

Genome sequence of the ubiquitous hydrocarbon-degrading marine bacterium *Alcanivorax borkumensis*

Susanne Schneiker^{1,3,8}, Vítor AP Martins dos Santos^{2,8}, Daniela Bartels³, Thomas Bekel³, Martina Brecht^{1,3}, Jens Buhrmester¹, Tatyana N Chernikova^{2,4}, Renata Denaro⁵, Manuel Ferrer^{2,7}, Christoph Gertler^{2,4}, Alexander Goesmann³, Olga V Golyshina^{2,4}, Filip Kaminski², Amit N Khachane², Siegmund Lang⁶, Burkhard Linke³, Alice C McHardy^{3,7}, Folker Meyer³, Taras Nechitaylo^{2,4}, Alfred Pühler^{1,3}, Daniela Regenhardt^{2,7}, Oliver Rupp³, Julia S Sabirova^{2,4}, Werner Selbitschka^{1,3}, Michail M Yakimov^{2,5}, Kenneth N Timmis^{2,4}, Frank-Jörg Vorhölter^{1,3}, Stefan Weidner^{1,3}, Olaf Kaiser^{1,3,8} & Peter N Golyshin^{2,8}

Alcanivorax borkumensis is a cosmopolitan marine bacterium that uses oil hydrocarbons as its exclusive source of carbon and energy. Although barely detectable in unpolluted environments, *A. borkumensis* becomes the dominant microbe in oil-polluted waters. *A. borkumensis* SK2 has a streamlined genome with a paucity of mobile genetic elements and energy generation-related genes, but with a plethora of genes accounting for its wide hydrocarbon substrate range and efficient oil-degradation capabilities. The genome further specifies systems for scavenging of nutrients, particularly organic and inorganic nitrogen and oligo-elements, biofilm formation at the oil-water interface, biosurfactant production and niche-specific stress responses. The unique combination of these features provides *A. borkumensis* SK2 with a competitive edge in oil-polluted environments. This genome sequence provides the basis for the future design of strategies to mitigate the ecological damage caused by oil spills.

Several million tons of oil are discharged into the oceans every year. Some of it seeps from natural oil fields, but the bulk of the discharge comes as a result of anthropogenic activities. One of the most dramatic polluting events was the spill of oil from the supertanker *Prestige*, which sank off the Galician coast of Spain, releasing an estimated 17,000 tons of heavy fuel oil¹. Maintenance of sustainable marine and coastal ecosystems urgently requires the development of effective measures to reduce oil pollution and mitigate its environmental impact.

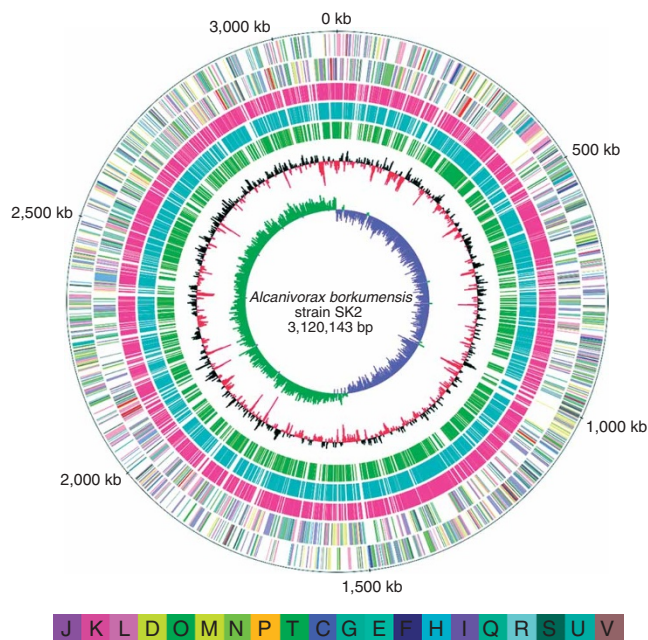
A. borkumensis is an unusual, rod-shaped marine γ -proteobacterium able to grow on a highly restricted spectrum of substrates, predominantly alkanes². This bacterium is found in low numbers in unpolluted environments, but it quickly becomes the dominant microbe in oil-polluted open ocean and coastal waters, where it may comprise up to 80–90% of the oil-degrading microbial community^{3–6}. Several recent field studies on bacterial community dynamics and hydrocarbon degradation in coastal systems have demonstrated the pivotal role of *A. borkumensis* in oil-spill bioremediation^{5,7}.

Bacteria of the genus *Alcanivorax* belong to a group of slow-growing marine hydrocarbonoclastic bacteria (HCB; 'clastic'

from the Greek word *klastes*, meaning breaker) comprising bacterial genera, such as *Cycloclasticus*, *Marinobacter*, *Neptunomonas*, *Oleiphilus*, *Oleispira* and *Thalassolituus*⁸, that preferentially use petroleum-derived aliphatic and aromatic hydrocarbons as carbon and energy sources. Since the first description of *A. borkumensis* in 1998, it has been detected in many marine and coastal habitats worldwide including the Mediterranean Sea, the Pacific Ocean, the Japanese and Chinese Seas and the Arctic Sea^{3,4,6–9}.

We report here the genome sequence of the marine, hydrocarbonoclastic bacterium *A. borkumensis* strain SK2, the first hydrocarbonoclastic bacterium to be sequenced. Genome analysis yielded unprecedented insights into the bacterium's capacity for (i) n-alkane degradation, including metabolism, biosurfactant production and biofilm formation, (ii) scavenging of nutrients and cofactors in the oligotrophic marine environment and (iii) coping with various habitat-specific stress factors. The genome data and their functional analysis provide us with an invaluable knowledge base essential to design, develop, test and optimize rational strategies to mitigate the ecological damage caused by oil spills in marine systems.

¹Lehrstuhl für Genetik, Fakultät für Biologie, Universität Bielefeld, D-33594 Bielefeld, Germany. ²Division of Microbiology, German Research Center for Biotechnology, D-38124 Braunschweig, Germany. ³Center for Biotechnology (CeBiTec), Universität Bielefeld, D-33594 Bielefeld, Germany. ⁴Institute for Microbiology, Technical University of Braunschweig, D-38106 Braunschweig, Germany. ⁵Istituto per l'Ambiente Marino Costiero (CNR), I-98122 Messina, Italy. ⁶Institute of Biochemistry and Biotechnology, Technical University of Braunschweig, D-38106 Braunschweig, Germany. ⁷Present addresses: Institute of Catalysis, Campus UAM, E-28049 Madrid, Spain (M.F.), Bioinformatics & Pattern Discovery Group, IBM Thomas J Watson Research Center, Yorktown Heights, New York 10598, USA (A.C.McH.), Division Biogeochemistry, Research Centre Rossendorf, D-01314 Dresden, Germany (D.R.). ⁸These authors contributed equally to this work. Correspondence should be addressed to V.A.P.M.d.S. (vds@gbf.de).



RESULTS

General features of the *A. borkumensis* SK2 genome

The *A. borkumensis* SK2 genome consists of a single circular chromosome of 3,120,143 base pairs (bp) with an average G+C content of 54.7% (Fig. 1). The assembly of the sequence was validated by a complete bacterial artificial chromosome (BAC) map (Supplementary Fig. 1 online). Biological roles were assigned to 2,241 of the 2,755 predicted coding sequences. The remaining 514 coding sequences comprise 316 conserved hypothetical and 198 of unknown function (Fig. 1 and Table 1).

Genomic comparison of *A. borkumensis* SK2 to related species

A. borkumensis SK2 is the first sequenced member within the order Oceanospirillales of the γ -Proteobacteria (Fig. 2). Most of the *A. borkumensis* SK2-encoded proteins show highest similarity to proteins of γ - (1,897; 68.9%) and β -Proteobacteria (302; 11.0%) (Supplementary Fig. 2 online). BLAST comparisons to nonmarine members of the Pseudomonaceae, marine γ - and α -Proteobacteria as well as to selected marine Gram-positive bacteria revealed that the proportion of homologous proteins roughly decreases as phylogenetic distance increases (Supplementary Fig. 2 online).

Genomic islands and mobile genetic elements

The *A. borkumensis* SK2 genome has a number of regions with atypical GC content. One such island (1,011,000 bp to 1,058,000 bp) comprises a cluster of 40 genes whose products are predicted to be involved in cell-surface polysaccharide biosynthesis. A second prominent island (3,057,000 bp to 3,075,000 bp) carries a complete gene cluster for alkane degradation (*alkSB₁GJH*)¹⁰. This unusual nucleotide composition may be indicative of horizontal gene transfer or result from a multitude of factors such as selection or mutation bias.

The *A. borkumensis* SK2 genome harbors only a small number of mobile genetic elements such as transposons, insertion (IS) elements or their remnants. Interestingly, various elements of heavy metal cation efflux transport systems were identified within a putative transposon (Supplementary Results 1 online). This paucity of mobile elements may be explained by counterselection against variants with

Figure 1 Circular representations of the *A. borkumensis* SK2 chromosome displaying relevant genome features. From the outer to the inner concentric circle: circle 1, genomic position in kb; circles 2 and 3, predicted protein-coding sequences (CDS) on the forward (outer wheel) and the reverse (inner wheel) strands colored according to the assigned COG classes; circles 4, 5 and 6, CDS with homologs in *P. aeruginosa*⁵⁰ (pink), *P. putida*⁵¹ (light blue) and *I. loihensis*³⁰ (green), respectively; circle 7, G+C content showing deviations from the average (54.7%); circle 8, GC skew. The bar below the plot represents the COG colors for the functional groups (C, energy production and conversion; D, cell cycle control, mitosis and meiosis; E, amino acid transport and metabolism; F, nucleotide transport and metabolism; G, carbohydrate transport and metabolism; H, coenzyme transport and metabolism; I, lipid transport and metabolism; J, translation; K, transcription; L, replication, recombination and repair; M, cell wall/membrane biogenesis; N, cell motility; O, post-translational modification, protein turnover, chaperones; P, inorganic ion transport and metabolism; Q, secondary metabolites biosynthesis, transport and catabolism; R, general function prediction only; S, function unknown; T, signal transduction mechanisms; U, intracellular trafficking and secretion; V, defense mechanisms).

increased numbers of mobile elements, thus avoiding the detrimental effects of transposition events on the fitness of bacteria. Alternatively, it may reflect a minor impact of horizontal gene transfer on the *A. borkumensis* SK2 genome.

Genetic basis of alkane degradation in *A. borkumensis* SK2

A. borkumensis SK2's most distinctive feature is its ability to grow efficiently and almost exclusively on alkanes^{2,5,11,12}. Its genome specifies a number of systems for the catabolism of hydrocarbons (Figs. 3 and 4). The proteins encoded by the SK2 *alkSB₁GJH* operon genes (ABO_2707 ABO_2710, Fig. 3a) have, with the exception of *alkS*, > 80% amino acid similarity to those of the corresponding, well-characterized alkane degradation components in *Pseudomonas putida* strains^{10,13–15}. The AlkS of *A. borkumensis* SK2 shows 48% amino acid identity with AlkS of *P. putida*, which has been grouped within the MalT subfamily of LuxR transcriptional activators¹⁶. The *A. borkumensis* SK2 AlkB₁ alkane hydroxylase efficiently oxidizes medium-chain alkanes (C5 to C12). Its expression, as revealed by quantitative real time (qRT)-PCR, was strongly (44-fold) induced in the exponential growth phase compared to the stationary phase when grown on alkanes (Supplementary Fig. 3 online). Other proteins involved in

Table 1 General features of the *A. borkumensis* SK2 genome

| Feature | |
|--------------------------------|------------------|
| Size, bp | 3,120,143 |
| G+C content, % | 54.7 |
| Coding sequences | 2,755 |
| Function assigned | 2,241 |
| Conserved hypothetical protein | 316 |
| Hypothetical protein | 198 |
| % of genome coding | 87.4 |
| Average length, bp | 993 |
| Maximal length, bp | 10,803 |
| % ATG initiation codons | 82.1 |
| % GTG initiation codons | 11.3 |
| % TTG initiation codons | 6.6 |
| RNA elements | |
| rRNA | 3 × (16S-23S-5S) |
| tRNAs | 42 |

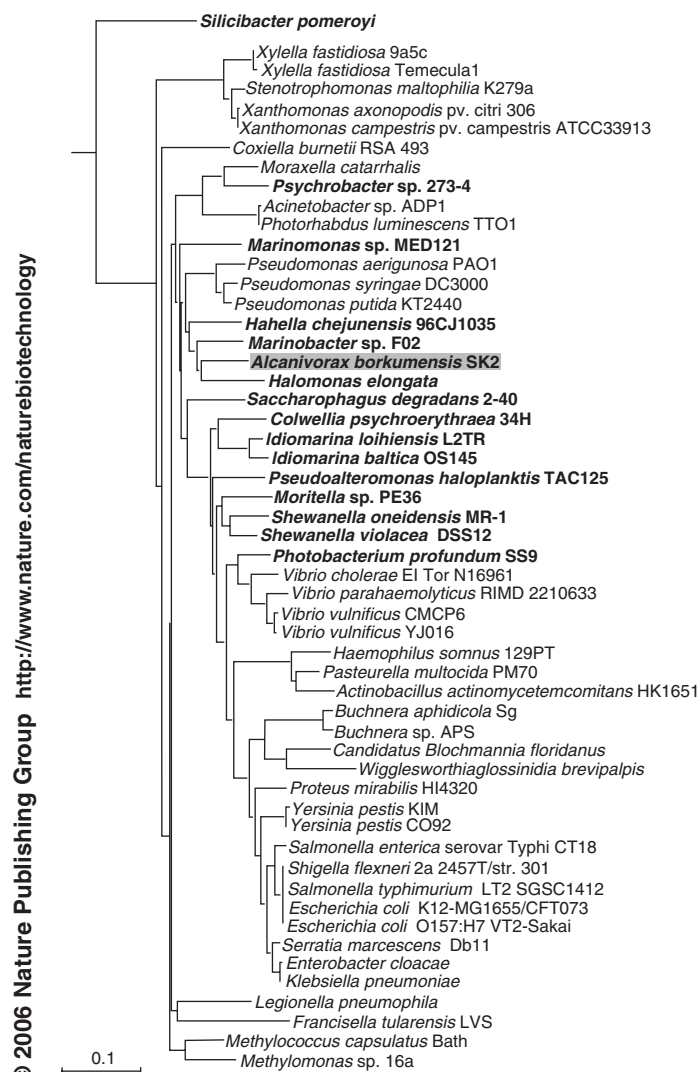


Figure 2 Phylogenetic affiliation of *A. borkumensis* SK2 based on 16S rDNA sequence comparison. Neighborhood-joining tree showing a 16S rDNA gene phylogeny-based placement of *A. borkumensis* among the γ -Proteobacteria. Sequences were taken from the type strains of the given organisms. Marine organisms with completely sequenced genomes are highlighted in bold. The tree is rooted with *Aquifex pyrophilus* M83548. The α -proteobacterial strain *Silicibacter pomeroyi* served as outgroup.

have multiple copies (**Supplementary Fig. 4** and **Supplementary Table 1** online).

A. borkumensis SK2 is able to degrade a large range of alkanes up to C32 and branched aliphatic, as well as isoprenoid hydrocarbons (e.g., phytane), alkylarenes and alkylcycloalkanes (ref. 20 and M.M.Y., unpublished data). This spectrum is much broader than expected based on current knowledge about alkane hydroxylase complexes. Thus, the *A. borkumensis* SK2 genome must specify a wider range of systems involved in hydrocarbon catabolism. It encodes three cytochromes: ABO_2384 for P450(a), and two paralogous genes ABO_2288 P450(b) and ABO_0201 P450(c). ABO_0201 (P450(c)) clusters in an operon-like structure with genes encoding a ferredoxin (*fdx*, ABO_0200), an alcohol dehydrogenase (*alkJ*₂, ABO_0202), a FAD-dependent oxidoreductase (ABO_0203) and an AraC transcriptional regulator gene (ABO_0199) (**Fig. 3a**). Proteome²¹ and qRT-PCR expression experiments revealed that all the enzymes encoded by this gene cluster were upregulated on hexa- and tetradecane. Furthermore, all three genes for P450 (ABO_0201, ABO_2288, ABO_2384) were specifically expressed in the presence of isoprenoid hydrocarbons, whereas at exponential growth phase, substantially lower levels of mRNA, or none at all, were demonstrated for *alkB*₂ and *alkB*₁ genes, respectively (**Supplementary Fig. 3** online). The two alkane-induced cytochromes of *A. borkumensis* SK2 are phylogenetically grouped in one branch with the P450 cytochrome of *Acinetobacter sp.* EB104 (**Fig. 3b**), which catalyzes the terminal oxidation of alkanes²². Thus, the three P450 systems are likely to account for an important part of the expanded hydrocarbon degradation capacity of *A. borkumensis* SK2. A metabolic reconstruction of these various alkane degradation pathways is depicted in **Figures 3c** and **4**. It is anticipated that a fraction of the large number of enzymes, especially oxidoreductases with unknown specificity (see Biotechnological potential below), might be involved in hydrocarbon catabolism. The versatility and wider spectrum for hydrocarbon utilization clearly provides *A. borkumensis* SK2 a competitive advantage over other members of oil-degrading marine microbial communities.

Emulsification of hydrocarbons by *A. borkumensis* SK2

The production of biosurfactants facilitates emulsification of alkanes, enhances their bioavailability and increases the degradation rate of these hydrophobic organic substrates. *A. borkumensis* forms stable emulsions of hydrocarbon in water and produces biosurfactants²³. These biosurfactants are anionic glucolipids carrying four fatty acids of varying chain lengths¹¹. The genetic organization of the glucolipid biosynthesis remains unclear in *A. borkumensis* SK2, but genome annotation revealed candidate genes potentially involved in biosurfactant production. ABO_1783 and ABO_2215 encode glycosyltransferases, exhibiting significant homology with RhlB from *Pseudomonas aeruginosa*²⁴ and glycosyltransferase protein family 9, respectively. These gene products possibly provide the sugar moiety of the glucolipids, yielding glucose lipid surfactants. *A. borkumensis* SK2 also expresses an OprF/OmpA protein encoded by ABO_0822, which is upregulated when grown on alkane²⁰.

ngp this alkane degradation pathway, namely, AlkK (ABO_2748), AlkL (ABO_1922) and AlkN (ABO_0106), have lower percentage similarities to the orthologs in *P. putida*, whereas no homolog could be detected for the *alkF* gene encoding rubredoxin 1 in *P. putida* GPo1. However, two genes, *rubA* and *rubB* (ABO_0163 and ABO_0162) encode a rubredoxin and a rubredoxin reductase, respectively, and are likely to be involved in alkane catabolism. Rubredoxin reductase genes map in many alkane-degrading bacteria separately from the alkane hydroxylase genes^{10,17}.

A second alkane hydroxylase system is specified by ABO_0121, encoding a GntR transcriptional negative regulator, and ABO_0122, encoding the alkane hydroxylase AlkB₂, which oxidizes medium-chain alkanes in the range C8 to C16 (ref. 10). Although both *alkB*₁ and *alkB*₂ in *A. borkumensis* SK2 exhibit similar expression patterns, being simultaneously induced on *n*-alkanes (**Supplementary Fig. 3** online), their underlying regulation is certainly different and possibly responds complementarily under different conditions¹⁸. Importantly, both alkane hydroxylase systems are located close to the origin of replication of the chromosome (**Supplementary Fig. 1** online), which ensures a high dosage of alkane hydroxylation genes. A search for *alkB* genes in sequenced marine organisms and the only marine metagenomic database available¹⁹ revealed that only a comparatively small number of bacteria harbor *alkB* homologs and even fewer

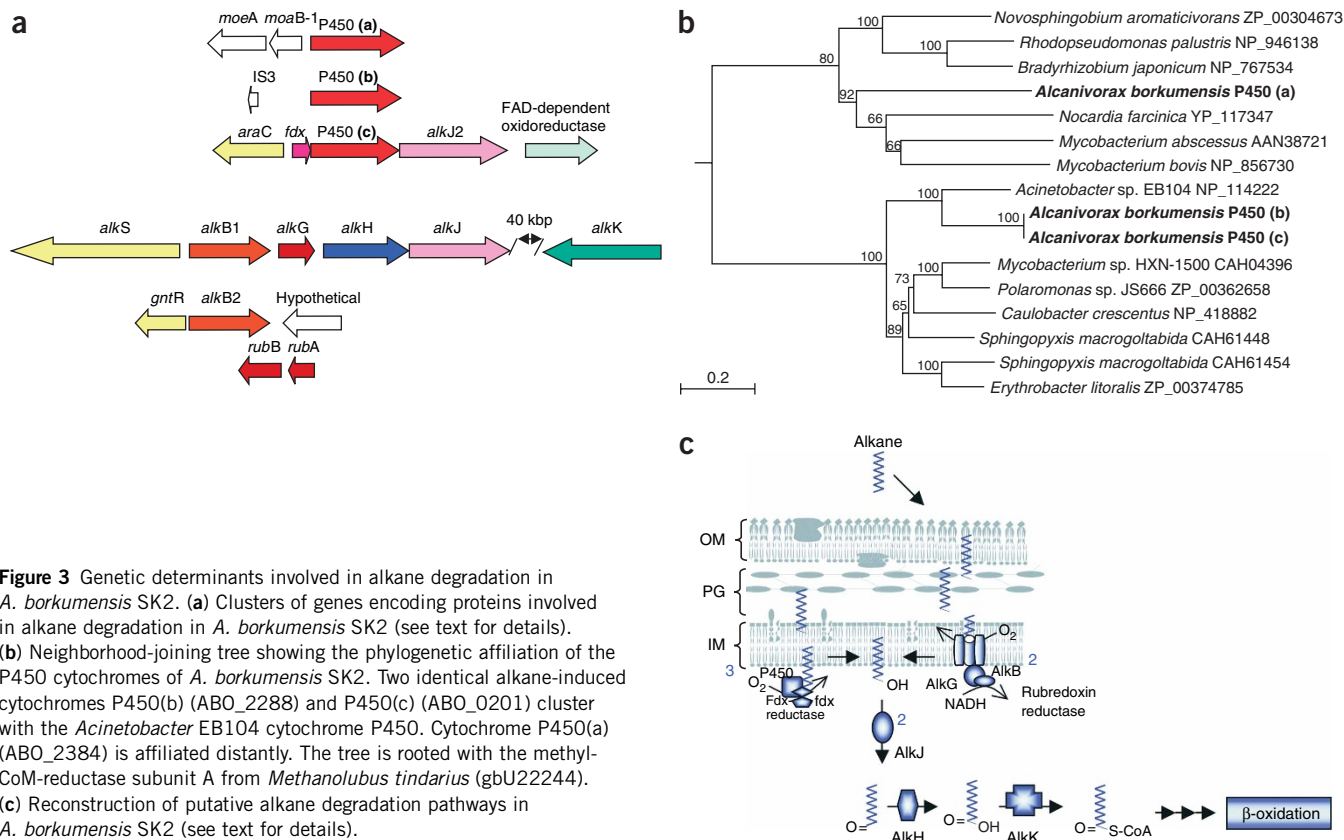


Figure 3 Genetic determinants involved in alkane degradation in *A. borkumensis* SK2. (a) Clusters of genes encoding proteins involved in alkane degradation in *A. borkumensis* SK2 (see text for details). (b) Neighborhood-joining tree showing the phylogenetic affiliation of the P450 cytochromes of *A. borkumensis* SK2. Two identical alkane-induced cytochromes P450(b) (ABO_2288) and P450(c) (ABO_0201) cluster with the *Acinetobacter* EB104 cytochrome P450. Cytochrome P450(a) (ABO_2384) is affiliated distantly. The tree is rooted with the methyl-CoM-reductase subunit A from *Methanolobus tindarius* (gbU22244). (c) Reconstruction of putative alkane degradation pathways in *A. borkumensis* SK2 (see text for details).

OmpA proteins, the active constituents of the biosurfactant alasan, have been demonstrated to efficiently emulsify hydrocarbons²⁵ and to be produced by a number of oil-degrading γ -Proteobacteria²⁶. *A. borkumensis* SK2 harbors determinants of OprG/OmpW (ABO_1922) and OmpH (ABO_1152), which are also possibly involved in emulsifier production.

Secretion and biofilm formation in *A. borkumensis* SK2

Owing to the low solubility of hydrocarbons in water, bacterial degradation takes place mainly at the hydrocarbon-water interface. *A. borkumensis* SK2 attaches readily to the oil-water interface of hydrocarbon droplets in salt water (Fig. 4). The *A. borkumensis* SK2 genome has many determinants for the biosynthesis, export, modification and polymerization of exopolysaccharides putatively involved in biofilm formation. Forty of these determinants are identified in a 50-kb region of lower G+C content (1,011,000 bp to 1,058,000 bp), which codes for putative glycosyltransferases, for proteins predicted to be involved in sugar nucleotide biosynthesis and addition of noncarbohydrate decorations, as well as export and polymerization of polysaccharides. A cluster putatively specifying alginate biosynthesis is located elsewhere (432,500 to 448,500, Supplementary Results 2 online). The *A. borkumensis* SK2 chromosome harbors 16 *pil* genes related to Type IV pili, which mediate the formation of biofilms and are involved in surface translocations²⁷. A number of determinants for a Type II secretion system, Sec translocon and twin-arginin translocation (Tat) have been identified, as well as five genes putatively encoding secretion proteins of the HlyD family (Fig. 4). Most of the above genes are likely to be involved in the formation of biofilms at oil-water interfaces.

Nutrient transport and sodium dependency

The success of a bacterium in the generally oligotrophic marine environment depends on effective scavenging of elements such as N and P, S and various oligo-elements such as Fe, Zn, Co, Mg, Mn and Mo. *A. borkumensis* SK2 encodes genes for a broad range of transport proteins (Fig. 4, see also Supplementary Results 2 online). These comprise determinants for about 50 permeases, of which roughly half belong to high affinity ABC-transport systems, five major facilitator superfamily transport systems and two TRAP C4-dicarboxylate transporters (DctP, ABO_2146 and ABO_0688). No other carbohydrate transporters, such as ABC-type transport systems for ribose, xylose, arabinose and galactoside, that are usually present in other bacteria, were identified. This is consistent with *A. borkumensis* SK2's reported inability to grow on or to use many common, simple carbohydrates².

Many of the transport systems in *A. borkumensis* SK2, consistent with its marine lifestyle, are linked to Na⁺ pumps (Fig. 4). The *A. borkumensis* SK2 genome encodes six subunits of a Na⁺-dependent NADH-quinone dehydrogenase (*nqrABCDE*, ABO_1032-1037), enabling it to use the sodium gradient as a source of energy for nutrient uptake. Export of sodium ions at the expense of the proton gradient is performed by a variety of sodium/proton antiporters, including the multisubunit Na⁺/H⁺ antiporter (*mnhABCDEFG* genes, ABO_2653-2658), as well as NhaD (ABO_0238), NhaB (ABO_1226), NhaP (ABO_1404) and NhaC (ABO_2329) (Fig. 4). *A. borkumensis* SK2 also encodes a symporter for Na⁺/alanine (*dagA*, ABO_0618), Na⁺/sulfate (ABO_0929), Na⁺/glutamate (ABO_1478 and ABO_1616, see above) as well as several other Na⁺-dependent symporters (ABO_0344, ABO_1913, ABO_2155, ABO_2158, ABO_2678).

The *A. borkumensis* SK2 genome encodes diverse and alternative systems for the uptake of both dissolved organic and inorganic nitrogen²⁸. Phosphate uptake is mediated by a high-affinity ABC-type Pst system encoded by *phoU-pstBACS*, ABO_2681-2685 and *phoBR* (ABO_0166, 0167), as well as by a low-affinity transporter protein (encoded by ABO_2305). In addition to this array of N, P and S uptake systems (for details see **Supplementary Results 2** online), SK2 specifies the import of various oligo-elements, such as molybdate (*modABC*, ABO_1254-1256), zinc (*znuAB*, ABO_0155-0157), magnesium and cobalt (ABO_1335), and magnesium (*mgtE*, ABO_0541). This wealth of scavenging abilities enables *A. borkumensis* SK2 to use its alkane-degrading potential to tackle quickly and effectively the sudden increase in carbon availability and consequent carbon/nutrient imbalance that results from typical oil spills. Indeed, although many

constituents in crude oil are biodegradable, the main limitation to their actual biodegradation is nutrient availability, particularly nitrogen and phosphorus³. These features thus provide the genomic basis underlying the overwhelming ecological success of *A. borkumensis* and underpin its potential for bioremediation of oil-contaminated marine environments.

Metabolic specialization and biotechnological potential

A. borkumensis SK2, like other hydrocarbonoclastic bacteria, has a limited substrate range for its growth². The absence of a functional PEP-dependent sugar/phosphotransferase system or other types of sugar transporters, as well as the lack of several determinants for key enzymes of the glycolytic, pentose phosphate shunt and Entner-Doudoroff pathways (**Supplementary Results 1** online), is consistent

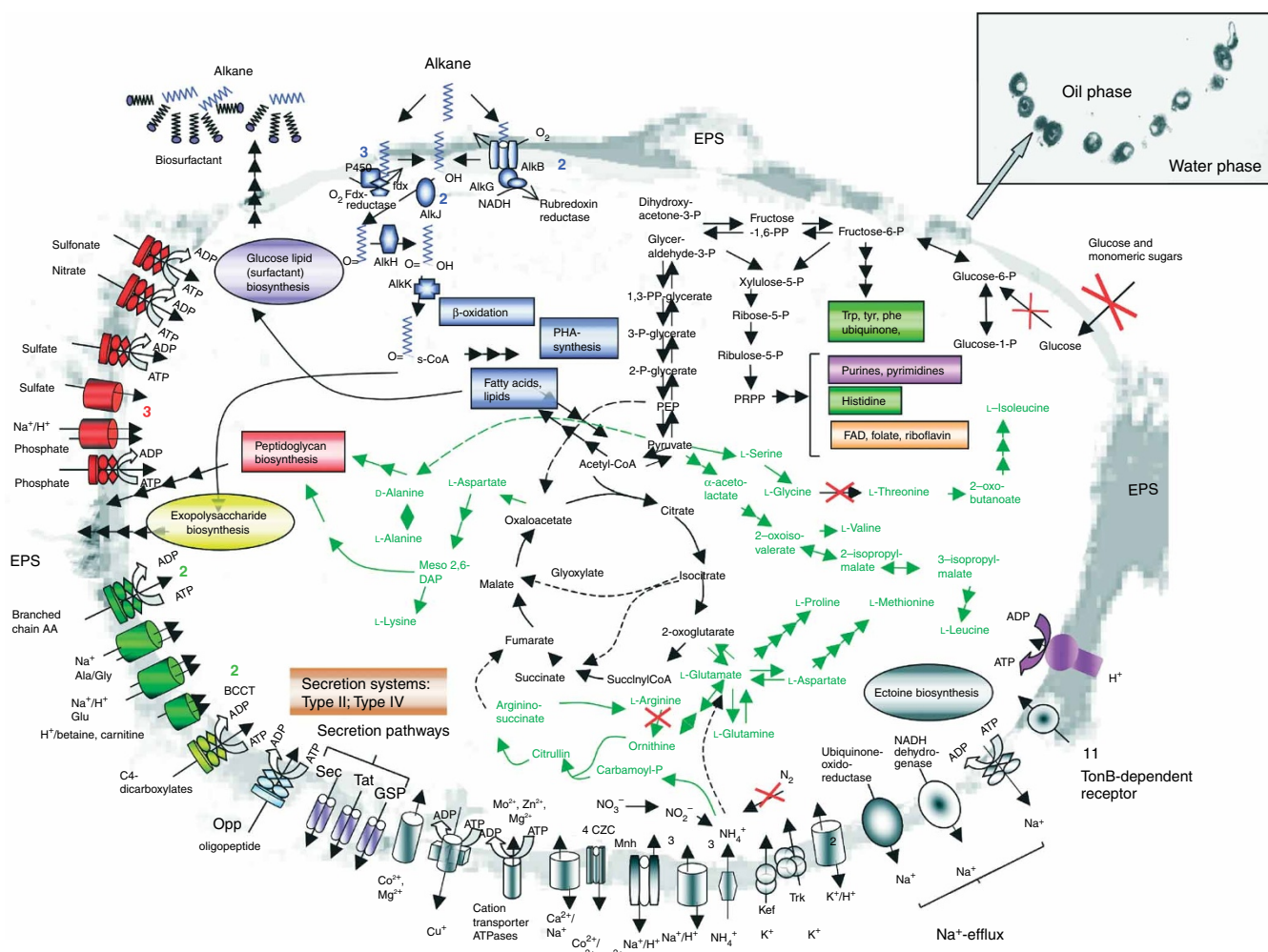


Figure 4 Schematic overview of metabolism and transport in *A. borkumensis* SK2. The background is a transmission electron micrograph (TEM) of an *A. borkumensis* cell grown on hexadecane (courtesy of H. Lünsdorf). The insert in the right upper corner shows a TEM of *A. borkumensis* SK2 cells at the oil-water interface of hydrocarbon droplets in salt water. Predicted pathways for alkane degradation are depicted in marine blue. Predicted transporters are grouped by substrate specificity: inorganic cations (gray), inorganic anions (dark orange), amino acids/peptides/amines/purines/pyrimidines and other nitrogenous compounds (dark green), carboxylates (light green), drug efflux and other (dark gray). Export or import of solutes is designated by the direction of the arrow through the transporter. The energy coupling mechanisms of the transporters are also shown: solutes transported by channel proteins are shown with a double-headed arrow; secondary transporters are shown with two arrowed lines indicating both the solute and the coupling ion; ATP-driven transporters are indicated by the ATP hydrolysis reaction; transporters with an unknown energy-coupling mechanism are shown with only a single arrow. The P-type ATPases are shown with a double-headed arrow to indicate they include both uptake and efflux systems. Where multiple homologous transporters with similar substrate predictions exist, the number of that type of protein is indicated in parentheses. Abbreviations of less common terms: EPS, extracellular polysaccharides; AA, amino acids; BCCT, betaine/carnitine/choline transporters; GSP, general secretion pathways; PRPP, 5'-phospho-alpha-D-ribose 1-diphosphate; Mhn, complex sodium/proton antiporter involved in sodium excretion.

with the inability of *A. borkumensis* SK2 to utilize simple hexoses and other simple carbohydrates for growth.

The *A. borkumensis* SK2 genome encodes a number of proteins putatively involved in metabolic reactions of potential biotechnological interest, including eight hydrolases of the haloacid dehydrogenase/epoxide family, two determinants for dienelactone hydrolases (ABO_1618 and ABO_1886), three for deacylases, 36 for various cytochrome proteins and 30 for oxidoreductases. Various oxidoreductase genes are clustered together or in operon-like structures with determinants for monooxygenases, aldehyde dehydrogenases, decarboxylases, Rieske 2Fe-2S family proteins, transcription regulators and/or transporter-coding genes. The genome also includes 11 genes coding for different lipases/esterases of unknown specificity. Two of these esterases were purified and functionally characterized. They exhibit excellent enzymatic activity up to two orders of magnitude higher than typical esterases, have a wide substrate profile, remarkable enantioselectivity and chemical resistance, which underscores their potential for the resolution of chiral mixtures in biocatalysis (M.F. *et al.*, unpublished data). Thus, overall, *A. borkumensis* SK2 exhibits a biotechnological potential that goes well beyond its application to marine oil degradation.

Stress responses and osmoregulation

In addition to general regulation and signal transduction mechanisms (Supplementary Results 2 online), *A. borkumensis* SK2 specifies a number of habitat-related stress-response systems. Being a marine organism that thrives mostly in the upper layers of the ocean, *A. borkumensis* SK2 harbors a number of genes involved in reducing the damaging effect of UV radiation. The full gene complement for DNA alkylation and base excision repair, recombinational and nucleotide excision repair, and those for the RecF pathway and SOS response are shown in Supplementary Table 2 online. *A. borkumensis* specifies as well a variety of chaperone-like activity proteins, many of which are commonly found in all bacterial genomes (Supplementary Results 2 online). Interestingly, however, and like *Idiomarina loihiensis*²⁹ and other marine bacteria, *A. borkumensis* SK2 specifies a relatively low number of cold-shock proteins as compared to bacteria such as *Escherichia coli*. *A. borkumensis* SK2 encodes three Na⁺-driven multidrug efflux pumps (ABO_0158, ABO_1554 and ABO_2623), as well as several systems for detoxification of compounds like arsenate, mercury, copper and other heavy metals. (Fig. 4 and Supplementary Results 1 online). Halotolerant bacteria generally accumulate K⁺, glutamate and the compatible solutes ectoine and betaine as osmoprotectants. Accordingly, and in addition to the uptake systems for glutamate (*gltP*, ABO_1478, *gltS*, ABO_1616) and choline/betaine (ABO_0232, ABO_0637 and ABO_0808), the *A. borkumensis* SK2 genome contains the genetic determinants for the biosynthesis of glutamate (*gltA*, ABO_1501, *gltB*, ABO_2229, *gltD*, ABO_2228 and *gltX*, ABO_1509), ectoine (*ectABC*, ABO_2150_2152) and (putatively) betaine (ABO_0886) (Fig. 4). However, genes for the synthesis of other organic compatible solutes, such as trehalose, di-inositol phosphate or mannosylglycerate, were not identified. Determinants for ATP-dependent K⁺ transporters are absent, which may indicate a strategy for minimization of ATP expenditure in response to consistently high salt concentrations in marine environments.

DISCUSSION

The analysis of the genome sequence of *A. borkumensis* SK2 has provided us with valuable insights into the genomic basis of its (i) unusual metabolic capability and cellular composition,

(ii) oligotrophic lifestyle and high affinity for hydrocarbon substrates, (iii) genomic responses to the signals and environmental stresses it faces in its natural environment and (iv) ability to overcome carbon/nutrient imbalances typical of oil spills to degrade a range of oil hydrocarbons and to dominate oil-degrading microbial communities. *A. borkumensis* attains its ecologically competitive advantage to both adapt rapidly to the presence of oil and thrive in most oceans of the world, by relying on a streamlined and efficient combination of essential, central metabolic functions with remarkable hydrocarbon degradation and emulsification abilities and a broad nutrient uptake complement, in particular for organic and inorganic nitrogen. The ubiquity of *A. borkumensis* reflects its highly developed ability to adapt to the varying conditions it faces in different unpolluted and polluted marine environments. Future functional studies involving spatiotemporal proteomic, transcriptional and metabolomic profiling of *A. borkumensis* under a wide range of hydrocarbon-relevant conditions (including pilot- and field studies), complemented with thorough, systematic and targeted functional investigations, will largely expand our insights into marine hydrocarbon degradation. The availability of this expanded knowledge base, in conjunction with the growing wealth of (meta)genomic, microbial diversity and population dynamics data³⁰, provides a framework within which to study the interactions among hydrocarbon-degrading communities and the impact of environmental conditions on the genetic programs underlying oil degradation, stress responses and ecological fitness. This framework will thus undoubtedly contribute to the development of new strategies for the bioremediation of petroleum-contaminated marine sites.

METHODS

Whole genome shotgun sequencing. Genomic DNA of *A. borkumensis* SK2 (ref. 2) was extracted from cells grown at 30 °C in ONR7a medium³¹ by using the GNOME DNA isolation kit from Qbiogene. The isolated DNA was used for construction of DNA shotgun clone libraries with insert sizes of 1 kb, 2–3 kb and 8 kb in pUC19 vector³² by Qiagen GmbH. Plasmid clones were end sequenced on ABI 3700 capillary sequencing machines (ABI) by Qiagen GmbH. Base calling was carried out using PHRED^{33,34}. High-quality reads were defined by a minimal length of 250 base pairs with an average quality value of \geq phred 20. While sequencing the *A. borkumensis* genome, we had to readjust the estimated genome size of 2.1 Mbp (as measured by pulse field gel electrophoresis) to a real size of \sim 3.1 Mbp³⁵. Finally, 32,428 high-quality reads, a total of 11,472 (2.0 X), 13,052 (2.3 X), and 7,904 (1.4 X) end sequences from libraries of 1 kb, 2–3 kb and 8 kb, respectively, were established (X's indicate genome equivalents).

Sequence assembly and assembly validation. Base calling, quality control and vector sequence elimination of the sequences were performed using the software package BioMake (Bielefeld University, unpublished) as described³⁵. Sequence assembly was performed by using the PHRAP assembly tool (<http://www.phrap.org>). The CONSED/AUTOFINISH software package^{36,37}, supplemented by the in-house tool BACCard1³⁸, was used for the finishing of the genome sequence.

For gap closure and assembly validation, a BAC library with inserts of \sim 100 kb was constructed by Molecular Engines Laboratories, as described³⁹. Sequencing of BAC ends was carried out on ABI 3100 and ABI 377 sequencing machines by Integrated Genomics GmbH and IIT GmbH, respectively. Remaining gaps of the whole genome shotgun assembly were closed by sequencing on shotgun and BAC clones carried out by IIT GmbH on LI-COR 4200L (LI-COR) and ABI 377 sequencing machines and by direct sequencing with chromosomal DNA as a template on an ABI 3730xl DNA analyzer by the sequencing group of the Max Planck Institute for Developmental Biology. To obtain a high quality genome sequence, we polished all regions of the consensus sequence to at least phred 40 quality by primer walking. Collectively, 1,471 sequencing reads were added to the shotgun

assembly for finishing and polishing of the genomic sequence. Repetitive elements, that is, rRNA operons, were sequenced completely by primer walking on BACs and, in some cases, on shotgun clones. For assembly validation, BAC end sequences were mapped onto the genome sequence using the BACCardI tool³⁸.

Assembly and genome analysis. Curation and annotation of the genome was done by using the annotation system GenDB⁴⁰. Briefly, a combined gene prediction strategy⁴¹ was applied on the assembled sequences using GLIMMER and CRITICA. Putative ribosomal binding sites and transfer RNA genes were identified with RBSFINDER⁴² and tRNAscan-SE⁴³, respectively. Before the manual annotation of each predicted gene, an automatic annotation was computed based on different tool results as follows: similarity searches were performed against different databases including SWISS-PROT and TrEMBL, KEGG, Pfam, TIGRFAM and InterPro. Additionally, SignalP, helix-turn-helix and TMHMM were applied. Finally, each gene was functionally classified by assigning a Clusters of Orthologous Groups (COG) number and its corresponding COG category⁴⁴ and GeneOntology numbers⁴⁵.

Genomic comparisons. For comparative analyses, the annotated genome sequences of the following bacteria were imported into GenDB: *Pseudomonas aeruginosa* PAO1 (gb|AE004091), *Pseudomonas putida* KT2440 (gb|AE015451), *Idiomarina loihiensis* L2TR (gb|NC_006512). Homology searches were conducted against the genomes and the plasmids on the nucleotide and amino acid sequence level by using BLAST⁴⁶. Comparisons of chromosomal sequences were performed with GenDB.

Detection of regions with atypical GC content. Genomic regions with atypical G+C content were identified using the 'sliding window' technique with a window size of 2,000 bp and a step size of 1,000 bp. For this purpose, the G+C content was assumed to follow a gaussian distribution and regions with at least 1.5 s.d. difference from the mean were calculated.

Construction of phylogenetic trees. Homologs were identified using BLAST analysis with a cutoff *E*-value of 1e-45 and >50% sequence alignment. Rooted neighbor-joining phylogenetic trees were constructed using the ClustalX 1.83 tool⁴⁷, edited and visualized by NJPLOT⁴⁸ and TreeExplorer software MEGA package⁴⁹.

Database submission. The nucleotide sequence of the *A. borkumensis* SK2 genome was submitted to GenBank under accession number AM286690.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

The authors thank all the people involved in this project. We are also grateful to Heinrich Lünsdorf for the electron microscope pictures of *A. borkumensis* SK2 in different settings. This work was supported by grants of the German Federal Ministry of Education and Research (031U113D, 031U213D) and 0313105 in the frame of the GenoMik network "Genome Research on Bacteria Relevant for Agriculture, Environment and Biotechnology." K.T. and A.P. thank the Fonds der Chemischen Industrie for generous support.

AUTHOR CONTRIBUTIONS

S.S. coordinated and contributed to sequencing and finishing of the genome, analysis of genome data and to the writing of the manuscript, V.A.P.M.d.S wrote the manuscript, analyzed data and coordinated the annotation strategy and the analysis of genome information, D.B. contributed to the genome annotation and data analysis, T.B. contributed to the processing of sequence reads and quality assessment, M.B. contributed to the data analysis, J.B. contributed to the annotation, T.N.C. contributed to the annotation, R.D. produced the qRT-PCR data, M.F. contributed to the annotation and produced data on enzyme characterization, C.G. contributed to the annotation, A.G. contributed to the automatic and manual annotation and the coordination of software development for genome analysis, O.V.G. contributed to the annotation, F.K. contributed to the annotation, A.N.K. contributed to the annotation, data analysis and carried out genome surveys, S.L. contributed to data analysis and the writing of the manuscript, B.L. contributed to the maintenance and improvement of GenDB web interface, A.C.McH. contributed to the gene prediction, F.M. contributed to the annotation, T.N. contributed to the annotation, A.P. was co-responsible for the project as a whole and

contributed to the writing of the manuscript, D.R. contributed to the annotation and preliminary proteomic data, O.R. contributed to the detailed analysis of genome features and production of genome visualizations, J.S.S. contributed to the analysis of data and annotation, W.S. contributed to data analysis and writing of the manuscript, M.M.Y. contributed to data analysis and produced supporting gene expression data, K.N.T. initiated the project and contributed to conception of the work and manuscript writing, F.-J.V. contributed to the annotation, S.W. contributed to the annotation, O.K. coordinated the sequencing and finishing of the genome and analysis of the data, P.N.G. coordinated the project and contributed to the analysis of the data and annotation, and the writing of the manuscript.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at <http://www.nature.com/naturebiotechnology/>

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>

1. Medina-Bellver, J.I. *et al.* Evidence for *in situ* crude oil biodegradation after the Prestige oil spill. *Environ. Microbiol.* **7**, 773–779 (2005).
2. Yakimov, M.M. *et al.* *Alcanivorax borkumensis* gen. nov., sp. nov., a new, hydrocarbon-degrading and surfactant-producing marine bacterium. *Int. J. Syst. Bacteriol.* **48**, 339–348 (1998).
3. Harayama, S., Kishira, H., Kasai, Y. & Syutsubo, K. Petroleum biodegradation in marine environments. *J. Mol. Microbiol. Biotechnol.* **1**, 63–70 (1999).
4. Kasai, Y., Kishira, H., Syutsubo, K. & Harayama, S. Molecular detection of marine bacterial populations on beaches contaminated by the Nakhodka tanker oil-spill accident. *Environ. Microbiol.* **3**, 246–255 (2001).
5. Hara, A., Syutsubo, K. & Harayama, S. *Alcanivorax* which prevails in oil-contaminated seawater exhibits broad substrate specificity for alkane degradation. *Environ. Microbiol.* **5**, 746–753 (2003).
6. Yakimov, M.M. *et al.* Natural microbial diversity in superficial sediments of Milazzo Harbor (Sicily) and community successions during microcosm enrichment with various hydrocarbons. *Environ. Microbiol.* **7**, 1426–1441 (2005).
7. Röling, W.F.M. *et al.* Bacterial community dynamics and hydrocarbon degradation during a field-scale evaluation of bioremediation on a mudflat beach contaminated with buried oil. *Appl. Environ. Microbiol.* **70**, 2603–2613 (2004).
8. Golyshin, P.N. *et al.* Genome sequence completed of *Alcanivorax borkumensis*, a hydrocarbon-degrading bacterium that plays a global role in oil removal from marine systems. *J. Biotechnol.* **106**, 215–220 (2003).
9. Syutsubo, K., Kishira, H. & Harayama, S. Development of specific oligonucleotide probes for the identification and *in situ* detection of hydrocarbon-degrading *Alcanivorax* strains. *Environ. Microbiol.* **3**, 371–379 (2001).
10. van Beilen, J.B. *et al.* Characterization of two alkane hydroxylase genes from the marine hydrocarbonoclastic bacterium *Alcanivorax borkumensis*. *Environ. Microbiol.* **6**, 264–273 (2004).
11. Abraham, W.R., Meyer, H. & Yakimov, M. Novel glycine containing glucolipids from the alkane using bacterium *Alcanivorax borkumensis*. *Biochim. Biophys. Acta* **1393**, 57–62 (1998).
12. Hara, A. *et al.* Cloning and functional analysis of *alkB* genes in *Alcanivorax borkumensis* SK2. *Environ. Microbiol.* **6**, 191–197 (2004).
13. Kok, M. *et al.* The *Pseudomonas oleovorans* alkane hydroxylase gene – sequence and expression. *J. Biol. Chem.* **264**, 5435–5441 (1989).
14. van Beilen, J.B. *et al.* Analysis of *Pseudomonas putida* alkane-degradation gene clusters and flanking insertion sequences: evolution and regulation of the *alk* genes. *Microbiology* **147**, 1621–1630 (2001).
15. van Beilen, J.B., Li, Z., Duetz, W.A., Smits, T.H.M. & Witholt, B. Diversity of alkane hydroxylase systems in the environment. *Oil Gas Sci. Technol.* **58**, 427–440 (2003).
16. Danot, O. A complex signalling module governs the activity of MalT, the prototype of an emerging transactivator family. *Proc. Natl. Acad. Sci. USA* **98**, 435–440 (2001).
17. Smits, T.H.M., Balada, S.B., Witholt, B. & van Beilen, J.B. Functional analysis of alkane hydroxylases from Gram-negative and Gram-positive bacteria. *J. Bacteriol.* **184**, 1733–1742 (2002).
18. van Beilen, J.B. *et al.* Identification of an amino acid position that determines the substrate range of integral membrane alkane hydroxylases. *J. Bacteriol.* **187**, 85–91 (2005).
19. Venter, J.C. *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66–74 (2004).
20. Dutta, T.K. & Harayama, S. Biodegradation of *n*-alkylcycloalkanes and *n*-alkylbenzenes via new pathways in *Alcanivorax* sp. strain MBIC 4326. *Appl. Environ. Microbiol.* **67**, 1970–1974 (2001).
21. Sabirova, J.S., Ferrer, M., Regenhardt, D., Timmis, K.N. & Golyshin, P.N. Proteomic insights into metabolic adaptations in *Alcanivorax borkumensis* induced by alkane utilization. *J. Bacteriol.* **188**, 3763–3773 (2006).
22. Maier, T., Förster, H.H., Asperger, O. & Hahn, U. Molecular characterization of the 56-kDa CYP153 from *Acinetobacter* sp. EB104. *Biochem. Biophys. Res. Commun.* **286**, 652–658 (2001).

23. Passeri, A. *et al.* Marine biosurfactants. 4. Production, characterization and biosynthesis of an anionic glucose lipid from the marine bacterial strain MM1. *Appl. Microbiol. Biotechnol.* **37**, 281–286 (1992).
24. Ochsner, U.A., Fiechter, A. & Reiser, J. Isolation, characterization, and expression in *Escherichia coli* of the *Pseudomonas aeruginosa rhlAB* genes encoding a rhamnolytransferase involved in rhamnolipid biosurfactant synthesis. *J. Biol. Chem.* **269**, 19787–19795 (1994).
25. Toren, A., Orr, E., Paitan, Y., Ron, E.Z. & Rosenberg, E. The active component of the bioemulsifier alasin from *Acinetobacter radioresistens* KA53 is an OmpA-like protein. *J. Bacteriol.* **184**, 165–170 (2002).
26. Ron, E.Z. & Rosenberg, E. Natural roles of biosurfactants. *Environ. Microbiol.* **3**, 229–236 (2001).
27. Averhoff, B. & Friedrich, A. Type IV pili-related natural transformation systems: DNA transport in mesophilic and thermophilic bacteria. *Arch. Microbiol.* **180**, 385–393 (2003).
28. Berman, T. & Bronk, D.A. Dissolved organic nitrogen: a dynamic participant in aquatic ecosystems. *Aquat. Microb. Ecol.* **31**, 279–305 (2003).
29. Hou, S. *et al.* Genome sequence of the deep-sea gamma-proteobacterium *Idiomarina loihiensis* reveals amino acid fermentation as a source of carbon and energy. *Proc. Natl. Acad. Sci. USA* **101**, 18036–18041 (2004).
30. Head, I.M., Jones, D.M. & Røling, F.M. Marine microorganisms make a meal of oil. *Nat. Rev. Microbiol.* **4**, 173–182 (2006).
31. Dyksterhouse, S.E., Gray, J.P., Herwig, R.P., Lara, J.C. & Staley, J.T. *Cycloclastis pugetii* gen. nov., sp. nov., an aromatic hydrocarbon-degrading bacterium from marine sediments. *Int. J. Syst. Bacteriol.* **45**, 116–123 (1995).
32. Yanisch-Perron, C., Vieira, J. & Messing, J. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103–119 (1985).
33. Ewing, B., Hillier, L., Wendl, M. & Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**, 175–185 (1998).
34. Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**, 186–194 (1998).
35. Kaiser, O. *et al.* Whole genome shotgun sequencing guided by bioinformatics pipelines—an optimized approach for an established technique. *J. Biotechnol.* **106**, 121–133 (2003).
36. Gordon, D., Abajian, C. & Green, P. Consed: A graphical tool for sequence finishing. *Genome Res.* **8**, 195–202 (1998).
37. Gordon, D., Desmarais, C. & Green, P. Automated finishing with Autofinish. *Genome Res.* **11**, 614–625 (2001).
38. Bartels, D. *et al.* BACCardI—a tool for the validation of genomic assemblies, assisting genome finishing and intergenome comparison. *Bioinformatics* **21**, 853–859 (2005).
39. Tauch, A. *et al.* Strategy to sequence the genome of *Corynebacterium glutamicum* ATCC 13032: use of a cosmid and a bacterial artificial chromosome library. *J. Biotechnol.* **95**, 25–38 (2002).
40. Meyer, F. *et al.* GenDB—an open source genome annotation system for prokaryote genomes. *Nucleic Acids Res.* **31**, 2187–2195 (2003).
41. McHardy, A.C., Goesmann, A., Pühler, A. & Meyer, F. Development of joint application strategies for two microbial gene finders. *Bioinformatics* **20**, 1622–1631 (2004).
42. Suzek, B.E., Ermolaeva, M.D., Schreiber, M. & Salzberg, S.L. A probabilistic method for identifying start codons in bacterial genomes. *Bioinformatics* **17**, 1123–1130 (2001).
43. Lowe, T.M. & Eddy, S.R. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964 (1997).
44. Tatusov, R.L. *et al.* The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4**, 41 (2003).
45. Harris, M.A. *et al.* The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* **32**, D258–D261 (2004).
46. Altschul, S.F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
47. Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. & Higgins, D.G. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876–4882 (1997).
48. Perrière, G. & Gouy, M. WWW-Query: an on-line retrieval system for biological sequence banks. *Biochimie* **78**, 364–369 (1996).
49. Kumar, S., Tamura, K., Jakobsen, I.B. & Nei, M. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**, 1244–1245 (2001).
50. Stover, C.K. *et al.* Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* **406**, 959–964 (2000).
51. Nelson, K.E. *et al.* Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environ. Microbiol.* **4**, 799–808 (2002).