The ISME Journal (2012) 6, 2229–2244

www.nature.com/ismej

ORIGINAL ARTICLE

Phaeobacter gallaeciensis genomes from globally opposite locations reveal high similarity of adaptation to surface life

Sebastian Thole[1], Daniela Kalhoefer[1], Sonja Voget[2], Martine Berger[1], Tim Engelhardt[1], Heiko Liesegang[2], Antje Wollherr[2], Staffan Kjelleberg[3,4], Rolf Daniel[2], Meinhard Simon[1], Torsten Thomas[3] and Thorsten Brinkhoff[1]

[1]*Institute for Chemistry and Biology of the Marine Environment, University of Oldenburg, Oldenburg, Germany;* [2] *Department of Genomic and Applied Microbiology and Göttingen Genomics Laboratory, Institute of Microbiolgy and Genetics, University of Göttingen, Göttingen, Germany;* [3] *Centre of Marine Bio-Innovation and School of Biotechnology and Biomolecular Sciences, The University of New South Wales, Sydney, New South Wales, Australia and* [4]*Singapore Centre on Environmental Life Sciences Engineering, Nanyang Technological University, Singapore, Singapore*

***Phaeobacter gallaeciensis*, a member of the abundant marine *Roseobacter* clade, is known to be an effective colonizer of biotic and abiotic marine surfaces. Production of the antibiotic tropodithietic acid (TDA) makes *P. gallaeciensis* a strong antagonist of many bacteria, including fish and mollusc pathogens. In addition to TDA, several other secondary metabolites are produced, allowing the mutualistic bacterium to also act as an opportunistic pathogen. Here we provide the manually annotated genome sequences of the *P. gallaeciensis* strains DSM 17395 and 2.10, isolated at the Atlantic coast of north western Spain and near Sydney, Australia, respectively. Despite their isolation sites from the two different hemispheres, the genome comparison demonstrated a surprisingly high level of synteny (only 3% nucleotide dissimilarity and 88% and 93% shared genes). Minor differences in the genomes result from horizontal gene transfer and phage infection. Comparison of the *P. gallaeciensis* genomes with those of other roseobacters revealed unique genomic traits, including the production of iron-scavenging siderophores. Experiments supported the predicted capacity of both strains to grow on various algal osmolytes. Transposon mutagenesis was used to expand the current knowledge on the TDA biosynthesis pathway in strain DSM 17395. This first comparative genomic analysis of finished genomes of two closely related strains belonging to one species of the *Roseobacter* clade revealed features that provide competitive advantages and facilitate surface attachment and interaction with eukaryotic hosts.**

The ISME Journal (2012) **6,** 2229–2244; doi:10.1038/ismej.2012.62; published online 21 June 2012
Subject Category: integrated genomics and post-genomics approaches in microbial ecology
Keywords: antibiotic; comparative genomics; prophages; *Roseobacter* clade; siderophores; symbiosis

Introduction

Phaeobacter gallaeciensis belongs to the *Roseobacter* clade (Martens *et al.*, 2006), and members of this genus are present in various marine habitats in temperate regions (Pommier *et al.*, 2005). *Phaeobacter* were found in turbot larva rearings (Hjelm *et al.*, 2004), cutaneous mucus of seahorses (Balcázar *et al.*, 2010) and on cephalopods (Grigioni *et al.*, 2000; Barbieri *et al.*, 2001). In the Chinese Changjiang estuary, close relatives of *P. gallaeciensis* are among the most abundant bacterial groups (Sekiguchi *et al.*, 2002). *Phaeobacter* spp. are effective surface colonizers and antagonistic against other bacteria due to the antibiotic tropodithietic acid (TDA), which is produced by several strains affiliated with the genera *Phaeobacter*, *Ruegeria* and *Pseudovibiro* (Porsby *et al.*, 2008; Gram *et al.*, 2010). Because of the competitive success of *Phaeobacter* spp. against fish and mollusc pathogens, ongoing research focuses on these organisms as probiotic agents in aquaculture (Rao *et al.*, 2007; Porsby *et al.*, 2008; Prado *et al.*, 2009). Most recently, it was shown that *P. gallaeciensis* also produces algicides upon sensing the lignin-derived breakdown product p-coumaric acid from a co-cultivated microalga (Seyedsayamdost *et al.*, 2011). Thus, the mutualistic

Correspondence: T Thomas, Centre of Marine Bio-Innovation and School of Biotechnology and Biomolecular Sciences, The University of New South Wales, Sydney, New South Wales 2052 Australia.
E-mail: t.thomas@unsw.edu.au
Received 16 December 2011; revised 7 May 2012; accepted 14 May 2012; published online 21 June 2012

relation between this bacterium and its host can switch to parasitism. In addition, other secondary metabolites, like acylated homoserine lactones (Bruhn *et al.*, 2005; Berger *et al.*, 2011), are produced by *Phaeobacter* species and it was shown that *P. gallaeciensis* BS107 harbors a hybrid polyketide synthase/non-ribosomal peptide synthetase (NRPS) cluster (Martens *et al.*, 2007). However, none of the above-mentioned traits is unique to *Phaeobacter* strains, hence, limiting the knowledge of features delineating *P. gallaeciensis* from other roseobacters.

The *P. gallaeciensis* strains DSM 17395 and 2.10 were isolated from geographically very distant locations at the Atlantic coast of north western Spain (strain DSM 17395; scallop larvae hatchery; Ruiz-Ponte *et al.*, 1998) and Sydney, Australia (strain 2.10; surface of the alga *Ulva lactuca*; Rao *et al.*, 2005). Despite the great distance between the isolation sites, the organisms are very closely related and indistinguishable by their 16S rRNA gene sequence.

For more than 20 years, bacteria affiliated with the marine *Roseobacter* clade have been intensively studied because of their abundance in diverse marine habitats as well as their physiological and metabolic versatility. In addition, as part of the Marine Microbiology Initiative (http://camera.calit2.net/microgenome/) and other sequencing projects, the complete genomes of 40 members of this prominent bacterial cluster were sequenced (Brinkhoff *et al.*, 2008). Comparative analyses based on the (mostly unfinished) genomes of these isolates emphasized the heterogeneity of this bacterial clade and demonstrated the difficulty to classify and organize roseobacters in terms of ecological or biogeochemical functions (Newton *et al.*, 2010). In contrast, comparative analyses encompassing more closely related organisms have proven to be a valuable approach to assess patterns of differentiation and adaptive strategies (Rocap *et al.*, 2003; Pena *et al.*, 2010; Kalhoefer *et al.*, 2011). Our study thus focused on the genome comparison of two strains of the same species to gain insights into the genomic and phenotypic diversity below the species level (Schloter *et al.*, 2000).

Here we report the complete genome sequences and comparative analysis of the manually annotated genomes of the *P. gallaeciensis* strains DSM 17395 and 2.10. Differences between the genomes were examined and *Phaeobacter*-specific coding sequences were compared with other *Roseobacter* clade members. Additional genomic traits were deduced from the genomes and tested *in vitro*. This study compares, for the first time, two finished genomes of one species within the *Roseobacter* clade and provides insight into intraspecies diversity and mechanisms leading to genomic differentiation. A comparison with 35 *Roseobacter* genomes further revealed genomic and physiological traits that are characteristic for *P. gallaeciensis* strains.

Materials and methods

Genome sequencing, assembly, finishing and annotation

The genomes of the *P. gallaeciensis* strains DSM 17395 and 2.10 (DSM 24588) were sequenced as parts of the Microbial Genome Sequencing Project (http://camera.calit2.net/microgenome/). Gap closure and sequence editing were performed using the Staden software package (Staden, 1996). For further details see Supplementary Material S1. The sequences have been deposited in GenBank under the accession numbers CP002976, CP002977, CP002978, CP002979 (strain DSM 17395) and CP002972, CP002973, CP002974, CP002975 (strain 2.10).

Prediction of potential protein coding open reading frames (ORFs) was conducted using YACOP (Tech and Merkl, 2003). ORF calling was verified manually and expert annotation was performed with the ERGO software package (Overbeek *et al.*, 2003) (Supplementary Material S1).

Sequence analyses and comparative genomics

The BiBag software tool—combining reciprocal BLAST analyses and global sequence alignments—was applied to compare the genomes of the two *P. gallaeciensis* strains (DSM 17395 and 2.10), as well as 35 other roseobacters (Supplementary Material S1). Whole-genome alignments were performed with the MAUVE software (Darling *et al.*, 2004) and circular plots were generated with DNAPlotter (Carver *et al.*, 2009). Functional classes were assigned to the ORFs according to the cluster of orthologous groups classification (Tatusov *et al.*, 1997). Prediction of horizontally acquired genes was performed using the programs IslandViewer and COLOMBO (Waack *et al.*, 2006; Langille and Brinkman, 2009). Because different results were obtained with the two programs, predicted alien genes and genomic islands (GEIs) were manually checked for elements commonly associated with GEIs like transposases, integrases, tRNAs and GC-content deviations. The metabolic pathways were reconstructed using the Pathway-tools software (Karp *et al.*, 2002) from the Biocyc Database collection (Karp *et al.*, 2005). The pathways were checked and curated manually, if required.

The recently published genome of *P. gallaeciensis* ANG1 (Collins and Nyholm, 2011) could not be included in the analysis of unique gene content as it is in draft stage (135 contigs). Furthermore, the 16S rRNA gene of strain ANG1 is more similar to that of *Phaeobacter daeponensis* strain TF-218 (99%) than to those of strains DSM 17395 and 2.10 (97.32% identity), so that ANG1 is not suitable for strain-level analysis within the species *P. gallaeciensis*.

Growth experiments

For growth tests, cells of *P. gallaeciensis* DSM 17395 and 2.10 were cultivated in salt-water medium

(Zech *et al*., 2009). Media were supplemented with different carbon or nitrogen sources. For details see Supplementary Material S1.

Transposon Tn5 mutagenesis of P. gallaeciensis DSM 17395
Random insertion libraries were generated using EZ-Tn5 mutagenesis as described by Berger *et al.* (2011). TDA-negative mutants were screened for reduced pigmentation, which is known to coincide with loss of antibiotic production. For details see Supplementary Material S1.

Induction of prophages with mitomycin C and virus analysis
The *Phaeobacter* genomes were checked for prophage-associated regions. A protocol by Chen *et al.* (2006) was adapted for induction of prophages with mitomycin C. Bacterial growth and abundance of induced phages were monitored by measuring the optical density or by epifluorescence microscopy. Phage lysates were concentrated and analyzed with pulsed field gel electrophoresis. For details see Supplementary Material S1.

Chrome Azurol S test for siderophore analysis
Secretion of iron-scavenging siderophores was determined on Chrome Azurol S agar plates, basically following the protocols by Schwyn and Neilands (1987), Alexander and Zuberer (1991) and Carson *et al.* (1992). For details see Supplementary Material S1.

Colonization experiments
P. gallaeciensis DSM 17395 was labeled with a green fluorescent protein color tag and co-cultivated with macroalgae, microalgae, pieces of driftwood or crab tissue. Growth and surface colonization was monitored by epifluorescence microscopy as described in the Supplementary Material S1.

Results and Discussion

Comparative genome analysis of P. gallaeciensis strains DSM 17395 and 2.10
The genomes of *P. gallaeciensis* DSM 17395 and 2.10 have a size of 4 227 134 and 4 160 916 bp, respectively, and are each organized in a circular chromosome and three plasmids (Table 1). On the nucleotide level, the genomes differ by only 3% and the chromosomes and plasmids are highly syntenic (Figures 1 and 2 and Supplementary Materials S2A and B). Only two loci at the start and end of the chromosomes are inverted. Highly similar gene order and orientation, as well as 94% similarity on the nucleotide and protein level, were also reported for two co-habitating strains of the halophilic bacterium *Salinibacter ruber* (Pena *et al*., 2010). In contrast, two strains of *Alteromonas macleodii* (Ivars-Martinez *et al*., 2008), isolated from either the Mediterranean Sea or close to Hawaii, revealed a low level of synteny, an average nucleotide identity of only 81.25% and 65% of shared genes. This comparison shows that the two *Phaeobacter* strains have maintained a level of genomic congruence, similar to what has been seen in co-habitating strains and higher than that of other geographically separated strains.

The two *P. gallaeciencsis* strains share a total of 3438 coding sequences (Needleman–Wunsch similarity $\geqslant$30%), accounting for 88% in strain DSM 17395 and 93% of all predicted genes of strain 2.10. Strain DSM 17395 harbors 437 genes not present in strain 2.10, whereas 285 genes are unique in strain 2.10 (Supplementary Material S4). In both

Table 1 General genomic features of *P. gallaeciensis* strains DSM 17395 and 2.10.

P. gallaeciensis *strain DSM 17395*	*Chromosome*	*pPGA1_262*	*pPGA1_78*	*pPGA1_65*
Length (bp)	3 821 831	261 820	78 238	65 245
G+C content (%)	59.84	57.99	62.67	62.72
Number of CDSs	3528	239	61	47
Number of pseudogenes	16			
Number of unique[a] genes	407	48	1	3
rRNAs	4			
tRNAs	56	1		
P. gallaeciensis *strain 2.10*	*Chromosome*	*pPGA2_239*	*pPGA2_95*	*pPGA2_71*
Length (bp)	3 758 273	2 37 765	94 494	70 384
G+C content (%)	59.78	58.4	61.1	62.78
Number of CDSs	3396	208	71	49
Number of pseudogenes	4	1		
Number of unique[a] genes	234	19	11	4
rRNAs	4			
tRNAs	56	1		

Abbreviations: CDSs, coding sequences; rRNA, ribosomal RNA; tRNA, transfer RNA.
[a]Not present in the other *P. gallaeciensis* strain.

strains, unique genes code for different types of transporters, transcriptional regulators and two-component signal transduction systems beside many hypothetical genes. Among the unique features in strain DSM 17395 are a complete type IV secretion system (PGA1_c22830-PGA1_c22980) as well as three genes that are part of the biosynthetic pathway for acetoin and 2,3-butanediol, indicating the possible production of these compounds by the strain. These compounds can act as nutrient source as well as plant growth-promoting molecules (Ryu *et al.*, 2003; Rudrappa *et al.*, 2010), which might facilitate the interaction with algal hosts. Genes for the complete multi-peptide urease (*ureA-G*; PGA2_c31970-c32030) and three genes probably coding for surface antigens (PGA2_c18820-PGA2_c18840) are also unique in strain 2.10.

Plasmid pPGA2_239 of strain 2.10 has a large insertion (~20 kb, see Figure 1 and Supplementary Material S2B) compared with DSM 17395 through an integron containing an ABC transporter with predicted specificity for spermidine/putrescine, a glutathione-S-transferase and an helix-turn-helix-type transcriptional regulator. An additional insertion was found on pPGA2_95 of strain 2.10, containing one integrase and eight proteins of unknown function with no homologs in the *Roseobacter* clade. The only difference in the smallest plasmids is the presence of a complete NRPS in strain 2.10 (PGA2_71p110), which is located between three putative transposases and a two-component regulatory system. Evidence for mobility of this 'NRPS transposon' is given by its location on the chromosome of strain DSM 17395 (PGA1_c28490). This ORF is similar (E value 3×10^{-80}) to a subunit of the *Bacillus subtilis* surfactin synthase, thus indicating a similar function in *Phaeobacter*. Furthermore, the NRPS sequence appears to be rare within the *Roseobacter* clade, as we could only find ORFs with weak similarity (<30%) in six other roseobacters.

Comparison of the chromosomes of the *Phaeobacter* strains illustrates the high level of genetic uniformity. This is remarkable, given the large geographic distance of the isolation sites. The overall synteny of the plasmids of strains 2.10 and DSM 17395 is striking. They harbor *Phaeobacter*-characteristic traits (see below), thus indicating plasmid-located genes being mediators of adaptation and heredity transmission. This is supported by recent studies on the plasmids and replication machineries of *P. gallaeciensis* and other roseobacters (Petersen, 2011). There, a new replicase (*dnaA*-like) on pPGA1_262 was found to be essential for stable coexistence of this plasmid in the bacterium. Transformation analyses revealed that the presence of an incompatible construct resulted in loss of the 262-kb replicon and led to white colonies of strain DSM 17395, thus indicating that the plasmid is essential for the production of the characteristic brown pigment produced by *P. gallaeciensis* (Petersen *et al.*, 2011).

Several of the strain-specific genes have no homologs in any database ('orphan genes') and are interesting targets for further research, because they may account for *P. gallaeciensis* strain-specific metabolic traits and may also influence the host specificity (Mira, 2002; Silva *et al.*, 2002). Our survey also suggests strain differentiation in the regulatory networks of extracytoplasmic function σ factors and small RNAs (Supplementary Materials S12 and 13).

GEIs, prophages and gene-transfer agents (GTAs)

Many of the unique genes in both strains are located in GEIs. We found 98 putative alien genes that are organized in three GEIs in the genome of strain DSM 17395 (Figure 2a) and 126 ORFs in five GEIs in that of strain 2.10 (Supplementary Material S2A, circle 4). The GEIs are inserted in the vicinity of tRNAs and several mobility-related elements, such

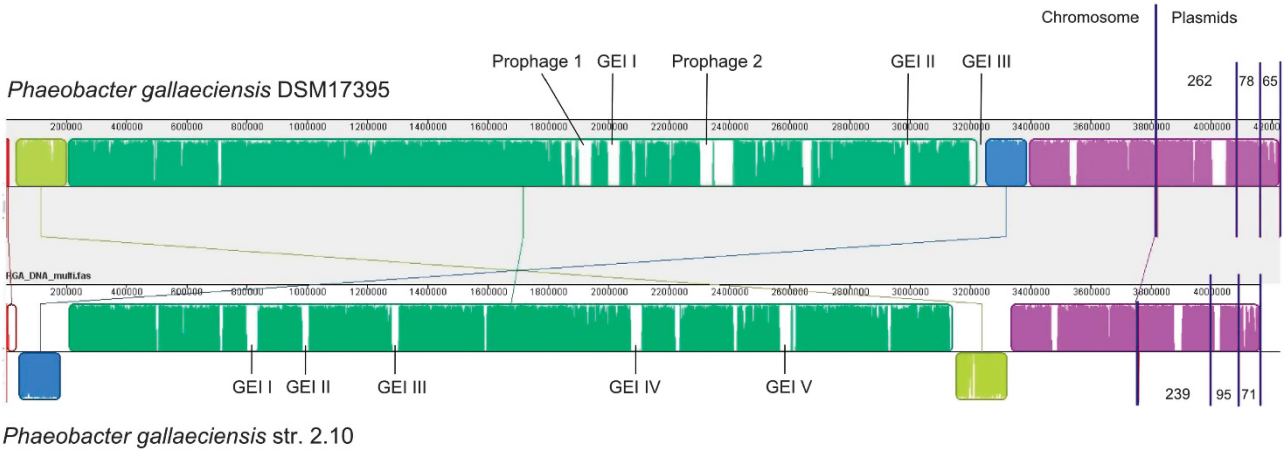

Figure 1 Whole-genome alignment of *P. gallaeciensis* strains DSM 17395 and 2.10. The progressive MAUVE alignment tool identifies nucleotide matches, which are depicted as linked boxes with the same color. Inverted regions are identified by boxes below the centre line. The four replicons are separated by red lines. Non-matching regions that were identified as GEI or prophage are marked and labeled.

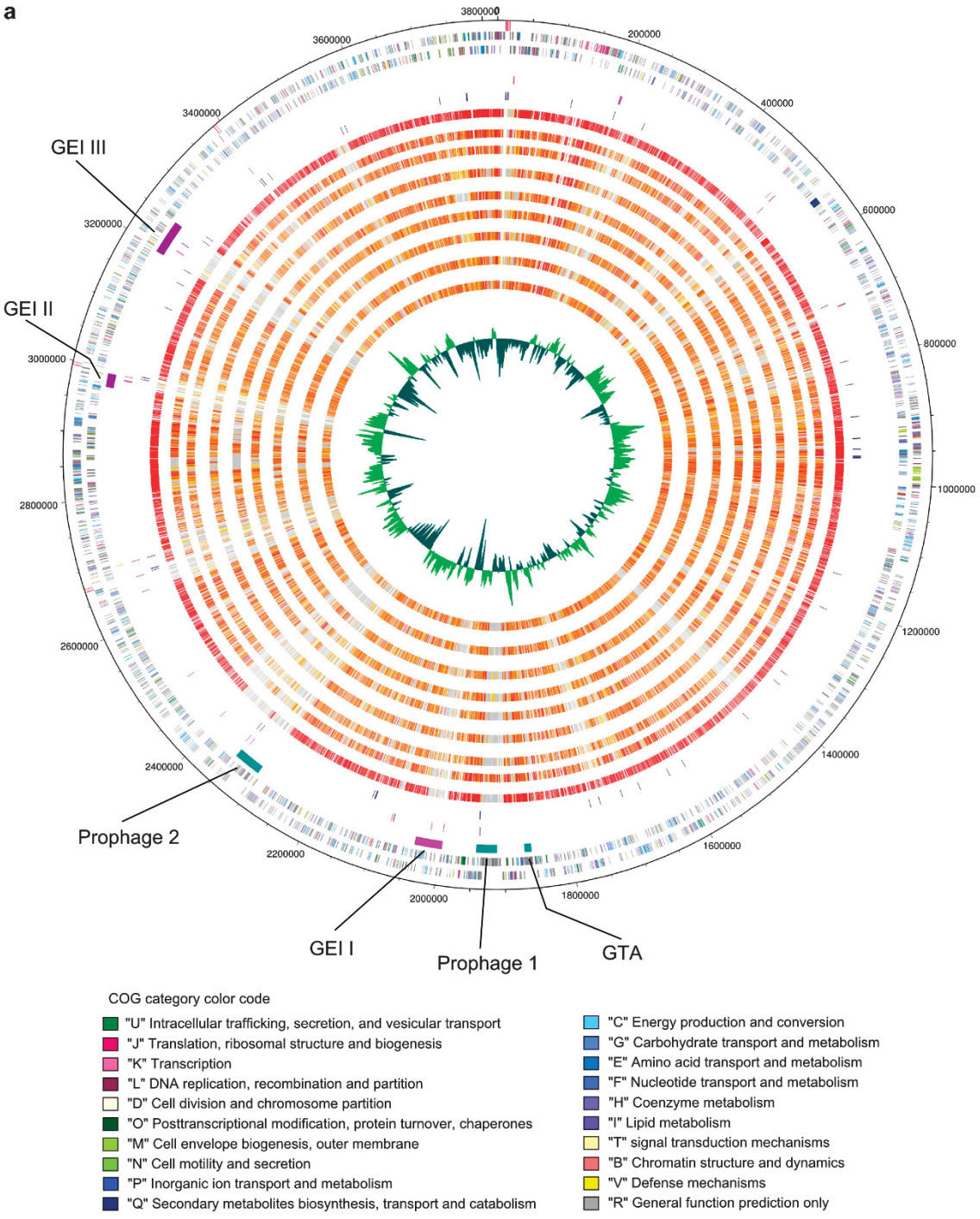

Figure 2 (**a**) Circular representation of the genome of *P. gallaeciensis* DSM 17395. Circles (from outside to inside): 1: rRNA cluster (red); 2 and 3: open ORFs on the leading and lagging strand, respectively. ORFs are colored according to the clusters of orthologous group (COG) categories; 4: regions that were identified as GEI (GEI I-III; purple), as prophages (turquoise) or as GTA (turquoise); 5: transposases (pink); 6: tRNAs (blue); circles 7–15 represent orthologous ORFs in members of the *Roseobacter* clade: *P. gallaeciensis* strain 2.10, *Ruegeria* spp. TM1040, *Roseobacter litoralis* Och149, *Sulfitobacter* NAS-14.1, *Roseovarius nubinhibens* ISM, *Sagittula stellata* E-37, *Octadecabacter arcticus* 238, *Roseobacter* spp. CCS2, *Dinoroseobacter shibae* DFL 12. The similarity of orthologous genes (based on the Needleman–Wunsch algorithm) is indicated by differend shades of red, with red bars representing ORFs with best conformity to the corresponding ORF in strain DSM 17395. Gray bars indicate that no orthologous gene exists in the respective organism and 16: G+C content of the chromosome of *P. gallaeciensis* DSM 17395; dark green: below average, light green: above average. (**b**) (see next page) Circular representation of the plasmids of *P. gallaeciensis* DSM 17395. Circles (from outside to inside): 1 and 2: open ORFs on the leading and lagging strand, respectively. ORFs that encode functions of interest in this study are colored: red, TDA biosynthesis; brown, siderophore synthesis; green, cell envelope/outer-membrane synthesis. Circle 3: orthologous ORFs in *P. gallaeciensis* strain 2.10. The similarity of orthologous genes (based on the Needleman–Wunsch algorithm) is indicated by different shades of red, with red bars representing ORFs with best conformity to the corresponding ORF in strain DSM 17395. Gray bars indicate that no orthologous gene exists in the respective organism. Circle 4: G+C content of the plasmid: purple, below average and green, above average.

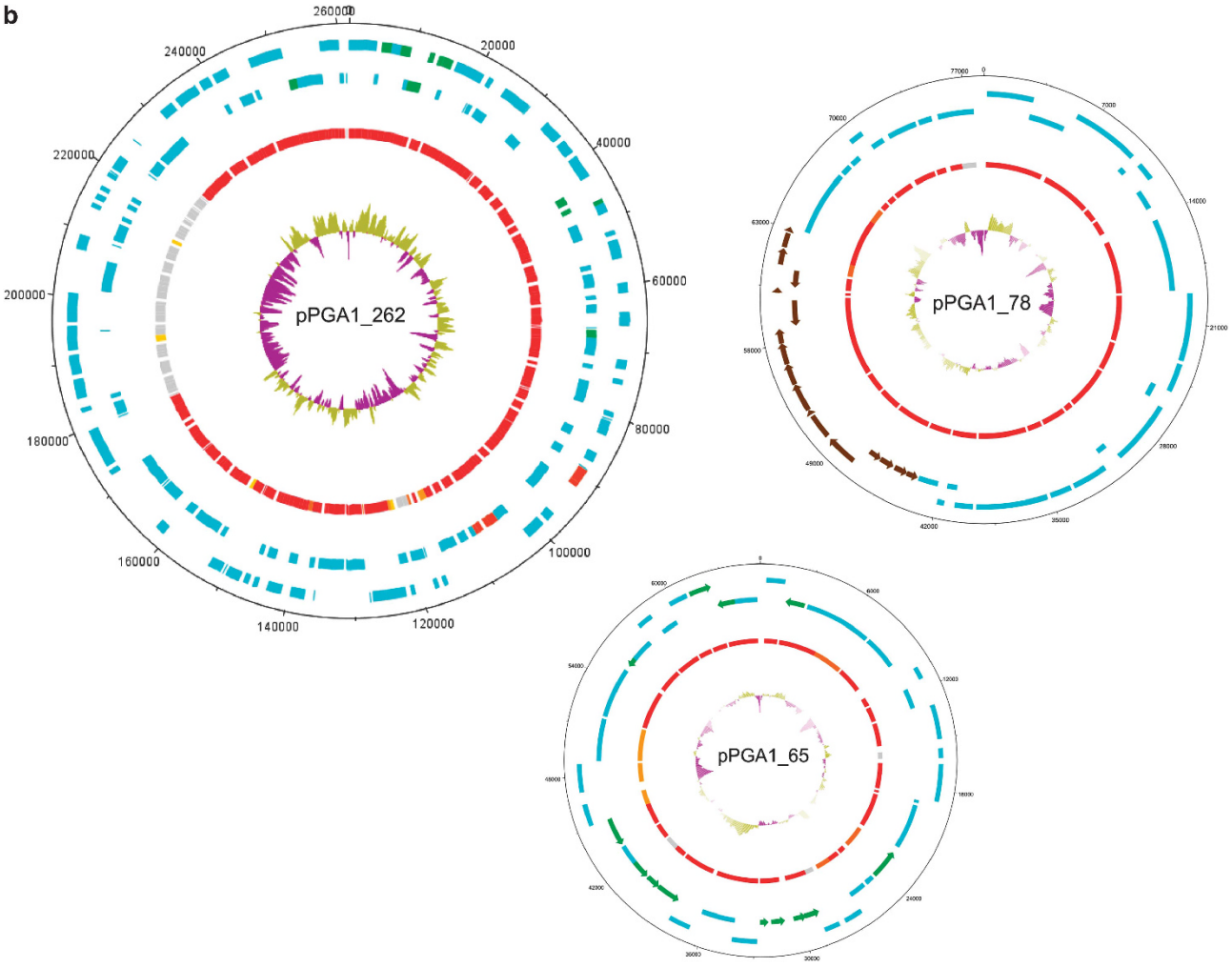

Figure 2 Continued.

as transposases or integrases, and their G+C content deviates clearly from the mean value. Several genes in the GEIs encode for transport systems as well as enzymes with regulatory functions. In GEI II of strain DSM 17395, a two-component system was found, which probably regulates the expression of the adjacent NRPS gene. The third island of strain DSM 17395 encodes a complete type I restriction-modification system (RMS) (PGA1_c30690-_c30710), probably protecting the cell from invasion of foreign DNA. RMSs may also cause cell death through restriction of genomic DNA or are involved in genomic rearrangements (Kobayashi, 2001). Such RMS genes are also located in GEI I of strain 2.10, and a putative anti-restriction protein (PGA2_c11710; ArdA-like) in GEI III might modulate the activity of the type I RMS (Nekrasov *et al.*, 2007). Other genes found in GEIs are likely involved in cell envelope synthesis (*rfbFGH*; PGA2_c18820-_c18840; GEI IV), resistance to heavy metals (PGA2_c08910-PGA2_c08950; GEI II) and metabolism of amino acids and amines (most genes located in GEI V; for example, PGA2_c23460-PGA2_c23500).

Two complete prophages are encoded in the genome of strain DSM 17395 (Supplementary Material S5). Inducible prophages have been found in several marine bacteria (Williamson *et al.*, 2002; Mobberley *et al.*, 2008) and are known to influence structure, biomass and genetic diversity of marine bacterial communities (Fuhrman, 1999). The genome of *P. gallaeciensis* DSM 17395 encodes an additional prophage-like element, which strongly resembles a so-called GTA (Supplementary Material S5), a virus-like particle that mediates transfer of genomic DNA between prokaryotes without negative effects on the host cell (Lang and Beatty, 2001). Comparative genome analysis revealed that GTA-like structures are widespread among genomes of organisms affiliated with the *Rhodobacterales* (Lang and Beatty, 2007). The GTA cluster has a length of ~15 kb (18 ORFs), is identically organized in the two *Phaeobacter* strains and exhibits a high similarity to GTAs of other *Rhodobacterales* species, for example, *Silicibacter* spp. TM1040 or *Rhodobacter capsulatus* (Lang and Beatty, 2007). The other two prophages of strain DSM 17395 consist of 48 ORFs (prophage 1; ~33.8 kb) and 39 ORFs (prophage 2;

~39 kb), respectively. Prophage 1 shares similarity with prophage 3 present in the genome of *Silicibacter* spp. TM1040. Several characteristic ORFs are present in both prophage genomes, namely structural elements like tail-, head- and baseplate-assembly proteins (Supplementary Materials S5 and S6). The coding sequences of prophage 1 are located upstream of a $tRNA^{LEU}$, a common integration site for bacteriophages (Cheetham and Katz, 1995). The flanking regions are identical in strain DSM 17395 and 2.10, thus indicating a non-disruptive integration of the prophage. The same applies to prophage 2, where the ORF encoding the phage integrase (PGA1_c22580) is located next to a $tRNA^{GLY}$.

Except for the GTA-like locus, no other complete prophage was found in the genome of *P. gallaeciensis* 2.10. To confirm these differences in prophage content for strains DSM 17395 and 2.10, we treated cultures of both with mitomycin C. As expected from the genomic data, no induction of prophages was observed for strain 2.10 (data not shown). Shortly after exposure to mitomycin C, strain DSM 17395 was inhibited in growth for a period of approximately 8 h (Figure 3). The number of virus-like particles in the non-induced control steadily increased from 10^6 to 10^8 per ml within 14 h and remained at that level. Virus-like particle numbers in the mitomycin-C-treated culture increased rapidly from 10^6 to a maximum of 10^9 per ml within 20 h post induction during which time the bacterial growth also nearly stayed stagnant. Presence of the free-living forms of both prophages in the lysate was confirmed by PCR with primers targeting the terminases and subsequent sequencing of the PCR products (data not shown). Pulsed field gel electrophoresis analysis of the lysates showed two bands of approximately 35 and 70 kb (Figure 3). The 35-kb band corresponds well to the predicted genome size of prophage 1, whereas the 70-kb band exceeds the expected size of prophage 2. The difference in size might be due to a concomitant excision of host genomic DNA and could include parts of the 62-kb upstream of prophage 2, which are unique in strain DSM 17395 compared with strain 2.10 (Figure 2a).

Together, these data show clear differences in the active prophage profile of strains DSM 17395 and 2.10, which might define their abilities to exchange genetic material via generalized transduction. The lack of the two prophages might also explain the lower number of unique genes in strain 2.10 compared with strain DSM 17395.

Unique genes in P. gallaeciensis compared with other Roseobacter clade genomes

The two *Phaeobacter* genomes possess 74 orthologous genes that are not present in other *Roseobacter* clade organisms using a 30% similarity cut-off based on a Needleman–Wunsch alignment (Supplementary Material S7). The majority of these genes have no functional annotation; however, two functions can potentially be used as unique chemotaxonomic markers for the species. These two gene markers were not found in the ANG1 draft genome, further supporting their potential uniqueness to *P. gallaeciensis*. The first refers to two copies of a

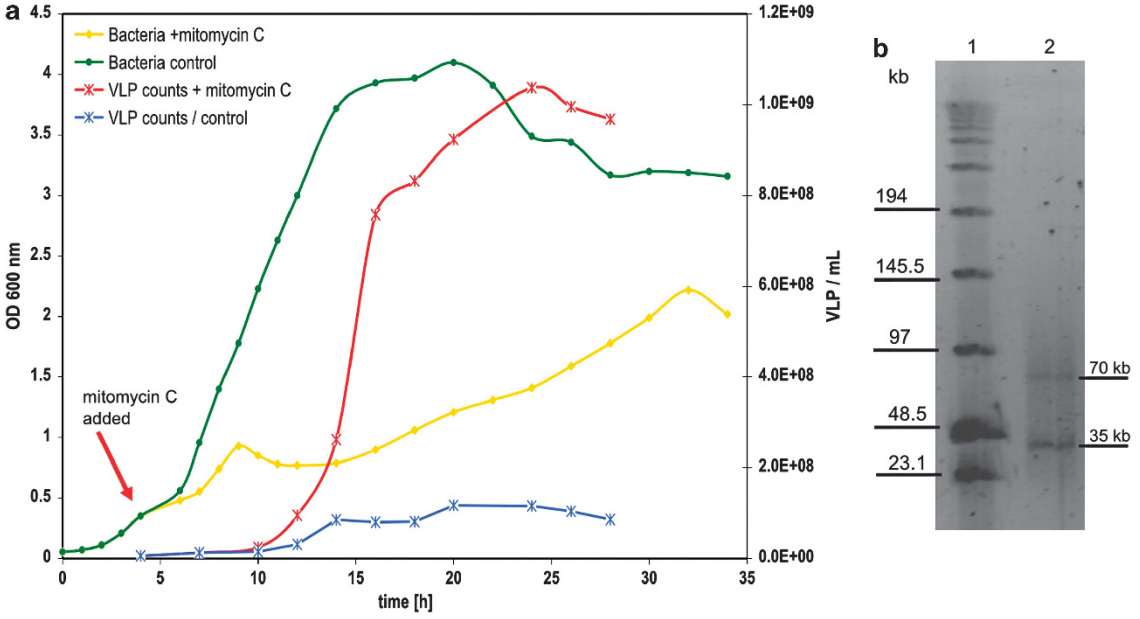

Figure 3 Induction of prophages in P. gallaeciensis DSM 17395. (**a**) Growth of *P. gallaeciensis* DSM 17395 and virus-like particle (VLP) counts in bacterial cultures treated with mitomycin C (0.5 μg ml^{-1}) and untreated controls. Bacterial growth was determined by measuring the optical density at 600 nm The VLP yield was also monitored (by fluorescence microscopy) in a mitomycin-treated and control culture. (**b**) Pulsed field gel electrophoresis of prophage DNA isolated from the induced viral lysate. Lane 1: size marker (λ, low range size marker; New England Biolabs, Ipswich, MA, USA) and lane 2: prophage DNA.

chromosomally encoded D-alanine-poly(phosphoribitol) ligase (*dltA*), which is involved in biosynthesis of D-alanyl-lipoteichoic acid. This compound is a well-known constituent in the cell wall of Gram-positive bacteria (Neuhaus and Baddiley, 2003) and has been reported to have a role in surface attachment of *Staphylococcus aureus* (Gross *et al*., 2001). Teichoic acid was also found in the Gram-negative bacterium *Sulfitobacter brevis* KMM 6006 (Gorshkova *et al*., 2007), which belongs to the *Roseobacter* clade, too. The *dltA* genes are part of a larger cluster of 19 *Phaeobacter* unique genes (PGA1_c13610-PGA1_c13830 in DSM 17395). PFAM (protein family) motifs were identified in several of the proteins that are present in the *dlt* operon of Gram-positive bacteria (Glaser, 1995). Presence of these gene clusters in the two *Phaeobacter* strains indicates an uncommon cell envelope composition.

The second unique genomic feature is a cluster for biosynthesis and transport of an iron-chelating siderophore, located on the plasmids pPGA1_78 and pPGA2_95 (Figure 4, Supplementary Material S7). A Chrome Azurol S assay showed that both *Phaeobacter* strains excrete siderophores under iron-limiting conditions, whereas *Roseobacter denitrificans*, which lacks the gene cluster, shows no excretion (Figure 4). The organization of the gene cluster is identical to that of the *asbABCDEF* petrobactin biosynthesis locus of *Bacillus anthracis* and *Marinobacter aquaeoeli* VT8 (Pfleger *et al*., 2007; Homann *et al*., 2009). The only difference is that *asbC* and *asbD* are fused in *P. gallaeciensis* (Figure 4). Consequently, the siderophore produced by *P. gallaeciensis* might be related to petrobactin. Upstream of the biosynthetic enzymes, the genes PGA1_78p00510, 78p00520 and 78p00530 encode two RND-type multidrug exporter and a putative outer-membrane protein similar to proteins mediating export of the pyoverin siderophores (Poole *et al*., 1993). In addition, two systems for the import of Fe^{3+}-siderophore complexes were found in the *Phaeobacter* genomes. One system is located on the chromosome (PGA1_c25890-PGA1_c25980) and comprises an ABC transporter and an energy-transducing system composed of TonB and ExbBD, which are typical for a ferric enterobactin transport system (Shea and McIntosh, 1991). The other system is plasmid-encoded and located adjacent to the siderophore biosynthesis cluster (PGA1_78p00360-PGA1_78p00390, Figure 4). The genes in this cluster resemble those of the *B. anthracis fatBCD/fhuC*

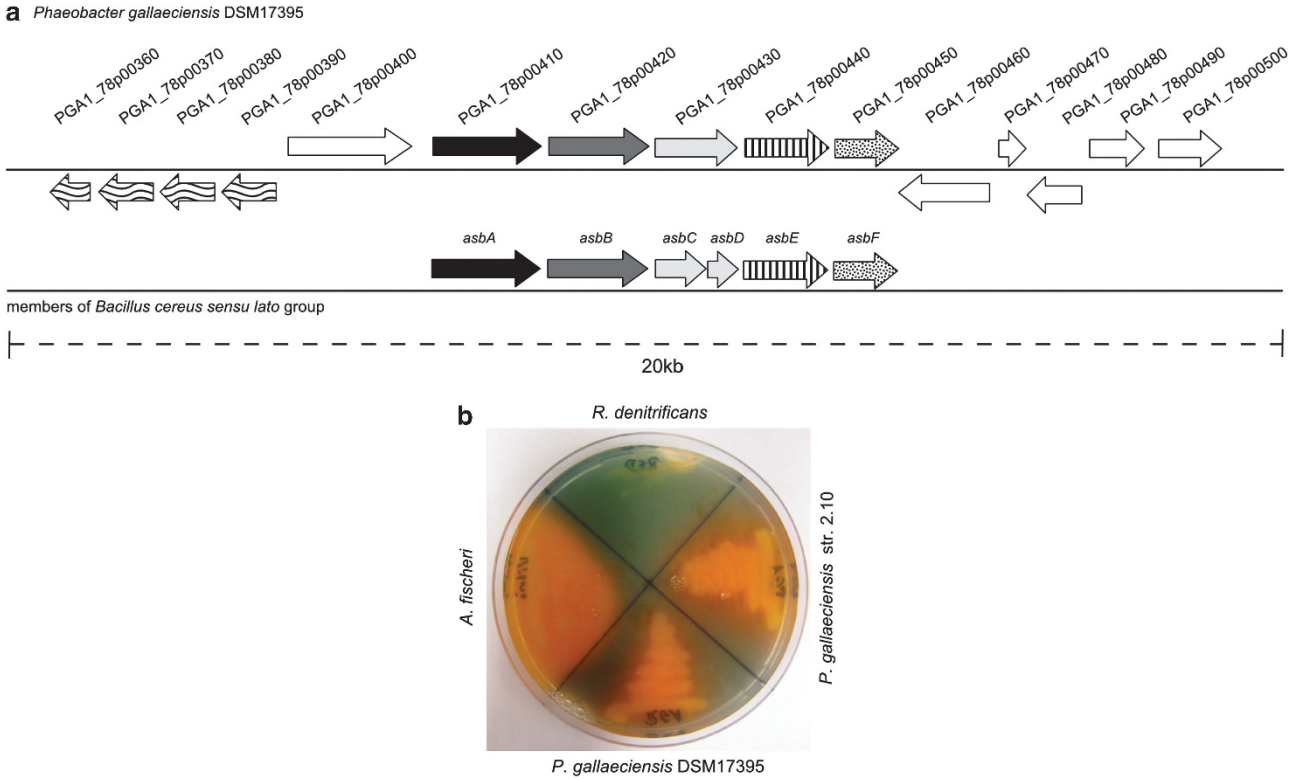

Figure 4 Siderophore biosynthesis gene cluster and determination of siderophore production in *P. gallaeciensis*. (**a**) The *P. gallaeciensis* plasmid-encoded siderophore biosynthesis cluster and the syntenic petrobactin synthesis cluster *asbABCDEF* of the *Bacillus cereus* group. Homology between genes of the organisms is indicated by shaded or hatched areas. The gene cluster is identical in both *Phaeobacter* strains (**b**) Chrome Azurol S (CAS) assay for the determination of siderophore production. The strains were grown on iron-depleted medium and overlaid with the blue CAS solution. Excreted siderophores remove iron from the CAS complex and the color changes from blue to orange.

cluster for petrobactin uptake (Hotta *et al.*, 2010). The presence of several siderophore transport systems might confer an advantage, as iron-siderophore complexes produced by other organisms can be acquired without the need for energy-consuming siderophore synthesis by the organism itself. It has also been shown that some marine bacteria produce siderophores facilitating iron uptake of the algal host (Soria-Dengg *et al.*, 2001; Amin *et al.*, 2009), which may also apply to *P. gallaeciensis*.

Environmental adaptation of Phaeobacter

P. gallaeciensis is known to be an effective colonizer of marine surfaces and has the ability to outcompete other microbiota, possibly due to the production of TDA (Rao *et al.*, 2006; Porsby *et al.*, 2008). A survey of the isolation sources of other *P. gallaeciensis* strains (16S rRNA gene identity $\geqslant$99%) clearly reflects a surface-associated lifestyle of these bacteria (Supplementary Material S8). Thus, the genomes were analyzed for features supporting this lifestyle strategy.

Attachment to and interaction with surfaces

The *Phaeobacter* genomes harbor several genes sharing similarity with the *exo* genes of *Sinorhizobium meliloti* (Supplementary Material S9), which mediate production of extracellular polysaccharides. Extracellular polysaccharides represent a major factor contributing to surface attachment (Danhorn and Fuqua, 2007; Vu *et al.*, 2009). Variation in its composition, as indicated by unique genes and gene combination, might facilitate host-specific association of *P. gallaeciensis* (Skorupska *et al.*, 2006). Most of these genes are located on the plasmids of the *Phaeobacter* strains (pPGA1_262, pPGA1_65 and pPGA2_239, pPGA2_71) (Supplementary Material S3). Some homologous genes occur on plasmids in several other *Roseobacter* clade members (Kalhoefer *et al.*, 2011); however, some genes are *P. gallaeciensis* unique (Supplementary Material S7). These include a glycosyltransferase-like protein (PGA1_65p00370), putatively involved in polysaccharide biosynthesis, and two ORFs (PGA1_65p00320 and 65p00330) related to a type I secretion system, which is known to export exopolysaccharides in *S. meliloti* (Moreira *et al.*, 2000). Moreover, gene clusters were found that share homology with some *rkp* genes of *S. meliloti* (Supplementary Material S9), which are required for production and export of capsular polysaccharides (Kereszt *et al.*, 1998). Similar findings were recently reported for *Roseobacter litoralis, R. denitrificans* and several other roseobacters, where plasmid localization of *rfb* genes was shown to correlate with a host-associated lifestyle (Kalhoefer *et al.*, 2011).

Despite this genomic evidence for a surface-attached lifestyle (Supplementary Material S13), colonization of *P. gallaeciensis* has only been investigated for one natural surface type, the green alga *U. lactuca* (Rao *et al.*, 2007). Therefore, green fluorescent protein (GFP)-labeled cells of

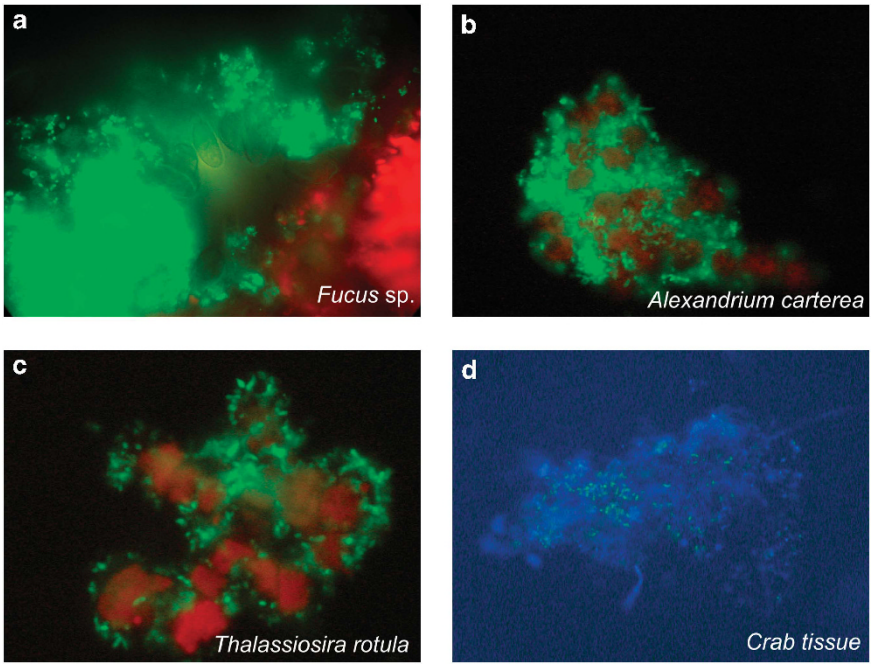

Figure 5 Colonization of green fluorescent protein (gfp)-labeled *P. gallaeciensis* DSM 17395 cells. Fluorescence microscopic images of gfp-labeled *P. gallaeciensis* cells proliferating on different surfaces. (**a**) Tissue sample of the macro alga *Fucus* spp. (**b**) Aggregated cells of the dinoflagellate *Alexandrium carterea.* (**c**) Culture of the diatom *Thalassiosira rotula.* (**d**) Slice of crab tissue magnification $\times$1000.

P. gallaeciensis DSM 17395 were added to a variety of micro- and macroalgae, crustacean tissue and driftwood. Growth and colonization, albeit weak, was observed on *Ulva* spp., driftwood and *Rhodomonas baltica*. Strongest growth and colonization was observed on *Fucus vesiculosus*, *Alexandrium carterea*, *Thalassiosira rotula* and crab tissue (Figure 5a–c). *P. gallaeciensis* formed a number of microcolonies on crab tissue, whereas dense biofilms were established on the surface of *F. vesiculosus*. Planktonic *Phaeobacter* cells that were not attached to a surface exhibited a tendency to form aggregates. The latter was also found when GFP-labeled *P. gallaeciensis* cells grew in co-culture with *T. rotula*. Proliferation of *P. gallaeciensis* cells seemed to be initiated by damage of the microalgae, which is consistent with previous observations that colonization by *Phaeobacter* spp. is dependent on the physiological state of *T. rotula* (Grossart *et al.*, 2005). Consistent with this is also our observation that strain DSM 17395 was able to form microcolonies on autoclaved samples of *R. baltica*, *U. lactuca* and *Ulva intestinalis*, which released a number of breakdown products and metabolites into the medium that would facilitate bacterial growth.

Nutrient utilization of P. gallaeciensis

The *in vitro* tests for growth of *P. gallaeciensis* on several carbohydrates and amino acids used as single-carbon sources (Ruiz-Ponte *et al.*, 1998) are in general agreement with the genomic predictions (Supplementary Material S10). Both *Phaeobacter* strains grew well with creatine, sarcosine and taurine and the key enzymes for their degradation were found in the genomes (Supplementary Materials S10 and 11). One enzyme in the degradation pathway of putrescine could not be predicted from the genome, which was also reported for *R. litoralis* (Kalhoefer *et al.*, 2011). However, because growth was observed on this substance, the missing reaction is likely to be complemented by a yet unknown enzyme.

N,N-dimethylglycine was neither used as a carbon nor as a nitrogen source, even though several ORFs coding for dimethylglycine dehydrogenases, highly similar to mitochondrial enzymes, were found in the genomes. These dehydrogenases might mediate the decrease of intracellular *N,N*-dimethylglycine produced from the degradation of choline and glycine betaine (Supplementary Material S11). Only weak growth was observed with choline or glycine betaine as sole source of carbon (Supplementary Material S10). Enzymes catalyzing the synthesis of glycine betaine from choline-O-sulfate and choline were found in both genomes (*betABC*). The enzyme converting glycine betaine to *N,N*-dimethylglycine (betaine-homocysteine methyltransferase, BHMT), however, could not be annotated. Only one ORF (PGA1_c13370/PGA2_c13310) was found in each of the *Phaeobacter* genomes encoding a homocysteine S-methyltransferase domain (PF02574), which is required for the enzymatic activity of BHMT (Garrow, 1996). These ORFs share homology with the *metH* gene of *Escherichia coli* and with the *S. meliloti* betaine-homocysteine methyltransferase gene *bmt*, which represents a link between methionine biosynthesis and glycine betaine degradation in this organism (Barra *et al.*, 2006). The *Phaeobacter* ORFs have only one-third of the size of *metH* and lack the cobalamin binding, the activation and the pterin-binding domains. Nevertheless, the vitamin B12-binding domains are encoded in ORFs (PGA1_c13350/PGA2_c13290) in the vicinity. The activation domain might be substituted by PGA1_c13360 (PGA2_c13300) that encodes a radical SAM domain (PF04055), which is thought to catalyze unusual methylation reactions (Sofia *et al.*, 2001). Together, these three ORFs might code for enzymes that substitute both MetH, known from *E. coli*, and Bmt from *S. meliloti*. These genes and their arrangement are also present in other *Roseobacter* species and might explain why glycine betaine and choline are only used to some extent in the *Phaeobacter* strains, namely as a byproduct during the synthesis of methionine.

Both genomes lack enzymes for the utilization of creatinine, an important and abundant energy-transfer molecule in eukaryotic tissue. The inability to use this compound as a carbon source was experimentally confirmed (Supplementary Material S10). Nevertheless, creatinine could be used as N-source due to the unspecific reaction of a cytosine deaminase (Wyss & Kaddurah-Daouk, 2000), present in both *Phaeobacter* genomes (PGA1_c26940/PGA2_c24980).

TDA biosynthesis in P. gallaeciensis DSM 17395

In order to elucidate the pathways leading to TDA synthesis in strain DSM 17395, we conducted transposon insertion mutagenesis and screened for mutants with reduced pigmentation, which was shown to correlate with TDA production (Brinkhoff *et al.*, 2004; Geng *et al.*, 2008; Berger *et al.*, 2011). Within the framework of this study, we identified 26 genes involved in TDA biosynthesis and pigment production, of which 18 are essential (Table 2). The identified genes are located in a cluster on pPGA1_262, including the well-known key TDA production genes *tdaABCEF* (Geng *et al.*, 2008), the newly discovered *paaZ2* and a putative Na-dependent transporter. The remaining 19 genes are scattered over the chromosome and affiliated with different pathways of the primary metabolism (for example, phenylacetate and sulfur metabolism). A proposed model for the biosynthesis of TDA in DSM 17395 is presented in Figure 6, combining the results of this study with those of other publications (Thiel *et al.*, 2010; Teufel *et al.*, 2011). The Paa enzymes of the phenylacetyl-CoA pathway produce compound 8 (Figure 6), which has been proposed as

Table 2 List of mutants derived from Tn5 mutagenesis of *P. gallaeciensis* DSM 17395

Strain	*Genotype*	*Phenotype*		*Locus tag*	*Gene*	*Function*	*Homologous ORF in* P. gallaeciensis *str. 2.10*	*Source or reference*
paa + ior1		TDA	PIGMENT					
CP26	C3 *paaJ*::EZTn5, Km^r Cm^r Gm^r	(+)	—	PGA1_c04090	*paaJ*	Beta-ketoadipyl-CoA thiolase	PGA2_c03640	This study
CP51	C3 *paaA*::EZTn5, Km^r Cm^r Gm^r	—	—	PGA1_c04080	*paaA*	ring-1,2-phenylacetyl-CoA epoxidase	PGA2_c03630	Berger *et al*., 2012
WP3	DSM 17395 *paaC*::EZTn5, Gm^r	—	—	PGA1_c04060	*paaC*	ring-1,2-phenylacetyl-CoA epoxidase	PGA2_c03610	Berger *et al*., 2012
CP65	C3 *paaD*::EZTn5, Km^r Cm^r Gm^r	—	—	PGA1_c04050	*paaD*	ring-1,2-phenylacetyl-CoA epoxidase	PGA2_c03600	Berger *et al*., 2012
CP33	C3 *paaE*::EZTn5, Km^r Cm^r Gm^r	—	—	PGA1_c04040	*paaE*	ring-1,2-phenylacetyl-CoA epoxidase	PGA2_c03590	Berger *et al*., 2012
CP4	C3 *paaZ2*::EZTn5, Km^r Cm^r Gm^r	(+)	—	PGA1_262p00800	*paaZ2*	enoyl-CoA hydratase	PGA2_239p0790	This study
WP56	DSM 17395 *iorAB1*::EZTn5, Gm^r	(+)	(+)	PGA1_c04490	*ior1*	Indole pyruvate oxidoreductase	PGA2_c04040	Berger *et al*., 2012
Regulation								
WP52	DSM 17395 *raiR*::EZTn5, Gm^r	—	—	PGA1_c03880	*pgaR*	Transcriptional activator protein		Berger *et al*., 2011
CP20	C3 PGA1_c02470::EZTn5, Km^r Cm^r Gm^r	—	—	PGA1_c02470		Putative serine-protein kinase, prkA type	PGA2_c02060	This study
WP69	DSM 17395 PGA1_c02470::EZTn5, Gm^r	(+)	(+)	PGA1_c04480	*iorR*	Transcriptional regulator, tetR family	PGA2_c04030	Berger *et al*., 2012
WP37	DSM 17395 PGA1_c02470::EZTn5, Gm^r	—	—	PGA1_c20730		Transcriptional regulator, AsnC family	PGA2_c19570	This study
CP21	C3 PGA1_c02850::EZTn5, Km^r Cm^r Gm^r	(+)	—	PGA1_c02850		Putative signal peptide peptidase	PGA2_c02440	This study
WP75	C3 *tdaA*::EZTn5, Km^r Cm^r Gm^r	—	—	PGA1_262p00980	*tdaA*	Transcriptional regulator, LysR family	PGA2_239p0970	Berger *et al*., 2011
Sulfur metabolism								
WP50	DSM 17395 *cobA2*::EZTn5, Gm^r	—	—	PGA1_c20740	*cobA2*	Uroporphyrinogen-III C-methyl transferase	PGA2_c19580	This study
WP45	DSM 17395 *cysI*::EZTn5, Gm^r	—	—	PGA1_c20760	*cysI*	Putative nitrite/sulfite reductase	PGA2_c19600	This study
WP73	DSM 17395 *sat/cysC*::EZTn5, Gm^r	—	—	PGA1_c24800	*cysC, sat*	Bifunctional SAT/APS kinase	PGA2_c22800	This study
CP58	C3 *patB*::EZTn5, Km^r Cm^r Gm^r	—	—	PGA1_c00860	*patB*	Cystathionine beta-lyase	PGA2_c30030	This study
Transport								
CP22	C3 PGA1_262p00850::EZTn5, Km^r Cm^r Gm^r	(+)	(+)	PGA1_262p00850		putative sodium-dependent symporter	PGA2_239p0840	This study
WP44	DSM 17395 PGA1_c20660::EZTn5, Gm^r	—	—	PGA1_c20660		TRAP transporter, DctM subunit	PGA2_c19500	This study
Other proteins								
CP14	C3 PGA1_c04310::EZTn5, Km^r Cm^r Gm^r	(+)	—	PGA1_c04310		Protein of unknown function (DUF1045)	PGA2_c03860	This study
WP49	DSM 17395 *thiG*::EZTn5, Gm^r	—	—	PGA1_c09070	*thiG*	Thiazole biosynthesis protein	PGA2_c08800	This study
CP103	C3 PGA1_c23490::EZTn5, Km^r Cm^r Gm^r	—	—	PGA1_c23490		NADH:flavin oxidoreductase/NADH oxidase, containing additional FAD-binding domain	PGA2_c21440	This study
tda *genes*								
CP15	C3 *tdaB*::EZTn5, Km^r Cm^r Gm^r	—	—	PGA1_262p00970	*tdaB*	ß-etherase , GST	PGA2_239p0960	This study
CP29	C3 *tdaC*::EZTn5, Km^r Cm^r Gm^r	(+)	—	PGA1_262p00960	*tdaC*	Prephenate dehydratase domain protein	PGA2_239p0950	This study
WP14	DSM 17395 *tdaE*::EZTn5, Gm^r	—	—	PGA1_262p00940	*tdaE*	Acyl-CoA dehydrogenase	PGA2_239p0930	This study
CP18	C3 *tdaF*::EZTn5, Km^r Cm^r Gm^r	—	—	PGA1_262p00810	*tdaF*	Putative flavoprotein, HFCD family	PGA2_239p0800	This study

Cm^r, chloramphenicol resistance; C3 = DSM 17395 Δ*paaK1*::km *ΔpaaK2*::cm (Berger *et al*, 2012); Gm^r, gentamicin resistance; Km^r, kanamycin resistance.

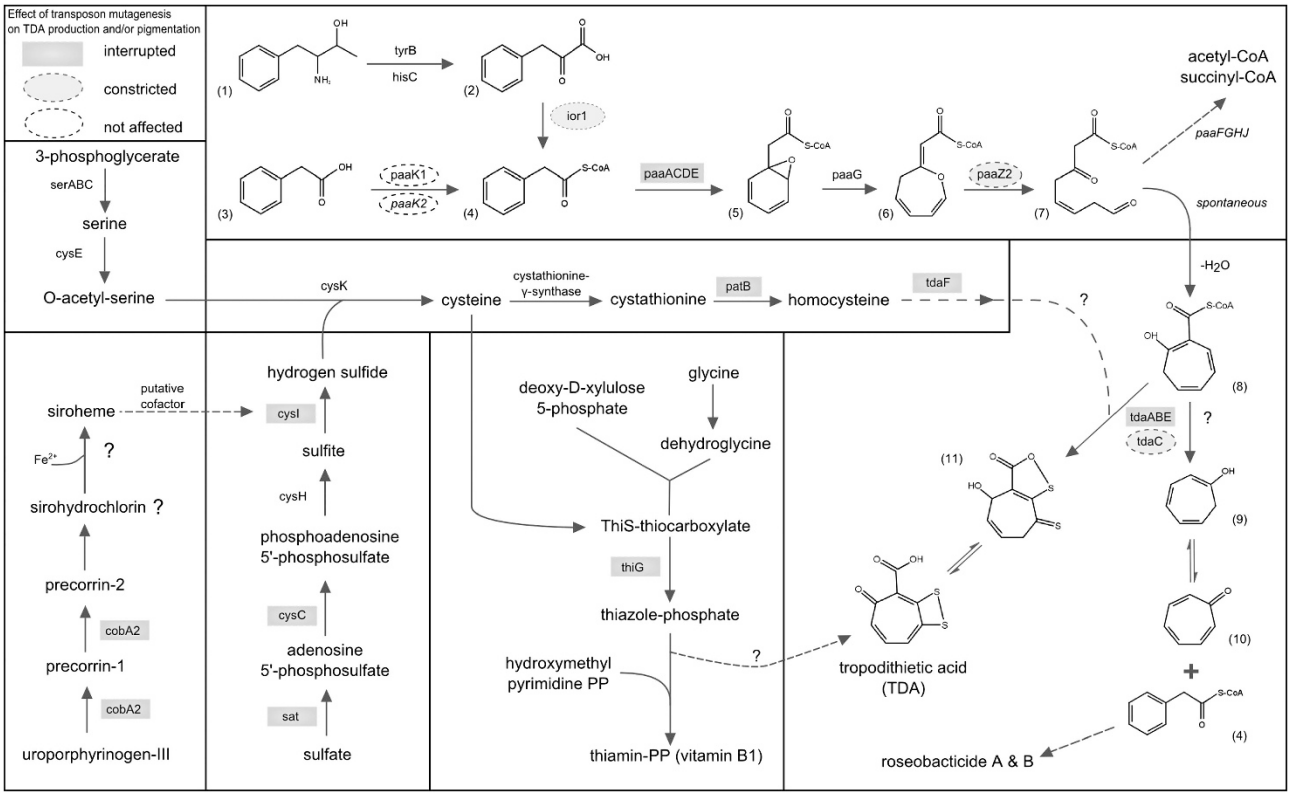

Figure 6 Proposed model for the biosynthesis of TDA in *P. gallaeciensis* DSM 17395. Integrated are the combined results of the transposon mutagenesis and the genome analysis presented in this study, as well as previously published data. For details, see text and references therein. Unknown reactions or ambiguities with respect to enzyme functions are indicated by question marks. Chemical structures: (1) phenylalanine; (2) phenylpyruvate; (3) phenylacetate; (4) phenyacetyl-CoA; (5) ring-1,2-epoxyphenylacetyl-CoA; (6) 2-oxepin-2(3H)-ylideneacetyl-CoA (oxepin-CoA); (7) 3-oxo-5,6-dehydrosuberyl-CoA semialdehyde; (8) 2-hydroxycyclohepta-1,4,6-triene-1-formyl-CoA; (9) tropolone; (10) tropone and (11) thiotropocin. Gene and protein names: *cobA2*: uroporphyrinogen-III C-methyltransferase; *cysE*: serine acetyltransferase; *cysH*: phosphoadenosine phosphosulfate reductase; *cysI*: putative sulfite reductase; *cysK*: cysteine synthase; *hisC*: histidinol-phosphate aminotransferase; *ior1*: indole pyruvate oxidoreductase (fused); *paaA*: ring-1,2-phenylacetyl-CoA epoxidase; *paaC*: ring-1,2-phenylacetyl-CoA epoxidase; *paaD*: ring-1,2-phenylacetyl-CoA epoxidase; *paaE*: ring-1,2-phenylacetyl-CoA epoxidase; *paaF*: 2,3-dehydroadipyl-CoA hydratase; *paaG*: ring-1,2-epoxyphenylacetyl-CoA isomerase (oxepin-CoA forming)/postulated 3,4-dehydroadipyl-CoA isomerase; *paaH*: 3-hydroxyadipyl-CoA dehydrogenase (NAD+); *paaJ*: 3-oxoadipyl-CoA/3-oxo-5,6-dehydrosuberyl-CoA thiolase; *paaK1*, *paaK2*: phenylacetate-CoA ligase; *paaZ2*: enoyl-CoA hydratase; *patB*: cystathionine beta-lyase; *sat/cysC*: putative bifunctional SAT/APS kinase; *serB*: phosphoserine phosphatase; *serC*: phosphoserine aminotransferase; *tdaA*: transcriptional regulator, LysR family; *tdaB*: ß-etherase; *tdaC*: prephenate dehydratase domain protein; *tdaE*: acyl-CoA dehydrogenase; *tdaF*: putative flavoprotein, HFCD family; *thiG*: thiazole biosynthesis protein and *tyrB*: aromatic amino-acid aminotransferase.

a precursor of TDA (Cane *et al.*, 1992; Teufel *et al.*, 2010; Teufel *et al.*, 2011). Disruption of the genes *paaACDE* led to complete loss of TDA production, which corresponds to data obtained for *Silicibacter* spp. TM1040 (Geng *et al.*, 2008).

The first enzyme of the phenylacetyl-CoA pathway *paaK1* (phenylacetyl-CoA ligase) was shown to be non-essential for TDA production and this part of the pathway is probably substituted by conversion of phenylalanine via phenylpyruvate to phenylacetyl-CoA through the enzymatic action of either TyrB or HisC and Ior1 (Berger *et al.*, 2012). Interruption of the plasmid-encoded *paaZ2* gene, whose product mediates ring cleavage (Teufel *et al.*, 2011), led to reduced TDA production. The remaining TDA synthesis is likely a result of partial complementation by the chromosomal PaaZ1 enzyme.

Phenylacetyl-CoA as precursor for TDA biosynthesis and tropolone as an intermediate product in the TDA synthesis were both recently proposed as precursor of the newly discovered roseobacticides (Seyedsayamdost *et al.*, 2011), which would link the synthesis of these secondary metabolites in *P. gallaeciensis*. The exact mechanism for the addition of sulfur to the TDA precursor has not been elucidated yet. Nevertheless, the assimilatory sulfate reduction pathway is essential for TDA biosynthesis, shown for *Silicibacter* spp. TM1040 by Geng *et al.* (2008) and confirmed for strain DSM 17395 in our analysis as a result of interruption of the sulfite reductase (*cysI*) and of the bifunctional SAT/APS kinase (*sat/cysC*) (Figure 6).

The enzyme PatB (cystathionine ß-lyase) converts cystathionine into homocysteine in *B. subtilis*

(Auger *et al.*, 2005) and was found to be essential for TDA synthesis in strain DSM 17395. Thus, we postulate that homocysteine acts as a source of sulfur atoms for TDA. In support of this, the essential *tdaF* gene most probably encodes a homo-oligomeric flavin-containing cysteine decarboxylase, which catalyzes the oxidative decarboxylation of cysteine residues for several substrates (Majer *et al.*, 2002). Therefore, TdaF could provide a reactive, sulfur-containing precursor for the final steps of TDA synthesis (Figure 6). Disruption of the *cobA2* gene, involved in the synthesis of siroheme, also led to a phenotype deficient in TDA production and we propose that siroheme or a derivative acts as prosthetic group of the essential sulfite reductase. *P. gallaeciensis* harbors two copies of genes coding for uroporphyrinogen-III methyltransferases (*cobA1*; PGA1_c08070 and *cobA2*; PGA1_c20740) with *cobA1* being co-located with genes involved in vitamin B12 synthesis and *cobA2* clustered with genes for siroheme synthesis. In addition, siroheme is known to act as prosthetic group of the sulfite reductase CysI (Murphy and Siegel, 1973). One TDA-negative mutant had an insertion in the *thiG* gene of the *de novo* synthesis pathway of thiamine (vitamin B1, Table 2). ThiG catalyzes formation of the thiazole phosphate ring (Park *et al.*, 2003). The *P. gallaeciensis thiG* mutant was unable to grow on mineral medium without thiamine, whereas addition of thiamine restored growth but not production of TDA. It is therefore unlikely that thiamine acts as a direct cofactor in TDA synthesis, but it is possible that an intermediate of the thiamine synthesis is involved in TDA synthesis.

In the present study, we also identified six genes involved in the regulation of TDA synthesis. Among those is *tdaA*, which was shown to regulate the expression of *tdaBEF and paaZ2* (Berger *et al.*, 2011). The gene coding for the regulatory protein IorR (PGA1_c04480) is located adjacent to *ior1* and was shown to be essential for the transcription of *ior1* and the phenylalanine catabolism (Berger *et al.*, 2012). The regulator PgaR (PGA1_c03880) is part of the PgaRI quorum-sensing system, which is essential for the transcription of *tdaA* (Berger *et al.*, 2011). Taken as a whole, our analysis confirmed previous results from other TDA-producing organisms and revealed further genes, which have previously been unknown to be involved in TDA metabolism.

Conclusions

The *P. gallaeciensis* strains analyzed in this study were isolated from different habitats (fish aquaculture and macroalgal thallus) at almost opposite locations of the globe (distance: ~18 000 km). However, their genomes exhibit a high level of synteny and genetic conformity. The first comparative analysis of two finished genomes of organisms of the *Roseobacter* clade, which goes beyond the species level, gives an insight into the core and flexible genome of the strains. Only subtle differences in the genetic equipment—mostly mediated by lateral gene transfer and prophages—point to strain-specific features for coping with individual environmental conditions. Several genomic traits seem to be more common among roseobacters, for example, TDA biosynthesis or quorum sensing. Only two genomic features likely delineate *P. gallaeciensis* from other roseobacters. Although these findings support the 'mix and match' hypothesis (Moran *et al.*, 2007) for the heterogeneity of genomic content of roseobacters, they also underline the importance of intra-species comparisons to detect genomic traits that could result in strain-level specificity.

Our survey also suggests strain differentiation due to changes in the regulatory network. The genome analyses support a surface-associated lifestyle of the two strains, which includes mechanisms to defend themselves against both, the host's innate immune response and other, competing bacteria. Several of these *Phaeobacter*-characteristic features, such as the biosynthesis of TDA, siderophore and extracellular polysaccharide production, are encoded on the plasmids, underlining the importance of mobile genetic elements in organisms of the *Roseobacter* clade. Association with eukaryotes is furthermore supported by the ability to grow on host-derived substances. Although a variety of biotic and abiotic surfaces is colonized by *P. gallaeciensis*, some exceptions imply that some degree of host specificity exists. The type of microbe–host interaction of *P. gallaeciensis* may be mutualistic as well as pathogenic, as suggested by Seyedsayamdost *et al.* (2011). The latter is supported by several virulence-associated features like the type IV secretion system or siderophore production and was also stated for other marine bacteria (Thomas *et al.*, 2008). The apparently highly defined habitat specificity of the two strains has selected for a very similar genomic design despite the great geographic distance.

Acknowledgements

We thank Frauke-Dorothee Meyer, Anne Buthoff (Göttingen) and Renate Gahl-Janßen (Oldenburg) for excellent technical assistance. We also thank Stephan Schuster for support during the 454 pyrosequencing. This study was supported by the niedersächsisches VW-Vorab 'Comparative and functional genome analysis of representative members of the *Roseobacter* clade' (ZN2235), Germany, the Collaborative Research Center TRR 51 funded by Deutsche Forschungsgemeinschaft, Center for Marine Bio-Innovation, Australia, and the Marine Microbiology Initiative of the Gordon and Betty Moore Foundation, USA.

References

Alexander DB, Zuberer DA. (1991). Use of chrome azurol S reagents to evaluate siderophore production by rhizosphere bacteria. *Biol Fert Soils* **12**: 39–45.

Amin SA, Green DH, Hart MC, Küpper FC, Sunda WG, Carrano CJ. (2009). Photolysis of iron-siderophore chelates promotes bacterial-algal mutualism. *Proc Natl Acad Sci USA* **106**: 17071–17076.

Auger S, Gomez MP, Danchin A, Martin-Verstraete I. (2005). The PatB protein of *Bacillus subtilis* is a C-S-lyase. *Biochimie* **87**: 231–238.

Balcázar JL, Lee NM, Pintado J, Planas M. (2010). Phylogenetic characterization and in situ detection of bacterial communities associated with seahorses (*Hippocampus guttulatus*) in captivity. *Syst Appl Microbiol* **33**: 71–77.

Barbieri E, Paster BJ, Hughes D, Zurek L, Moser DP, Teske A *et al.* (2001). Phylogenetic characterization of epibiotic bacteria in the accessory nidamental gland and egg capsules of the squid *Loligo pealei* (*Cephalopoda:Loliginidae*). *Environ Microbiol* **3**: 151–167.

Barra L, Fontenelle C, Ermel G, Trautwetter A, Walker GC, Blanco C. (2006). Interrelations between glycine betaine catabolism and methionine biosynthesis in *Sinorhizobium meliloti* strain 102F34. *J Bacteriol* **188**: 7195–7204.

Berger M, Brock NL, Liesegang H, Dogs M, Preuth I, Simon M *et al.* (2012). Genetic analysis of the upper phenylacetate catabolic pathway in the production of tropodithietic acid by *Phaeobacter gallaeciensis*. *Appl Environ Microbiol* **78**: 3539–3551.

Berger M, Neumann A, Schulz S, Simon M, Brinkhoff T. (2011). Tropodithietic acid production in *Phaeobacter gallaeciensis* is regulated by N-Acyl Homoserine Lactone-mediated Quorum Sensing. *J Bacteriol* **193**: 6576–6585.

Brinkhoff T, Bach G, Heidorn T, Liang L, Schlingloff A, Simon M. (2004). Antibiotic production by a *Roseobacter* clade-affiliated species from the German Wadden Sea and its antagonistic effects on indigenous isolates. *Appl Environ Microbiol* **70**: 2560–2565.

Brinkhoff T, Giebel HA, Simon M. (2008). Diversity, ecology, and genomics of the *Roseobacter* clade: a short overview. *Arch Microbiol* **189**: 531–539.

Bruhn JB, Nielsen KF, Hjelm M, Hansen M, Bresciani J, Schulz S *et al.* (2005). Ecology, inhibitory activity, and morphogenesis of a marine antagonistic bacterium belonging to the *Roseobacter* clade. *Appl Environ Microbiol* **71**: 7263–7270.

Cane DE, Wu Z, Van Epp JE. (1992). Thiotropocin biosynthesis. Shikimate origin of a sulfur-containing tropolone derivative. *J Am Chem Soc* **114**: 8479–8483.

Carson KC, Holliday S, Glenn AR, Dilworth MJ. (1992). Siderophore and organic acid production in root nodule bacteria. *Arch Microbiol* **157**: 264–271.

Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. (2009). DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* **25**: 119–120.

Cheetham BF, Katz ME. (1995). A role for bacteriophages in the evolution and transfer of bacterial virulence determinants. *Mol Microbiol* **18**: 201–208.

Chen F, Wang K, Stewart J, Belas R. (2006). Induction of multiple prophages from a marine bacterium: a genomic approach. *Appl Environ Microbiol* **72**: 4995–5001.

Collins AJ, Nyholm SV. (2011). Draft genome of *Phaeobacter gallaeciensis* ANG1, a dominant member of the accessory nidamental gland of *Euprymna scolopes*. *J Bacteriol* **193**: 3397–3398.

Danhorn T, Fuqua C. (2007). Biofilm formation by plant-associated bacteria. *Annu Rev Microbiol* **61**: 401–422.

Darling ACE, Mau B, Blattner FR, Perna NT. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* **14**: 1394–1403.

Fuhrman JA. (1999). Marine viruses and their biogeochemical and ecological effects. *Nature* **399**: 541–548.

Garrow TA. (1996). Purification, kinetic properties, and cDNA cloning of mammalian betaine-homocysteine methyltransferase. *J Biol Chem* **271**: 22831–22838.

Geng H, Bruhn JB, Nielsen KF, Gram L, Belas R. (2008). Genetic dissection of tropodithietic acid biosynthesis by marine roseobacters. *Appl Environ Microbiol* **74**: 1535–1545.

Glaser P. (1995). Incorporation of D-Alanine into lipoteichoic acid and wall teichoic acid in *Bacillus subtilis*. *J Biol Chem* **270**: 15598–15606.

Gorshkova RP, Isakov VV, Shevchenko LS, Ivanova EP, Denisenko VA, Nazarenko EL. (2007). Structure of teichoic acid from the marine proteobacterium *Sulfitobacter brevis* KMM 6006. *Chem Nat Compd* **43**: 643–647.

Gram L, Melchiorsen J, Bruhn JB. (2010). Antibacterial activity of marine culturable bacteria collected from a global sampling of ocean surface waters and surface swabs of marine organisms. *Mar Biotechnol* **12**: 439–451.

Grigioni S, Boucher-Rodoni R, Demarta A, Tonolla M, Peduzzi R. (2000). Phylogenetic characterisation of bacterial symbionts in the accessory nidamental glands of the sepioid *Sepia officinalis* (*Cephalopoda: Decapoda*). *Mar Biol* **136**: 217–222.

Gross M, Cramton SE, Götz F, Peschel A. (2001). Key role of teichoic acid net charge in *Staphylococcus aureus* colonization of artificial surfaces. *Infect Immun* **69**: 3423–3426.

Grossart HP, Levold F, Allgaier M, Simon M, Brinkhoff T. (2005). Marine diatom species harbour distinct bacterial communities. *Environ Microbiol* **7**: 860–873.

Hjelm M, Riaza A, Formoso F, Melchiorsen J, Gram L. (2004). Seasonal incidence of autochthonous antagonistic *Roseobacter* spp. and *Vibrionaceae* strains in a turbot larva (*Scophthalmus maximus*) rearing system. *Appl Environ Microbiol* **70**: 7288–7294.

Homann VV, Edwards KJ, Webb EA, Butler A. (2009). Siderophores of *Marinobacter aquaeolei*: petrobactin and its sulfonated derivatives. *Biometals* **22**: 565–571.

Hotta K, Kim CY, Fox DT, Koppisch AT. (2010). Siderophore-mediated iron acquisition in *Bacillus anthracis* and related strains. *Microbiology* **156**: 1918–1925.

Ivars-Martinez E, Martin-Cuadrado AB, D'Auria G, Mira A, Ferriera S, Johnson J *et al.* (2008). Comparative genomics of two ecotypes of the marine planktonic copiotroph *Alteromonas macleodii* suggests alternative lifestyles associated with different kinds of particulate organic matter. *ISME J* **2**: 1194–1212.

Kalhoefer D, Thole S, Voget S, Lehman R, Liesegang H, Wollherr A *et al.* (2011). Comparative genome analysis and genome-guided physiological analysis of *Roseobacter litoralis*. *BMC Genomics* **12.**

Karp PD, Ouzounis CA, Moore-Kochlacs C, Goldovsky L, Kaipa P, Ahrén D *et al.* (2005). Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. *Nucleic Acids Res* **33**: 6083–6089.

Karp PD, Paley S, Romero P. (2002). The Pathway Tools software. *Bioinformatics* **18**: S225–S232.

Kereszt a, Kiss E, Reuhs BL, Carlson RW, Kondorosi A, Putnoky P. (1998). Novel rkp gene clusters of *Sinorhizobium meliloti* involved in capsular polysaccharide production and invasion of the symbiotic nodule: the rkpK gene encodes a UDP-glucose dehydrogenase. *J Bacteriol* **180**: 5426–5431.

Kobayashi I. (2001). Behavior of restriction–modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Res* **29**: 3742–3756.

Lang AS, Beatty JT. (2001). The gene transfer agent of *Rhodobacter capsulatus* and ''constitutive transduction'' in prokaryotes. *Arch Microbiol* **175**: 241–249.

Lang AS, Beatty JT. (2007). Importance of widespread gene transfer agent genes in alpha-proteobacteria. *Trends Microbiol* **15**: 54–62.

Langille MGI, Brinkman FSL. (2009). IslandViewer: an integrated interface for computational identification and visualization of genomic islands. *Bioinformatics* **25**: 664–665.

Majer F, Schmid DG, Altena K, Bierbaum G, Kupke T. (2002). The flavoprotein MrsD catalyzes the oxidative decarboxylation reaction involved in formation of the peptidoglycan biosynthesis inhibitor mersacidin. *J Bacteriol* **184**: 1234–1243.

Martens T, Gram L, Grossart HP, Kessler D, Muller R, Simon M *et al.* (2007). Bacteria of the *Roseobacter* clade show potential for secondary metabolite production. *Microb Ecol* **54**: 31–42.

Martens T, Heidorn T, Pukall R, Simon M, Tindall BJ, Brinkhoff T. (2006). Reclassification of *Roseobacter gallaeciensis* Ruiz-Ponte et al. 1998 as *Phaeobacter gallaeciensis* gen. nov., comb. nov., description of *Phaeobacter inhibens* sp. nov., reclassification of *Ruegeria algicola* (Lafay et al. 1995) Uchino et al. 1999 as *Marinovum*. *Int J Syst Evol Microbiol* **56**: 1293–1304.

Mira A. (2002). Microbial genome evolution: sources of variability. *Curr Opin Microbiol* **5**: 506–512.

Mobberley JM, Authement RN, Segall AM, Paul JH. (2008). The temperate marine phage PhiHAP-1 of *Halomonas aquamarina* possesses a linear plasmid-like prophage genome. *J Virol* **82**: 6618–6630.

Moran MA, Belas R, Schell MA, González JM, Sun F, Sun S *et al.* (2007). Ecological genomics of marine roseobacters. *Appl Environ Microbiol* **73**: 4559–4569.

Moreira LM, Becker JD, Puhler A, Becker A. (2000). The *Sinorhizobium meliloti* ExpE1 protein secreted by a type I secretion system involving ExpD1 and ExpD2 is required for biosynthesis or secretion of the exopolysaccharide galactoglucan. *Microbiology* **146**: 2237–2248.

Murphy MJ, Siegel LM. (1973). Siroheme and sirohydrochlorin. The basis for a new type of porphyrin-related prosthetic group common to both assimilatory and dissimilatory sulfite reductases. *J Biol Chem* **248**: 6911–6919.

Nekrasov SV, Agafonova OV, Belogurova NG, Delver EP, Belogurov AA. (2007). Plasmid-encoded antirestriction protein ArdA can discriminate between type I methyltransferase and complete restriction-modification system. *J Mol Biol* **365**: 284–297.

Neuhaus FC, Baddiley J. (2003). A continuum of anionic charge: structures and functions of D-alanyl-teichoic acids in gram-positive bacteria. *Microbiol Mol Biol Rev* **67**: 686–723.

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.

Overbeek R, Larsen N, Walunas T, DSouza M, Pusch G, Selkov EJ *et al.* (2003). The ERGO genome analysis and discovery system. *Nucleic Acids Res* **31**: 164–171.

Park J-H, Dorrestein PC, Zhai H, Kinsland C, McLafferty FW, Begley TP. (2003). Biosynthesis of the thiazole moiety of thiamin pyrophosphate (vitamin B1). *Biochemistry* **42**: 12430–12438.

Pena A, Teeling H, Huerta-Cepas J, Santos F, Yarza P, Brito-Echeverría J *et al.* (2010). Fine-scale evolution: genomic, phenotypic and ecological differentiation in two coexisting *Salinibacter ruber* strains. *ISME J* **4**: 882–895.

Petersen J. (2011). Phylogeny and compatibility: plasmid classification in the genomics era. *Arch Microbiol* **193**: 313–321.

Petersen J, Brinkmann H, Berger M, Brinkhoff T, Päuker O, Pradella S. (2011). Origin and evolution of a novel DnaA-like plasmid replication type in Rhodobacterales. *Mol Biol Evol* **28**: 1229–1240.

Pfleger BF, Lee JY, Somu RV, Aldrich CC, Hanna PC, Sherman DH. (2007). Characterization and analysis of early enzymes for petrobactin biosynthesis in *Bacillus anthracis*. *Biochemistry* **46**: 4147–4157.

Pommier T, Pinhassi J, Hagström Å. (2005). Biogeographic analysis of ribosomal RNA clusters from marine bacterioplankton. *Aquat Microb Ecol* **41**: 79–89.

Poole K, Heinrichs DE, Neshat S. (1993). Cloning and sequence analysis of an EnvCD homologue in *Pseudomonas aeruginosa*: regulation by iron and possible involvement in the secretion of the siderophore pyoverdine. *Mol Microbiol* **10**: 529–544.

Porsby CH, Nielsen KF, Gram L. (2008). *Phaeobacter* and *Ruegeria* species of the *Roseobacter* clade colonize separate niches in a Danish Turbot (*Scophthalmus maximus*)-rearing farm and antagonize *Vibrio anguillarum* under different growth conditions. *Appl Environ Microbiol* **74**: 7356–7364.

Prado S, Montes J, Romalde JL, Juan LB. (2009). Inhibitory activity of *Phaeobacter* strains against aquaculture pathogenic bacteria. *Int Microbiol* **12**: 107–114.

Rao D, Webb JS, Holmstrom C, Case R, Low A, Steinberg P *et al.* (2007). Low densities of epiphytic bacteria from the marine alga *Ulva australis* inhibit settlement of fouling organisms. *Appl Environ Microbiol* **73**: 7844–7852.

Rao D, Webb JS, Kjelleberg S. (2005). Competitive interactions in mixed-species biofilms containing the marine bacterium *Pseudoalteromonas tunicata*. *Appl Environ Microbiol* **71**: 1729–1736.

Rao D, Webb JS, Kjelleberg S. (2006). Microbial colonization and competition on the marine alga *Ulva australis*. *Appl Environ Microbiol* **72**: 5547–5555.

Rocap G, Larimer FW, Lamerdin J, Malfatti S, Chain P, Ahlgren NA *et al.* (2003). Genome divergence in two *Prochlorococcus ecotypes* reflects oceanic niche differentiation. *Nature* **424**: 1042–1047.

Rudrappa T, Biedrzycki ML, Kunjeti SG, Donofrio NM, Czymmek KJ, Paré PW *et al.* (2010). The rhizobacterial elicitor acetoin induces systemic resistance in *Arabidopsis thaliana*. *Commun Integr Biol* **3**: 130–138.

Ruiz-Ponte C, Cilia V, Lambert C, Nicolas JL. (1998). *Roseobacter gallaeciensis* sp. nov., a new marine

bacterium isolated from rearings and collectors of the scallop *Pecten maximus*. *Int J Syst Bacteriol* **48**: 537–542.

Ryu C-M, Farag MA, Hu C-H, Reddy MS, Wei H-X, Paré PW *et al.* (2003). Bacterial volatiles promote growth in Arabidopsis. *Proc Natl Acad Sci USA* **100**: 4927–4932.

Schloter M, Lebuhn M, Heulin T, Hartmann A. (2000). Ecology and evolution of bacterial microdiversity. *FEMS Microbiol Rev* **24**: 647–660.

Schwyn B, Neilands JB. (1987). Universal chemical assay for the detection and determination of siderophores. *Anal Biochem* **160**: 47–56.

Sekiguchi H, Koshikawa H, Hiroki M, Murakami S, Xu K, Watanabe M *et al.* (2002). Bacterial distribution and phylogenetic diversity in the Changjiang estuary before the construction of the Three Gorges Dam. *Microb Ecol* **43**: 82–91.

Seyedsayamdost MR, Case RJ, Kolter R, Clardy J. (2011). The Jekyll-and-Hyde chemistry of *Phaeobacter gallaeciensis*. *Nat Chem* **3**: 331–335.

Shea CM, McIntosh MA. (1991). Nucleotide sequence and genetic organization of the ferric enterobactin transport system: homology to other periplasmic binding protein-dependent systems in *Escherichia coli*. *Mol Microbiol* **5**: 1415–1428.

Silva ACR da, Ferro JA, Reinach FC, Farah CS, Furlan LR, Quaggio RB *et al.* (2002). Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature* **417**: 459–463.

Skorupska A, Janczarek M, Marczak M, Mazur A, Król J. (2006). Rhizobial exopolysaccharides: genetic control and symbiotic functions. *Microb Cell Fact* **5**: 7.

Sofia HJ, Chen G, Beth GH, Reyes-Spindola JF, Miller NE. (2001). Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: functional characterization using new analysis and information visualization methods. *Nucleic Acids Res* **29**: 1097–1106.

Soria-Dengg S, Reissbrodt R, Horstmann U. (2001). Siderophores in marine coastal waters and their relevance for iron uptake by phytoplankton: experiments with the diatom *Phaeodactylum tricornutum*. *Mar Ecol Prog Ser* **220**: 73–82.

Staden R. (1996). The Staden sequence analysis package. *Mol Biotechnol* **5**: 233–241.

Tatusov RL, Koonin EV, Lipman DJ. (1997). A genomic perspective on protein families. *Science* **278**: 631–637.

Tech M, Merkl R. (2003). YACOP: Enhanced gene prediction obtained by a combination of existing methods. *In Silico Biol* **3**: 441–451.

Teufel R, Gantert C, Voss M, Eisenreich W, Haehnel W, Fuchs G. (2011). Studies on the mechanism of ring-hydrolysis in phenylacetate degradation - A metabolic branching point. *J Biol Chem* **286**: 11021–11034.

Teufel R, Mascaraque V, Ismail W, Voss M, Perera J, Eisenreich W *et al.* (2010). Bacterial phenylalanine and phenylacetate catabolic pathway revealed. *Proc Natl Acad Sci USA* **107**: 14390–14395.

Thiel V, Brinkhoff T, Dickschat JS, Wickel S, Grunenberg J, Wagner-Döbler I *et al.* (2010). Identification and biosynthesis of tropone derivatives and sulfur volatiles produced by bacteria of the marine *Roseobacter* clade. *Org Biomol Chem* **8**: 234–246.

Thomas T, Evans FF, Schleheck D, Mai-Prochnow A, Burke C, Penesyan A *et al.* (2008). Analysis of the *Pseudoalteromonas tunicata* genome reveals properties of a surface-associated life style in the marine environment. *PLoS One* **3**: e3252.

Vu B, Chen M, Crawford RJ, Ivanova EP. (2009). Bacterial extracellular polysaccharides involved in biofilm formation. *Molecules* **14**: 2535–2554.

Waack S, Keller O, Asper R, Brodag T, Damm C, Fricke WF *et al.* (2006). Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. *BMC Bioinformatics* **7**: 142.

Williamson SJ, Houchin LA, McDaniel L, Paul JH. (2002). Seasonal variation in lysogeny as depicted by prophage induction in Tampa Bay, Florida. *Appl Environ Microbiol* **68**: 4307–4314.

Wyss M, Kaddurah-Daouk R. (2000). Creatine and creatinine metabolism. *Physiol Rev* **80**: 1107–1213.

Zech H, Thole S, Schreiber K, Kalhoefer 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)