UNIVERSITY OF NAPOLI FEDERICO II

Doctorate School in Molecular Medicine

Doctorate Program in
Genetics and Molecular Medicine
Coordinator: Prof. Lucio Nitsch
XXIV Cycle

"Core and variable components in prokaryotic genomes"

FRANCESCO ROCCO

Napoli 2011

TABLE OF CONTENTS

PUBLICATIONS

***Stenotrophomonas maltophilia* genomes: a start-up comparison.**

Rocco F, De Gregorio E, Colonna B, Di Nocera PP.

Int J Med Microbiol. **2009** 299:535-46.

A giant family of short palindromic sequences in *Stenotrophomonas maltophilia*.

Rocco F, De Gregorio E, Di Nocera PP.

FEMS Microbiol Lett. **2010** 308:185-92.

Genome sequences of three *Acinetobacter baumannii* strains assigned to the multilocus sequence typing genotypes ST2, ST25, and ST78.

Zarrilli R, Giannouli M, Rocco F, Loman NJ, Haines AS, Constantinidou C, Pallen MJ, Triassi M, Di Nocera PP.

J Bacteriol. **2011**, 193:2359-60.

Genome organization of epidemic *Acinetobacter baumannii* strains.

Di Nocera PP, Rocco F, Giannouli M, Triassi M, Zarrilli R.

BMC Microbiol. **2011** 11:224.

ABSTRACT

Acinetobacter baumannii and *Stenotrophomonas maltophilia* are opportunistic pathogens responsable for several hospital-acquired infections. We performed Extensive comparative analyses were performed in silico to elucidate the cromosomal organization of both species, and identify core and variable genome component in each. In the case of *S. maltophilia*, the genomes of strains isolated from the blood of a cancer patient (K279a strain) and the poplar *Populus trichocarpa* (R551-3 strain) have been compared. The analyzed chromosomes exhibit extensive synteny and 3620 homologous genes are located at the same relative position in the two strains. However, the R551-3 and the K279a chromosomes vary significantly, as they contain 13% and 19% of specific DNAs sequences, respectively. This DNA fraction is largely represented by unrelated foreign DNA segments or genomic islands (GEIs), which vary in size from 3 to 70 kb. Similar work was done by comparing of the genomes of seven clinical isolates of *A. baumannii*. The analyzed DNAs also showed extensive synteny. We have identified 3068 conserved coding regions in all isolates, and 63 variable genomic loci containing GEIs ranging in size from 4 o 126 kb. GEIs found in *A. baumannii* and *S. maltophilia* GEIs are, except for some resistance islands, unrelated.

BACKGROUND

1.1. Dynamic organization of microbial genomes

The first bacterial genome to be completed was that of *Haemophilus influenzae*, sequenced by a team at The Institute for Genomic Research in 1995. In a few years, the complete sequence of genomes of different pathogenic bacteria were subsequently reported. At the moment, the number of wholly sequenced genomes approaches 2000, and the advent of ultra-high throughput next generation sequencing technologies tremendously increased the number of bacterial genomes sequenced. Large-scale sequencing made available whole genome sequences of many strains of the same bacterial species.

Comparative analyses carried out with multiple genomes of one species have revealed extensive, unexpected intra-species diversity of *Streptococcus agalactiae* (Tettelin *et al.*,2005) and *Haemophilus influenzae* (Hogg *et al.*, 2007) genomes allowed to separate genes of both species into two main categories. Some genes are shared by all the strains of the species analyzed, and constitute the so called core genome. Other genes are present in some isolates, but not in anothers, and make up the dispensable or variable genome component. The sum of core and dispensable genes makes up in turn the so called pan-genome, which represents the full species gene repertoire. The size of pangenome and core genomes is closely correlated. Thus for example, the analysis of 61 *E. coli* genomes brought to a pangenome made up by approximatively 16000 genes, and a core genome restricted to less than 1000 genes (Lukjancenko *et al.*, 2010; Fig. 1).

The dispensable genome contributes to the species diversity and provides functions that are not essential to its basic life-style, but confer selective

advantages such as niche adaptation, antibiotic resistance, the ability to colonize new hosts (Dobrindt *et al.*, 2004).

Many dispensable genes have been acquired by horizontal gene transfer (HGT), a process in which DNA segments which may vary in size from 1 to 100 kb, or more, can be transferred from one species to another, and be incorporated in the recipient genome.

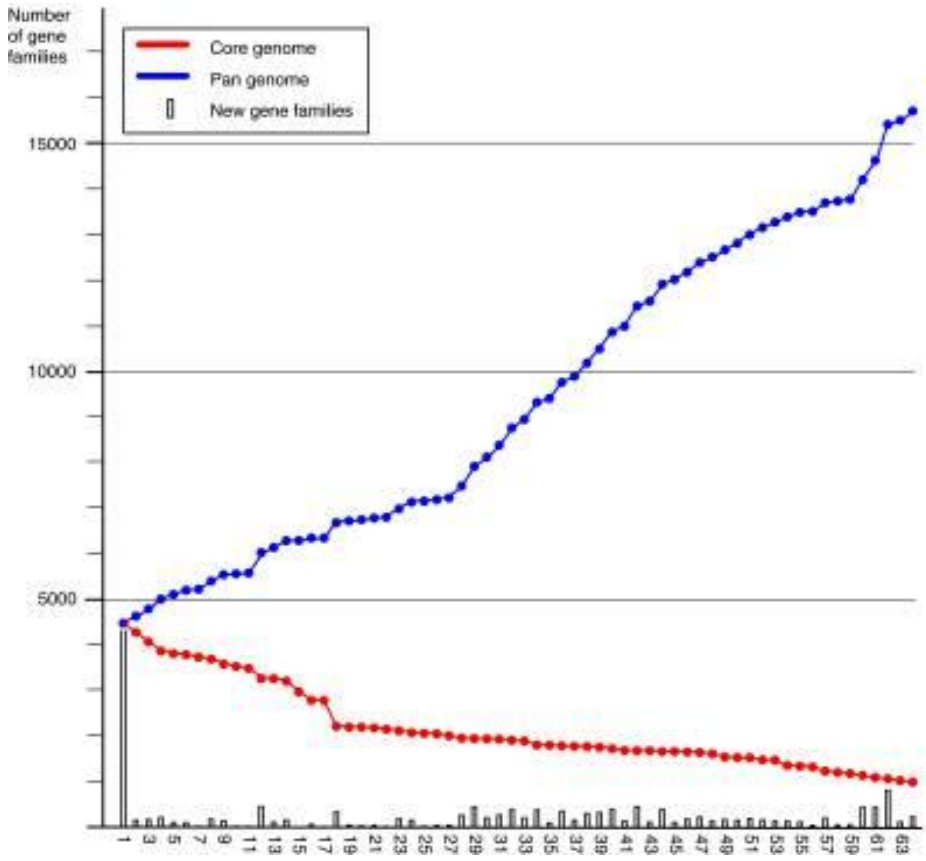

Fig. 1: Pan- and core genome plot of the genomes of *E. coli*. The blue pan-genome curve connects the cumulative number of gene families present in the analyzed genomes. The red core genome curve connects the conserved number of gene families. The gray bars show the numbers of novel gene families identified in each genome.

Mobile genetic entities, known as genomic (GEIs) or pathogenic (PAIs) islands are typical dispensable genome components acquired via HGT. The size and coding capacity of GEIs or PAIs may eventually be remodeled by a variety of mutations and rearrangements (Fig. 2). Most GEIs are relatively large segments of DNA and may be recognized by nucleotide composition (e.g. GC content) that usually differ from the rest of the chromosome. Genomic islands are often inserted at tRNA genes and are flanked by 16–30 bp direct repeats (DR), that arise by the duplication of the integration site. DRs may act as recognition sequences for GEIs enzymatic excision (Juhas *et al.* 2009). Genes encoding integrases, enzymes involved in GEI transfer, are often found in genomic islands (Fig. 3).

Depending on the functions they encode and the advantages they may confer to the host genome, GEIs can be distinguished in pathogenicity, symbiosis, fitness, metabolic or resistance islands. GEIs might provide a selective advantage under specific growth conditions as they can enhance adaptability and competitiveness within a niche.

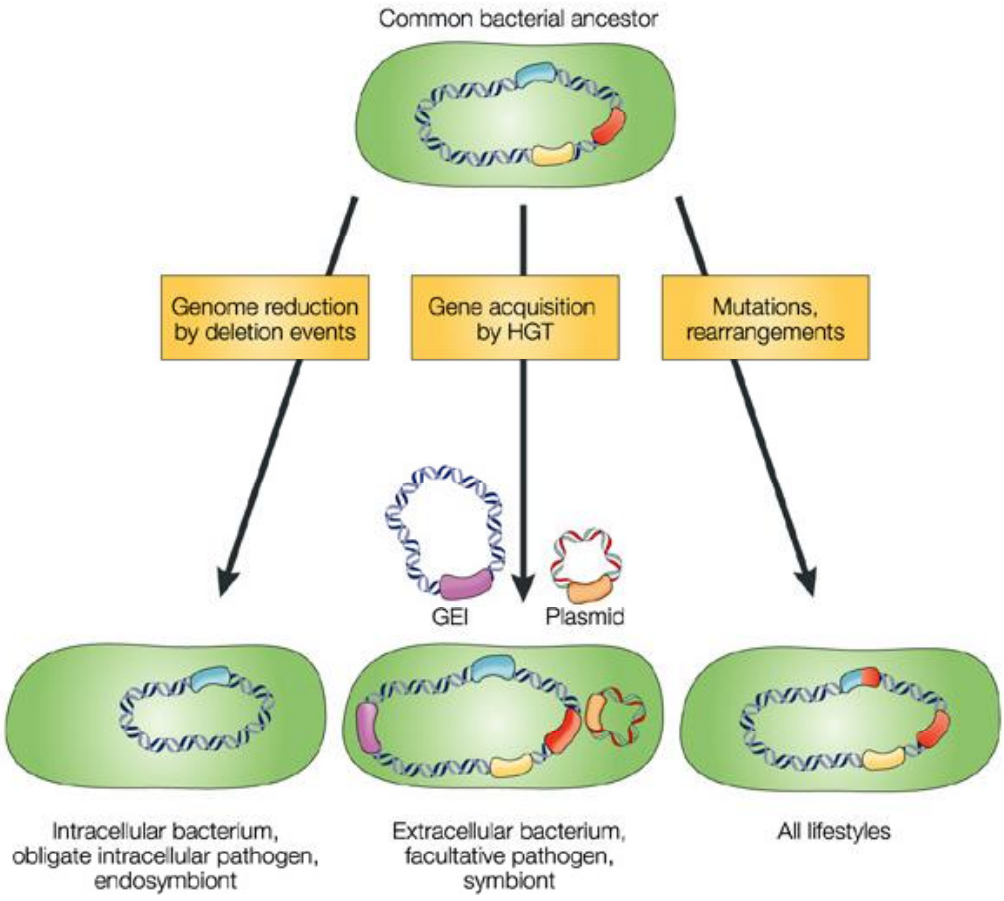

Fig. 2: **Evolution of bacterial variants by acquisition and loss of genitic information.** Gene acquisition by horizontal transfer between different species, which involves mobile genetic elements, such as plasmids, genomic islands (GEIs), increases the versatility and adaptability of the recipient. This is common in extracellular bacteria, such as facultative pathogens and symbionts, and the acquisition of genes in this way allows bacteria to adapt to a new or changing environment. In addition to these processes, point mutations and genetic rearrangements constantly contribute to evolution of new gene variants in all types of bacteria. HGT, horizontal gene transfer.

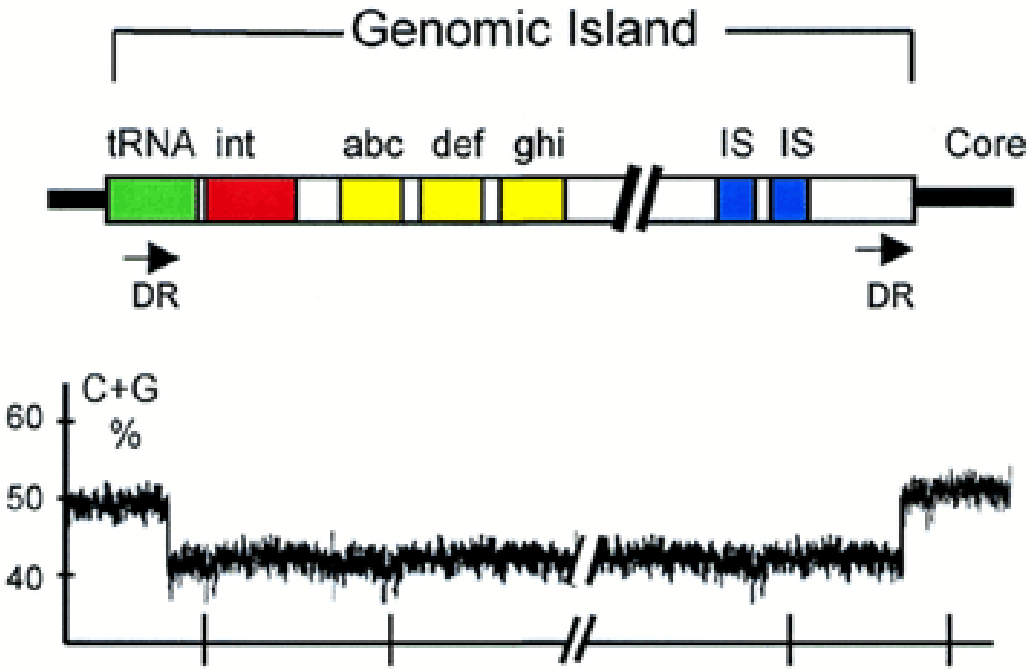

Fig. 3: General features of GEIs: Schematic model of a genomic island of bacteria (upper part). The formerly transferred DNA block is linked to a tRNA gene and flanked by direct repeats (DR). The guanine plus cytosine (G+C) content of the genomic island is different from that of the core genome (lower part). Other abbreviations: int, integrase gene; *abc*, *def* and *ghi*, genes encoding specific functions; IS, insertion sequence element; bp, base pair.

1.2. *Stenotrophomonas maltophilia*

Stenotrophomonas maltophilia is an aerobic, non-fermentative Gram-negative bacterium widespread in the environment. This species constitutes one of the dominant rhizosphere inhabitants. *S. maltophilia* is able to degrade xenobiotic compounds, as detoxify high molecular weight polycyclic aromatic hydrocarbons, and is therefore considered with interest for its potential use in bioremediation (Page *et al.*, 2008). *S. maltophilia* is also increasingly described as an important nosocomial pathogen in debilitated and immunodeficient patients and has been associated with a broad spectrum of clinical syndromes, e.g. bacteraemia, endocarditis, respiratory tract infections (Looney, 2005). Over the last decade, *S. maltophilia* has been frequently isolated from cystic fibrosis patients (Denton and Kerr, 1998), and turned out to be a serious pathogen in cancer patients (Safdar and Rolston, 2007). *S. maltophilia* displays intrinsic resistance to many antibiotics, making selection of optimal therapy difficult. Several factors confer *S. maltophilia* a role as emerging pathogen, most notably the ability to elaborate a wide range of extracellular enzymes, such as lipases, fibrolysin, and proteases, potentially involved in the colonization process (Denton and Kerr, 1998), the ability to adhere to and form biofilm on epithelial cells (Di Bonaventura *et al.*, 2007), and the ability to stimulate factors involved in the inflammatory process (Waters *et al.*, 2007).

1.3. *S. maltophilia* genomes

The complete nucleotide sequence of the genome of K279a, a *S. maltophilia* strain isolated from the blood of a cancer patient, has been determined (Crossman *et al.*, 2008). The whole nucleotide sequence of R551-3, a *S. maltophilia* strain isolated from the poplar *Populus trichocarpa*, has been completed at the DOE Joint Genome Institute (Lucas *et al.*, 2008). Evaluating

peculiarities of the genetic organization of *Stenotrophomonas* chromosomes is an essential step to shed light on unexplored aspects of *S. maltophilia*, with special emphasis on the possible correlation between pathogenic traits and specific genes or gene clusters. In this thesis, results emerging from various comparisons of the K279a and R551-3 DNA are reported.

1.4. *Acinetobacter baumannii*

The genus *Acinetobacter* comprises 26 species with valid names and nine genomic species with provisional designations that were defined by DNA-DNA hybridization. *Acinetobacter baumannii*, is the species that is more frequently isolated in hospitalized patients, especially in intensive-care-unit (ICU) wards. The capability to survive in dry conditions and resistance to disinfectants and antimicrobial agents contribute to the selection of *A. baumannii* in the hospital setting (Dijkshoorn *et al.*, 2007).

Epidemics caused by multidrug-resistant (MDR) strains of *A. baumannii* were reported in several hospitals worldwide and shown to be caused by *A. baumannii* strains resistant to all classes of antimicrobials including carbapenems (Durante-Mangoni and Zarrilli, 2011). Outbreaks were caused by clusters of highly similar *A. baumannii* strains that were assigned by several genotypic methods to three main international clonal lineages referred to as international clones I, II and III, respectively (Diancourt *et al.*, 2010). The predominance of international clone II lineage world-wide and the occurrence of hospital outbreaks caused by MDR strains belonging to novel genotypes not related to the three main clonal complexes have been reported during the last few years (Diancourt *et al.*, 2010).

1.5. *A. baumannnii* genomes

The draft genome sequences of three *A. baumannii* strains, 3990, 4190 and 3909, respectively assigned to ST (sequence types) 2, 25 and 78, which are representative of the most frequent genotypes responsible for epidemics occurred in Mediterranean hospitals have been recently reported. The genomes of the 3990, 4190 and 3909 strains have been compared to the genomes of four wholly sequenced MDR *A. baumannii* strains, two assigned to ST1 (AB0057 and AYE strains) one each to ST2 (ACICU strain) and ST77 (ATCC17978 strain).

AIMS OF THE STUDY

Taking into account the degree of genetic variation exhibited by isolates of different pathogens, decypher the complexity of bacterial species, comparative analyses carried out on multiple genomes are needed to distinguish the core and variable genome components, and define at the molecular level the basic organization of a given species.

Considering the interest paid to *S. maltophilia* and *A. baumannii* as pathogens associated with nosocomial infections, genomic comparative analyses have been performed. The overall scaffold of the *S. maltophilia* chromosome has been defined by whole comparison of the coding regions of a clinical and an environmental strain of *S. maltophlia* isolates. More accurate work done by the comparison of the gene products encoded by the genomes of seven clinical MDR isolates of *A. baumannii* helped to elucidate the basic organization of the *A. baumannii* chromosome. The two lines of research brought to the identification of hundred of mobile genetic regions or GEIs.

MATERIALS AND METHODS

3.1. In silico analyses

a) *S. maltophilia* genomes. The whole genomes of the *S. maltophilia* strains K279a (EMBL/GeneBank access no. AM743169) and R551-3 (NCBI, locus CP001111) were aligned by using mVISTA, a set of programs for comparing DNA sequences from 2 or more species and visualizing the obtained alignments (Frazer *et al.*, 2004; http://genome.lbl.gov/vista/index.shtml). EMBL/GeneBank and NCBI gene annotations were integrated with annotations available at the KEGG resource (Kyoto Encyclopedia of Genes and Genomes, http://www.genome.jp/kegg).

b) *A. baumannii* genomes. Comparative genome analysis were performed on whole genome sequences of *A. baumannii* strains AB0057 (GenBank:NC_011586) (Adams *et al.*,2008), ACICU (GenBank:NC_010611) (Iacono *et al.*, 2008), ATCC17978 (GenBank:NC_009085) (Smith *et al.*, 2007) and AYE (GenBank:NC_010410) (Vallenet *et al.*, 2008) and draft genome sequences of *A. baumannii* strains ST2 3990 (GenBank:AEOY00000000), ST25 4190 (GenBank:AEPA00000000) and ST78 3909 (GenBank:AEOZ00000000) strains (Zarrilli *et al.*, 2011). The GenBank:CP000521 file, which contains 436 hypothetical proteins putatively encoded by ATCC17978 early annotated as AS1, but not included in the GenBank:NC_009085 file, was also used for comparisons. Gene products putatively encoded by the ST25 4190, ST78 3909 and ST2 3990 strains were identified using xBASE2, comparing the draft genome sequences to the genome of the *A. baumannii* strain AB0057 used as reference template (Zarrilli *et al.*, 2011). Predicted ORFs were subsequently compared to the gene products of the wholly sequenced *A. baumannii* AB0057, ACICU, ATCC and ABAYE strains using MAUVE (Darling *et al.*, 2010). Homologies under

looked by MAUVE were detected by BLAST and tBLASTn analyses. Gene products encoded by aligned coding regions exhibited at least 50% identity.

3.2. *S. maltophilia* strains and DNA analyses

The *S. maltophilia* clinical strains used in this study were obtained from the Bambino Gesù Hospital (Rome, Italy) and from the University Federico II Hospital (Naples, Italy) and have been described previously (Di Bonaventura *et al.*, 2007). The K279a strain (Crossman *et al.*, 2008) was a kind gift of M.B. Avison (University of Bristol, UK), while the LMG strains were purchased at the LMG/BMCC collection. The *S. maltophilia* strains analyzed in this study are listed in Table 1. All strains were routinely grown in brain heart infusion at 37 °C, except for the environmental strains (LMG959, LMG10871, LMG10879, LMG11104, and LMG11108) which were grown at 30 °C. Genomic DNA was extracted by routine procedure.

PCR amplifications. PCR reactions were carried out by incubating 20 ng of DNA with 160 ng of each primer in the presence of dXTPs (200 nM), 1.5 mM MgCl2, and the Taq DNA polymerase Recombinant (Invitrogen). Samples were incubated first at 95 °C for 5 min. The amplification programme included 1 min at 95 °C, 1 min at the annealing temperature, and 1 min at 72 °C for a total of 30 cycles. At the end of the cycle, samples were kept at 72 °C for 7 min before harvesting. PCR products were electrophoresed on 1.5–2% agarose gels in 0.5X TBE buffer (45 mM Tris pH 8, 45 mM borate, 0.5 mM EDTA) at 120 V (constant voltage). The 100-bp ladder (Fermentas) was used as molecular weight marker. Primers measured all 25 nt, and annealing temperatures varied from 57 to 64 °C.

Slot-blot hybridizations. One microgram of DNA from each strain was loaded onto Hybond filters and cross-linked by UV treatment. The filters were

hybridized to 32P-radiolabelled PCR products amplified from orfs belonging to specific GEIs.

Source, origin, and reference of the *S. maltophilia* strains analyzed

strain name	source	location	reference
92	bronchial aspirate (ICU)	UFH, Naples	Roscetto et al. 2008
262	bronchial aspirate (CF)	UFH, Naples	Roscetto et al. 2008
527	pharyngeal swab (H)	UFH, Naples	Roscetto et al. 2008
528	bronchial aspirate (ICU)	UFH, Naples	Roscetto et al. 2008
545	bronchial aspirate (CF)	UFH, Naples	Roscetto et al. 2008
549	bronchial aspirate (ICU)	UFH, Naples	Roscetto et al. 2008
571	bronchial aspirate (ICU)	UFH, Naples	Roscetto et al. 2008
598	pharyngeal swab (H)	UFH, Naples	Roscetto et al. 2008
616	pharyngeal swab (H)	UFH, Naples	Roscetto et al. 2008
707	bronchial aspirate (CF)	UFH, Naples	Roscetto et al. 2008
714	pharyngeal swab (H)	UFH, Naples	Roscetto et al. 2008
915	bronchial aspirate (ICU)	UFH, Naples	Roscetto et al. 2008
916	bronchial aspirate (ICU)	UFH, Naples	Roscetto et al. 2008
1019	bronchial aspirate (ICU)	UFH, Naples	Roscetto et al. 2008
1029	bronchial aspirate (ICU)	UFH, Naples	Roscetto et al. 2008
1039	bronchial aspirate (ICU)	UFH, Naples	Roscetto et al. 2008
1053	urine (ICU)	UFH, Naples	Roscetto et al. 2008
1054	bronchial aspirate (ICU)	UFH, Naples	Roscetto et al. 2008
OBGTC3	pharyngeal swab (CF)	BGH, Rome	Roscetto et al. 2008
OBGTC9	bronchial aspirate (CF)	BGH, Rome	Di Bonaventura et al. 2007
OBGTC13	bronchial aspirate (CF)	BGH, Rome	Roscetto et al. 2008
OBGTC10	bronchial aspirate (CF)	BGH, Rome	Di Bonaventura et al. 2007
OBGTC16	pharyngeal swab (CF)	BGH, Rome	Di Bonaventura et al. 2007
OBGTC20	bronchial aspirate (CF)	BGH, Rome	Di Bonaventura et al. 2007
OBGTC22	bronchial aspirate (CF)	BGH, Rome	Roscetto et al. 2008
OBGTC26	pharyngeal swab (CF)	BGH, Rome	Di Bonaventura et al. 2007
OBGTC28	bronchial aspirate (CF)	BGH, Rome	Di Bonaventura et al. 2007
OBGTC29	bronchial aspirate (CF)	BGH, Rome	Di Bonaventura et al. 2007
OBGTC30	bronchial aspirate (CF)	BGH, Rome	Roscetto et al. 2008
STM2	emocolture (H)	BGH, Rome	Roscetto et al. 2008
K279a	emocolture (C)	BOU, Bristol	Crossman et al. 2008
LMG959	rice paddy	Japan	Hauben et al. 1999
LMG10879	rice paddy		Hauben et al. 1999
LMG10871	soil		Hauben et al. 1999
LMG11104	Cichorium intybus,	France	Hauben et al. 1999
LMG11108	Triticum, roots	France	Hauben et al. 1999
LMG10851	human, blood culture	Belgium	Hauben et al. 1999
LMG10853	sputum	Belgium	Hauben et al. 1999
LMG10873	Case of conjunctivitis		Hauben et al. 1999
LMG10874	Human, blood culture		Hauben et al. 1999
LMG10889			Hauben et al. 1999
LMG10991	leg, pus	Belgium	Hauben et al. 1999

Table 1: CF, cystic fibrosis; ICU, intensive care unit; H, haematology; C, cancer; UHF, University Federico II hospital; BGH, Bambino Gesù Hospital; BOU, Bristol oncology unit; LMG, Laboratorium voor Microbiologie Gent culture collection, Belgium.

RESULTS AND DISCUSSION

4.1. Core and variable genome components in *S. maltophilia*

The chromosomes of *S. maltophilia* K279a and R551-3 strains have the same GC content (67%), but differ significantly in length, since K279a DNA measures 4,851,126 bp, and R551-3 DNA 4,573,969 bp. However, the 2 chromosomes share a common scaffold, as they are widely collinear, and can be easily aligned throughout their lengths (Fig. 4). Differences between the genomes of the 2 strains are correlated primarily to specific DNA sequences present exclusively in K279a. Accordingly, the number of potential gene products is higher in K279a (4386 annotated orfs) than in R551-3 (4041 annotated orfs). Nonetheless, several DNA segments are unique to the R551-3 genome. Sequence alignments allowed to build up a comparative map of *S. maltophilia* orfs, and to easily identify strain-specific gene products. Orfs annotated as hypothetical proteins shorter than 100 aminoacids were not considered for comparisons.

Alignment of K279a and R551-3 genomes. VISTA plots showing similarity of K279a vs. R551-3 (A) and R551-3 vs. K279a (B) DNAs. The percent identity of the aligned genomes is shown.

Sequence alignments allowed to identify about 200 regions in which shared chromosome synteny is interrupted in one strain, or in both. A major source of variation between the 2 *S. maltophilia* chromosomes is represented by a multitude of strain-specific genomic islands (GEIs), ranging in size from approximately 3 to 70 kb. Routinely, islands denote chromosomal DNA regions, plausibly acquired by horizontal gene transfer (HGT), larger than 10 kb, smaller regions being referred to as islets (Dobrindt *et al.*, 2004). For sake of simplicity, as already done in other genome comparisons (see Myers *et al.*, 2006), all DNA segments >3 kb present in only one of the 2 strains were

referred to as islands. *S. maltophilia* GEIs were marked by a K or R to designate K279a or R551-3 DNA, respectively, and numbered progressively according to chromosome position. Fourty-one GEIs have been identified in K279a, and 36 in R551-3 DNA. Size, orfs, GC content, and chromosomal location of all GEIs are shown in Fig. 5. In both strains, half of GEIs measure less than 5 kb. Of the remaining islands, most measure more than 15 kb in K279a, but only 4 exceed such size in R551-3. On the whole, K and R GEIs constitute 12.1% and 6.6% of the genome and encode 597 and 249 orfs, respectively. GEIs often have a GC content different from bulk chromosomal value (Dobrindt *et al.*, 2004). The GC content of many GEIs is lower than the average 66.7% value of *S. maltophilia* DNA. Values range between 63 and 66% in 16 K-GEIs and 15 R-GEIs and are lower than 63% in 16 K-GEIs and 7 R-GEIs. Nine islands were inserted at the same relative chromosomal position in the 2 strains (Fig. 5). At least in one case, this is correlated to the use of the same chromosomal entry site.

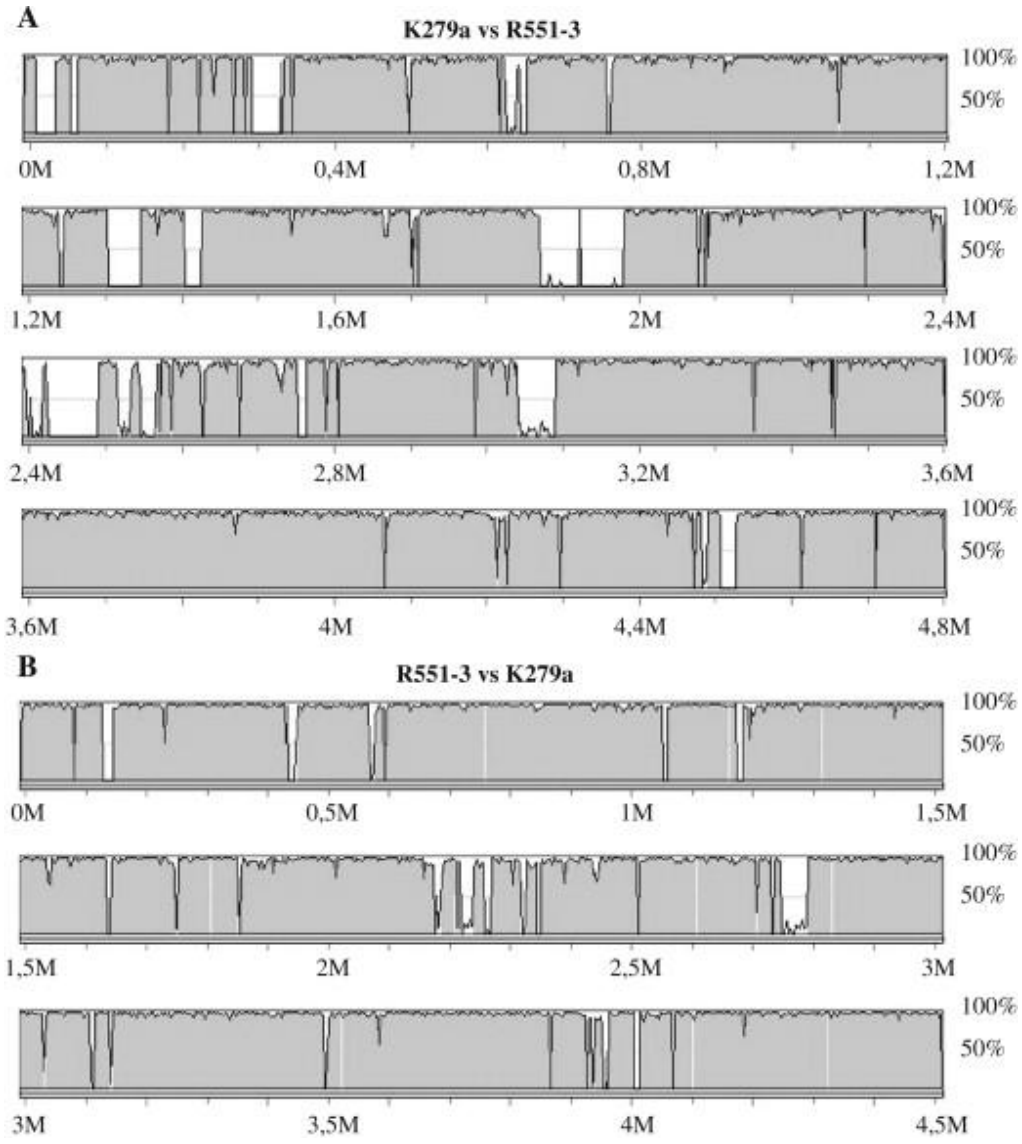

Fig 4: Alignment of K279a and R551-3 genomes. VISTA plots showing similarity of K279a vs. R551-3 (A) and R551-3 vs. K279a (B) DNAs. The percent identity of the aligned genomes is shown.

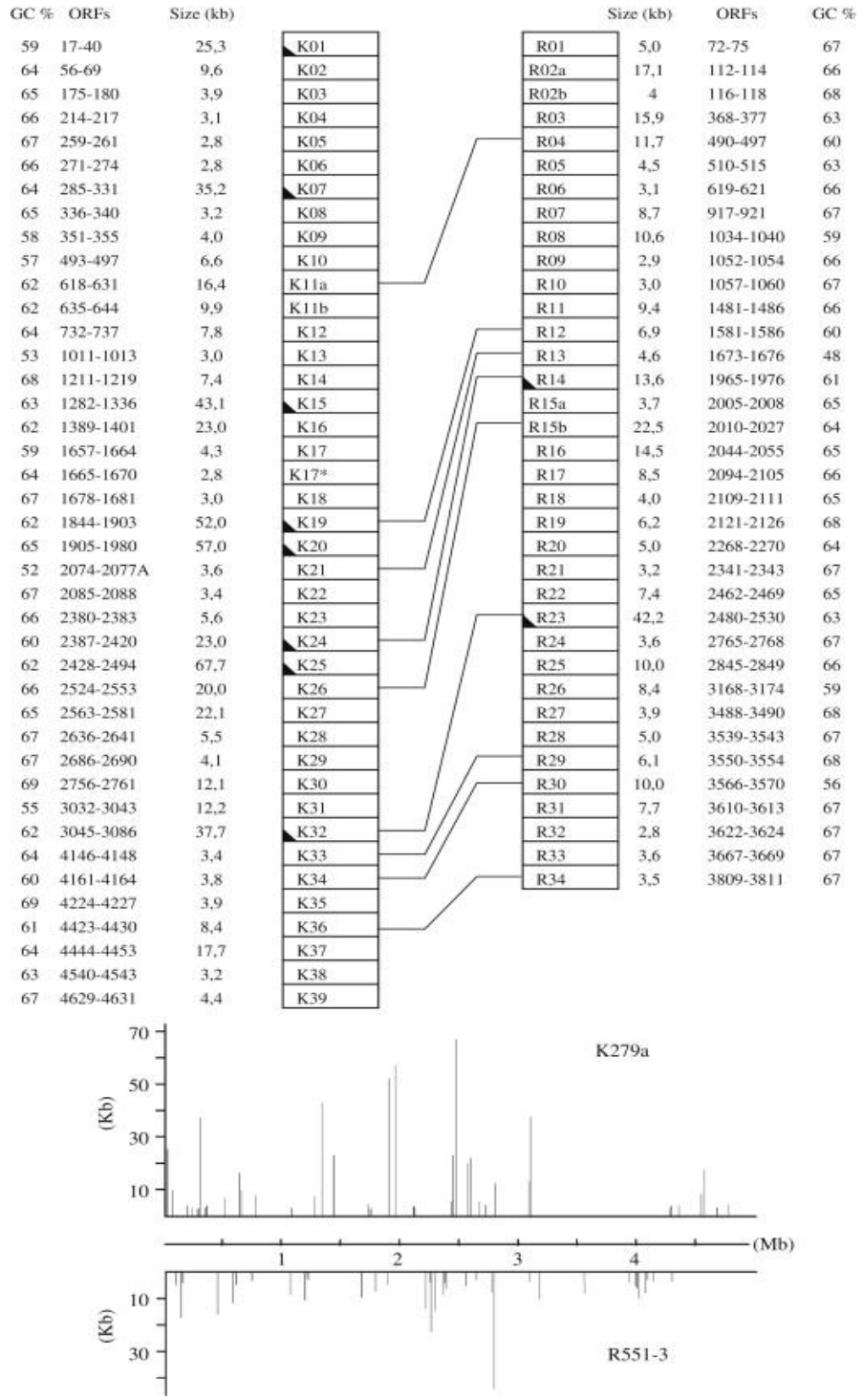

Fig 5: K279a and R551-3 GEIs. Size, orfs interval and G+C content of each GEI are shown. Islands inserted at the same relative chromosomal position in the 2 strains are connected by lines. Islands potentially encoding one or more integrases are marked by a triangle. The chromosomal distribution of GEIs is diagrammed at the bottom. The height of the lines denotes GEI size.

4.2. *S. maltophilia* GEIs boundaries

Sequence alignments provided unequivocal information on the ends of GEIs in a few cases only. Consequently, the length of most GEIs is underestimated, the start/stop codons of terminal orfs arbitrarily functioning as provisional GEI ends. However, the ends of 8 GEIs were exactly defined by sequence comparisons. Relatively to the mode of integration, the 8 GEIs can be sorted into 2 groups (Fig. 6). The first one includes K01, K02, and R19, which are not flanked by duplication of bases at the insertion site. The second group includes K07, K17, K20, K32, and R23. All these GEIs are integrated within tRNA or tm-RNA sequences, which are a preferential GEI target (Mantri and Williams, 2004). The termini of the 5 GEIs are flanked by target site duplications (TSDs) 14–53 bp long, in which segments of both target tRNA/tm-RNA and flanking DNA are duplicated. TSDs flank K20, but not the adjacent K19 island, which plausibly also used the tm-RNA as insertion target. The lack of TSDs has been observed for GEIs inserted at the tm-RNA locus (Williams, 2003).

Many large islands potentially encode integrases or functionally related enzymes. Both K32 and R23 are inserted at the serine tRNA (Fig. 6), but encode non-homologous integrases (orfs 3086 and 2480, respectively). K24 and R14 are inserted at the same relative chromosomal position. While actual boundaries have not been defined, it is plausible that also the integration of these 2 GEIs had been mediated by non-homologous proteins. These may correspond to one of the 2 gene products (orfs 2416 and 2417) encoded by Sm3, an insertion sequence located at one terminus of K24, and one of the 2 R14 integrases (orfs 1971 and 1972).

no target site duplications

K01 16017-41601

ATCGGTGAATCCGTACCGCGGTTGTGAGCAT [ggtatc //gtgtt] GCTGCTCCTACTGCTTCGCGCGCCCCTCGCA

K02 60294-70902

GGCGTGCAGGTCGGTGTGGCAGACGCCACAG [ctgcc // cacag] GCCTCGATCCTGACCAGCACCTCGCCCGCCC

R19 2377995-2384559

GATACGGAGCACAAGGACGCCGCTTGACCGG [tagat // tttgg] TAGCTGCCCACCTTGGTGGGCACCGATGGAA

tRNA/tmRNA targets

K07 299849-335431 (thr-tRNA, 47-53 bp TSD)

AACAACCCATGCCGCTTTAGCTCAGTTGGTAGAGCAGTTGGTAGAGCCAACTGTCTTGTAAACAGTAGGTCATCCGTTCG
ATTCGGATAAGCGGCACCAT [ttcgg // cagtt] CAACTATCTTGTAAACAGTAGGTCATCCGTTCGATTCGGATA
AGCGGCACCATCCCCATGA

K17 1720094-1724413 (cys-tRNA, 48 bp TSD)

TGGCGCGCAGTACGTTGGCCTCATGGCAGAGTGGCTATGCACCGGATTGCAAATCCGTTTACAGCGGTTCGATTCCGCTT
GAGGCCTCCA [attga // aggtg] CGGATTGCAAATCCGTTTACAGCGGTTCGATTCCGCTTGAGGCCTCCATTTG
AAGAGCCCTGACTCCGGCCAGGGCTTTTTTTGCGG

K20 1945413-2002722 (tmRNA, 34 bp TSD)

GGTGGAGGTGGGCGGAATTGAACCGCCGTCCGAAGGCACTCCATCCCCAGCACTACATGCTTAGCTCACCGTTGGATCTC
GTCCCCGAACAGCACGGCGTGCAAAGCGCATCCGGGAACCAGCCTGTTGTGTTCTAGTGCCGGACTGACAGGCAGCCACC
CAGCGCGATTCCATGATAGTGACTCTACACCGCGAGCATGGACACAAGCGGTTTCGAGGCTTAGGCCTTAAGCGGCCAGA
GCGTAGTTGTCGTCGTTGGCAACTAGAGTTTTGCAGCTGGATTTACGAGGAAAGCTACCCCCTCGGCATGCGCCAGGCGA
CTTCACAACCCCCGTCGAAACCAATGCACCCCCGGTTTCTTCAAGTGCTGCAAGG [ctttt // gggga] CAATGCA
CCCCCGGTTTCT-ACAGTGCTGCAAGGTACAGGGCCG

K32 3089420-3127152 (ser-tRNA, 18 bp TSD)

TTGCGATTATTTGGCGGAGAGAGTGGGA [cttcc // aagga] TTGGCGGAGAGAGTGGGATTCGAACCCACGGAAG
GTTTAACCCTTCGCCGGTTTTCAAGACCGGTGCCTTAAACCGCTCGGCCATCTCTCCAATCGGGTCCC

R23 2782893-2825334 (ser-tRNA, 14 bp TSD)

CTGAACATCTTGGCGGAGAGAGTG [tctga // aagac] TGGCGGAGAGAGTGGGATTCGAACCCACGGAAGGTTTA
ACCCTTCGCCGGTTTTCAAGACCGGTGCCTTAAACCGCTCGGCCATCTCTCCATCGGGTCCCG

Fig. 6: GEIs target sites. Sequences at the boundaries of specific GEIs are shown. GEI chromosomal coordinates and TSD lengths are shown. Islands are in brackets, and only the 5 terminal nucleotides at either side are shown in lowercase letters. tRNA and tm-RNA targets are underlined, TSDs are highlighted.

4.3. *S. maltophilia* GEIs gene products

The content of GEIs is highly variable. About 1/2 of island-encoded proteins to which a function could be assigned is represented by molecules mediating interactions with the environment. Gene products of interest encoded by GEIs are listed in Table 2 and are briefly discussed here.

(i) Metal resistance genes. *S. maltophilia* hosts several operons involved in import, storage, and efflux of metals. Some of these gene clusters are conserved at the same relative chromosomal position in the 2 strains, others are found in K279a only, on specific GEIs. As summarized in Table 2, K01 and K25 carry *czc* (cobalt–zinc–cadmium resistance) genes, K25 carries also genes involved in copper metabolism and homoeostasis (*cop* and *cus* operons), and K03 arsenic-resistance genes (*ars* operon). Mercury resistance genes are present in K279a but not in R551-3, and are located on K24.

(ii) T1SS (type I secretion system). Type I pili consist of a rod composed by a major fimbrial protein and 2 or more ancillary proteins. Such structures are assembled and secreted by the chaperon/usher pathway (Nishiyama *et al.*, 2008). T1SS genes are conserved at 2 sites in K279a and R551-3. A third, K279a-specific cluster, is encoded by K12.

(iii) T4SS (type IV secretion system). T4SS genes conserved in K279a and R551-3 are in the same order (*vir*D4, B8, B9, B10, B11, B1, B2, B3, B4, B6) of *X. axonopodis* T4SS genes (Alegria *et al.*, 2005). A second, K279a-specific T4SS cluster is encoded by K15. Here, genes are in the same order (*vir*D4, B11, B2, B3, B4, B5, B6, B8, B9, B10) of T4SS genes encoded by the PA7 strain of *Pseudomonas aeruginosa* (orfs 3708–3697). The relatedness of the 2 T4SS gene clusters is reinforced by the observation that the hypothetical proteins which separate VirB11 and VirD4 encoded by PA7 (orf 3707) and K279a (orf 1292) genomes are homologous.

(iv) Filamentous haemagglutinin (FHA) genes. FHA is a major colonization factor in *Bordetella pertussis* (Locht *et al.*, 1993), and FHA proteins have been

found in species as diverse as *Clostridium perfringens* (Myers *et al.*, 2006) and *Moraxella catarrhalis* (Balder *et al.*, 2007). FHAs are encoded along with transporter proteins (Jacob-Dubuisson *et al.*, 2001). Three *fha* and related transporter genes are found in *S. maltophilia*, and all are encoded by GEIs (K16, K37, R02a). The 3 FHAs vary in size, and this is correlated, as shown for other FHAs (Kajava *et al.*, 2001), to the number of repeats fitting the consensus LDNGGGX13–22.

(v) LPS genes. Variation at the interstrain level is common in LPS biosynthetic gene clusters inserted between the conserved *metB* and *etfA* loci in bacteria (Patil *et al.*, 2007). Alternative LPS gene sets, both carried by GEIs, are found in K279a and R551-3. K11 includes 2 gene clusters, K11a and K11b. The former contains genes playing a role in the O-chain synthesis found also in R04. K11b contains 7 genes known as *pmrIHFJLMK* or *arnABCDEFT*, encoding proteins which act in a coordinate manner to ultimately modify the lipid A by the addition of 4-amino-4-deoxy-L-arabinose. This modification causes resistance to polymyxin and cationic antimicrobial peptides (Yan *et al.*, 2007). The genes separating K11a and K11b (orfs 632–634) are homologous to genes located downstream from R04 (orfs 498–500). This suggests that K11 derives from R04 by the acquisition of K11b genes.

Additional proteins of interest potentially encoded by genomic islands are the homologue of McrBC (K10), a restriction endonuclease which cuts DNA containing modified cytosines (Panne *et al.*, 2001), the Ssu proteins (K27), involved in organosulfur metabolism (Kahnert *et al.*, 2000), several hypothetical proteins, all encoded by K26, which feature the BLUF (for sensors of blue-light using FAD) domain, a novel FAD-binding domain plausibly involved in sensory transduction in microorganisms (Gomelsky and Klug, 2002). Finally, K30 includes a cluster of genes involved in glycogen biosynthesis and catabolism. The *glg* gene products have a role in the formation of biofilms (Jackson *et al.*, 2002) and capsular layers (Sambou *et al.*,

2008), and hence may be catalogued as potential pathogenic genetic determinants.

***S. maltophilia* GEIs potential gene products**

Genes	Chromosome orfs		GEI orfs	
	K279a	R551-3	K279a	R551-3
czc	2697-2699 4606-4608	2169-2171 3956-3958	**K01** (0036-0038), **K25** (2457-2461)	-
cop	3691b2-3692 2691-2692	3105-3106 2163-2164	**K25** (2440-2449)	-
cus	2693-2694	2165-2166	**K25** (2433-2434)	-
ars	2421-2425	1977-1981	**K03** (0176-0179)	-
mer	-	-	**K24** (2409-2412)	-
T1SS	0706-0709 1508-1512	0561-0564 1267-1271	**K12** (0732-0736)	-
T4SS	2997-3008	2439-2451	**K15** (1283-1293)	-
fha	-	-	**K16** (1389-1390) **K37** (4452-4453)	**R02a** (0112-0113)
LPS	-	-	**K11a** (0618-0631) **K11b** (0636-0642)	**R04** (0490-0497)
mcrBC	-	-	**K10** (0497)	-
ssu	-	-	**K27** (2572-2574)	-
BLUF	-	-	**K26** (2528,2535,2541)	-
glg	-	-	**K30** (2756-2761)	-

Table 2: *S. maltophilia* GEIs potential gene products.

4.4. *S. maltophilia* GEIs and prophages

Five GEIs (K02, K07, K19, K20, and R23) encode mostly phage-related products. Homology searches enabled to correlate 2 of these islands to known prophages.

K07. This island encodes several phage-like products, including an integrase (orf 56) which is highly homologous (54% identity) to the analogous protein encoded by Xcc37 K, a 37,309-bp island identified in *Xanthomonas campestris* (Mantri and Williams, 2004). K07 and Xcc37 K are comparable in size and are closely related (Fig. 7A). Comparisons carried out at the prophage database (http://bicmku.in:8082/prophagedb) revealed that Xcc37 K indeed corresponds to the *X. campestris* P-like prophage PH138. K07 is also highly related to the *P. aeruginosa* cytotoxin-converting P2-like phage CTX (Nakayama *et al.*, 1999). K07 and CTX share a common structural scaffold, but the 2 phages diverge because each carries a specific set of genes at one end. K07 plausibly corresponds to the 'phage II cluster' described by Crossman *et al.* (2008).

K19. K19 and K20 are closely located GEIs, spanning together a 109-kb chromosome segment, and plausibly correspond to the region defined as 'phage I cluster' by Crossman *et al.* (2008). Both GEIs potentially encode phage-like products but can be distinguished as individual entities, because they are separated by the tm-RNA target (residues 1945038–1945390). K19 exhibited significant homology to a 41-kb prophage-like sequence identified in the *Minibacterium massiliensis* genome (residues 3029900–3070592, see Audic *et al.*, 2007). K19 and *M. massiliensis* sequences are largely collinear (Fig. 7B), but diverge because K19 carries 10 kb of non-phage DNA at the 5' end side.

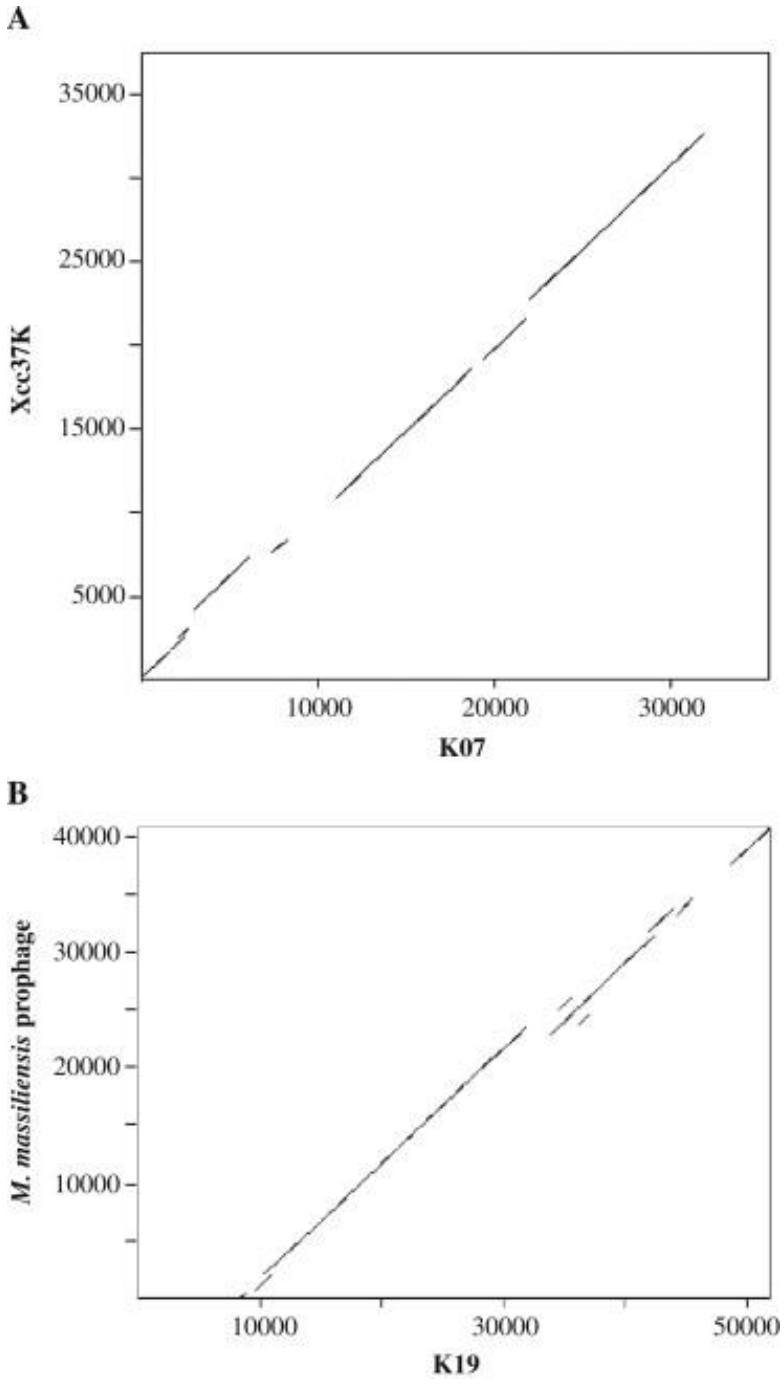

Fig. 7: Phage GEIs. Dot plot alignments of the nucleotide sequences of K07 and Xcc37 K islands (panel A), and K19 and a *M. massiliensis* prophage (panel B), are shown. Sizes are in nt.

4.5. *S. maltophilia* Solo ORFs

At a first glance, K279a and R551-3 genomes are collinear throughout and differ because equipped with alternative sets of GEIs. A closer look at the chromosome backbone, however, revealed that the genome of each strain is punctuated by 'solo' orfs, for which homologous potential gene products could not be identified at the corresponding chromosomal position in the other strain. 172 and 160 orfs, measuring 1–3 kb, are scattered along the genomes of K279a and R551-3, respectively. Most are single, many come in pairs, a few in small clusters (3–4 orfs) not catalogued as GEIs because of their small size. The origin of these orfs is unclear. Some are found at positions marked in the other strain by GEIs, and may plausibly represent sequences removed upon island insertion. Others are found only in one strain because of mutations affecting homologous sequences in the other strain. Orf 3990 in K279a and orf 3400 in R551-3 are encoded by homologous DNA segments, but differ because translated from initiating GTG and ATG triplets on different frames. About 60% of strain-specific orfs is constituted, in both genomes, by hypothetical proteins. Interestingly, membrane proteins account for 15% of the 'solo' orfs in K279a, but only for 6% in R551-3.

4.6. GEIs in the *S. maltophilia* population

To check whether islands identified in K279a are present in the population, the DNAs of 41 *S. maltophilia* strains were analyzed by PCR and dot-blot hybridizations. (Fig. 8). The distribution of GEIs greatly varied among strains, the degree of conservation ranging from zero to 90%. Strains may be tentatively assigned to 3 groups. 13/41 strains contain the largest number (7–10) of tested islands, 21/41 strains contain 1–6 islands, and 7 carry no islands. GEIs were distributed apparently at random. However, it can be noticed that

K11, K15, K16, K23, K25, and K32 are over-represented, K07, K19, K20, and K24 are under-represented.

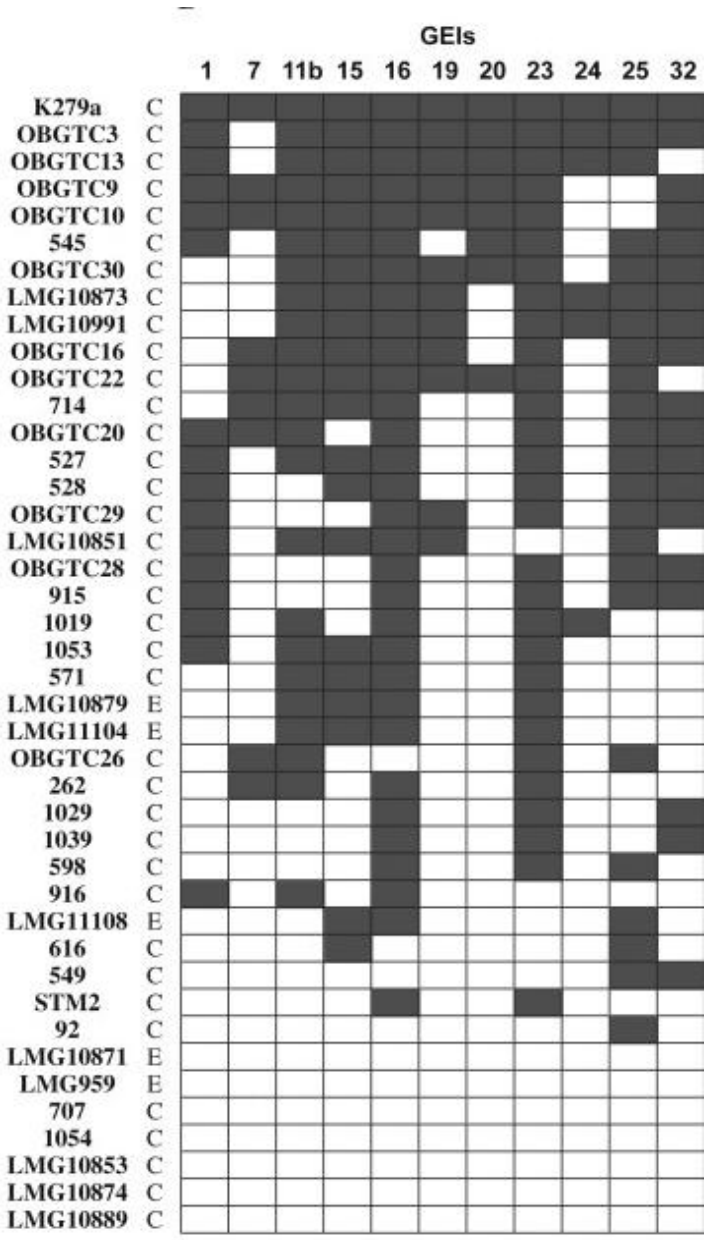

Fig. 8: Distribution of GEIs among *S. maltophilia* isolates. One microgram of total DNA from the indicated strains was hybridized to 32-P radiolabelled, 300–600 bp island-specific DNA probes. Probes were amplified from K279a orfs 32 (K01), 285 (K07), 638 (K11b), 1290 (K15), 1389 (K16), 1901 (K19), 1906 (K20), 2382 (K23), 2412 (K24), 2447 (K25), 3086 (K32). C, clinical isolates; E, environmental isolates.

4.7. The *A. baumannii* chromosome

Making use of the Mauve software (Darling *et al.*, 2010), the proteins putatively encoded by the draft genomes of the *A. baumannii* strains 3990, 3909 and 4190 (Zarrilli *et al.*, 2011) were compared to the ORFs encoded by the wholly sequenced genomes of the *A. baumannii* AB0057 and AYE strains assigned to ST1, ACICU strain assigned to ST2, ATCC17978 strain assigned to ST77 (Iacono *et al.*, 2008; Adams *et al.*, 2008).

A. baumannii genomes exhibit extensive synteny. Sequence comparisons revealed that 3068 coding regions are conserved, at the same chromosomal position, in the compared *A. baumannii* genomes. Genes encoding proteins shown or hypothesized to be important for pathogenicity are conserved in the analyzed strains at the same relative chromosomal position (Table 3). The set includes *OmpA*, the outer membrane protein which has role in biofilm formation (Gaddy *et al.*, 2009) and induces, when secreted, death of epithelial and dendritic cells (Lee *et al.*, 2010), the DD-endopeptidase, which contributes to the resistance of *A. baumannii* to bactericidal activity presumably by remodelling the cell surface (Russo *et al.*, 2009), phospholipase D, an enzyme crucial for proliferation in human serum (Jacobs *et al.*, 2010), proteins involved in the formation of capsule (Russo *et al.*, 2010), type I pili (Tomaras *et al.*, 2003), and iron metabolism (Zimbler *et al.*, 2009). According to the published annotation, OmpA, DD-endopeptidase, phospholipase D, and many other deduced gene products are smaller in ATCC 17978 as compared to their orthologs. Size differences do not denote allelic variation, but are determined by the criteria adopted to select the initiating methionine in ATCC17978 ORFs.

Multidrug resistance is a key feature of *A. baumannii* and several genes have a role in establishing a MDR phenotype. Genes encoding efflux pumps and resistance proteins shown or hypothesized (Coyne *et al.*, 2011) to be involved in the process are conserved in all strains (not shown). In contrast, genes

encoding drug-inactivating and drug-resistant enzymes reside in accessory DNA regions which are present only in some strains, in specific genomic islands (see below).

Shared synteny lets to represent the *A. baumannii* chromosomes as ˜4 Mb long DNA segments homologous to each other throughout their lengths (Fig. 9). DNA tracts, ranging in size from 4 to 126 kb, are present in one or more strains, but missing or replaced by alternative DNA segments in others (see vertical bars in Fig. 9). Some of these regions correspond to DNA sequences earlier suspected to be mobile because found in *A. baumannii* but not in *A. baylyi* DNA or vice versa (Smith *et al.*, 2007). Specific 15-36 kb regions are missing in all strains but AB0057 (see triangles in Fig. 9), and may therefore plausibly correspond to strain-specific deletions. Many of the accessory genomic DNA segments exhibit characteristic features of genomic islands, such as the presence of insertion sequences at one end, a GC content different from the bulk chromosome, insertion within tRNA or non-coding RNA genes, target site duplications (TSDs) at the ends formed upon genome integration (Dobrindt *et al.*, 2004). For sake of simplicity, all the accessory DNA regions have been called GEnomic Islands (GEIs). GEIs found at the 63 variable loci identified in the *A. baumannii* genomes, and some of their properties, are diagrammatically reported in Fig. 10. In text and figures individual GEIs are referred by the locus number and the strain acronym used in Fig. 10. Core and accessory chromosomal DNAs are fully conserved in ACICU and 3990 strains. Because of this, only the ACICU GEIs are shown in Fig. 10. In draft genomes some GEIs reside in different contigs. The colinearity of the contigs and the GEI DNA content of the corresponding chromosomal regions were assessed by sequencing PCR products bridging contigs ends.

A close look at *A. baumannii* chromosomes further identified about one hundred DNA regions encoding 1-2 ORFs smaller than 4 kb conserved in one or more strains, but missing, or replaced by non homologous DNA of

comparable length in others. These smaller accessory regions correspond to the "solo" ORFs described in *S. maltophilia* genomes.

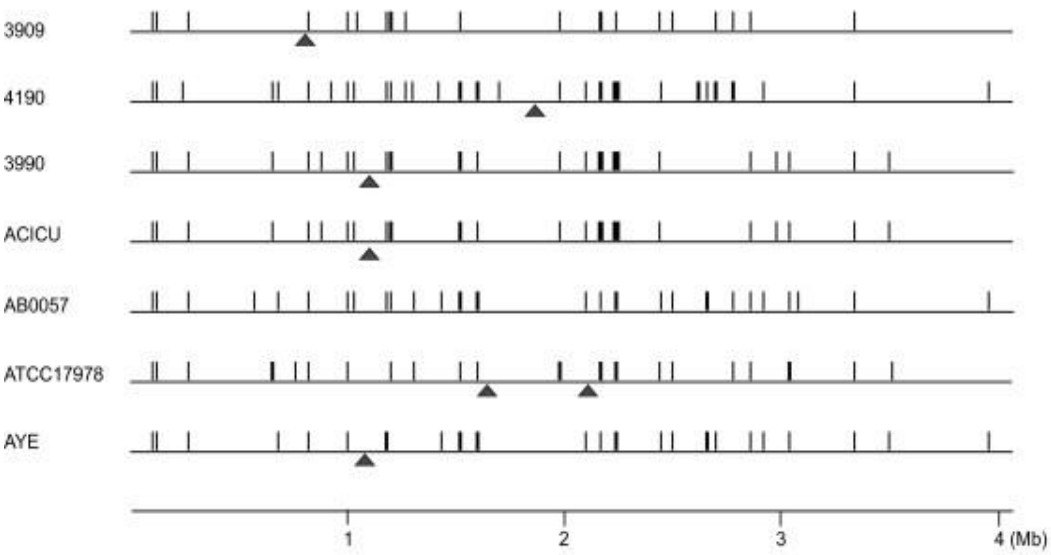

Fig. 9: Comparison of *A. baumannii* genomes. The seven *A. baumannii* genomes analyzed have been aligned. Accessory regions are denoted by vertical bars. Strain-specific deletions are marked by triangles.

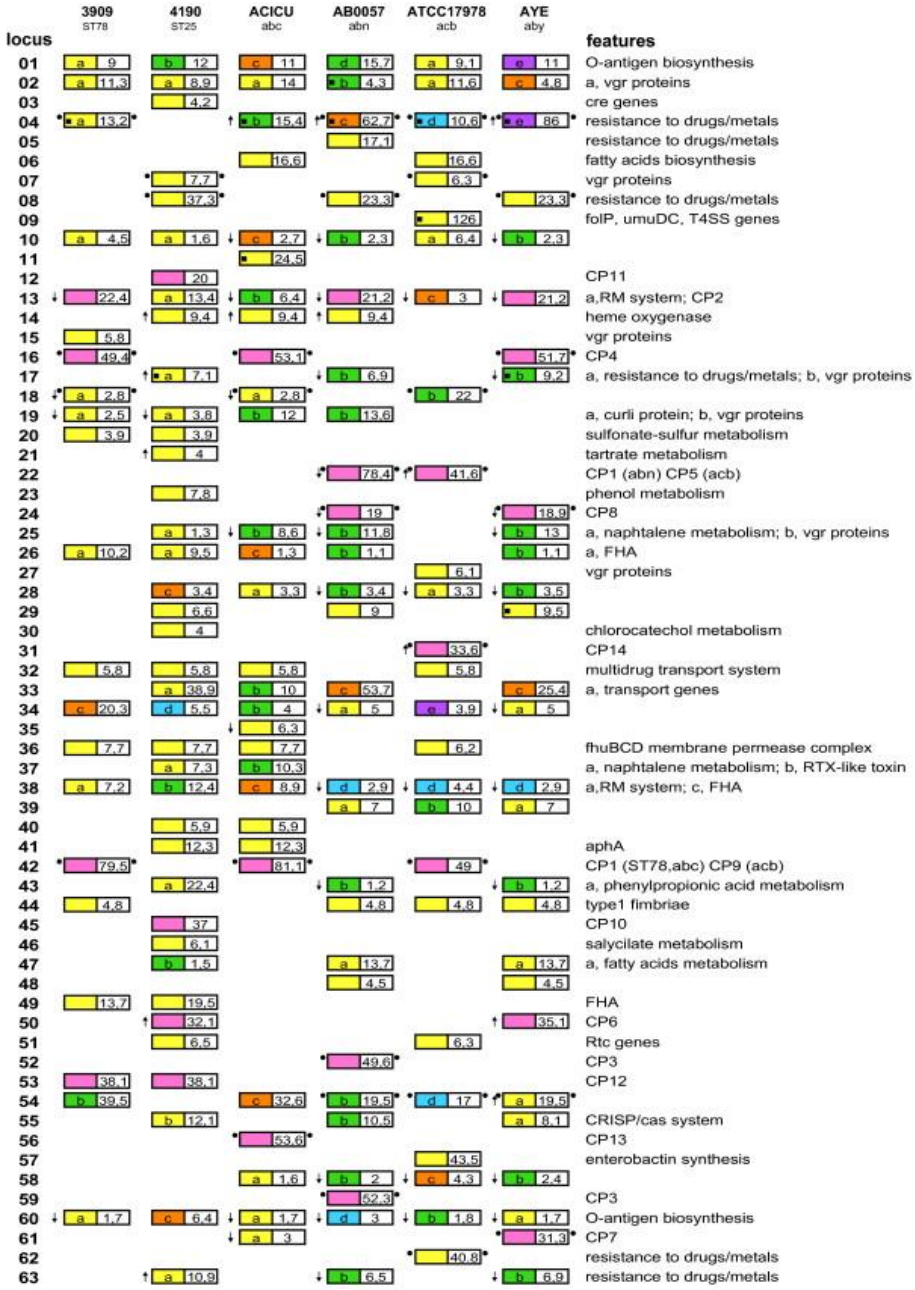

Fig. 10: Variable regions in *A. baumannii* genomes. A chart of the genomic islands (GEIs) depicted as bars in figure 9 is displayed. Each line corresponds to a chromosomal locus. Different GEIs inserted at the same locus in different strains are marked by different colours and lower case letters. Sizes of GEIs are given in kb. Black boxes within GEIs denote mobile sequences, down and up arrows to the left indicate that the GEI G+C content is lower than 36% or higher than 42%, respectively. Dots flanking GEIs denote TSDs. The strain names and relative acronyms used throughout the text are given at the top. Acronyms below complete genomes are those used at Kyoto Encyclopaedia of Genes and Genomes (KEGG).

Gene products involved in pathogenicity in *A.baumannii* genomes

Gene products	Strains						
	AB0057	**AYE**	**3990**	**ACICU**	**4190**	**ATCC17978**	**3909**
capsule formation							
tyrosine kinase Ptk	91	3818	936	71	3295	49	2600
Tyrosine phosphatase Ptp	92	3817	935	72	3296	50	2601
type I pili formation							
CsuE	2565	1324	787	2414	3382	2213	744
CsuD	2566	1323	786	2415	3383	2214	745
CsuC	2567	1322	785	2416	3384	2215	746
CsuB	2568	1321	784	2417	3385	2216	747
CsuA	2569	1320	783	2418	3386	2217	748
CsuA/B	2570	1319	782	2420	3387	2218	3415
iron metabolism							
nonribosomal peptide synthetase BasD	2811	1095	2421	2579	tblastn	2383	1389
nonribosomal peptide synthetase BasC	2812	1094	2420	2580	3813	2384	tblastn
ferric acinetobactin receptor	2813	1093	2419	2581	3814	2385	3376
ferric acinetobactin transport system periplasmic binding protein	2814	1092	2418	2582	3815	2386	3375
erric acinetobactin transport system ATP-binding protein	2815	1091	2417	2583	3816	2387	3374
ferric acinetobactin transport system permease	2816	1090	2416	2584	3817	2388	3373
ferric acinetobactin transport system permease	2817	1089	2415	2585	3818	2389	3372
hemin utilization							
biopolymer transport protein ExbD/TolR	1827	2051	351	1629	227	1063	1994
biopolymer transport protein ExbD/TolR	1828	2050	352	1630	228	1064	1993
biopolymer transport protein	1829	2049	353	1631	229	1065	1992
TonB family protein	1830	2047	354	1632	230, 231	3708*	1991
TonB-dependent receptor	1831	2046	355	1633	232	1606, 1607	1990, 1989
heme-binding protein A	1832	2045	358	1634	234	1608	1987
heme-binding protein A	1833	2044	359	1635	235	1609	1986
Zn-dependent oligopeptidase	1834	2043	360	1636	236	1610	1985
ABC-type dipeptide/oligopeptide/nickel transport system permease component	1835	2042	361	1637	237, 238	1611	1984
ABC-type dipeptide/oligopeptide/nickel transport system permease component	1836	2041	362	1638	239	1612	1983
glutathione import ATP-binding protein GsiA	1837	2040	363	1639	3719	1613	1982

Table 3: * The asterisk indicates one of the 436 proteins putatively encoded by ATCC17978 not included in the GenBank:NC_009085 file. tblastn refer to unannotated 4190 and 3909 proteins identified by tblastn searches.

4.8. Categories of genomic islands of *A. baumannii*

Some islands are strain-specific; others are completely or partially conserved in more than one strain. Non homologous islands are inserted at the same locus in different strains, and some loci are extremely heterogeneous, featuring up to 4-5 alternative islands. Some islands are composite, and changes in their organization among strains are correlated to changes in the number and association of specific DNA segment.
On the basis of the putative gene products, GEIs can be broadly sorted into a few categories. Properties and overall organization of relevant GEIs are below discussed.

4.8.1. Resistance islands

Many of the accessory drug resistance determinants found in AB0057 and AYE are encoded by genes located within G4aby, G4abn and G5abn, which correspond to the resistance regions previously described as AbaR1, AbaR3, and AbaR4 (Adams *et al.*, 2008), respectively. G4aby and G4abn are both inserted in the *comM* gene, and result from the association of the 16 kb Tn6019 transposon with multiple antibiotic resistance regions (MARR), which are delimited by Tn6018 elements (Post *et al.*, 2010). Tn6019 features genes involved in transposition (*tniA, tniB*), an arsenate resistance operon, a universal stress protein gene (*uspA*), and a sulphate permease gene (sup). MARR are inserted within *uspA* and vary in length and composition (Post *et al.*, 2010). The G4abc island of the ACICU genome corresponds to the AbaR2 region (Post *et al.*, 2010), which carries few resistance genes and lacks Tn6019 sequences (Fig. 11A). G4ST78 is similarly inserted in the *comM* gene, and features genes homologous to *tniA* and *tniB* (38-40% identity of the gene products), but lacks resistance genes and encodes a set of hypothetical proteins

(Fig. 11A). G4 is missing in strain 4190. However, resistance genes are scattered in different GEIs of this strain (Fig. 11B). The *aadA1* (streptomycin 3"-adenylyltransferase) gene, flanked by *satR* (streptothricin acetyltransferase) and *dhfr* (dihydrofolate reductase) genes are found in G63ST25. Genes involved in resistance to mercury (*merRCAD* cluster) are located in G17ST25, and a 4.5 kb DNA segment containing *feoAB* (ferrous iron transport operon), *czc* (tricomponent proton/cation antiporter efflux system) and *ars* (arsenite transporters) genes are found in G8ST25, next to the *cus* (copper resistance) genes conserved in all G8 (Fig. 11B). The G62acb region also contains *cus*, *feo* and *czc* genes involved in heavy metal resistance. These genes differ in sequence and overall arrangement from G8ST25 homologs. This supports the notion that the set of accessory genes had been independently acquired by the strains 4190 and ATCC17978.

Additional resistance genes found in GEIs include an aminoglycoside phosphotransferase gene (G41ST25, G41abc), a dihydropteroate synthase gene (G9acb), and an ABC-type multidrug transport system, conserved in all the G32 islands.

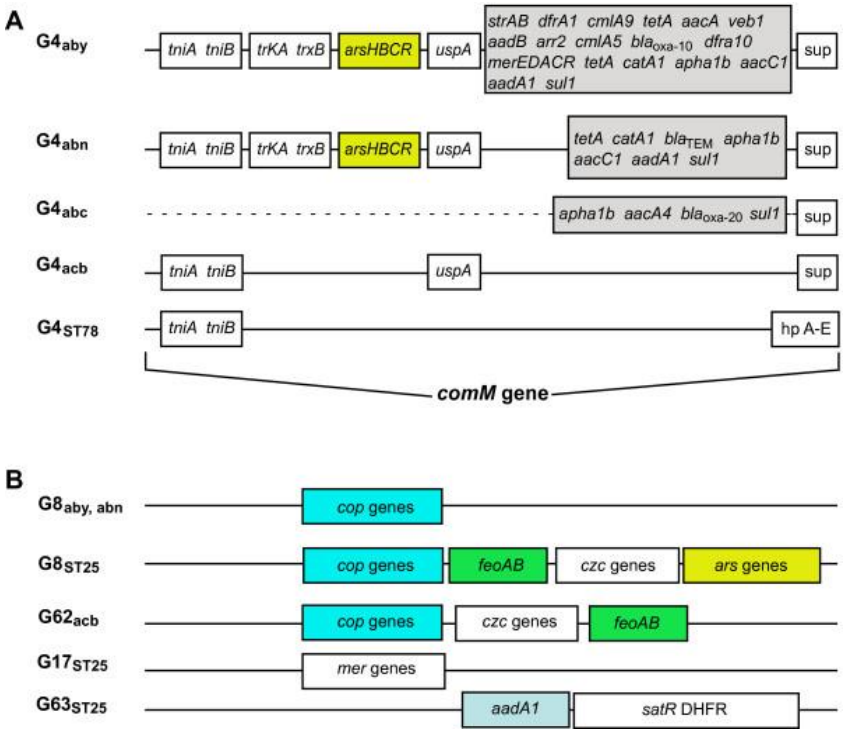

Fig. 11: Resistance gene islands. A) Diagrammatic representation of G4 islands. The structure of the resistance islands and gene symbols are as in reference 30. Grey boxes represent MARR. Deleted DNA in G4abc is marked by a dotted line. B) Resistance genes in other GEIs.

4.8.2. Surface components and transport systems

GEI-1 and GEI-60 host genes involved in cell envelope. Heterogeneity among *A. baumannii* strains at the level of O-antigen biosynthetic genes was already noticed (Adams *et al*., 2008), and is correlated to the presence of alternative glycosylases. The G44 island, present in all strains but ACICU, 3990 and 4190, is a four gene operon involved in the assembly of fimbriae (type I pili) by the chaperone/usher pathway (Nuccio et Bäumler, 2007). G44aby corresponds to the surface adhesion protein region annotated as Cus1R in the AYE genome (Vallenet *et al.*, 2008). G19ST25 and G19ST78 are related islands which both carry an operon encoding three hypothetical lipoproteins. Of these, one exhibits homology to *CsgG*, the key factor in the secretion of curli, the proteinaceous component having a role in host cell adhesion and biofilm formation in many Enterobacteriaceae (Barnhart et Chapman, 2006). Purified CsgG forms ring-shaped complexes analogous to those formed by outer membrane channel-forming proteins (Barnhart et Chapman, 2006). The CsgG-like protein, in association with the two co-expressed lipoproteins, may influence the permeability of the outer membrane of *A. baumannii*.

Filamentous haemagglutinin (FHA) is a major virulence factor in Bordetella pertussis (Locht *et al.*, 1993). *fhaB* and *fhaC* genes, respectively encoding the haemagglutinin and the transporter protein, have been identified in many pathogens (Rocco *et al.*, 2009). *fhaBC* gene clusters are found at the same loci in strains 4190 and 3909 (islands G26ST25, G26ST78, G49ST25 and G49ST78), and strains ACICU and 3990(islands G38abc and G38ST2). The transporter proteins are highly conserved in the four clusters, whereas FHAs vary in length (1834 to 4812 amino acids), mostly because of changes in the number and organization of body sequence repeats (Locht *et al.*, 1993). A 3216 amino acids long calcium binding hemolysin protein, unrelated to FHAs, is encoded by G18acb.

Cyclopropane fatty acids (CFA) are phospholipids found in the bacterial membranes in the late exponential and early stationary phases of cell growth (Zhang et Rock, 2008), which derive from the corresponding unsaturated fatty acid (UFA) phospholipids. The synthesis of CFA is catalyzed by the enzyme CFA synthase, the substitution of a saturated by an unsaturated fatty acid by the enzyme delta-9 acyl-lipid desaturase. CFA synthase and delta-9 acyl-lipid desaturase are both encoded by G47abn and G47aby.

G33ST25 is a large island which encodes four different transport and translocation systems: i) Tat (twin-arginine translocation) proteins, involved in the translocation of folded proteins to the cell envelope or the extracellular space ii) a TonB/ExbBD complex iii) a Opp (oligopeptide transport proteins) complex iv) a sulfur utilization system, made by a FMNH2-dependent sulfonatase and three ABC-type transporters, which resemble the products of the *E. coli* ssu gene cluster (van der Ploeg *et al.*, 1999). Two unlinked copies of the sulfonatase gene are also present. Genes involved in the capture and intracellular transport of iron are found in different islands. G57abc carries a gene cluster involved in the synthesis of the high-affinity siderophore enterobactin. Heme oxygenase is an alternative to siderophores to capture iron from the environment (Frankenberg-Dinkel, 2004). G14, an island which is conserved in 4190, ACICU and AB0057, carries an operon encoding a heme oxygenase, an outer membrane and a TonB family protein. The presence of a flanking *fecIR* gene cluster suggests that heme internalization may be regulated by the Fec transduction system (Braun *et al.*, 2006). The *fhuBCD* genes, which catalyze the internalization of iron III hydroxamate compounds, are located on G36, an island conserve in all strains but AB0057 and AYE.

4.8.3. Metabolic islands

Many GEIs carry genes encoding proteins involved in specific metabolic pathways. G23ST25 carries a *mph* (multi component phenol hydroxylase) gene complex, involved in the conversion of phenol to cathecol, flanked by a sigma54-dependent activator gene. It has been shown that the expression of *mph* gene complex described in *Acinetobacter sp.* PHAE-2 is dependent on the alternative sigma factor RpoN (Xu *et al.*, 2003). G37ST25 carries *nag* genes, involved in the metabolism of naphthalene. In *Ralstonia* (Zhou *et al.*, 2001), *nag* genes are arranged in two separate clusters, involved in the conversion of naphthalene to gentisate (*nagAGHBFCQED* genes), and gentisate to pyruvate and fumarate (*nagIKL genes*), respectively. In G37ST25 *nagIKL* genes and *nagGH,* encoding the salicylate 5-hydroxylase, are linked, and flanked by benzoate transport genes.

G43ST25 carries genes involved in the catabolism of 3HPP (3-hydroxyphenylpropionic acid) and PP (phenylpropionic acid). In *E. coli*, the dioxygenase complex *(hcaEFCD* genes), and the dihydrodiol dehydrogenase (*hcaB* gene) oxidize PP (phenylpropionic acid) and CI (cinnamic acid) to DHPP (2,3-dihydroxyphenylpropionate) and DHCI (2,3-dihydroxycinnamic acid), respectively. These substrates are subsequently converted to citric acid cycle intermediates by the *mhp* genes products (Díaz *et al.*, 2001). The *hca* and *mhp* genes, separated in *E. coli*, are linked and interspersed with additional genes in G43ST25. G21ST25 potentially encodes 4 proteins (tartrate dehydratase subunits alpha and beta, a MFS transporter and a transcriptional regulator) possibly involved in the metabolism of tartrate. Proteins exhibiting homology to the dienelactone hydrolase, an enzyme which plays a crucial role in the degradation of chloro-aromatic compounds, are encoded by the islands G30ST25, G34abn and G34aby. G46ST25 is made by an operon including the salicylate 1-monooxygenase (*salA*), a benzoate transporter (*benK*) and the *salA* regulator (*salR*) genes. A salicylate 1-monooxygenase is also encoded by

G25ST25. The genes *fabA, fabB, fabG, fabF, acpP, pslB, acsA*, involved in the biosynthesis of fatty acids (Zhang et Rock, 2008) are conserved in all *A. baumannii* strains, at separate loci. Orthologues of all these genes are clustered in G6abc and G6acb.

4.8.4. Phage islands

Many variable genomic regions are relatively large (19 to 82 kb) DNA blocks which potentially encode typical phage products. These regions have all been classified as cryptic prophages (CPs; see Fig. 10). Three to six CPs were identified in each strain. Six of the different 14 CPs identified are present in two or more strains, the remaining 8 are strain-specific. CPs characteristically carries at one end an integrase gene, and many are sharply defined by flanking TSDs induced upon insertion. CPs are poorly related to each other, and even CPs of the same type differ in size and coding ability. Ten of 14 CPs were assigned to four groups on the basis of sequence homologies. CPs found at the same locus encode identical or highly homologous (> 80% identity) integrases. CP1 encode different integrases, which are homologous to CP5- or CP9-encoded enzymes. This explains why CP1 and CP5 in AB0057 and ATCC17978 (G22abn and G22acb, respectively), and CP1 in 3909 and ACICU (G42ST78 and G42abc), and CP9 in ATCC 17978 (G42acb), are inserted at the same locus. CP3 are integrated at different sites of the AB0057 genome (G52abn and G59abn), but the target in both is an *arg-tRNA* gene.
Remnants of prophage sequences are found in G33abn and G33aby. These islands share the G33abc backbone, but contain also large DNA segments, reiterated in a head-to-tail configuration, in which genes encoding phage and hypothetical proteins are variously interleaved. G33abn and G33aby hypothetical gene products exhibit poor homology to all CPs gene products, and therefore were not included among CPs.

Phages may acquire ORFs named morons (Hendrix *et al.*, 2000) by lateral gene transfer. The PapS reductase (3'-phosphoadenosine 5'-phosphosulfate sulfotransferase) encoded by CP13 (G56abc), the toxin-antitoxin (TA) system encoded by CP1 (G42abc and G42ST78), the proofreading 3'-5' exonuclease epsilon subunit of the DNA polymerase III in the above mentioned CPs, the *umuDC* gene products, which are the components of the error-prone DNA polymerase V, again in CP1 (G22abn and G42ST78) and CP5 (G22abc) can all be considered morons. Not surprisingly, these enzymes are frequently associated with mobile genome elements (Permina *et al.*, 2002). Unlinked *umuD* and *umuC* genes are conserved in all *A. baumannii* strains.

G9acb also contains an *umuDC* cluster. This 126 kb region, found only in the ATCC 17978 strain, is a composite genomic island, carrying at one end a dihydropteroate synthase gene, at the other a DNA mismatch repair enzyme. G9acb carries a complete set of type IV secretion system (T4SS) genes, arranged in the same order in which T4SS homologs are found on the 153 Kb plasmid of *Yersinia pseudotuberculosis* IP31758 strain (Eppinger *et al.*, 2007). Because *umuDC* genes are carried by this plasmid, one may hypothesize that raises G9acb had been imported from Yersinia. In addition, a G9acb gene cluster, including an integrase, a DNA helicase and a TrbL/VirB6 conjugal transfer protein is highly homologous to a gene cluster from *Enterobacter cloacae*.

4.8.5. Additional islands

G3ST25 carries a cre genes cluster. In *E. coli* the *cre* locus includes a response regulator (*creB*) a sensor kinase (*creC*) and an inner membrane protein (*creD*). The corresponding two-component regulatory system *CreB-CreC* controls the expression of a variety of genes, among which the *creD*

regulator. Overexpression of *CreBC* causes modification of the envelope, inducing the colicin E2 tolerance phenotype (Cariss *et al.*, 2010).

G51ST25 and G51acb carry the *rtcA* and *rntZ* genes, encoding the RNA 3'-terminal phosphate cyclase and the RNAseZ, respectively. The cyclase catalyzes the ATP-dependent conversion of the 3'-phosphate to the 2', 3'-cyclic phosphodiester at the end of various RNA substrates (Genschik *et al.*, 1998); RNAseZ is responsible for the maturation of the 3'-end of a large family of transfer RNAs (Redko *et al.*, 2007). In *E. coli* the 3'-terminal phosphate cyclase *rtcA* gene forms an operon with the upstream *rtcB* gene. Expression of *rtcAB* is regulated by *rtcR*, a gene positioned upstream of *rtcAB*, but transcribed in the opposite direction, encoding a sigma54-dependent regulator (Genschik *et al.*, 1998). *rtcBA* and *rtcR* genes are conserved in both G51ST25 and G51acb islands, separated by *rntZ*. Interestingly, only *rntZ* is present at the corresponding chromosomal position in strains lacking G51.

In type I restriction systems the three subunits S, M and R, which may variably associate to form a modification methylase or a restriction endonuclease, are encoded by *hsd* (host specificity of DNA) genes. Alternative *hsd* genes reside in G13ST25 and G13ST78. The former are clustered in one operon, whereas *hsdSM* and *hsdR* genes in G13ST78 are at distance, as frequently found in other species.

Homologs of a cytosine DNA methyltransferase and a restriction endonuclease, which may constitute a type II restriction modification system, are encoded by genes residing in G38ST78.

The G55 islands found in strains 4190, AB0057 and AYE are closely related, and all include a CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) block, flanked by a *cas* (CRISPR-associated) gene cluster. CRISPRs are repeated DNA sequence blocks found in the genomes of approximately 40% of bacteria, often next to a cluster of *cas* genes. The CRISPR/Cas system provides a form of acquired immunity against exogenous DNA, foreign DNA sequences being first integrated at the CRISPR locus and eventually degraded

by Cas proteins (Horvath et Barrangou, 2010). Horizontal transfer of CRISPRs and associated genes among prokaryotes is documented (Godde et Bickerton, 2006).

Gram-negative bacteria contain a variety of genes encoding proteins enriched in dipeptide motifs (valine-glycine repeats) hence called Vgr. Islands encoding Vgr-like proteins are found inserted at eight genome variable loci (loci 2, 7, 15, 17, 19, 25, 27 of Fig. 10). Vgr proteins are associated with ligand-binding proteins at the bacterial surface, and are involved in biofilm formation and swarming and swimming motility in *Burholderia* (Bernier et Sokol, 2011). Intriguingly, Vgr proteins, along with Hcp (hemolysin co-regulated) proteins, are components of the type VI (T6SS) secretion apparatus, a transport system extensively conserved among Gram-negative bacteria (Bingle *et al.*, 2008). Secreted Vgr proteins assemble a cell-puncturing device analogous to phage tail spikes to deliver effector proteins, and are also able to covalently cross-link host cell actin contributing to T6SS pathogenicity (Pukatzki *et al.*, 2007). A T6SS gene cluster is conserved in all the analyzed *A. baumannii* strains.

CONCLUSION

The comparison of the whole DNA sequences of different strains of *S. maltophilia* and *A. baumannii* provided valuable informations on the nature of genomic variation in either species. The information gleaned from this study helped to build chromosomes scaffolds to be exploited to rapidly evaluate the degree of genomic variations in clinical isolates. Changes in island profiling will be useful in genomic epidemiology studies. Analysis carried out on novel sequenced genomes will help to define more rigorously the core genome component in both species, and hopefully link the presence of specific islands to selective pathogenic traits.

ACKNOWLEDGMENTS

I am indebted to my supervisor Pier Paolo Di Nocera, for continuous support in all this period, and Eliana De Gregorio for sincere friendship.

REFERENCES

Adams MD, Goglin K, Molyneaux N, Hujer KM, Lavender H, Jamison JJ, MacDonald IJ, Martin KM, Russo T, Campagnari AA, Hujer AM, Bonomo RA, Gill SR: Comparative genome sequence analysis of multidrug-resistant *Acinetobacter baumannii*. J Bacteriol 2008, 190:8053-8064.

Alegria MC, Souza DP, Andrade MO, Docena C, Khater L, CH Ramos, AC Silva and Farah CS: Identification of new protein-protein interactions involving the products of the chromosome- and plasmid-encoded type IV secretion loci of the phytopathogen Xanthomonas axonopodis pv. citri. J. Bacteriol. 2005, 187: 2315–2325.

Audic S, Robert C, Campagna B, Parinello H, Claverie JM, Raoult D and Drancourt M: Genome analysis of Minibacterium massiliensis highlights the convergent evolution of water-living bacteria. PLoS Genet. 2007, 3:e138.

Balder R, Hassel J, Lipski S and Lafontaine ER: Moraxella catarrhalis strain O35E expresses two filamentous hemagglutinin-like proteins that mediate adherence to human epithelial cells. Infect. Immun. 2007, 75:2765–2775.

Barnhart MM, Chapman MR: Curli biogenesis and function. Annu Rev Microbiol 2006, 60:131-147.

Bernier SP, Sokol PA: Use of suppression-subtractive hybridization to identify genes in the Burkholderia cepacia complex that are unique to Burkholderia cenocepacia. J Bacteriol 2005, 187:5278-5291.

Bingle LEH, Bailey CE, Pallen MJ: Type VI secretion: a beginner's guide. Curr Opin Microbiol 2008, 11:3-8.

Braun V, Mahren S, Sauter A: Gene regulation by transmembrane signaling. Biometals 2006, 19:103-113.

Cariss SJ, Constantinidou C, Patel MD, Takebayashi Y, Hobman JL, Penn CW, Avison MB: YieJ (CbrC) mediates CreBC-dependent colicin E2 tolerance in *Escherichia coli*. J Bacteriol 2010, 192:3329-3336.

Coyne S, Courvalin P, Périchon B: Efflux-mediated antibiotic resistance in Acinetobacter spp. Antimicrob Agents Chemother 2011, 55:947-953.

Crossman LC, Gould VC, Dow JM, Vernikos GS, Okazaki A, Sebaihia M, Saunders D, Arrowsmith C, Carver T, Peters N, Adlem E, Kerhornou A, Lord A, Murphy L, Seeger K, Squares R, Rutter S, Quail MA, Rajandream MA, Harris D, Churcher C, Bentley SD, Parkhill J, Thomson NR and Avison MB: The complete genome, comparative and functional analysis of *Stenotrophomonas maltophilia* reveals an organism heavily shielded by drug resistance determinants. Genome Biol. 2008, 9:R74.

Darling AE, Mau B, Perna NT: progressiveMauve: multiple genome alignment with gene gain, loss, and rearrangement. PLoS One 2010, 5: e11147.

Denton M, Kerr KG: Microbiological and clinical aspects of infection associated with *Stenotrophomonas maltophilia.* Clin Microbiol Rev 1998, 11:57-80

Di Bonaventura G, Prosseda G, Del Chierico F, Cannavacciuolo S, Cipriani P, Petrucca A, Superti F, Ammendolia MG, Concato C, Fiscarelli E, Casalino M,

Piccolomini R, Nicoletti M, Colonna B: Molecular characterization of virulence determinants of *Stenotrophomonas maltophilia* strains isolated from patients affected by cystic fibrosis. Int J Immunopathol Pharmacol 2007, 20:529-537.

Diancourt L, Passet V, Nemec A, Dijkshoorn L, Brisse S: The population structure of *Acinetobacter baumannii*: expanding multiresistant clones from an ancestral susceptible genetic pool. PLoS One 2010, 5:e10034.

Díaz E, Ferrández A, Prieto MA, García JL: Biodegradation of aromatic compounds by *Escherichia coli*. Microbiol Mol Biol Rev 2001, 65:523-569.

Dijkshoorn L, Nemec A, Seifert H: An increasing threat in hospitals: multidrugresistant *Acinetobacter baumannii*. Nat Rev Microbiol 2007, 5:939-951.

Dobrindt U, Hochhut B, Hentschel U, Hacker J: Genomic islands in pathogenic and environmental microorganisms. Nat Rev Microbiol 2004, 2:414-424

Durante-Mangoni E, Zarrilli R: Global spread of drug-resistant *Acinetobacter baumannii*: molecular epidemiology and management of antimicrobial resistance. Future Microbiol 2011, 6:407-422.

Eppinger M, Rosovitz MJ, Fricke WF, Rasko DA, Kokorina G, Fayolle C, Lindler LE, Carniel E, Ravel J: The complete genome sequence of *Yersinia pseudotuberculosis* IP31758, the causative agent of Far East scarlet-like fever. PLoS Genet 2007, 3:e142.

Frankenberg-Dinkel N: Bacterial heme oxygenases. Antioxidants & Redox Signaling 2004, 6:825-834.

Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I: VISTA, computational tools for comparative genomics Nucleic Acids Res 2004, 32:W273-279.

Gaddy JA, Tomaras AP, Actis LA: The *Acinetobacter baumannii* 19606 OmpA protein plays a role in biofilm formation on abiotic surfaces and in the interaction of this pathogen with eukaryotic cells. Infect Immun 2009, 77:3150-3160.

Genschik P, Drabikowski K, Filipowicz W: Characterization of the *Escherichia coli* RNA 3'-terminal phosphate cyclase and its sigma54-regulated operon. J Biol Chem 1998, 273:25516-25526.

Godde JS, Bickerton A: The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. J Mol Evol 2006, 62:718-729.

Gomelsky M, Klug G: BLUF, a novel FAD-binding domain involved in sensory transduction in microorganisms Trends Biochem Sci 2002, 27:497-500

Hendrix RW, Lawrence JG, Hatfull GF, Casjens S: The origins and ongoing evolution of viruses. Trends Microbiol 2000, 8:504-508.

Hogg JS, Hu FZ, Janto B, Boissy R, Hayes J, Keefe R, Post JC, Ehrlich GD. Characterization and modeling of the *Haemophilus influenzae* core and supragenomes based on the complete genomic sequences of Rd and 12 clinical nontypeable strains. Genome Biol. 2007, 8:R103

Horvath P, Barrangou R: CRISPR/Cas, the immune system of Bacteria and Archaea. Science 2010, 327:167-170.

Iacono M, Villa L, Fortini D, Bordoni R, Imperi F, Bonnal RJ, SicheritzPonten T, De Bellis G, Visca P, Cassone A, Carattoli A: Whole-genome pyrosequencing of an epidemic multidrug-resistant *Acinetobacter baumannii* strain belonging to the European clone II group. Antimicrob Agents Chemother 2008, 52:2616-2625.

Jackson DW, Suzuki K, Oakford L, Simecka JW, Hart ME, Romeo T: Biofilm formation and dispersal under the influence of the global regulator CsrA of *Escherichia coli*. Journal of Bacteriology 2002, 184:290-301

Jacobs AC, Hood I, Boyd KL, Olson PD, Morrison JM, Carson S, Sayood K, Iwen PC, Skaar EP, Dunman PM: Inactivation of phospholipase D diminishes *Acinetobacter baumannii* pathogenesis. Infect Immun 2010, 78:1952-1962.

Juhas M, van der Meer JR, Gaillard M, Harding RM, Hood DW, Crook DW. Genomic: islands: tools of bacterial horizontal gene transfer and evolution. FEMS Microbiol Rev. 2009, 33:376-93.

Kahnert A, Vermeij P, Wietek C, James P, Leisinger T, Kertesz MA: The ssu locus plays a key role in organosulfur metabolism in *Pseudomonas putida* S-313 J Bacteriol 2000, 182:2869-2878

Kajava AV, Cheng N, Cleaver R, Kessel M, Simon MN, Willery E, Jacob-Dubuisson F, Locht C, Steven AC: Beta-helix model for the filamentous haemagglutinin adhesin of *Bordetella pertussis* and related bacterial secretory proteins Mol Microbiol 2001, 42:279-292.

Lee JS, Choi CH, Kim JW, Lee JC: *Acinetobacter baumannii* outer membrane protein A induces dendritic cell death. J Microbiol 2010, 48:387-392.

Locht C, Bertin P, Menozzi FD, Renauld G: The filamentous haemagglutinin, a multifaceted adhesion produced by virulent *Bordetella* spp. Mol Microbiol 1993, 9:653-660.

Looney WJ: Role of *Stenotrophomonas maltophilia* in hospital-acquired infection Br J Biomed Sci 2005, 62:145-154.

Lucas S, Copeland A, Lapidus A, Glavina del Rio T, Dalin E, Tice H, Pitluck S, Chain P, Malfatti S, Shin M, Vergez L, Lang D, Schmutz J, Larimer F, Land M, Hauser L, Kyrpides N, Mikhailova N, Taghavi S, Monchy S, Newman L, Vangronsveld J, van der Lelie D, Richardson P ;"Complete sequence of *Stenotrophomonas maltophilia* R551-3"; Submitted (JUN-2008) to the EMBL/GenBank/DDBJ databases)

Lukjancenko O, Wassenaar TM, Ussery DW. Comparison of 61 sequenced *Escherichia coli* genomes. Microb Ecol. 2010, 60:708-20.

Mantri Y, Williams KP: Islander, a database of integrative islands in prokaryotic genomes, the associated integrases and their DNA site specificities Nucleic Acids Res 2004, 32:D55-58.

Myers GS, Rasko DA, Cheung J K, Ravel J, Seshadri R, DeBoy RT, Ren Q, Varga J, Awad MM, Brinkac LM, Daugherty SC, Haft DH, Dodson RJ, Madupu R, Nelson WC, Rosovitz MJ, Sullivan SA, Khouri H, Dimitrov GI, Watkins KL, Mulligan S, Benton J, Radune D, Fisher DJ, Atkins HS, Hiscox T, Jost BH, Billington SJ, Songer, JG, McClane, BA, Titball, RW, Rood, JI, Melville, SB, Paulsen, IT:Skewed genomic variability in strains of the

toxigenic bacterial pathogen, *Clostridium perfringens* Genome Res 2006, 16:1031-1040.

Nakayama, K, Kanaya, S, Ohnishi, M, Terawaki, Y, Hayashi, T:The complete nucleotide sequence of phi CTX, a cytotoxin-converting phage of *Pseudomonas aeruginosa*, implications for phage evolution and horizontal gene transfer via bacteriophages Mol Microbiol 1999, 31:399-419. |

Nishiyama, M, Ishikawa, T, Rechsteiner, H, Glockshuber, R: Reconstitution of pilus assembly reveals a bacterial outer membrane catalyst Science 2008, 320:376-379.

Nuccio SP, Bäumler AJ: Evolution of the chaperone/usher assembly pathway: fimbrial classification goes Greek. Microbiol Mol Biol Rev 2007, 71:551-575.

Page D, Rose J, Conrod S, Cuine S, Carrier P, Heulin T, Achouak W: Heavy metal tolerance in *Stenotrophomonas maltophilia* PLoS ONE 2008, 3:e1539.

Panne D, Müller SA, Wirtz S, Engel A, Bickle TA, The McrBC restriction endonuclease assembles into a ring structure in the presence of G nucleotides EMBO J 2001, 20:3210-3217.

Patil PB, Bogdanove AJ, Sonti RV, The role of horizontal transfer in the evolution of a highly variable lipopolysaccharide biosynthesis locus in *xanthomonads* that infect rice citrus and crucifers BMC Evol Biol 2007, 7:243.

Permina EA, Mironov AA, Gelfand MS: Damage-repair error-prone polymerases of eubacteria: association with mobile genome elements. Gene 2002, 293:133-140.

Post V, White PA, Hall R: Evolution of AbaR-type genomic resistance islands in multiply antibiotic-resistant *Acinetobacter baumannii*. J Antimicrob Chemother 2010, 65:1162-1170.

Pukatzki S, Ma AT, Revel AT, Sturtevant D, Mekalanos JJ: Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. Proc Natl Acad Sci USA 2007, 104:15508-15513.

Redko Y, de la Sierra-Gallay , Condon C: When all's zed and done: the structure and function of RNase Z in prokaryotes. Nature Rev Microbiol 2007, 5:278-286.

Rocco F, De Gregorio E, Colonna B, Di Nocera PP: *Stenotrophomonas maltophilia* genomes: a start-up comparison. Int J Med Microbiol 2009, 299:535-546.

Russo TA, Luke NR, Beanan JM, Olson R, Sauberan SL, MacDonald U, Schultz LW, Umland TC, Campagnari AA: The K1 capsular polysaccharide of *Acinetobacter baumannii* strain 307-0294 is a major virulence factor. Infect Immun 2010, 78:3993-4000.

Russo TA, MacDonald U, Beanan JM, Olson R, MacDonald IJ, Sauberan SL, Luke NR, Schultz LW, Umland TC: Penicillin-binding protein 7/8 contributes to the survival of *Acinetobacter baumannii* in vitro and in vivo. J Infect Dis 2009, 199:513-521.

Safdar A, Rolston KV, *Stenotrophomonas maltophilia* changing spectrum of a serious bacterial pathogen in patients with cancer Clin Infect Dis 2007, 45:1602-1609.

Sambou T, Dinadayala P, Stadthagen G, Barilone N, Bordat Y, Constant P, Levillain F, Neyrolles O, Gicquel B, Lemassu A, Daffé, M, Jackson M, Capsular glucan and intracellular glycogen of *Mycobacterium tuberculosis* biosynthesis and impact on the persistence in mice Mol Microbiol 2008, 70:762-774.

Smith MG, Gianoulis TA, Pukatzki S, Mekalanos JJ, Ornston LN, Gerstein M, Snyder M: New insights into *Acinetobacter baumannii* pathogenesis revealed by high-density pyrosequencing and transposon mutagenesis. Genes Dev 2007, 21:601-614.

Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, Deboy RT, Davidsen TM, Mora M, Scarselli M, Margarit y Ros I, Peterson JD, Hauser CR, Sundaram JP, Nelson WC, Madupu R, Brinkac LM, Dodson RJ, Rosovitz MJ, Sullivan SA, Daugherty SC, Haft DH, Selengut J, Gwinn ML, Zhou L, Zafar N, Khouri H, Radune D, Dimitrov G, Watkins K, O'ConnorKJ, Smith S, Utterback TR, White O, Rubens CE, Grandi G, Madoff LC, Kasper DL, Telford JL, Wessels MR, Rappuoli R, Fraser CM. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". Proc Natl Acad Sci U S A. 2005, 102:13950-5.

Tomaras AP, Dorsey CW, Edelmann RE, Actis L: Attachment to and biofilm formation on abiotic surfaces by *Acinetobacter baumannii*: involvement of a novel chaperone-usher pili assembly system. Microbiology 2003, 149:3473-3484.

Vallenet D, Nordmann P, Barbe V, Poirel L, Mangenot S, Bataille E, Dossat C, Gas S, Kreimeyer A, Lenoble P, Oztas S, Poulain J, Segurens B, Robert C, Abergel C, Claverie JM, Raoult D, Médigue C, Weissenbach J, Cruveiller S:

Comparative analysis of Acinetobacters: three genomes for three lifestyles. PLos ONE 2008, 3:e1805.

van der Ploeg JR, Iwanicka-Nowicka R, Bykowski T, Hryniewicz M, Leisinger T: The *Escherichia coli* ssuEADCB gene cluster is required for the utilization of sulfur from aliphatic sulfonates and is regulated by the transcriptional activator Cbl. J Biol Chem 1999, 274:29358-29365.

Waters VJ, Gómez M I, Soong G, Amin S, Ernst R, Prince A, Immunostimulatory properties of the emerging pathogen *Stenotrophomonas maltophilia* Infect Immun 2007, 75:1698-1703.

Williams KP: Traffic at the tmRNA gene J Bacteriol 2003, 185:1059-1070.

Xu Y, Chen M, Zhang W, Lin M: Genetic organization of genes encoding phenol hydroxylase, benzoate 1,2-dioxygenase alpha subunit and its regulatory proteins in *Acinetobacter calcoaceticus* PHEA-2. Curr Microbiol 2003, 46:235-240.

Yan A, Guan Z, Raetz C R: An undecaprenyl phosphate-aminoarabinose flippase required for polymyxin resistance in *Escherichia coli* J Biol Chem 2007, 282:36077-36089.

Zarrilli R, Giannouli M, Rocco F, Loman NJ, Haines AS, Constantinidou C, Pallen MJ, Triassi M, Di Nocera PP: Genome sequences of three *Acinetobacter baumannii* strains assigned to ST2, ST25 and ST78 multilocus sequencing typing genotypes. J Bacteriol 2011, 193:2359-2360.

Zhang YM, Rock CO: Membrane lipid homeostasis in bacteria. Nature Rev Microb 2008, 6:222-232.

Zhou NY, Fuenmayor SL, Williams PA: nag genes of Ralstonia (formerly *Pseudomonas*) sp. strain U2 encoding enzymes for gentisate catabolism. J Bacteriol 2001, 183:700-708.

Zimbler DL, Penwell WF, Gaddy JA, Menke SM, Tomaras AP, Connerly PL, Actis LA: Iron acquisition functions expressed by the human pathogen *Acinetobacter baumannii*. Biometals 2009, 22:23-32.

Available online at www.sciencedirect.com
ScienceDirect
International Journal of Medical Microbiology 299 (2009) 535–546
IJMM
www.elsevier.de/ijmm

Stenotrophomonas maltophilia genomes: A start-up comparison

Francesco Rocco[a,1], Eliana De Gregorio[a,1], Bianca Colonna[b], Pier Paolo Di Nocera[a,*]

[a] *Dipartimento di Biologia e Patologia Cellulare e Molecolare, Università Federico II, Via S. Pansini 5, 80131 Napoli, Italy*
[b] *Istituto Pasteur-Fondazione Cenci Bolognetti, Dipartimento di Biologia Cellulare e dello Sviluppo Università Roma La Sapienza, Via dei Sardi 70 Roma, Italy*

Received 25 March 2009; received in revised form 6 May 2009; accepted 21 May 2009

Abstract

The whole DNA sequences of 2 *Stenotrophomonas maltophilia* strains isolated from the blood of a cancer patient (K279a) and the poplar *Populus trichocarpa* (R551-3) have been compared. The 2 chromosomes exhibit extensive synteny, but each is punctuated by about 40 genomic islands (GEIs), which vary in size from 3 to 70 kb, and may encode up to about 50 proteins. A large set of smaller DNA sequences, encoding strain-specific 'solo' orfs, contributes to genetic heterogeneity in a significant manner. *S. maltophilia* GEIs potentially encode several proteins mediating interactions with the environment such as transmembrane proteins, haemagglutinins, components of type I and IV secretion systems, and efflux proteins having a role in metal and/or drug resistance. The presence of specific GEIs in the *S. maltophilia* population was monitored by PCR and slot-blot analyses. Data suggest that some islands are present at sites different from those identified in K279a and that alternative islands may be integrated at mapped sites.

Keywords: Genomic islands; Horizontal gene transfer

Introduction

Stenotrophomonas maltophilia is an aerobic, non-fermentative Gram-negative bacterium widespread in the environment. This species constitutes one of the dominant rhizosphere inhabitants. *S. maltophilia* is able to degrade xenobiotic compounds, as detoxify high molecular weight polycyclic aromatic hydrocarbons, and is therefore considered with interest for its potential use in bioremediation (Antonioli et al., 2007; Page et al., 2008). *S. maltophilia* is also increasingly described as an important nosocomial pathogen in debilitated and immunodeficient patients and has been associated with a broad spectrum of clinical syndromes, e.g. bacteraemia, endocarditis, respiratory tract infections (Senol, 2004; Looney, 2005). Over the last decade, *S. maltophilia* has been frequently isolated from cystic fibrosis patients (Denton and Kerr, 1998; Di Bonaventura et al., 2007), and turned out to be a serious pathogen in cancer patients (Safdar and Rolston, 2007). *S. maltophilia* displays intrinsic resistance to many antibiotics, making selection of optimal therapy difficult. Several factors confer *S. maltophilia* a role as emerging pathogen, most notably the ability to elaborate a wide range of extracellular enzymes, such as lipases, fibrolysin, and proteases, potentially involved in the colonization process (Denton and Kerr, 1998), the ability to adhere to and form biofilm on epithelial cells (Di Bonaventura

*Corresponding author. Tel.: +39 081 7462059; fax: +39 081 7703285.
E-mail address: dinocera@unina.it (P.P. Di Nocera).
[1] The first two authors contributed equally to this work and should be regarded as joint first authors.

1438-4221/$ - see front matter © 2009 Elsevier GmbH. All rights reserved.
doi:10.1016/j.ijmm.2009.05.004

et al., 2007), and the ability to stimulate factors involved in the inflammatory process (Waters et al., 2007).

The complete nucleotide sequence of the genome of K279a, a *S. maltophilia* strain isolated from the blood of a cancer patient, has been recently determined (Crossman et al., 2008). The analysis of K279a genome has allowed to correlate the typical multi-drug resistance (MDR) of the bacterium to the occurrence of several resistance-nodulation-division (RND) efflux pumps, as well as to the presence of enzymes playing a role in drug resistance, among which 2 beta-lactamases and various aminoglycoside phosphotransferases (Crossman et al., 2008). *Xanthomonas* are the closest sequenced relatives of *S. maltophilia* (Palleroni and Bradbury, 1993). Comparing K279a and *X. campestris* 8004 DNA sequences revealed that several genes encoding transport or adhesion proteins are missing in *X. campestris* 8004, but failed to identify the acquisition or the loss of large genetic islands (Crossman et al., 2008).

The whole nucleotide sequence of R551-3, a *S. maltophilia* strain isolated from the poplar *Populus trichocarpa*, has been recently completed at the DOE Joint Genome Institute (Lucas et al., 2008), and its genome has been compared to the genomes of other endophytic bacteria with the aim to start deciphering the mechanisms which underlie the promotion of plant growth (Taghavi et al., 2009). Evaluating peculiarities of the genetic organization of *Stenotrophomonas* chromosomes is an essential step to shed light on unexplored aspects of *S. maltophilia*, with special emphasis on the possible correlation between pathogenic traits and specific genes or gene clusters. A close look at the *S. maltophilia* K279a genome has recently led us to set up a rapid PCR-based typing protocol for an efficient classification of clinical *S. maltophilia* isolates (Roscetto et al., 2008). In this manuscript, results emerging from various comparisons of the K279a and R551-3 DNA are reported.

Materials and methods

In silico analyses

The whole genomes of the *S. maltophilia* strains K279a (EMBL/GeneBank access no. AM743169) and R551-3 (NCBI, locus CP001111) were aligned by using mVISTA, a set of programs for comparing DNA sequences from 2 or more species and visualizing the obtained alignments (Frazer et al., 2004; http://genome.lbl.gov/vista/index.shtml). EMBL/GeneBank and NCBI gene annotations were integrated with annotations available at the KEGG resource (Kyoto Encyclopedia of Genes and Genomes, http://www.genome.jp/kegg). *S. maltophilia* IS families were defined on the basis of BLAST analyses carried out against sequences deposited at the IS database (www.is.biotoul.fr/is.html).

Bacterial strains and genomic DNA extraction

The *S. maltophilia* clinical strains used in this study were obtained from the Bambino Gesù Hospital (Rome, Italy) and from the University Federico II Hospital (Naples, Italy) and have been described previously (Di Bonaventura et al., 2007; Roscetto et al., 2008). K279a strain (Crossman et al., 2008) was a kind gift of M.B. Avison (University of Bristol, UK), while the LMG strains were purchased at the LMG/BMCC collection. The *S. maltophilia* strains analyzed in this study are listed in Supplemental Table 1. All strains were routinely grown in brain heart infusion at 37 °C, except for the environmental strains (LMG959, LMG10871, LMG10879, LMG11104, and LMG11108) which were grown at 30 °C. Genomic DNA was extracted as described (De Gregorio et al., 2006).

PCR amplifications. PCR reactions were carried out by incubating 20 ng of DNA with 160 ng of each primer in the presence of dXTPs (200 nM), 1.5 mM $MgCl_2$, and the Taq DNA polymerase Recombinant (Invitrogen). Samples were incubated first at 95 °C for 5 min. The amplification programme included 1 min at 95 °C, 1 min at the annealing temperature, and 1 min at 72 °C for a total of 30 cycles. At the end of the cycle, samples were kept at 72 °C for 7 min before harvesting. PCR products were electrophoresed on 1.5–2% agarose gels in 0.5X TBE buffer (45 mM Tris pH 8, 45 mM borate, 0.5 mM EDTA) at 120 V (constant voltage). The 100-bp ladder (Fermentas) was used as molecular weight marker. Primers measured all 25 nt, and annealing temperatures varied from 57 to 64 °C. Primer sequences are available upon request.

Slot-blot hybridizations. One microgram of DNA from each strain was loaded onto Hybond filters and cross-linked by UV treatment as described (Venditti et al., 2007). The filters were hybridized to 32P-radiolabelled PCR products amplified from orfs belonging to specific GEIs.

Results

The chromosomes of *S. maltophilia* K279a and R551-3 strains have the same GC content (67%), but differ significantly in length, since K279a DNA measures 4,851,126 bp, and R551-3 DNA 4,573,969 bp. However, the 2 chromosomes share a common scaffold, as they are widely collinear, and can be easily aligned throughout their lengths (Fig. 1). Differences between the genomes of the 2 strains are correlated primarily to specific DNA sequences present exclusively in K279a. Accordingly, the number of potential gene products is higher in K279a (4386 annotated orfs) than in R551-3 (4041 annotated orfs). Nonetheless, several DNA segments are unique to the R551-3 genome. Sequence

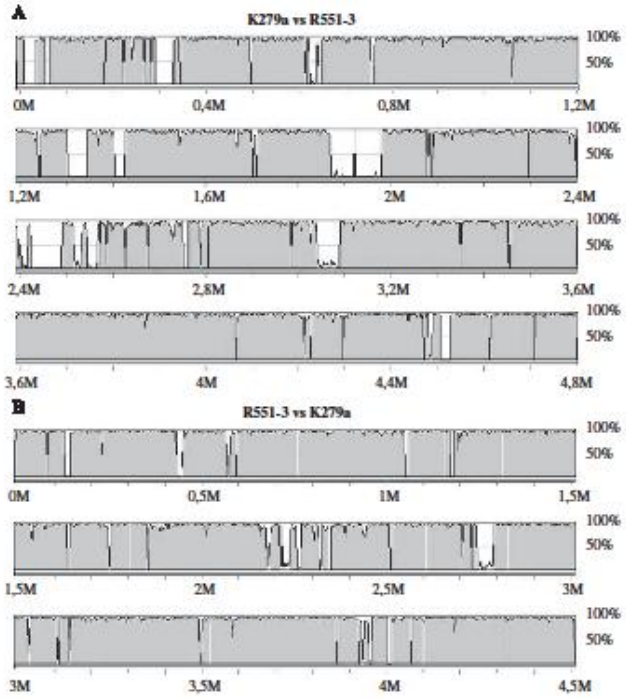

Fig. 1. Alignment of K279a and R551-3 genomes. VISTA plots showing similarity of K279a vs. R551-3 (A) and R551-3 vs. K279a (B) DNAs. The percent identity of the aligned genomes is shown.

alignments allowed to build up a comparative map of *S. maltophilia* orfs, and to easily identify strain-specific gene products. Orfs annotated as hypothetical proteins shorter than 100 aminoacids were not considered for comparisons.

Large DNA insertions

Sequence alignments allowed to identify about 200 regions in which shared chromosome synteny is interrupted in one strain, or in both. A major source of variation between the 2 *S. maltophilia* chromosomes is represented by a multitude of strain-specific genomic islands (GEIs), ranging in size from approximately 3 to 70 kb. Routinely, islands denote chromosomal DNA regions, plausibly acquired by horizontal gene transfer (HGT), larger than 10 kb, smaller regions being referred to as islets (Hacker and Carniel, 2001; Dobrindt et al., 2004). For sake of simplicity, as already done in genome comparisons (see Myers et al., 2006), all DNA segments >3 kb present in only one of the 2 strains were referred to as islands. *S. maltophilia* GEIs were marked by a K or R to designate K279a or R551-3 DNA, respectively, and numbered progressively according to chromosome position. Fourty-one GEIs have been identified in K279a, and 36 in R551-3 DNA. Size, orfs, GC content, and chromosomal location of all GEIs are shown in Fig. 2; chromosome coordinates and orf products are reported in Supplemental Table 2. In both strains, half of GEIs measure less than 5kb. Of the remaining islands, most measure more than 15 kb in K279a, but only 4 exceed such size in R551-3. On the whole, K and R GEIs constitute 12.1% and 6.6% of the genome and encode 597 and 249 orfs, respectively. GEIs often have a GC content different from bulk chromosomal value (Dobrindt et al., 2004). The GC content of many GEIs is lower than the average 66.7% value of *S. maltophilia* DNA. Values range between 63 and 66% in 16 K-GEIs and 15 R-GEIs and are lower than 63% in 16 K-GEIs and 7 R-GEIs. Nine islands were inserted at the same relative chromosomal position in the 2 strains (Fig. 2).

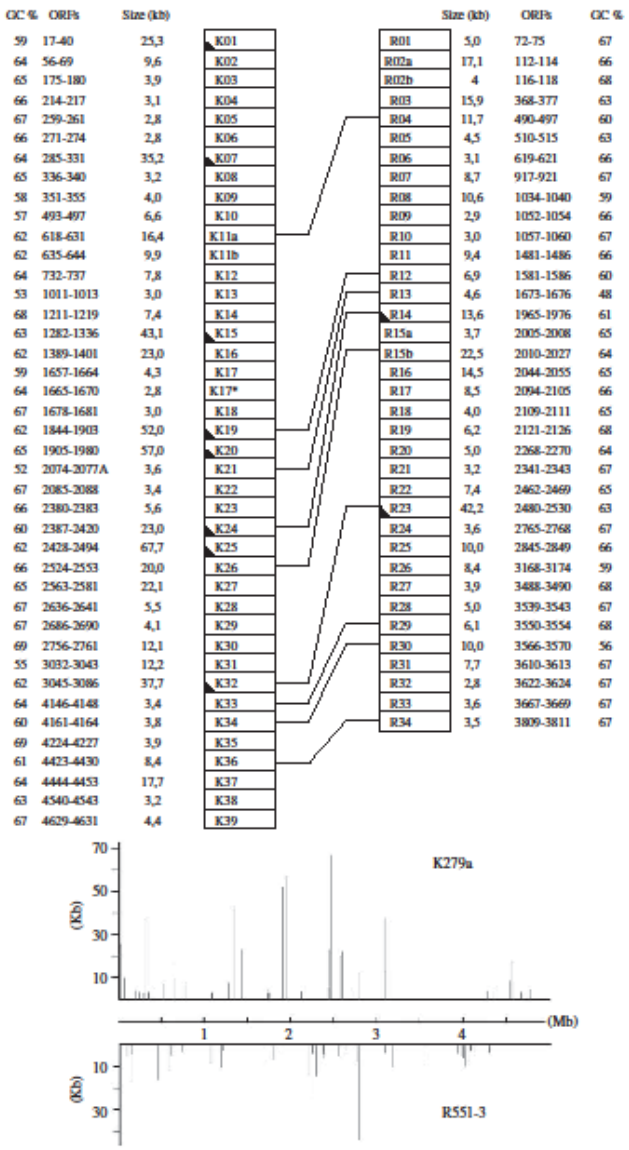

Fig. 2. K279a and R551-3 GEIs. Size, orfs interval and G + C content of each GEI are shown. Islands inserted at the same relative chromosomal position in the 2 strains are connected by lines. Islands potentially encoding one or more integrases are marked by a triangle. The chromosomal distribution of GEIs is diagrammed at the bottom. The height of the lines denotes GEI size.

At least in one case, this is correlated to the use of the same chromosomal entry site.

GEIs boundaries

Sequence alignments provided unequivocal information on the ends of GEIs in a few cases only. Consequently, the length of most GEIs is underestimated, the start/stop codons of terminal orfs arbitrarily functioning as provisional GEI ends (see Supplemental Table 2). However, the ends of 8 GEIs were exactly defined by sequence comparisons. Relatively to the mode of integration, the 8 GEIs can be sorted into 2 groups (Fig. 3). The first one includes K01, K02, and R19, which are not flanked by duplication of bases at the insertion site. The second group includes K07, K17, K20, K32, and R23. All these GEIs are integrated within tRNA or tm-RNA sequences, which are a preferential GEI target (Williams, 2002; Mantri and Williams, 2004). The termini of the 5 GEIs are flanked by target site duplications (TSDs) 14–53 bp long, in which segments of both target tRNA/tm-RNA and flanking DNA are duplicated. TSDs flank K20, but not the adjacent K19 island, which plausibly also used the tm-RNA as insertion target. The lack of TSDs has been observed for GEIs inserted at the tm-RNA locus (Williams, 2003).

Many large islands potentially encode integrases or functionally related enzymes. Both K32 and R23 are inserted at the serine tRNA (Fig. 3), but encode non-homologous integrases (orfs 3086 and 2480, respectively). K24 and R14 are inserted at the same relative chromosomal position. While actual boundaries have not been defined, it is plausible that also the integration of these 2 GEIs had been mediated by non-homologous proteins. These may correspond to one of the 2 gene products (orfs 2416 and 2417) encoded by Sm3, an insertion sequence located at one terminus of K24, and one of the 2 R14 integrases (orfs 1971 and 1972).

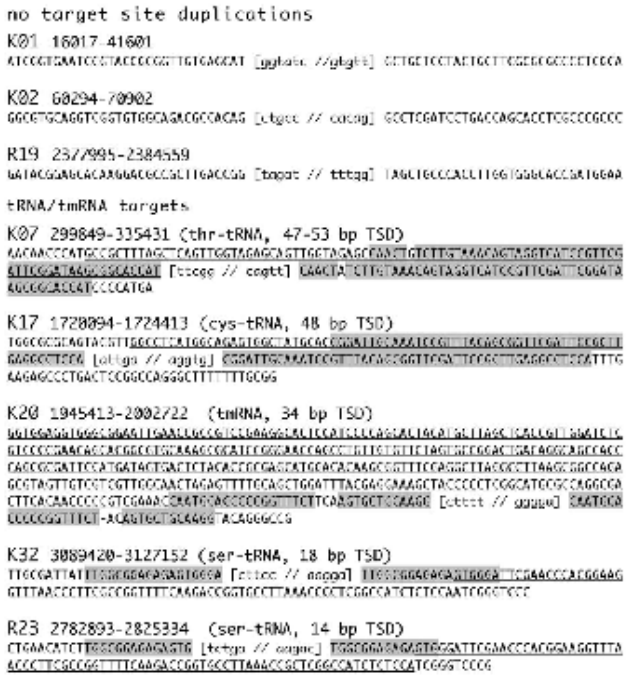

Fig. 3. GEIs target sites. Sequences at the boundaries of specific GEIs are shown. GEI chromosomal coordinates and TSD lengths are shown. Islands are in brackets, and only the 5 terminal nucleotides at either side are shown in lowercase letters. tRNA and tm-RNA targets are underlined, TSDs are highlighted.

GEIs gene products

The content of GEIs is highly variable. About 1/2 of island-encoded proteins to which a function could be assigned is represented by molecules mediating interactions with the environment. Gene products of interest encoded by GEIs are listed in Table 1 and are briefly discussed here.

(i) **Metal resistance genes.** *S. maltophilia* hosts several operons involved in import, storage, and efflux of metals. Some of these gene clusters are conserved at the same relative chromosomal position in the 2 strains, others are found in K279a only, on specific GEIs. As summarized in Table 1, K01 and K25 carry *czc* (cobalt–zinc–cadmium resistance) genes, K25 carries also genes involved in copper metabolism and homoeostasis (*cop* and *cus* operons), and K03 arsenic-resistance genes (*ars* operon). Mercury resistance genes are present in K279a but not in R551-3, and are located on K24.

(ii) **T1SS (type I secretion system).** Type I pili consist of a rod composed by a major fimbrial protein and 2 or more ancillary proteins. Such structures are assembled and secreted by the chaperon/usher pathway (Nishiyama et al., 2008). T1SS genes are conserved at 2 sites in K279a and R551-3. A third, K279a-specific cluster, is encoded by K12.

(iii) **T4SS (type IV secretion system).** T4SS genes conserved in K279a and R551-3 are in the same order (*vir*D4, B8, B9, B10, B11, B1, B2, B3, B4, B6) of *X. axonopodis* T4SS genes (da Silva et al., 2002; Alegria et al., 2005). A second, K279a-specific T4SS cluster is encoded by K15. Here, genes are in the same order (vir D4, B11, B2, B3, B4, B5, B6, B8, B9, B10) of T4SS genes encoded by the PA7 strain of *Pseudomonas aeruginosa* (orfs 3708–3697). The relatedness of the 2 T4SS gene clusters is reinforced by the observation that the hypothetical proteins which separate VirB11 and VirD4 encoded by PA7 (orf 3707) and K279a (orf l292) genomes are homologous.

(iv) **Filamentous haemagglutinin (FHA) genes.** FHA is a major colonization factor in *Bordetella pertussis* (Locht et al., 1993), and FHA proteins have been found in species as diverse as *Clostridium perfringens* (Myers et al., 2006) and *Moraxella catarrhalis* (Balder et al., 2007). FHAs are encoded along with transporter proteins (Jacob-Dubuisson et al., 2001). Three *fha* and related transporter genes are found in *S. maltophilia*, and all are encoded by GEIs (K16, K37, R02a). The 3 FHAs vary in size, and this is correlated, as shown for other FHAs (Kajava et al., 2001), to the number of repeats fitting the consensus $LDNGGGX_{13–22}$.

(v) **LPS genes.** Variation at the interstrain level is common in LPS biosynthetic gene clusters inserted between the conserved *met*B and *etf*A loci in bacteria (Patil et al., 2007). Alternative LPS gene sets, both carried by GEIs, are found in K279a and R551-3. K11 includes 2 gene clusters, K11a and K11b. The former contains genes playing a role in the O-chain synthesis found also in R04. K11b contains 7 genes known as *pmr*IHFJLMK or

Table 1. GEI potential gene products.

Genes	Chromosome orfs		GEI orfs	
	K279a	R551-3	K279a	R551-3
czc	2697–2699 4606–4608	2169–2171 3956–3958	K01 (0036–0038), K25 (2457–2461)	–
cop	3691b2–3692 2691–2692	3105–3106 2163–2164	K25 (2440–2449)	–
cus	2693–2694	2165–2166	K25 (2433–2434)	–
ars	2421–2425	1977–1981	K03 (0176–0179)	–
mer	–	–	K24 (2409–2412)	–
T1SS	0706–0709 1508–1512	0561–0564 1267–1271	K12 (0732–0736)	–
T4SS	2997–3008	2439–2451	K15 (1283–1293)	–
fha	–	–	K16 (1389–1390) K37 (4452–4453)	R02a (0112–0113)
LPS	–	–	K11a (0618–0631) K11b (0636–0642)	R04 (0490–0497)
mcrBC	–	–	K10 (0497)	–
ssu	–	–	K27 (2572–2574)	–
BLUF	–	–	K26 (2528, 2535, 2541)	–
glg	–	–	K30 (2756–2761)	–

*arn*ABCDEFT, encoding proteins which act in a coordinate manner to ultimately modify the lipid A by the addition of 4-amino-4-deoxy-L-arabinose. This modification causes resistance to polymyxin and cationic antimicrobial peptides (Yan et al., 2007). The genes separating K11a and K11b (orfs 632–634) are homologous to genes located downstream from R04 (orfs 498–500). This suggests that K11 derives from R04 by the acquisition of K11b genes.

Additional proteins of interest potentially encoded by genomic islands are the homologue of McrBC (K10), a restriction endonuclease which cuts DNA containing modified cytosines (Panne et al., 2001), the Ssu proteins (K27), involved in organosulfur metabolism (Kahnert et al., 2000), several hypothetical proteins, all encoded by K26, which feature the BLUF (for sensors of blue-light using FAD) domain, a novel FAD-binding domain plausibly involved in sensory transduction in microorganisms (Gomelsky and Klug, 2002). Finally, K30 includes a cluster of genes involved in glycogen biosynthesis and catabolism. The *glg* gene products have a role in the formation of biofilms (Jackson et al., 2002) and capsular layers (Sambou et al., 2008), and hence may be catalogued as potential pathogenic genetic determinants.

GEIs and prophages

Five GEIs (K02, K07, K19, K20, and R23) encode mostly phage-related products. Homology searches enabled to correlate 2 of these islands to known prophages.

K07. This island encodes several phage-like products, including an integrase (orf 56) which is highly homologous (54% identity) to the analogous protein encoded by Xcc37K, a 37,309-bp island identified in *X. campestris* (Mantri and Williams, 2004). K07 and Xcc37K are comparable in size and are closely related (Fig. 4a). Comparisons carried out at the prophage database (http://bicmku.in:8082/prophagedb) revealed that Xcc37K indeed corresponds to the *X. campestris* P-like prophage PH138. K07 is also highly related to the *P. aeruginosa* cytotoxin-converting P2-like phage CTX (Nakayama et al., 1999). K07 and CTX share a common structural scaffold, but the 2 phages diverge because each carries a specific set of genes at one end. K07 plausibly corresponds to the 'phage II cluster' described by Crossman et al. (2008).

K19. K19 and K20 are closely located GEIs, spanning together a ~109-kb chromosome segment, and plausibly correspond to the region defined as 'phage I cluster' by Crossman et al. (2008). Both GEIs potentially encode phage-like products but can be distinguished as individual entities, because they are separated by the tm-RNA target (residues 1,945,038–1,945,390). K19 exhibited significant homology to a 41-kb prophage-like sequence identified in the *Minibacterium massiliensis* genome (residues 3,029,900–3,070,592, see Audic et al., 2007). K19 and *M. massiliensis* sequences are largely collinear (Fig. 4b), but diverge because K19 carries ~10 kb of non-phage DNA at the 5' end side.

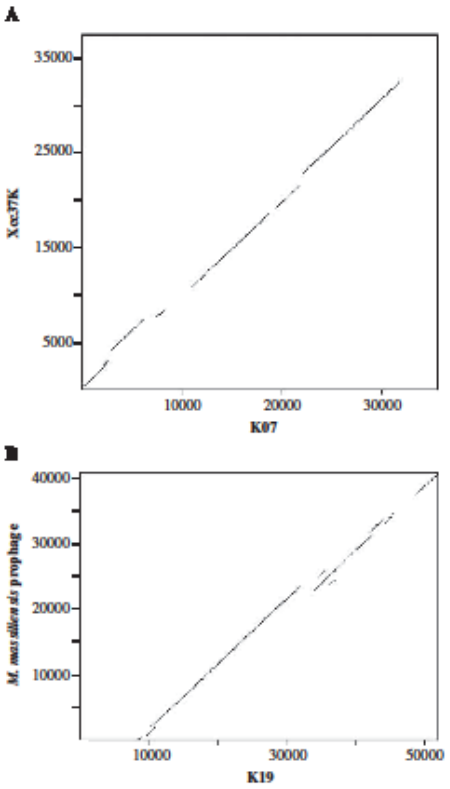

Fig. 4. Phage GEIs. Dot plot alignments of the nucleotide sequences of K07 and Xcc37K islands (panel A), and K19 and a *M. massiliensis* prophage (panel B), are shown. Sizes are in nt.

Insertion sequences

In addition to integrases, GEIs encode a variety of transposases and transposition-related proteins. Members of 3 major IS families (IS3, IS481, and IS110) are present in K279a and R551-3 genomes. Comparative

Table 2. *S. maltophilia* insertion sequences.

Name	Family	Related ISs	Copy number	
			R551-3	K279a
Sm1	IS3	ISXc8 (*X. campestris*)	–	4
Sm2		ISXac3 (*X. axonopodis*)	–	4
Sm3		IS1404 (*X. campestris*)	–	1
Sm4		ISPst9 (*P. stutzeri*)	–	1
Sm5		ISAzo10 (*Azoarcus* sp. EbN1)	–	1
Sm6		ISPsy9 (*P. syringae*)	3	4
Sm7	IS481	ISBcen21 (*B. cenocepacia*)	5	5
Sm8	IS110	ISBcen8 (*B. cenocepacia*)	1	12
Sm9		IS621 (*E. coli* HS)	7	–

analyses, carried out at the ISfinder database, allowed to sort the 32 and 16 complete ISs, respectively, found in K279a and R551-3 into 9 sub-families (Table 2). Some ISs are found in K279a only, remarkably all in GEIs. The 4 Sm1 elements are inserted in K01, K25, K31, and K32, the 4 Sm2 elements in K10, K17, K19, and K25. Sm3, Sm4, and Sm5 are single-copy ISs. The first 2 are found, next to each other, at one end of K24, Sm5 is inserted in K19.

Strain-specific orfs not encoded by GEIs

At a first glance, K279a and R551-3 genomes are collinear throughout and differ because equipped with alternative sets of GEIs. A closer look at the chromosome backbone, however, revealed that the genome of each strain is punctuated by 'solo' orfs, for which homologous potential gene products could not be identified at the corresponding chromosomal position in the other strain (see Supplemental Table 3). 172 and 160 orfs, measuring 1–3 kb, are scattered along the genomes of K279a and R551-3, respectively. Most are single, many come in pairs, a few in small clusters (3–4 orfs) not catalogued as GEIs because of their small size. The origin of these orfs is unclear. Some are found at positions marked in the other strain by GEIs, and may plausibly represent sequences removed upon island insertion. Others are found only in one strain because of mutations affecting homologous sequences in the other strain. Orf 3990 in K279a and orf 3400 in R551-3 are encoded by homologous DNA segments, but differ because translated from initiating GTG and ATG triplets on different frames. About 60% of strain-specific orfs is constituted, in both genomes, by hypothetical proteins. Interestingly, membrane proteins account for 15% of the 'solo' orfs in K279a, but only for 6% in R551-3.

GEIs in the *S. maltophilia* population

To check whether islands identified in K279a are present in the population, the DNAs of 41 *S. maltophilia* strains were analyzed by PCR and dot-blot hybridizations. PCR experiments were designed to test whether 2 islands selected by chance, K01 and K32, occupy the chromosomal site identified in K279a also in other strains. To this end, island/chromosome junctions were amplified with specific L (left) and R (right) in-out primer pairs (Fig. 5A). According to this scheme, products of amplification obtained with primers Lin and Rin denote the absence of the island. Data support the notion that K01 is present in 8 strains at the same position found in K279a DNA (amplimers obtained with both left and right primer pairs) but is either missing or inserted in another location in 21 strains. Sequence analyses confirmed that PCR products obtained with Lin and Rin primers indeed corresponded to chromosomal empty sites. Of the remaining DNAs analyzed, 5 yielded PCR products corresponding to only one island/chromosome junction, 7 could not be amplified by PCR. Different results were obtained for K32. The island is present only in 2 strains, and PCR products corresponding to only one island/chromosome junction were detected in 13 DNAs. In contrast to K01, chromosomal empty sites were not detected, and 26 strains were negative to PCR (Fig. 5A). Sequence heterogeneity may account for the inability to amplify some DNA regions, and also explain why only one island/chromosome junction was detected in some strains. To circumvent the problem, islands were monitored by dot-blot analyses (Fig. 5B). The distribution of GEIs greatly varied among strains, the degree of conservation ranging from zero to 90%. Strains may be tentatively assigned to 3 groups. 13/41 strains contain the largest number (7–10) of tested islands, 21/41 strains contain 1–6 islands, and 7 carry no islands. GEIs were distributed apparently at random. However, it can be noticed that K11, K15, K16, K23, K25, and K32 are over-represented, K07, K19, K20, and K24 are under-represented.

Discrepancies can be noticed by comparing slot blot and PCR data. According to hybridization data, the DNAs of strains 528, OBTGC29, 545, and LMG10851 contain K01 sequences. In contrast, 'empty site' products denoting the absence of K01 were obtained by PCR for all of them. This suggests that K01 is located in the 4 strains at a novel site. For 7 strains negative for K01 at the blot level (1054, OBGTC22, OBGTC28, STM2, LMG10879, LMG11104, and LMG10853), chromosomal empty sites were not detected by PCR. Though not proven, data support the hypothesis that in these strains K01 is replaced by one (or more) alternative GEI(s).

Discussion

The results of whole-genome comparative analyses reported in this work add knowledge to an earlier report

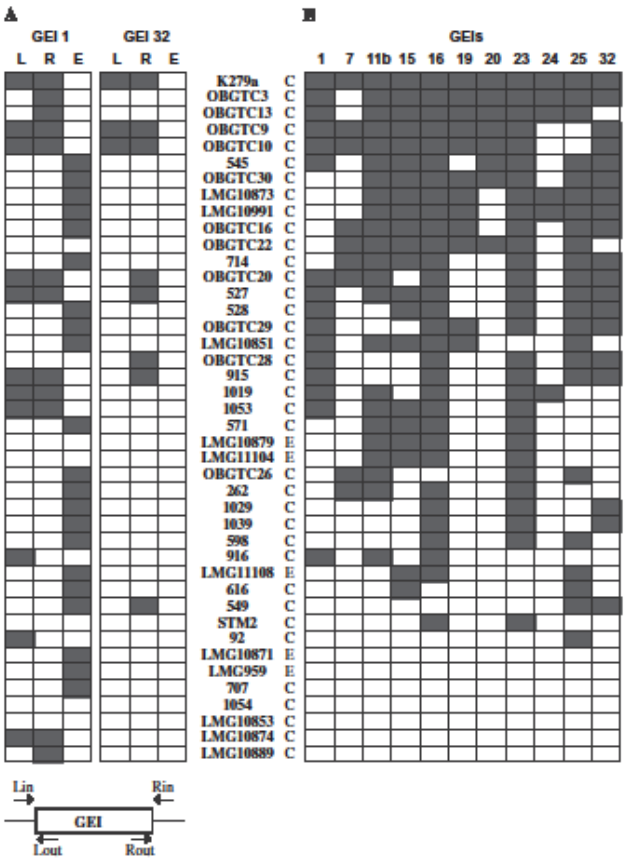

Fig. 5. Distribution of GEIs among *S. maltophilia* isolates. (A) PCR analyses. Chromosome/island junctions of K01 and K32 were amplified with specific Lin-Lout (lane L, left boundary) and Rin-Rout (lane R, right boundary) primer pairs. Products of amplification obtained with primer pairs Lin-Rin denoted the absence of the island (lane E, empty site). (B) Dot-blot analysis. One microgram of total DNA from the indicated strains was hybridized to 32-P radiolabelled, 300–600 bp island-specific DNA probes. Probes were amplified from K279a orfs 32 (K01), 285 (K07), 638 (K11b), 1290 (K15), 1389 (K16), 1901 (K19), 1906 (K20), 2382 (K23), 2412 (K24), 2447 (K25), 3086 (K32). C, clinical isolates; E, environmental isolates.

focused on the organization of K279a DNA (Crossman et al., 2008), and set the base for a start-up definition of the core genome component of *S. maltophilia*. In the light of results obtained by different typing procedures, *S. maltophilia* is hypothesized to be extremely heterogeneous at the genetic level (Hauben et al., 1999; Coenye et al., 2004). Data shown here corroborate in part such opinion. Strains isolated from different environments, which likely epitomize *S. maltophilia* causing infection to men (K279a) or living in the wild (R551-3), share extensive chromosomal synteny, and their DNAs can be easily aligned throughout their lengths. The absence of major rearrangements (loss/acquisition of very large DNA regions, strain-specific dissemination of mobile DNA elements) suggests that the set of pathogenic gene products and/or programmes crucial to infect plants or men may substantially be conserved in *S. maltophilia* cells from different habitats. At once, whole-genome comparisons lighted up a high degree of genetic heterogeneity, and let to define, ~200 sites (for

each strain) in which chromosome collinearity is interrupted.

A main source of variation is represented by the presence/absence of specific GEIs. The chromosome of either strain is punctuated by ~40 GEIs. None is common to K279a and R551-3, but some carry genes having the same function. Large sets of strain-specific sequences smaller than GEIs also contribute to genetic heterogeneity. Altogether, strain-specific orfs represent ~17.5 and 10.1% of the potential gene products of K279a and R551-3, respectively. Accordingly, the number of potential gene products encoded by both strains is approximately 3620. Thus *S. maltophilia* strains are less heterogeneous than *E. coli* strains at the genome level. The pathogenic CFT073 and EDL933 and the apathogenic MG1655 *E. coli* strains share only 39.2% of the proteins (Welch et al., 2002), and strain-specific orfs represent 23.1 and 17.2% of the potential gene products encoded by the uropathogenic 536 and MG1655 strains, respectively (Brzuszkiewicz et al., 2006).

Many GEI proteins mediate interactions with the environment. Some may assist in the colonization and survival of *S. maltophilia* in the host, and contribute to pathogenicity. Several proteins involved in import, storage, and efflux of metals are encoded by K279a islands. Compared to R551-3, K279a has an additional cluster of T1SS (island K12) and T4SS (island K15) genes. By looking at two-partner secretion systems, R551-3 DNA contains one *fha* operon (island R02A), K279a 2 (islands K16 and K37). It is difficult, however, to correlate these findings to the pathogenic potential of K279a. Moreover, proteins having a role in the life cycle of *S. maltophilia* as a pathogen may not be encoded by GEIs. In this respect, it is worth noting that membrane proteins encoded by 'solo' orfs are twice as much abundant in K279a than R551-3.

In contrast to what was observed in many species, ISs contribute to genetic variation of *S. maltophilia* to a limited extent. Many ISs have been likely imported by HGT, because found within or at the border of specific GEIs. Consequently, islands containing Sm1 (K01, K25, K31, K32) and Sm2 (K10, K17, K19, K25) elements may have been acquired from *X. campestris* or *X. axonopodis*, which feature ISs homologous to Sm1 and Sm2 (Table 2). The same holds for K24, which contains Sm4, a IS homologous to the *X. campestris* IS1404 element.

Large GEIs plausibly arose by step-wise acquisition of DNA segments in different species prior to be inserted into the *S. maltophilia* chromosome.

In addition to Sm2, K19 contains a prophage likely imported from *M. massiliensis*. The latter, in turn, is interrupted by Sm5, an insertion element homologous to ISs found in *Azoarcus* sp. The K01 region spanning *czc* and flanking genes (orfs 33–38) is homologous to a strain-specific DNA segment of the *P. aeruginosa* PA14 strain (PA14 orfs 30,980–31,040; see Lee et al., 2006). Based on both gene order and sequence homologies, we suggest that the T4SS genes found in K15 also derive from *P. aeruginosa*.

The degree of conservation of virulence chromosomal regions among *P. aeruginosa* strains recovered from environmental or clinical sources may significantly vary (Finnan et al., 2004). Several *S. maltophilia* isolates differ from K279a for the lack of a few GEIs. OBGTC9 and OBGTC10 lack K24 and K25. The 2 clones are indistinguishable from K279a when tested by MLVA-like analyses, but display a PFGE profile quite different from K279a (Roscetto et al., 2008). These observations suggest that both strains may carry GEIs missing in K279a. Seven strains (LMG959, LMG10871, LMG10853, LMG10874, LMG10889, 707, 1054) were negative to all GEI probes. Of these, the first 2 are environmental, the others are clinical isolates. All, according to different genotyping protocols, belong to different subgroups (Hauben et al., 1999; Roscetto et al., 2008). Thus environmental and clinical isolates may retain, or lack, specific GEIs, and analyses performed do not allow to draw correlations between types of isolates and GEI subsets. DNA amplification and hybridization data suggest that specific islands may be present in some isolates at sites different from those mapped in K279a as well that novel islands may be integrated at known sites.

A deeper knowledge of the genome organization of *S. maltophilia* could contribute to a better understanding of the rapid adaptation of this bacterium to the human host, making it an emergent pathogen in nosocomial infections. It will be of interest to ascertain whether strains isolated from different patient cohorts may fit subgroups displaying specific GEI profiles, and to evaluate whether the genome of *S. maltophilia* clinical isolates has been remodelled by the acquisition of DNA blocks from coinfecting pathogenic bacteria.

Acknowledgments

We thank Emanuela Roscetto and Tullia Bertocco, who performed some of the PCR and hybridization experiments. We are also grateful to M.B. Avison, E. Fiscarelli, and R. Zarilli for making *S. maltophilia* strains available. We are indebted to Jean Gilder for reviewing the manuscript. Research has been supported from grants of the Italian Cystic Fibrosis Research Foundation (FFC) to B.C. and P.P.D.N.

Appendix A. Supporting Information

Supplementary data associated with this article can be found in the online version at: doi:10.1016/j.ijmm.2009.05.004.

References

Alegria, M.C., Souza, D.P., Andrade, M.O., Docena, C., Khater, L., Ramos, C.H., Silva, A.C., Farah, C.S., 2005. Identification of new protein-protein interactions involving the products of the chromosome- and plasmid-encoded type IV secretion loci of the phytopathogen *Xanthomonas axonopodis pv. citri*. J. Bacteriol. 187, 2315–2325.

Antonioli, P., Lampis, S., Chesini, I., Vallini, G., Rinalducci, S., Zolla, L., Righetti, P.G., 2007. *Stenotrophomonas maltophilia* SeITE02, a new bacterial strain suitable for bioremediation of selenite-contaminated environmental matrices. Appl. Environ. Microbiol. 73, 6854–6863.

Audic, S., Robert, C., Campagna, B., Parinello, H., Claverie, J.M., Raoult, D., Drancourt, M., 2007. Genome analysis of *Minibacterium massiliensis* highlights the convergent evolution of water-living bacteria. PLoS Genet. 3, e138.

Balder, R., Hassel, J., Lipski, S., Lafontaine, E.R., 2007. *Moraxella catarrhalis* strain O35E expresses two filamentous hemagglutinin-like proteins that mediate adherence to human epithelial cells. Infect. Immun. 75, 2765–2775.

Brzuszkiewicz, E., Brüggemann, H., Liesegang, H., Emmerth, M., Ölschläger, T., Nagy, G., Albermann, K., Wagner, C., Buchrieser, C., Emody, L., Gottschalk, G., Hacker, J., Dobrindt, U., 2006. How to become a uropathogen: comparative genomic analysis of extraintestinal pathogenic *Escherichia coli* strains. Proc. Natl. Acad. Sci. USA 103, 12879–12884.

Coenye, T., Vanlaere, E., LiPuma, J.J., Vandamme, P., 2004. Identification of genomic groups in the genus *Stenotrophomonas* using *gyr*B RFLP analysis. FEMS Immunol. Med. Microbiol. 40, 181–185.

Crossman, L.C, Gould, V.C., Dow, J.M., Vernikos, G.S., Okazaki, A., Sebaihia, M., Saunders, D., Arrowsmith, C., Carver, T., Peters, N., Adlem, E., Kerhornou, A., Lord, A., Murphy, L., Seeger, K., Squares, R., Rutter, S., Quail, M.A., Rajandream, M.A., Harris, D., Churcher, C., Bentley, S.D., Parkhill, J., Thomson, N.R., Avison, M.B., 2008. The complete genome, comparative and functional analysis of *Stenotrophomonas maltophilia* reveals an organism heavily shielded by drug resistance determinants. Genome Biol. 9, R74.

da Silva, A.C., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R., Quaggio, R.B., Monteiro-Vitorello, C.B., Sluys, M.A., Almeida, N.F., Alves, L.M., Amaral, A.M., Bertolini, M.C., Camargo, L.E., Camarotte, G., Cannavan, F., Cardozo, J., Chambergo, F., Ciapina, L.P., Cicarelli, R.M., Coutinho, L.L., Cursino-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J., Ferreira, R.C., Ferro, M.I., Formighieri, E.F., Franco, M.C., Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite, R.P., Lemos, E.G., Lemos, M.V., Locali, E.C., Machado, M.A., Madeira, A.M., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J., Menck, C.F., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T., Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira, H.A., Rossi, A., Sena, J.A., Silva, C., Souza, R.F., Spinola, L.A., Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I., Trindade dos Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C., Kitajima, J.P., 2002. Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. Nature 417, 459–463.

De Gregorio, E., Silvestro, G., Venditti, R., Carlomagno, M.S., Di Nocera, P.P., 2006. Structural organization and functional properties of miniature DNA insertion sequences in *Yersiniae*. J. Bacteriol. 188, 7876–7884.

Denton, M., Kerr, K.G., 1998. Microbiological and clinical aspects of infection associated with *Stenotrophomonas maltophilia*. Clin. Microbiol. Rev. 11, 57–80.

Di Bonaventura, G., Prosseda, G., Chierico, F., Cannavacciuolo, S., Cipriani, P., Petrucca, A., Superti, F., Ammendolia, M.G., Concato, C., Fiscarelli, E., Casalino, M., Piccolomini, R., Nicoletti, M., Colonna, B., 2007. Molecular characterization of virulence determinants of *Stenotrophomonas maltophilia* strains isolated from patients affected by cystic fibrosis. Int. J. Immunopathol. Pharmacol. 20, 529–537.

Dobrindt, U., Hochhut, B., Hentschel, U., Hacker, J., 2004. Genomic islands in pathogenic and environmental microorganisms. Nat. Rev. Microbiol. 2, 414–424.

Finnan, S., Morrissey, J.P., O'Gara, F., Boyd, E.F., 2004. Genome diversity of *Pseudomonas aeruginosa* isolates from cystic fibrosis patients and the hospital environment. J. Clin. Microbiol. 42, 5783–5792.

Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., Dubchak, I., 2004. VISTA, computational tools for comparative genomics. Nucleic Acids Res. 32, W273–W279.

Gomelsky, M., Klug, G., 2002. BLUF, a novel FAD-binding domain involved in sensory transduction in microorganisms. Trends Biochem. Sci. 27, 497–500.

Hacker, J., Carniel, E., 2001. Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes. EMBO Rep. 2, 376–381.

Hauben, L., Vauterin, L., Moore, E.R.B., Hoste, B., Swings, J., 1999. Genomic diversity of the genus *Stenotrophomonas*. Int. J. Syst. Bacteriol. 49, 1749–1760.

Jackson, D.W., Suzuki, K., Oakford, L., Simecka, J.W., Hart, M.E., Romeo, T., 2002. Biofilm formation and dispersal under the influence of the global regulator CsrA of *Escherichia coli*. J. Bacteriol. 184, 290–301.

Jacob-Dubuisson, F., Locht, C., Antoine, R., 2001. Two-partner secretion in Gram-negative bacteria, a thrifty, specific pathway for large virulence proteins. Mol. Microbiol. 40, 306–313.

Kahnert, A., Vermeij, P., Wietek, C., James, P., Leisinger, T., Kertesz, M.A., 2000. The ssu locus plays a key role in organosulfur metabolism in *Pseudomonas putida* S-313. J. Bacteriol. 182, 2869–2878.

Kajava, A.V., Cheng, N., Cleaver, R., Kessel, M., Simon, M.N., Willery, E., Jacob-Dubuisson, F., Locht, C., Steven, A.C., 2001. Beta-helix model for the filamentous haemagglutinin adhesin of *Bordetella pertussis* and related bacterial secretory proteins. Mol. Microbiol. 42, 279–292.

Lee, D.G., Urbach, J.M., Wu, G., Liberati, N.T., Feinbaum, R.L., Miyata, S., Diggins, L.T., He, J., Saucier, M., Déziel, E., Friedman, L., Li, L., Grills, G., Montgomery, K., Kucherlapati, R., Rahme, L.G., Ausubel, F.M., 2006. Genomic analysis reveals that *Pseudomonas aeruginosa* virulence is combinatorial. Genome Biol. 7, R90.

Locht, C., Bertin, P., Menozzi, F.D., Renauld, G., 1993. The filamentous haemagglutinin, a multifaceted adhesion produced by virulent *Bordetella* spp. Mol. Microbiol. 9, 653–660.

Looney, W.J., 2005. Role of *Stenotrophomonas maltophilia* in hospital-acquired infection. Br. J. Biomed. Sci. 62, 145–154.

Lucas, S., Copeland, A., Lapidus, A., Glavina del Rio, T., Dalin, E., Tice, H., Pitluck, S., Chain, P., Malfatti, S., Shin, M., Vergez, L., Lang, D., Schmutz, J., Larimer, F., Land, M., Hauser, L., Kyrpides, N., Mikhailova, N., Taghavi, S., Monchy, S., Newman, L., Vangronsveld, J., van der Lelie, D., Richardson, P., Complete sequence of *Stenotrophomonas maltophilia* R551-3. Submitted (JUN-2008) to the EMBL/GenBank/DDBJ databases.

Mantri, Y., Williams, K.P., 2004. Islander, a database of integrative islands in prokaryotic genomes, the associated integrases and their DNA site specificities. Nucleic Acids Res. 32, D55–D58.

Myers, G.S., Rasko, D.A., Cheung, J.K., Ravel, J., Seshadri, R., DeBoy, R.T., Ren, Q., Varga, J., Awad, M.M., Brinkac, L.M., Daugherty, S.C., Haft, D.H., Dodson, R.J., Madupu, R., Nelson, W.C., Rosovitz, M.J., Sullivan, S.A., Khouri, H., Dimitrov, G.I., Watkins, K.L., Mulligan, S., Benton, J., Radune, D., Fisher, D.J., Atkins, H.S., Hiscox, T., Jost, B.H., Billington, S.J., Songer, J.G., McClane, B.A., Titball, R.W., Rood, J.I., Melville, S.B., Paulsen, I.T., 2006. Skewed genomic variability in strains of the toxigenic bacterial pathogen, *Clostridium perfringens*. Genome Res. 16, 1031–1040.

Nakayama, K., Kanaya, S., Ohnishi, M., Terawaki, Y., Hayashi, T., 1999. The complete nucleotide sequence of phi CTX, a cytotoxin-converting phage of *Pseudomonas aeruginosa*, implications for phage evolution and horizontal gene transfer via bacteriophages. Mol. Microbiol. 31, 399–419.

Nishiyama, M., Ishikawa, T., Rechsteiner, H., Glockshuber, R., 2008. Reconstitution of pilus assembly reveals a bacterial outer membrane catalyst. Science 320, 376–379.

Page, D., Rose, J., Conrod, S., Cuine, S., Carrier, P., Heulin, T., Achouak, W., 2008. Heavy metal tolerance in *Stenotrophomonas maltophilia*. PLoS ONE 3, e1539.

Palleroni, N.J., Bradbury, J.F., 1993. *Stenotrophomonas*, a new bacterial genus for *Xanthomonas maltophilia* (Hugh 1980) Swings et al. 1983. Int. J. Syst. Bacteriol. 43, 606–609.

Panne, D., Müller, S.A., Wirtz, S., Engel, A., Bickle, T.A., 2001. The McrBC restriction endonuclease assembles into a ring structure in the presence of G nucleotides. EMBO J. 20, 3210–3217.

Patil, P.B., Bogdanove, A.J., Sonti, R.V., 2007. The role of horizontal transfer in the evolution of a highly variable lipopolysaccharide biosynthesis locus in *xanthomonads* that infect rice, citrus and crucifers. BMC Evol. Biol. 7, 243.

Roscetto, E., Rocco, F., Carlomagno, M.S., Casalino, M., Colonna, B., Zarrilli, R., Di Nocera, P.P., 2008. PCR-based rapid genotyping of *Stenotrophomonas maltophilia* isolates. BMC Microbiol. 8, 202.

Safdar, A., Rolston, K.V., 2007. *Stenotrophomonas maltophilia*, changing spectrum of a serious bacterial pathogen in patients with cancer. Clin. Infect. Dis. 45, 1602–1609.

Sambou, T., Dinadayala, P., Stadthagen, G., Barilone, N., Bordat, Y., Constant, P., Levillain, F., Neyrolles, O., Gicquel, B., Lemassu, A., Daffé, M., Jackson, M., 2008. Capsular glucan and intracellular glycogen of *Mycobacterium tuberculosis*, biosynthesis and impact on the persistence in mice. Mol. Microbiol. 70, 762–774.

Senol, E., 2004. *Stenotrophomonas maltophilia*, the significance and role as a nosocomial pathogen. J. Hosp. Infect. 57, 1–7.

Taghavi, S., Garafola, C., Monchy, S., Newman, L., Hoffman, A., Weyens, N., Barac, T., Vangronsveld, J., Lelie, D., 2009. Genome survey and characterization of endophytic bacteria exhibiting a beneficial effect on growth and development of poplar trees. Appl. Environ. Microbiol. 75, 748–757.

Venditti, R., De Gregorio,, E., Silvestro, G., Bertocco, T., Salza, M.F., Zarrilli, R., Di Nocera,, P.P., 2007. A novel class of small repetitive DNA sequences in *Enterococcus faecalis*. FEMS Microbiol. Lett. 271, 193–201.

Waters, V.J., Gómez, M.I., Soong, G., Amin, S., Ernst, R., Prince, A., 2007. Immunostimulatory properties of the emerging pathogen *Stenotrophomonas maltophilia*. Infect. Immun. 75, 1698–1703.

Welch, R.A., Burland, V., Plunkett III, G., Redford, P., Roesch, P., Rasko, D., Buckles, E.L., Liou, S.R., Boutin, A., Hackett, J., Stroud, D., Mayhew, G.F., Rose, D.J., Zhou, S, Schwartz, D.C., Perna, N.T., Mobley, H.L., Donnenberg, M.S., Blattner, F.R., 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. Proc. Natl. Acad. Sci. USA 99, 17020–17024.

Williams, K.P., 2003. Traffic at the tmRNA gene. J. Bacteriol. 185, 1059–1070.

Williams, K.P., 2002. Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes, sublocation preference of integrase subfamilies. Nucleic Acids Res. 30, 866–875.

Yan, A., Guan, Z., Raetz, C.R., 2007. An undecaprenyl phosphate-aminoarabinose flippase required for polymyxin resistance in *Escherichia coli*. J. Biol. Chem. 282, 36077–36089.

FEMS MICROBIOLOGY LETTERS

RESEARCH LETTER

A giant family of short palindromic sequences in *Stenotrophomonas maltophilia*

Francesco Rocco, Eliana De Gregorio and Pier Paolo Di Nocera

Dipartimento di Biologia e Patologia Cellulare e Molecolare, Università FEDERICO II, Napoli, Italy

Correspondence: Pier Paolo Di Nocera, Dipartimento di Biologia e Patologia Cellulare e Molecolare, Università FEDERICO II, Via S. Pansini 5, 80131 Napoli, Italy. Tel.: +39 81 746 2059; fax: +39 81 770 3285; e-mail: dinocera@unina.it

Received 9 March 2010; revised 30 April 2010; accepted 30 April 2010.
Final version published online 26 May 2010.

DOI: 10.1111/j.1574-6968.2010.02010.x

Editor: Roger Buxton

Keywords
repeated DNA sequences; palindromic DNA; stem-loop structures; whole-genome analysis; RNA hairpins; microbiological diagnostic.

Abstract

The genome of *Stenotrophomonas maltophilia* is peppered with palindromic elements called SMAG (*Stenotrophomonas maltophilia* GTAG) because they carry at one terminus the tetranucleotide GTAG. The repeats are species-specific variants of the superfamily of repetitive extragenic palindromes (REPs), DNA sequences spread in the intergenic space in many prokaryotic genomes. The genomic organization and the functional features of SMAG elements are described herein. A total of 1650 SMAG elements were identified in the genome of the *S. maltophilia* K279a strain. The elements are 22–25 bp in size, and can be sorted into five distinct major subfamilies because they have different stem and loop sequences. One fifth of the SMAG family is comprised of single units, 2/5 of elements located at a close distance from each other and 2/5 of elements grouped in tandem arrays of variable lengths. Altogether, SMAGs and intermingled DNA occupy 13% of the intergenic space, and make up 1.4% of the chromosome. Hundreds of genes are immediately flanked by SMAGs, and the level of expression of many may be influenced by the folding of the repeats in the mRNA. Expression analyses suggested that SMAGs function as RNA control sequences, either stabilizing upstream transcripts or favoring their degradation.

Introduction

Stenotrophomonas maltophilia is a nonfermentative Gram-negative bacterium that is ubiquitous in nature. It constitutes one of the dominant rhizosphere inhabitants (Ryan *et al.*, 2009; Taghavi *et al.*, 2009), but is also increasingly being described as an important nosocomial pathogen in debilitated and immunodeficient patients, and has been associated with a broad spectrum of clinical syndromes. It has been isolated frequently from cystic fibrosis patients, and has emerged as a serious pathogen in cancer patients (Looney *et al.*, 2009). *Stenotrophomonas maltophilia* displays an intrinsic resistance to many antibiotics, making the selection of optimal therapy difficult (Crossman *et al.*, 2008). Whether the bacterium is a mere colonizer or an infectious agent often remains unresolved, and virulence factors are still ill-defined. The chromosomes of the clinical K279a (Crossman *et al.*, 2008) and the environmental R551-3 (Taghavi *et al.*, 2009) strains exhibit extensive synteny, but each is punctuated by about 40 different GEIs or genomic islands (Rocco *et al.*, 2009). Whether pathogenicity may be associated in part with the maintenance of specific GEIs in the *S. maltophilia* population remains to be established.

Stenotrophomonas maltophilia is extremely heterogeneous at the genetic level (Coenye *et al.*, 2004; Kaiser *et al.*, 2009). We described a procedure to obtain a rapid genotyping of *S. maltophilia* isolates based on the measurement of length variations of genomic regions marked by arrays of palindromic sequences (Roscetto *et al.*, 2008). In this paper, we describe the organization and the features of this peculiar class of repeats, called SMAG (*Stenotrophomonas maltophilia* GTAG), because they carry at one terminus the tetranucleotide GTAG. SMAGs are species-specific members of the superfamily of repetitive extragenic palindromes (REPs), sequences described earlier in *Escherichia coli* and other *Enterobacteriaceae* (Higgins *et al.*, 1988; Versalovic *et al.*, 1991; Bachellier *et al.*, 1999) and later on in other prokaryotes (Aranda-Olmedo *et al.*, 2002; Feil *et al.*, 2005; Tobes & Pareja, 2005; Tobes & Ramos, 2005). SMAGs constitute the largest family of REPs described so far. A look at the structure and organization of SMAG elements provides information on the processes underlying the expansion and

remodeling of REP families, and the functional role that REPs may play.

Materials and methods

In silico analyses

Searches were carried out on the genomes of the *S. maltophilia* strains K279a (http://www.ncbi.nlm.nih.gov/nuccore/NC_010943) and R551-3 (http://www.ncbi.nlm.nih.gov/nuccore/NC_011071) and the 50 contigs of the strain SKA14 (http://www.ncbi.nlm.nih.gov/nuccore/NZ_ACDV00000000). The K279a genome was searched for SMAG sequences using the FUZZNUC program (http://mobyle.pasteur.fr/cgi-bin/portal.py?form=fuzznuc). Initial searches were performed using as a query the sequence described in Roscetto *et al.* (2008), and selecting homologous sequences containing up to four mismatches. Sequence variants were subsequently used as queries for refined searches. Regions of interest in the R551-3 and SKA14 genomes were identified by BLAST.

Bacterial strains and PCR analyses

SMAG-negative regions were searched in the DNA of 25 *S. maltophilia* strains (92, 262, 527, 545, 549, 598, 616, 707, 714, 915, 1019, 1029, 1039, 1054, STM2, OBGTC3, OBGTC13, OBGTC16, OBGTC22, OBGTC28, OBGTC29, OBGTC30, LMG959, LMG10851 and LMG10871) by PCR and sequence analyses. The strains and PCR conditions were described previously (Roscetto *et al.*, 2008).

RNA analyses

Reverse transcriptase-PCR (RT-PCR) analyses were carried out by reverse transcribing total *S. maltophilia* RNA by random priming, and amplifying the resulting cDNA using pairs of gene-specific oligonucleotides as described (De Gregorio *et al.*, 2005). RNAse protection and primer extension assays were carried out as described (De Gregorio *et al.*, 2005). The sequences of all the primers used are available upon request.

Results

The SMAG family

A thorough analysis of the chromosome of the *S. maltophilia* K279a strain revealed that the SMAG family is much wider than postulated initially (Roscetto *et al.*, 2008). K279a DNA hosts 1650 SMAG repeats, all constituted by a stem-loop sequence (SLS) flanked, at one side, by the tetranucleotide GTAG. The genomic coordinates of all SMAGs are reported in the Supporting Information, Table S1. The elements can be sorted, on the basis of changes in the stem and loop residues, into 40 variants. For the sake of simplicity, they have been assigned to five major subfamilies (Fig. 1a). The large SMAG-1 subfamily includes all the repeats used for genotyping (Roscetto *et al.*, 2008). SMAG-1 to SMAG-4 repeats have 8 bp stems and SMAG-5 repeats have 9 bp stems. The *S. maltophilia* genome contains hundreds of DNA tracts that partly resemble SMAG sequences. We discarded complementary sequences fitting the consensuses shown in Fig. 1a, but either located 5 bp away or more, or containing more than two mismatches. In the selection scheme adopted, GT pairing was allowed, because SMAGs may fold into secondary structures at the DNA as at the RNA level. In most repeats, stem sequences are fully complementary (Fig. 1b). An exception is SMAG-2 units, many of which have stems with one to two mismatches. In 50% of the stems with one mismatch, the first base pair is mutated. The folding ability of these elements is therefore impaired only slightly.

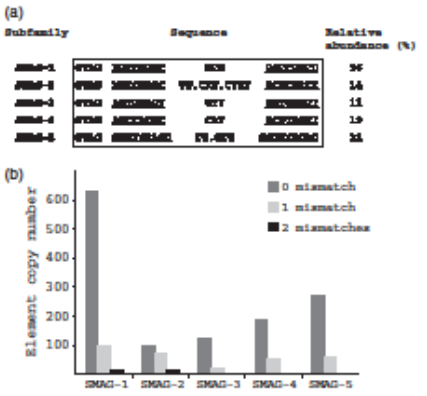

Fig. 1. The SMAG family. (a) The consensus sequences of the five subfamilies of SMAG repeats identified in the K279a genome, and the relative abundance of each subfamily, are shown. Sequences are according to the IUB codes: A, adenosine; C, cytidine; G, guanosine; T, thymidine; H, A or C or T; D, A or G or T; K, G or T; M, A or C; N, any base; R, A or G; S, G or C; Y, C or T. Complementary residues are underlined. (b) Abundance of SMAG-1 to SMAG-5 units with zero, one or two mismatches.

Genomic organization of SMAG repeats

Only 20% of the SMAG family is comprised of solitary elements. Most repeats are grouped into a few predominant arrangements, described below.

Dimers > 1/3 of the SMAG family is comprised of elements located at a close distance (<100 bp) from each other. On the basis of their relative position, these elements form head–head (HH) or head–tail (HT) or tail–tail (TT)

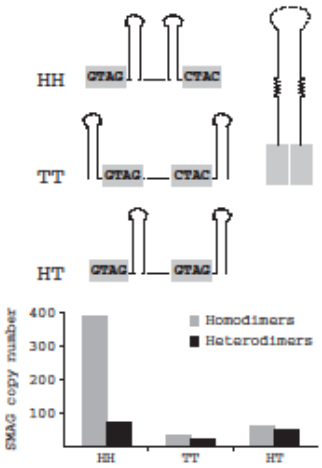

Fig. 2. SMAG dimers. The three classes of SMAG dimers, and the SLSs potentially formed by alternative folding of HH and TT dimers are depicted. The abundance of HH, TT and HT homodimers and heterodimers is shown.

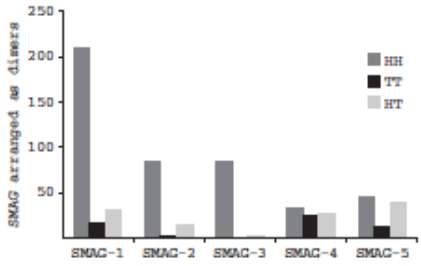

Fig. 3. Abundance of HH, TT and HT dimers among SMAG subfamilies.

dimers. Dimers range in size from 47 to 142 bp, the majority of them being ~70–90 bp in size. Paired repeats belong to the same (homodimers) or different (heterodimers) subfamilies. In total, 228 HH, 55 HT and 26 TT dimers were identified in the K279a chromosome (Fig. 2). HH homodimers represent the most abundant category of paired elements. The differences among dimer categories shown in Fig. 2 are statistically significant ($\chi^2 = 53.4$, $P = 2.5 \times 10^{-12}$). A main difference among the HH, TT and HT dimers is that repeats of the first two classes may fold, rather than into separate SLSs, into a large one (Fig. 2). According to analyses carried out at the MFOLD web server (Zuker, 2003), 70% of HH dimers may fold into large SLSs, with dG values ranging from −50 to −70 kcal mol^{-1}. In none of the three classes of heterodimers could a preferential combination of specific subfamilies repeats be observed. In terms of homodimers, HH dimers are predominantly comprised of SMAG-1, SMAG-2 and SMAG-3 sequences. In contrast, TT dimers are predominantly comprised of SMAG-4 (Fig. 3).

Spacer sequences that separate dimer repeats are poorly homologous. An exception is the spacers of SMAG-3 HH homodimers, most of which (30/40) fit the consensus sequence nnCGCGCGCAGCGCGGn$_{(16-19)}$GAAGAGC.

Trimers at 86 loci in the K279a genome, groups of three repeats can be found at a close distance from each other. Taking into account the relative position of each element, trimers can be viewed as dimers flanked by solo repeats. Twenty-eight trimers include SMAGs from one subfamily, 58 SMAGs belonging to two or three subfamilies.

Clusters 456 elements are clustered at 64 loci at a 10–150 bp distance from each other. Large clusters may include up to 22 repeats, and contain elements from different subfamilies. Most clusters contain 4–8 SMAGs, are comprised of repeats of one subfamily and result from tandem amplification of SMAGs (monomers or dimers), together with stretches of flanking DNA of variable lengths.

Interspersion of SMAGs with coding sequences

Many SMAG monomers, dimers and trimers are at a close distance from genes. We found 307 SMAGs located 1–20 bp from ORF stop codons, and 99 that overlap ORF stop codons. Nine of the overlapping repeats encode a few aminoacids and the stop codon; all the others provide only the stop codon, the terminal GTAG motif functioning as a UAG translational stop signal. Curiously, the stop codons of the convergently oriented ORFs Smlt0783–Smlt0784 and Smlt4197–Smlt4198, are contributed by interleaved SMAG dimers. The same holds for ORFs Smlt1380–Smlt1381 and Smlt0188–Smlt0189, the stop codons of each being contributed by interleaved SMAG trimers. Some SMAGs located between convergently oriented ORFs are at a close distance from the stop codons of both. Accordingly, the number of the ORFs immediately flanked by SMAGs is higher than the number of repeats (501 vs. 406). By contrast, we found only 81 SMAGs located 1–50 bp from ORF stop codons, and 16 that overlap ORF start codons and encode 4–29 aminoacids. About 1/3 of the ORFs flanked 5′ by SMAGs (26/97) carries SMAG sequences also at the 3′ end. K279a ORFs at a close distance from SMAGs are listed in Table S2.

Thirty SMAGs are entirely located within ORFs. These repeats can be sorted into two main groups. Sixteen out of 30 lie within ORFs encoding small hypothetical proteins that do not exhibit significant homology to ORFs encoded

by either the *S. maltophilia* R551-3 or other prokaryotic genomes, and thus plausibly do not correspond to authentic gene products. Similar conclusions were reached for short ORFs interrupted by REPs in *Pseudomonas syringae* (Tobes & Pareja, 2005). The remaining 14 repeats are found at the same relative genome coordinates in the R551-3 DNA. However, only six interrupted ORFs are conserved in the two strains. SMAGs within ORFs are listed in Table S3.

On the whole, intergenic SMAGs are found at 747 loci. Of these, 370 separate unidirectionally transcribed ORFs, 343 convergently transcribed ORFs and only 34 divergently transcribed ORFs.

Conservation of SMAG sequences in other *S. maltophilia* strains

The size of repeated DNA families may vary among isolates. To gain a rough estimate of the size of SMAG families scattered in the other two sequenced *S. maltophilia* genomes, repeats perfectly matching the 40 SMAG sequence variants found in K279a DNA were searched in R551-3 and SKA14 DNAs. The relative abundance of the five SMAG subfamilies is comparable in the three genomes. However, their sizes varied, SMAG-2 elements being more abundant in R551-3 and SKA14 and SMAG-3 being predominant in K279a DNA (Fig. 4). The degree of conservation of SMAG sequences was checked by direct sequence comparisons. Thirty-two regions of the K279a chromosome containing SMAG-3 dimers were analyzed in R551-3. Dimers were conserved in 10 regions, missing in nine and replaced in 13 by SMAG-1 or SMAG-2 sequences (monomers or dimers). Fifty K279a intergenic regions containing SMAG-1 HH dimers were also checked in R551-3 DNA. Most (91%) of the K279a SMAG-1 fit the consensus WGCCGGCCgctGGCCGCCW, and have been called α units, and only 4% fit the consensus CGCCGGGCcatGC CCGGCG, and have been called β units (lowercase letters denote loop sequences). Consequently, most (88/99) K279a SMAG-1 HH dimers are comprised of α units. α dimers were conserved in 32/50 regions. Yet, the significant difference in spacer sequences makes it likely that some K279a dimers had been replaced by homologous dimers in R551-3 DNA or vice versa. α dimers were replaced by single β repeats in four regions, β HH dimers in five regions, β TT dimers in three regions and an SMAG-5 TT dimer in one region. SMAG sequences were not found in five regions. In three of them, 40–90-bp-long tracts with an almost perfect dyad symmetry were found.

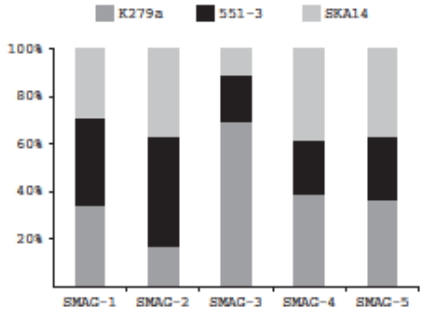

Fig. 4. Size variations of SMAG subfamilies in *Stenotrophomonas maltophilia* genomes. The height of stacked bars denotes the dimension of each subfamily in the three sequenced *S. maltophilia* genomes.

The changes observed arise from a recombination plausibly driven by the terminal GTAG sequences. The presence at several sites of either alternative SMAGs or unrelated palindromic sequences suggests that the functional role played by SMAG repeats is primarily associated with their ability to fold into secondary structures.

RNA analyses of SMAG-containing selected loci

The pattern of chromosomal interspersion suggests that many SMAG sequences may be passively transcribed into mRNA. Folding of these repeats into RNA hairpins may influence the level of expression of flanking genes. To investigate this issue, 14/50 K279a chromosomal regions containing SMAGs inserted between unidirectionally transcribed genes, and located at a short distance from both, were selected, and their lengths were measured in 25 *S. maltophilia* strains by PCR. The sizes of the amplimers suggested that SMAG sequences were conserved in most of the analyzed regions. Only two SMAG-negative regions were identified in two different strains, 545 and STM2, and the lack of SMAG DNA was confirmed by sequence analysis. Transcripts spanning the selected genes were detected by RT-PCR, and SMAG-negative regions functioned as a control. The detection of K279a transcripts encompassing both ORFs in each pair ensured that ORFs and interleaved SMAGs are transcribed from the same promoter (Fig. 5). For both gene pairs, upstream transcripts accumulated at higher levels than downstream transcripts in K279a, but not in the strains 545 and STM2 lacking SMAG sequences (Fig. 5). The 4076/4075 cDNA ratio did not change in strain 1029, in which ORFs are separated by a SMAG monomer (Fig. 5b). This suggests that, in a given RNA context, SMAG monomers and dimers function as RNA stabilizers with the same efficiency. We also analyzed a trimeric SMAG repeat located 4 bp downstream from the sensor kinase and the response regulator genes of the smeS–smeR two-component system (ORFs 4477 and 4478), and 13 bp upstream of ORF 4479, which encodes a hypothetical protein. The short distances suggest that the SMAG trimer is cotranscribed

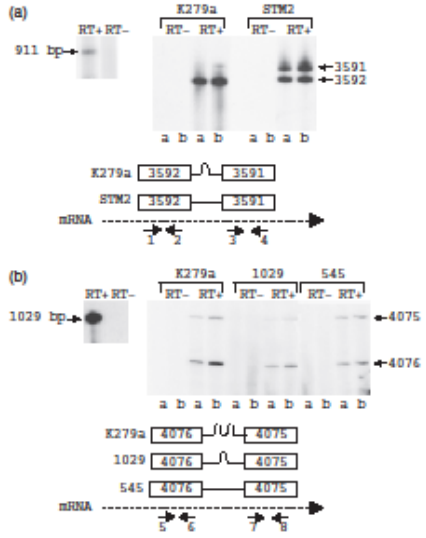

Fig. 5. RNA expression of genes flanked by SMAGs. Total RNAs (200 ng) derived from the K279a, STM2, 545 and 1029 *Stenotrophomonas maltophilia* strains were reverse transcribed using a mixture of random hexamers as primers. Transcripts homologous to ORFs 3592 and 3591 [(a)] and ORFs 4076 and 4075 [(b)] were measured by RT-PCR using pairs of gene-specific oligonucleotides. Lanes a and b show the reaction products obtained after 24 and 27 amplification cycles, respectively. Amplimers were detected only when samples were incubated with reverse transcriptase (RT+ lanes) before PCR. The 911- and 1029-bp amplimers in the small autoradiograms in the left side of (a) and (b) correspond to K279a transcripts spanning SMAG and flanking ORFs detected with primers 1 and 4 (ORFs 3592 and 3591) and 5 and 8 (ORFs 4076 and 4075) after 35 PCR cycles. Single and double hairpins indicate SMAG monomers and dimers, and dotted lines indicate SMAG-negative intergenic regions. ORFs 3592 and 3591 encode the adenylosuccinate synthetase and a putative transmembrane protein, and are at a distance of 1 and 3 bp from SMAG sequences, respectively. The two ORFs are the last two cistrons of an operon also including ORFs 3596, 3595, 3594 and 3593. ORFs 4076 and 4075 correspond to the heat shock proteins HslV and HslU, and are at a distance of 2 and 25 bp from SMAG sequences, respectively.

with flanking ORFs. We failed to identify strains lacking SMAG sequences in this region that could function as a control. RT-PCR experiments similar to those shown in Fig. 5 revealed that downstream 4479 transcripts accumulated at high levels, but upstream 4478 transcripts were almost undetectable (Fig. 6a). To clarify the issue, RNAse protection assays were carried out. Antisense RNAs of different lengths spanning 4478 and 4479 ORFs protected only 4479 transcripts (Fig. 6b). Intriguingly, protected bands included the SMAG repeat labeled as c in Fig. 6b. The same result was obtained in RNA extension experiments, in which bands of elongation extended over SMAG repeat c only (Fig. 6c). We hypothesize that repeats a and b fold into one large secondary structure, which is cleaved, and this promotes rapid 3′–5′ degradation of upstream 4478 transcripts.

Discussion

The number of predicted SLSs is significantly higher in prokaryotic genomes existing in nature than in random sequences of comparable GC content (Petrillo *et al.*, 2006). This implies that the ability of a variety of sequences to fold into secondary structures is positively selected in prokaryotic genomes and may have functional significance. A fraction of SLSs is represented by REPs, sequences shown or hypothesized to serve different functions. REPs are binding sites for the integration host factor, a protein required for site-specific recombination and DNA replication (Engelhorn *et al.*, 1995). REPs are targets for the DNA gyrase (Espéli & Boccard, 1997), and repeats located between convergent genes may be a privileged target for the enzyme, in order to counteract the excess of positive supercoiling induced in the chromosome by DNA transcription (Moulin *et al.*, 2005). As RNA elements, REPs may enhance the stability of 5′ proximal mRNA segments (Khemici & Carpousis, 2004). Finally, REPs induce innate immune system stimulation via TLR9, and could play a key role in the pathogenesis of Gram-negative septic shock (Magnusson *et al.*, 2007).

Tobes & Ramos (2005) established that, for a palindromic sequence to be considered as REP, the following criteria should be met: (a) be extragenic, (b) range in size from 21 to 65 bp and (c) constitute > 0.5% of the total intergenic space. SMAGs meet all these criteria, and constitute the largest set of REPs described so far. SMAGs correspond to the repeats identified by Nunvar *et al.* (2010). SMAGs can be sorted into five distinct subfamilies, and come in different genomic formats. Single units make up only 1/5 of the SMAG family. The remaining elements are organized as dimers or are grouped in tandem arrays of variable lengths. Altogether, SMAGs and intermingled DNA occupy 13% of the overall intergenic space, and make up 1.4% of the total chromosome.

SMAG families residing in the environmental R551-3 and SKA14 *S. maltophilia* strains are comparable in size to the repeat family found in K279a. Yet, the sizes of some subfamilies vary, and K279a is enriched in SMAG-3. Most SMAG-3 are organized as HH dimers that feature conserved spacers, and may thus represent a relatively young sequence family variant. Changes in the abundance and chromosomal

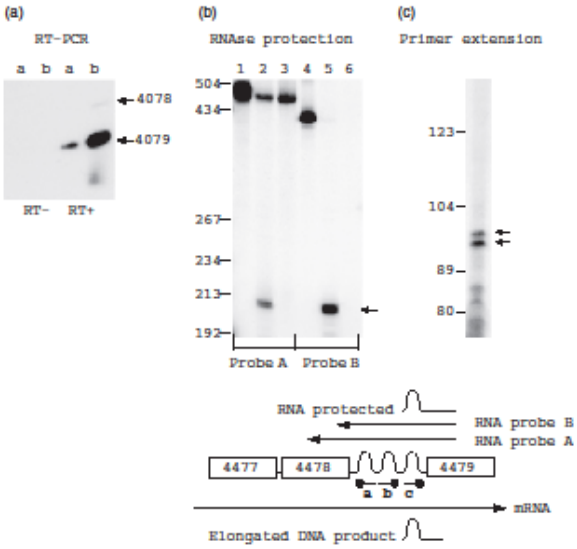

Fig. 6. Cleavage of SMAG RNA. (a) Transcripts corresponding to ORF 4479 (down) and ORF4478 (up) accumulated in K279a cells were detected as in Fig. 5. (b) Antisense RNA probes A and B (lanes 1 and 4) were hybridized to 20 μg of *Stenotrophomonas maltophilia* K279a (lanes 2 and 5) or yeast (lanes 3 and 6) RNA. T1 RNAse-resistant RNA hybrids were electrophoresed on a 6% polyacrylamide–8 M urea gel. (c) A primer complementary to ORF 4479 was hybridized to 10 μg of K279a RNA. Annealed primer moieties were extended by reverse transcriptase, and elongated products were electrophoresed on a 6% polyacrylamide-8 M urea gel. Major reaction products are marked by arrows. Numbers to the left of the autoradiograms indicate the size in nucleotides of coelectrophoresed DNA molecular size markers. SMAG sequences are shown as in Fig. 5. Filled circles indicate GTAG termini.

distribution may make SMAG-3 sequences suitable for use in accurate genotyping and epidemiological studies.

Also, the ~500 REPs identified in the *E. coli* MG1655 strain have been sorted into subfamilies. Similar to SMAGs, single REPs represent only 20% of the family, the other elements being grouped in various configurations, all denoted as bacterial interspersed mosaic elements (BIME; reviewed in Bachellier *et al.*, 1999). BIME-1 and BIME-2 correspond to SMAG TT and HH dimers. However, HH dimers are about 10 times more abundant than TT dimers. In contrast, BIME-1 (74 repeats) are three times more abundant than BIME-2 (24 repeats). Moreover, both BIME-1 and BIME-2 are invariably comprised of elements from different subfamilies (Bachellier *et al.*, 1999; see also http://www.pasteur.fr/recherche/unites/pmtg/repet/index.html). The predominance of TT over HH dimers, and the composite nature of dimers, is also a distinctive feature of the abundant REP families found in *Pseudomonas putida* (Aranda-Olmedo *et al.*, 2002) and *P. syringae* (Feil *et al.*, 2005).

It has been hypothesized that REPs are mobilized by a transposase of the IS200/IS605 family, and the corresponding genes have been shown to be flanked by REPs in many species (Nunvar *et al.*, 2010). Four genes encoding this transposase were identified in K279a DNA (ORFs 1101, 1152, 2816 and 4509), but only ORFs 1101 and 2816 are flanked by SMAGs. We believe that REPs are an ancient component of the genomes of *Proteobacteria*, which have been actively mobilized by transposition only early in their history. According to this view, REPs disappeared in time from most species, their dissemination being plausibly detrimental to the cell, and have been maintained only in species in which they could no longer transpose. This hypothesis is supported by the observation that SMAG sequences were found in none of the 41 species-specific GEIs, plausibly acquired by lateral gene transfer, which account for >10% of the K279a chromosome (Rocco *et al.*, 2009). REPs are similarly restricted to core genome regions in *P. syringae* (Tobes & Pareja, 2005). In contrast to what was observed for REPs in other species (Tobes & Pareja, 2006), SMAGs are not targeted by mobile DNA. However, it is worth noting that a K279a GEI encoding type 1 pili (Rocco *et al.*, 2009) is flanked by SMAG-2 dimers. About 1/7 of the ORFs of the K279a strain are flanked by SMAGs in a distance range that makes the presence of promoter or terminator sequences unlikely. It is plausible that most of these elements are transcribed into mRNA, and that their folding into RNA hairpins may influence the level of expression of flanking genes. The number of genes potentially controlled at the post-transcriptional level by SMAGs may be higher than estimated, because many repeats

are inserted either upstream (17 elements) or downstream (150 elements) or within (44 elements) known or putative operons.

We analyzed genes transcribed in the same direction intermingled with SMAG sequences, and found that the repeats influence the segmental mRNA stability. Both monomers and dimers function as stabilizers of upstream transcripts, and work with comparable efficiency when embedded in the same RNA context (Fig. 5). RNA expression data are in line with the results of *in silico* analyses, indicating that some of the genes separated by HH dimers in K279a are intermingled with monomers, or HH, or even TT dimers of the same or different SMAG subfamilies in R551-3 DNA. This varied scenario shows that recombination may extensively reshape SMAG-positive regions without substantially altering the regulatory role of SMAGs. The distance between ORFs and SMAGs increased 10–15 bp in some R551-3 regions. This suggests that SMAGs may function as RNA elements over a relatively flexible distance interval. Some SMAGs may favor the degradation of upstream transcripts. This may correlate to the cleavage of large SLSs formed by alternative folding of SMAG dimers (Fig. 6). These structures resemble RNA hairpins formed by 100–170 bp repeats found in *Neisseriae* (De Gregorio *et al.*, 2003) and *Yersiniae* (De Gregorio *et al.*, 2006), which may be cleaved by RNAse III. Whether the hypothesized structures may be formed, whether they are cut by specific endoribonucleases or are resistant to cleavage is likely determined by the overall mRNA context in which SMAG dimers are embedded. Thorough analyses may eventually establish how SMAG sequences regulate the level of expression of different sets of *S. maltophilia* genes.

The dimensions and the complexity of the SMAG family make *S. maltophilia* an ideal organism to gain knowledge of the universe of small palindromic sequences, and clarify the roles that they may play in the lifestyle of the organisms in which they reside.

Acknowledgements

We are indebted to Raffaele Zarrilli for critically reading the manuscript, and Sergio Cocozza for statistical analyses. We thank one of the referees for hints and suggestions. Research was supported by a grant from the Italian Cystic Fibrosis Research Foundation (FFC) to P.P.D.N.

References

Aranda-Olmedo I, Tobes R, Manzanera M, Ramos JL & Marques S (2002) Species-specific repetitive extragenic palindromic (REP) sequences in *Pseudomonas putida*. *Nucleic Acids Res* **30**: 1826–1833.

Bachellier S, Clement JM & Hofnung M (1999) Short palindromic repetitive DNA elements in enterobacteria: a survey. *Res Microbiol* **150**: 627–639.

Coenye T, Vanlaere E, LiPuma JJ & Vandamme P (2004) Identification of genomic groups in the genus *Stenotrophomonas* using gyrB RFLP analysis. *FEMS Immunol Med Mic* **40**: 181–185.

Crossman LC, Gould VC, Dow JM *et al.* (2008) The complete genome, comparative and functional analysis of *Stenotrophomonas maltophilia* reveals an organism heavily shielded by drug resistance determinants. *Genome Biol* **9**: R74.

De Gregorio E, Abrescia C, Carlomagno MS & Di Nocera PP. (2003) Ribonuclease III-mediated processing of specific Neisseria meningitidis mRNAs. *Biochem J.* **374**: 799–805.

De Gregorio E, Silvestro G, Petrillo M, Carlomagno MS & Di Nocera PP (2005) Enterobacterial repetitive intergenic consensus sequence repeats in *Yersiniae*: genomic organization and functional properties. *J Bacteriol* **187**: 7945–7954.

De Gregorio E, Silvestro G, Venditti R, Carlomagno MS & Di Nocera PP (2006) Structural organization and functional properties of miniature DNA insertion sequences in *Yersiniae*. *J Bacteriol* **188**: 7876–7884.

Engelhorn M, Boccard F, Murtin C, Prentki P & Geiselmann J (1995) *In vivo* interaction of the *Escherichia coli* integration host factor with its specific binding sites. *Nucleic Acids Res* **23**: 2959–2965.

Espéli O & Boccard F (1997) *In vivo* cleavage of *Escherichia coli* BIME-2 repeats by DNA gyrase: genetic characterization of the target and identification of the cut site. *Mol Microbiol* **26**: 767–777.

Feil H, Feil WS, Chain P *et al.* (2005) Comparison of the complete genome sequences of *Pseudomonas syringae* pv. syringae B728a and pv. tomato DC3000. *P Natl Acad Sci USA* **102**: 11064–11069.

Higgins CF, McLaren RS & Newbury SF (1988) Repetitive extragenic palindromic sequences, mRNA stability and gene expression: evolution by gene conversion? A review. *Gene* **72**: 3–14.

Kaiser S, Biehler K & Jonas D (2009) A *Stenotrophomonas maltophilia* multilocus sequence typing scheme for inferring population structure. *J Bacteriol* **191**: 2934–2943.

Khemici V & Carpousis AJ (2004) The RNA degradosome and poly(A) polymerase of *Escherichia coli* are required *in vivo* for the degradation of small mRNA decay intermediates containing REP-stabilizers. *Mol Microbiol* **51**: 777–790.

Looney WJ, Narita M & Mühlemann K (2009) *Stenotrophomonas maltophilia*: an emerging opportunist human pathogen. *Lancet Infect Dis* **9**: 312–323.

Magnusson M, Tobes R, Sancho J & Pareja E (2007) Cutting edge: natural DNA repetitive extragenic sequences from gram-negative pathogens strongly stimulate TLR9. *J Immunol* **179**: 31–35.

Moulin L, Rahmouni AR & Boccard F (2005) Topological insulators inhibit diffusion of transcription-induced positive

supercoils in the chromosome of *Escherichia coli*. *Mol Microbiol* **55**: 601–610.
Nunvar J, Huckova T & Licha I (2010) Identification and characterization of repetitive extragenic palindromes (REP)-associated tyrosine transposases: implications for REP evolution and dynamics in bacterial genomes. *BMC Genomics* **11**: 44.
Petrillo M, Silvestro G, Di Nocera PP, Boccia A & Paolella G (2006) Stem-loop structures in prokaryotic genomes. *BMC Genomics* **7**: 170.
Rocco F, De Gregorio E, Colonna B & Di Nocera PP (2009) *Stenotrophomonas maltophilia* genomes: a start-up comparison. *Int J Med Microbiol* **299**: 535–546.
Roscetto E, Rocco F, Carlomagno MS, Casalino M, Colonna B, Zarrilli R & Di Nocera PP (2008) PCR-based rapid genotyping of *Stenotrophomonas maltophilia* isolates. *BMC Microbiol* **8**: 202.
Ryan RP, Monchy S, Cardinale M, Taghavi S, Crossman L, Avison MB, Berg G, van der Lelie D & Dow JM (2009) The versatility and adaptation of bacteria from the genus *Stenotrophomonas*. *Nat Rev Microbiol* **7**: 514–525.
Taghavi S, Garafola C, Monchy S, Newman L, Hoffman A, Weyens N, Barac T, Vangronsveld J & van der Lelie D (2009) Genome survey and characterization of endophytic bacteria exhibiting a beneficial effect on growth and development of poplar trees. *Appl Environ Microb* **75**: 748–757.
Tobes R & Pareja E (2005) Repetitive extragenic palindromic sequences in the *Pseudomonas syringae* pv. tomato DC3000 genome: extragenic signals for genome reannotation. *Res Microbiol* **156**: 424–433.
Tobes R & Pareja E (2006) Bacterial repetitive extragenic palindromic sequences are DNA targets for insertion sequence elements. *BMC Genomics* **7**: 62.
Tobes R & Ramos JL (2005) REP code: defining bacterial identity in extragenic space. *Environ Microbiol* **7**: 225–228.
Versalovic J, Koeuth T & Lupski JR (1991) Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. *Nucleic Acids Res* **19**: 6823–6831.
Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**: 3406–3415.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Sequences and chromosomal coordinates of the 1650 SMAG sequences found in K279a DNA.
Table S2. SMAGs that are close to, or overlap K279a ORFs, are listed.
Table S3. K279a ORFs containing SMAG sequences.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

JOURNAL OF BACTERIOLOGY, May 2011, p. 2359–2360 Vol. 193, No. 9
0021-9193/11/$12.00 doi:10.1128/JB.00245-11

Genome Sequences of Three *Acinetobacter baumannii* Strains Assigned to the Multilocus Sequence Typing Genotypes ST2, ST25, and ST78[▽]

Raffaele Zarrilli,[1]* Maria Giannouli,[1] Francesco Rocco,[2] Nicholas J. Loman,[3] Anthony S. Haines,[3] Chrystala Constantinidou,[3] Mark J. Pallen,[3] Maria Triassi,[1] and Pier Paolo Di Nocera[2]

Department of Preventive Medical Sciences, University of Napoli Federico II, Naples, Italy[1]; Department of Cellular and Molecular Biology and Pathology, University of Napoli Federico II, Naples, Italy[2]; and Centre for Systems Biology, School of Biosciences, University of Birmingham, Birmingham B15 2TT, United Kingdom[3]

Received 21 February 2011/Accepted 28 February 2011

***Acinetobacter baumannii* is an emerging opportunistic Gram-negative pathogen responsible for hospital-acquired infections. *A. baumannii* epidemics described in Europe and worldwide were caused by a limited number of genotypic clusters of multidrug-resistant strains. Here, we report the availability of draft genome sequences for three multidrug-resistant *A. baumannii* strains assigned to multilocus sequence typing genotypes ST2, ST25, and ST78 that were more frequently isolated during outbreaks occurred in Greece, Italy, Lebanon, and Turkey.**

Acinetobacter baumannii is an emerging opportunistic pathogen, causing a variety of nosocomial infections (13). Outbreaks of *A. baumannii* were caused by few genotypic clusters of strains that were initially named European clones I, II, and III and are now regarded as international (7, 10, 13) and referred to as ST1, ST2, and ST3, respectively, according to multilocus sequence typing (7). Epidemiological studies showed the prevalence of the international clone II lineage during the last few years (7, 8, 10, 13), as well as the occurrence of epidemics caused by multidrug-resistant strains belonging to novel genotypes ST25 and ST78 in several Mediterranean hospitals (8, 9).

The whole-genome sequences of six ST1 (1, 2, 15) and individual ST2, ST3, and ST77 *A. baumannii* strains were available to date (2, 11, 14, 15). Here, we announce the availability of three draft genome sequences for carbapenem-resistant *A. baumannii* ST2 strain 3990, ST78 strain 3909, and ST25 strain 4190, isolated during cross-transmission episodes that occurred at the Monaldi Hospital, Naples, Italy, during 2006, 2007, and 2009, respectively (8, 9).

The genomes were sequenced to at least 10-fold coverage using 454 FLX Titanium emPCR pyrosequencing (Roche) according to the manufacturer's recommendations. Draft genomes were assembled using *Newbler* and automatically annotated using the xBASE2 bacterial genome annotation service (4). The draft genome sequences of the ST2, ST25, and ST78 strains, respectively, consisted of 96, 396, and 236 contigs, comprised 4,015,011 bases, 4,032,291 bases, and 3,954,832 bases, and generated 3,806, 3,910, and 3,721 protein coding sequences by automated annotation against the *A. baumannii* AB0057 genome. Comparative analysis of the ST2, ST25, and ST78 genomes with the ACICU, AB0057, ABAYE, and ATCC17978 genomes using Mauve software (6) identified 3,068 homologous protein coding sequences at the same relative positions in the seven genomes. Sixty-three DNA segments, ranging in size from 3 to 126 kb, were present only in some of the seven genomes and were either missing or replaced by nonhomologous DNA sequences in others. Consistent with their close genetic relatedness, both the ST2 3990 and the ACICU strain carried a 15.4-kb region containing antimicrobial resistance genes inserted in the ATPase gene locus. At the corresponding chromosomal location, a 12.7-kb region flanked by transposases but devoid of resistance genes was identified in the ST78 but not in the ST25 strain.

Complete plasmid sequences for the ST2, ST25, and ST78 strains were obtained by comparing the DNA sequences of contigs with those determined by a primer-walking technique on purified plasmid preparations. Genome annotations were manually verified and corrected using the Artemis viewer (3). The ST2 3990 strain contained two plasmids, p1-ABST2 (63,320 bp) carrying a complete *tra* locus and p2-ABST2 (21,846 bp) carrying one copy of the carbapenem-hydrolyzing oxacillinase (CHDL) bla_{OXA-58} gene, homologous to plasmids pACICU2 and pACICU1, respectively (11). The ST25 strain contained two distinct plasmids, p1-ABST25 (15,267 bp) and p2-ABST25 (8,970 bp), that both carried one copy of the CHDL bla_{OXA-72} gene and were homologous to plasmids carrying the bla_{OXA-24} gene (5, 12). The single plasmid p1-ABST78 (26,411 bp) identified in the ST78 strain contained one copy of the bla_{OXA-58} gene and was homologous to plasmids p2-ABST2 and pACICU1 (11).

Nucleotide sequence accession numbers. The draft genome sequences of strains 3990, 4190, and 3909 have been deposited at GenBank under accession numbers AEOY00000000, AEPA00000000, and AEOZ00000000, respectively.

* Corresponding author. Mailing address: Dipartimento di Scienze Mediche Preventive, Università di Napoli Federico II, Via Pansini 5, 80131 Napoli, Italy. Phone: 39-081-7463026. Fax: 39-081-7463352. E-mail: rafzarri@unina.it.

▽ Published ahead of print on 11 March 2011.

This work was supported in part by grants from Agenzia Italiana del Farmaco, Italy (AIFA2007 contract no. FARM7X9F8K) and from Ministero dell'Istruzione, dell'Università e della Ricerca, Italy (PRIN 2008 to R.Z.).

We thank people at the Birmingham Science City-funded 454 sequencing facility for genome sequencing.

REFERENCES

1. **Adams, M. D., E. R. Chan, N. D. Molyneaux, and R. A. Bonomo.** 2010. Genomewide analysis of divergence of antibiotic resistance determinants in closely related isolates of *Acinetobacter baumannii*. Antimicrob. Agents Chemother. **54**:3559–3567.
2. **Adams, M. D., et al.** 2008. Comparative genome sequence analysis of multidrug-resistant *Acinetobacter baumannii*. J. Bacteriol. **190**:8053–8064.
3. **Berriman, M., and K. Rutherford.** 2003. Viewing and annotating sequence data with Artemis. Brief. Bioinform. **4**:124–132.
4. **Chaudhuri, R. R., et al.** 2008. xBASE2: a comprehensive resource for comparative bacterial genomics. Nucleic Acids Res. **36**:D543–D546.
5. **D'Andrea, M. M., et al.** 2009. Characterization of pABVA01, a plasmid encoding the OXA-24 carbapenemase from Italian isolates of *Acinetobacter baumannii*. Antimicrob. Agents Chemother. **53**:3528–3533.
6. **Darling, A. C., B. Mau, F. R. Blattner, and N. T. Perna.** 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. **14**:1394–1403.
7. **Diancourt, L., V. Passet, A. Nemec, L. Dijkshoorn, and S. Brisse.** 2010. The population structure of *Acinetobacter baumannii*: expanding multiresistant clones from an ancestral susceptible genetic pool. PLoS One. **5**:e10034.
8. **Di Popolo, A., M. Giannouli, M. Triassi, S. Brisse, and R. Zarrilli.** 2011. Molecular epidemiology of multidrug-resistant *Acinetobacter baumannii* strains in four Mediterranean countries using a multilocus sequence typing scheme. Clin. Microbiol. Infect. **17**:197–201.
9. **Giannouli, M., et al.** 2010. Molecular epidemiology of multi-drug resistant *Acinetobacter baumannii* in a tertiary care hospital in Naples, Italy, shows the emergence of a novel epidemic clone. J. Clin. Microbiol. **48**:1223–1230.
10. **Higgins, P. G., C. Dammhayn, M. Hackel, and H. Seifert.** 2010. Global spread of carbapenem-resistant *Acinetobacter baumannii*. J. Antimicrob. Chemother. **65**:233–238.
11. **Iacono, M., et al.** 2008. Whole-genome pyrosequencing of an epidemic multidrug-resistant *Acinetobacter baumannii* strain belonging to the European clone II group. Antimicrob. Agents Chemother. **52**:2616–2625.
12. **Merino, M., et al.** 2010. OXA-24 carbapenemase gene flanked by XerC/XerD-like recombination sites in different plasmids from different *Acinetobacter* species isolated during a nosocomial outbreak. Antimicrob. Agents Chemother. **54**:2724–2727.
13. **Peleg, A. Y., H. Seifert, and D. L. Paterson.** 2008. *Acinetobacter baumannii*: emergence of a successful pathogen. Clin. Microbiol. Rev. **21**:538–582.
14. **Smith, M. G., et al.** 2007. New insights into *Acinetobacter baumannii* pathogenesis revealed by high-density pyrosequencing and transposon mutagenesis. Genes Dev. **21**:601–614.
15. **Vallenet, D., et al.** 2008. Comparative analysis of acinetobacters: three genomes for three lifestyles. PLoS ONE. **3**:e1805.

Di Nocera et al. BMC Microbiology 2011, 11:224
http://www.biomedcentral.com/1471-2180/11/224

BMC Microbiology

RESEARCH ARTICLE Open Access

Genome organization of epidemic *Acinetobacter baumannii* strains

Pier Paolo Di Nocera[1*], Francesco Rocco[1], Maria Giannouli[2], Maria Triassi[2] and Raffaele Zarrilli[2*]

Abstract

Background: *Acinetobacter baumannii* is an opportunistic pathogen responsible for hospital-acquired infections. *A. baumannii* epidemics described world-wide were caused by few genotypic clusters of strains. The occurrence of epidemics caused by multi-drug resistant strains assigned to novel genotypes have been reported over the last few years.

Results: In the present study, we compared whole genome sequences of three *A. baumannii* strains assigned to genotypes ST2, ST25 and ST78, representative of the most frequent genotypes responsible for epidemics in several Mediterranean hospitals, and four complete genome sequences of *A. baumannii* strains assigned to genotypes ST1, ST2 and ST77. Comparative genome analysis showed extensive synteny and identified 3068 coding regions which are conserved, at the same chromosomal position, in all *A. baumannii* genomes. Genome alignments also identified 63 DNA regions, ranging in size from 4 o 126 kb, all defined as genomic islands, which were present in some genomes, but were either missing or replaced by non-homologous DNA sequences in others. Some islands are involved in resistance to drugs and metals, others carry genes encoding surface proteins or enzymes involved in specific metabolic pathways, and others correspond to prophage-like elements. Accessory DNA regions encode 12 to 19% of the potential gene products of the analyzed strains. The analysis of a collection of epidemic *A. baumannii* strains showed that some islands were restricted to specific genotypes.

Conclusion: The definition of the genome components of *A. baumannii* provides a scaffold to rapidly evaluate the genomic organization of novel clinical *A. baumannii* isolates. Changes in island profiling will be useful in genomic epidemiology of *A. baumannii* population.

Background

The genus *Acinetobacter* comprises 26 species with valid names and nine genomic species with provisional designations that were defined by DNA-DNA hybridization. *Acinetobacter baumannii*, *A. pittii* and *A. nosocomialis* are the three species more frequently associated with human diseases [1-3]. *A. baumannii* is the species that is more frequently isolated in hospitalized patients, especially in intensive-care-unit (ICU) wards. The capability to survive in dry conditions and resistance to disinfectants and antimicrobial agents contribute to the selection of *A. baumannii* in the hospital setting [1,2].

Epidemics caused by multidrug-resistant (MDR) strains of *A. baumannii* were reported in several hospitals worldwide and shown to be caused by *A. baumannii* strains resistant to all classes of antimicrobials including carbapenems, exhibiting variable resistance to rifampicin and tigecycline, but still susceptible to colistin [2,4]. Outbreaks were caused by clusters of highly similar *A. baumannii* strains that were assigned by several genotypic methods to three main international clonal lineages initially named European clones I, II and III [1,2,4-6], and now are referred to as international clones I, II and III, respectively [7,8]. The predominance of international clone II lineage world-wide and the occurrence of hospital outbreaks caused by MDR strains belonging to novel genotypes not related to the three main clonal complexes have been reported during the last few years [4,8-10].

* Correspondence: dinocera@unina.it; rafzarri@unina.it
[1]Dipartimento di Biologia e Patologia Cellulare e Molecolare, Università Federico II, Via S. Pansini 5, 80131 Napoli, Italy
[2]Dipartimento di Scienze Mediche Preventive, Sezione di Igiene, Università Federico II, Via S. Pansini 5, 80131 Napoli, Italy
Full list of author information is available at the end of the article

We have recently reported [11] the draft genome sequences of three *A. baumannii* strains, 3990, 4190 and 3909, respectively assigned to ST (sequence types) 2, 25 and 78, which are representative of the most frequent genotypes responsible for epidemics occurred in Mediterranean hospitals [9]. Here we compare the genomes of the 3990, 4190 and 3909 strains and the genomes of four wholly sequenced MDR *A. baumannii* strains, two assigned to ST1, one each to ST2 and ST77. Data helped to define core and auxiliary genome components of the *A. baumannii* genomes.

Results

Features of the genome of ST2 3990, ST25 4190 and ST78 3909 strains

The draft genome sequences of the ST2 3990, ST25 4190 and ST78 3909 strains, isolated during cross-transmission episodes occurred at the Monaldi Hospital, Naples, Italy between 2006 and 2009, comprised 4,015,011 bases, 4,032,291 bases and 3,954,832 bases, and generated 3,806, 3,910 and 3,721 protein coding sequences by automated annotation against *A. baumannii* AB0057 genome, respectively [11]. The ST25 4190 strain contained two plasmids, p1-ABST25 and p2-ABST25, that both carry one copy of the carbapenem-hydrolyzing oxacillinase (CHDL) bla_{OXA-72} a gene which encodes a protein identical but for a single amino-acid substitution to the product of the bla_{OXA-24} gene. The ST2 3990 strain contained also two plasmids, p1-ABST2 carrying a complete *tra* locus, and p2-ABST2 carrying one copy of the CHDL bla_{OXA-58} gene. p1-ABST2 and p2-ABST2 were homologous to plasmids pACICU2 and pACICU1 identified in the ST2 ACICU strain [12], respectively. While p1-ABST2 and pACICU2 are almost identical, p2-ABST2 shares only two third of the coding sequences with pACICU1. The plasmid p1-ABST78 identified in the ST78 3909 strain shares approximately 80% of the coding sequences, including the bla_{OXA-58} gene, with plasmid pACICU1 (Additional files 1 and 2). The different plasmids were classified using the PCR-typing procedure recently described [13]. A conserved scaffold that includes four/five direct perfect repeats that can be defined as "iterons", and the gene encoding the replicase repAci1 belonging to the Rep-3 superfamily and assigned to the GR2 homology group, was found in plasmids pACICU1, p2ABST2, p2ABST25 and p1ABST78. The repAciX replicase (Rep-3 superfamily, GR10 homology group) is encoded by plasmids pACICU1 and p2ABST2, the Aci6 replicase (GR6 homology group) by pACICU2 and p1ABST2 plasmids. A protein identical to the replicase encoded by plasmid pMMA2 carrying the bla_{OXA-24} gene [14], is encoded by p1ABST25. While sharing common sequences, all plasmids exhibited a mosaic genetic structure that might have been generated by multiple recombination events. The hypothetical gene products encoded by the plasmids found in the *A. baumannii* strains 3990, 3909 and 4190 are listed in Additional file 2.

The *A. baumannii* chromosome

Making use of the Mauve software [15], the proteins putatively encoded by the draft genomes of the *A. baumannii* strains 3990, 3909 and 4190 [11] were compared to the ORFs encoded by the wholly sequenced genomes of the *A. baumannii* AB0057 and AYE strains assigned to ST1, ACICU strain assigned to ST2, ATCC17978 strain assigned to ST77 [12,16-18].

A. baumannii genomes exhibit extensive synteny. Sequence comparisons revealed that 3068 coding regions are conserved, at the same chromosomal position, in the compared *A. baumannii* genomes. A file including all conserved gene products is available upon request. Genes encoding proteins shown or hypothesized to be important for pathogenicity are conserved in the analyzed strains at the same relative chromosomal position (Table 1). The set includes OmpA, the outer membrane protein which has role in biofilm formation [19] and induces, when secreted, death of epithelial and dendritic cells [20], the DD-endopeptidase, which contributes to the resistance of *A. baumannii* to bactericidal activity presumably by remodelling the cell surface [21], phospholipase D, an enzyme crucial for proliferation in human serum [22], proteins involved in the formation of capsule [23], type I pili [24], and iron metabolism [25]. According to the published annotation, OmpA, DD-endopeptidase, phospholipase D, and many other deduced gene products are smaller in ATCC 17978 as compared to their orthologs. Size differences do not denote allelic variation, but are determined by the criteria adopted to select the initiating methionine in ATCC17978 ORFs.

Multidrug resistance is a key feature of *A. baumannii* and several genes have a role in establishing a MDR phenotype. Genes encoding efflux pumps and resistance proteins shown or hypothesized [26] to be involved in the process are conserved in all strains. In contrast, genes encoding drug-inactivating and drug-resistant enzymes reside in accessory DNA regions which are present only in some strains (Table 2). Among these, are worth of mention the extended spectrum beta-lactamase *VEB-1* gene, found in the AYE genome, the bla_{OXA-20} class D beta-lactamase gene, found in the ACICU and 3990 genomes, both assigned to ST2 genotype, the CHDL genes bla_{OXA-23}, found in the AB0057 genome, bla_{OXA-58}, found in the plasmids of 3990, ACICU and 3909 strains, and bla_{OXA-72} found in the plasmids of 4190 strain, respectively. Promoter sequences within flanking insertion sequences likely

Table 1 Gene products involved in pathogenicity in *A.baumannii* genomes

Gene products	Strains						
	AB0057	AYE	3990	ACICU	4190	ATCC17978	3909
capsule formation							
tyrosine kinase Ptk	91	3818	936	71	3295	49	2600
Tyrosine phosphatase Ptp	92	3817	935	72	3296	50	2601
type I pili formation							
CsuE	2565	1324	787	2414	3382	2213	744
CsuD	2566	1323	786	2415	3383	2214	745
CsuC	2567	1322	785	2416	3384	2215	746
CsuB	2568	1321	784	2417	3385	2216	747
CsuA	2569	1320	783	2418	3386	2217	748
CsuA/B	2570	1319	782	2420	3387	2218	3415
iron metabolism							
nonribosomal peptide synthetase BasD	2811	1095	2421	2579	tblastn	2383	1389
nonribosomal peptide synthetase BasC	2812	1094	2420	2580	3813	2384	tblastn
ferric acinetobactin receptor	2813	1093	2419	2581	3814	2385	3376
ferric acinetobactin transport system periplasmic binding protein	2814	1092	2418	2582	3815	2386	3375
ferric acinetobactin transport system ATP-binding protein	2815	1091	2417	2583	3816	2387	3374
ferric acinetobactin transport system permease	2816	1090	2416	2584	3817	2388	3373
ferric acinetobactin transport system permease	2817	1089	2415	2585	3818	2389	3372
hemin utilization							
biopolymer transport protein ExbD/TolR	1827	2051	351	1629	227	1063	1994
biopolymer transport protein ExbD/TolR	1828	2050	352	1630	228	1064	1993
biopolymer transport protein	1829	2049	353	1631	229	1065	1992
TonB family protein	1830	2047	354	1632	230, 231	3708*	1991
TonB-dependent receptor	1831	2046	355	1633	232	1606, 1607	1990, 1989
heme-binding protein A	1832	2045	358	1634	234	1608	1987
heme-binding protein A	1833	2044	359	1635	235	1609	1986
Zn-dependent oligopeptidase	1834	2043	360	1636	236	1610	1985
ABC-type dipeptide/oligopeptide/nickel transport system permease component	1835	2042	361	1637	237, 238	1611	1984
ABC-type dipeptide/oligopeptide/nickel transport system permease component	1836	2041	362	1638	239	1612	1983
glutathione import ATP-binding protein GsiA	1837	2040	363	1639	3719	1613	1982

* The asterisk indicates one of the 436 proteins putatively encoded by ATCC 17978 not included in the GenBank NC_009085 file. tblastn refer to unannotated 4190 and 3909 proteins identified by tblastn searches.

influence the expression of many of these resistance genes. Interestingly, the majority of the genomes harbour mutations in *gyrA* and/or *parC* genes.

Shared synteny lets to represent the *A. baumannii* chromosomes as ~4 Mb long DNA segments homologous to each other throughout their lengths (Figure 1). DNA tracts, ranging in size from 4 to 126 kb, are present in one or more strains, but missing or replaced by alternative DNA segments in others (see vertical bars in Figure 1). Some of these regions correspond to DNA sequences earlier suspected to be mobile because found in *A. baumannii* but not in *A. baylyi* DNA or vice versa [17,27]. Specific 15-36 kb regions are missing in all strains but AB0057 (see triangles in Figure 1), and may therefore plausibly correspond to strain-specific deletions. Many of the accessory genomic DNA segments exhibit characteristic features of genomic islands, such as the presence of insertion sequences at one end, a GC content different from the bulk chromosome, insertion within tRNA or non-coding RNA genes, target site duplications (TSDs) at the ends formed upon genome integration [28,29]. For sake of simplicity, all the accessory DNA regions have been called GEnomic Islands (GEIs). GEIs found at the 63 variable loci identified in the *A. baumannii* genomes, and some of their properties, are diagrammatically reported in Figure 2. TSDs flanking GEIs are reported in Additional file 3, and GEI gene products are listed in Additional file 4. In text and figures individual GEIs are referred by the locus number and the strain acronym used in Figure 2. Core and accessory chromosomal DNAs are fully conserved in ACICU and 3990 strains. Because of this, only the ACICU GEIs are shown in Figure 2. In draft genomes some GEIs reside in different contigs. The colinearity of

Table 2 Antimicrobial resistance gene products encoded by *A.baumannii* genomes

Gene Products	Strains						
	AB0057	**AYE**	**3990**	**ACICU**	**4190**	**ATCC17978**	**3909**
Class C β-lactamase	9, 2796	1110	2437	2564	2076	2367	1404
Class A β-lactamase	283 (TEM-1)	-	-	-	-	-	-
	-	3623 (VEB-1)	-	-	-	-	-
Class D β-lactamase	1757 (oxa-69)	2122 (oxa-69)	2827 (oxa-66)	1560 (oxa-66)	63 (oxa-64)	1517 (oxa-95)	1089 (oxa-90)
	-	-	3514 (oxa-20)	0226 (oxa-20)	-	-	-
	0551 (oxa-23)*	-	-	-	-	-	-
	-	-	p2ABST2 (oxa-58)*	pACICU1 (oxa-58)* (2X)	p1ABST25 (oxa-72)*	-	p1ABST78 (oxa-58)*
	-	-	-	-	p2ABST25 (oxa-72)*	-	-
AAC (3)-I aminoglycoside acetyltransferase	291	3573	-	-	-	-	-
AAC (6')-I aminoglycoside acetyltransferase	-	3630	3516	223	-	-	-
APH (3')-I aminoglycoside phosphotransferase	288	3578	-	-	-	-	-
	-	-	3897	1948	560	-	-
ANT (3")-I aminoglycoside adenylyltransferase	293	3570,3618	-	-	3268	-	-
	171	3739	1641	156	2954	131	2919
Chloramphenicol acetyl transferase	280	3587	-	-	-	-	-
	3104	798	3709	2932	1731	2691	1443
DNA topoisomerase II	3037 [R^1]	0867 [R^1]	0747 [R^1]	2869 [R^1]	2907 [R^1]	2626 [S]	0539 [R^1]
DNA topoisomerase IV	0232 [R^2]	3679 [R^2]	1415 [S]	0214 [S]	2382 [R^2]	-	3413 [R^2]
RNA polymerase β Subunit	0369 [S]	3489 [S]	2179 [R^3]	0303 [S]	3155 [S]	0287 [S]	0411 [S]
Dihydropteroate synthase	265, 294	3568,3616,3612	3142	228	-	675	-
	3095	807	3700	2923	2684	2680	1433
Dihydrofolate reductase type 1	-	3644	-	-	-	-	-
Dihydrofolate reductase type 3	540	3315	3351	467	3501	457	403

R, resistant; S, susceptible; R^1 81 Ser →Leu; R^2 84 Ser → Leu; R^3 535 His → Leu; * carbapenem-hydrolysing class D beta-lactamase; + ORFs identified by tBLASTn.

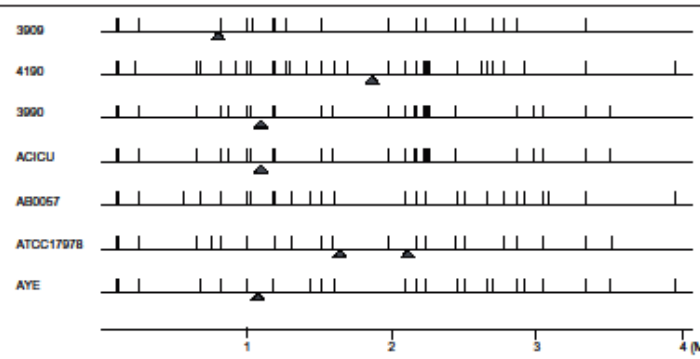

Figure 1 Comparison of *A. baumannii* genomes. The seven *A. baumannii* genomes analyzed have been aligned. Accessory regions are denoted by vertical bars. Strain-specific deletions are marked by triangles.

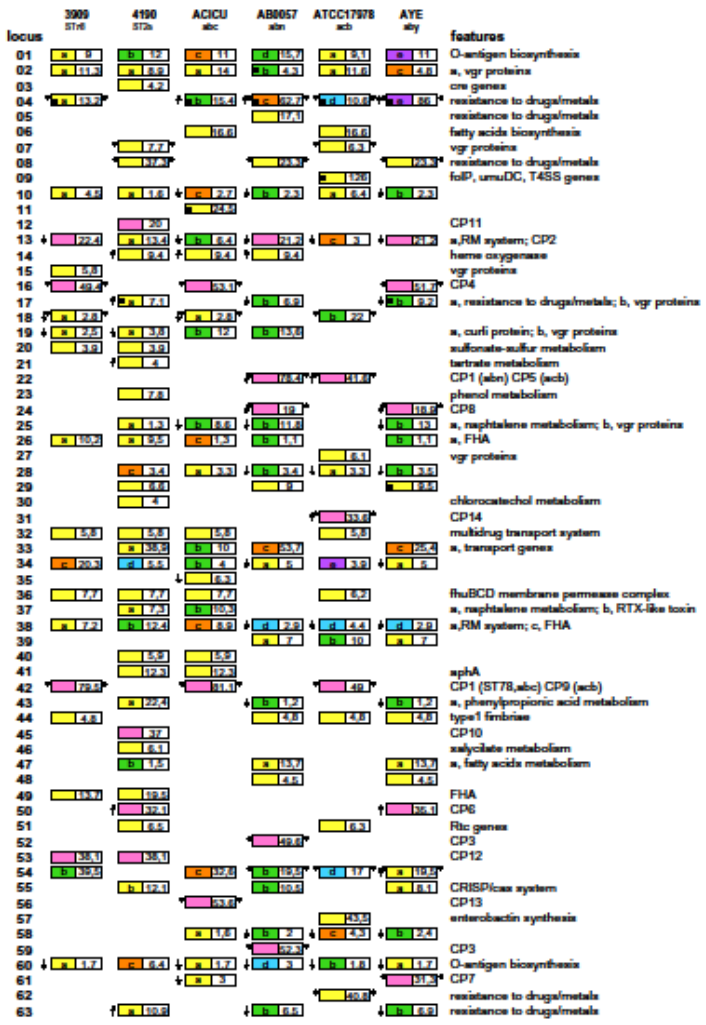

Figure 2 Variable regions in *A. baumannii* genomes. A chart of the genomic islands (GEIs) depicted as bars in Figure 1 is displayed. Each line corresponds to a chromosomal locus. Different GEIs inserted at the same locus in different strains are marked by different colours and lower case letters. Sizes of GEIs are given in kb. Black boxes within GEIs denote mobile sequences, down and up arrows to the left indicate that the GEI G+C content is lower than 36% or higher than 42%, respectively. Dots flanking GEIs denote TSDs. The strain names and relative acronyms used throughout the text are given at the top. Acronyms below complete genomes are those used at Kyoto Encyclopaedia of Genes and Genomes (KEGG).

the contigs and the GEI DNA content of the corresponding chromosomal regions were assessed by sequencing PCR products bridging contigs ends.

A close look at *A. baumannii* chromosomes further identified about one hundred DNA regions encoding 1-2 ORFs smaller than 4 kb conserved in one or more strains, but missing, or replaced by non homologous DNA of comparable length, in others. The potential gene products encoded by these smaller accessory regions, that we called *mhrs* (for micro-heterogeneity regions), are reported in Additional file 5.

Categories of genomic islands

Some islands are strain-specific; others are completely or partially conserved in more than one strain. Non homologous islands are inserted at the same locus in different strains, and some loci are extremely heterogeneous, featuring up to 4-5 alternative islands. Some islands are composite, and changes in their organization among strains are correlated to changes in the number and association of specific DNA segment. Thus, for example, $G54_{ST78}$ can be viewed as made by ABC segments. Segments AB are missing in $G54_{acb}$, segments AC in both $G54_{abn}$ and $G54_{aby}$, and segment C is replaced by a shorter DNA segment in $G54_{acb}$ (see Additional file 4 for a direct G54 islands comparison). On the basis of the putative gene products, GEIs can be broadly sorted into a few categories. Properties and overall organization of relevant GEIs are below discussed.

Resistance islands

Many of the accessory drug resistance determinants of Table 2 found in AB0057 and AYE are encoded by genes located within $G4_{aby}$, $G4_{abn}$ and $G5_{abn}$, which correspond to the resistance regions previously described as AbaR1, AbaR3, and AbaR4 [16,30], respectively. $G4_{aby}$ and $G4_{abn}$ are both inserted in the *comM* gene, and result from the association of the 16 kb Tn6019 transposon with multiple antibiotic resistance regions (MARR), which are delimited by Tn6018 elements [30]. Tn6019 features genes involved in transposition (*tniA*, *tniB*), an arsenate resistance operon, a universal stress protein gene (*uspA*), and a sulphate permease gene (*sup*). MARR are inserted within *uspA* and vary in length and composition [30]. The $G4_{abc}$ island of the ACICU genome corresponds to the AbaR2 region [30], which carries few resistance genes and lacks Tn6019 sequences (Figure 3A). $G4_{ST78}$ is similarly inserted in the *comM* gene, and features genes homologous to *tniA* and *tniB* (38-40% identity of the gene products), but lacks resistance genes and encodes a set of hypothetical proteins (Figure 3A). G4 is missing in strain 4190. However, resistance genes are scattered in different GEIs of this strain (Figure 3B). The *aad*A1 (streptomycin 3"-adenylyltransferase) gene, flanked by *satR* (streptothricin acetyltransferase) and *dhfr* (dihydrofolate reductase) genes are found in $G63_{ST25}$. Genes involved in resistance to mercury (*merRCAD* cluster) are located in $G17_{ST25}$, and a 4.5 kb DNA segment containing *feo*AB (ferrous iron transport operon), *czc* (tricomponent proton/cation antiporter efflux system) and *ars* (arsenite transporters) genes are found in $G8_{ST25}$, next to the *cus* (copper resistance) genes conserved in all G8 (Figure 3B). The $G62_{acb}$ region also contains *cus*, *feo* and *czc* genes involved in heavy metal resistance. These genes differ in sequence and overall arrangement from $G8_{ST25}$ homologs. This supports the notion that the set of accessory genes had been independently acquired by the strains 4190 and ATCC17978.

Additional resistance genes found in GEIs include an aminoglycoside phosphotransferase gene ($G41_{ST25}$, $G41_{abc}$), a dihydropteroate synthase gene ($G9_{acb}$), and an ABC-type multidrug transport system, conserved in all the G32 islands.

GEIs encoding surface components and transport systems

GEI-1 and GEI-60 host genes involved in cell envelope. Heterogeneity among *A. baumannii* strains at the level of O-antigen biosynthetic genes was already noticed (16), and is correlated to the presence of alternative glycosylases. The G44 island, present in all strains but ACICU, 3990 and 4190, is a four gene operon involved in the assembly of fimbriae (type I pili) by the chaperone/usher pathway [31]. $G44_{aby}$ corresponds to the surface adhesion protein region annotated as CusIR in the AYE genome [18]. $G19_{ST25}$ and $G19_{ST78}$ are related islands which both carry an operon encoding three hypothetical lipoproteins. Of these, one exhibits homology to CsgG, the key factor in the secretion of curli, the proteinaceous component having a role in host cell adhesion and biofilm formation in many *Enterobacteriaceae* [32]. Purified CsgG forms ring-shaped complexes analogous to those formed by outer membrane channel-forming proteins [32]. The CsgG-like protein, in association with the two co-expressed lipoproteins, may influence the permeability of the outer membrane of *A. baumannii*.

Filamentous haemagglutinin (FHA) is a major virulence factor in *Bordetella pertussis* [33]. *fhaB* and *fhaC* genes, respectively encoding the haemagglutinin and the transporter protein, have been identified in many pathogens [34]. *fhaBC* gene clusters are found at the same loci in strains 4190 and 3909 (islands $G26_{ST25}$, $G26_{ST78}$, $G49_{ST25}$ and $G49_{ST78}$), and strains ACICU and 3990 (islands $G38_{abc}$ and $G38_{ST2}$). The transporter proteins are highly conserved in the four clusters, whereas FHAs vary in length (1834 to 4812 amino acids), mostly because of changes in the number and organization of

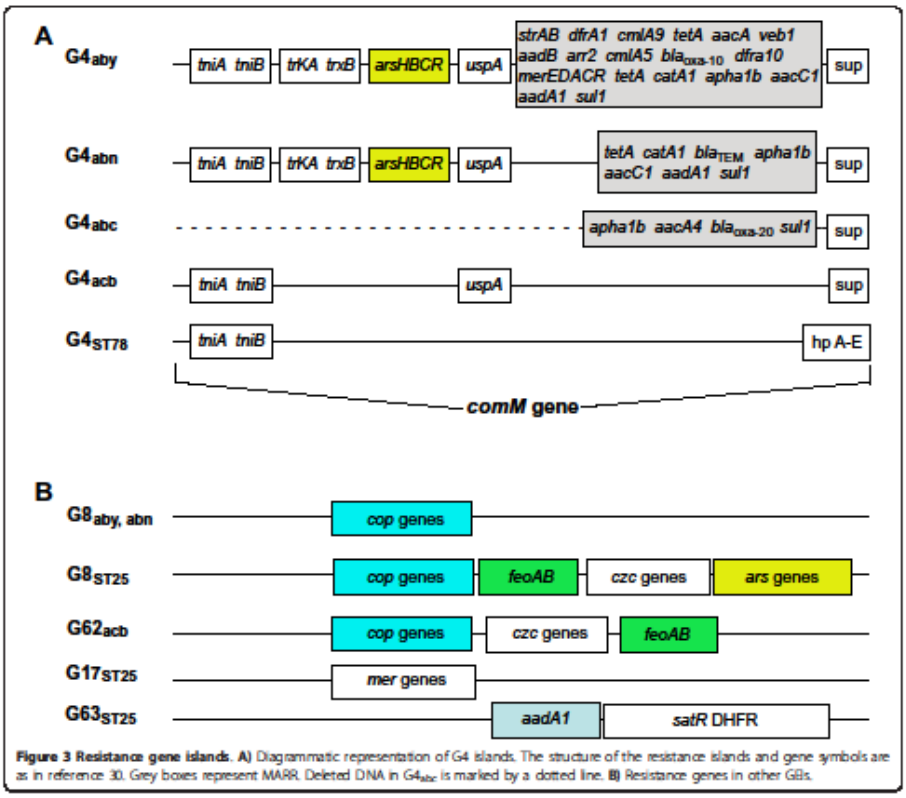

Figure 3 Resistance gene islands. A) Diagrammatic representation of G4 islands. The structure of the resistance islands and gene symbols are as in reference 30. Grey boxes represent MARR. Deleted DNA in $G4_{abc}$ is marked by a dotted line. **B)** Resistance genes in other GEIs.

body sequence repeats [33]. A 3216 amino acids long calcium binding hemolysin protein, unrelated to FHAs, is encoded by $G18_{acb}$.

Cyclopropane fatty acids (CFA) are phospholipids found in the bacterial membranes in the late exponential and early stationary phases of cell growth [35], which derive from the corresponding unsaturated fatty acid (UFA) phospholipids. The synthesis of CFA is catalyzed by the enzyme CFA synthase, the substitution of a saturated by an unsaturated fatty acid by the enzyme delta-9 acyl-lipid desaturase. CFA synthase and delta-9 acyl-lipid desaturase are both encoded by $G47_{abn}$ and $G47_{aby}$.

$G33_{ST25}$ is a large island which encodes four different transport and translocation systems: i) Tat (twin-arginine translocation) proteins, involved in the translocation of folded proteins to the cell envelope or the extracellular space ii) a TonB/ExbBD complex iii) a Opp (oligopeptide transport proteins) complex iv) a sulfur utilization system, made by a FMNH2-dependent sulfonatase and three ABC-type transporters, which resemble the products of the *E. coli ssu* gene cluster [36]. Two unlinked copies of the sulfonatase gene are also present. Genes involved in the capture and intracellular transport of iron are found in different islands. $G57_{abc}$ carries a gene cluster involved in the synthesis of the high-affinity siderophore enterobactin. Heme oxygenase is an alternative to siderophores to capture iron from the environment [37]. G14, an island which is conserved in 4190, ACICU and AB0057, carries an operon encoding a heme oxygenase, an outer membrane and a TonB family protein. The presence of a flanking *fecIR* gene cluster suggests that heme internalization may be regulated by the Fec transduction system [38]. The *fhu*BCD genes,

which catalyze the internalization of iron III hydroxamate compounds, are located on G36, an island conserve in all strains but AB0057 and AYE.

Metabolic islands

Many GEIs carry genes encoding proteins involved in specific metabolic pathways. $G23_{ST25}$ carries a *mph* (multi component phenol hydroxylase) gene complex, involved in the conversion of phenol to cathecol, flanked by a sigma54-dependent activator gene. It has been shown that the expression of *mph* gene complex described in *Acinetobacter sp.* PHAE-2 is dependent on the alternative sigma factor RpoN [39]. $G37_{ST25}$ carries *nag* genes, involved in the metabolism of naphthalene. In *Ralstonia* [40], *nag* genes are arranged in two separate clusters, involved in the conversion of naphthalene to gentisate (*nagAGHBFCQED* genes), and gentisate to pyruvate and fumarate (*nagIKL* genes), respectively. In $G37_{ST25}$ *nagIKL* genes and *nagGH*, encoding the salicylate 5-hydroxylase, are linked, and flanked by benzoate transport genes.

$G43_{ST25}$ carries genes involved in the catabolism of 3HPP (3-hydroxyphenylpropionic acid) and PP (phenylpropionic acid). In *E. coli*, the dioxygenase complex (*hcaEFCD* genes), and the dihydrodiol dehydrogenase (*hcaB* gene) oxidize PP (phenylpropionic acid) and CI (cinnamic acid) to DHPP (2,3-dihydroxyphenylpropionate) and DHCI (2,3-dihydroxycinnamic acid), respectively. These substrates are subsequently converted to citric acid cycle intermediates by the *mhp* genes products [41]. The *hca* and *mhp* genes, separated in *E. coli*, are linked and interspersed with additional genes (see Additional file 4) in $G43_{ST25}$. $G21_{ST25}$ potentially encodes 4 proteins (tartrate dehydratase subunits alpha and beta, a MFS transporter and a transcriptional regulator) possibly involved in the metabolism of tartrate. Proteins exhibiting homology to the dienelactone hydrolase, an enzyme which plays a crucial role in the degradation of chloro-aromatic compounds, are encoded by the islands $G30_{ST25}$, $G34_{abn}$ and $G34_{aby}$. $G46_{ST25}$ is made by an operon including the salicylate 1-monooxygenase (*salA*), a benzoate transporter (*benK*) and the *sal*A regulator (*salR*) genes. A salicylate 1-monooxygenase is also encoded by $G25_{ST25}$. The genes *fabA*, *fabB*, *fabG*, *fabF*, *acpP*, *pslB*, *acsA*, involved in the biosynthesis of fatty acids [35] are conserved in all *A. baumannii* strains, at separate loci. Orthologues of all these genes are clustered in $G6_{abc}$ and $G6_{acb}$.

Phage islands

Many variable genomic regions are relatively large (19 to 82 kb) DNA blocks which potentially encode typical phage products. These regions have all been classified as cryptic prophages (CP; see Figure 2). Three to six CPs were identified in each strain. Six of the different 14 CPs identified are present in two or more strains, the remaining 8 are strain-specific. CPs characteristically carries at one end an integrase gene, and many are sharply defined by flanking TSDs induced upon insertion. CPs are poorly related to each other, and even CPs of the same type differ in size and coding ability. Ten of 14 CPs were assigned to four groups on the basis of sequence homologies (Additional file 6). CPs found at the same locus encode identical or highly homologous (> 80% identity) integrases. CP1 encode different integrases, which are homologous to CP5- or CP9-encoded enzymes. This explains why CP1 and CP5 in AB0057 and ATCC17978 ($G22_{abn}$ and $G22_{acb}$, respectively), and CP1 in 3909 and ACICU ($G42_{ST78}$ and $G42_{abc}$), and CP9 in ATCC 17978 ($G42_{acb}$), are inserted at the same locus. CP3 are integrated at different sites of the AB0057 genome ($G52_{abn}$ and $G59_{abn}$), but the target in both is an *arg-tRNA* gene.

Remnants of prophage sequences are found in $G33_{abn}$ and $G33_{aby}$. These islands share the $G33_{abc}$ backbone, but contain also large DNA segments, reiterated in a head-to-tail configuration, in which genes encoding phage and hypothetical proteins are variously interleaved. $G33_{abn}$ and $G33_{aby}$ hypothetical gene products exhibit poor homology to all CPs gene products, and therefore were not included among CPs.

Phages may acquire ORFs named morons [42] by lateral gene transfer. The PapS reductase (3'-phosphoadenosine 5'-phosphosulfate sulfotransferase) encoded by CP13 ($G56_{abc}$), the toxin-antitoxin (TA) system encoded by CP1 ($G42_{abc}$ and $G42_{ST78}$), the proofreading 3'-5' exonuclease epsilon subunit of the DNA polymerase III in the above mentioned CPs, the *umuDC* gene products, which are the components of the error-prone DNA polymerase V, again in CP1 ($G22_{abn}$ and $G42_{ST78}$) and CP5 ($G22_{abc}$) can all be considered morons. Not surprisingly, these enzymes are frequently associated with mobile genome elements [43]. Unlinked *umuD* and *umuC* genes are conserved in all *A. baumannii* strains, and an *umuDC* cluster resides on the 64 Kb pACICU2 plasmid.

$G9_{acb}$ also contains an *umuDC* cluster. This 126 kb region, found only in the ATCC 17978 strain, is a composite genomic island, carrying at one end a dihydropteroate synthase gene, at the other a DNA mismatch repair enzyme. $G9_{acb}$ carries a complete set of type IV secretion system (T4SS) genes, arranged in the same order in which T4SS homologs are found on the 153 Kb plasmid of *Yersinia pseudotuberculosis* IP31758 strain [44]. Because *umuDC* genes are carried by this plasmid, one may hypothesize that raises $G9_{acb}$ had been imported from *Yersinia*. In addition, a $G9_{acb}$ gene cluster, including an integrase, a DNA helicase and a TrbL/

VirB6 conjugal transfer protein is highly homologous to a gene cluster from *Enterobacter cloacae*.

Additional islands

$G3_{ST25}$ carries a *cre* genes cluster. In *E. coli* the *cre* locus includes a response regulator (*creB*) a sensor kinase (*creC*) and an inner membrane protein (*creD*). The corresponding two-component regulatory system CreB-CreC controls the expression of a variety of genes, among which the *creD* regulator. Overexpression of CreBC causes modification of the envelope, inducing the colicin E2 tolerance phenotype [45].

$G51_{ST25}$ and $G51_{acb}$ carry the *rtcA* and *rntZ* genes, encoding the RNA 3'-terminal phosphate cyclase and the RNAseZ, respectively. The cyclase catalyzes the ATP-dependent conversion of the 3'-phosphate to the 2', 3'-cyclic phosphodiester at the end of various RNA substrates [46]; RNAseZ is responsible for the maturation of the 3'-end of a large family of transfer RNAs [47]. In *E. coli* the 3'-terminal phosphate cyclase *rtcA* gene forms an operon with the upstream *rtcB* gene. Expression of *rtcAB* is regulated by *rtcR*, a gene positioned upstream of *rtcAB*, but transcribed in the opposite direction, encoding a sigma54-dependent regulator [46]. *rtcBA* and *rtcR* genes are conserved in both $G51_{ST25}$ and $G51_{acb}$ islands, separated by *rntZ*. Interestingly, only *rntZ* is present at the corresponding chromosomal position in strains lacking G51.

In type I restriction systems the three subunits S, M and R, which may variably associate to form a modification methylase or a restriction endonuclease, are encoded by *hsd* (host specificity of DNA) genes. Alternative *hsd* genes reside in $G13_{ST25}$ and $G13_{ST78}$. The former are clustered in one operon, whereas *hsdSM* and *hsdR* genes in $G13_{ST78}$ are at distance, as frequently found in other species.

Homologs of a cytosine DNA methyltransferase and a restriction endonuclease, which may constitute a type II restriction modification system, are encoded by genes residing in $G38_{ST78}$.

The G55 islands found in strains 4190, AB0057 and AYE are closely related, and all include a CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) block, flanked by a *cas* (CRISPR-associated) gene cluster. CRISPRs are repeated DNA sequence blocks found in the genomes of approximately 40% of bacteria, often next to a cluster of *cas* genes. The CRISPR/Cas system provides a form of acquired immunity against exogenous DNA, foreign DNA sequences being first integrated at the CRISPR locus and eventually degraded by Cas proteins [48]. Horizontal transfer of CRISPRs and associated genes among prokaryotes is documented [49].

Gram-negative bacteria contain a variety of genes encoding proteins enriched in dipeptide motifs (valine-glycine repeats) hence called Vgr. Islands encoding Vgr-like proteins are found inserted at eight genome variable loci (loci 2, 7, 15, 17, 19, 25, 27 of Figure 2). Vgr proteins are associated with ligand-binding proteins at the bacterial surface [50], and are involved in biofilm formation and swarming and swimming motility in *Burholderia* [51]. Intriguingly, Vgr proteins, along with Hcp (hemolysin co-regulated) proteins, are components of the type VI (T6SS) secretion apparatus, a transport system extensively conserved among Gram-negative bacteria [52]. Secreted Vgr proteins assemble a cell-puncturing device analogous to phage tail spikes to deliver effector proteins, and are also able to covalently cross-link host cell actin contributing to T6SS pathogenicity [53]. A T6SS gene cluster is conserved in all the analyzed *A. baumannii* strains.

A. baumannii GEIs in other species of the *Acinetobacter* genus

Acinetobacter baylyi is a non-pathogenic nutritionally versatile soil bacterium. The chromosome of the *A. baylyi* strain ADP1 carries metabolic genes involved in the utilization of a large variety of compounds. Most of these genes are clustered in five major catabolic islands, grouped in the so called archipelago of catabolic diversity [27]. The organization of the *A. baylyi* and *A. baumannii* chromosomes is different, and most catabolic islands of *A. baylyi* are conserved in all *A. baumannii* strains, although ungrouped, at separate loci (Figure 4). Interestingly, some archipelago genes were found in $G33_{ST25}$ and $G46_{ST25}$, two accessory DNA regions specific of the *A. baumannii* strain 4190. Prompted by this finding, we checked whether twenty GEIs, including $G33_{ST25}$ and $G46_{ST25}$, were present in *A. baylyi* (GenBank: NC_005966), in the complete genome of the diesel-degrading *Acinetobacter* sp. strain DR1 (GenBank: NC_014259) [54] and in the nine draft genomes of the *Acinetobacter* genus deposited at Genbank. GEIs encoding filamentous haemagglutin and vgr-proteins, as those corresponding to cryptic prophages were not searched because of their heterogeneity. The results of the survey are summarized in Table 3. Seven islands (GEIs 14, 20, 21, 23, 29, 44, 51) are conserved in one or more genomes, flanked at one or both sides by the same genes found in *A. baumannii*, but their dimensions vary, as consequence of gain/loss of DNA segments. As expected for mobile DNA, some islands were missing, and only flanking genes could be identified (genomic empty sites). Segments of $G13_{ST25}$ and $G43_{ST25}$ are spread among non-*baumannii Acinetobacter* genomes, thus suggesting that both GEIs might result from multiple recombination events. Recombination likely contributed to the formation of the large DR1 island encompassing genes found in $G37_{ST25}$ and $G37_{abc}$, two non-

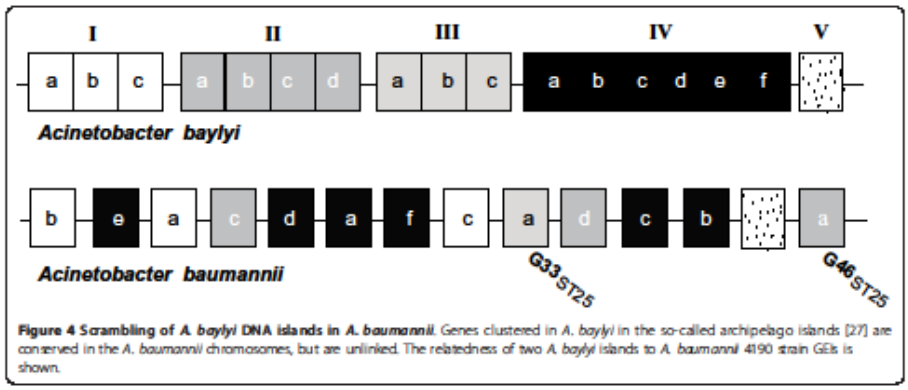

Figure 4 Scrambling of *A. baylyi* DNA islands in *A. baumannii*. Genes clustered in *A. baylyi* in the so-called archipelago islands [27] are conserved in the *A. baumannii* chromosomes, but are unlinked. The relatedness of two *A. baylyi* islands to *A. baumannii* 4190 strain GEIs is shown.

homologous GEIs encoding enzymes involved in naphthalene degradation and a RTX-type toxin. Curiously, the two *A. baumannii* islands are separated in the DR1 island by 10 kb DNA homologous to *fha*BC genes found in $G38_{abc}$.

A high number of GEIs is conserved in the genome of the *Acinetobacter* sp. strain DR1. Interestingly, dot plot analyses showed that gene order is more similar between *A. baumannii* AB0057 strain and *Acinetobacter* sp. strain DR1 than between the same *A. baumannii* strain and *A. baylyi* (Figure 5). According to *rpoB* sequence analysis, DR-1 strain belongs to the *A. calcoaceticus-A. baumannii* complex, and is closely related (99.7% identity) to gen. sp. "Between 1 and 3" [3].

Genomic regions in *A. baumannii* strains of different genotypes

The distribution of 18 genomic islands in the *A. baumannii* population was monitored by PCR analyses. Coding DNA regions of 600-1500 bp, representative of each GEI, were amplified from the DNA of 23 *A. baumannii* strains associated with 21 epidemics that occurred in 14 hospitals of the Mediterranean area from 1999 to 2009, including the sequenced 3909 and 4190 strains used as control. Nearly all the strains were representative of cross-transmission episodes, and were isolated with identical PFGE types from more than two patients of the same or different institutions [9]. Strains belong to eight different STs, and 10/23 strains are ST2. PCR data are summarized in Table 4. Taking into account that negative data may denote partial island deletion or polymorphism in sequences targeted by the primers, the conservation of islands seems to vary significantly among the analyzed strains. G43 and G51 had been found in most strains but not in the two strains assigned to ST78 and some strains assigned to ST2. In contrast, G18 is missing in all except one ST25 strain (4190), and G57 is found only in 3 strains of different STs. G47 seems to be a marker of ST1 strains, being found in all 5 strains assigned to ST1, as well in two strains assigned to ST20, which is a single locus variant of ST1. Similarly, G6 and G11 seem to be markers of ST2 strains, being found in all 10 ST2 strains. Interestingly, the three islands are also present in the single ST3 strain analyzed. G37 is also found in all ST2 strains and also in strains assigned to ST3, ST15 and ST84. G32 is found in all but not ST1 and ST20 strains. All the eleven islands found in the genome of the 4190 strain are conserved in the other two ST25 strains analyzed, with the exception of G8 and G63, both missing in the 3890 strain. Of the eleven islands, two (G23 and G46) are found only in the 3 ST25 strains, six (G3, G8, G63, G43, G21, G51) are also present in strains assigned to other STs. No correlation was found between the pattern of island distribution and PFGE profile among strains.

Discussion

Data reported are in line with the results of previous analyses [16], indicating that the genomes of *A. baumannii* strains isolated from geographically different regions are closely related and share the same overall organization. Shared synteny made possible to align the seven *A. baumannii* genomes throughout, and obtain a robust chromosomal scaffold by which easily distinguish core and accessory genome components in each strain.

The ST2 strains ACICU and 3990 exhibit 99.9% sequence identity, and share the same core and variable genome components. Mapped differences are restricted to size changes of ~40 intergenic regions, which vary in

Table 3 Distribution of genomic regions in non-*baumannii* *Acinetobacter* species

A. baumannii GEIs	ORF contained	*A.baylyi* ADP1	*A.calcoaceticus* RUH2202	*A. haemolyticus* ATCC19194	*A. johnsonii* SH046	*A. junii* SH205	*A. lwoffii* SH145	*A. radioresistens* SK82	*Acinetobacter sp.* ATCC27244	*Acinetobacter sp.* DR1	*A. nosocomialis* RUH2624	*A. pittii* SH024
G13 (ST25)	**[A to L]**	-	C	HL	HL	AB	HL	-	EFG_HL	C	C	CDE HL
G14	**[A to H]**	**[A to H]**	**[A to H]**	-	-	-	-	-	**[A to H]**	**[A to H]**	[ ]	**[A to H]**
G18 (ST78)	**[AB]**	-	-	-	-	-	-	-	-	**[AB**	-	**[AB**
G20 (ST78)	**[A to C]**	[ ]	-	BC	-	-	-	-	[ ]	**[A to C]**	[ ]	**[A to C]**
G21 (ST25)	**[A to E]**	-	[ ]	-	-	-	-	-	-	**[A to E]**	[ ]	**[A to E]**
G23 (ST25)	**[A to H]**	-	[ ]	-	-	-	-	A to G	-	**[A to H]**	**[A to H]**	**[A to H]**
G29 (ST25)	**[A to D]**	-	-	-	-	-	-	-	-	**[A to D]**	**[A to D]**	[ ]
G33 (ST25)	**[A to AF]**	**[A to AF]**#	-	-	-	-	-	-	-	**[A to AF]**#	-	-
G35 (abc)	**[A to N]**	-	-	-	-	-	-	-	-	[novel GEI]	-	C, N
G36	**[A to I]**	-	**[A to I]**#	-	-	-	-	-	-	**[A to I]**#	[A to F	[C to F
G37 (ST25)	**[A to H]**	-	-	B to E	-	-	-	-	-	**[A to H]**#	[ ]	[ ]
G37 (abc)	**[A to G]**	-	-	-	-	-	-	-	-	**[A to G]**#	[ ]	**[A to G]**#
G43 (ST25)	**[A to V]**	-	-	-	-	-	FG_H to V	FG_H to V	-	-	A to C	-
G44	**[A to D]**	-	[A to D]	-	-	-	-	-	-	[A to D	A to D	**[A to D]**
G46 (ST25)	**[A to E]**	CDE	CDE	-	-	-	-	-	-	CDE	[ ]	[ ]
G47 (abn, aby)	**[A to R]**	BL	-	L	-	BL	-	-	-	**[B to R]**	**[B to R]**	**[B to R]**
G51 (abc)	**[A to G]**	-	**[A to G]**	**[A to G]**	-	**[A to G]**	-	-	B to L	**[A to G]**	C	**[A to G]**
G57 (acb)	**[A to H]**	M to AG	-	-	-	[ ]	-	-	-	[ ]	[ ]	-

ORFs in each island are referred to by capital letters. Brackets denote ORFs flanking genomic islands. Conserved genomic regions are highlighted in bold. Dots between letters denote that corresponding ORFs are not contiguous. #Genomic regions larger than those identified in *A. baumannii*.

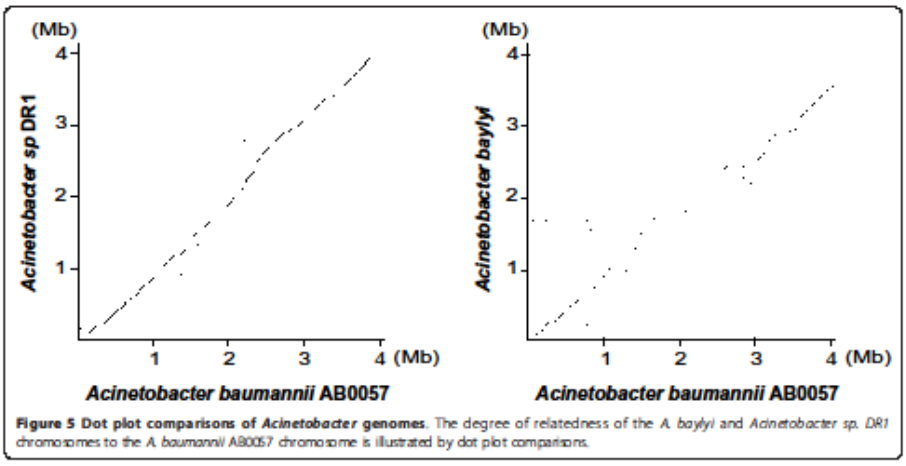

Figure 5 Dot plot comparisons of *Acinetobacter* genomes. The degree of relatedness of the *A. baylyi* and *Acinetobacter sp. DR1* chromosomes to the *A. baumannii* AB0057 chromosome is illustrated by dot plot comparisons.

the two strains because they contain a different number of short sequence repeats. A major difference can be ascribed to a > 36 kb CP3-like element, found in the 3990 strain only, the chromosomal location of which has not yet been determined. Two CP3-like prophages specific of strains 3909 and 4190 have not yet been mapped as well. The ACICU and 3990 strains are however phenotypically distinguishable, since the his-leu replacement at residue 535 of the *beta* subunit of the RNA polymerase made the 3990 strain not susceptible to rifampicin (MIC > 500 mg/L). Sequence comparisons revealed that 3068 coding regions are conserved, at the same chromosomal position, in all *A. baumannii* genomes. Accessory coding regions, including both GEI- and mhr-encoded ORFs, varies from 433 (3909 strain) to 707 (AB0057 strain). In estimating the number of conserved coding regions, it was taken into account that many correspond to a single ORF in one genome, but to two or even three adjacent ORFs in others, and vice versa. Likely most "double ORFs" are artifactual, since mutations are known to be introduced by PCR amplification of DNA samples prior to sequencing. Accessory DNA regions correspond to 12% of the 3909 genome, 19% of the AB0057 genome, and to 14-16% of all other genomes analysed. Although closure of draft genomes and addition of whole genome sequences of other strains may lead to the definition of a few additional GEIs, data clearly indicate that *A. baumannii* strains exhibit less variation than *E. coli* strains, which may share only 60-70% of their coding capacity [55].

Many *A. baumannii* GEIs have a role in drug resistance, biosynthesis of surface components, iron metabolism, and this may confer advantage in the course of an infection, since successful pathogens encode multiple adhesins, are equipped to sequester iron from the environment and can escape therapy. Less clear is the advantage conferred to *A. baumannii* by other islands. The functional role of the RNA 3'-terminal phosphate cyclase, an enzyme conserved among Bacteria, Archaea and Eucarya, encoded by $G51_{ST25}$ and $G51_{acb}$, is debated. The same holds for vgr-like proteins, encoded by several GEIs, though it is worth noting that six of the ten genomic islands identified in the pathogenic *P. aeruginosa* PA01 strain [56] encode vgr-like proteins. Some GEIs carry genes involved in lipid metabolism. $G47_{abn}$ and $G47_{aby}$ carry genes controlling the formation of CFA and UFA phospholipids. Cyclopropanation plays a role in the pathogenesis of *Mycobacterium tuberculosis*, a specific CFA synthase being required to modify the alpha mycolates on the cell envelope, and pathogenic *E. coli* strains have higher CFA contents and are more resistant to acid shock than non-pathogenic strains [57]. $G6_{abc}$ and $G6_{acb}$ carry homologues of genes involved in fatty acid metabolism (Fab genes) conserved at multiple loci in all *A. baumannii* strains. Additional Fab genes may confer metabolic advantage, and is worth noting that Fab and other GEI-6 genes reside in OI-47, a genomic island conserved in all O157:H7 *E. coli* strains [58]. Finally, Many GEIs, most of which unique to the 4190 strain, carry genes and/or operons controlling specific metabolic pathways, such as naphthalene and phenyl-propionic acid degradation.

Several GEIs correspond to cryptic prophages. Of these, a few may have conserved the ability to replicate

Table 4 Distribution of genomic regions in *A.baumannii* strains of different genotypes

Strain	ST type	PFGE type	G47	G37	G11	G6	G57	G18	G51	G32	G20	G43	G3	G21	G33	G23	G46	G63	G8
AB0057	1	nd	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AYE	1	nd	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
700	1	A	1	0	0	0	0	0	1	0	0	1	1	1	0	0	0	1	0
3891	1	B	1	0	0	0	0	0	1	0	0	1	1	1	0	0	0	1	1
3887	1	C	1	0	0	0	0	0	1	0	0	1	1	1	1	0	1	1	0
2979	20	D	1	0	0	0	0	0	1	0	0	1	1	0	0	0	0	1	0
3130	20	E	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
ACICU	2	nd	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0
2105	2	F	0	1	1	1	0	0	1	1	0	1	0	0	0	0	0	1	1
2638	2	F	0	1	1	1	0	0	1	1	0	1	0	0	0	0	0	1	1
3892	2	F	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0
3990	2	F	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0
2735	2	F1	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0
3858	2	F2	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0
3889	2	G	0	1	1	1	0	0	0	1	0	0	0	0	1	0	0	0	0
4026	2	H	0	1	1	1	1	0	0	1	0	1	0	1	0	0	0	0	1
4030	2	I	0	1	1	1	0	0	0	1	1	0	0	0	0	0	0	0	0
4009	2	J	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0
4025	3	K	1	1	1	1	0	0	1	1	0	1	0	0	0	0	0	1	1
3890	25	L	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0
3865	25	M	0	0	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1
4190	25	N	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1
ATCC17978	77	nd	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0
3909	78	O	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
3911	78	O1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
3868	15	P	0	1	0	0	1	0	1	1	1	1	1	1	0	0	0	0	1
3871	84	P1	0	1	0	0	0	0	1	1	0	1	1	1	0	0	0	0	1

Positive or negative PCR amplification are indicated by 1 or 0, respectively; nd, not done.

as phages upon appropriate stimuli, and CP3, CP9 and CP14 encode lysozyme. However, none exhibited homology to bacteriophages so far identified in *A. baumannii* [59,60]. Few CPs are decorated by morons, accessory genes unnecessary for the virus, which may be helpful for the host bacteria when the prophage is integrated in its genome. Advantage conferred by morons is debated. PapS reductase functions in the assimilatory sulphate reduction pathway, and could serve as a fitness factor under conditions of iron limitation [61], *umuDC* gene could convey a mutator phenotype on the host [62]. As previously noted [16], the high variability exhibited by prophage sequences suggests recent insertion/and or rapid loss, and a large pool of phage genomes.

Genotypic characterization of *A. baumannii* isolates during outbreaks occurred in different geographical locations showed the prevalence of clusters of highly similar strains [4,10]. Data presented suggest that strains assigned to distinct genotypes according to MLST analysis may harbour specific GEIs. However, variability exists in the distribution of other genomic regions between *A. baumannii* strains assigned to the same genotypes, thus suggesting that horizontal gene transfer and recombination may occur between strains of different genotypes.

The identification of sequences homologous to several GEIs suggests that the genomes of non-*baumannii Acinetobacter* spp. may function as reservoirs of accessory *A. baumannii* DNA. Bacteria of the genus *Acinetobacter*, including *A. baumannii* isolates, are naturally competent [63] and have likely exchanged DNA in evolution. A few GEIs are perfectly conserved in different *Acinetobacter* species, but many vary in size and content, and have been plausibly remodelled both by recombination and insertional events. Comparative analyses also demonstrated a marked difference in the genome organization of the non-*baumannii Acinetobacter* sp. *baylyi* and DR1 relatively to *A. baumannii*.

Differences among *A. baumannii* genomes are also correlated to large strain-specific deletions, which are interestingly associated to selective loss of function. The 3909 strain lacks *mucK* and *tcu* genes which enable the growth on cis, cis-muconate and tricarballylate as sole

carbon sources [64,65]. The 4190 strain lacks *tau* genes, needed to utilize taurine as a sulphur source in sulphate starvation conditions [36], the AYE and ACICU strains lack genes enabling growth on d-glucarate as sole carbon source [66], the ATCC17978 lacks genes involved in the metabolism of anthranilate, molybdate transport, biosynthesis of the pyrroloquinoline quinone cofactor, chaperone-usher pathway, growth on dicarboxylic acids as the only carbon source [67]. All these large deleted regions can alternatively be viewed as GEIs conserved in the population but missing in one or a few isolates. Sequencing of additional *A. baumannii* isolates will set the issue.

Conclusions

The definition of the genome components of *A. baumannii* provides a scaffold to rapidly evaluate the genomic organization of novel clinical *A. baumannii* isolates. Distinguishing conserved from accessory components in *A. baumannii* chromosomes is a functional framework useful for further investigations on the biology and the genetic organization of this species. Changes in island profiling will be useful in genomic epidemiology of *A. baumannii* population. Data provided in this work will facilitate comparisons of *A. baumannii* isolates, and help to define the features of *A. baumannii* as species as to pin down its pathogenic traits.

Methods

A. baumannii strains

Comparative genome analysis were performed on whole genome sequences of *A. baumannii* strains AB0057 [GenBank:NC_011586] [16], ACICU [GenBank:NC_010611] [12], ATCC17978 [GenBank:NC_009085] [17] and AYE [GenBank:NC_010410] [18] and draft genome sequences of *A. baumannii* strains ST2 3990 [GenBank:AEOY00000000], ST25 4190 [GenBank:AEPA00000000] and ST78 3909 [GenBank:AEOZ00000000] strains [11]. The GenBank:CP000521 file, which contains 436 hypothetical proteins putatively encoded by ATCC17978 early annotated as AS1, but not included in the GenBank:NC_009085 file, was also used for comparisons. The genome sequences of non-*baumannii Acinetobacter* species *A. baylyi* ADP1 [GenBank:NC_011586], *Acinetobacter sp.* DR1 [GenBank:NC_014259], *A. calcoaceticus* RUH2202 [GenBank:ACPK00000000], *A. haemolyticus* ATCC19194 [GenBank:ADMT00000000], *A. johnsonii* SH046 [GenBank:ACPL00000000], *A. junii* SH205 [GenBank:ACPM00000000], *A. lwoffii* SH145 [GenBank:ACPN00000000], *A. radioresistens* SK82 [GenBank:ACVR00000000], *Acinetobacter sp.* ATCC27244 [GenBank:ABYN00000000], *A. nosocomialis* RUH2624 [GenBank:ACQF00000000] and *A. pittii* SH024 [GenBank:ADCH00000000] were also used for comparison. The *A. baumannii* strains used in PCR analyses of GEIs have been previously described [10].

Genome analyses

Gene products putatively encoded by the ST25 4190, ST78 3909 and ST2 3990 strains were identified using xBASE2, comparing the draft genome sequences to the genome of the *A. baumannii* strain AB0057 used as reference template [11]. The corresponding amino acid sequences are listed in Additional file 7. Predicted ORFs were subsequently compared to the gene products of the wholly sequenced *A. baumannii* AB0057, ACICU, ATCC and ABAYE strains using MAUVE [15]. Homologies under looked by MAUVE were detected by BLAST and tBLASTn analyses. Gene products encoded by aligned coding regions exhibited at least 50% identity. *rpoB* gene sequence analysis for genomic species identification was performed as previously described [3].

PCR analyses

The conservation of specific GEIs in a set of *A. baumannii* strains was assessed by PCR amplification. PCR reactions were carried out by incubating 20 ng of genomic DNA with 160 ng of each primer in the presence of dXTPs (200 nanomoles), 1.5 mM magnesium chloride and the Taq DNA polymerase Recombinant (Invitrogen). The sequences of the oligomers used as primers, the experimental conditions, the length of the amplimers, the coding regions amplified are all listed in Additional file 8. PCR products were electrophoresed on 1.5-2% agarose gels in 0.5×TBE buffer (45 mM Tris pH 8, 45 mM Borate, 0.5 mM EDTA) at 120 V (constant voltage). The 100 bp ladder (Promega) was used as molecular weight marker.

The co-linearity of contigs and the DNA content of the corresponding chromosomal regions were assessed by sequencing PCR products bridging contig ends.

Additional material

Additional file 1: Structures of plasmids identified in ST2 3990, ST25 4190 and ST78 3909 strains. the figure shows the circular maps of plasmids p1AB5T2, p2AB5T2, p1AB5T25, p2AB5T25 and p1AB5T78 with relevant features. ORFs and direction of the transcription are represented by arrow-shaped boxes. Plasmid sizes and names of various features are reported.

Additional file 2: Coding capacity of plasmids carried by strains 3909 3990 and 4190. the table lists ORFs of plasmids p1AB5T2, p2AB5T2, p1AB5T25, p2AB5T25 and p1AB5T78. Position, number of amino acids and putative function are reported for each ORF.

Additional file 3: Target site duplications. sequences duplicated at the ends of GEIs upon genome integration are listed in the table. Base changes in left and right TSDs are marked according to IUB codes. Residues missing in one TSD are in parenthesis. Known target genes are indicated.

Additional file 4: GEIs organization and ORFs content. the 63 sheets of the EXCEL file correspond to the 63 genomic loci carrying GEIs shown

In Figure 2. The ORF number, the amino acid length and the hypothesized function are given in each sheet. For draft genomes, the corresponding contigs are indicated. Identical or closely related ORFs present in different GEIs are positioned in the same row and labelled by the same colour to facilitate view. ORFs denoted as tb were identified by tBLASTn analyses. Grey and orange bars denote closely located ORFs putatively co-expressed. Homologous coding regions are boxed when a single ORF in one strain corresponds to two or more contiguous ORFs in others.

Additional file 5: Micro-heterogeneity regions. coding regions present/absent in the compared *A. baumannii* genomes, denoted in the text as *mhrs* (micro-heterogeneity regions), and their hypothetical function, are listed in the table. Alternative regions present at the same locus are marked by different colour characters. *mhrs* containing two or more ORFs are boxed.

Additional file 6: Cryptic prophages. structures of cryptic prophages identified in *A. baumannii* genomes. Prophage types are boxed to highlight their relatedness as resulting from MAUVE alignment. Different CP1 and CP2 are shown to illustrate the degree of genetic variation of *A. baumannii* prophage families.

Additional file 7: Gene products putatively encoded by strains 4190, 3909 and 3990. ORFs of strains 4190, 3909 and 3990 and the corresponding contig number are shown.

Additional file 8: Genomic regions, amplified genes, primers, amplicon sizes and cycling conditions used in PCR surveys. (none, title sufficiently describes data).

Acknowledgements
We thank all colleagues who generously provided strains included in the study: Antonella Agodi, Matteo Bassetti, Susanna Cuccurullo, Ziad Daoud, Athanasios Tsakris, and Haluk Vahaboglu.
This work was supported in part by grants from Agenzia Italiana del Farmaco, Italy (AIFA2007 contract no. *FARMTX9FBK*) and from Ministero dell'Istruzione, dell'Universitaìe della Ricerca, Italy (PRIN 2008 to RZ, PRIN 2009 to PPDN). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author details
[1]Dipartimento di Biologia e Patologia Cellulare e Molecolare, Università Federico II, Via S. Pansini 5, 80131 Napoli, Italy. [2]Dipartimento di Scienze Mediche Preventive, Sezione di Igiene, Università Federico II, Via S. Pansini 5, 80131 Napoli, Italy.

Authors' contributions
Conceived and designed the experiments: PPDN, FR, MG, MT, and RZ. Performed the experiments and analyzed the data: FR, PPDN, and MG. Wrote the paper: PPDN and RZ. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Received: 4 May 2011 Accepted: 10 October 2011
Published: 10 October 2011

References

1. Dijkshoorn L, Nemec A, Seifert H: **An increasing threat in hospitals: multidrug-resistant *Acinetobacter baumannii*.** *Nat Rev Microbiol* 2007, 5:939-951.
2. Durante-Mangoni E, Zarrilli R: **Global spread of drug-resistant *Acinetobacter baumannii*: molecular epidemiology and management of antimicrobial resistance.** *Future Microbiol* 2011, 6:407-422.
3. Nemec A, Krizova L, Maixnerova M, van der Reijden TJ, Deschaght P, Passet V, Vaneechoutte M, Brisse S, Dijkshoorn L: **Genotypic and phenotypic characterization of the *Acinetobacter calcoaceticus-Acinetobacter baumannii* complex with the proposal of *Acinetobacter pittii* sp. nov. (formerly *Acinetobacter* genomic species 3) and *Acinetobacter nosocomialis* sp. nov. (formerly *Acinetobacter* genomic species 13TU).** *Res Microbiol* 2011, 162:393-404.
4. Higgins PG, Dammhayn C, Hackel M, Seifert H: **Global spread of carbapenem-resistant *Acinetobacter baumannii*.** *J Antimicrob Chemother* 2010, 65:233-238.
5. Van Dessel H, Dijkshoorn L, van der Reijden T, Bakker N, Paauw A, van den Broek P, Verhoef J, Brisse S: **Identification of a new geographically widespread multiresistant *Acinetobacter baumannii* clone from European hospitals.** *Res Microbiol* 2004, 155:105-112.
6. Dijkshoorn L, Aucken H, Gerner-Smidt P, Janssen P, Kaufmann ME, Garaizar J, Ursing J, Pitt TL: **Comparison of outbreak and nonoutbreak *Acinetobacter baumannii* strains by genotypic and phenotypic methods.** *J Clin Microbiol* 1996, 34:1519-1525.
7. Wisplinghoff H, Hippler C, Bartual SG, Haefs C, Stefanik D, Higgins PG, Seifert H: **Molecular epidemiology of clinical *Acinetobacter baumannii* and *Acinetobacter* genomic species 13TU isolates using a multilocus sequencing typing scheme.** *Clin Microbiol Infect* 2008, 14:708-715.
8. Diancourt L, Passet V, Nemec A, Dijkshoorn L, Brisse S: **The population structure of *Acinetobacter baumannii*: expanding multiresistant clones from an ancestral susceptible genetic pool.** *PLoS One* 2010, 5:e10034.
9. Turton JF, Gabriel SN, Valderrey C, Kaufmann ME, Pitt TL: **Use of sequence-based typing and multiplex PCR to identify clonal lineages of outbreak strains of *Acinetobacter baumannii*.** *Clin Microbiol Infect* 2007, 13:807-815.
10. Di Popolo A, Giannouli M, Triassi M, Brisse S, Zarrilli R: **Molecular epidemiological investigation of multidrug-resistant *Acinetobacter baumannii* strains in four Mediterranean countries with a multilocus sequence typing scheme.** *Clin Microbiol Infect* 2011, 17:197-201.
11. Zarrilli R, Giannouli M, Rocco F, Loman NJ, Haines AS, Constantinidou C, Pallen MJ, Triassi M, Di Nocera PP: **Genome sequences of three *Acinetobacter baumannii* strains assigned to ST2, ST25 and ST78 multilocus sequencing typing genotypes.** *J Bacteriol* 2011, 193:2359-2360.
12. Iacono M, Villa L, Fortini D, Bordoni R, Imperi F, Bonnal RJ, Sicheritz-Ponten T, De Bellis G, Visca P, Cassone A, Carattoli A: **Whole-genome pyrosequencing of an epidemic multidrug-resistant *Acinetobacter baumannii* strain belonging to the European clone II group.** *Antimicrob Agents Chemother* 2008, 52:2616-2625.
13. Bertini A, Poirel L, Mugnier PD, Villa L, Nordmann P, Carattoli A: **Characterization and PCR-based replicon typing of resistance plasmids in *Acinetobacter baumannii*.** *Antimicrob Agents Chemother* 2010, 54:4168-4177.
14. Merino M, Acosta J, Poza M, Sanz F, Beceiro A, Chaves F, Bou G: **OXA-24 carbapenemase gene flanked by XerC/XerD-like recombination sites in different plasmids from different *Acinetobacter* species isolated during a nosocomial outbreak.** *Antimicrob Agents Chemother* 2010, 54:2724-2727.
15. Darling AE, Mau B, Perna NT: **progressiveMauve: multiple genome alignment with gene gain, loss, and rearrangement.** *PLoS One* 2010, 5:e11147.
16. Adams MD, Goglin K, Molyneaux N, Hujer KM, Lavender H, Jamison JJ, MacDonald IJ, Martin KM, Russo T, Campagnari AA, Hujer AM, Bonomo RA, Gill SR: **Comparative genome sequence analysis of multidrug-resistant *Acinetobacter baumannii*.** *J Bacteriol* 2008, 190:8053-8064.
17. Smith MG, Gianoulis TA, Pukatzki S, Mekalanos JJ, Ornston LN, Gerstein M, Snyder M: **New insights into *Acinetobacter baumannii* pathogenesis revealed by high-density pyrosequencing and transposon mutagenesis.** *Genes Dev* 2007, 21:601-614.
18. Vallenet D, Nordmann P, Barbe V, Poirel L, Mangenot S, Bataille E, Dossat C, Gas S, Kreimeyer A, Lenoble P, Oztas S, Poulain J, Segurens B, Robert C, Abergel C, Claverie JM, Raoult D, Médigue C, Weissenbach J, Cruveiller S: **Comparative analysis of *Acinetobacters*: three genomes for three lifestyles.** *PLoS ONE* 2008, 3:e1805.
19. Gaddy JA, Tomaras AP, Actis LA: **The *Acinetobacter baumannii* 19606 OmpA protein plays a role in biofilm formation on abiotic surfaces and in the interaction of this pathogen with eukaryotic cells.** *Infect Immun* 2009, 77:3150-3160.
20. Lee JS, Choi CH, Kim JW, Lee JC: ***Acinetobacter baumannii* outer membrane protein A induces dendritic cell death.** *J Microbiol* 2010, 48:387-392.
21. Russo TA, MacDonald U, Beanan JM, Olson R, MacDonald II, Sauberan SL, Luke NR, Schultz LW, Umland TC: **Penicillin-binding protein 7/8 contributes to the survival of *Acinetobacter baumannii* in vitro and in vivo.** *J Infect Dis* 2009, 199:513-521.

22. Jacobs AC, Hood I, Boyd KL, Olson PD, Morrison JM, Carson S, Sayood K, Iwen PC, Skaar EP, Dunman PM: **Inactivation of phospholipase D diminishes *Acinetobacter baumannii* pathogenesis.** *Infect Immun* 2010, **78**:1952-1962.
23. Russo TA, Luke NR, Beanan JM, Olson R, Sauberan SL, MacDonald U, Schultz LW, Umland TC, Campagnari AA: **The K1 capsular polysaccharide of *Acinetobacter baumannii* strain 307-0294 is a major virulence factor.** *Infect Immun* 2010, **78**:3993-4000.
24. Tomaras AP, Dorsey CW, Edelmann RE, Actis L: **Attachment to and biofilm formation on abiotic surfaces by *Acinetobacter baumannii*: involvement of a novel chaperone-usher pili assembly system.** *Microbiology* 2003, **149**:3473-3484.
25. Zimbler DL, Penwell WF, Gaddy JA, Menke SM, Tomaras AP, Connerly PL, Actis LA: **Iron acquisition functions expressed by the human pathogen *Acinetobacter baumannii*.** *Biometals* 2009, **22**:23-32.
26. Coyne S, Courvalin P, Périchon B: **Efflux-mediated antibiotic resistance in *Acinetobacter* spp.** *Antimicrob Agents Chemother* 2011, **55**:947-953.
27. Barbe V, Vallenet D, Fonknechten N, Kreimeyer A, Oztas S, Labarre L, Cruveiller S, Robert C, Duprat S, Wincker P, Ornston LN, Weissenbach J, Marlière P, Cohen GN, Médigue C: **Unique features revealed by the genome sequence of *Acinetobacter* sp. ADP1, a versatile and naturally transformation competent bacterium.** *Nucleic Acids Res* 2004, **32**:5766-5779.
28. Dobrindt U, Hochhut B, Hentschel U, Hacker J: **Genomic Islands in pathogenic and environmental microorganisms.** *Nat Rev Microbiol* 2004, **2**:414-424.
29. Sridhar J, Narmada SR, Sabarinathan R, Ou HY, Deng Z, Sekar K, Rafi ZA, Rajakumar K: **sRNAscanner: a computational tool for intergenic small RNA detection in bacterial genomes.** *PLoS One* 2010, **5**:e11970.
30. Post V, White PA, Hall R: **Evolution of AbaR-type genomic resistance islands in multiply antibiotic-resistant *Acinetobacter baumannii*.** *J Antimicrob Chemother* 2010, **65**:1162-1170.
31. Nuccio SP, Bäumler AJ: **Evolution of the chaperone/usher assembly pathway: fimbrial classification goes Greek.** *Microbiol Mol Biol Rev* 2007, **71**:551-575.
32. Barnhart MM, Chapman MR: **Curli biogenesis and function.** *Annu Rev Microbiol* 2006, **60**:131-147.
33. Locht C, Bertin P, Menozzi FD, Renauld G: **The filamentous haemagglutinin, a multifaceted adhesion produced by virulent Bordetella spp.** *Mol Microbiol* 1993, **9**:653-660.
34. Rocco F, De Gregorio E, Colonna B, Di Nocera PP: ***Stenotrophomonas maltophilia* genomes: a start-up comparison.** *Int J Med Microbiol* 2009, **299**:535-546.
35. Zhang YM, Rock CO: **Membrane lipid homeostasis in bacteria.** *Nature Rev Microb* 2008, **6**:222-232.
36. van der Ploeg JR, Iwanicka-Nowicka R, Bykowski T, Hryniewicz M, Leisinger T: **The *Escherichia coli ssuEADCB* gene cluster is required for the utilization of sulfur from aliphatic sulfonates and is regulated by the transcriptional activator Cbl.** *J Biol Chem* 1999, **274**:29358-29365.
37. Frankenberg-Dinkel N: **Bacterial heme oxygenases.** *Antioxidants & Redox Signaling* 2004, **6**:825-834.
38. Braun V, Mahren S, Sauter A: **Gene regulation by transmembrane signaling.** *Biometals* 2006, **19**:103-113.
39. Xu Y, Chen M, Zhang W, Lin M: **Genetic organization of genes encoding phenol hydroxylase, benzoate 1,2-dioxygenase alpha subunit and its regulatory protein in *Acinetobacter calcoaceticus* PHEA-2.** *Curr Microbiol* 2003, **46**:235-240.
40. Zhou NY, Fuenmayor SL, Williams PA: ***nag* genes of *Ralstonia* (formerly *Pseudomonas*) sp. strain U2 encoding enzymes for gentisate catabolism.** *J Bacteriol* 2001, **183**:700-708.
41. Díaz E, Fernández A, Prieto MA, García JL: **Biodegradation of aromatic compounds by *Escherichia coli*.** *Microbiol Mol Biol Rev* 2001, **65**:523-569.
42. Hendrix RW, Lawrence JG, Hatfull GF, Casjens S: **The origins and ongoing evolution of viruses.** *Trends Microbiol* 2000, **8**:504-508.
43. Permina EA, Mironov AA, Gelfand MS: **Damage-repair error-prone polymerases of eubacteria: association with mobile genome elements.** *Gene* 2002, **293**:133-140.
44. Eppinger M, Rosovitz MJ, Fricke WF, Rasko DA, Kokorina G, Fayolle C, Lindler LE, Carniel E, Ravel J: **The complete genome sequence of *Yersinia pseudotuberculosis* IP31758, the causative agent of Far East scarlet-like fever.** *PLoS Genet* 2007, **3**:e142.
45. Carlss SJ, Constantinidou C, Patel MD, Takebayashi Y, Hobman JL, Penn CW, Avison MB: **YieJ (CbrC) mediates CreBC-dependent colicin E2 tolerance in *Escherichia coli*.** *J Bacteriol* 2010, **192**:3329-3336.
46. Genschik P, Drabikowski K, Filipowicz W: **Characterization of the *Escherichia coli* RNA 3'-terminal phosphate cyclase and its sigma54-regulated operon.** *J Biol Chem* 1998, **273**:25516-25526.
47. Redko Y, de la Sierra-Gallay I, Condon C: **When all's zed and done: the structure and function of RNase Z in prokaryotes.** *Nature Rev Microbiol* 2007, **5**:278-286.
48. Horvath P, Barrangou R: **CRISPR/Cas, the immune system of Bacteria and Archaea.** *Science* 2010, **327**:167-170.
49. Godde JS, Bickerton A: **The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes.** *J Mol Evol* 2006, **62**:718-729.
50. Wang YD, Zhao S, Hill CW: ***Rhs* elements comprise three subfamilies which diverged prior to acquisition by *Escherichia coli*.** *J Bacteriol* 1998, **180**:4102-4110.
51. Bernier SP, Sokol PA: **Use of suppression-subtractive hybridization to identify genes in the *Burkholderia cepacia* complex that are unique to *Burkholderia cenocepacia*.** *J Bacteriol* 2005, **187**:5278-5291.
52. Bingle LEH, Bailey CE, Pallen MJ: **Type VI secretion: a beginner's guide.** *Curr Opin Microbiol* 2008, **11**:3-8.
53. Pukatzki S, Ma AT, Revel AT, Sturtevant D, Mekalanos JJ: **Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin.** *Proc Natl Acad Sci USA* 2007, **104**:15508-15513.
54. Jung J, Baek JH, Park W: **Complete genome sequence of diesel-degrading *Acinetobacter* sp. DR1.** *J Bacteriol* 2010, **192**:4794-4795.
55. Dobrindt U, Chowdary MG, Krumbholz G, Hacker J: **Genome dynamics and its impact on evolution of *Escherichia coli*.** *Med Microbiol Immunol* 2010, **199**:145-154.
56. Ernst RK, D'Argenio DA, Ichikawa JK, Bangera MG, Selgrade S, Jane L, Burns JL, Hiatt P, McCoy K, Bittner M, Kas A, Spencer DH, Olson MV, Ramsey BW, Lory S, Miller SI: **Genome mosaicism is conserved but not unique in *Pseudomonas aeruginosa* isolates from the airways of young children with cystic fibrosis.** *Environ Microbiol* 2003, **5**:1341-1349.
57. Cronan JE: **Phospholipid modifications in bacteria.** *Curr Opin Microbiol* 2002, **5**:202-205.
58. Shen S, Mascarenhas M, Morgan R, Rahn K, Karmali MA: **Identification of four fimbria-encoding genomic islands that are highly specific for verocytotoxin-producing *Escherichia coli* serotype O157 strains.** *J Clin Microbiol* 2005, **43**:3840-3850.
59. Yang H, Liang L, Lin S, Jia S: **Isolation and characterization of a virulent bacteriophage AB1 of *Acinetobacter baumannii*.** *BMC Microbiol* 2010, **10**:131.
60. Lin NT, Chiou PY, Chang KC, Chen LK, Lai MJ: **Isolation and characterization of phi AB2: a novel bacteriophage of *Acinetobacter baumannii*.** *Res Microbiol* 2010, **161**:308-314.
61. Romling CM, Losada L, Brinkac L, Inman J, Ulrich RL, Schell M, Nierman WC, DeShazer D: **Genetic and phenotypic diversity in *Burkholderia*: contributions by prophage and phage-like elements.** *BMC Microbiol* 2010, **10**:202.
62. Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W, Fouts DE, Levy S, Knap AH, Lomas MW, Nealson K, White O, Peterson J, Hoffman J, Parsons R, Baden-Tillson H, Pfannkoch C, Rogers YH, Smith HO: **Environmental genome shotgun sequencing of the Sargasso Sea.** *Science* 2004, **304**:66-74.
63. Ramirez MS, Don M, Merkier AK, Bistué AJ, Zorreguieta A, Centrón D, Tolmasky ME: **Naturally competent *Acinetobacter baumannii* clinical isolate as a convenient model for genetic studies.** *J Clin Microbiol* 2010, **48**:1488-1490.
64. Williams PA, Shaw LE: ***mucK*, a gene in *Acinetobacter calcoaceticus* ADP1 (BD413), encodes the ability to grow on exogenous cis, cis-muconate as the sole carbon source.** *J Bacteriol* 1997, **179**:5935-5942.
65. Lewis JA, Horswill AR, Schwem BE, Escalante-Semerena JC: **The tricarballylate utilization (tcuRABC) genes of *Salmonella enterica* serovar Typhimurium LT2.** *J Bacteriol* 2004, **186**:1629-1637.
66. Aghaie A, Lechaplais C, Sirven P, Tricot S, Besnard-Gonnet M, Muselet D, de Berardinis V, Kreimeyer A, Gyapay G, Salanoubat M, Perret A: **New insights into the alternative D-glucarate degradation pathway.** *J Biol Chem* 2008, **283**:15638-15646.

67. Parke D, Garcia MA, Ornston LN: **Cloning and genetic characterization of dca genes required for beta-oxidation of straight-chain dicarboxylic acids in *Acinetobacter* sp. strain ADP1.** *Appl Environ Microbiol* 2001, 67:4817-4827.

doi:10.1186/1471-2180-11-224
Cite this article as: Di Nocera *et al.*: Genome organization of epidemic *Acinetobacter baumannii* strains. *BMC Microbiology* 2011 11:224.