

ORIGINAL ARTICLE

Proteomic insights into the lifestyle of an environmentally relevant marine bacterium

Joseph Alexander Christie-Oleza¹, Bernard Fernandez¹, Balbina Nogales², Rafael Bosch² and Jean Armengaud¹

¹Laboratoire de Biochimie des Systèmes Perturbés, CEA Marcoule, DSV, IBEB, Bagnols-sur-Cèze, France and ²Departament de Biologia, Microbiologia, Universitat de les Illes Balears, Palma de Mallorca, Spain

In terms of lifestyle, free-living bacteria are classified as either oligotrophic/specialist or opportunist/generalist. Heterogeneous marine environments such as coastal waters favour the establishment of marine generalist bacteria, which code for a large pool of functions. This is basically foreseen to cope with the heterogeneity of organic matter supplied to these systems. Nevertheless, it is not known what fraction of a generalist proteome is needed for house-keeping functions or what fraction is modified to cope with environmental changes. Here, we used high-throughput proteomics to define the proteome of *Ruegeria pomeroyi* DSS-3, a model marine generalist bacterium of the *Roseobacter* clade. We evaluated its genome expression under several natural environmental conditions, revealing the versatility of the bacterium to adapt to anthropogenic influence, poor nutrient concentrations or the presence of the natural microbial community. We also assayed 30 different laboratory incubations to increase proteome coverage and to dig further into the functional genomics of the bacterium. We established its core proteome and the proteome devoted to adaptation to general cellular physiological variations (almost 50%). We suggest that the other half of its theoretical proteome is the opportunist genetic pool devoted exclusively to very specific environmental conditions.

The ISME Journal (2012) 6, 124–135; doi:10.1038/ismej.2011.86; published online 21 July 2011

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

Keywords: proteomics; core proteome; specialist; generalist; *Roseobacters*; functional genomics

Introduction

Lifestyle dichotomies such as specialist/generalist or oligotroph/opportunitroph have been proposed to simplify the classification of microorganisms based on the adaptive strategies they apply in marine environments (Polz *et al.*, 2006; Newton *et al.*, 2010; Singer *et al.*, 2011). Although specialist or oligotroph terms refer to microorganisms adapted to use low nutrient concentrations efficiently, generalists or opportunitrophs are predicted to be versatile microorganisms that are able to exploit variable resources spatially and temporally (Polz *et al.*, 2006). Nevertheless, the genome usage of these organisms has not yet been clearly established. Oligotrophic oceanic waters are characterised by being largely colonised by specialised oligotrophic organisms, for example, the SAR11 cluster (Biers *et al.*, 2009). A cultured representative of this group, *Candidatus pelagibacter ubique*, is a free-living microorganism with one of the smallest genomes

known to date (Giovannoni *et al.*, 2005). Despite this, it has the ability to use several different energy and carbon sources typically found in its environment. On the other hand, in surface and coastal marine environmental niches, a higher predominance of bacteria with a generalist-like lifestyle is generally found, owing to the higher concentration of nutrients (Lauro *et al.*, 2009) and greater heterogeneity of dissolved organic carbon in these waters (Mou *et al.*, 2008). In agreement with the generalist lifestyle, microorganisms in these environments show no specialisation to a specific metabolic compound (Mou *et al.*, 2008). Instead, bacteria adopt the strategy of accumulating a large pool of catabolic functions to face the diversity of dissolved organic carbon. To understand the functioning of marine communities and their impact on global ecosystem processes, numerous studies have been performed (reviewed in Strom, 2008; DeLong, 2009; Fuhrman, 2009). In order to help with the interpretation of marine metagenomic data and the physiology of relevant microbial groups, more than 150 marine bacterial genomes are currently being sequenced (<http://www.moore.org>). Despite this, little is known about what fraction of the encoded proteome has an impact on global ecosystem processes. Metaproteomics is becoming a key tool in revealing these functional profiles and determining

Correspondence: J Armengaud, Laboratoire de Biochimie des Systèmes Perturbés, CEA Marcoule, DSV, IBEB, SBTN, LBSP, F-30207 Bagnols-sur-Cèze, France.

E-mail: jean.armengaud@cea.fr

Received 29 March 2011; revised 30 May 2011; accepted 30 May 2011; published online 21 July 2011

the roles of each component in complex microbial communities (Wilmes and Bond, 2006, 2009; Keller and Hettich, 2009; VerBerkmoes *et al.*, 2009).

The *Roseobacter* clade is a key component of surface marine ecosystems, as over 10% of bacterial cells in the open ocean and 20% in coastal waters are members of this group (Buchan *et al.*, 2005). Their large inventory of encoded functions, in order to obtain carbon and energy, has established *Roseobacters* as generalists (Newton *et al.*, 2010). Isolates show an enormously versatile metabolism, having important roles in the global carbon and sulphur cycles, and potentially modifying the climate, as they process important greenhouse gases such as carbon monoxide and dimethylsulphide (Wagner-Dobler and Biebl, 2006; Newton *et al.*, 2010). *Ruegeria pomeroyi* DSS-3 (formerly *Silicibacter pomeroyi*) was the first member of the *Roseobacter* clade to have its genome completely sequenced and annotated (Moran *et al.*, 2004). Inspection of the genome of *R. pomeroyi* revealed a wide gene pool, which is supposed to be helpful for adaptation to variations in the marine environment (Moran *et al.*, 2004; Newton *et al.*, 2010). Its genome comprises a 4.11-Mbp chromosome and a 0.49-Mbp megaplasmid. This genome size is consistent with a generalist free-living bacterial lifestyle with a large number of regulatory and signal transduction proteins (Koonin and Wolf, 2008). *R. pomeroyi* has become a model organism for studies of the ecological and physiological strategies of heterotrophic marine bacteria (Burgmann *et al.*, 2007).

The large number of sequenced marine bacterial genomes has led to several comparative genomic studies (Moran *et al.*, 2007; Scanlan *et al.*, 2009). It is known that oligotrophic marine bacteria present 1.5–2 Mbp-sized genomes, whereas generalist genomes are much larger. Nevertheless, there is no experimental evidence of the fraction of their proteome needed for living in the different ecological niches found in the oceans. Interestingly, we have noted that genome-wide proteome analyses always show high coverage in oligotroph-like bacteria, for example, 88% coverage of the 777-Mbp minimal organism *Mycoplasma mobile* (Jaffe *et al.*, 2004), but much lower values when generalists are scrutinised, for example, 40%, 54% and 53% of the genes encoded by the 5.13-Mbp genome of *Shewanella oneidensis*, 3.4-Mbp genome of *Sphingopyxis alaskensis* and 3.9-Mbp genome of *Synechocystis* sp. PCC 6803, respectively (Gupta *et al.*, 2008; Ting *et al.*, 2009; Wegener *et al.*, 2010). In the present study, we carried out a comprehensive proteomic analysis on a member of the marine *Roseobacter* clade that will give an overview of the functional genomics in this kind of organism. The proteomic patterns obtained after subjecting *R. pomeroyi* to different natural conditions showed only small variations, although these highlighted key proteins used by the bacterium in its environment. The extra 30 different physiological conditions tested enabled

higher coverage of the organism's theoretical proteome. By comparing the high-throughput nano-liquid chromatography–mass spectrometry (LC-MS/MS) proteomic data obtained throughout this study, we observed a large number of proteins detected under all conditions and those set for general physiological changes in the cell. Nevertheless, a larger number of proteins (over 50% of the theoretical proteome) were not detected, suggesting that this fraction is probably dedicated exclusively to allow adaptation to very specific physiological conditions or a generalist lifestyle.

Materials and methods

R. pomeroyi culture conditions

R. pomeroyi DSS-3 (DSM15171) was grown on marine agar plates (DIFCO/Becton Dickinson, Le Pont de Claix, France) before inoculating liquid cultures. This was performed using either marine broth (MB, Difco) or minimal marine medium (Lanfranconi *et al.*, 2010). Different Marseille (France) coastal winter surface seawater samples were obtained to simulate natural conditions (2011-02-16): Bassin d'Arcenc (ferry basin, 43.32567719N–5.35845279E); Vieux Port (marina, 43.29532433N–5.36682128E); Parc Borély (beach, 43.25770573N–5.37553310E); and Cap Croisette (cape, 43.21496607N–5.33602952E). The four different samples were processed immediately as follows: duplicated 750-ml quantities of natural seawater (presence of the microbial community) or seawater filtered through a 0.22- μ m-diameter pore filter (absence of the natural community) were placed in 1-l Erlenmeyer flasks. Each Erlenmeyer flask was inoculated with MB-pre-grown *R. pomeroyi* cells at a final concentration of 10^5 cells ml⁻¹. To avoid any addition of dissolved organic carbon to the seawater milieu, *R. pomeroyi* cells were washed twice with minimal marine medium before inoculation. The resulting 16 flasks (four different origins with the presence/absence of the natural microbial community, all duplicated) were agitated at 200 r.p.m. for 20 h at 25 °C. Similarly, community samples (without *R. pomeroyi* inoculum) were processed in order to assay for blanks, giving as a result no significant MS/MS-detected protein. For laboratory stress conditions, cells were grown and treated as shown in Table 1 and Supplementary Information. Whatever the conditions, cells were harvested by centrifugation for 10 min at 6000 g at 4 °C. Pelleted cells were washed twice with cold 20 mM Tris-HCl buffered at pH 8.0. Pellets were frozen at –80 °C until protein extraction took place. Protein extraction, SDS-PAGE and proteome fractionation are detailed in the Supplementary Information. Short electrophoresis migrations were carried out in order to detect the most abundant proteins and compare the different conditions assayed. To detect the lower abundant proteins, we simplified the bacterial protein mix with long electrophoresis migrations.

Table 1 Laboratory growth and stress conditions applied to *Ruegeria pomeroyi*

Sample	Duplicated	Medium	Incubation/stress ^a	Growth stage ^b	Other details
C01	A and B	MB		Exponential phase	1 day until peak growth
C02		MB		Late stationary phase	1 day until peak growth
C03	A and B	MMM YE succinate		Exponential phase	2 day until peak growth
C04		MMM YE succinate		Late stationary phase	2 day until peak growth
C05		MA	Cells grown on agar plates (3 days)	Stationary phase	
C06	A and B	MA	Cells grown on agar plates (10 days)	Late stationary phase	
C07		MB	Microaerobic conditions	Early stationary phase	5 days until peak growth (OD ₆₀₀ 0.4)
C08		MB	40 °C	Mid stationary phase	4 days until peak growth
C09		MB	12 °C	Early stationary phase	3 days until peak growth
C10		MB	7.5% sea salts	Mid stationary phase	4 days until peak growth
C12		MB	UV stress (15 s)	Mid stationary phase	2% CFU survival
C13		MB	UV stress (1 min)	Mid stationary phase	0,2% CFU survival
C15	A and B	MB	UV stress (5 min)	Mid stationary phase	0,02% CFU survival
C16		MB grown cells ^c	Pelleted cells incubated at 26 °C	Stationary phase	1 day incubation
C17		MB grown cells ^c	Pelleted cells incubated at 4 °C	Stationary phase	1 day incubation
C18		MB	0.2% sea salts	Early stationary phase	5 days until peak growth
C19		MB grown cells ^c	Incubated in Tris-HCl 20 mM pH 8.2	Stationary phase	1 day incubation
C21		MMM succinate	No YE added	Mid stationary phase	4 days until peak growth (OD ₆₀₀ 0.2)
C25		MB grown cells ^c	Incubated in autoclaved SW	Stationary phase	1 day incubation
C28		SW succinate		Mid stationary phase	4 days until peak growth
C31		MB-grown cells ^c	Incubated in MMM naphthalene 0.5%	Stationary phase	1 day incubation
C32		MB	Diesel 0.3 ml l ⁻¹	Early stationary phase	1 day until peak growth
C33		MB	Diesel 3 ml l ⁻¹	Early stationary phase	1 day until peak growth
C34		MB grown cells ^c	High cell concentration (10 ¹¹ CFU ml ⁻¹) ^d	Stationary phase	1 day incubation in high density cell suspension
C35		MB	Diesel 30 ml l ⁻¹	Early stationary phase	1 day until peak growth
C36		MMM YE Benzoate		Mid stationary phase	4 day until peak growth
C38		MB	Naphthalene 0.5%	Mid stationary phase	4 day until peak growth (brown culture)
C41		MB	UV stress (15 s)	Early stationary phase	3% CFU survival
C43		MB	Heat shock (42 °C; 30 min)	Early stationary phase	30% CFU survival
C46		MB	Cold shock (-20 °C until freezing)	Early stationary phase	80% CFU survival

Abbreviations: MA, marine agar; MB, marine broth; MMM, minimal marine media; SW, sea water; YE, yeast extract.

^aIf not mentioned incubation/stress was carried out at 26 °C under agitation at 180 r.p.m. *R. pomeroyi* is unable to grow using naphthalene or diesel as unique sources of energy and carbon, hence their inclusion in this column.

^bExponential phase (OD₆₀₀ 0.7), early stationary phase (when culture reached peak growth), mid and late stationary phase (1 and 3 days, respectively, after the culture reached stationary phase).

^c'MB grown cells' is indicated under those conditions where *R. pomeroyi* was not able to grow due to the lack of a useful source of energy and carbon. Cells had to be inoculated in sufficient number for the MS/MS analysis after the incubation/stress.

^dA 100 ml grown culture was centrifuged and the pellet was resuspended in 1 ml of the remaining MB supernatant.

Trypsin in-gel proteolysis, nano-LC-MS/MS, database searches and spectral count

Polyacrylamide gel bands were cut and processed for in-gel proteolysis with trypsin (Roche, Meylan, France) following the ProteaseMAX protocol (Promega, Charbonnières-les-Bains, France) as described (Clair *et al.*, 2010). Nano-LC-MS/MS experiments were performed using the LTQ-Orbitrap XL hybrid mass spectrometer (ThermoFisher, Courtaboeuf, France) coupled to an UltiMate 3000 LC system (Dionex-LC Packings, Voisins le Bretonneux, France) as previously described (de Groot *et al.*, 2009; Baudet *et al.*, 2010). Database and peak list generation, as well as the parameters for MS/MS database, are described in the Supplementary Information searches. Protein semi-quantitation by spectral abundance took place as previously described (Liu *et al.*, 2004). For determining the normalised spectral abundance factors of each protein, spectral counts assigned to each polypep-

ptide were divided by its molecular weight (Paoletti *et al.*, 2006). Values were then normalised by the total sum corresponding to all the polypeptides detected with two or more non-redundant peptides. Clustering and sequence analysis is detailed in the Supplementary Information. The statistical protein variation among the duplicated samples of the natural conditions was calculated with the ACFold option of the PatternLab software (Carvalho *et al.*, 2008). This module combines average fold changes with statistics (AC-test) and a theoretical false-positive estimator to properly select the differentially expressed proteins. We chose stringent parameters: significance threshold at $P < 0.01$ and fold change > 2.0 . Protein levels detected in cells subjected to laboratory conditions (without duplicates) were compared with the mean levels of these proteins considering all the conditions. In this case, only fold changes > 4.0 were considered.

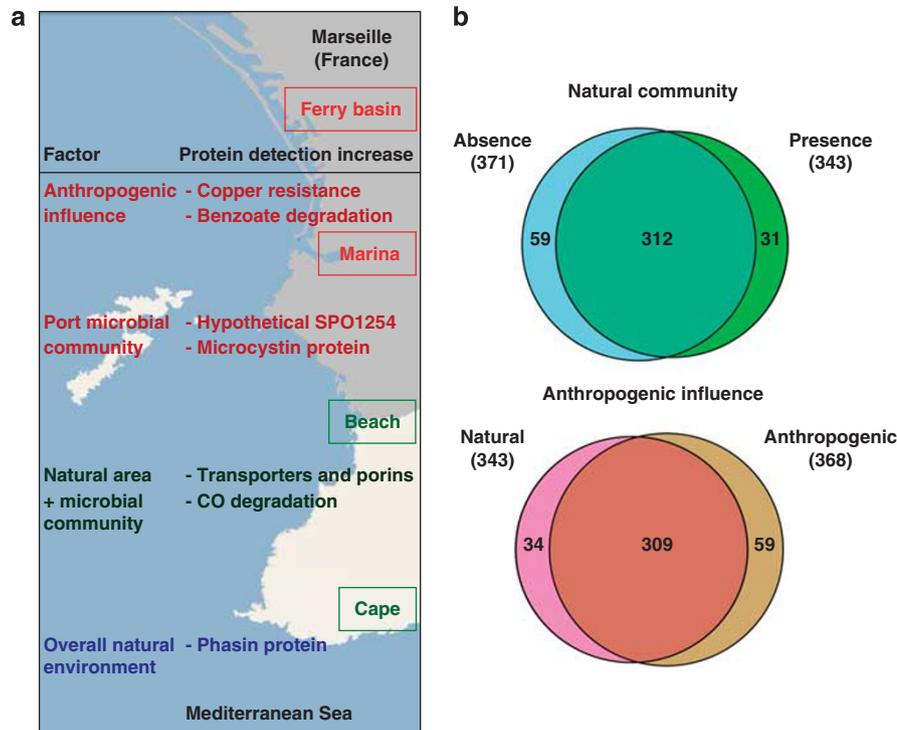


Figure 1 Polypeptides detected from *R. pomeroyi* cells incubated under different natural conditions. (a) Sampling areas where the seawater for natural-condition simulation was obtained. Anthropogenically perturbed and natural areas are in red and green, respectively. The most statistically differing proteins detected in *R. pomeroyi* in each environmental condition compared with other conditions are indicated (statistical analysis shown in Supplementary Table S2). ‘Anthropogenic influence’ represents the main protein differences between the ferry basin and marina to cape conditions (in the presence and absence of the natural community); ‘Port microbial community’ compares the overproduced proteins of *R. pomeroyi* in the presence of the natural community in both port systems; ‘Natural area + microbial community’ highlights those proteins overproduced in the cape and beach conditions with the natural community when compared with the port systems; ‘Overall natural environment’ compares the natural conditions tested (polypeptide list of Supplementary Table S1) with the laboratory conditions assayed (polypeptide list of Supplementary Table S3). (b) Both Venn diagrams show the number of shared and specific polypeptides detected in the absence and presence of the natural community (top) and based on the anthropogenic influence (bottom). For this, the data were grouped and compared as follows: the four sampling areas in the absence (blue) versus presence (green) of the natural community; natural (beach and cape in the presence/absence of the natural community, pink) versus anthropogenically perturbed (marina and ferry basin in the presence/absence of the natural community, brown) seawater areas. In brackets are the total non-redundant polypeptides corresponding to each group.

Results

Proteome variations of R. pomeroyi in its natural coastal environment

Cells of *R. pomeroyi* were incubated in natural seawater taken from four different sampling areas (open cape, beach, ferry basin and marina) in the presence or absence of the natural microbial community (Figure 1a). *R. pomeroyi* cells were inoculated at a concentration of 10^5 cells ml⁻¹. Cells from each of the duplicated culture conditions were harvested 20 h later. Their protein contents were analysed by means of shotgun nano-LC-MS/MS proteomic analysis in order to detect only the more abundant proteins of *R. pomeroyi* (Supplementary Table S1). We defined the abundant proteins as those accounting for at least 50% of the total quantities (here estimated with spectral count) detected in an exhaustive proteomic analysis (when considering the most minimal set). Figure 1b shows that the presence of the microbial community or the origin of the seawater does not result in a strong

variation in the detected abundant-protein pattern of *R. pomeroyi*. Nevertheless, the number of proteins with a significant variation is slightly higher when the seawater origin is considered (Supplementary Table S2). It is noteworthy that no significant false-positive attributions were obtained from the natural community samples (blank samples detailed in the Materials and methods section).

Protein profiling in the presence/absence of the microbial natural communities. In the natural seawater samples (cape and beach), where microbial density and nutrient concentration are expected to be lower, we found no significant protein variation when *R. pomeroyi* was incubated in the presence or absence of the natural community ($P < 0.01$, Supplementary Table S2, panels A and B). Nevertheless, we noted that transporters and porins were possibly more abundant in cape water when the natural community is present (fold change 2–5 ×), but with a low statistical confidence (P -value < 0.06), plausibly due to competition for nutrients. Meanwhile

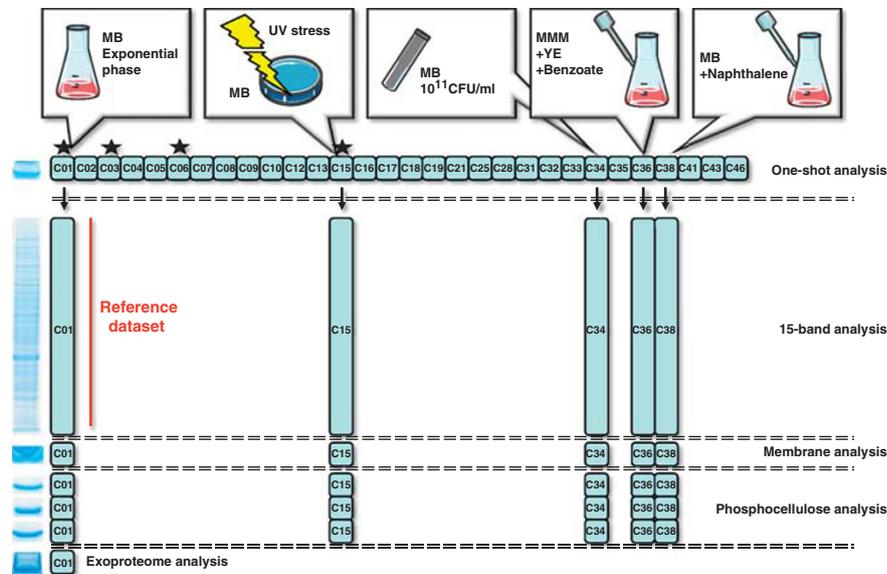


Figure 2 Strategy used to obtain the broadest proteome coverage for *R. pomeroyi*. An image of the Coomassie-stained SDS-PAGE gel is shown for the reference data set (far left) to illustrate the extent of migration. A star indicates those conditions for which biological and technical duplicates were performed. Conditions are numbered from C01 to C46 as indicated in Table 1.

this did not happen in the anthropogenically perturbed seawater samples (ferry basin and marina), where transporters are less detected with the presence of the natural community (Supplementary Table S2, panels C and D). Two proteins were significantly more abundant in the presence of port communities: the hypothetical orfan protein SPO1254 (fold change 16 ×) and the microcystin-dependent protein SPO0483 (10 ×) that presents a conserved phage tail collar domain, which could be involved in pathogenicity or plant–microbe interactions.

Protein profiling between the different sampling areas. Beach, ferry basin and marina samples were compared with the cape samples in the absence (Supplementary Table S2, panels E–G) and presence of the natural community (Supplementary Table S2, panels H–J). The plausible higher concentration of nutrients in the anthropogenically perturbed seawater samples may be the reason for the observed decrease in the abundance of many transporters in these conditions. A strong increase (between 12–35 ×) in copper resistance proteins (SPOA0369 and SPOA0370) was evidenced in both port samples. This could be a consequence of the environmental use of copper antifouling paints for boat hulls. The benzoyl-coA oxygenase (SPO3701) detected in the marina samples (4 ×) was also remarkable and probably illustrates anthropogenic influence in these environments. We also found that *R. pomeroyi* has a lithotrophic lifestyle, where dissolved organic carbon is possibly less abundant (that is, cape seawater samples), through the detection of carbon monoxide degradation determinants. Another interesting feature of *R. pomeroyi* in its

natural environment is the production of phasin, a protein involved in the accumulation of polyoxoesters (York *et al.*, 2001). This protein is not usually detected under laboratory conditions (see below).

Proteomic-based screening for the most divergent laboratory conditions

Figure 2 shows the strategy followed to determine the most divergent physiological conditions for *R. pomeroyi* in terms of its proteome. Briefly, we assayed the most abundant proteins present in cells grown under 30 different laboratory conditions by means of short-migrated 1D-gel stacking. As shown in Table 1, we tried out exponential and stationary culture phases in rich and poor media, with different incubation parameters (salt concentration, temperature, microaerobiosis, pelleted cells and plate growth) and aggressive stresses (UV illumination, the presence of aromatic compounds). Biological (A and B) and technical (B1 and B2) duplicates were performed for four conditions to estimate the variability of our experimental procedure. A total of 878 non-redundant proteins were catalogued by their abundance in the whole screening (Supplementary Table S3). A hierarchical cluster analysis revealed the relationships between the protein profiles from the 30 conditions assayed (Figure 3). As expected, the 8 samples that corresponded to either technical or biological replicates grouped together. Furthermore, we chose the five growth conditions that generated the most divergent proteomes for an in-depth analysis (Figures 2 and 3). We resolved the total extracted proteins on a longer-gel electrophoresis run and analysed 15 fractions. We also enriched lower abundant proteins such as

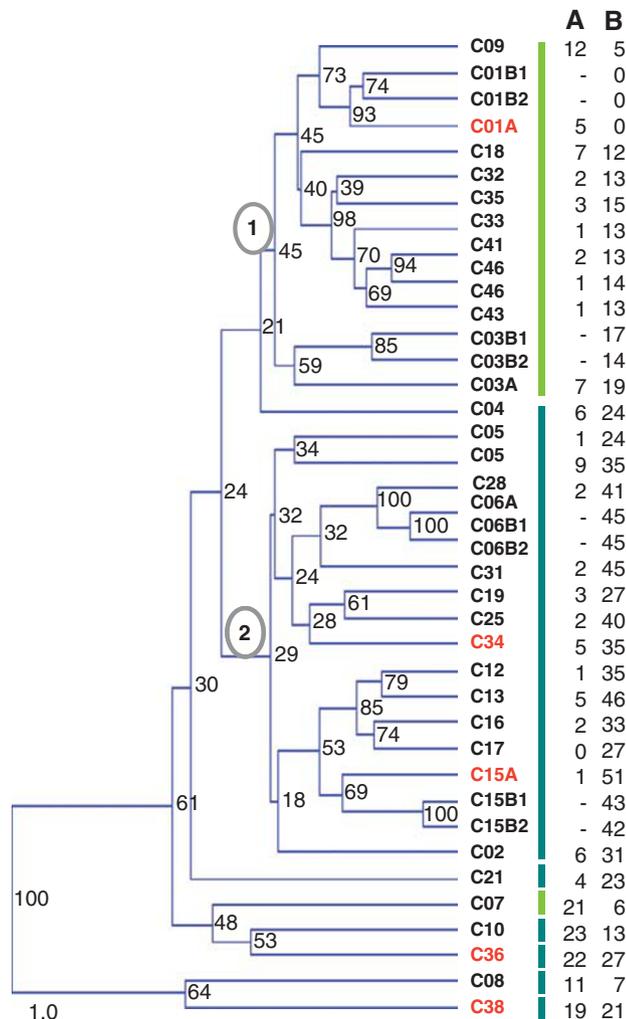


Figure 3 Hierarchical clustering of the protein profiles corresponding to 30 different growth conditions and 8 replicates based on MS/MS one-shot analysis. The tree was produced using UPGMA and Euclidean distance analysis. Bootstrap values (1000 replicates) are indicated. Clusters 1 and 2 indicate protein profiles grouped according to growth phase, exponential–early stationary-phase and mid–late stationary-phase, respectively. Light green lines indicate exponential–early stationary-phase cultures, while dark green lines indicate mid–late stationary-phase cultures. The number of detected and validated polypeptides that are unique to each condition compared with all the others is given in column A. Column B represents the polypeptides not seen in the reference dataset (C01 15-band analysis).

phosphate-binding proteins (phosphocellulose P11 chromatography) and membrane proteins, and included data from its exoproteome (Christie-Oleza and Armengaud, 2010). Supplementary Table S4 compiles all the non-redundant peptides and proteins detected by MS/MS under the five conditions. From the 831 851 spectra recorded in the whole study, a total of 1963 polypeptides (Supplementary Table S5) were detected with high confidence (global false positive rate of 1.2%) covering 46% of the theoretical *R. pomeroyi* proteome (4252 annotated CDS).

Proteomic trends among the different physiological conditions

General protein pattern detected between the physiological conditions tested. Ribosomal proteins represented over 26% of the normalised spectral abundance factor for the total proteins detected from cells obtained from exponential and early stationary phase conditions, but only 14% in those harvested at the mid–late stationary-phase. The abundance of these proteins is directly related to cellular activity. For this reason, protein profile grouping related to the culture growth phase (exponential–early stationary (1) versus mid–late stationary (2) phases in Figure 3) prevails over the other parameters. Two main branches, which group 24 of the 30 conditions, clearly correspond to the distinct growth phases. The profile of condition C04 is the only stationary-phase growth that groups slightly with the exponential phase cluster, probably owing to the similarities to condition C03 and to a plausible metabolic ‘stand-by’ effect already reported in other *Roseobacter* members (Zech *et al.*, 2009). This adaptation strategy is foreseen to maintain central carbon catabolism unchanged during the initial transition from exponential to stationary growth phase. In contrast, the presence of the aromatic compound naphthalene (C38), high salt concentrations (C10), O₂ limitation (C07), high temperatures (C08) or growth using benzoate (C36), are conditions that present the most protein differences. In these cases, proteins abundantly produced to cope with the corresponding stresses prevail over the protein pattern changes due to culture phase transition.

We noted the presence of the hypothetical protein SPOA0343 among the most detected proteins under almost all the growth conditions tested, except those grouping outside clusters 1 and 2 (Supplementary Table S3). Figure 4 shows the levels of detection for the proteins coded by the SPOA0339–SPOA0340–SPOA0341–SPOA0342–SPOA0343 genes, which form a probable operon on the pDSS-3 megaplasmid. The products of the two genes SPOA0342 and SPOA0341 (annotated as hypothetical genes), and of SPOA0340 and SPOA0339 (with a functional annotation as a putative tautomerase and a HAD family hydrolase, respectively) were also detected. As SPOA0342 is a distantly related homologue of the AfsA protein from *Streptomyces griseus* (Kato *et al.*, 2007) involved in A factor biosynthesis, we may infer a role in microbial hormone biosynthesis for this plausible operon. Such a role is in agreement with the fact that detection of these five proteins strongly decreases under adverse stress conditions.

Specific proteins overproduced under the different physiological conditions tested. We observed some differences according to whether the cells had to face a sudden heat shock or whether they were allowed a longer adaptation time to high temperatures. On the one hand, under conditions C08 and C43 (growth at 40 °C and heat shock at 42 °C,

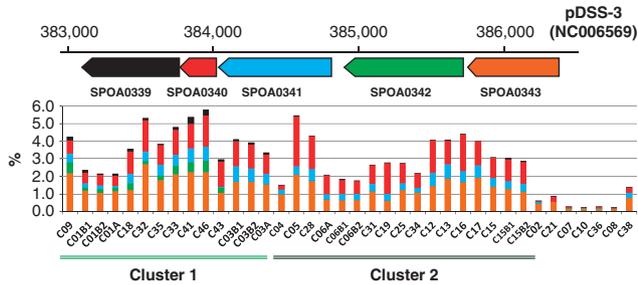


Figure 4 Detection of proteins SPOA0339–SPOA0340–SPOA0341–SPOA0342–SPOA0343 among the different physiological conditions tested. The putative operon structure on the pDSS-3 megaplasmid is shown. The stacked bar graph shows the accumulation of normalised spectral abundance factor values for the five detected proteins under each of the 30 conditions and eight replicates. Conditions are sorted as in the hierarchical clustering shown in Figure 3. Clusters 1 and 2 are indicated.

respectively) we observed a response probably to avoid protein denaturation and aggregation that comprises the ClpB chaperone (SPO3276, 15 × higher than the average of detection under all conditions) and the La protease (SPO2613, 7 ×). On the other hand, for long-term adaptation to high temperatures (only under condition C08), *R. pomeroyi* produces the GroEL chaperonin (SPO0887, 4 ×), a membrane integrated protein with lipid raft-associated functions (SPO2617, 13 ×), and a hypothetical transmembrane protein (SPO2567, 11 ×). At low temperatures (C09, growth at 12 °C), *R. pomeroyi* increases the production of specific proteins: the cold shock protein CspA (SPO3625, 18 ×), the DEAD-box helicase (SPO3833, 20 ×), which increases the cold tolerance of transcriptional regulators CBF/DREB, the transcriptional regulator CarD (SPO1420, 7 ×), the elongation factor TypA (SPO2031, 10 ×) and the FKBP isomerase involved in protein folding (SPOA0438, 8 ×). The cellular responses observed under these temperature-related conditions are highly consistent with those described for another marine bacterium, *Sphingopyxis alaskensis* (Ting *et al.*, 2009). At high salt concentrations near *R. pomeroyi*'s growth limit (condition C10, 7.5% sea salts), a Na-translocating complex comprising SPOA0028 (14 ×), SPOA0029 (7 ×) and SPOA0030 (9 ×) is overproduced. When *R. pomeroyi* is grown in natural sea water supplemented with 0.5% succinate (C28), nitrogen-regulatory proteins P-II (SPO2294, 11 × and SPO3724, 16 ×) and nitrogen-regulation protein NtrC (SPO2087, 9 ×) have a higher detection rate than under any other conditions. These proteins are probably overproduced to cope with an expected nitrogen imbalance in this environment after adding an excess carbon source. In condition C28, we also noted an accumulation of the phasin protein (8 ×) as reported in the natural seawater conditions described here, SPO1710 (the periplasmic component of an active transport system for short-chain amides and urea, 19 ×) and SPO2261 (an HK97 capsid-like protein,

11 ×). On the other hand, when comparing agar plate conditions (C5 and C6, plausible simulation of biofilm formation) to their respective liquid culture conditions (C01 and C02), we observed a decrease in the levels of detection of RNA polymerase (6 × for SPO3507 and SPO3508), ribosomal proteins (from 26% to 13%) and central metabolic enzymes, typical of stationary-phase cells, and a steep increase in carbon monoxide dehydrogenase (SPO1519 44 × and SPO1520 16 ×) and metallo-beta-lactamase enzyme levels (SPO0429 9 × and SPO2170 24 ×).

It is also interesting to highlight those proteins that were overproduced in the presence of diesel and naphthalene. Although it has not been demonstrated that *R. pomeroyi* can assimilate or mineralise these compounds, several proteins are detected as overproduced. They may be involved in the bacterium response to the plausible toxic effects of these compounds using mechanisms as reviewed by Ramos *et al.* (2001). *R. pomeroyi* tolerates high concentrations of diesel added to MB (C35, 3% of diesel) well, as shown by its normal growth (1 day until peak growth) such as that seen under condition C01 (Table 1). Based on the proteomic results, addition of diesel gives rise to several similar processes to those observed under the microaerobic condition C07. We noted a likely increase in nitrogen reduction triggered by the production of SPOA0050 (a nitrous-oxide reductase, 9 ×) and SPOA0220 (a cytochrome cd1 nitrite reductase, 12 ×), as well as a need for ATP synthesis from ADP with the production of polyphosphate kinase 2 (SPO1256, 8 ×) to cope with a plausible lower redox potential. In this sense, diesel seems to limit oxygen exchange, producing a microoxygenic state for the bacterium. The presence of diesel also requires membrane stabilisation with the production of the transmembrane protein SPO2567 (highly detected under naphthalene condition C38, see below) and with the biogenesis of rhamnose for the outer cellular envelope (SPO0212, 16 ×). On the other hand, naphthalene seems to have a much more toxic effect on *R. pomeroyi* (4 days until peak growth, Table 1). In the presence of this aromatic compound, membrane proteins that could have a role in stabilising the cell membrane, SPO2567 (13 ×), SPO2568 (8 ×) and SPO3325 (11 ×), were highly detected. These proteins were also strongly produced under condition C08 (growth at 40 °C). We also detected strong production of proteins SPO2757 (EF hand domain-containing protein, 9 ×), SPO0115 (bifunctional enoyl-CoA hydratase/phosphate acetyltransferase, 14 ×) and SPO1625 (periplasmic serine protease, 19 ×) that have plausible detoxifying roles, and three proteins from efflux pumps SPO1497 (16 ×), SPO0111 (16 ×) and SPO1094 (12 ×).

Differential protein detection among the five conditions used for in-depth analysis. Proteins related to translation, ribosomal structure and

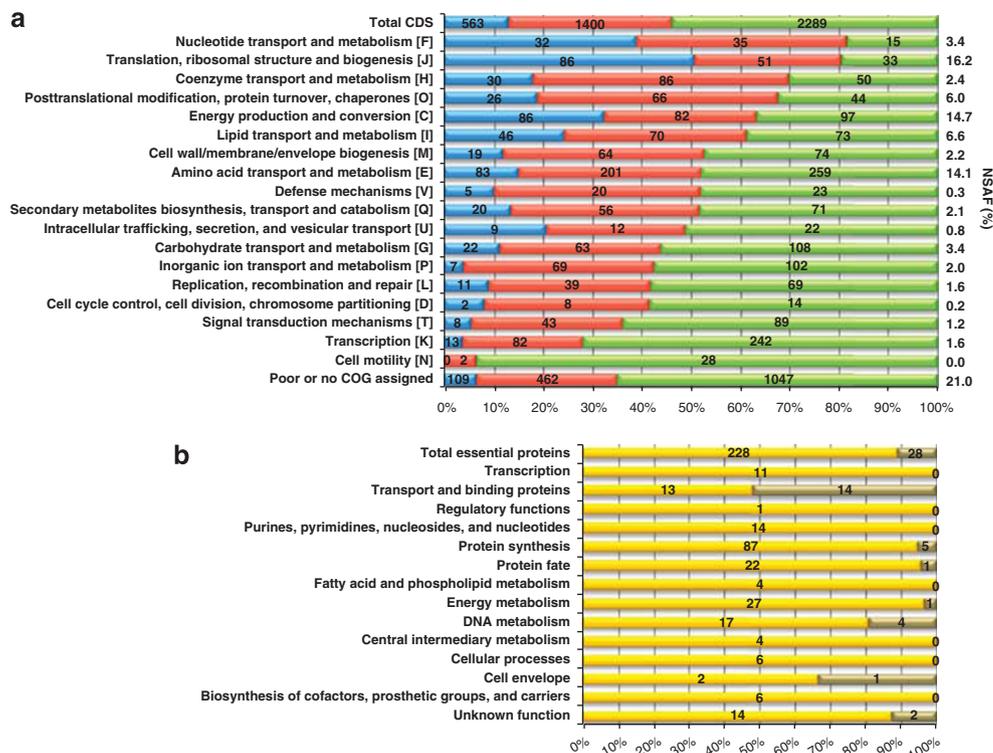


Figure 5 Total and putative essential proteome from *R. pomeroyi* sorted by functional categories. Total proteins in each category are graphed by 100% stacked bars. (a) Proteins belonging to the detected core proteome (blue), accessory-detected proteome (red) and non-detected proteome (green) for each of the COG categories. (b) Detected (yellow) and non-detected (grey) proteins of the putative essential proteome of *R. pomeroyi* sorted by the categories assigned by Glass *et al.* (2006). The number of proteins corresponding to each of the bars is shown. The normalised spectral abundance factor (NSAF) values are the accumulated protein abundance values observed in the 5 in-depth 15-band analyses (Supplementary Table S4).

biogenesis (16%), energy production and conversion (15%), and amino-acid transport and metabolism (14%) were among the most abundant proteins detected irrespective of the growth conditions (Figure 5). Interestingly, the four conditions that had entered the stationary phase (C15, C34, C36 and C38) presented a considerable increase in transporters for the uptake of nutrients remaining in the medium: 98 polypeptides detected in C34 (14% of the total normalised spectral abundant factor), while only 55 in C01 (< 2% of the total). Again, the overall relative abundance of ribosomal proteins decreased slightly under all stationary conditions (8–12%) when compared with the exponential phase sample C01 (18%). The transmembrane protein SPO2567 commented above represented over 19% of the proteins detected during growth in the presence of naphthalene (C38), while it was detected in much lower quantities (0.02–0.34%) under the other four conditions. As *R. pomeroyi* shows a slower growth in the presence of naphthalene, we may suggest a possible function related to specific stress tolerance or membrane stabilisation for this protein. Homologues of SPO2567 are systematically found in 29 of the 32 sequenced *Roseobacter* genomes currently deposited in the Rosebase database. We also noted that SPO2568, annotated as ‘hypothetical protein’ and encoded in the close neighbourhood of the

SPO2567 gene, has a similar detection profile. Other proteins also became abundant in specific samples, which were consistent with the growth conditions applied. For example, benzoate degradation enzymes that are coded in a probable operon (benzoate-coA ligase, benzoate-coA oxygenase A and B subunits, hydrolase and an enoyl-coA hydratase) were abundantly detected only under condition C36. Stress response proteins were observed under condition C15, that is, two different polypeptides, namely SPO3527 and SPO2371, annotated as universal stress proteins. These polypeptides have an Usp-like conserved sequence that corresponds to a putative nucleotide binding domain and may carry out a function resulting in enhancing cellular survival in the presence of stress agents (Gustavsson *et al.*, 2002).

Products of putative essential genes and core MS-detected proteome

In the overall analysis, the COG ‘nucleotide transport and metabolism’ and ‘translation, ribosomal structure and biogenesis’ groups displayed the highest detection ratio (Figure 5a). On the other hand, proteins from ‘signal transduction mechanisms’ and ‘transcription’ (generally transcriptional regulators) COG categories were less detected and found to be

condition-specific. We analysed whether the products of already known essential genes in bacteria were listed in our MS/MS-detected catalogue. For this, we identified *R. pomeroyi* proteins that are homologues to the products of the 382 genes listed as essential in *Mycoplasma genitalium* (Glass et al., 2006) by PSI-BLAST and manual curation. Supplementary Table S6 lists the 256 genes that could be confidently identified in the *R. pomeroyi* genome on this basis. A total of 228 (89%) of the corresponding proteins were detected by MS/MS (Figure 5b). The only abundant group of proteins that presented a low detection ratio (48%) was the one corresponding to 'transport and binding proteins'. This was expected owing to the exclusive growth conditions applied to *M. genitalium* (Glass et al., 2006) and the wide range of ABC and TRAP transporters observed in *R. pomeroyi* (Moran et al., 2007). In the whole study, 145 polypeptides related to these transporters were detected, out of a total of 368 (39%) annotated in the genome. Excluding this specific group of proteins from the putative essential gene analysis, we covered 94% of the total essential genes (215 of the 229) within the whole detected proteome reported here.

The proteins were systematically detected, no matter what culture conditions were listed and designated as the core MS-detected proteome. A total of 563 polypeptides were found (shaded items in Supplementary Table S5) when compiling those detected under all 5 conditions analysed in depth. This core MS-detected proteome could define the main physiological basic traits of *R. pomeroyi*. Indeed, we observed a high predominance of proteins belonging to the 'translation, ribosomal structure and biogenesis', 'energy production and conversion' and 'amino-acid transport and metabolism' COG groups. From the core-proteome list, 318 proteins were found to be strictly conserved (sequence similarities with *E* values below 10^{-20}) out of the 32 currently sequenced *Roseobacter* genomes. This number notably increases when considering a less strict degree of occurrence, this being 481 polypeptides conserved in 27 out of 32 genomes (Supplementary Table S5). Interestingly, 48 'hypothetical' proteins were evidenced in this core MS-detected proteome. Their function should be important as they are detected systematically under all 5 conditions. Of these proteins, 31 are relatively well conserved among the 32 sequenced *Roseobacter* genomes (seen in over 23 of these isolates), whereas 5 are exclusive to *R. pomeroyi* (SPOA0341 and SPOA0343, two proteins with a plausible role in hormone production commented above, SPO1254, SPO1996 and SPO2776). When comparing the MS-detected core proteome of *R. pomeroyi* with the *Roseobacter* core genome defined previously by Moran et al. (Moran et al., 2007), we found 66% of the MS-detected proteins (370 out of 563) among the 1310 proteins established by comparative genomics (Supplementary Table S5). When comparing our

core proteome with a more recent genomic analysis of marine *Roseobacters* that restricted the core-genome set (Tang et al., 2010) to 694 items, we found 68% (475) of these as MS-detected in our whole study.

Confines of identifiable proteome

Figure 6 shows the protein discovery saturation curves in terms of novel detected proteins as we added new growing conditions up to 30 (panel a) and the 5 conditions analysed in depth (panel b). In both cases, the rapid increase in terms of identification of proteins observed after analysing the first conditions was followed by apparent non-saturation. This indicates that under each specific condition assayed, a few novel genes were being expressed in order for the bacterium to cope with the novel environment conditions or stress. Figure 6c, shows that for abundant proteins, identification seems to saturate below 1500 polypeptides if the logarithmical tendency is extrapolated to 1000 theoretically different physiological conditions. Remarkably, the limit would be the total theoretical proteome (4252 polypeptides) through a 15-band fully resolved gel-based approach (Figure 6d). Although the data correlate well with a logarithmic regression tendency (r^2 is over 0.99 in both cases), these considerations should be taken with caution, as experimental data do not cover the inflection point of the curve. At this stage, another aspect to take into consideration is when novel polypeptide detection overlaps false positive increase (Figure 6e). In this case, the powered algorithm tendency of novel polypeptide detections among the total and the rate of predicted false positive accumulation is shown. The false positive detection would exceed novel polypeptide detection after analysis of 38 in-depth conditions, being in this case above 5.2% (Figure 6e).

Discussion

Here, we present the first study describing the proteome variations of a marine bacterium in its natural environment and the influence exerted by the natural community. The four different seawater samples tested in the presence or absence of a natural seawater microbial community revealed interesting variations regarding the production of several key proteins (Figure 1a). Nevertheless, most of the proteins detected (over 95%) showed no significant variation between the different conditions compared. In order to stress the limits of the expressed genome of the bacterium we processed 30 different laboratory conditions. This allowed us to establish the broadest proteome compilation of a *R. clade* member, covering 46% of its total theoretical proteome. We confidently validated 1963 *R. pomeroyi* DSS-3 polypeptides after detecting 23 957 non-redundant peptides by MS/MS. In this way we

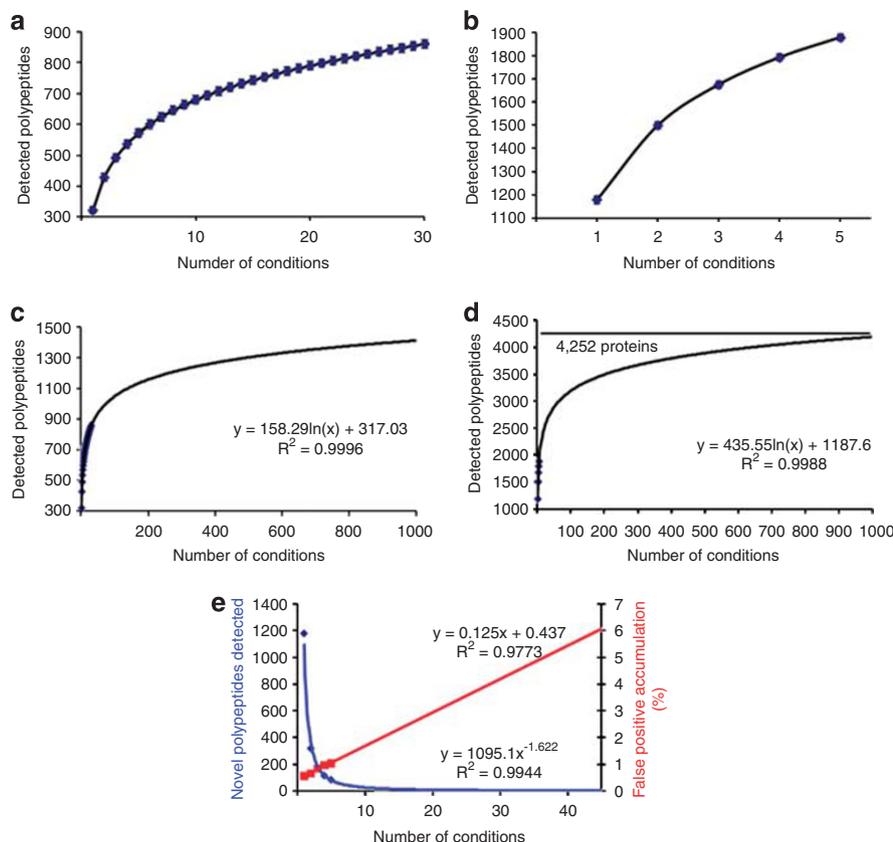


Figure 6 Protein discovery saturation curves. Polypeptide accumulation among the 30 growth conditions (a) and 5 in-depth analyses including 15-band, phosphocellulose and membrane-bound polypeptide analysis (b) are represented by means of sample rarefaction. S.d. bars are represented. Logarithmic tendency of polypeptide accumulation over 1000 hypothetical conditions following the one-band abundant protein (c) or in-depth analysis (d) strategies are shown. The top line in the (d) graph shows the limit corresponding to the total 4252 theoretical proteome of *R. pomeroyi*. (e) Powered tendency of new polypeptide detection in novel in-depth analysis (blue, left axis) and the false-positive percentage increase and extrapolated tendency when considering additional conditions (red, right axis).

covered over 48% of the theoretical chromosomally encoded proteome and 32% of the theoretical pDSS-3 megaplasmid-encoded proteome.

In essence, a complete functional genomic analysis of *R. pomeroyi* was possible using the extensive proteomic data compiled in this report. By comparing the proteomes detected for each of the conditions, we have defined essential proteins to face specific stresses. This includes not only the proteins already expected (for example, cold-shock proteins at low temperatures or chaperonin-like proteins at high temperatures), but also proteins with unknown functions that may have specific roles under given conditions, for example, SPO2567 in the presence of naphthalene and SPO1254 with port microbial communities. Some of these detected proteins could be used as bioindicators for contamination, such as benzoate determinants or copper resistance proteins detected in the port seawaters. Data obtained from the whole analysis also showed general variations due to different cellular physiological states or mechanisms to face common stresses (for example, ribosomal proteins or the proteins encoded on the pDSS3 megaplasmid in Figure 4). We found that *R. pomeroyi*'s genome encodes for (i) proteins

dedicated to specific conditions and (ii) others predicted for general physiological states and core proteome functions of the bacterium. In order to estimate the ratio of each group, we listed the proteins that were detected under all the conditions tested, those detected throughout the study and those never detected (Figure 5a). In the first place we found that around 600 proteins are always detected among the different in-depth conditions tested (~15% of the theoretical proteome). This so-called core MS-detected proteome includes most of the house-keeping genes needed for marine *Roseobacters* (Supplementary Table S5, Moran *et al.*, 2007; Tang *et al.*, 2010). However, most free-living marine bacteria require a larger pool of functions to face their environment. We have shown that the number of proteins produced by *R. pomeroyi* to face a given environmental condition is around 1200–1400 polypeptides (each of the five in-depth analyses, Supplementary Table S4 (35% of its theoretical proteome)). It is noteworthy that this number of proteins would be very close to the whole proteome of the minimal oligotrophic marine bacterium *C. pelagibacter ubique* (Giovannoni *et al.*, 2005). This bacterium is highly adapted to grow under very

specific conditions found in the oligotrophic ocean. When we added the MS/MS data obtained from all the different physiological conditions tested, we compiled 1963 proteins (~50% of the theoretical proteome of *R. pomeroyi*). Interestingly, Wegener *et al* obtained the same value (1955 polypeptides) when analysing the proteome of the freshwater cyanobacterium *Synechocystis* sp. PCC 6803 under 12 different environmental conditions (Wegener *et al.*, 2010). This indicates that 30–35% of the theoretical proteome of *R. pomeroyi* when not including the 600 core proteome from the 1963 total detected) might be dedicated to adaptation to general physiological changes (that is, exponential-stationary transition), environmental changes (that is, rich-poor media) or common stress response factors. Thus, we suggest that the remaining 50% of the theoretical proteome of *R. pomeroyi* seems to be dedicated to very specific functions and will be detected only if specific growth conditions are attempted (Figures 6b and d).

The analysis carried out here has allowed us to define the genomic fractions of *R. pomeroyi*. Although the oligotrophic waters of the vast oceans are mainly dominated by oligotrophic-like organisms (Biers *et al.*, 2009), in surface and coastal waters a higher predominance of bacteria with a generalist lifestyle is found (Mou *et al.*, 2008; Lauro *et al.*, 2009). The use of high throughput proteomics to analyse the detectable proteome of *R. pomeroyi* has revealed the basic need for a marine generalist in terms of proteome. Our data suggest that ~30% of its theoretically encoded proteome is dedicated to what could be considered as the bare minimum proteome for living in a virtual environment without stress, energy and specific metabolite requirements, that is, what is generally considered as house-keeping functions. Whether the equivalent bare minimum proteome could be the same in specialist organisms remains to be established. Most but not all of these proteins are indeed orthologs and conserved among microorganisms. An additional ~20% of the *R. pomeroyi* proteome is dedicated to adaptation to general cellular physiological variations and main specific physiological traits. In essence, this pool is also probably required in specialist organisms but could be much diverse in nature and proportion. This fraction could also be drastically reduced or modified in species such as *C. pelagibacter ubique*, where transitions from exponential to stationary phase have shown very small protein variations (Sowell *et al.*, 2008). The remaining 50% of *R. pomeroyi* theoretically encoded proteome would be plausibly required to face very specific environmental changes as shown here. This 50%-opportunistic gene pool is possibly lacking or extremely reduced in specialist-like bacteria. Part of the 50%-opportunistic fraction could even just represent a gene pool in generalists for further long-term mutation-based adaptation of the population.

Our proteomic study shows that covering the full theoretical proteome of such generalist microorganisms is today a difficult challenge. Hundreds of different growth and stress conditions are probably necessary to detect the whole theoretical proteome of *R. pomeroyi*. This marine bacterium comprises a large toolbox to take advantage of a large number of lithoheterotrophic energy sources, interact with its environmental and cope with the potential toxic effects associated to different life conditions. We have shown here that half of its large genetic pool is probably used in very specific situations. Nevertheless, the products encoded by this fraction give enough adaptive advantage in the marine environment to be maintained in the genome.

Acknowledgements

JA Christie-Oleza was supported by a fellowship from the Fundación Ramón Areces. We thank the Commissariat à l'Énergie Atomique et aux Énergies Alternatives, the Fundación Ramón Areces and the Ministère des Affaires Étrangères et Européennes, République Française (Partenariat Hubert Curien, Picasso Program) for financial support. We also thank C Bruley (CEA-Grenoble) for providing the IRMa parser, Jean-Charles Gaillard and Olivier Pible (both from CEA-Marcoule) for help with data treatment. B Nogales and R Bosch acknowledge the financial support of the Spanish MICINN through project CTM2008-02574/MAR (with FEDER co-funding) and the Acciones Integradas program (FR2009-0106).

References

- Baudet M, Ortet P, Gaillard JC, Fernandez B, Guerin P, Enjalbal C *et al.* (2010). Proteomics-based refinement of *Deinococcus deserti* genome annotation reveals an unwonted use of non-canonical translation initiation codons. *Mol Cell Proteomics* **9**: 415–426.
- Biers EJ, Sun S, Howard EC. (2009). Prokaryotic genomes and diversity in surface ocean waters: interrogating the global ocean sampling metagenome. *Appl Environ Microbiol* **75**: 2221–2229.
- Buchan A, Gonzalez JM, Moran MA. (2005). Overview of the marine roseobacter lineage. *Appl Environ Microbiol* **71**: 5665–5677.
- Burgmann H, Howard EC, Ye W, Sun F, Sun S, Napierala S *et al.* (2007). Transcriptional response of *Silicibacter pomeroyi* DSS-3 to dimethylsulfoniopropionate (DMSP). *Environ Microbiol* **9**: 2742–2755.
- Carvalho PC, Fischer JS, Chen EI, Yates JR, Barbosa VC. (2008). PatternLab for proteomics: a tool for differential shotgun proteomics. *BMC Bioinformatics* **9**: 316.
- Christie-Oleza JA, Armengaud J. (2010). In-depth analysis of exoproteomes from marine bacteria by shotgun liquid chromatography-tandem mass spectrometry: the *Ruegeria pomeroyi* DSS-3 case-study. *Mar Drugs* **8**: 2223–2239.
- Clair G, Roussi S, Armengaud J, Duport C. (2010). Expanding the known repertoire of virulence factors produced by *Bacillus cereus* through early secretome profiling in three redox conditions. *Mol Cell Proteomics* **9**: 1486–1498.

- de Groot A, Dulermo R, Ortet P, Blanchard L, Guerin P, Fernandez B *et al.* (2009). Alliance of proteomics and genomics to unravel the specificities of Sahara bacterium *Deinococcus deserti*. *PLoS Genet* **5**: e1000434.
- DeLong EF. (2009). The microbial ocean from genomes to biomes. *Nature* **459**: 200–206.
- Fuhrman JA. (2009). Microbial community structure and its functional implications. *Nature* **459**: 193–199.
- Giovanoni 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.
- Glass JI, Assad-Garcia N, Alperovich N, Yooseph S, Lewis MR, Maruf M *et al.* (2006). Essential genes of a minimal bacterium. *Proc Natl Acad Sci USA* **103**: 425–430.
- Gupta N, Benhamida J, Bhargava V, Goodman D, Kain E, Kerman I *et al.* (2008). Comparative proteogenomics: combining mass spectrometry and comparative genomics to analyze multiple genomes. *Genome Res* **18**: 1133–1142.
- Gustavsson N, Diez A, Nystrom T. (2002). The universal stress protein paralogues of *Escherichia coli* are co-ordinately regulated and co-operate in the defence against DNA damage. *Mol Microbiol* **43**: 107–117.
- Jaffe JD, Stange-Thomann N, Smith C, DeCaprio D, Fisher S, Butler J *et al.* (2004). The complete genome and proteome of *Mycoplasma mobile*. *Genome Res* **14**: 1447–1461.
- Kato JY, Funa N, Watanabe H, Ohnishi Y, Horinouchi S. (2007). Biosynthesis of gamma-butyrolactone autoregulators that switch on secondary metabolism and morphological development in *Streptomyces*. *Proc Natl Acad Sci USA* **104**: 2378–2383.
- Keller M, Hettich R. (2009). Environmental proteomics: a paradigm shift in characterizing microbial activities at the molecular level. *Microbiol Mol Biol Rev* **73**: 62–70.
- Koonin EV, Wolf YI. (2008). Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res* **36**: 6688–6719.
- Lanfranconi MP, Bosch R, Nogales B. (2010). Short-term changes in the composition of active marine bacterial assemblages in response to diesel oil pollution. *Microbial Biotechnol* **3**: 607–621.
- Lauro FM, McDougald D, Thomas T, Williams TJ, Egan S, Rice S *et al.* (2009). The genomic basis of trophic strategy in marine bacteria. *Proc Natl Acad Sci USA* **106**: 15527–15533.
- Liu H, Sadygov RG, Yates 3rd JR. (2004). A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* **76**: 4193–4201.
- Moran MA, Buchan A, Gonzalez JM, Heidelberg JF, Whitman WB, Kiene RP *et al.* (2004). Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment. *Nature* **432**: 910–913.
- Moran MA, Belas R, Schell MA, Gonzalez JM, Sun F, Sun S *et al.* (2007). Ecological genomics of marine *Roseobacters*. *Appl Environ Microbiol* **73**: 4559–4569.
- Mou X, Sun S, Edwards RA, Hodson RE, Moran MA. (2008). Bacterial carbon processing by generalist species in the coastal ocean. *Nature* **451**: 708–711.
- Newton RJ, Griffin LE, Bowles KM, Meile C, Gifford S, Givens CE *et al.* (2010). Genome characteristics of a generalist marine bacterial lineage. *ISME J* **4**: 784–798.
- Paoletti AC, Parmely TJ, Tomomori-Sato C, Sato S, Zhu D, Conaway RC *et al.* (2006). Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors. *Proc Natl Acad Sci USA* **103**: 18928–18933.
- Polz MF, Hunt DE, Preheim SP, Weinreich DM. (2006). Patterns and mechanisms of genetic and phenotypic differentiation in marine microbes. *Philosophical Transactions of the Royal Society B-Biological Sciences* **361**: 2009–2021.
- Ramos JL, Gallegos MT, Marques S, Ramos-Gonzalez MI, Espinosa-Urgel M, Segura A. (2001). Responses of Gram-negative bacteria to certain environmental stressors. *Curr Opin Microbiol* **4**: 166–171.
- Scanlan DJ, Ostrowski M, Mazard S, Dufresne A, Garczarek L, Hess WR *et al.* (2009). Ecological genomics of marine picocyanobacteria. *Microbiol Mol Biol Rev* **73**: 249–299.
- Singer E, Webb EA, Nelson WC, Heidelberg JF, Ivanova N, Pati A *et al.* (2011). The genomic potential of *Marinobacter aquaeolei*—A biogeochemical opportunist. *Appl Environ Microbiol* **77**: 2763–2771.
- Sowell SM, Norbeck AD, Lipton MS, Nicora CD, Callister SJ, Smith RD *et al.* (2008). Proteomic analysis of stationary phase in the marine bacterium ‘*Candidatus* Pelagibacter ubique’. *Appl Environ Microbiol* **74**: 4091–4100.
- Strom SL. (2008). Microbial ecology of ocean biogeochemistry: a community perspective. *Science* **320**: 1043–1045.
- Tang K, Huang H, Jiao N, Wu CH. (2010). Phylogenomic analysis of marine *Roseobacters*. *PLoS One* **5**: e11604.
- Ting L, Cowley MJ, Hoon SL, Guilhaus M, Raftery MJ, Cavicchioli R. (2009). Normalization and statistical analysis of quantitative proteomics data generated by metabolic labeling. *Mol Cell Proteomics* **8**: 2227–2242.
- VerBerkmoes NC, Deneff VJ, Hettich RL, Banfield JF. (2009). Systems biology: Functional analysis of natural microbial consortia using community proteomics. *Nat Rev Microbiol* **7**: 196–205.
- Wagner-Dobler I, Biebl H. (2006). Environmental biology of the marine *Roseobacter* lineage. *Annu Rev Microbiol* **60**: 255–280.
- Wegener KM, Singh AK, Jacobs JM, Elvitigala TR, Welsh EA, Keren N *et al.* (2010). Global proteomics reveal an atypical strategy for carbon/nitrogen assimilation by a cyanobacterium under diverse environmental perturbations. *Mol Cell Proteomics* **9**: 2678–2689.
- Wilmes P, Bond PL. (2006). Metaproteomics: studying functional gene expression in microbial ecosystems. *Trends Microbiol* **14**: 92–97.
- Wilmes P, Bond PL. (2009). Microbial community proteomics: elucidating the catalysts and metabolic mechanisms that drive the Earth’s biogeochemical cycles. *Curr Opin Microbiol* **12**: 310–317.
- York GM, Junker BH, Stubbe J, Sinskey AJ. (2001). Accumulation of the PhaP phasin of *Ralstonia eutropha* is dependent on production of polyhydroxybutyrate in cells. *J Bacteriol* **183**: 4217–4226.
- Zech H, Thole S, Schreiber K, Kalhofer D, Voget S, Brinkhoff T *et al.* (2009). Growth phase-dependent global protein and metabolite profiles of *Phaeobacter gallaeciensis* strain DSM 17395, a member of the marine *Roseobacter*-clade. *Proteomics* **9**: 3677–3697.

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