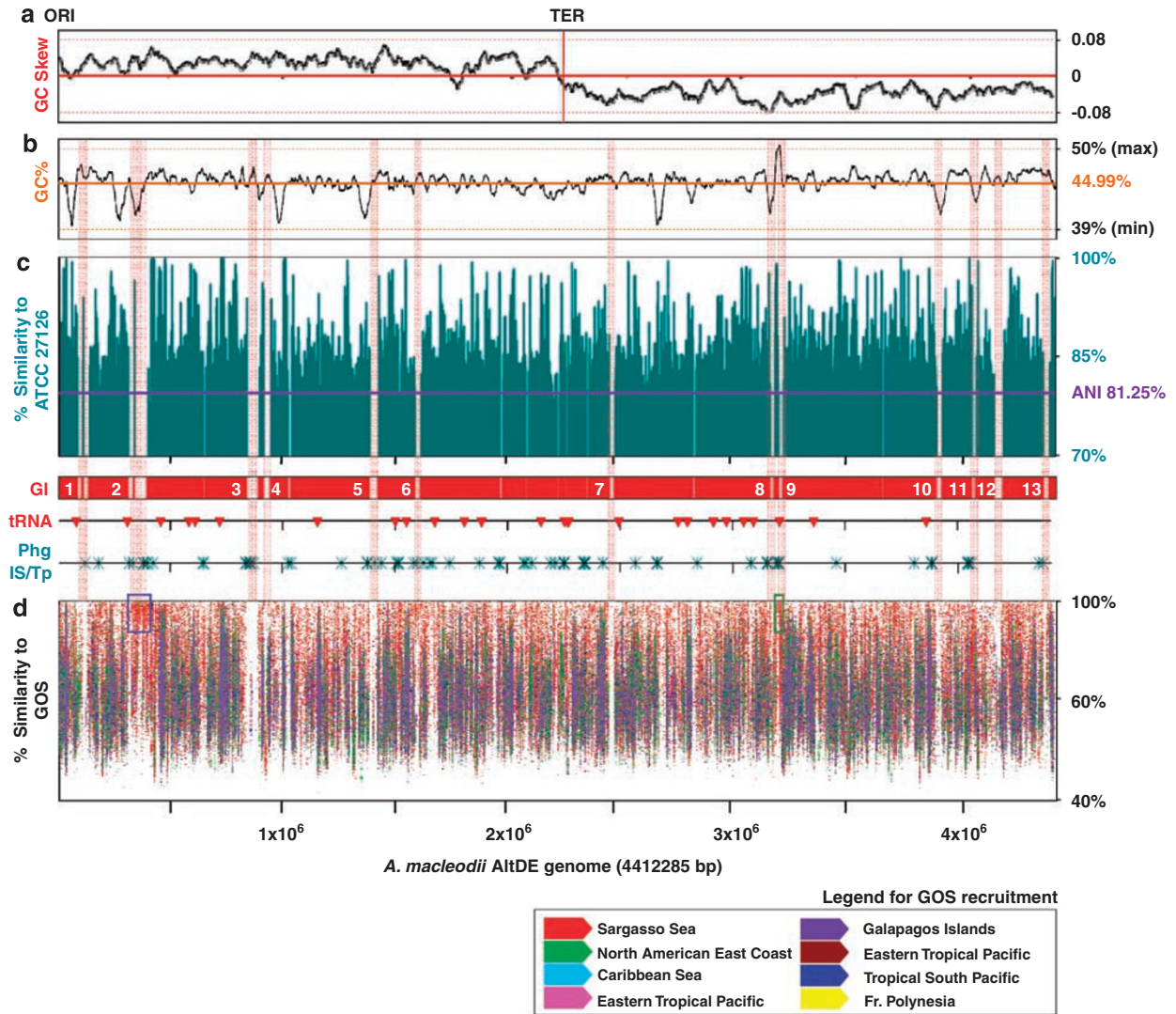
Brackets show the number of genes found only in one of the strains, that is, without ortholog in the other (see Materials and methods).

low similarity to any other transporters in the database, it was difficult to predict the substrate specificity in many cases.

#### Genome recruitment in the GOS database

We have analyzed the presence of both genomes in the large marine metagenomic databases of GOS (Venter *et al.*, 2004; Rusch *et al.*, 2007) (Figure 2d). A BLASTN comparison was performed between a local database constructed with the two *A. macleodii*

genomes and the complete GOS database (cutoff: 70% of the length of the GOS sequence with a similarity larger than 50%, see Materials and methods). This allowed us to distinguish the preferential recruitment of each environmental marine sequence for either AltDE or ATCC 27126 genome (Figure 3). Although the number of hits was not as large as those of extremely prevalent marine microbes like *Candidatus Pelagibacter* or *Prochlorococcus* sp., the number of sequences with high similarities to either of the genomes were sufficient to indicate that *A. macleodii* is indeed an abundant marine bacterium of global distribution. One thing of particular interest was that using similar parameters, other marine bacteria of assumed prevalence such as *Roseobacter* or *Rhodopyrellula* recruited significantly less (data not shown). In fact, considering that *A. macleodii* has relatively large cells (Baumann *et al.*, 1972) that are frequently associated to aggregates, its representation in most GOS samples (that represent only the 0.8  $\mu\text{m}$ –0.1  $\mu\text{m}$  size range) is remarkable. One single sample, GS025 (from a fringing reef in Costa Rica), that corresponded to the 3  $\mu\text{m}$ –0.8  $\mu\text{m}$  fraction, produced many more hits than the rest (25% of the sequences for ATCC 27126) confirming that the standard filtration protocols for marine bacteria selectively leave out *A. macleodii* as described previously (Acinas *et al.*, 1999; Garcia-Martinez *et al.*, 2002). BLASTN analysis showed that of a total of 15 029 hits retrieved using both genomes, 11 520 were more similar to ATCC 27126, while only 3509 had higher similarity to



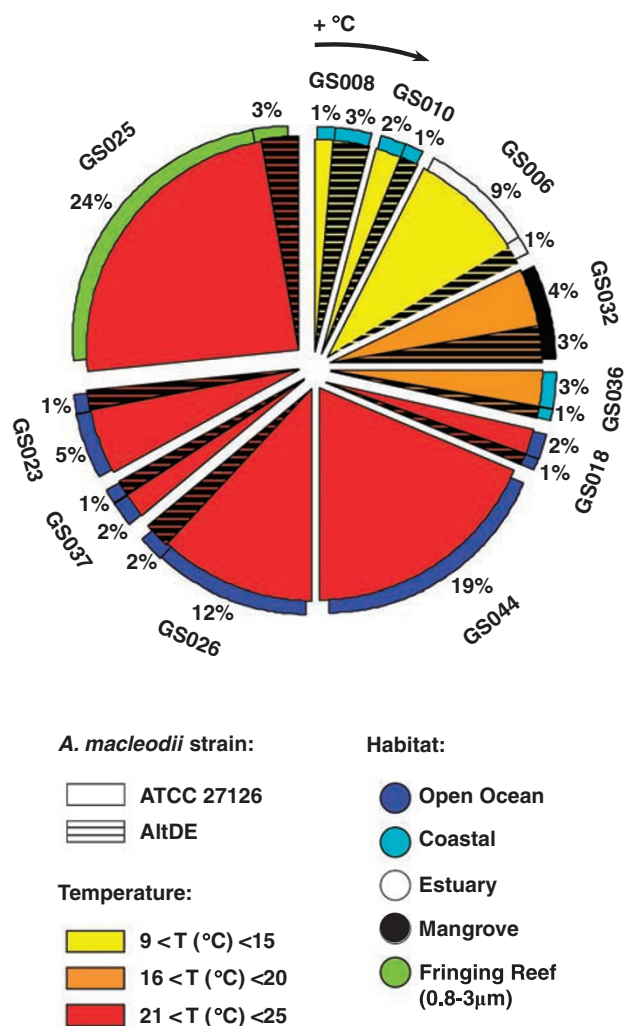
**Figure 2** *Alteromonas macleodii* AltDE genome and genomic islands. (a) GC-skew of *A. macleodii* AltDE genome plotted with a sliding window of 50 000 nucleotides. Origin (ORI) and terminus (TER) are shown. (b) GC-content plotted with a sliding window of 50 000 nucleotides. Average percentage of GC is shown in orange. (c) Individual ATCC 27126 ORFs aligned with the AltDE genome. Each vertical line on the graph represents an individual ORF sequence aligned along with its homologous region in AltDE, and its height indicates nucleotide similarity. Genomic islands present only in AltDE and larger than 20 kbp are shaded and numbered. Average nucleotide identity (ANI) is shown as purple line. tRNAs, integrases, IS transposases and phage-related proteins are also shown. (d) Mummerplot showing the recruitment of AltDE genome in the GOS database (see Materials and methods). Geographic origin is shown by a color code. AltDE, *Alteromonas macleodii* ‘deep ecotype’; GOS, Global Ocean Sampling.

AltDE (including Sargasso Sea collections). This is consistent with the nature of GOS samples as they were recovered in superficial waters (1 m–30 m deep). Other tendencies detected were a clear association of the species with warm temperatures, more pronounced in the ATCC 27126. In some samples below 20 °C, AltDE recruited more hits than ATCC 27126. Finally, there seems to also be a preference to estuaries and coastal areas for the ecotype represented by ATCC 27126 (Figure 3).

#### Genomic islands in *A. macleodii* deep ecotype

Of the 1242 AltDE genes without orthologs in ATCC 27126, 472 were clustered in GIs of more than

20 kbp. Using this threshold, it was possible to identify up to 13 GIs (Figure 2c). Table 3 provides general information about the GIs detected. The presence of unique genes in AltDE was confirmed by BLAST analysis not only against the 711 assembled contigs from the surface ecotype but also against the original unassembled sequences that provided over  $\times 12$  coverage of the genome. Most of the inferred differences between both strains at the level of metabolism and cell structure were located in islands and are described below. When we examined the recruitment of the complete AltDE genome in the GOS database, it was clear that some of these GIs were under-represented (Figure 2d), a phenomenon that has been shown for a number of prokaryotes when



**Figure 3** Distribution of the best BLASTN hits of different GOS database sequence collections to either AltDE (hatched) or ATCC 27126 (full) genomes (see Materials and methods). Wedges are ordered clockwise according to increasing sample temperature. AltDE, *Alteromonas macleodii* ‘deep ecotype’; GOS, Global Ocean Sampling.

compared with metagenomic databases in which they are well represented (Coleman *et al.*, 2006; Legault *et al.*, 2006; Cuadros-Orellana *et al.*, 2007). Among the ORFs present in the islands, 154 were hypothetical proteins and could not be assigned any function, but most GIs contained enough information to infer reliable annotation of functions and sometimes phenotypic properties that could be verified by laboratory experiments. Many islands contain genes that evoke phage-related ancestry (Supplementary Figure S3), but only GI 13 shows clear traits of a recently lysogenized phage (Supplementary Figure S3d).

#### Alternative catabolic pathways

The first part of GI 1 contains genes for a cytochrome D ubiquinol oxidase subunit II, a cytochrome BD-II

oxidase subunit I, an ABC cytochrome efflux transporter (MdlB) and an ATP-binding component of cytochrome-related transport (pink rectangle in Supplementary Figure S4a). This cluster has been found to be conserved in other genomes of marine bacteria (*Chromohalobacter salexigens* DSM3043, three genomes of *Marinobacter* (*M. sp* ELBB17, *M. algicola* DG893 and *M. aqueolei* VT8) and *Idiomarina baltica* OS145). The cytochrome BD complex is considered an alternative oxidase and oxidizes ubiquinol-reducing oxygen as part of the aerobic respiratory electron transport chain. In *Escherichia coli*, it is not expressed under normal conditions of growth but it is induced by low oxygen tension, in a shift from aerobic to anaerobic conditions, phosphate starvation and when bacterial cultures reach stationary phase (Sturr *et al.*, 1996). Also, in some nitrogen fixing bacteria, for example, *Klebsiella pneumoniae*, this oxidase is responsible for removing oxygen in microaerobic conditions and it is required for nitrogen fixation (Juty *et al.*, 1997). These findings suggest that this cluster of genes could provide an alternative respiratory chain in microaerobic conditions. On the contrary, in *Shewanella sp.* DSS12, a moderately piezophilic strain, this complex is expressed only at high pressure, 60 Mpa (Chikuma *et al.*, 2007). Therefore, it could be used preferentially at high pressure in *A. macleodii* as well. Following this cluster, there is a thiol-disulfide interchange protein, an alkyl-hydroperoxide-reductase, which is responsible for directly reducing organic hyperoxides in its reduced form. Thiol-specific antioxidant is a physiologically important antioxidant, which constitutes an enzymatic defense against sulphur-containing radicals. Therefore, these genes could be involved in preventing oxidative damage catalyzed by the microaerophilic respiratory chain located upstream when oxygen levels rise abruptly or even during its normal operation. We have also found very divergent copies of the cytochrome *c* oxidase subunits in GI6 (pink rectangle in Supplementary Figure S4b) with the best similarities: 76% with subunit I of *Pseudomonas stutzeri* A1501 and 66% with subunit II of *Azoarcus sp.* EbN1. None of these hits were at a similarity over 50% to any marine bacteria, what may explain why these two genes recruit very poorly in GOS, in spite of the widespread occurrence of cytochrome *c* oxidases in aerobic bacteria (data not shown). There is a more standard (phylogenetically coherent) set of cytochrome *c* oxidases in the genome (MADE\_04012 and MADE\_04013) and they resemble other marine  $\gamma$ -proteobacteria cytochrome *c* oxidases (the best matches with *P. atlantica* T6c: 96% subunit I and 86% subunit II, and *I. baltica* OS145: 93% subunit I and 81% subunit II). The paralogous set found in GI6 could also provide a wider adaptation at the respiratory level for AltDE.

One of the most common alternatives to oxygen respiration found in aerobic bacteria is nitrate respiration. A prominent feature of GI12 (Supplementary



**Table 3** Characteristics of genomic islands

GI	Start	End	Size (kpb)	Number of genes	GC content (%)	Hypothetical proteins	GOS recruitment	Inferred character	Codon usage deviation <sup>a</sup>	Phenotypic confirmation
1	85 980	139 932	53.9	50	42.51	15	Yes	Microaerobic respiration	1.366	ND
2	310 219	405 744	95.5	101	43.13	36	No <sup>b</sup>	Metal resistance Hydrogenases	1.515	+ ND
3	836 134	890 184	54	55	43.86	40	Yes	Putative phage	2.735	ND
4	909 176	941 608	32.4	29	43.59	3	Yes	CRISPR Urease	1.555	ND +
5	137 6741	1 415 870	39.1	37	40.54	16	Yes	Alternative LPS O-chain	1.611	ND
6	1 579 182	1 603 448	24.2	29	43.91	10	No	—	1.543	ND
7	2 438 494	2 462 099	23.6	7	45.09	1	Yes	Giant protein	3.391	ND
8	3 148 826	3 169 706	20.8	21	40.94	1	No	Flagellin	2.214	ND
9	3 187 850	3 208 697	20.8	21	48.45	0	No <sup>c</sup>	Mercury resistance	2.875	+
10	3 884 054	3 911 448	27.3	25	40.75	13	Yes	Putative phage	2.156	ND
11	4 040 946	4 060 115	19.1	30	40.39	9	No	Putative phage	2.511	ND
12	4 139 622	4 182 541	42.9	44	45.02	5	No	Nitrate respiration	1.7881	+
13	4 360 178	4 386 638	26.4	23	46.05	5	Yes	Lysogenic phage	2.859	ND

Abbreviations: GOS, Global Ocean Sampling; LPS, lipopolysaccharide; ND, not determined.

<sup>a</sup>Sum of the differences in codon use for each triplet between an Island and the Genome.

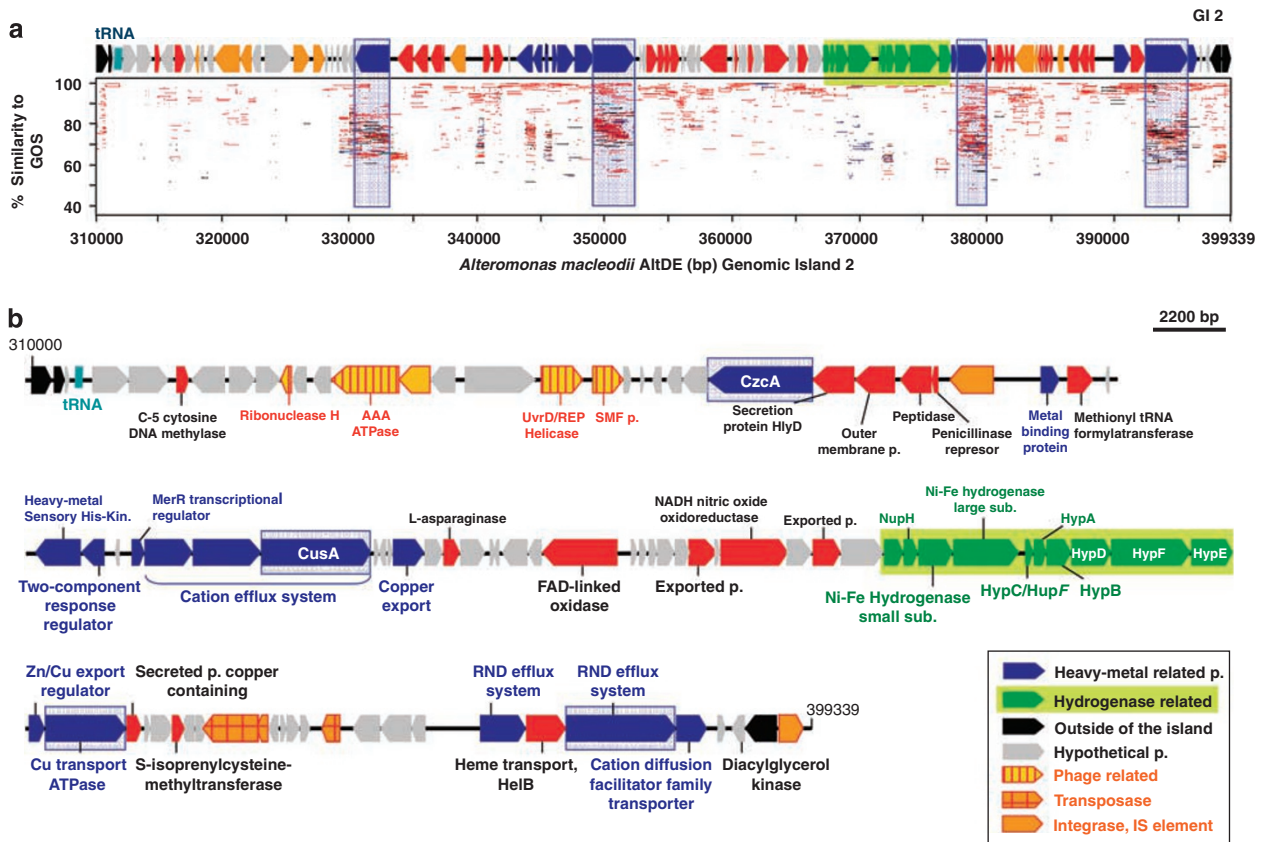
<sup>b</sup>With the exceptions of ORFs MADE\_00305, 0320, 0355 and 0374.

<sup>c</sup>With the exceptions of genes of the *mer* operon and ORFs MADE\_01413, 01414, 01415 and 01416.

Figure S4c) is a cluster for the nitrate reductase ( $\alpha$ -,  $\beta$ - and  $\gamma$ -subunits) or *nar* system for nitrate respiration. Also, many of the genes found here are related to the molybdopterin cofactor biosynthesis, which is required for Nar-complex function, and several transporters for molybdenum and nitrate. Curiously, the best hits in BLASTX comparisons for the Nar subunits were with the marine-sediment isolate *Hahella chejuensis* KCTC 2396 (similarities of  $\alpha$ -subunit: 89%,  $\beta$ -subunit: 89%,  $\gamma$ -subunit: 79% and chaperone NarJ: 72%). This taxonomic affiliation was also shared by the parvulin-like protein, the histidine-kinase nitrate specific and one of the three nitrate/nitrite transporters (MADE\_03852). However, the molybdenum transporter system was not found in the same gene cluster in *H. chejuensis* KCTC 2396 and the best matches of these genes were with *Methylococcus capsulatus* str. Bath. The lack of *nir* and *nor/nos* genes in the genome suggests that only reduction to nitrite rather than  $N_2$  is present. This feature was checked phenotypically (data not shown) and AltDE was able to reduce nitrate without gas production, while ATCC 27126 did not reduce nitrate at all. In ATCC 27126, only the *narA* gene, which codes for the catalytic subunit, was found without any other of the rest of required subunits of the complex. In GI2 (see below), there is also a gene presumably involved in nitric oxide detoxification, which may increase survival odds in a microaerophilic/anaerobic interphase where denitrification by other members of the particle consortium is actively taking place (Figure 4b). Similar genes are found in many other proteobacteria including *E. coli* (Gardner and Gardner, 2002). Taken together, all the above properties of the AltDE genome reflect a better

adaptation to reduced oxygen availability, either permanent or cyclic, and could be connected to a specialization in degrading larger organic aggregates that would often contain anaerobic microniches (Shanks, 1983).

A hydrogenase cluster was discovered in GI 2 (Figure 4b). Oxidation of molecular hydrogen to water catalyzed by hydrogenases is a widespread mechanism of energy generation among prokaryotes, regulated negatively by organic energy sources (Vignais and Billoud, 2007). Many bacteria and archaea species harbor multiple hydrogenases, which mediate heterolytic cleavage of  $H_2$  into  $2 H^+$  and  $2 e^-$ . The energy yielded is recovered in the form of ATP by the chemiosmotic mechanism of oxidative phosphorylation. Ni-Fe hydrogenases are the most common enzymes, representing a fairly conserved family of proteins, composed of (at least) a large active site-containing subunit and a small electron-transferring subunit, which bears one to three FeS clusters. The hydrogenase cluster found in GI2 is a multicomponent system formed by two hydrogenases (large and small subunit) grouped together with a set of auxiliary proteins involved in assembly (HypC/HupF), enzyme expression (HypE and NupH) and formation-maturation (HynD/HyaD, HypA, HypB, HypD and HypF). These genes could have been acquired through an ancient horizontal gene transfer event from a  $\beta$ -proteobacterium since at least 7 of the 10 proteins have the best hit with their homologs in *Thiobacillus denitrificans* ATCC 25259. It has been described previously that this cluster has been horizontally transferred through some large plasmids in *Alcaligenes* (now *Ralstonia*) *eutrophus* H16 and *A. hydrogenophilus* (Lenz and Friedrich, 1998). Interestingly, a cluster of conserved



**Figure 4** Genomic island 2 (GI2). (a) Mummerplot recruitment of GI2 in the GOS database. Blue rectangles indicate the four cation-efflux pumps with an over-representation of GOS database sequences. (b) Detailed schematic representation of the ORFs depicted in GI2. GOS, Global Ocean Sampling.

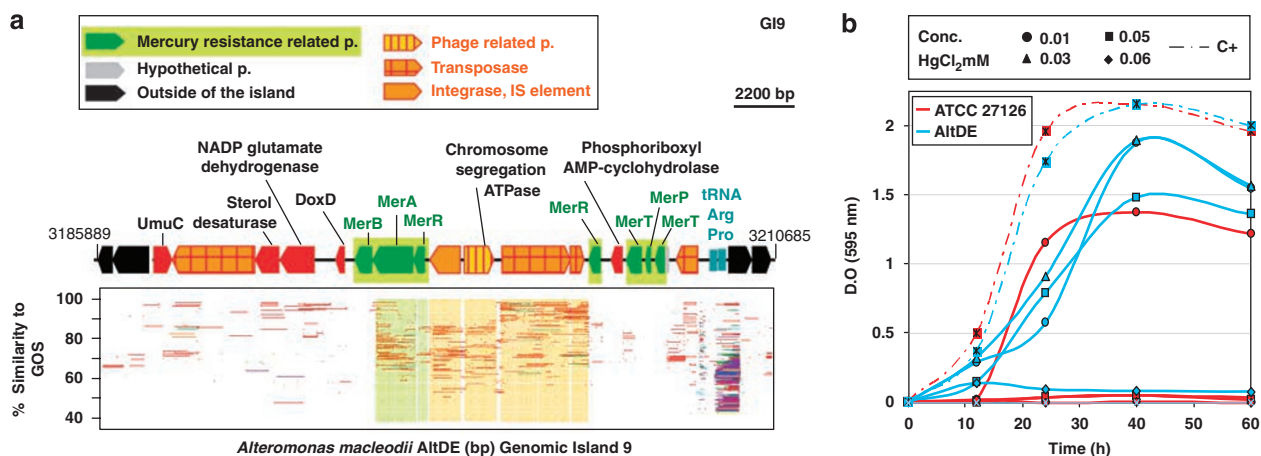
sinteny and sequence similarity to that of AltDE is found in the pBBta01 plasmid of the  $\alpha$ -proteobacterium *Bradyrhizobium* sp. BTAi1. Although hydrogen is generated in aerobic conditions, for example, during photosynthesis, it might also be an advantage in the aerobic/anaerobic interphase as it is often generated by fermentation of organic matter.

Finally, a cluster of genes involved in urea transport and utilization was found in GI4 (Figure 6). This is consistent with the ureolytic properties of the AltDE strain. Urea is a common nitrogen excretion product of animals and can be found in living and decaying zooplankton (Alldredge, 1976; Pomeroy 1980; Bruland, 1981). The capabilities of AltDE regarding this compound could be important to life associated to their excreta or bodies, either dead or alive.

#### Heavy metal resistance

Unexpectedly, many GI genes were related to heavy metal resistance, but this genomic feature is consistent with the phenotypic findings (see below). GI2 is very special in showing punctual over-recruitment of GOS metagenome fragments, or more similar fragments at higher similarities, than the average core genome (blue rectangle in Figure 2d). A finer dissection of the island shows that this reflects

the extremely efficient recruitment of four genes present in the island that were all annotated as metal efflux pumps (Figure 4a, blue rectangles, MADE\_00305, 0320, 0355 and 0374). Metal detoxification seems to be the common motive of the gene clusters in GI2, including putative efflux pumps for cobalt, zinc, cadmium, silver or copper. There is also a metal sensor histidine kinase that may possibly be involved in the regulation of the metal response. The region between the first metal efflux pump, CzcA, and the phage integrase (Figure 4a) has been found in several *Shewanella* genomes (*S. putrefaciens* CN-32, *Shewanella* sp. W3-18-1 and *Shewanella* sp. ANA-3). Along the same lines, GI9 mainly comprises a mercury resistance operon, which may be enough in itself to explain the higher mercury resistance found experimentally in AltDE (Table 2 and Figure 5b). Mercury is a highly toxic metal and its toxicity can result from three different mercurial forms: elemental, inorganic ions and organomercurial compounds. The ability of bacteria to detoxify mercurial compounds by reduction and volatilization is conferred by the *mer* genes, which are usually plasmid encoded (Kiyono *et al.*, 1997). The *mer* genes of GI9 code for a mercury ion reductase (MerA), an organomercurial lyase (MerB) and two transcriptional regulators (MerR) 77% similar to each other (Figure 5a). Three more MerA



**Figure 5** Genomic island 9 (GI9). (a) Schematic representation of the ORFs. Genes in green are related to mercury resistance. (b) Growth curve of AltDE and ATCC 27126 isolates at different concentrations of mercury (growth curves at different concentrations of Zinc are shown in Supplementary Figure S6). AltDE, *Alteromonas macleodii* ‘deep ecotype’.

proteins (MADE\_O1787: 21.2%, MADE\_O2041: 50.8% and MADE\_O1039: 55%) can be found outside this island. MerA is the key enzyme in detoxification of mercury in bacteria ( $\text{Hg}^{2+}$  is reduced to  $\text{Hg}^0$  gas) but cannot function without the lyase subunit MerB that mediates the first of the two steps in the detoxification of organomercurial salts (protonolysis of the C-Hg bond) (Huang *et al.*, 1999). The cluster of GI9 encodes a complete common molecular mechanism for mercury resistance and its volatilization in aerobic heterotrophic aquatic communities (Barkay *et al.*, 1990). AltDE MerA–MerB proteins are very well conserved in several marine genomes such as *Marinobacter aquaeoli* VT8, in the  $\gamma$ -proteobacterium KT71, and *Marinobacter* ELB17. The AltDE mer cluster is enclosed between two inverted Tn3 family transposases and could be a transposon. Downstream from it and also surrounded by two other transposases, there is a mercuric transport and sensing system, which comprises two putatively paralogous MerT genes (41% similarity between them), a periplasmic MerP subunit and one MerR transcriptional regulator. Finally, there is also a protein implied in histidine metabolism (a phosphoriboxyl-AMP-cyclohydrolase) that, even if it is not necessarily related to mercury resistance, in *Alkalilimnicola ehrlichei*, MLHE-1 appears next to an ABC transporter system that could be coupled to the transporter MerT. In conclusion, GI9 seems to be a collection of mer determinants that have been acquired by transposition in this region of the AltDE genome. This region shows one of the highest differences in codon usage (Table 3 and Supplementary Figure S7b) and tetranucleotide frequency (Supplementary Figure S7a) and has all the hallmarks of having been acquired by horizontal gene transfer. This island, like GI2, is over-represented in the GOS metagenome (green rectangles in Figure 2d and in Figure 5a), mainly due to the region between merA and the 3' copy of the Tn3-like transposase. This might

indicate that this putative transposable element is widespread in marine bacteria.

The heavy metal resistance of the 23 strains of *A. macleodii* present in our laboratory (Ivars-Martinez *et al.*, 2008) was assessed by determining the minimal inhibitory concentration for zinc, mercury and lead. We could prove that most deep isolates were more resistant (data not shown). Specifically, AltDE is significantly more resistant to zinc and mercury (Figure 5b and Supplementary Figure S6). Although the ecological meaning of the increased mercury resistance shown by most of the strains belonging to the genotypic cluster represented by AltDE is not obvious, there could be a connection to the particulate organic matter that offers sites for adsorption of heavy metals and trace elements (Hebel *et al.*, 1986).

#### Phage resistance and cell surface

Many genes of the islands may be related, or reflect, interactions with phages. GI3, 10 and 13 have all the hallmarks of lysogenic phages or phage remnants (Supplementary Figure S3) and shown important codon usage deviations (Table 3, Supplementary Figure S7). GI3, 10 and 11 have large noncoding spacers and small hypothetical proteins, and the few that could be annotated (a DNA polymerase, an ATPase AAA superfamily and a transcriptional regulator) were clearly phage related. Both GIs 10 and 13 contain restriction modification systems. GI13 has high coding density and most genes are clearly associated to phages indicating that it could be in fact a recently lysogenized phage. On the other hand, GI10 is highly degraded and the restriction modification is nearly the only gene cluster that appears functional. If it was a phage, it would degrade beyond recognition.

GI4 contains two parts, the first half is a set of genes of the phage immunity system known as CRISPR with the associated cas/cse genes and the

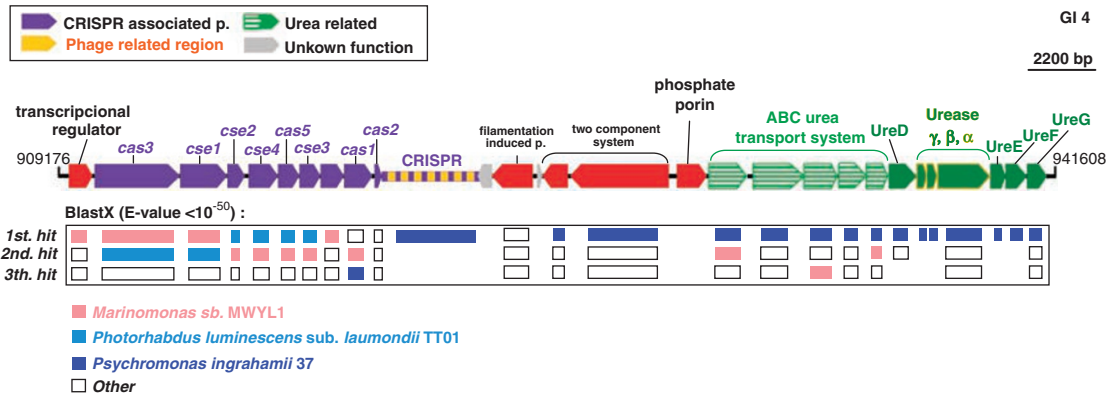
tandem repeats (Figure 6) (Sorek *et al.*, 2008). Neither ATCC 27126 nor *P. atlantica* T6c has the CRISPR feature in their genomes. Interestingly, this part of *A. macleodii* AltDE genome shows a mosaic relationship mainly with the marine  $\gamma$ -proteobacteria *Marinomonas* sp. MWYL1 and *Psychromonas ingrahamii* 37. The dynamic nature of this genome feature is shown in Figure 6. The Cas/Cse proteins of AltDE are very similar to the ones of *Marinomonas* sp. MWYL1 (which has two CRISPR systems), ranging from 67% (Cse2) to 92% (Cas1). The similarity of AltDE Cse2 and Cse3 is even higher to *Photorhabdus luminescens* subunit *laumondii* TT01, a nematode-associated enterobacterium, which has lost the tandem repeats (Duchaud *et al.*, 2003). The psychrophilic  $\gamma$ -proteobacterium *P. ingrahamii* 37 was isolated from Arctic waters and has two CRISPR-protein related systems in its genome, but only one of them seems to be functional as the other has lost the tandem repeats region. Interestingly, AltDE palindromic tandem repeats are 100% identical to those of *P. ingrahamii* 37. On the other hand, the CRISPR spacers are completely different. However, of the associated proteins, only Cas1 of *P. ingrahamii* 37 is similar to that of AltDE (91%). The urease cluster also found in GI4 (see above) and the preceding genes (Figure 6; with the exception of the phosphate porine), also show the highest similarity to *P. ingrahamii* 37, ranging from 96% (urease  $\alpha$ -subunit) to 70% (UreD). All these data seem to indicate that the whole urea-CRISPR cluster repetitions-Cas1 has been acquired by AltDE from a relative of *Psychromonas* or vice versa. These observations may suggest that the CRISPR modules could be horizontally transferred between largely divergent groups (Godde and Bickerton, 2006) and that due to different biogeography, they evolve differently, changing the variable part according to dominating viral sequences. In the GOS metagenome, the urea cluster recruited well, while the *cas-cse* genes and the repeats recruited very little and with low similarity (data not shown). This is another example of the extremely dynamic nature of CRISPR elements.

GI6 is probably a composite transposon with two identical transposases at its ends (Supplementary Figure S4b). On both sides of the putative transposon, we found genes of related function, involved in carbohydrate metabolism, reinforcing the impression that a large and recent insertion has occurred in this strain. GI6 contains also the stress-induced toxin-antitoxin system MazEF, which has its closest homologs in *Shewanella putrefaciens* 200. In *E. coli*, MazEF-mediated cell death has been described as triggered by stress by antibiotics, ultraviolet irradiation or high population density (Kolodkin-Gal and Engelberg-Kulka, 2006; Kolodkin-Gal *et al.*, 2007). It can prevent the spread of phage infection as it causes an infected cell to commit suicide (Hazan *et al.*, 2004). Thus, it could be an efficient phage defence system. Alternatively, and given the role of

toxin-antitoxin systems in plasmids maintenance (Hayes 2003), GI6 may just reflect the chromosomal integration of a plasmid fragment.

GI5 contains many genes involved in extracellular polysaccharide biosynthesis, probably the outer membrane lipopolysaccharide (LPS), as O-antigen- and LPS-related genes appear nearby (Supplementary Figure S5c). The abundant transposases present at both ends indicate potential mechanisms of insertion and deletion of this cluster, in spite of its large size (40.4 kbp). The island contains seven different glycosyl-transferases, which belong to several mannose transferases, LPS core biosynthesis and galactosyltransferases families. However, proteins located between the glycosyltransferase MADE\_O1298 and MADE\_O1309 are also present with nearly identical synteny, although at lower similarities, in *P. atlantica* T6c, so this region appears to have been deleted in ATCC 27126. Also related with the LPS are the genes comprised between MADE\_00965 and MADE\_00985 (Supplementary Figure S5a). This part of the genome, not defined here as an island (as it is smaller than 20 kbp) contains unique genes of AltDE: an LPS A protein, together with two proteins for manose modifications and an iduronate sulfatase. The population genomics of the LPS O-chain polysaccharide synthesis genes has been shown before to be extremely complex and dynamic (Samuel and Reeves, 2003; Giovannoni *et al.*, 2005; Coleman *et al.*, 2006) and is most likely related to its major role as phage receptor (Samuel and Reeves, 2003; Sharma *et al.*, 2008).

In addition to the differences in the LPS O-chain, there are abundant evidences showing the presence of a capsular or slime layer exopolysaccharide (EPS) and the potentiality for biofilm formation in both strains (Flemming *et al.*, 2007). The mucous consistence of growth on solid media of AltDE and many other deep isolates have been known for many years (Rougeaux *et al.*, 1998). ATCC 27126 and other surface isolates on the other hand have much less mucous colonies (data not shown). Consistently, the genomes show many differences regarding capsular EPS biosynthesis gene content (although they were not localized in islands). For example, the cluster found from MADE\_02585 to MADE\_02592 (Supplementary Figure S5b) has at least 18 EPS biosynthesis proteins in AltDE (including an epimerase (WcaG), an isomerase (GutQ), an acetyltransferase (WbbT), CapC, seven glycosyltransferases, ExoD, two GumC and other four putative-related proteins, whereas we found only six in ATCC 27126 (CapC, two sugar phosphoisomerases and other three putative capsular polysaccharide biosynthesis-related proteins). In addition, AltDE possess a larger number of proteins related to the export of polysaccharides (five vs three found in ATCC 27126). This suggests a more complex matrix for the deep isolate and agrees with the higher mucosity observed in AltDE colonies. Although there is genetic evidence that EPS production might be important for growth at



**Figure 6** Genomic island 4 (GI4). Schematic representations of the ORFs. Clustered regularly interspaced short palindromic repeats (CRISPR) and their intergenic regions are shown in purple striped squares. The taxonomic affiliation of the best hits for each gene is shown below.

low temperatures (Lauro *et al.*, 2008), we could not detect differences in growth rate or range at different temperatures between AltDE and ATCC 27126 (data not shown). However, the carbohydrate composition of the EPS described in *A. macleodii* subsp. *fijiensis*, (isolated at 2000 m near a hydrothermal vent in the North Fiji Basin) is rich in glucuronic acid and, by its chemical characteristics, could be expected to have some heavy metal-binding ability. AltDE EPS could have similar characteristics and would help explain the higher number of metal detoxifiers in this strain compared with the surface type.

GI7 was composed mostly of a gene coding for a 6492 aminoacids 'giant protein', (Reva and Tummeler, 2008), with the highest anomaly of codon usage found in the genome (Table 3) and with no significant similarity in the metagenome or ATCC 27126 (Supplementary Figure S5d). This protein contains a VCBS motif found in other giant proteins and could be involved in adhesion. Although the role of giant proteins in prokaryotes is far from being fully understood, one of their proposed functions is that it is involved in cell adhesion, or is a constituent of the extracellular matrix (Reva and Tummeler, 2008). Only two other described giant proteins, found in *Roseobacter litoralis* and *Roseobacter denitrificans*, showed significant similarity (*ca* 51%) to the one of AltDEs. Giant proteins are known to be strain specific, and again, their availability as phage recognition targets could be the driving force behind this high variability.

GI8 contains a cluster (20 kbp) of flagellin genes that comprise of proteins from the exposed section of the flagellum (Supplementary Figure S5e). In AltDE, the motility genes (flagellins and capping protein, basal body, hook-length regulator, switch, export apparatus, flagellum placement determinant, flagellum number regulator, chemotaxis and motor proteins) are distributed in six chromosomal regions; however, a large cluster contains most of them (MADE\_02870-MADE\_02935, comprising 61.7 kbp). GI8 is located in the middle of this large cluster and includes Maf1, (a motility accessory

factor) FliS (a chaperone), FliD (which forms the distal end of the flagella and plays a role in mucin-specific adhesion in some pathogenic bacteria), FlaG (a flagellar structural protein) and two flagellin FliC (the mayor filament protein), thus, this island contains genes coding for the major exposed components of the flagellum. The ATCC 27126 strain contains homologs to all of these genes, but with lower similarities (to Maf1: 44%, to FliS: 62%, to FliD: 60%, to two copies of FliC: 59% and 64%). The major structural component of the flagellum, FliC, is present in at least four copies in ATCC 27126 and six in AltDE. This again illustrates the requirement for high variability in outside structures potentially exposed to phage recognition (Zhilenkov *et al.*, 2006; Coward *et al.*, 2006). Along the same lines, GI8 also contains genes related with sialic acid synthesis genes, CMP-N-acetylneuraminic acid synthase (NeuA) and a sialic acid synthase (NeuB). Sialic acid has been shown in  $\alpha$ -proteobacteria to influence phage adsorption (Defives *et al.*, 1996). In *Aeromonas*, it has been described that Maf1, together with Maf-2, NeuB-like, FlmD and NeuA-like proteins are responsible for the glycosilation of polar and lateral flagella (Canals *et al.*, 2007). In AltDE, *neuA* is separated from *neuB* by an insertion of six transposases together with an integrase and a methionyl-tRNA-formyltransferase. In *Alteromonadales bacterium* TW-7 and other genomes, genes *neuA* and *neuB* are contiguous; thereby indicating that AltDE has acquired this transposase-rich insertion. If this insertion affects the expression or activity of the *neu* genes, it could also alter the glycosilation pattern of the flagellin, altering again a potentially important phage target.

## Discussion

Here, we present the genome of isolate AltDE characterized as a Mediterranean deep ecotype representative (Lopez-Lopez *et al.*, 2005). It was isolated from 1000 m deep in the South Adriatic basin, although we have shown that the close

relatives are found in the deep Mediterranean, down to 3000 m. In fact, the common retrieval of sequences related to this microbe from the deep Mediterranean metagenome indicates that it might represent a significant fraction of the microbiota in this habitat (Zaballos *et al.*, 2006; Martin-Cuadrado *et al.*, 2007).

For comparative purposes, an unassembled genome sequence has been obtained by pyrosequencing for the surface isolate ATCC 27126 isolated by Baumann *et al.* (1972). These are representatives of two subspecies or ecotypes (Lopez-Lopez *et al.*, 2005; Ivars-Martinez *et al.*, 2008), with the so-called surface ecotype being found worldwide in surface waters, while the deep ecotype representatives are found nearly exclusively in deep Mediterranean samples. Comparative genomics of both strains show that they represent very divergent lineages and could be considered borderline members of the same species (ANI 81.25%). The comparison of the two genomes described here with the GOS database permitted us to assess the relative contributions of both genomes to the prokaryotic biomass in a significant section of the world oceans, although only at shallow depth. This comparison indicated that *A. macleodii* species seems to have a predilection for warmer waters and for particulate organic matter, supporting previous evidence that this microbe is an *r*-strategist with preference to the nutrient rich particulate fraction microniche (Acinas *et al.*, 1999; Lopez-Lopez *et al.*, 2005). When the fragment recruitment of both genomes were compared, it was clear that the 'surface' ATCC 27126 genome was largely dominant, particularly at higher temperatures. However, AltDE relatives seem to be also present in the wide set of surface habitats represented in the GOS, although in consistently smaller amounts, and with their relative contribution increased at lower temperatures. In a recent work, close relatives of the AltDE strain were isolated from surface waters in the English channel (Ivars-Martinez *et al.*, 2008) and one isolate in the original study was retrieved at the surface in the eastern end of the Mediterranean (Lopez-Lopez *et al.*, 2005). This all leads to the question, which are the specific adaptations that relegate these organisms to the brink of becoming a different species? We know that there are no sexual barriers, because evidence of recent recombination among representatives of both ecotypes has been found frequently (Lopez-Lopez *et al.*, 2005; Ivars-Martinez *et al.*, 2008). In fact, the comparative genomics provides some clues of very different lifestyles.

*A. macleodii* deep ecotype has GIs and features that are consistent with deep dwellers, at least the easily cultured types, mostly the high abundance of transposases and phage-related elements. However, other than the slightly larger set of chaperones, other differences associated to life in the deep were not found in AltDE (Lauro and Bartlett, 2008). Particularly contradictory in this sense was the

finding of photoreactivation genes, which indicate that, at least cyclically, this microbe can be found in the photic zone and it is not an obligate bathytype (Lauro and Bartlett, 2008). The recruitment by AltDE genome of high-similarity hits in the GOS surface metagenome also supports this view. Thus, this ecotype is adapted to cool deep waters, but can also survive on surface waters. Conversely, many features indicate that AltDE has a preference to live within aggregates of high population density. The abundant exopolysaccharide that confers the colonies of AltDE and all the deep ecotype isolates, a very mucous morphology, is consistent with life in biofilms or polymer aggregates. At the physiological level, AltDE seems to be much better prepared to withstand microaerophilic and even transient anaerobic conditions, as should be expected in larger aggregates and, for example, zooplankton fecal pellets (Shanks and Reeder, 1993). In addition, the degradative properties are consistent with this niche differentiation. The larger number of dioxygenases in AltDE and the presence of urease suggest a specialization on degrading recalcitrant organic matter.

ATCC 27126, then again, appears to have a more adaptable biology with more complex regulation and environmental sensing, and uses a wider assortment of common rapidly utilized nutrients such as sugars and amino acids. These differences at the genomic level might be a reflection of having larger periods of free-living existence as would be expected from inhabiting smaller particles whose nutrient content is rapidly depleted (Karl *et al.*, 1988). In other words, smaller organic aggregates of relatively short lifespan would be more successfully colonized by the ecotype represented by ATCC 27126, while larger, rapidly sinking long-lived particles would favor the AltDE strategy (Karl *et al.*, 1988; Simon *et al.*, 2002; Grossart *et al.*, 2003; Grossart *et al.*, 2007). A remarkable and totally unexpected difference found was the higher tolerance of AltDE to heavy metals. Actually, some of the transporters annotated as cation efflux pumps could be involved in detoxifying other substances (Silver and Phung, 1996). In any case, their presence is not inconsistent with association to larger particles. The long and more complex degradative processes that would take place within these particles would in the long term accumulate toxic metabolites. It may also attract metallic ions that would form complexes with the recalcitrant compounds generated by them. Furthermore, biofilms and capsules required for this lifestyle would tend to accumulate toxic metals (Hebel *et al.*, 1986).

Why is AltDE found to be more abundant in deeper waters? First of all, sampling methodologies that screen out larger aggregates would bias the results toward increasing the apparent weight of the ecotype represented by ATCC 27126. Secondly, larger particles, regardless of where they are generated, will sink more rapidly than smaller ones

(Simon *et al.*, 2002). If this hypothesis is true, the proper ecotype denomination should be 'small aggregates' for the 'surface' and 'large aggregates' for the 'deep'. Living in populations of much higher density, the accessibility to phage predation should be higher in the larger aggregates (Riemann and Grossart, 2008). In fact, the chances of a phage particle to find prey in the highly diluted offshore waters appear slim, and the origin of the large oceanic phage populations has always been somewhat of a puzzle. However, if the phages predate mostly at the level of aggregated communities, they might be extremely successful (Riemann and Grossart, 2008). Interaction with phages seems to be a fundamental factor in preserving microdiversity among prokaryotic groups, and the enormous diversity and variability of exposed components found in these and other genomes strengthen this view (Coleman and Chisholm, 2007; Cuadros-Orellana *et al.*, 2007; Kettler *et al.*, 2007; Kunin *et al.*, 2008). The genomes studied here had clearly different O-chains in the LPS as well as different external flagellar components. Besides, AltDE contains several extra protection features indicating that for this, ecotype surviving phage predation might be even more important. Specific examples include the CRISPR modules, the toxin-antitoxin systems, the restriction-modification enzymes or even the giant protein of GI7. In addition, the genome shows abundant evidence of recent and older phage integration.

One interesting phenomenon that we have found is that GIs detected when comparing different strains do not always coincide with 'meta-GIs', in the sense of recruiting more poorly metagenomic fragments. It has been a rather puzzling discovery (Coleman *et al.*, 2006; Legault *et al.*, 2006; Cuadros-Orellana *et al.*, 2007) that some parts of the genomes of strains recruit much less than others even in metagenomic databases where the strain providing the scaffold genome is very abundant and even predominant. Generally, these meta-GIs overlap with the islands found when comparing different strain genomes. As mentioned above, the exposed parts of the cell such as the O-chain of the LPS have genes in clusters that form both genomic and meta-GIs, as lineage diversity at these regions guarantee that any strain-specific cluster is poorly represented in the lineage ensemble that the metagenome may represent. However, we have also found a few instances where parts of the GI over-recruited in the metagenome. One example is found in GI2 where some efflux pumps recruited many highly similar sequences, much more so than most core genes. The most obvious explanation for this is that such genes are present in many different species with a high level of conservation, perhaps because they are often exchanged by horizontal gene transfer and/or because their function is extremely sensitive to sequence variation (Zhaxybayeva *et al.*, 2004). These genes could have

an important role for survival in the marine environment.

The genomic data presented here are consistent with a system of resource use diversification generating two different bacterial ecotypes that nevertheless maintain genetic recombination. In bacteria, the genetic distance between two populations increases if mutation is higher than recombination and some authors view speciation as the outcome of recombination incompatibility (Fraser *et al.*, 2007). The reduction in recombination frequency may be caused by allopatry or niche specialization and although physical barriers may apply for host-associated bacteria, there is substantial evidence that geographic barriers do not exist for most free-living prokaryotes. Previous housekeeping gene sequence data for *A. macleodii* shows that there is not recombinational (that is, sexual) isolation (Lopez-Lopez *et al.*, 2005; Ivars-Martinez *et al.*, 2008). Instead, our genomic data draws a picture of niche specialization by metabolic diversification in the two ecotypes. Given that many genomic differences found among the strains of other free-living bacteria related to resource utilization (Cuadros-Orellana *et al.*, 2007; Kettler *et al.*, 2007) and that experimental evolution of an ancestral lineage may lead to population splitting by adaptive diversification (Friesen *et al.*, 2004; Tyerman *et al.*, 2005), the importance of sympatric speciation in bacteria should be re-evaluated.

It is noteworthy that many of the genes that allow the metabolic diversification of the two ecotypes (for example, anaerobic respiration, aggregation or metal detoxification) show evidence of recent horizontal transfer. This suggests that horizontal gene transfer could be an important mechanism underlying population splitting by resource specialization, as it has previously been suggested (De la Cruz and Davies, 2000). Whether horizontal gene transfer promotes resource exploitation diversification in bacterial populations should be further studied, as this peculiarity of prokaryotes could enable sympatric speciation processes more frequently than previously anticipated.

## Acknowledgements

We thank Mila Zaballos for providing the AltDE isolate. Sequencing, assembly and annotation efforts were supported by the GBMF as part of its Marine Microbial Sequencing Project ([www.moore.org/marinemicro](http://www.moore.org/marinemicro)). We thank Granger Sutton and his team for the ongoing development and maintenance of the Celera Assembler and related tools and Aaron Halpern for the help with specific analysis issues. The JCVI software team, under the leadership of Saul A Kravitz, manages assembly, annotation and web delivery of data for the GBMF funded project (<http://research.venterinstute.org/moore>). We thank Robert Friedman for his leadership of the GBMF funded project. We acknowledge the J Craig Venter Institute (JCVI) Joint Technology Center, under the leadership of Yu-Hui

Rogers, for producing the genomic libraries and the sequence data. EIM acknowledges Caja de Ahorros del Mediterráneo (CAM) for the predoctoral Fellowship. ABMC was supported by an MEC (Spain) Postdoctoral Fellowship. GD was supported by the GEMINI (QLK3-CT\_2002-02056). Work at FRV laboratory was funded by the Spanish Ministry of Education and Science Grant CTM2005-04564.

**Authors' Contributions:** FRV conceived and led the work. EIM did all the experimental work and assembled the 111 AltDE contigs into one. EIM and ABMC carried out the manual annotation. EIM, ABMC and GD performed computational analyses. EIM and FRV analyzed the general features of the genome. FRV, ABMC and AM wrote the manuscript. ABMC and EIM prepared the figures and tables. SF, JJ and RF sequencing assembly and autoannotation. All authors read and approved the final manuscript.

This manuscript is dedicated to Dr Paul Baumann, who recently retired, in appreciation for his inspirational work in the field.

## References

- Acinas SG, Anton J, Rodriguez-Valera F. (1999). Diversity of free-living and attached bacteria in offshore Western Mediterranean waters as depicted by analysis of genes encoding 16S rRNA. *Appl Environ Microbiol* **65**: 514–522.
- Allredge AL. (1976). Discarded appendicularian houses as sources of food, surface habitats and particulate organic matter in planktonic environments. *Limnol Oceanogr* **21**: 14–23.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990). Basic local alignment search tool. *J Mol Biol* **215**: 403–410.
- Barkay T, Gillman M, Liebert C. (1990). Genes encoding mercuric reductases from selected gram-negative aquatic bacteria have a low degree of homology with merA of transposon Tn501. *Appl Environ Microbiol* **56**: 1695–1701.
- Baumann L, Baumann P, Mandel M, Allen RD. (1972). Taxonomy of aerobic marine eubacteria. *J Bacteriol* **110**: 402–429.
- Brundland SKWMWS. (1981). Sinking rates of fecal pellets from gelatinous zooplankton (salps, pteropods, doliolids). *Mar Biol* **63**: 295–300.
- Campanaro S, Vezzi A, Vitulo N, Lauro FM, D'Angelo M, Simonato F et al. (2005). Laterally transferred elements and high pressure adaptation in *Photobacterium profundum* strains. *BMC Genomics* **6**: 122.
- Canals R, Vilches S, Wilhelms M, Shaw JG, Merino S, Tomas JM. (2007). Non-structural flagella genes affecting both polar and lateral flagella-mediated motility in *Aeromonas hydrophila*. *Microbiology* **153**(Pt 4): 1165–1175.
- Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J. (2005). ACT: the Artemis Comparison Tool. *Bioinformatics* **21**: 3422–3423.
- Chikuma S, Kasahara R, Kato C, Tamegai H. (2007). Bacterial adaptation to high pressure: a respiratory system in the deep-sea bacterium *Shewanella violacea* DSS12. *FEMS Microbiol Lett* **267**: 108–112.
- Christian WB. (1946). Urea descomposition as a means of differentiating *Proteus* and paracolon cultures from each other and from *Salmonella* and *Shigella* types. *J Bacteriol* **52**: 461–466.
- Coleman ML, Chisholm SW. (2007). Code and context: *Prochlorococcus* as a model for cross-scale biology. *Trends Microbiol* **15**: 398–407.
- 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.
- Costa-Ramos C, Rowley AF. (2004). Effect of extracellular products of *Pseudoalteromonas atlantica* on the edible crab *Cancer pagurus*. *Appl Environ Microbiol* **70**: 729–735.
- Coward C, Grant AJ, Swift C, Philp J, Towler R, Heydarian M et al. (2006). Phase-variable surface structures are required for infection of *Campylobacter jejuni* by bacteriophages. *Appl Environ Microbiol* **72**: 4638–4647.
- Cuadros-Orellana S, Martin-Cuadrado A-B, Legault B, DAuria G, Zhaxybayeva O, Thane Papke T et al. (2007). Genomic plasticity in prokaryotes: the case of the square haloarchaeon. *ISME J* **1**: 235–245.
- De la Cruz F, Davies J. (2000). Horizontal gene transfer and the origin of species: lessons from bacteria. *Trends Microbiol* **8**: 128–133.
- Defives C, Werquin M, Mary P, Hornez JP. (1996). Roles of Exopolysaccharides and Lipopolysaccharides in the Adsorption of the Siphovirus Phage NM8 to *Rhizobium meliloti* M11S Cells. *Curr Microbiol* **33**: 371–376.
- Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. (1999). Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* **27**: 4636–4641.
- DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard NU et al. (2006). Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311**: 496–503.
- Dempsey MP, Nietfeldt J, Ravel J, Hinrichs S, Crawford R, Benson AK. (2006). Paired-end sequence mapping detects extensive genomic rearrangement and translocation during divergence of *Francisella tularensis* subsp. *tularensis* and *Francisella tularensis* subsp. *holarctica* populations. *J Bacteriol* **188**: 5904–5914.
- Dobzhansky T. (1970). *Genetics of the evolutionary process*. Columbia University Press: New York.
- Duchaud E, Rusniok C, Frangeul L, Buchrieser C, Givaudan A, Taourit S et al. (2003). The genome sequence of the entomopathogenic bacterium *Photobacterium luminescens*. *Nat Biotechnol* **21**: 1307–1313.
- Edgar RC. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**: 113.
- Eilers H, Perenthaler J, Glockner FO, Amann R. (2000). Culturability and *in situ* abundance of pelagic bacteria from the North Sea. *Appl Environ Microbiol* **66**: 3044–3051.
- Flemming HC, Neu TR, Wozniak DJ. (2007). The EPS matrix: the 'house of biofilm cells'. *J Bacteriol* **189**: 7945–7947.
- Franklin L. (2007). Bacteria, sex, and systematics. *Philos Sci* **74**: 69–95.
- Fraser C, Hanage WP, Spratt BG. (2007). Recombination and the nature of bacterial speciation. *Science* **315**: 476–480.
- Friesen ML, Saxer G, Travisano M, Doebeli M. (2004). Experimental evidence for sympatric ecological diversification due to frequency-dependent competition in *Escherichia coli*. *Evolution Int J Org Evolution* **58**: 245–260.



- Garcia-Martinez J, Acinas SG, Massana R, Rodriguez-Valera F. (2002). Prevalence and microdiversity of *Alteromonas macleodii*-like microorganisms in different oceanic regions. *Environ Microbiol* **4**: 42–50.
- Gardner AM, Gardner PR. (2002). Flavohemoglobin detoxifies nitric oxide in aerobic, but not anaerobic, *Escherichia coli*. Evidence for a novel inducible anaerobic nitric oxide-scavenging activity. *J Biol Chem* **277**: 8166–8171.
- Ghai R, Hain T, Chakraborty T. (2004). GenomeViz: visualizing microbial genomes. *BMC Bioinformatics* **5**: 198.
- 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.
- Glockner FO, Fuchs BM, Amann R. (1999). Bacterioplankton compositions of lakes and oceans: a first comparison based on fluorescence *in situ* hybridization. *Appl Environ Microbiol* **65**: 3721–3726.
- Godde JS, Bickerton A. (2006). The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. *J Mol Evol* **62**: 718–729.
- Goldberg SM, Johnson J, Busam D, Feldblyum T, Ferreira S, Friedman R et al. (2006). A Sanger/pyrosequencing hybrid approach for the generation of high-quality draft assemblies of marine microbial genomes. *Proc Natl Acad Sci USA* **103**: 11240–11245.
- Grossart HP, Kiorboe T, Tang K, Ploug H. (2003). Bacterial colonization of particles: growth and interactions. *Appl Environ Microbiol* **69**: 3500–3509.
- Grossart HP, Tang KW, Kiorboe T, Ploug H. (2007). Comparison of cell-specific activity between free-living and attached bacteria using isolates and natural assemblages. *FEMS Microbiol Lett* **266**: 194–200.
- Hall TA. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* **41**: 95–98.
- Hanage WP, Spratt BG, Turner KM, Fraser C. (2006). Modelling bacterial speciation. *Philos Trans R Soc Lond B Biol Sci* **361**: 2039–2044.
- Hayes F. (2003). Toxins–antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest. *Science* **301**: 1496–1499.
- Hazan R, Sat B, Engelberg-Kulka H. (2004). *Escherichia coli* mazEF-mediated cell death is triggered by various stressful conditions. *J Bacteriol* **186**: 3663–3669.
- Hebel D, Knauer GA, Martin JH. (1986). Trace metals in large agglomerates (marine snow). *J Plankton Res* **8**: 819–824.
- Hochhut B, Wilde C, Balling G, Middendorf B, Dobrindt U, Brzuszkiewicz E et al. (2006). Role of pathogenicity island-associated integrases in the genome plasticity of uropathogenic *Escherichia coli* strain 536. *Mol Microbiol* **61**: 584–595.
- Huang CC, Narita M, Yamagata T, Endo G. (1999). Identification of three merB genes and characterization of a broad-spectrum mercury resistance module encoded by a class II transposon of *Bacillus megaterium* strain MB1. *Gene* **239**: 361–366.
- Ivars-Martinez E, D’Auria G, Rodriguez-Valera F, Sanchez-Porro C, Ventosa A, Joint I et al. (2008). Biogeography of the ubiquitous marine bacterium *Alteromonas macleodii* determined by multi-locus sequence analysis (MLSA). *Mol Ecol* (in press).
- Juty NS, Moshiri F, Merrick M, Anthony C, Hill S. (1997). The *Klebsiella pneumoniae* cytochrome bd’ terminal oxidase complex and its role in microaerobic nitrogen fixation. *Microbiology* **143**(Pt 8): 2673–2683.
- Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. (2004). The KEGG resource for deciphering the genome. *Nucleic Acids Res* **32** (Database issue): D277–D280.
- Karl DM, Knauer GA, Martin JH. (1988). Downward flux of particulate organic matter in the ocean: a particle decomposition paradox. *Nature* **332**: 438–441.
- Kettler GC, Martiny AC, Huang K, Zucker J, Coleman ML, Rodrigue S et al. (2007). Patterns and implications of gene gain and loss in the evolution of prochlorococcus. *PLoS Genet* **3**: e231.
- Kiyono M, Omura T, Inuzuka M, Fujimori H, Pan-Hou H. (1997). Nucleotide sequence and expression of the organomercurial-resistance determinants from a *Pseudomonas* K-62 plasmid pMR26. *Gene* **189**: 151–157.
- Kolodkin-Gal I, Engelberg-Kulka H. (2006). Induction of *Escherichia coli* chromosomal mazEF by stressful conditions causes an irreversible loss of viability. *J Bacteriol* **188**: 3420–3423.
- Kolodkin-Gal I, Hazan R, Gaathon A, Carmeli S, Engelberg-Kulka H. (2007). A linear pentapeptide is a quorum-sensing factor required for mazEF-mediated cell death in *Escherichia coli*. *Science* **318**: 652–655.
- Konstantinidis KT, Tiedje JM. (2005). Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci USA* **102**: 2567–2572.
- Kunin V, He S, Warnecke F, Peterson SB, Garcia Martin H, Haynes M et al. (2008). A bacterial metapopulation adapts locally to phage predation despite global dispersal. *Genome Res* **18**: 293–297.
- Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C et al. (2004). Versatile and open software for comparing large genomes. *Genome Biol* **5**: R12.
- Lauro FM, Bartlett DH. (2008). Prokaryotic lifestyles in deep sea habitats. *Extremophiles* **12**: 15–25.
- Lauro FM, Chastain RA, Blankenship LE, Yayanos AA, Bartlett DH. (2007). The unique 16S rRNA genes of piezophiles reflect both phylogeny and adaptation. *Appl Environ Microbiol* **73**: 838–845.
- Lauro FM, Tran K, Vezzi A, Vitulo N, Valle G, Bartlett DH. (2008). Large-scale transposon mutagenesis of *Photobacterium profundum* SS9 reveals new genetic loci important for growth at low temperature and high pressure. *J Bacteriol* **190**: 1699–1709.
- Legault BA, Lopez-Lopez A, Alba-Casado JC, Doolittle WF, Bolhuis H, Rodriguez-Valera F et al. (2006). Environmental genomics of ‘*Haloquadratum walsbyi*’ in a saltern crystallizer indicates a large pool of accessory genes in an otherwise coherent species. *BMC Genomics* **7**: 171.
- Lenz O, Friedrich B. (1998). A novel multicomponent regulatory system mediates H<sub>2</sub> sensing in *Alcaligenes eutrophus*. *Proc Natl Acad Sci USA* **95**: 12474–12479.
- Lopez-Lopez A, Bartual SG, Stal L, Onyshchenko O, Rodriguez-Valera F. (2005). Genetic analysis of house-keeping genes reveals a deep-sea ecotype of *Alteromonas macleodii* in the Mediterranean Sea. *Environ Microbiol* **7**: 649–659.
- Lowe TM, Eddy SR. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**: 955–964.
- Martin-Cuadrado AB, Lopez-Garcia P, Alba JC, Moreira D, Monticelli L, Strittmatter A et al. (2007).

- Metagenomics of the deep Mediterranean, a warm bathypelagic habitat. *PLoS ONE* **2**: e914.
- Martin-Cuadrado AB, Rodriguez-Valera F, Moreira D, Alba JC, Ivars-Martinez E, Henn MR *et al.* (2008). Hindsight in the relative abundance, metabolic potential and genome dynamics of uncultivated marine *archaea* from comparative metagenomic analyses of bathypelagic plankton of different oceanic regions. *The ISME J* (advance online publication, 8 May 2008; doi:10.1038/ismej.2008.40).
- Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP, Flanigan MJ *et al.* (2000). A whole-genome assembly of *Drosophila*. *Science* **287**: 2196–2204.
- Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M *et al.* (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* **33**: 5691–5702.
- Pernthaler A, Pernthaler J, Eilers H, Amann R. (2001). Growth patterns of two marine isolates: adaptations to substrate patchiness? *Appl Environ Microbiol* **67**: 4077–4083.
- Petrosino JF, Xiang Q, Karpthy SE, Jiang H, Yerrapragada S, Liu Y *et al.* (2006). Chromosome rearrangement and diversification of *Francisella tularensis* revealed by the type B (OSU18) genome sequence. *J Bacteriol* **188**: 6977–6985.
- Pomeroy LRDD. (1980). Aggregation of organic matter by pelagic tunicates. *Limnol Oceanogr* **25**: 643–652.
- Pukall R, Buntfuss D, Fruhling A, Rohde M, Kroppenstedt RM, Burghardt J *et al.* (1999). *Sulfitobacter mediterraneus* sp. nov, a new sulfite-oxidizing member of the alpha-proteobacteria. *Int J Syst Bacteriol* **49**(Pt 2): 513–519.
- Reva O, Tummeler B. (2008). Think big—giant genes in bacteria. *Environ Microbiol* **10**: 768–777.
- Rice P, Longden I, Bleasby A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **16**: 276–277.
- Richerson P, Armstrong R, Goldman CR. (1970). Contemporaneous disequilibrium, a new hypothesis to explain the ‘paradox of the plankton’. *Proc Natl Acad Sci USA* **67**: 1710–1714.
- Riemann L, Grossart HP. (2008). Elevated Lytic Phage Production as a Consequence of Particle Colonization by a Marine Flavobacterium (*Cellulophaga* sp.). *Microb Ecol* (advance online publication, 18 March 2008; doi:10.1007/s00248-008-9369-8).
- Rougeaux H, Talaga P, Carlson RW, Guenzennec J. (1998). Structural studies of an exopolysaccharide produced by *Alteromonas macleodii* subsp. *fijiensis* originating from a deep-sea hydrothermal vent. *Carbohydr Res* **312**: 53–59.
- Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S *et al.* (2007). The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biol* **5**: e77.
- Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA *et al.* (2000). Artemis: sequence visualization and annotation. *Bioinformatics* **16**: 944–945.
- Samuel G, Reeves P. (2003). Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydr Res* **338**: 2503–2519.
- Sass AM, Sass H, Coolen MJ, Cypionka H, Overmann J. (2001). Microbial communities in the chemocline of hypersaline deep-sea basin (Uranian Basin, Mediterranean Sea). *Appl Environ Microbiol* **67**: 3083–3091.
- Schafer H, Servais P, Muyzer G. (2000). Successional changes in the genetic diversity of a marine bacterial assemblage during confinement. *Arch Microbiol* **173**: 138–145.
- Shanks AL, Reeder ML. (1993). Reducing microzones and sulphide production in marine snow. *Mar Ecol Prog Ser* **96**: 43–47.
- Sharma RS, Mishra V, Mohammed A, Babu CR. (2008). Phage specificity and lipopolysaccharides of stem- and root-nodulating bacteria (*Azorhizobium caulinodans*, *Sinorhizobium* spp., and *Rhizobium* spp.) of *Sesbania* spp. *Arch Microbiol* **189**: 411–418.
- Silver S, Phung LT. (1996). Bacterial heavy metal resistance: new surprises. *Annu Rev Microbiol* **50**: 753–789.
- Simon M, Grossart HP, Schweitzer B, Plough H. (2002). Microbial ecology of organic aggregates in aquatic ecosystems. *Aquat Microb Ecol* **28**: 175–211.
- Simonato F, Campanaro S, Lauro FM, Vezzi A, D’Angelo M, Vitulo N *et al.* (2006). Piezophilic adaptation: a genomic point of view. *J Biotechnol* **126**: 11–25.
- Skerman VBD. (1967). *A Guide to the Identification of the Genera Bacteria*, 2nd edn. Williams and Willkins: Baltimore, USA.
- Sorek R, Kunin V, Hugenholtz P. (2008). CRISPR—a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat Rev Microbiol* **6**: 181–186.
- Spencer CC, Bertrand M, Travisano M, Doebeli M. (2007). Adaptive diversification in genes that regulate resource use in *Escherichia coli*. *PLoS Genet* **3**: e15.
- Sturr MG, Krulwich TA, Hicks DB. (1996). Purification of a cytochrome bd terminal oxidase encoded by the *Escherichia coli* app locus from a delta cyo delta cyd strain complemented by genes from *Bacillus firmus* OF4. *J Bacteriol* **178**: 1742–1749.
- Tamura K, Dudley J, Nei M, Kumar S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**: 1596–1599.
- Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS *et al.* (2001). The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res* **29**: 22–28.
- Tettelin H, Massignani 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.
- Thompson JD, Higgins DG, Gibson TJ. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680.
- Ting CS, Rocap G, King J, Chisholm SW. (2002). Cyanobacterial photosynthesis in the oceans: the origins and significance of divergent light-harvesting strategies. *Trends Microbiol* **10**: 134–142.
- Tyerman J, Havard N, Saxer G, Travisano M, Doebeli M. (2005). Unparallel diversification in bacterial microcosms. *Proc Biol Sci* **272**: 1393–1398.
- Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA *et al.* (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**: 66–74.

- Vezi A, Campanaro S, D'Angelo M, Simonato F, Vitulo N, Lauro FM *et al.* (2005). Life at depth: *Photobacterium profundum* genome sequence and expression analysis. *Science* **307**: 1459–1461.
- Vignais PM, Billoud B. (2007). Occurrence, classification, and biological function of hydrogenases: an overview. *Chem Rev* **107**: 4206–4272.
- Willenbrock H, Petersen A, Sekse C, Kiil K, Wasteson Y, Ussery DW. (2006). Design of a seven-genome *Escherichia coli* microarray for comparative genomic profiling. *J Bacteriol* **188**: 7713–7721.
- Xu K, Ma BG. (2007). Comparative analysis of predicted gene expression among deep-sea genomes. *Gene* **397**: 136–142.
- Zaballos M, Lopez-Lopez A, Ovreas L, Bartual SG, D'Auria G, Alba JC *et al.* (2006). Comparison of prokaryotic diversity at offshore oceanic locations reveals a different microbiota in the Mediterranean Sea. *FEMS Microbiol Ecol* **56**: 389–405.
- Zhaxybayeva O, Lapierre P, Gogarten JP. (2004). Genome mosaicism and organismal lineages. *Trends Genet* **20**: 254–260.
- Zhilentkov EL, Popova VM, Popov DV, Zaval'sky LY, Svetoch EA, Stern NJ *et al.* (2006). The ability of flagellum-specific *Proteus vulgaris* bacteriophage PV22 to interact with *Campylobacter jejuni* flagella in culture. *Virol J* **3**: 50.

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