ELSEVIER
CrossMark
Institut Pasteur
Research in Microbiology 164 (2013) 439–449
www.elsevier.com/locate/resmic

The genome sequence of the hydrocarbon-degrading *Acinetobacter venetianus* VE-C3

Marco Fondi [a,b,1], Ermanno Rizzi [c,1], Giovanni Emiliani [a], Valerio Orlandini [a], Luisa Berna [d,e], Maria Cristiana Papaleo [a], Elena Perrin [a], Isabel Maida [a], Giorgio Corti [c,i], Gianluca De Bellis [c], Franco Baldi [f], Lenie Dijkshoorn [g], Mario Vaneechoutte [h], Renato Fani [a,*]

[a] *Laboratory of Microbial and Molecular Evolution, Dept. of Biology, University of Florence, Via Madonna del Piano 6, Sesto Fiorentino (FI), Italy*
[b] *Computer Laboratory, Cambridge University, William Gates Building 15, JJ Thomson Avenue, Cambridge, United Kingdom*
[c] *Istituto di Tecnologie Biomediche, Consiglio Nazionale delle Ricerche (ITB-CNR), Segrate (MI), Italy*
[d] *Sección Biomatematica, Facultad de Ciencias, Universidad de la República, Iguá 4225, Montevideo, Uruguay*
[e] *Department of Preclinical and Clinical Pharmacology, University of Florence, viale G. Pieraccini 6, 50139 Firenze, Italy*
[f] *Dipartimento di Scienze Molecolari e Nanosistemi (DSMN), Cà Foscari, Università di Venezia, 30123 Venezia, Italy*
[g] *Department of Infectious Diseases, Leiden University Medical Center, PO Box 9600, 2300 RC Leiden, The Netherlands*
[h] *Laboratory Bacteriology Research, Faculty Medicine & Health Sciences, University of Ghent, Belgium*
[i] *Institute for Cancer Research and Treatment, Candiolo (TO), Italy*

Received 27 July 2012; accepted 8 March 2013
Available online 23 March 2013

Abstract

Here we report the genome sequence of *Acinetobacter venetianus* VE-C3, a strain isolated from the Venice Lagoon and known to be able to degrade *n*-alkanes. Post sequencing analyses revealed that this strain is relatively distantly related to the other *Acinetobacter* strains completely sequenced so far as shown by phylogenetic analysis and pangenome analysis (1285 genes shared with all the other *Acinetobacter* genomes sequenced so far). *A. venetianus* VE-C3 possesses a wide range of determinants whose molecular functions are probably related to the survival in a strongly impacted ecological niche. Among them, genes probably involved in the metabolism of long-chain *n*-alkanes and in the resistance to toxic metals (e.g. arsenic, cadmium, cobalt and zinc) were found. Genes belonging to these processes were found both on the chromosome and on plasmids. Also, our analysis documented one of the possible genetic bases underlying the strategy adopted by *A. venetianus* VE-C3 for the adhesion to oil fuel droplets, which could account for the differences existing in this process with other *A. venetianus* strains. Finally, the presence of a number of DNA mobilization-related genes (i.e. transposases, integrases, resolvases) strongly suggests an important role played by horizontal gene transfer in shaping the genome of *A. venetianus* VE-C3 and in its adaptation to its special ecological niche.

Keywords: *Acinetobacter*; Alkane metabolism; Microbial genomics; Bioremediation

1. Introduction

The marine environment is subjected to the contamination by organic pollutants from a variety of sources, with crude oil being one of the most important substances (Head and Swannell, 1999). Alkanes are the major components of crude oils and are commonly found in oil-contaminated environments (Feng et al., 2007). Aerobic *n*-alkane degradation is a widespread phenomenon in nature, and several microbial species/strains and enzymes involved in *n*-alkane degradation

* Corresponding author.
E-mail addresses: marco.fondi@unifi.it (M. Fondi), ermanno.rizzi@itb.cnr.it (E. Rizzi), giovanni.emiliani@unifi.it (G. Emiliani), valerio.orlandini@gmx.com (V. Orlandini), luisa.berna@unifi.it (L. Berna), cristiana.papaleo@unifi.it (M.C. Papaleo), elena.perrin@unifi.it (E. Perrin), isabel.maida@unifi.it (I. Maida), giorgio.corti@ircc.it (G. Corti), gianluca.debellis@itb.cnr.it (G. De Bellis), baldi@unive.it (F. Baldi), L.Dijkshoorn@lumc.nl (L. Dijkshoorn), mario.vaneechoutte@ugent.be (M. Vaneechoutte), renato.fani@unifi.it (R. Fani).
[1] Equal contributors.

0923-2508/$ - see front matter © 2013 Institut Pasteur. Published by Elsevier Masson SAS. All rights reserved.
http://dx.doi.org/10.1016/j.resmic.2013.03.003

have been identified, isolated and studied in detail (Throne-Holst et al., 2006).

In most described cases, the *n*-alkane is oxidized to the corresponding primary alcohol by substrate-specific terminal monooxygenases/hydroxylases. Two unrelated classes of enzymes for long-chain *n*-alkane oxidation have been proposed: (1) cytochrome P450-related enzymes in both yeasts and bacteria, e.g., bacterial CYP153 enzymes, and (2) bacterial alkane hydroxylases (pAHs) (Wentzel et al., 2007). The latter class of integral membrane non-heme diiron alkane monooxygenases of the AlkB-type allows a wide range of microorganisms to grow on *n*-alkanes with carbon chain lengths from C5 to C16 (van Beilen and Funhoff, 2007). AlkB-type enzymes function in complex with two electron transfer proteins, a dinuclear iron rubredoxin, and a mononuclear iron rubredoxin reductase channeling electrons from NADH to the active site of the alkane hydroxylase (van Beilen and Funhoff, 2007). After the initial oxidation of the *n*-alkane, the corresponding alcohol is oxidized step by step by alcohol dehydrogenase and aldehyde hydrogenase to the corresponding aldehyde and carboxylic acid, respectively. The carboxylic acid then serves as a substrate for acyl-CoA-synthase, and the resulting acyl-CoA enters the β-oxidation pathway (van Beilen and Funhoff, 2007).

Degradation of *n*-alkanes through this kind of catabolic pathway has been extensively studied in *Pseudomonas putida* GPo1 [formerly *Pseudomonas oleovorans* (Baptist et al., 1963; van Beilen et al., 2001)]. Several bacterial strains able to degrade C5–C10 alkanes contain alkane hydroxylases that belong to a distinct family of soluble cytochrome P450 monooxygenases (Wentzel et al., 2007) as, for example, *Acinetobacter* sp. EB104 (Maier et al., 2001) and representatives from mycobacteria, rhodococci and proteobacteria (Sekine et al., 2006; van Beilen et al., 2005, 2006). Alternative alkane hydroxylases have been found in those microorganisms capable of degrading alkanes longer than C_{20}. Usually, these enzymes are not evolutionary related to known AlkB- and P450-like sequences and include AlmA (a flavin binding monooxygenase able to oxidize C_{20}–C_{32} alkanes) from *Acinetobacter* strain DSM 17874 (Throne-Holst et al., 2006) and LadA from *Geobacillus thermodenitrificans* (Feng et al., 2007), able to generate primary alcohols from C_{15}–C_{36} alkanes.

Finally, the cell contact with hydrophobic substrates is crucial because the initial step of alkane degradation is usually carried out by oxidation reactions catalyzed by cell-surface associated oxygenases (Foster, 1962; Wentzel et al., 2007). The solubility of low-molecular-weight alkanes is sufficient to mediate the uptake of the alkane from water, whereas uptake of medium- and long-chain-length *n*-alkanes occurs by either adhesion to hydrocarbon droplets or by a surfactant-facilitated process (Rojo, 2009).

Many alkane-degrading bacteria secrete diverse surfactants that facilitate emulsification of hydrocarbons (Hommel, 1990; Ron and Rosenberg, 2002). In particular, among surfactant producers, *Acinetobacter venetianus* RAG-1^T (Reisfeld et al., 1972; Vaneechoutte et al., 1999) has been shown to produce an extracellular anionic lipoheteropolysaccharide, known as emulsan, to aid in the capture and transport of the carbon sources to the cell (Mercaldi et al., 2008; Pines et al., 1983; Zuckerberg et al., 1979). A 27 kbp cluster of genes responsible for the biosynthesis of this amphipathic, polysaccharide bioemulsifier from the oil-degrading *A. venetianus* RAG-1^T was isolated and characterized (Nakar and Gutnick, 2001). The draft genome of this strain was recently obtained (Fondi et al., 2012) and is likely to provide further insight into the genetic basis of its alkane degradation and emulsan production in this strain. Genomes of other oil-degrading bacteria have been obtained in recent years, including those of *Acinetobacter* sp. DR1 (Kang et al., 2011), *Alcanivorax borkumensis* (Schneiker et al., 2006) and *Marinobacter aquaeolei* VT8 (Kostka et al., 2011).

Some light has been shed on the different strategies for diesel fuel degradation adopted by different *Acinetobacter* strains, suggesting a good efficiency in this process by *A. venetianus* (Mara et al., 2012). Thus, use of strains of this species in bioremediation might provide valuable advances in this important biological/biotechnological field. One *A. venetianus* strain, VE-C3, was isolated in 1993 (Baldi et al., 1997) from the superficial waters of the former industrialized Marghera Port in the Venice lagoon. This area has been polluted due to oil refineries for decades up to late nineties although, today, it is a dismissed and remediated area. VE-C3 strain has been shown to grow on C10 and C14 (Mara et al., 2012). The overall genetic and functional understanding of its biologically and biotechnologically relevant phenotype, including its adhesion to hydrocarbon molecules, its subsequent metabolism and its adaption to its peculiar ecological niche is far from complete. To address these points whole genome sequencing of the marine, hydrocarbon-degrading bacterium *A. venetianus* VE-C3 was performed by the use of a comprehensive approach that combined the next-generations sequencing (NGS) platforms Roche/454 and Illumina with the classical Sanger sequencing of PCR products. Genome analysis yielded interesting insights into the biology of this strain, allowing the identification of putative niche-adaptation and bioremediation-related specific gene sets.

2. Material and methods

2.1. Sample preparation and genome sequencing

Genomic DNA extraction was carried out as previously described (Giovannetti et al., 1990). A first single stranded Roche/454 library was then prepared starting from 5 μg of *A. venetianus* VE-C3 DNA and used to perform the shotgun sequencing, following the procedure as reported in the Roche/454 standard protocol. A second library was prepared in order to obtain paired ends reads. For this purpose, another aliquot of 5 μg of genomic DNA was fragmented to obtain fragments of an average size of 3 kb using the HydroShear apparatus (Digilab Inc., Holliston, MA, USA). These fragments were converted into a paired ends single stranded library following the Roche/454 procedure as reported in the 3 kb paired ends library preparation method manual. Both libraries were

quantitated by the Ribo Green assay (Invitrogen Inc, Carlsbad, CA, USA), amplified by emulsion PCR as reported in the Roche/454 procedure and sequenced using the Titanium version of the Genome Sequencer FLX System. A total of about 1.2×10^5single shotgun reads and about 1.5×10^5 paired ends reads were obtained. Illumina sequencing of *A. venetianus* VE-C3 was carried out by IGA (Istituto di Genomica Applicata, Udine, Italy) with Illumina HiSeq2000.

2.2. Genome assembly

The Roche/454 paired ends reads were assembled with Roche assembler (Newbler). From this assembly, 111 contigs embedded in 14 different scaffolds were generated. Two scaffolds corresponded to already sequenced *A. venetianus* VE-C3 plasmids, namely pAV1 and pAV2 (Mengoni et al., 2007), and where therefore not considered in the further stages of genome assembly.

Illumina GA1 reads were first trimmed to eliminate low quality base callings. Trimming was performed adopting the dynamic trimming algorithm embedded in the SolexaQA suite (Cox et al., 2010), selecting a Phred score threshold value of 30. Further on, reads were assembled using the Abyss assembler v. 1.2.7 (Simpson et al., 2009) with a *k*-mer size of 51. This resulted in a preliminary assembly of 438 contigs.

In order to integrate the two different assemblies, we combined the contigs generated by Illumina and Roche/454 technologies in a hybrid assembly using Phrap assembler (de la Bastide and McCombie, 2007). This resulted in an improved assembly embedding 12 scaffold and 27 contigs. Part of the remaining gaps, within and among the different scaffolds were closed adopting a computational approach based on the mapping of original Roche/454 reads at the extremities of scaffolds (Fondi et al., submitted for publication).

The closure of the gaps among the different scaffolds was then validated with PCR amplification and Sanger sequencing of the overlapping regions. PCR amplification coupled with Sanger sequencing was also performed in those cases in which the adopted computational gap closure strategy (described above) failed to reconstruct the correct order of the scaffolds.

2.3. Genome annotation

Genome annotation was performed using the Rapid Annotation by Subsystem Technology (RAST) pipeline (Aziz et al., 2008). Additional functional annotation was performed querying other functional databases, including KAAS (Moriya et al., 2007), Interpro (*via* Interproscan (Quevillon et al., 2005)) and COG (Clusters of Orthologous Groups) (Tatusov et al., 2003). Atypical chromosomal regions were identified with AlienHunter tool (Vernikos and Parkhill, 2006) using default parameters.

2.4. Genomes retrieval and orthologs identification

Acinetobacter genomes were downloaded from NCBI database as on November 2011 and included *Acinetobacter baumannii* ATCC_17978 (NC_009085), *Acinetobacter calcoaceticus* PHEA-2 (CP002177), *Acinetobacter* sp. DR1 (NC_014259), *Acinetobacter baylyi* ADP1 (NC_014259) and the newly sequenced *A. venetianus* RAG-1^T (AKIQ01000000). When comparing the different *Acinetobacter* genomes (including the newly sequenced *A. venetianus* VE-C3), the groups of orthologous genes were identified with Inparanoid and Multiparanoid softwares (Alexeyenko et al., 2006; Remm et al., 2001). Amino acid sequences of the proteins used to build *Acinetobacter* reference phylogeny were retrieved adopting the Bidirectional Best Hit (BBH) criterion and using the *A. baylyi* ADP1 sequences (retrieved from the Ribosomal Database Project (RDP), http://rdp.cme.msu.edu/) as queries.

2.5. Sequence alignment and phylogenetic tree construction

Multiple sequence alignments were performed using Muscle (Edgar, 2004) and misaligned regions were visually inspected and removed where necessary.

Maximum Likelihood (ML) analysis was carried out using Phyml (Guindon et al., 2005, 2009), with a WAG model of amino acid substitution, including a gamma function with 6 categories to take into account differences in evolutionary rates at sites. Statistical support at nodes was obtained by non-parametric bootstrapping on 1000 re-sampled datasets.

2.6. Permutation tests

To assess whether the accessory genome was enriched in a particular functional category, the proportions of the COGs (Tatusov et al., 2003) in the core and accessory genome were compared. Statistical significance to the enrichment analysis was gained through permutation tests on the original gene sets, i.e. one million random samplings were performed and the COG proportions of each sample were compared to a sample from the whole genome. *P*-values below 0.05 were considered to be significant.

2.7. Genome accession numbers

This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession ALIG00000000. The version described in this paper is the first version, ALIG01000000.

3. Results and discussion

3.1. Genome overview

After Roche/454 and Illumina sequencing of the *A. venetianus* VE-C3 genome, a total of 3,564,836 bp bases were assembled; 10,820 bp and 15,135 bp corresponded to the already known pAV1 and pAV2 plasmids (Mengoni et al., 2007) respectively, whereas 186.446 corresponded to the newly identified pAV3 (large) plasmid (Fig. 1). The remaining

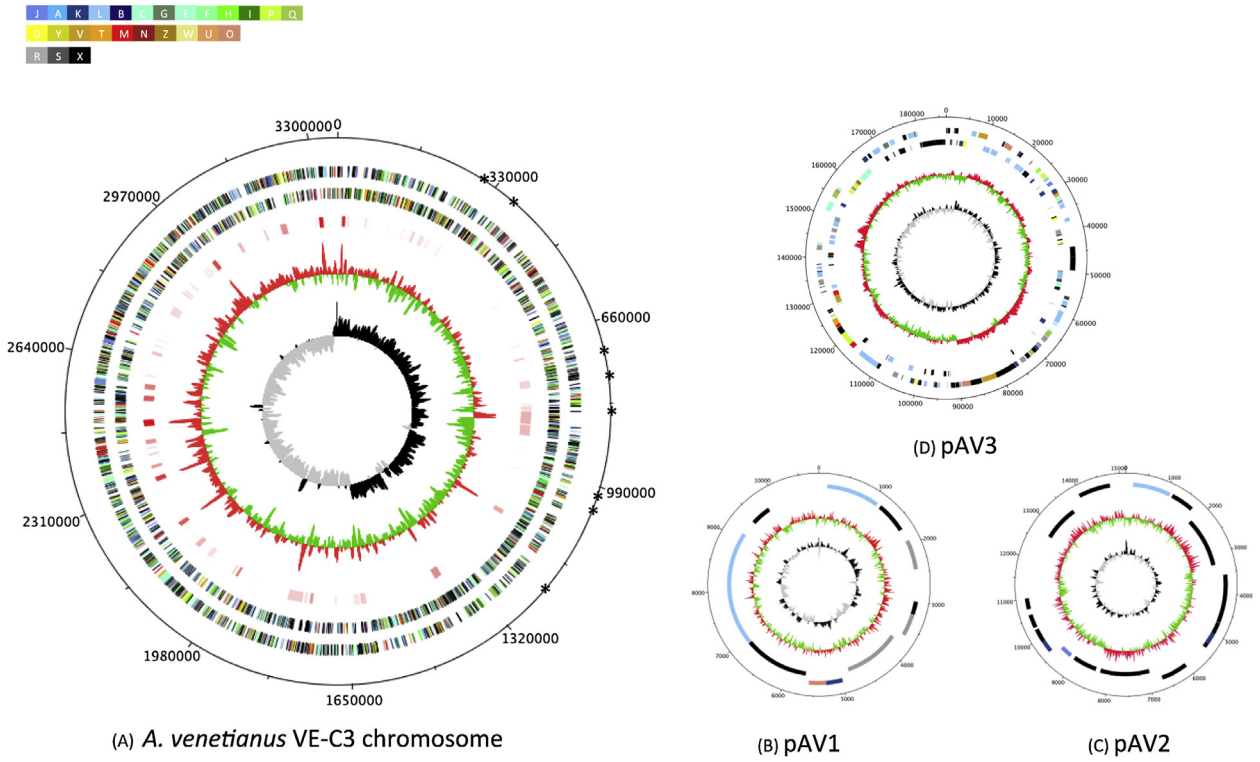

Fig. 1. Circular representations of *A. venetianus* VE-C3 chromosome and plasmids displaying relevant genome features. From the outer to the inner concentric circle: circle 1, genomic position in kb; circles 2 and 3, predicted protein coding sequences (CDS) on the forward (outer wheel) and the reverse (inner wheel) strands coloured according to the assigned COG classes; circle 4 represents atypical chromosomal regions identified by AlienHunter (see Material and methods); circle 5, G + C content showing deviations from the average (39%); circle 6, GC skew. The bar below the plot represents the COG colours 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, defence mechanisms; X, unknown function). Asterisks in (A) represent the regions of the scaffold still containing gaps.

3,352,435 bp were assembled in a single scaffold (embedding 8 contigs) and represented the *A. venetianus* VE-C3 chromosome. A total of 3472 coding sequences (CDS) were identified and a putative function was assigned to about 2675 of them (77%) (see Supplementary Material 1 for the complete list of encoded functions). Additionally, *A. venetianus* VE-C3 encodes 6 rDNA operons (16S-23S-5S) and 74 tRNAs. The likely origin of replication (*oriC*) on the chromosome was inferred with OriFinder tool (Gao and Zhang, 2008) and confirmed by GC skew analysis (Fig. 1); six DnaA boxes were found in the inferred *oriC* region. The general features of the *A. venetianus* VE-C3 are reported in Table 1.

A summary of the functional capabilities of *A. venetianus* VE-C3 as a result of a BLAST search of its ORFs only against the COG database (Tatusov et al., 2003) is reported in Supplementary Material 1. Remarkably, with the exception of sequences without a clear assigned function (embedded in functional COG categories X, R and S and representing 26.30%, 7.94% and 7.34% of the total gene content, respectively), those embedded in functional category L (replication, recombination and repair) are the most abundant. This general COG category embeds, among the others, those sequences involved in the recombination (e.g. COG1381), transposition (e.g. COG3676, COG3547) and integration (e.g. COG0582) of DNA fragments, thus revealing that the exogenous acquisition (or loss) of genes might have played a role in shaping the genome of this strain, possibly with the intervention of mobile genetic elements. Interestingly, the proportion of genes

Table 1
General features of the *A. venetianus* VE-C3 genome.

DNA molecule	Chromosome	pAV1	pAV2	pAV3
Size (nucleotides)	3,352,435	10,820	15,135	186,446
GC-content (%) of chromosome	39.11	34.56	36.41	39.56
Protein coding genes	3255	12	16	189
Hypothetical proteins	732	3	11	51
Functions assigned	2523	9	5	138
Average protein length (amino acids)	292.77	210.92	217.41	260.80
Maximum protein length (amino acids)	1797	738	446	1863
rRNA operons (16S-23S-5S)	6	0	0	0
tRNAs	74	0	0	0

involved in this process in *A. venetianus* VE-C3 is larger than that observed in genomes of other bacteria, oil degrading as well as non-degrading. Indeed, the number of genes belonging to the COG category of transposition/mobilization/integration (COG category "L") was calculated also in the genomes of 3 arbitrarily chosen oil-degrading bacteria (i.e.: *A. borkumensis* SK2, *Acinetobacter oleivorans* DR1, and *M. aquaeolei* VT8) and 3 non-degrading bacteria (*Acinetobacter baumanii* ATCC_17978, *Bacillus subtilis* subsp. *subtilis* str. 168 and *Escherichia coli* K12). Results showed that *A. venetianus* VE-C3 possesses, on average, more genes belonging to the L COG category than both oil-degrading bacteria (7.10% against 4.46%, 6.5% and 2.71% in *A. borkumensis* SK2, *M. aquaeolei* VT8 and *Acinetobacter* sp. DR1, respectively) and non-degrading bacteria (7.10% against 3.62%, 3.1% and 5.2% in *A. baumanii* ATCC_17978, *Bacillus subtilis* subsp. *subtilis* str. 168 and *E. coli* K12, respectively). Comparison between plasmids and chromosome-encoded functions (Supplementary Material 2) shows, as expected, an overall predominance of transposition/integration/recombination related genes in plasmids rather than in the chromosome.

The massive presence of recombination related genes in *A. venetianus* VE-C3 is also confirmed by the comparative analysis with the genomes of other representatives of this genus (see below) and from previous work that showed a high level of horizontal gene transfer (HGT) and recombination events (also occurring within the same cell) in the *Acinetobacter* genus (Fondi et al., 2010). Interestingly, *A. venetianus* RAG-1^T possesses about half of the amount of the recombination related genes in respect to *A. venetianus* VE-C3 [117 (3.4%) and 254 (7.3%), respectively] and to other *A. venetianus* strains (Fondi et al., manuscript in preparation), suggesting that the abundance of this particular class of genes might be a peculiarity of VE-C3 strain.

For a deeper inspection of the *A. venetianus* VE-C3 chromosome we adopted the computational strategy implemented in the AlienHunter software, which exploits compositional biases using variable order motif distributions and captures the local composition of a sequence compared with fixed-order methods (Vernikos and Parkhill, 2006). This allowed the identification, although with different confidence values, of 66 atypical regions (Fig. 1), probably the outcome of chromosomal recombination and/or HGT events. This observation is supported by the fact that, as shown in Fig. 2, most of the genes found in these regions encode proteins that either have no homologs in the COG database or code for proteins likely involved in the recombination/transposition/integration of DNA fragments (58.7% of the proteins found in atypical regions). Indeed, it is known that the pool of genes responsible in the horizontal flow of genetic information, usually encodes proteins whose function is unknown yet (Bosi et al., 2011; Brilli et al., 2008; Tamminen et al., 2012). Nevertheless, genes belonging to other functional categories not strictly associated to the process of HGT itself [e.g. transcription factors, inorganic ion transport and metabolism (e.g. arsenic resistance), and defence mechanisms (e.g. ABC-type multidrug transport system)] were found within these regions, supporting the idea that HGT might have been a key player in the adaptation of this microorganism to the heavily polluted ecosystem of Venice Lagoon.

3.2. Genetic basis of alkane degradation in A. venetianus *VE-C3*

3.2.1. Adhesion to oil fuel

It has been suggested that *A. venetianus* VE-C3 is capable of two types of adhesion, i.e. (i) cell-to-cell interactions, preceding the cell adhesion to the *n*-alkane molecules, and (ii)

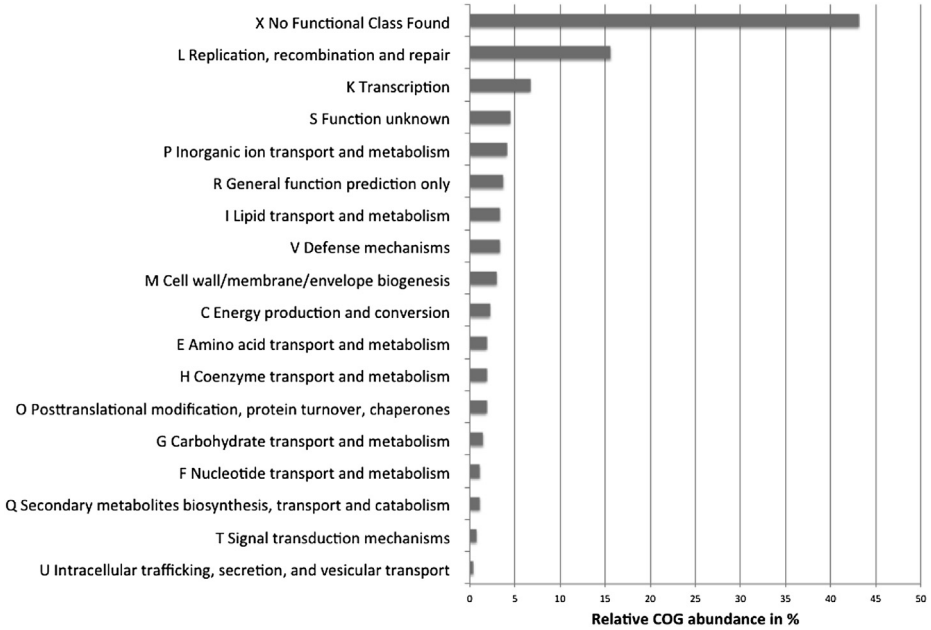

Fig. 2. Clusters of orthologous groups (COG) functional categories distribution of the genes that were found inside putative atypical regions.

n-alkane nano-micelles embedding into the capsular polysaccharide (Baldi et al., 2003), a mechanism for controlling the carbon source uptake, avoiding direct contact of *n*-alkanes with the membrane structures, which would be disruptive.

The genome sequence of *A. venetianus* VE-C3 offers the possibility to investigate the genetic basis of these two different *n*-alkane adhesion strategies, for which the genes and/or the pathways are still unknown.

Accordingly, the *wee* gene cluster of *A. venetianus* VE-C3 was identified (the complete annotation of the *wee*-like cluster is available in Supplementary Material 3) and compared to the one previously described in *A. venetianus* RAG-1^T. Results of this analysis (shown in Fig. 3) revealed that strains VE-C3 and RAG-1^T share two large portions (14 out of 22 genes are orthologous between RAG-1^T and VE-C3) of their *wee* gene cluster, i.e. its first part (including genes *mip*, *wzc*, *wzb*, *wza*, *weeB*, but lacking *weeA*) and the final one (from *weeH* to *pgm*) (Fig. 3). However, the central part of the cluster partially differs in the two strains: in the central region of its cluster, RAG-1^T harbors genes *weeA/C*, *wzx*, *wzy* and *weeD/E/F/G* while VE-C3 harbors a set of genes that are not orthologous but probably code for proteins which perform similar reactions, at least on the basis of *in silico* annotation, such as the *A. venetianus* VE-C3 *mviM*, encoding an oxidoreductase with similarity to RAG-1^T WeeA, or the three VE-C3 *rfaG* genes (not sharing significant sequence identity among each other), encoding a glucosyl transferase, in correspondence with the *weeG* gene of RAG-1^T. Anyway, it is important to stress that some genes of the RAG-1^T *wee* cluster, such as *wzx* and *wzy*, responsible for polymerisation of the apoemulsan (Nakar and Gutnick, 2001) were not found in the genome of *A. venetianus* VE-C3, explaining the different strategies adopted by these strains for the interaction with *n*-alkanes.

Interestingly, the central portions of both *A. venetianus* RAG-1^T and VE-C3 *wee* clusters are also different in terms of GC-content from the rest of the cluster in which they are embedded and also from the average GC-content of the corresponding genome (red dashed line in Fig. 3), indicating that past recombination and/or transposition event(s) might have shaped the *wee* clusters in both microorganisms.

3.2.2. Metabolism of n-alkanes

To search for the genetic determinants likely responsible for the catabolism of hydrocarbons, a set of sequences known to be involved in this pathway was assembled through bibliographical data mining. This seeds dataset included the complete set of Alk sequences (AlkB, G, H, L, J, K, S, T and N) from *P. putida* GPo1 (van Beilen et al., 2001), soluble cytochrome P450 monooxygenases from *Acinetobacter* sp. EB104 (Maier et al., 2001), AlmA from *Acinetobacter* sp. DSM 17874 (Throne-Holst et al., 2007) and LadA from *G. thermodenitrificans* (Feng et al., 2007).

The *A. venetianus* VE-C3 genome encodes a smaller set of *alk*-like sequences compared to *P. putida* GPo1. Moreover, unlike *P. putida* GPo1, in which all the genes are embedded in a single operon, these genes are scattered throughout the genome of VE-C3. *A. venetianus* VE-C3 encodes two paralogous copies of *alkB*, *alkH* (sharing 38% identity among themselves at the amino acid level) and *alkJ* (sharing 37% identity at the amino acid level), whereas AlkK was retrieved in a single copy. Conversely, no ortholog was retrieved from *A. venetianus* VE-C3 when probing its genome with AlkG, AlkT, AlkN and AlkS sequences. AlkG and AlkT (coding for rubredoxin and rubredoxin reductase, respectively) could be replaced by the *rubA-rubB* operon, as already suggested for *A. borkumensis* (Schelstraete et al., 2010) (Supplementary Material 4). Indeed rubredoxin reductase genes map in many alkane-degrading bacteria separately from the alkane hydroxylase genes (Abraham et al., 1998; Schneiker et al., 2006; Smits et al., 2002). Interestingly, a sequence embedding both a rubredoxin and a rubredoxin reductase domain was identified in the genome of *A. venetianus* VE-C3, suggesting its possible role in the alkane degradation process. Furthermore, the absence of ALkS, known to regulate the expression of the *alkBFGHJKL* operon in *P. putida* GPo1 (Eggink et al., 1987; Kok et al., 1989; Panke et al., 1999), raises the intriguing issue of how *alk*-like genes are regulated in *A. venetianus* VE-C3. Finally, *A. venetianus* VE-C3 lacks AlkN, a gene involved in chemotaxis transduction in *P. putida* GPo1 (van Beilen et al., 2001).

A. venetianus VE-C3 encodes a single cytochrome P450 in an operon-like structure with genes encoding a ferredoxin, an

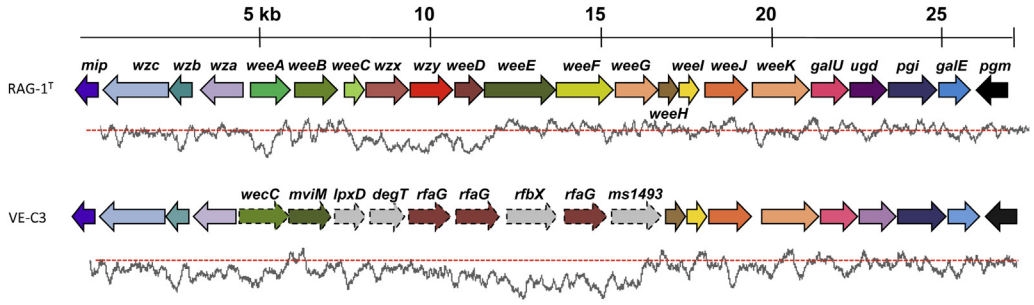

Fig. 3. Schematic representation of the *wee* cluster involved in emulsan production in *A. venetianus* RAG-1^T and VE-C3 strains. Solid line arrows represent orthologous genes between *A. venetianus* VE-C3 and RAG-1^T; dashed line arrows represent genes that are not othologous between the two strains, a similar colour indicates functional analogy. Grey lines below the gene clusters represent GC-content (calculated using a sliding window approach with steps of 100 nucleotides) in respect to the average of the corresponding genome (red dashed line). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

FAD-dependent oxyreductase and a gene encoding an AraC transcriptional regulator (Supplementary Material 4). Interestingly, these genes are encoded by the largest plasmid harbored by *A. venetianus* VE-C3 (pAV3) and are flanked by others encoding a transposase and a resolvase. This, in addition to the observation that the organization of this cluster resembles that found, for example, in *A. borkumensis* SK2 (Schneiker et al., 2006), points to the possible acquisition of these genes by *A. venetianus* VE-C3 through one (or more) HGT event(s). Orthologs of LadA and AlmA encoding genes were also found when probing the genome of VE-C3 (Supplementary Material 4).

The presence/absence pattern of the genes coding for emulsan production (*A. venetianus* RAG-1 wee cluster) and for alkane degradation (*alk* genes from *P. putida* GPo1 (van Beilen et al., 2001), was also compared against the genomes of a set of 3 well-characterised oil-degrading bacteria: *A. borkumensis* SK2 (Schneiker et al., 2006), *Acinetobacter* sp. DR1 (Kang et al., 2011) and *M. aquaeolei* VT8 (Kostka et al., 2011) (Supplementary material 5). Concerning the *wee* cluster involved in fuel oil adhesion, the analysis showed a pattern for *Acinetobacter* sp. DR1 that was quite similar to the one observed in *A. venetianus* VE-C3 (lack of genes orthologous to their counterparts in *A. venetianus* RAG-1^T in the central part of the cluster); for *A. borkumensis* SK2 and *M. aquaeolei* VT8, only the genes from *weeJ* to *pgm* resulted to be present also in their genomes. Also in the case of *alk* genes set *A. venetianus* VE-C3 and *Acinetobacter* sp. DR1 showed a very similar pattern of presence/absence (except for *alkT*, absent in the first and present in latter). *A. borkumensis* SK2, showed a conserved *alkSB$_1$GJH* cluster, as already reported by Scheneiker et al. (2006), which, on the contrary, is absent in *M. aquaeolei* VT8.

3.3. Resistance to heavy metal

In the Venice Lagoon metal contamination by As, Cd, Co, Cr, Cu, Hg, Pb, Zn and others has been reported since decades and has been recently reviewed and elaborated based on the hazard quotients of sediments (Apitz et al., 2007). The *A. venetianus* VE-C3 genome encodes a high number of proteins potentially involved in the resistance to these heavy metals. Indeed, at least two gene clusters coding for CzcCBA complex systems [an RND family system involved in the efflux of such compounds, see (Silver and Phung le, 2005) for a review] were detected, the first also comprising a CzcD-like sequence, presumably involved in the efflux of Cu^{2+} and Zn^{2+} from the cell. Interestingly, two additional copies of CzcD-like coding genes were identified in the genome of *A. venetianus* VE-C3, both of them embedded in a bicistronic cluster together with a MerR-like transcriptional regulator. Besides, the first of these clusters is encoded by the pAV3 plasmid.

A. venetianus VE-C3 harbors three gene clusters potentially involved in arsenic resistance. The 5 genes embedded in each of the three clusters share the very same organization and encode for ArsH (NADPH-dependent FMN reductase), ACR3 (arsenite export protein), ArsC (arsenate reductase), ArsR (arsenic resistance operon repressor) and an additional copy of ArsC. Finally, an extra stand-alone copy of an ArsC coding gene was found in the genome of *A. venetianus* VE-C3. Interestingly, as in the case of cobalt–cadmium–zinc resistance related genes, one of the clusters is located on the major plasmid (pAV3).

Genes potentially involved in the resistance/tolerance to copper and chromium were also identified in the genome of *A. venetianus* VE-C3. In particular, two clusters carrying genes presumably involved in copper homeostasis were identified; one coding for CutE (copper homeostasis protein) and CorC (efflux protein) and another embedding multicopper oxidase and copper resistance protein encoding genes. Concerning chromium resistance, four chromate transport proteins (ChrA-like) coding genes were identified in the genome of *A. venetianus* VE-C3. Two of them are organized in cluster together with a LysR family transcriptional regulator. Another ChrA-like coding gene is embedded in a cluster together with a gene coding for ChrB (chromate resistance signal peptide protein). An additional stand-alone copy of a ChrA-like sequence was also identified. Finally, no genes encoding chromate reduction [from Cr(VI) to the less toxic and less insoluble Cr(III)] were identified in the genome of VE-C3 strain.

It must be stated clearly that all the genes mentioned above have been assigned to a particular functional category (i.e. resistance to a specific metal) only on the basis of their sequence similarity with other, better characterized, sequences. Thus, although in some cases the degree of sequence similarity is relatively high (i.e. above 50% at amino acid level) in cannot be excluded that their role might be slightly different (e.g. conferring resistance to a different compound).

3.4. Genomic comparison of A. venetianus VE-C3 to related species and the Acinetobacter pangenome

To establish the phylogenetic relationship existing between *A. venetianus* VE-C3 and the other representatives of the *Acinetobacter* genus sequenced so far, extensive phylogenetic analysis was conducted using a concatamer of a set of conserved proteins (FusA, IleS, LepA, LeuS, PyrG, RecA, RecG, RplB, RpoB). An alignment was built using these sequences, manually removing ambiguous positions. The result of this analysis (Fig. 4) reveals that *A. venetianus* is only distantly related to representatives of *A. baumannii*, *A. calcoaceticus* and *Acinetobacter* sp. strain DR1. Indeed, the clade embedding *A. venetianus* VE-C3, *A. venetianus* RAG-1^T and *A. baylyi* ADP1 is strongly supported as a sister group to the one embedding the aforementioned species, although the length of their branches suggests the presence of a massive evolutionary divergence between VE-C3 and ADP1 strains. This result was confirmed also when whole genome BLAST comparisons of the strains belonging to the different *Acinetobacter* species (and whose genome was completely sequenced) were carried out (*E*-value threshold: $1e^{-100}$). In this case, to avoid redundancy, only one representative from *A. baumannii* was maintained, namely strain ATCC_17978.

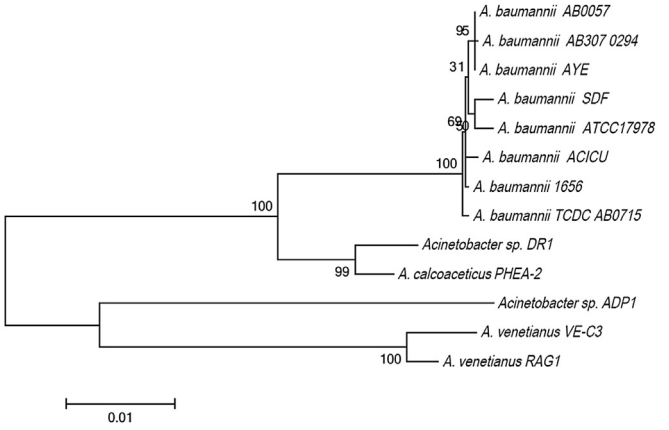

Fig. 4. Maximum likelihood phylogenetic tree showing the evolutionary relationships between *A. venetianus* VE-C3 and the other representatives of the *Acinetobacter* genus available in NCBI database.

BLAST comparisons were then transformed into a circular ideogram in which each genome is represented by an arc and the different genomes (arcs) are connected by vertices accounting for their shared sequence similarity (Fig. 5). It can be noted that *A. venetianus* VE-C3 and *A. baylyi* ADP1 are clearly less interconnected to the other *Acinetobacter* genomes analyzed, suggesting that these strains represent a distinct component of the *Acinetobacter* genus, at least among those strains whose genomes have been completely sequenced.

The genome sequence of *A. venetianus* VE-C3 genome allowed extending the analysis of the whole *Acinetobacter* genus pangenome compared to previous studies (Imperi et al., 2011; Peleg et al., 2012; Vallenet et al., 2008; Zhan et al., 2012). Accordingly, the *in silico* proteome of *A. venetianus* VE-C3 was compared with those of the other *Acinetobacter* representatives whose genome has been completely sequenced and that were previously used to build *Acinetobacter* reference phylogeny (Fig. 5). Also in this case only strain ATCC_17978 from *A. baumannii* was maintained to avoid possible biases due to overrepresentation of genomes belonging to the same species.

By comparing the 3472 CDSs found in the genome of *A. venetianus* VE-C3 with those of the other completely sequenced genomes, a set of orthologous groups was identified. A subset of 1940 CDSs was conserved across all the five genomes and, accordingly, was defined as the *core* genome of the strains belonging to the *Acinetobacter* genus completely sequenced so far.

The remaining 4200 orthologous groups were defined as members of the accessory genome for the five completely sequenced genomes and included both the unique genomes (that is the set of genes peculiar to each strain) and the sets of genes common to only few groups of strains (Fig. 6). In order to define possible differences in functions encoded by the core and/or the accessory genomes of the *Acinetobacter* genus, each protein was assigned to a COG category and the abundance of each COG category was plotted for both core and accessory genomes (Fig. 7). Statistically significant differences between core and accessory genome (computed as described in Material and methods) were found only for COG category L (DNA replication, recombination and repair), V (Defence mechanisms), and for proteins with no assigned COG (X): in these three categories, the accessory genome is enriched. In all the other COG categories (with the exception of functional category K, whose analysis resulted to be statistically unsound) the genes belonging to the core genome are more abundant than those belonging to the accessory genome.

It is interesting to notice that the genome of *A. venetianus* VE-C3 is the one possessing the highest number of unique genes (954, most of which do not have homologs in COG database (51.4% of the unique genes) or only have homologs with no assigned function (9.6%). Furthermore, genes belonging to COG categories L and V were those found to be

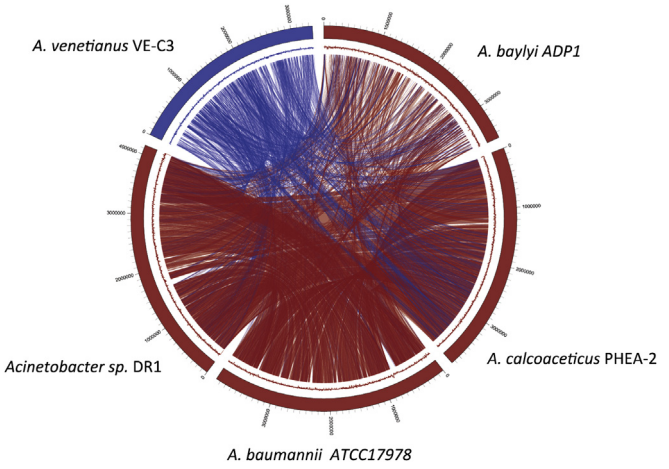

Fig. 5. Circular ideogram representing the comparison among *A. venetianus* VE-C3 genome and those of the other *Acinetobacter* representatives available in NCBI database. Each genome is represented as a bar on the outside of the ideogram together with its GC% content. Links connecting the different bars represent BLAST hits among the different genomes (threshold, *E*-value $1e^{-20}$).

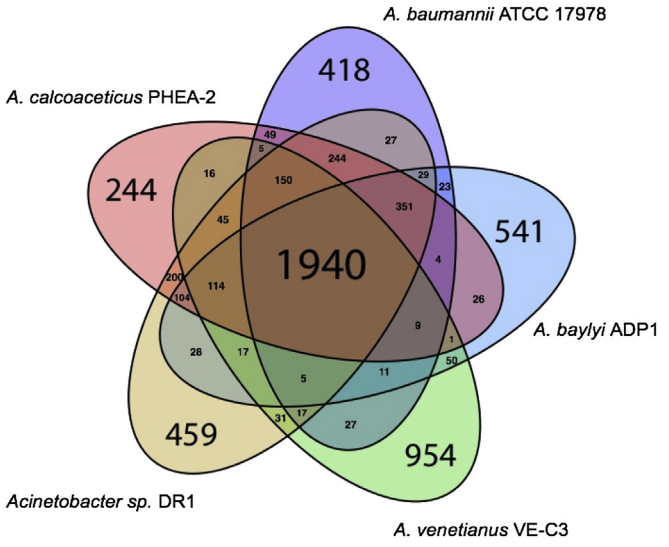

Fig. 6. The core, accessory and unique genomes of the *Acinetobacter* representatives.

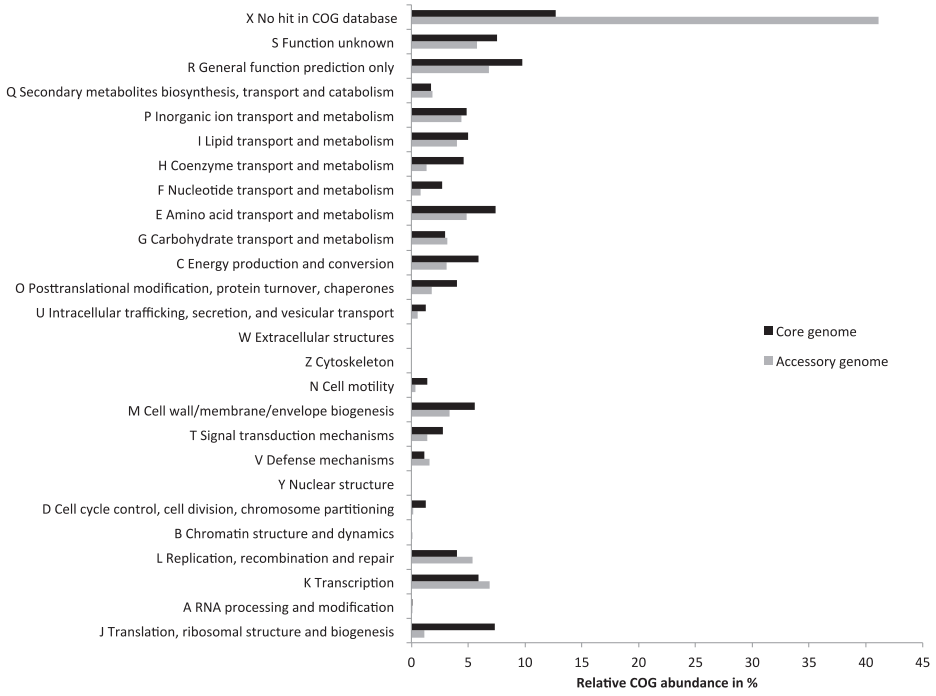

Fig. 7. Cluster of orthologous groups (COG) functional categories distribution of the genes of accessory and core genomes. Asterisks indicate statistically significant differences between accessory and core genomes.

significantly enriched (p-value < 0.05) in the unique genome of *A. venetianus* VE-C3.

Similar enrichment in genes with no assigned function has been previously reported in the accessory genome of other organisms (Bottacini et al., 2010; Galardini et al., 2011).

Given the small proportion of completely sequenced *Acinetobacter* genomes in respect to draft ones, the set of core genes found in this analysis involving only 5 strains, probably represents an overestimation of the actual *Acinetobacter* core genome. To overcome this, we also sampled a larger panel of *Acinetobacter* genomes, including all the complete and incomplete genomes available in NCBI database as on November 1st 2012 (37 genomes in total) and repeated the analysis for the identification of the core/accessory gene pool. In this case, as expected, the size of the core genome (1285 genes) is sensibly lower than in the analysis performed on the completely sequenced genomes only (1940 genes), probably representing a good approximation to the real universally shared gene pool of the *Acinetobacter* strains sequenced so far (Supplementary Material 6). Overall, we observed a general trend (Supplementary Material 6) towards the decrease of the core genome size parallel to the increase of the analyzed genomes, as recently observed by other authors (Imperi et al., 2011), when analyzing the pangenome size/dynamics of the *A. baumannii* species. Conversely, as expected, the size of the accessory genome is shown to increase as long as more strains are added to the analysis (Supplementary Material 6), indicating an open structure of the *Acinetobacter* pangenome as already shown for *A. baumannii* species (Imperi et al., 2011).

Besides providing interesting insight into bioremediation-related genes, data gained through our comparative genomics approach may also provide a basis for further analyses aimed at elucidating other important features of the overall *Acinetobacter* genus as, for example, pathogenicity (Peleg et al., 2012).

4. Conclusions

In this work we have reported the genome sequence of the strain *A. venetianus* VE-C3. In line with its environmentally strongly impacted ecological niche, the genome of this strain harbors a complete set of determinants whose functions are related to tolerance to various stresses. These include the genes probably involved in the metabolism of long-chain *n*-alkanes and in the resistance to toxic metals such as arsenic, cadmium, cobalt and zinc. We have also shown one of the possible genetic bases underlying the different strategies adopted by *A. venetianus* RAG-1^T and VE-C3 for the adhesion to oil fuel droplets. Furthermore, the presence of a number of DNA mobilization-related genes (i.e. transposases, integrases, resolvases) clearly points to a deep influence of HGT in shaping the genome of *A. venetianus* VE-C3 and in its adaptation to its special ecological niche.

Finally, the findings reported in this work provide a valuable background for future biotechnological applications as well as for deeper *in silico* analyses (e.g. metabolic network reconstruction and functional modelling) of *A. venetianus* VE-C3 metabolism.

Acknowledgments

Marco Fondi and Valerio Orlandini are financially supported by an FEMS advanced Fellowship (FAF2012) and an

Italian Cystic Fibrosis Research Foundation fellowship (grant 12#2011), respectively.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.resmic.2013.03.003.

References

Abraham, W.R., Meyer, H., Yakimov, M., 1998. Novel glycine containing glucolipids from the alkane using bacterium *Alcanivorax borkumensis*. Biochim. Biophys. Acta 1393, 57–62.

Alexeyenko, A., Tamas, I., Liu, G., Sonnhammer, E.L., 2006. Automatic clustering of orthologs and inparalogs shared by multiple proteomes. Bioinformatics 22, 9–15.

Apitz, S.E., Barbanti, A., Bocci, M., Carlin, A., Montobbio, L., Bernstein, A.G., 2007. The sediments of the Venice Lagoon (Italy) evaluated in a screening risk assessment approach: part I – application of international sediment quality guidelines. Integr. Environ. Assess. Manag. 3, 393–414.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., et al., 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9, 75.

Baldi, F., Pepi, M., Fani, R., Di Cello, F., Da Ros, L., Fossato, V.U., 1997. Complementary degradation of *n*-paraffins by aerobic Gram-negative bacteria isolated from Venice Lagoon. Croat. Chem. Acta 70, 333–346.

Baldi, F., Pepi, M., Capone, A., della Giovampaola, C., Milanesi, C., Fani, R., Focarelli, R., 2003. Envelope glycosylation determined by lectins in microscopy sections of *Acinetobacter venetianus* induced by diesel fuel. Res. Microbiol. 154, 417–424.

Baptist, J.N., Gholson, R.K., Coon, M.J., 1963. Hydrocarbon oxidation by a bacterial enzyme system. I. Products of octane oxidation. Biochim. Biophys. Acta 69, 40–47.

de la Bastide, M., McCombie, W.R., 2007. Assembling genomic DNA sequences with PHRAP. Curr. Protoc. Bioinformatics. (Chapter 11, Unit 11.14).

van Beilen, J.B., Funhoff, E.G., 2007. Alkane hydroxylases involved in microbial alkane degradation. Appl. Microbiol. Biotechnol. 74, 13–21.

van Beilen, J.B., Panke, S., Lucchini, S., Franchini, A.G., Rothlisberger, M., Witholt, B., 2001. Analysis of *Pseudomonas putida* alkane-degradation gene clusters and flanking insertion sequences: evolution and regulation of the alk genes. Microbiology 147, 1621–1630.

van Beilen, J.B., Holtackers, R., Luscher, D., Bauer, U., Witholt, B., Duetz, W.A., 2005. Biocatalytic production of perillyl alcohol from limonene by using a novel *Mycobacterium* sp. cytochrome P450 alkane hydroxylase expressed in Pseudomonas putida. Appl. Environ. Microbiol. 71, 1737–1744.

van Beilen, J.B., Funhoff, E.G., van Loon, A., Just, A., Kaysser, L., Bouza, M., Holtackers, R., Rothlisberger, M., Li, Z., Witholt, B., 2006. Cytochrome P450 alkane hydroxylases of the CYP153 family are common in alkane-degrading eubacteria lacking integral membrane alkane hydroxylases. Appl. Environ. Microbiol. 72, 59–65.

Bosi, E., Fani, R., Fondi, M., 2011. The mosaicism of plasmids revealed by atypical genes detection and analysis. BMC Genomics 12, 403.

Bottacini, F., Medini, D., Pavesi, A., Turroni, F., Foroni, E., Riley, D., Giubellini, V., Tettelin, H., van Sinderen, D., Ventura, M., 2010. Comparative genomics of the genus *Bifidobacterium*. Microbiology 156, 3243–3254.

Brilli, M., Mengoni, A., Fondi, M., Bazzicalupo, M., Lio, P., Fani, R., 2008. Analysis of plasmid genes by phylogenetic profiling and visualization of homology relationships using Blast2 Network. BMC Bioinformatics 9, 551.

Cox, M.P., Peterson, D.A., Biggs, P.J., 2010. SolexaQA: at-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinformatics 11, 485.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.

Eggink, G., van Lelyveld, P.H., Arnberg, A., Arfman, N., Witteveen, C., Witholt, B., 1987. Structure of the *Pseudomonas putida* alkBAC operon. Identification of transcription and translation products. J. Biol. Chem. 262, 6400–6406.

Feng, L., Wang, W., Cheng, J., Ren, Y., Zhao, G., Gao, C., Tang, Y., Liu, X., Han, W., Peng, X., Liu, R., Wang, L., 2007. Genome and proteome of long-chain alkane degrading *Geobacillus thermodenitrificans* NG80-2 isolated from a deep-subsurface oil reservoir. Proc. Natl. Acad. Sci. U S A 104, 5602–5607.

Fondi, M., Bacci, G., Brilli, M., Papaleo, M.C., Mengoni, A., Vaneechoutte, M., Dijkshoorn, L., Fani, R., 2010. Exploring the evolutionary dynamics of plasmids: the *Acinetobacter* pan-plasmidome. BMC Evol. Biol. 10, 59.

Fondi, M., Orlandini, V., Emiliani, G., Papaleo, M.C., Maida, I., Perrin, E., Vaneechoutte, M., Dijkshoorn, L., Fani, R., 2012. Draft genome sequence of the hydrocarbon-degrading and emulsan-producing strain *Acinetobacter venetianus* RAG-1^{T}. J. Bacteriol. 194, 4771–4772.

Foster, J.W., 1962. Bacterial oxidation of hydrocarbons. In: Foster, J.W. (Ed.), Oxygenases. Academic, New York, pp. 241–261.

Galardini, M., Mengoni, A., Brilli, M., Pini, F., Fioravanti, A., Lucas, S., Lapidus, A., Cheng, J.F., et al., 2011. Exploring the symbiotic pangenome of the nitrogen-fixing bacterium *Sinorhizobium meliloti*. BMC Genomics 12, 235.

Gao, F., Zhang, C.T., 2008. Ori-Finder: a web-based system for finding *oriC*s in unannotated bacterial genomes. BMC Bioinformatics 9, 79.

Giovannetti, L., Ventura, S., Bazzicalupo, M., Fani, R., Materassi, R., 1990. DNA restriction fingerprint analysis of the soil bacterium *Azospirillum*. J. Gen. Microbiol. 136, 1161–1166.

Guindon, S., Lethiec, F., Duroux, P., Gascuel, O., 2005. PHYML online – a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res. 33, W557–W559.

Guindon, S., Delsuc, F., Dufayard, J.F., Gascuel, O., 2009. Estimating maximum likelihood phylogenies with PhyML. Meth. Mol. Biol. 537, 113–137.

Head, I.M., Swannell, R.P., 1999. Bioremediation of petroleum hydrocarbon contaminants in marine habitats. Curr. Opin. Biotechnol. 10, 234–239.

Hommel, R.K., 1990. Formation and physiological role of biosurfactants produced by hydrocarbon-utilizing microorganisms. Biosurfactants in hydrocarbon utilization. Biodegradation 1, 107–119.

Imperi, F., Antunes, L.C., Blom, J., Villa, L., Iacono, M., Visca, P., Carattoli, A., 2011. The genomics of *Acinetobacter baumannii*: insights into genome plasticity, antimicrobial resistance and pathogenicity. IUBMB Life 63, 1068–1074.

Kang, Y.S., Jung, J., Jeon, C.O., Park, W., 2011. *Acinetobacter oleivorans* sp. nov. is capable of adhering to and growing on diesel–oil. J. Microbiol. 49, 29–34.

Kok, M., Oldenhuis, R., van der Linden, M.P., Raatjes, P., Kingma, J., van Lelyveld, P.H., Witholt, B., 1989. The *Pseudomonas oleovorans* alkane hydroxylase gene. Sequence and expression. J. Biol. Chem. 264, 5435–5441.

Kostka, J.E., Prakash, O., Overholt, W.A., Green, S.J., Freyer, G., Canion, A., Delgardio, J., Norton, N., Hazen, T.C., Huettel, M., 2011. Hydrocarbon-degrading bacteria and the bacterial community response in Gulf of Mexico beach sands impacted by the deepwater horizon oil spill. Appl. Environ. Microbiol. 77, 7962–7974.

Maier, T., Forster, H.H., Asperger, O., Hahn, U., 2001. Molecular characterization of the 56-kDa CYP153 from *Acinetobacter* sp. EB104. Biochem. Biophys. Res. Commun. 286, 652–658.

Mara, K., Decorosi, F., Viti, C., Giovannetti, L., Papaleo, M.C., Maida, I., Perrin, E., Fondi, M., Vaneechoutte, M., Nemec, A., van den Barselaar, M., Dijkshoorn, L., Fani, R., 2012. Molecular and phenotypic characterization of *Acinetobacter* strains able to degrade diesel fuel. Res. Microbiol. 163, 161–172.

Mengoni, A., Richi, S., Brilli, M., Baldi, F., Fani, R., 2007. Sequencing and analysis of plasmids pAV1 and pAV2 of *Acinetobacter venetianus* VE-C3 involved in diesel fuel degradation. Ann. Microbiol. 57, 521–526.

Mercaldi, M.P., Dams-Kozlowska, H., Panilaitis, B., Joyce, A.P., Kaplan, D.L., 2008. Discovery of the dual polysaccharide composition of emulsan and the isolation of the emulsion stabilizing component. Biomacromolecules 9, 1988–1996.

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., Kanehisa, M., 2007. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35, W182–W185.

Nakar, D., Gutnick, D.L., 2001. Analysis of the wee gene cluster responsible for the biosynthesis of the polymeric bioemulsifier from the oil-degrading strain *Acinetobacter lwoffii* RAG-1. Microbiology 147, 1937–1946.

Panke, S., Meyer, A., Huber, C.M., Witholt, B., Wubbolts, M.G., 1999. An alkane-responsive expression system for the production of fine chemicals. Appl. Environ. Microbiol. 65, 2324–2332.

Peleg, A.Y., de Breij, A., Adams, M.D., Cerqueira, G.M., Mocali, S., Galardini, M., Nibbering, P.H., Earl, A.M., Ward, D.V., Paterson, D.L., Seifert, H., Dijkshoorn, L., 2012. The success of Acinetobacter species; genetic, metabolic and virulence attributes. PLoS One 7 (10), e46984.

Pines, O., Bayer, E.A., Gutnick, D.L., 1983. Localization of emulsan-like polymers associated with the cell surface of *Acinetobacter calcoaceticus*. J. Bacteriol. 154, 893–905.

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., Lopez, R., 2005. InterProScan: protein domains identifier. Nucleic Acids Res. 33, W116–W120.

Reisfeld, A., Rosenberg, E., Gutnick, D., 1972. Microbial degradation of crude oil: factors affecting the dispersion in sea water by mixed and pure cultures. Appl. Microbiol. 24, 363–368.

Remm, M., Storm, C.E., Sonnhammer, E.L., 2001. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. J. Mol. Biol. 314, 1041–1052.

Rojo, F., 2009. Degradation of alkanes by bacteria. Environ. Microbiol. 11, 2477–2490.

Ron, E.Z., Rosenberg, E., 2002. Biosurfactants and oil bioremediation. Curr. Opin. Biotechnol. 13, 249–252.

Schelstraete, P., Deschaght, P., Van Simaey, L., Van Daele, S., Haerynck, F., Vaneechoutte, M., De Baets, F., 2010. Genotype based evaluation of *Pseudomonas aeruginosa* eradication treatment success in cystic fibrosis patients. J. Cyst. Fibros 9, 99–103.

Schneiker, S., Martins dos Santos, V.A., Bartels, D., Bekel, T., Brecht, M., Buhrmester, J., Chernikova, T.N., Denaro, R., et al., 2006. Genome sequence of the ubiquitous hydrocarbon-degrading marine bacterium *Alcanivorax borkumensis*. Nat. Biotechnol. 24, 997–1004.

Sekine, M., Tanikawa, S., Omata, S., Saito, M., Fujisawa, T., Tsukatani, N., Tajima, T., Sekigawa, T., et al., 2006. Sequence analysis of three plasmids harboured in *Rhodococcus erythropolis* strain PR4. Environ. Microbiol. 8, 334–346.

Silver, S., Phung le, T., 2005. A bacterial view of the periodic table: genes and proteins for toxic inorganic ions. J. Ind. Microbiol. Biotechnol. 32, 587–605.

Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J., Birol, I., 2009. ABySS: a parallel assembler for short read sequence data. Genome Res. 19, 1117–1123.

Smits, T.H., Balada, S.B., Witholt, B., van Beilen, J.B., 2002. Functional analysis of alkane hydroxylases from gram-negative and gram-positive bacteria. J. Bacteriol. 184, 1733–1742.

Tamminen, M., Virta, M., Fani, R., Fondi, M., 2012. Large-scale analysis of plasmid relationships through gene-sharing networks. Mol. Biol. Evol. 29, 1225–1240.

Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., et al., 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics 4, 41.

Throne-Holst, M., Markussen, S., Winnberg, A., Ellingsen, T.E., Kotlar, H.K., Zotchev, S.B., 2006. Utilization of *n*-alkanes by a newly isolated strain of *Acinetobacter venetianus*: the role of two AlkB-type alkane hydroxylases. Appl. Microbiol. Biotechnol. 72, 353–360.

Throne-Holst, M., Wentzel, A., Ellingsen, T.E., Kotlar, H.K., Zotchev, S.B., 2007. Identification of novel genes involved in long-chain *n*-alkane degradation by Acinetobacter sp. strain DSM 17874. Appl. Environ. Microbiol. 73, 3327–3332.

Vallenet, D., Nordmann, P., Barbe, V., Poirel, L., Mangenot, S., Bataille, E., Dossat, C., Gas, S., et al., 2008. Comparative analysis of Acinetobacters: three genomes for three lifestyles. PLoS One 3, e1805.

Vaneechoutte, M., Tjernberg, I., Baldi, F., Pepi, M., Fani, R., Sullivan, E.R., van der Toorn, J., Dijkshoorn, L., 1999. Oil-degrading *Acinetobacter* strain RAG-1 and strains described as '*Acinetobacter venetianus* sp. nov.' belong to the same genomic species. Res. Microbiol. 150, 69–73.

Vernikos, G.S., Parkhill, J., 2006. Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the *Salmonella* pathogenicity islands. Bioinformatics 22, 2196–2203.

Wentzel, A., Ellingsen, T.E., Kotlar, H.K., Zotchev, S.B., Throne-Holst, M., 2007. Bacterial metabolism of long-chain *n*-alkanes. Appl. Microbiol. Biotechnol. 76, 1209–1221.

Zhan, Y., Yan, Y., Zhang, W., Chen, M., Lu, W., Ping, S., Lin, M., 2012. Comparative analysis of the complete genome of an *Acinetobacter calcoaceticus* strain adapted to a phenol-polluted environment. Res. Microbiol. 163, 36–43.

Zuckerberg, A., Diver, A., Peeri, Z., Gutnick, D.L., Rosenberg, E., 1979. Emulsifier of *Arthrobacter* RAG-1: chemical and physical properties. Appl. Environ. Microbiol. 37, 414–420.