GENOME BIOLOGY AND EVOLUTION

SMBE

The Genomic Diversification of the Whole *Acinetobacter* Genus: Origins, Mechanisms, and Consequences.

Marie Touchon[1,2], Jean Cury[1,2] , Eun-Jeong Yoon[3], Lenka Krizova[4], Gustavo C. Cerqueira[5],

Cheryl Murphy[5], Michael Feldgarden[5], Jennifer Wortman[5], Dominique Clermont[6], Thierry

Lambert[3], Catherine Grillot-Courvalin[3], Alexandr Nemec[4,*], Patrice Courvalin[3,*], Eduardo PC

Rocha[1,2,*]

[1,2] Microbial Evolutionary Genomics, Institut Pasteur, 75724 Paris Cedex 15, France; CNRS, UMR3525, 75724 Paris, France

[3] Unité des Agents Antibactériens, Institut Pasteur, 75724 Paris Cedex 15, France.

[4] Laboratory of Bacterial Genetics, National Institute of Public Health, Šrobárova 48, 100 42 Prague 10, Czech Republic.

[5] Broad Institute of Harvard and MIT, Cambridge, MA 02142, Massachusetts, USA.

[6] Collection de l'Institut Pasteur, Institut Pasteur, 75724 Paris Cedex 15, Paris, France.

* Corresponding authors.

Author for Correspondence: Eduardo PC Rocha, Institut Pasteur, 25 R Dr Roux, 75724 Paris Cedex 15, France, telephone number +33 140613353, erocha@pasteur.fr.

Abstract

Bacterial genomics has greatly expanded our understanding of micro-diversification patterns within a species, but analyses at higher taxonomical levels are necessary to understand and predict the independent rise of pathogens in a genus. We have sampled, sequenced, and assessed the diversity of genomes of validly named and tentative species of the *Acinetobacter* genus, a clade including major nosocomial pathogens and biotechnologically important species. We inferred a robust global phylogeny and delimited several new putative species. The genus is very ancient and extremely diverse: genomes of highly divergent species share more orthologs than certain strains within a species. We systematically characterized elements and mechanisms driving genome diversification such as conjugative elements, insertion sequences, and natural transformation. We found many error-prone polymerases that may play a role in resistance to toxins, antibiotics, and in the generation of genetic variation. Surprisingly, temperate phages, poorly studied in *Acinetobacter*, were found to account for a significant fraction of most genomes. Accordingly, many genomes encode CRISPR-Cas systems with some of the largest CRISPR-arrays found so far in bacteria. Integrons are strongly over-represented in *Acinetobacter baumannii*, which correlates with its frequent resistance to antibiotics. Our data suggests that *A. baumannii* arose from an ancient population bottleneck followed by population expansion under strong purifying selection. The outstanding diversification of the species occurred largely by horizontal transfer, including some allelic recombination, at specific hotspots preferentially located close to the replication terminus. Our work sets a quantitative basis to understand the diversification of *Acinetobacter* into emerging resistant and versatile pathogens.

Keywords: Comparative genomics, bacterial genus, evolution, mobile genetic elements, nosocomial pathogens.

Introduction

In the last few years a number of studies have harnessed the power of high-throughput sequencing to study epidemiological and evolutionary patterns within bacterial species. Such studies have uncovered patterns of transmission of multidrug resistant clones, the emergence of virulent strains, and their within-host evolution (e.g., (Harris et al. 2010; Kennemann et al. 2011; Mather et al. 2013; McGann et al. 2014)). There has been considerably less emphasis on the large-scale sampling and sequencing of broader taxonomic units such as genera. Yet, this is an important level of analysis to understand the emergence of pathogenic species, because genera often include pathogens, commensals, and free-living (environmental) bacteria. The genus *Acinetobacter* is a good example of this since it includes a broad group of biochemically and physiologically versatile bacteria that occupy different natural ecosystems and play an increasing causative role in opportunistic human infections. The taxonomy of the genus consists currently of 33 distinct validly named species (http://www.bacterio.net/acinetobacter.html), about 10 provisionally termed genomic species delineated by DNA-DNA hybridization (DDH) (Dijkshoorn et al. 2007), several putative new species but not validly named, and a number of strains of an as-yet unknown taxonomic status (Rodriguez-Bano et al. 2006; Yamahira et al. 2008; Smet et al. 2012). Following seminal DDH measures of relatedness (Bouvet, Jeanjean 1989; Tjernberg, Ursing 1989), several distinct phylogenies of the *Acinetobacter* genus have been estimated using one or a few phylogenetic marker genes (Rainey et al. 1994; Yamamoto et al. 1999; Krawczyk et al. 2002; La Scola et al. 2006; Diancourt et al. 2010). These analyses suggested the existence of several groups encompassing phylogenetically close species, including the species of the *Acinetobacter calcoaceticus-Acinetobacter baumannii* (ACB) complex (Nemec et al. 2011), proteolytic genomic species (Bouvet, Jeanjean 1989; Nemec et al. 2009), *Acinetobacter guillouiae* and *Acinetobacter bereziniae* (Nemec et al. 2010), or *Acinetobacter nectaris* and *Acinetobacter boissieri* (Alvarez-Perez et al. 2013). However, these previous studies showed very diverse phylogenetic scenarios at higher taxonomic levels with weakly supported internal nodes (Yamamoto et al. 1999; Krawczyk et al. 2002; La Scola et al. 2006; Diancourt et al. 2010). Recent genome-wide analyses produced phylogenies showing that *A. baumannii* is well separated from the other strains of the ACB complex and these from other species in the genus (Chan et al. 2012). Nevertheless, the lack of a robust phylogenetic scenario encompassing all known *Acinetobacter* spp. and the unknown position of their last common

ancestor (root of the tree) seriously hampers the understanding of the diversification of this genus.

Acinetobacter spp. are among the most frequent causes of hospital-acquired bacterial infections (Peleg et al. 2008). Community-acquired infections are less frequent but have also been reported (Falagas et al. 2007; Eveillard et al. 2013). While *A. baumannii* is the most frequently identified nosocomial pathogen in the genus, several other species cause occasionally infections in humans including *Acinetobacter nosocomialis*, *Acinetobacter pittii*, and less frequently *Acinetobacter ursingii*, *Acinetobacter haemolyticus*, *Acinetobacter lwoffii*, *Acinetobacter parvus*, and *Acinetobacter junii* (Nemec et al. 2001; Nemec et al. 2003; Dijkshoorn et al. 2007; Peleg et al. 2008; Turton et al. 2010; Karah et al. 2011). Clinical cases typically involve ventilator-associated pneumonia and septicemia, but also endocarditis, meningitis, burn and surgical wound infections, and urinary tract infections. Importantly, *Acinetobacter* spp. are isolated in the environment and asymptomatically associated with humans, albeit the precise environmental reservoirs are unknown. Some species, notably *Acinetobacter baylyi*, are becoming emerging model organisms because of their transformability, metabolic versatility, and genome plasticity (Barbe et al. 2004; Metzgar et al. 2004; de Berardinis et al. 2008). Interestingly, these same traits, together with intrinsic resistance to toxins and antibiotics, are thought to cause the increasing frequency of *Acinetobacter* spp. as agents of nosocomial infections (Peleg et al. 2012).

In the last decade a small number of complete genome sequences and a large number of draft sequences of *Acinetobacter* spp. have become available (Barbe et al. 2004; Fournier et al. 2006; Antunes et al. 2013). These studies focused heavily on *A. baumannii* and the ACB complex and on what distinguishes them. While the diversity of the genus remains largely unexplored, several studies have shown that *A. baumannii* gene repertoires are very diverse, with fewer than half of the genes being part of the species' core genomes (Adams et al. 2008; Imperi et al. 2011; Sahl et al. 2011; Farrugia et al. 2013). Like for many bacterial clades, a large fraction of the genes of the pan-genome are of unknown function. Accessory functions involved in transport and genetic regulation are also highly variable between strains (Adams et al. 2008; Imperi et al. 2011). Finally, genome-wide comparisons between *A. baumannii* and other species suggest the existence of high genetic diversity in the genus (Vallenet et al. 2008; Fondi et al. 2013).

Somewhat surprisingly, most known virulence factors of *A. baumannii* are found in its core genome and are present in other species of the genus (Antunes et al. 2011). The multifactorial

basis of virulence and the independent emergence of pathogenic strains in the genus suggest that the genetic background has an important role in *Acinetobacter* evolution. To understand the emergence of pathogenic *Acinetobacter* it is therefore important to study the mechanisms of genetic diversification of the genus. For this, we have carefully sampled a large number of representative strains. Following the sequencing of their genomes, we characterized their genetic diversity and built a robust phylogeny of the entire genus. This information was used to assess pending taxonomic issues, to guide evolutionary studies, and to sample the genus for key mechanisms generating genetic variability. With these data at hand we focused on the genome dynamics of *A. baumannii*.

Results and discussion

Genetic diversity of the genus

We analyzed a panel of 133 genomes of *Acinetobacter* spp. covering the breadth of the known taxonomic diversity of the genus (Table S1). All validly named species were represented by the respective type strains. The provisionally designated genomic species or tentative novel species were represented by strains selected based on polyphasic analysis of multiple strains belonging to the given taxon. Additional strains of *A. baumannii* representing different genotypes were identified based on multilocus sequence typing (MLST) and other typing methods (Diancourt et al. 2010). In the case of clinically relevant species other than *A. baumannii*, one or two additional strains were included to study intra-species variation. These strains were selected to differ from the type strains as much as possible in terms of relevant genotypic/phenotypic properties and origin (*e.g.*, clinical versus environmental). Some strains were added to study their taxonomic position. We sequenced 120 genomes with high coverage to which we added the complete sequences of the 13 genomes available from GenBank (see Methods and Figure S1). The average genome in the dataset had 13 scaffolds for 31 contigs and an average size of 3.87 Mb (range between 2.7 and 4.9 Mb). Genomic GC content showed little variation around the average value of 39.6%. The density of protein coding sequences was found to be homogeneous between genomes at an average of 94%. To test the completeness of the incompletely assembled genomes we carried out two tests. First, we merged the lists of essential genes reported in *Escherichia coli* (Baba et al. 2006) and *A. baylyi* (de Berardinis et al. 2008) (including "double band" mutants, which may not be essential). A total of 414 of the resulting 533 putatively essential genes had homologs in all *Acinetobacter* genomes (78%). Only 38 of these genes were present in less than 121 genomes and more than 9 (7%), among which only one deemed essential in both *A. baylyi* and *E. coli* (*ligA*, absent in two genomes) (Table S2). Second, we compared the fully and partially assembled *A. baumannii* genomes in terms of genome size and number of genes. The two groups of genomes were indistinguishable on both accounts (both p>0.5, Wilcoxon tests), suggesting that not fully assembled genomes miss very few genes. Hence, our collection provides a unique and comprehensive reference dataset of high quality genomes of the genus *Acinetobacter*.

To quantify the diversity of the gene repertoires we computed the set of ubiquitous genes (core-genome) and the set of different homologous gene families (pan-genome) in the genus

and in the 34 genomes of *A. baumannii* (Figure 1, Table S3 and S4). The *Acinetobacter* core-genome contained 950 orthologous protein families corresponding to 37% of the size of the smallest proteome (*A. nectaris* CIP 110549^{T}) and more than twice the number of essential genes in *A. baylyi*. The *A. baumannii* core-genome had 1590 orthologous protein families corresponding to 44% of the size of the smallest proteome of the species (*i.e. A. baumannii* AB307 0294). Gene rarefaction analyses showed that core-genomes vary little with the addition of the last genomes (Figure 1), suggesting that our estimate of the core-genome is robust. Both the pan-genomes of the *Acinetobacter* genus and of *A. baumannii* were very large with, respectively, 26660 and 9513 gene families. Gene rarefaction analyses showed that in both cases the addition of new genomes to the analysis still significantly increases the size of the pan-genome. This was confirmed by the spectrum of gene frequencies for the *A. baumannii* pan-genome (Figure 1), which showed that the vast majority of gene families were either encoded in a few genomes (54% in three or less) or in most of them (27% in more than 31 genomes). Over a third of the pan-genome (40%) corresponded to gene families observed in a single genome, *i.e.*, strain-specific genes. These analyses confirm that the genus and *A. baumannii* have an extremely large pan-genome. Furthermore, the shape of the accumulation curves showed that we are yet very far from having sampled it enough. Further work will therefore be necessary to characterize the genetic diversity of the species in the genus.

We took advantage of the possibility offered by our dataset to compute the genetic variability of all the other validly names species or genomic species in the genus (Table 1 for clades with at least three genomes and Table S5 for all). With the exception of *A. baylyi*, *Acinetobacter indicus* and *Acinetobacter brisouii*, for which the genomes are very similar, all genomes showed large variability of gene repertoires. The comparison of the size of the pan-genomes of these species with the distribution of random samplings of the pan-genome of *A. baumannii* with an equivalent number of genomes showed systematic larger values for the latter (Figure 2, p<0.001, Binomial test). Hence, although we have chosen the genomes outside *A. baumannii* to maximize known biochemical and ecological diversity within these species, these were found to be less diverse than *A. baumannii*.

Phylogeny and systematics of the genus

Despite recent progress, the understanding of the *Acinetobacter* evolution is incomplete at the phylogenetic and taxonomic levels. We used the 950 core protein families of the genus to build its phylogeny (see Methods). The resulting genus phylogenetic tree is extremely well-

supported from the statistical point of view, showing only one bifurcation with a bootstrap support lower than 95% (67%, see Figure 3). To root this tree and thus infer the order of evolutionary events in the genus we used two genomes from the two most closely related genera for which complete genomes were available (*Moraxella* and *Psychrobacter*). This tree (topologically very similar to the previous one, see Figure S2), positions *A. brisouii* and *A. nectaris* as the taxa branching deeper in the genus, *i.e.*, the taxa most distantly related with the remaining *Acinetobacter* spp. Relative to the rest of the genus, the small (~3.2 Mb) genomes of *A. brisouii* showed average G+C content (41.5%). The genome of *A. nectaris* was among the smallest in size (2.9 Mb) and lowest in GC content (36.6%). The small size and extreme GC content of the genome and the very long terminal branch of *A. nectaris* in the phylogenetic tree suggest rapid evolution for this species. This feature is typical of bacteria enduring strong ecological niche contractions (Ochman, Moran 2001). Following this split, the phylogeny separates two very large groups of taxa: one including species such as *A. baumannii*, *A. parvus,* and *A. baylyi*; the other including *A. lwoffii*, *Acinetobacter johnsonii and Acinetobacter guillouiae*. Among these taxa, two are more isolated in the phylogenetic tree. *Acinetobacter radioresistens* is believed to be highly resistant to gamma-ray irradiation and might be the origin of the OXA-23 carbapenem resistance determinant in *A. baumannii* (Poirel et al. 2008; Perichon et al. 2014). It branched deep in the tree and lacked closely related species (Nishimura et al. 1988; Poirel et al. 2008; Sahl et al. 2013). *Acinetobacter rudis* showed a long branch in the tree, even if its position is very well supported, suggesting higher evolutionary rates of this bacterium (isolated from raw milk and wastewater) (Vaz-Moreira et al. 2011).

To assess the age of the genus, we computed the average protein similarity of positional orthologs of the core-genome between the earlier branching species (*A. brisouii*) and *A. baumannii*. We did not use *A. nectaris* for this analysis because its long external branch would lead to an over-estimate of the distances within the genus. The orthologs between *A. brisouii* and *A. baumannii* show an average sequence similarity of 80.1% (IC95%:79.5-80.7%). As a matter of comparison, the same analysis between the orthologs of the core-genomes of *E. coli* and *Yersinia pestis* - placed in extreme opposites of the *Enterobacteriaceae* (after removing the fast-evolving *Buchnera* clade) (Williams et al. 2010) - shows an average protein similarity of 80% (IC95%:79.3-80.7%). Hence, the genus of *Acinetobacter* is very ancient and its last common ancestor was close contemporary of the last common ancestor of *Enterobacteriaceae*. Accurate dating of bacterial genus is impossible

given the lack of fossil records. Nevertheless, *Enterobacteriaceae* are thought to have diverged from the *Pasteurellaceae* over 500 MY ago (Battistuzzi, Hedges 2009) and *Acinetobacter* might therefore be as ancient. The ancient history of the *Acinetobacter* genus contributes to explain its metabolic and ecological diversity.

The taxonomy of *Acinetobacter* still suffers from unclear taxonomic position and/or confusing nomenclature of some provisional species, high number of unidentifiable environmental strains (Nemec and Krizova, unpublished), and a number of controversial interpretations of taxonomic data (Nemec et al. 2008; Vaneechoutte et al. 2008; Nemec et al. 2011). The availability of complete genomes and a robust phylogenetic history allow the identification of these taxonomical problems and the preliminary identification of taxa that might be regarded as good candidates for new species. We therefore put together the information on the core-genome tree and the average nucleotide identity (ANI). Values of ANI between 94% and 96% have been proposed to be a good threshold for the definition of a bacterial species and to replace DDH measurements in preliminary identification of bacterial species from genome data (Konstantinidis et al. 2006; Richter, Rossello-Mora 2009). Most named species and genomic species in this analysis showed intra-species ANI values higher than 95%. This is in close agreement with the idea that they represent *bona fide* species (Figure 4). ANI analysis also shed some light on the taxonomical status of several strains that were previously the subject of taxonomic controversies (see values in Table S1). Firstly, high (>97%) ANI values unambiguously corroborated that *Acinetobacter grimontii* CIP 107470[T] belongs to *A. junii* and *'Acinetobacter septicus'* ANC 3649 to *Acinetobacter ursingii* as previously suggested based on DDH data (Nemec et al. 2008; Vaneechoutte et al. 2008). On the other hand, strain CIP 64.10 which was believed to be derived from *A. lwoffii* NCTC 5866[T] (Bouvet, Grimont 1986) is clearly distinct from it (ANI of 88.3%). This finding explains previous controversial DDH results for these organisms (Tjernberg, Ursing 1989). Previous studies have also pointed out taxonomic problems with some closely related provisional species, notably among proteolytic and hemolytic strains (Bouvet, Jeanjean 1989). DDH values found by these authors for genomic sp. 15BJ and 16 are in agreement with the observed ANI reference strains of these species (92.6 %), suggesting that the genetic distance between these taxa is lower but close to the thresholds underlying species definition. Genomic sp. 13BJ (Bouvet, Jeanjean 1989) and 14TU (Tjernberg, Ursing 1989) have also been considered as a single species based on DDH data, while their ANI values (~94.5%) are close to the threshold used to define a species. In such cases, rare in our dataset, clear taxonomic

conclusions will require analyses of biochemical and genetic data from more comprehensive sets of strains (to be published separately).

We studied the association between phylogenetic distance and the gene repertoire relatedness (GRR) (Figure 5). GRR was defined for each pair of genomes as the number of orthologs present in two genomes divided by the number of genes of the smallest genome (Snel et al. 1999) (see Methods). It is close to 100% if the gene repertoires are very similar (or one is a subset of the other) and lower otherwise. Consistent with the large pan-genome of most species in the genus, we observed highly variable gene repertoires for genomes within the same species (short phylogenetic distances). The most extreme differences were found when comparing the susceptible *A. baumannii* SDF strain with other strains of the same species. This strain endured a process of genome reduction concomitant with proliferation of insertion sequences (Vallenet et al. 2008). After its removal from the dataset, the lowest within-species GRR (78%) was still found between *A. baumannii* strains (ATCC 17978 and ANC 4097), in line with our previous observation that this species is particularly diverse. As expected, pairs of genomes of the same species tended to have higher GRR than distantly related genomes. Some of the latter had only around 60% GRR. Yet, there were many exceptions to this average trend, and comparisons between distant genomes often showed higher GRR than comparisons between closely related strains of the same species (insert in Figure 5). For example, *A. baumannii* NIPH 146 and *Acinetobacter soli* CIP 110264 were very distant in the phylogenetic tree and have more than 82% GRR, which is more than many within-species comparisons. This shows the importance of sampling bacterial diversity using complete genome sequences and not just using MLST or core genome-based analyses. In fact, given these patterns, some strains of distantly related *Acinetobacter* species might have more similar phenotypes than strains within the same species.

Mechanisms of genomic diversification

The emergence of antibiotic resistance genes in *Acinetobacter* is facilitated by conjugative elements (Goldstein et al. 1983), integrons (Ploy et al. 2000; Hujer et al. 2006), insertion sequences (IS) (Turton et al. 2006), and natural transformation (Wright et al. 2014). Accordingly, we searched for the genes associated with horizontal gene transfer or its control (Figure 6, Figure 7, Table S6). We have found 23 proteins matching the profiles of tyrosine recombinases and the specific profiles for integrases of integrons. Interestingly, 17 of these were of the type *intI1* (associated with a 3′-conserved segment) and were found in 13 strains

of *A. baumannii* (4 strains had two copies). The abundance of integrons in this species was much higher than would be expected if it were random in the genus (p<0.0001, χ^2 test). The 6 hits for integrases of integrons in other species do not match IntI1, IntI2, or IntI3 and require further functional study. Analysis of the integron cassette contents in *A. baumannii* showed an In0 structure, *i.e.*, no cassettes, in one strain (Bissonnette, Roy 1992), whereas the others contained 2 to 5 cassettes. We observed an atypical inverted organization where *sul1* was upstream from *intI1* in *A. baumannii* 1656-2.

Insertion sequences (IS) have also been implicated in antibiotic resistance. Most notably, IS*Aba1* provides a promoter allowing expression of a downstream carbapenemase gene (Turton et al. 2006). IS are very diverse in type and abundance in the genus (from 0 to ~400 per genome), even when comparing closely related strains. The genomes of some species are particularly enriched in IS - *A. lwoffii*, *A. parvus*, *A. junii, A. ursingii* - and these may have contributed to their reduced size. We found at least one copy of IS*Aba1* (IS*4* family, group IS*10*) in 36% of all genomes and at least ten copies in 10%, suggesting an important role of this type of IS in the genus.

While some *Acinetobacter* lytic phages have been studied for typing and phage therapy purposes (Bouvet et al. 1990; Ackermann et al. 1994; Shen et al. 2012), there is very little information in the literature on temperate phages infecting *Acinetobacter*. We found 260 prophages of dsDNA phages in the genomes of *Acinetobacter*. We have made a classification of these elements based on the available phages of gamma-proteobacteria (given the lack of information on *Acinetobacter* temperate phages) as in (Bobay et al. 2013). More than 98% of the prophages were classified as Caudovirales, among which *Siphoviridae* (41%) and *Myoviridae* (37%) were by far the most abundant. The prophages accounted for a total of 10.4 Mb of genomic sequence in our dataset, *i.e.*, an average of 2% of the genomes. Only 18 genomes lacked prophages. Hence, most *Acinetobacter* are lysogens. Among the 72 genomes with more than one prophage, *Acinetobacter* ANC 3929 stood out with six prophages (Table S1). Only a minority of these prophages (51) integrated next to tRNAs, as is common in other clades (Williams 2002), even if all identified phage integrases were Tyrosine recombinases. The genes of these prophages will be studied in detail in a subsequent work. Nevertheless, since these data suggest an unsuspected role of transduction in driving horizontal gene transfer in the genus we have searched for putatively adaptive traits among prophages. Among other genes, we found a beta-lactamase in *A. baumannii* ANC 4097 and a chloramphenicol resistance protein in *A. baumannii* BM4587. Recent findings suggest that

phages favour the horizontal transfer of antibiotic resistance determinants (Muniesa et al. 2013; Billard-Pomares et al. 2014). These results suggest they may indeed contribute to antibiotic resistance in *Acinetobacter*.

Many conjugative elements have been described in association with the spread of antibiotic resistance genes between distant species (Doucet-Populaire et al. 1992; Juhas et al. 2008). We scanned genomes for genes encoding components of the conjugation machinery: relaxases, coupling proteins, and type 4 secretion systems (T4SS) (Guglielmini et al. 2014). We identified 23 putative conjugation systems in the genus, of which 11 were classified as MPF_F (family of the F plasmid), 4 MPF_I (family of the R64 plasmid), and 8 MPF_T (family of the Ti plasmid). Since most genomes in our sample were not in a single contig, and breakpoints were typically found at mobile genetic elements, it is difficult to unambiguously distinguish integrative (ICE) from extra-chromosomal (plasmids) conjugative elements. Yet, some information can be retrieved from the abundance of MPF_F and MPF_I types. These systems are much more frequently associated with plasmids than with ICE (Guglielmini et al. 2011) and they typically correspond to narrow host-range mobile genetic elements (Encinas et al. 2014). Interestingly, the long flexible pili of these two families of elements endow them with the ability to engage in conjugation at high frequency in liquid (Bradley 1984). This suggests that liquid media may be relevant for the spread of genetic information in *Acinetobacter*. We identified 211 relaxases distant from any T4SS, mostly MOBQ (140) and MOBP1 (47) (Garcillan-Barcia et al. 2009), which presumably are part of elements mobilizable by conjugation in *trans*. Mobilizable elements are particularly abundant in the genome of *Acinetobacter gerneri* (16) and in certain strains of *A. junii* (up to 10 in a genome) and *A. lwoffii* (up to 9). They are less frequent in *A. baumannii* (between 0 and 3 per genome). Mobilizable plasmids are in general smaller than conjugative plasmids (Smillie et al. 2010) and it has been observed that *A. baumannii* plasmids tend to be small (Gerner-Smidt 1989; Fondi et al. 2010). Our observations suggest that mobilizable small elements may predominate over large conjugative elements in the genus, a much stronger over-representation than in other prokaryotes (Smillie et al. 2010).

Competence for natural transformation has been described in a few strains of *A. baylyi* (Gerischer, Ornston 2001), *A. baumannii* (Harding et al. 2013; Wilharm et al. 2013), and *A. calcoaceticus* (Nielsen et al. 1997) but is thought to be rare in the genus (Towner 2006). In *A. baumannii*, competence and twitching motility are tightly linked and depend on the same type 4-pilus (T4P) (Harding et al. 2013; Wilharm et al. 2013). We searched for the 13 key T4P and

competence-associated components and found most of them in all genomes. Only 16 genomes lacked one of the components, of which 13 lacked the *comP* gene that encodes a pilin (Table S7). These absences probably result from recent gene losses, since they are scattered in the phylogenetic tree of the genus. For example, *comP* is missing in only 3 of the 34 *A. baumannii* strains. It was suggested that *comP* was also absent from *A. baumannii* ATCC 17978 (Smith et al. 2007) but our re-annotation procedure revealed a very good hit to the corresponding PFAM domain (profile coverage 99%, e-value<10^{-21}). The very frequent specific deletion of *comP* is intriguing since it is one of the essential components of the natural transformation machinery in *A. baylyi* (Porstendorfer et al. 2000). It is tempting to speculate that the important antigenic potential of pilins (Nassif et al. 1993; Miller et al. 2014) might frequently favor selection for *comP* loss in host-associated bacteria. The conservation of the entire transformation machinery in the vast majority of the genomes suggests that most bacteria in the genus are naturally transformable under certain conditions. Further work will be required to understand the conditions leading to the expression of this trait.

Clustered regularly interspaced short palindromic repeats (CRISPR), together with associated sequences (*cas* genes and Cas proteins) form the CRISPR-Cas adaptive immune system against transmissible genetic elements such as plasmids and viruses (Sorek et al. 2013; Barrangou, Marraffini 2014). Fifty-one of the genomes encoded CRISPR-Cas systems (Figure 8). A type III-A system was associated with a cluster of 18 genes and a 37-bp repeat sequence and was present in only 2 distant strains (*A. junii* NIPH 182 and genomic sp. *Acinetobacter* 15TU NIPH 899). In both cases the CRISPR-arrays located at each side of the *cas* genes clusters were very small (Figure 8). These traits, strain-specific small CRISPR-arrays found in few strains, suggest that the type III-A system has been recently acquired and/or accumulates few spacers. Most CRISPR-Cas systems in the genus were of type I-F. They included 37% of all genomes. The *cas* operon was composed of 6 to 7 genes and the CRISPR repeat was 28 nt long. Based on the phylogenetic tree of Cas1 and the organization of the cluster of *cas* genes, we identified two I-F subtypes (I-Fa and I-Fb) that correspond to the subtypes identified in *A. baumannii* strains ADP and AYE, respectively (Hauck et al. 2012). Interestingly the I-Fb subtype is probably very ancient since it was integrated at the same genomic locus in very distant species, *e.g.*, *A. parvus*, *A. junii*, *A. ursingii*. This CRISPR-Cas system contained some very large CRISPR-arrays in some genomes, up to 304 repeats, with highly conserved repeat sequences and highly variable spacers. This suggests the existence of a strong selective pressure on the activity of CRISPR-Cas systems of this type.

We identified ~3000 spacers in CRISPR-arrays, most of which are unique (80%), *i.e.*, they are strain-specific. The vast majority of these spacers (88%) do not match other sequences in the *Acinetobacter* genomes. We found very few genomes having similar spacers and most of these cases corresponded to genomes of the same species among *A. baumannii*, *A. parvus* or *A. baylyi* (Figure 8). Most spacers matched genes of unknown function, but some matched phage-related functions. Since prophages and integrative elements are hard to delimit precisely, we searched for similarity between the spacers and the 831 complete phage and 3861 complete plasmid genomes available in Genbank. Only 2% of the spacers showed sequence similarity with elements of this dataset. This is not surprising, given the paucity of *Acinetobacter* phages in GenBank. Notwithstanding, we identified 10 spacers that match bacteriophages infecting *Acinetobacter* species (Bphi-B1251, AP22 phage), and 47 matching *Acinetobacter* plasmids (*e.g.*, pABTJ1, pNDM-BJ0, pNDM-BJ02). Interestingly, among the few spacers matching known genes, we found homologs of VirB4, VirB5, VirB8, and resolvase proteins, which are all key components of the conjugation machinery. Nevertheless, we found no significant statistical association between the presence of the CRISPR-Cas system (or the number of repeat sequences) and the number of prophages, mobilizable, and conjugative elements in the genus (all p>0.1 Spearman rho associations). The same negative results were obtained when the analysis was restricted to *A. baumannii*. The variability of CRISPR spacers might be used to type certain species, but only in combination with other markers, since many strains are devoid of such systems. Some of the CRISPR-arrays we have identified are amongst the largest ever found among bacteria. CRISPR-Cas systems are therefore likely to have an important role in the genome dynamics of the genus and in particular in controlling the transfer of conjugative elements.

Point mutations also account for the emergence of new traits in *Acinetobacter*, including antibiotic resistance (Yoon et al. 2013). The dynamics of adaptation by point mutations is accelerated when bacteria endure hyper-mutagenesis, for example following the implication of error-prone DNA polymerases in replicating damaged DNA (Tenaillon et al. 2004). The SOS-response of *A. baumannii* does not involve LexA, the typical key regulator of this response (Robinson et al. 2010). Accordingly, we searched and found no ortholog of LexA in the genus. It has been suggested that error-prone polymerases, which have multiple homologs in certain genomes, facilitate the rapid emergence of antibiotic resistance in *Acinetobacter* spp. by stress-induced mutagenesis (Norton et al. 2013). We screened the genomes for homologs of PolB and Y-polymerases and found no homolog of PolB, nor of the *imuABC* operon, which

is implicated in damage-induced mutagenesis (Galhardo et al. 2005). In contrast, we identified 345 Y-polymerases in the genus, *i.e.*, an average of almost three polymerases per genome (Figure 6). No single genome lacked Y-polymerases and certain harbored up to five copies of the gene. The pair *umuCD* (encoding PolV in *E. coli*) was present in nearly all genomes, often in multiple copies. The multiplicity of genes encoding Y-polymerases in these genomes is intriguing and suggests that they play important roles in *Acinetobacter*, *e.g.*, in acquiring tolerance to toxins and antibiotics and/or in their genetic diversification.

Origin and diversification of *A. baumannii*

The large pan-genome of *A. baumannii* showed that this species has highly diverse gene repertoires suggestive of frequent horizontal gene transfer (Figure 1). Genetic diversification can also result from allelic exchange by homologous recombination in the core-genome. We estimated the impact of this type of recombination in *A. baumannii* with Phi, a conservative and robust method to detect recombination (Bruen et al. 2006). We found that 32% of the core gene families are significantly affected by recombination ($p < 0.05$) (see Methods). To quantify the number and size of recombination tracts we concatenated the multiple alignments following the order of the *A. baumannii* ATCC 17978 strain and fetched 688 recombination events significantly highlighted by three procedures (RDP3, CHI2, and GENECONV, see Methods). We were able to precisely delimit the tracts for 526 events of recombination. Their size averaged 2.1 kb (95% of the tracts were between 367 bp and 16 kb long). This size is an underestimate because of the presence of sequences separating core genes and because multiple events of recombination lead to shorter tracts. We also confirmed the presence of recombination using ClonalFrame (Didelot, Falush 2007). This program estimated that recombination contributed to the observed polymorphisms more than mutations (1.37 times). This value is very close to the one observed for MLST data (1.3) (Diancourt et al. 2010). Homologous recombination near the origin of replication was recently associated with the diversification of three outbreak strains of *A. baumannii* (Snitkin et al. 2011). We therefore quantified the distribution of recombination rates along the chromosome of *A. baumannii*. The highest density of recombining genes among the 34 genomes was indeed found close to the origin of replication, but only on the counter-clockwise sense (end of the published sequence). Several other regions showed high frequency of recombination whereas others were nearly clonal (Figure 9). These results showed that a large fraction of the genes in *A. baumannii* are significantly affected by recombination, that rates of recombination vary along the chromosome, and that recombination tracts tend to be small.

A. baumannii has become a significant clinical problem in the 1970s (Bergogne-Berezin, Towner 1996), but whether this reflects adaptation of a small number of clones to hospital environments or population expansion is not known. The presence of short internal nodes close to the last common ancestor of the species and its large pan-genome have led to suggestions that *A. baumannnii* might have endured one wave of population expansion during the diversification of the species and another very recently after the introduction of antibiotics at the hospital (Diancourt et al. 2010; Antunes et al. 2013). The assessment of the hypothesis for a recent population expansion will require a larger sample of closely related genomes. To test the hypothesis of an ancient population expansion, we computed Tajima D in sliding windows along the genome of *A. baumannii* (see Methods) (Tajima 1989). We observed systematically negative values of D (average D=-0.50, p<0.001, Wilcoxon signed-rank test). Tajima D is affected by recombination (Thornton 2005), but purging the alignments of genes for which Phi identified significant evidence of recombination resulted in even more negative values (average D=-0.9, p<0.001, same test). Negative D is consistent with population expansion and/or purifying selection. To separate between these two possibilities we analyzed separately fourfold degenerate synonymous (D_4) and strictly non-synonymous (D_0) positions (Figure S3). The two measures are equally affected by sampling biases, recombination and population expansion. Differences between D_4 and D_0 pinpoint selective processes because non-synonymous changes are much more deeply imprinted by natural selection than synonymous ones. The D_0 values are significantly lower than those of D_4 (resp. average D_0=-1.5 and D_4=-0.4, difference significant p<0.001 Wilcoxon signed-rank test) even if both are significantly negative (p<0.001). This suggests that negative values of Tajima's D are driven by selection against non-synonymous substitutions, a clear sign of purifying selection. To further test this conclusion, we measured the ratio of non-synonymous and synonymous substitutions (dN/dS). The average within-species dN/dS was only 0.05 (p<0.001) and even very closely related strains (dS<0.001) showed dN/dS lower than 0.2 (Figure S4). This confirms that natural selection purges the vast majority of non-synonymous mutations in the genome (Rocha et al. 2006). We then computed Fay and Wu H at fourfold degenerate positions (H_4) (Fay, Wu 2000). We found very low H_4 average values (-51, p<0.001). Negative H_4 is an indication of selective sweeps or ancient population bottlenecks and negative D_4 suggests population expansion. These results are thus consistent with the hypothesis of a population bottleneck in *A. baumannii* in the early stages of speciation with subsequent population expansion under a regime dominated by purifying selection.

Acquisition of resistance often results from the transfer of a mobile element encoding several resistance genes. For example, the AYE multi-resistant strain has a genomic island (AbaR1) containing 45 resistance genes including numerous determinants of antibiotic resistance (Fournier et al. 2006). This island was probably acquired in multiple steps of accretion and deletion of genetic material (Sahl et al. 2011). We studied the general patterns of integration of horizontally acquired genes in *A. baumannii* to quantify how many regions in the genome were integration/deletion hotspots. We identified 1083 regions in the genomes that were flanked by two consecutive core genes and included more than 10 genes inserted or deleted (indel) in at least one genome (Figure 10). These loci were not distributed randomly. Instead, the 1083 regions with indels occurred at the same 78 hotspot regions (5% of all possible loci), *i.e.*, they were flanked by the same 78 pairs of core gene families. A third of these loci corresponded to a single indel in one single genome, typically a strain-specific deletion (indicated by light colors and a large number of families with low diversity in Figure 10). Other loci included many different protein families in different genomes. These corresponded to hotspots that endured multiple integrations/deletions in different lineages. Hotspots tended to be concentrated closer to the terminus of replication and symmetrically distributed around this position. This tendency has previously been observed in other species (Bobay et al. 2013), and might result from a compromise between selection for genome plasticity and organization (Rocha 2004). Intriguingly, some large regions of the chromosome showed no signs of genome plasticity suggesting that they are less plastic (Figure 10). The 78 hotspots contain 5203 families, *i.e.*, 5% of locations in the genome accumulated 66% of the accessory genome. Hence, most genetic diversification takes place at very few loci in the genome. Querying these regions might be an efficient means of typing *Acinetobacter* strains for specific genotypes. Understanding the mechanisms leading to hotspots should enlighten how new genetic information is accommodated in the genome of *A. baumannii*.

Conclusions

We have proceeded to an extensive characterization of the molecular and evolutionary mechanisms driving the genetic diversification of *Acinetobacter*. Interestingly, we observed that temperate phages are much more abundant than conjugative elements, even though their role as vectors for horizontal transfer has been neglected in the past. Accordingly, we observed the presence of very complex and fast-evolving CRISPR-Cas systems in the genomes of *Acinetobacter*. Population genetics analyses are consistent with the notion that *A. baumannii* arose from an ancient population bottleneck. Nevertheless, this species is extremely diverse in terms of gene repertoires and shows strong effects of natural selection on protein evolution. Our study sets a solid basis for the understanding of the evolution of the *Acinetobacter* genus. Further work will be necessary to understand how genetic diversification leads to the key features of the genus, notably high metabolic diversity, antibiotic resistance, and virulence. The confrontation between the genetic and the phenotypic data should facilitate predicting how multiple pathogens rise within a genus by virtue of their genetic backgrounds and genetic plasticity.

Methods

Choice of strains. We analyzed 13 complete genomes retrieved from GenBank RefSeq in February 2013 (Pruitt et al. 2007), two *A. baumannii* genomes sequenced at the Pasteur Institute and at Walter Read and 118 genomes sequences derived from 116 strains (two sequences were obtained from each of the type strains of *A. indicus* and *A. brisouii*) which were sequenced at the Broad Institute (see details in Table S1) (Perichon et al. 2014). The 116 strains were selected from the collections of A. Nemec (strains designated NIPH or ANC) or of the Institut Pasteur (CIP strains) based on polyphasic taxonomic analyses to reflect the currently known breadth of the diversity of the genus *Acinetobacter* at the species level. Overall, 83 strains belonged to 29 validly named species (including *Acinetobacter grimontii*, a junior synonym of *A. junii*, but devoid of *Acinetobacter boissieri*, *Acinetobacter harbinensis*, *Acinetobacter puyangensis*, and *Acinetobacter qingfengensis*, which were unavailable at the time), 16 strains to eight genomic species as defined by DDH, and 10 strains to seven tentative novel species termed *Acinetobacter* taxons 18-23 and 26. The name '*Acinetobacter bohemicus*' has recently been proposed for taxon 26 (Krizova et al. 2014). Seven remaining strains were closely related to one of the species/taxa but were considered taxonomically unique at the species level based on our previous taxonomic analysis and the ANI data obtained in this study (these strains are termed *A. calcoaceticus*-like, *A. brisouii*-like, *A. pittii*-like, or Taxon 18-like). The *Acinetobacter* taxons are working taxonomic groups as delineated at the Laboratory of Bacterial Genetics (National Institute of Public Heath, Prague) based on the comprehensive physiological/nutritional testing, *rpoB* and 16S rDNA phylogenies and on whole-cell MALDI-TOF profiling. Each of these taxa is, at the species level, clearly distinct from any of the known species with a valid name, genomic species, or species with effectively published names. All validly named species were represented by the respective type strains while each genomic species included a strain used as a reference in previous DDH experiments. If more strains per species or taxon were included, these differed from respective type/reference strains in their microbiological and ecological characteristics. Organisms were grown, according to their physiological requirements, at 30-37°C in brain heart infusion (BHI) broth and agar (Difco Laboratories, Detroit, MI).

Core-genomes. We built core-genomes for the genus and for *A. baumannii*. Orthologs were identified as bidirectional best hits, using end-gap free global alignment, between the proteome of *A. baumannii* AYE as a pivot and each of the other proteomes (133 for the genus and 34 for the species). Hits with less than 40% (genus) or 80% (species) similarity in amino

acid sequence or more than 20% difference in protein length were discarded. Genomes from the same species typically show low levels of genome rearrangements and this information can be used to identify orthologs more accurately (Dandekar et al. 1998; Rocha 2006). Therefore, the core-genome of the species was defined as the intersection of pairwise lists of strict positional orthologs (as in (Touchon et al. 2009)). The core-genomes consist in the genes present in all genomes of the two sets. They were defined as the intersection of the lists of orthologs between pairs of genomes.

Pan-genomes. Pan-genomes were built by clustering homologous proteins into families. We determined the lists of putative homologs between pairs of genomes with blastp (Altschul et al. 1997) and used the e-values ($<10^{-4}$) to cluster them by similarity with Silix v1.2 (Miele et al. 2011). A protein is thus included in the family if it shares a relation of homology to a protein already in the family. Silix parameters were set such that a protein is homolog to another in a given family if the aligned part has at least 35% of identity and represents more than 80% of the smallest protein. The pan-genomes are the full complement of genes in the genus and in the species. The pan genomes of the 133 *Acinetobacter* proteomes (470582 proteins) and of the 34 *A. baumannii* proteomes (128266 proteins) were determined independently. We used a more stringent criterion of protein similarity (60%) to compare the pan-genomes of different species of *Acinetobacter*.

Phylogenetic analyses. Each of the 950 families of proteins of the *Acinetobacter* core-genome was used to produce a multiple alignment with muscle v3.8 (default parameters) (Edgar 2004). Poorly aligned regions were removed with BMGE (Criscuolo, Gribaldo 2010). The phylogenetic tree was inferred using the approximated maximum-likelihood method implemented in FastTree v1.4 with the WAG matrix and a gamma correction for variable evolutionary rates (Price et al. 2009). We performed 100 bootstrap experiments on the concatenated sequences to assess the robustness of the topology. To root the genus phylogenetic tree we used the genomes of species distant from each other and as close to *Acinetobacter* as possible. They were identified using a phylogenetic tree built with the 16S rDNA sequences of all the species of γ-proteobacteria with complete genomes in NCBI RefSeq (February 2013). The selection of a single strain and a single 16S copy per species resulted in 189 16S rDNA sequences that were aligned using MAFFT v7.1 (Katoh, Toh 2008). Poorly aligned regions were removed with Gblocks 0.91b (default parameters) (Castresana 2000). The phylogenetic tree was inferred using PhyML v3.0 under the HKY model and a gamma correction for variable evolutionary rates with eight classes (Gascuel et al. 2010). This

tree highlighted *Moraxella catarrhalis* (GenBank ID NC_014147) and *Psychrobacter* species (GenBank ID NC_009524) as the two species closest to the *Acinetobacter* genus in our dataset. We therefore reconstructed a core-genome of the genus plus these two species (same method as above). This core-genome included 677 protein families (Table S3) and was used to build a tree of the 135 complete genomes using the method mentioned above for the genus.

The reference phylogenetic tree of *A. baumannii* was reconstructed from the concatenated alignments of the 1590 protein families of the core-genome obtained with muscle v3.8 (default parameters). Since at this evolutionary distance the DNA sequences provide more phylogenetic signal than protein sequences, we back-translated the alignments to DNA, as is standard usage. Poorly aligned positions were removed with BMGE. The tree was inferred with FastTree v1.4 under the GTR model and a gamma correction for variable evolutionary rates with eight classes. We performed 100 bootstrap experiments on the concatenated sequences to assess the robustness of the topology. To root the species tree we used two closely related strains in the genus: *A. nosocomialis* NIPH 2119[T] and *A. pitti*-like ANC 4052, re-built the core-genome of the species including these two outgroups and performed a similar analysis on this dataset.

Recombination and population genetics analyses. We analyzed recombination in *A. baumannii* with RDP version 4.24 with default parameters except that RDP3 was used with the option "internal references only" and Geneconv with a G-scale mismatch penalty of 3 (Martin et al. 2010). The Phi test was computed using Phi Pack with default parameters, except that we made 10,000 permutations (Bruen et al. 2006). Analysis with or without permutations revealed highly correlated p-values (Spearman's rho=0.96, p<0.0001). We used the results of the analysis with permutations and applied a sequential Bonferroni correction of the p-values. The analyses with ClonalFrame (Didelot, Falush 2007) were done using default options (except for the options -x 50000 -y 50000) on 30 non-overlapping sets of ~51 genes contiguous in terms of their position in the genome of *A. baumannii* AYE. Analysis of Tajima D and Fay and Wu H was done with DnaSP 5.10.1 (Librado, Rozas 2009), using a sliding window of 1000 positions and a step of 250. Statistical tests were done on the average of the values on non-overlapping windows. Analysis of dN and dS was done on the concatenate of the alignments of the genes of the core genome using codeML from the PAML package version 4.4 (runmode = -2, model = 2, CodonFreq = 2) (Yang 2007).

Evolutionary distances. For each pair of genomes we computed a number of measures of similarity. 1) The phylogenetic distance was computed from the length of branches in the

genus phylogenetic tree using the *cophenetic* function in APE package from R (Paradis et al. 2004). 2) The gene repertoire relatedness (GRR) was computed as the number of homologs shared by the two genomes divided by the number of genes in the smallest genome (Snel et al. 1999). 3) The average nucleotide identity (ANI) was computed using Jspecies (Richter, Rossello-Mora 2009). 4) The spacer repertoire relatedness (SRR) was computed as the number of similar spacers shared by the two genomes divided by the number of spacers in the smallest CRISPR-arrays of the two genomes.

Identification of specific genes/systems. CRISPR-arrays were identified with the CRISPR Recognition Tool (CRT) using default parameters (Bland et al. 2007). The clusters of *cas* genes were identified as in (Touchon, Rocha 2010) using the classification recently proposed (Makarova et al. 2011). Protospacers were identified using BLASTN to search for similarities between CRISPR spacer sequences and all the phage (831) and plasmid (3861) genomes available in Genbank (default settings, e-value < 10^{-5}). We retained the matches showing ≥90% of identity and ≤10% difference in sequence length with the query. Prophages were detected with Phage Finder v.2.1 (Fouts 2006), discarding prophages less than 18 kb long. Loci encoding conjugative or mobilizable elements were identified with CONJscan (default parameters) (Guglielmini et al. 2011; Guglielmini et al. 2014). Integrases were searched with the PFAM profile PF00589 for tyrosine recombinases and the pair PF00239 and PF07508 for Serine recombinases. The tyrosine recombinases of integrons were identified from the integrases using a protein profile for a region specific to these proteins that we built based on published data (Cambray et al. 2011). Error-prone polymerases were searched for with PFAM profiles: PF00136 (Pol2 like PolB), PF00817 (Y-Polymerases like UmuC, DinP or ImuB) and PF07733 (DnaE2), and PF00717 (UmuD). The cassettes *imuABC* were searched with the pair of profiles PF00817-PF07733 and *umuCD* using PF00817-PF00717 (Galhardo et al. 2005; Norton et al. 2013). All the protein profiles were searched with hmmer 3.0 with default parameters (Eddy 2011). We removed from further analysis the hits with low e-values (e>0.001) or those for which the alignment matched less than half of the protein profile.

Identification of integration/deletion hotspots. The *A. baumannii* core-genome was used to identify and locate large integration/deletion (indel) regions. All regions including more than 10 genes between two consecutive core genes of the species were considered as large indel regions. The relative positions of these regions were defined by the order of the core genes in *A. baumannii* AYE. This strain was used as a reference to order *A. baumannii* genes, because it represents the most likely configuration of the chromosome in the ancestor of the species.

Regions located between two non-successive core genes, that is, with rearrangements in between them were removed.

Funding

References

Ackermann, HW, G Brochu, HP Emadi Konjin. 1994. Classification of Acinetobacter phages. Arch Virol 135:345-354.

Adams, MD, K Goglin, N Molyneaux, et al. 2008. Comparative genome sequence analysis of multidrug-resistant Acinetobacter baumannii. J Bacteriol 190:8053-8064.

Altschul, SF, TL Madden, AA Schäfer, J Zhang, Z Zhang, W Miller, D Lipman. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 25:3389-3402.

Alvarez-Perez, S, B Lievens, H Jacquemyn, CM Herrera. 2013. Acinetobacter nectaris sp. nov. and Acinetobacter boissieri sp. nov., isolated from floral nectar of wild Mediterranean insect-pollinated plants. IJSEM 63:1532-1539.

Antunes, LC, F Imperi, A Carattoli, P Visca. 2011. Deciphering the multifactorial nature of Acinetobacter baumannii pathogenicity. PLoS ONE 6:e22674.

Antunes, LC, P Visca, KJ Towner. 2013. Acinetobacter baumannii: evolution of a global pathogen. Pathog Dis.

Baba, T, T Ara, M Hasegawa, Y Takai, Y Okumura, M Baba, KA Datsenko, M Tomita, BL Wanner, H Mori. 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol 2:0008.

Barbe, V, D Vallenet, N Fonknechten, et al. 2004. Unique features revealed by the genome sequence of Acinetobacter sp. ADP1, a versatile and naturally transformation competent bacterium. Nucleic Acids Res 32:5766-5779.

Barrangou, R, LA Marraffini. 2014. CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity. Mol Cell 54:234-244.

Battistuzzi, FU, SB Hedges. 2009. Eubacteria. In: SB Hedges, S Kumar, editors. The Timetree of Life. New York: Oxford University Press. p. 106-115.

Bergogne-Berezin, E, KJ Towner. 1996. Acinetobacter spp. as nosocomial pathogens: microbiological, clinical, and epidemiological features. Clin Microbiol Rev 9:148-165.

Billard-Pomares, T, S Fouteau, ME Jacquet, et al. 2014. Characterization of a P1-like bacteriophage encoding an SHV-2 extended-spectrum beta-lactamase from an Escherichia coli strain. Antimicrob Agents Chemother 58:AAC.03183-03114.

Bissonnette, L, PH Roy. 1992. Characterization of In0 of Pseudomonas aeruginosa plasmid pVS1, an ancestor of integrons of multiresistance plasmids and transposons of gram-negative bacteria. J Bacteriol 174:1248-1257.

Bland, C, TL Ramsey, F Sabree, M Lowe, K Brown, NC Kyrpides, P Hugenholtz. 2007. CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. BMC Bioinformatics 8:209.

Bobay, LM, EP Rocha, M Touchon. 2013. The Adaptation of Temperate Bacteriophages to Their Host Genomes. Mol Biol Evol 30:737-751.

Bouvet, PJ, S Jeanjean. 1989. Delineation of new proteolytic genomic species in the genus Acinetobacter. Res Microbiol 140:291-299.

Bouvet, PJ, S Jeanjean, JF Vieu, L Dijkshoorn. 1990. Species, biotype, and bacteriophage type determinations compared with cell envelope protein profiles for typing Acinetobacter strains. J Clin Microbiol 28:170-176.

Bouvet, PJM, PAD Grimont. 1986. Taxonomy of the genus Acinetobacter with the recognition of Acinetobacter baumannii sp-nov, Acinetobacter haemolyticus sp-nov, Acinetobacter johnsonii sp-nov, and Acinetobacter junii sp-nov and emended descriptions of Acinetobacter calcoaceticus and Acinetobacter lwoffii. International journal of systematic bacteriology 36:228-240.

Bradley, DE. 1984. Characteristics and function of thick and thin conjugative pili determined by transfer-derepressed plasmids of incompatability groups I1, I2, I5, B, K and Z. J Gen Microbiol 130:1489-1502.

Bruen, TC, H Philippe, D Bryant. 2006. A simple and robust statistical test for detecting the presence of recombination. Genetics 172:2665-2681.

Cambray, G, N Sanchez-Alberola, S Campoy, E Guerin, S Da Re, B Gonzalez-Zorn, MC Ploy, J Barbe, D Mazel, I Erill. 2011. Prevalence of SOS-mediated control of integron integrase expression as an adaptive trait of chromosomal and mobile integrons. Mob DNA 2:6.

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17:540-552.

Chan, JZ, MR Halachev, NJ Loman, C Constantinidou, MJ Pallen. 2012. Defining bacterial species in the genomic era: insights from the genus Acinetobacter. BMC Microbiol 12:302.

Criscuolo, A, S Gribaldo. 2010. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol Biol 10:210.

Dandekar, T, B Snel, M Huynen, P Bork. 1998. Conservation of gene order: a fingerprint of proteins that physically interact. Trends Biochem Sci 23:324-328.

de Berardinis, V, D Vallenet, V Castelli, et al. 2008. A complete collection of single-gene deletion mutants of Acinetobacter baylyi ADP1. Mol Syst Biol 4:174.

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

Didelot, X, D Falush. 2007. Inference of bacterial microevolution using multilocus sequence data. Genetics 175:1251-1266.

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

Doucet-Populaire, F, P Trieu-Cuot, A Andremont, P Courvalin. 1992. Conjugal transfer of plasmid DNA from Enterococcus faecalis to Escherichia coli in digestive tracts of gnotobiotic mice. Antimicrob Agents Chemother 36:502-504.

Eddy, SR. 2011. Accelerated Profile HMM Searches. PLoS Comput Biol 7:e1002195.

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

Encinas, D, MP Garcillan-Barcia, M Santos-Merino, L Delaye, A Moya, F de la Cruz. 2014. Plasmid conjugation from Proteobacteria as evidence for the origin of xenologous genes in Cyanobacteria. J Bacteriol.

Eveillard, M, M Kempf, O Belmonte, H Pailhories, ML Joly-Guillou. 2013. Reservoirs of Acinetobacter baumannii outside the hospital and potential involvement in emerging human community-acquired infections. Int J Infect Dis 17:e802-805.

Falagas, ME, EA Karveli, I Kelesidis, T Kelesidis. 2007. Community-acquired Acinetobacter infections. Eur J Clin Microbiol Infect Dis 26:857-868.

Farrugia, DN, LD Elbourne, KA Hassan, BA Eijkelkamp, SG Tetu, MH Brown, BS Shah, AY Peleg, BC Mabbutt, IT Paulsen. 2013. The complete genome and phenome of a community-acquired Acinetobacter baumannii. PLoS ONE 8:e58628.

Fay, JC, CI Wu. 2000. Hitchhiking under positive Darwinian selection. Genetics 155:1405-1413.

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

Fondi, M, E Rizzi, G Emiliani, et al. 2013. The genome sequence of the hydrocarbon-degrading Acinetobacter venetianus VE-C3. Res Microbiol 164:439-449.

Fournier, PE, D Vallenet, V Barbe, et al. 2006. Comparative genomics of multidrug resistance in Acinetobacter baumannii. PLoS Genet 2:e7.

Fouts, DE. 2006. Phage_Finder: automated identification and classification of prophage regions in complete bacterial genome sequences. Nucleic Acids Res 34:5839-5851.

Galhardo, RS, RP Rocha, MV Marques, CF Menck. 2005. An SOS-regulated operon involved in damage-inducible mutagenesis in Caulobacter crescentus. Nucleic Acids Res 33:2603-2614.

Garcillan-Barcia, MP, MV Francia, F de la Cruz. 2009. The diversity of conjugative relaxases and its application in plasmid classification. FEMS Microbiol Rev 33:657-687.

Gascuel, O, S Guindon, JF Dufayard, V Lefort, M Anisimova, W Hordijk. 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Systematic Biology 59:307-321.

Gerischer, U, LN Ornston. 2001. Dependence of linkage of alleles on their physical distance in natural transformation of Acinetobacter sp. strain ADP1. Arch Microbiol 176:465-469.

Gerner-Smidt, P. 1989. Frequency of plasmids in strains of Acinetobacter calcoaceticus. The Journal of hospital infection 14:23-28.

Goldstein, FW, A Labigne-Roussel, G Gerbaud, C Carlier, E Collatz, P Courvalin. 1983. Transferable plasmid-mediated antibiotic resistance in Acinetobacter. Plasmid 10:138-147.

Guglielmini, J, B Neron, SS Abby, MP Garcillan-Barcia, FD la Cruz, EP Rocha. 2014. Key components of the eight classes of type IV secretion systems involved in bacterial conjugation or protein secretion. Nucleic Acids Res 42:5715-5727.

Guglielmini, J, L Quintais, MP Garcillan-Barcia, F de la Cruz, EP Rocha. 2011. The Repertoire of ICE in Prokaryotes Underscores the Unity, Diversity, and Ubiquity of Conjugation. PLoS Genet 7:e1002222.

Harding, CM, EN Tracy, MD Carruthers, PN Rather, LA Actis, RS Munson, Jr. 2013. Acinetobacter baumannii strain M2 produces type IV pili which play a role in natural transformation and twitching motility but not surface-associated motility. MBio 4:e00360-00313.

Harris, SR, EJ Feil, MT Holden, et al. 2010. Evolution of MRSA during hospital transmission and intercontinental spread. Science 327:469-474.

Hauck, Y, C Soler, P Jault, et al. 2012. Diversity of Acinetobacter baumannii in four French military hospitals, as assessed by multiple locus variable number of tandem repeats analysis. PLoS ONE 7:e44597.

Hujer, KM, AM Hujer, EA Hulten, et al. 2006. Analysis of antibiotic resistance genes in multidrug-resistant Acinetobacter sp. isolates from military and civilian patients treated at the Walter Reed Army Medical Center. Antimicrob Agents Chemother 50:4114-4123.

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

Juhas, M, DW Crook, DW Hood. 2008. Type IV secretion systems: tools of bacterial horizontal gene transfer and virulence. Cell Microbiol 10:2377-2386.

Karah, N, B Haldorsen, K Hegstad, GS Simonsen, A Sundsfjord, O Samuelsen. 2011. Species identification and molecular characterization of Acinetobacter spp. blood culture isolates from Norway. J Antimicrob Chemother 66:738-744.

Katoh, K, H Toh. 2008. Recent developments in the MAFFT multiple sequence alignment program. Brief Bioinform 9:286-298.

Kennemann, L, X Didelot, T Aebischer, et al. 2011. Helicobacter pylori genome evolution during human infection. Proc Natl Acad Sci U S A 108:5033-5038.

Konstantinidis, KT, A Ramette, JM Tiedje. 2006. The bacterial species definition in the genomic era. Philos Trans R Soc Lond B Biol Sci 361:1929-1940.

Krawczyk, B, K Lewandowski, J Kur. 2002. Comparative studies of the Acinetobacter genus and the species identification method based on the recA sequences. Molecular and Cellular Probes 16:1-11.

Krizova, L, M Maixnerova, O Sedo, A Nemec. 2014. Acinetobacter bohemicus sp. nov. widespread in natural soil and water ecosystems in the Czech Republic. Syst Appl Microbiol in press doi 10.1016/j.syapm.2014.07.001.

La Scola, B, VA Gundi, A Khamis, D Raoult. 2006. Sequencing of the rpoB gene and flanking spacers for molecular identification of Acinetobacter species. J Clin Microbiol 44:827-832.

Librado, P, J Rozas. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451-1452.

Makarova, KS, DH Haft, R Barrangou, et al. 2011. Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol 9:467-477.

Martin, DP, P Lemey, M Lott, V Moulton, D Posada, P Lefeuvre. 2010. RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics 26:2462-2463.

Mather, AE, SW Reid, DJ Maskell, et al. 2013. Distinguishable epidemics of multidrug-resistant Salmonella Typhimurium DT104 in different hosts. Science 341:1514-1517.

McGann, P, P Courvalin, E Snesrud, et al. 2014. Amplification of Aminoglycoside Resistance Gene aphA1 in Acinetobacter baumannii Results in Tobramycin Therapy Failure. MBio 5:e00915.

Metzgar, D, JM Bacher, V Pezo, J Reader, V Doring, P Schimmel, P Marliere, V de Crecy-Lagard. 2004. Acinetobacter sp. ADP1: an ideal model organism for genetic analysis and genome engineering. Nucleic Acids Res 32:5780-5790.

Miele, V, S Penel, L Duret. 2011. Ultra-fast sequence clustering from similarity networks with SiLiX. BMC Bioinformatics 12:116.

Miller, F, G Phan, T Brissac, C Bouchiat, G Lioux, X Nassif, M Coureuil. 2014. The Hypervariable Region of Meningococcal Major Pilin PilE Controls the Host Cell Response via Antigenic Variation. MBio 5:e01024-01013.

Muniesa, M, M Colomer-Lluch, J Jofre. 2013. Potential impact of environmental bacteriophages in spreading antibiotic resistance genes. Future Microbiol 8:739-751.

Nassif, X, J Lowy, P Stenberg, P O'Gaora, A Ganji, M So. 1993. Antigenic variation of pilin regulates adhesion of Neisseria meningitidis to human epithelial cells. Mol Microbiol 8:719-725.

Nemec, A, T De Baere, I Tjernberg, M Vaneechoutte, TJ van der Reijden, L Dijkshoorn. 2001. Acinetobacter ursingii sp. nov. and Acinetobacter schindleri sp. nov., isolated from human clinical specimens. Int J Syst Evol Microbiol 51:1891-1899.

Nemec, A, L Dijkshoorn, I Cleenwerck, T De Baere, D Janssens, TJ Van Der Reijden, P Jezek, M Vaneechoutte. 2003. Acinetobacter parvus sp. nov., a small-colony-forming species isolated from human clinical specimens. Int J Syst Evol Microbiol 53:1563-1567.

Nemec, A, L Krizova, M Maixnerova, TJ van der Reijden, P Deschaght, V Passet, M Vaneechoutte, S Brisse, L Dijkshoorn. 2011. 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 162:393-404.

Nemec, A, M Musilek, M Maixnerova, T De Baere, TJ van der Reijden, M Vaneechoutte, L Dijkshoorn. 2009. Acinetobacter beijerinckii sp. nov. and Acinetobacter gyllenbergii sp. nov., haemolytic organisms isolated from humans. IJSEM 59:118-124.

Nemec, A, M Musilek, O Sedo, T De Baere, M Maixnerova, TJ van der Reijden, Z Zdrahal, M Vaneechoutte, L Dijkshoorn. 2010. Acinetobacter bereziniae sp. nov. and Acinetobacter guillouiae sp. nov., to accommodate Acinetobacter genomic species 10 and 11, respectively. IJSEM 60:896-903.

Nemec, A, M Musilek, M Vaneechoute, E Falsen, L Dijkshoorn. 2008. Lack of evidence for "Acinetobacter septicus" as a species different from Acinetobacter ursingii? J Clin Microbiol 46:2826-2827; author reply 2827.

Nielsen, KM, AM Bones, JD Van Elsas. 1997. Induced Natural Transformation of Acinetobacter calcoaceticus in Soil Microcosms. Appl Environ Microbiol 63:3972-3977.

Nishimura, Y, T Ino, H Iizuka. 1988. Acinetobacter radioresistens sp. nov. Isolated from Cotton and Soil. Int J Syst Bact 38:209-211.

Norton, MD, AJ Spilkia, VG Godoy. 2013. Antibiotic resistance acquired through a DNA damage-inducible response in Acinetobacter baumannii. J Bacteriol 195:1335-1345.

Ochman, H, NA Moran. 2001. Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. Science 292:1096-1099.

Paradis, E, J Claude, K Strimmer. 2004. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics 20:289-290.

Peleg, AY, A de Breij, MD Adams, et al. 2012. The success of acinetobacter species; genetic, metabolic and virulence attributes. PLoS ONE 7:e46984.

Peleg, AY, H Seifert, DL Paterson. 2008. Acinetobacter baumannii: emergence of a successful pathogen. Clin Microbiol Rev 21:538-582.

Perichon, B, S Goussard, V Walewski, et al. 2014. Identification of 50 Class D beta-Lactamases and 65 Acinetobacter-Derived Cephalosporinases in Acinetobacter spp. Antimicrob Agents Chemother 58:936-949.

Ploy, MC, F Denis, P Courvalin, T Lambert. 2000. Molecular characterization of integrons in Acinetobacter baumannii: description of a hybrid class 2 integron. Antimicrob Agents Chemother 44:2684-2688.

Poirel, L, S Figueiredo, V Cattoir, A Carattoli, P Nordmann. 2008. Acinetobacter radioresistens as a silent source of carbapenem resistance for Acinetobacter spp. Antimicrob Agents Chemother 52:1252-1256.

Porstendorfer, D, O Gohl, F Mayer, B Averhoff. 2000. ComP, a pilin-like protein essential for natural competence in Acinetobacter sp. Strain BD413: regulation, modification, and cellular localization. J Bacteriol 182:3673-3680.

Price, MN, PS Dehal, AP Arkin. 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol 26:1641-1650.

Pruitt, KD, T Tatusova, DR Maglott. 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res 35:D61-65.

Rainey, FA, E Lang, E Stackebrandt. 1994. The phylogenetic structure of the genus Acinetobacter. FEMS Microbiol Lett 124:349-353.

Richter, M, R Rossello-Mora. 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U S A 106:19126-19131.

Robinson, A, AJ Brzoska, KM Turner, R Withers, EJ Harry, PJ Lewis, NE Dixon. 2010. Essential biological processes of an emerging pathogen: DNA replication, transcription, and cell division in Acinetobacter spp. Microbiol Mol Biol Rev 74:273-297.

Rocha, EPC. 2004. Order and disorder in bacterial genomes. Curr Op Microbiol 7:519–527.

Rocha, EPC. 2006. Inference and Analysis of the Relative Stability of Bacterial Chromosomes. Mol Biol Evol 23:513–522.

Rocha, EPC, J Maynard Smith, LD Hurst, MT Holden, JE Cooper, NH Smith, E Feil. 2006. Comparisons of dN/dS are time-dependent for closely related bacterial genomes. J Theor Biol 239:226-235.

Rodriguez-Bano, J, S Marti, A Ribera, F Fernandez-Cuenca, L Dijkshoorn, A Nemec, M Pujol, J Vila. 2006. Nosocomial bacteremia due to an as yet unclassified acinetobacter genomic species 17-like strain. J Clin Microbiol 44:1587-1589.

Sahl, JW, JD Gillece, JM Schupp, VG Waddell, EM Driebe, DM Engelthaler, P Keim. 2013. Evolution of a pathogen: a comparative genomics analysis identifies a genetic pathway to pathogenesis in Acinetobacter. PLoS ONE 8:e54287.

Sahl, JW, JK Johnson, AD Harris, AM Phillippy, WW Hsiao, KA Thom, DA Rasko. 2011. Genomic comparison of multi-drug resistant invasive and colonizing Acinetobacter baumannii isolated from diverse human body sites reveals genomic plasticity. BMC Genomics 12:291.

Shen, GH, JL Wang, FS Wen, KM Chang, CF Kuo, CH Lin, HR Luo, CH Hung. 2012. Isolation and characterization of phikm18p, a novel lytic phage with therapeutic potential against extensively drug resistant Acinetobacter baumannii. PLoS ONE 7:e46537.

Smet, A, F Boyen, F Pasmans, P Butaye, A Martens, A Nemec, P Deschaght, M Vaneechoutte, F Haesebrouck. 2012. OXA-23-producing Acinetobacter species from horses: a public health hazard? The Journal of antimicrobial chemotherapy 67:3009-3010.

Smillie, C, MP Garcillan-Barcia, MV Francia, EP Rocha, F de la Cruz. 2010. Mobility of plasmids. Microbiol Mol Biol Rev 74:434-452.

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

Snel, B, P Bork, MA Huynen. 1999. Genome phylogeny based on gene content. Nat Genet 21:108-110.

Snitkin, ES, AM Zelazny, CI Montero, F Stock, L Mijares, PR Murray, JA Segre. 2011. Genome-wide recombination drives diversification of epidemic strains of Acinetobacter baumannii. Proc Natl Acad Sci U S A 108:13758-13763.

Sorek, R, CM Lawrence, B Wiedenheft. 2013. CRISPR-mediated adaptive immune systems in bacteria and archaea. Annu Rev Biochem 82:237-266.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585-595.

Team, RC. 2014. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Tenaillon, O, E Denamur, I Matic. 2004. Evolutionary significance of stress-induced mutagenesis in bacteria. Trends Microbiol 12:264-270.

Thornton, K. 2005. Recombination and the properties of Tajima's D in the context of approximate-likelihood calculation. Genetics 171:2143-2148.

Tjernberg, I, J Ursing. 1989. Clinical strains of Acinetobacter classified by DNA-DNA hybridization. APMIS : acta pathologica, microbiologica, et immunologica Scandinavica 97:595-605.

Touchon, M, C Hoede, O Tenaillon, et al. 2009. Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths. PLoS Genet 5:e1000344.

Touchon, M, EP Rocha. 2010. The small, slow and specialized CRISPR and anti-CRISPR of Escherichia and Salmonella. PLoS ONE 5:e11126.

Towner, K. 2006. The Genus Acinetobacter.746-758.

Turton, JF, J Shah, C Ozongwu, R Pike. 2010. Incidence of Acinetobacter species other than A. baumannii among clinical isolates of Acinetobacter: evidence for emerging species. J Clin Microbiol 48:1445-1449.

Turton, JF, ME Ward, N Woodford, ME Kaufmann, R Pike, DM Livermore, TL Pitt. 2006. The role of ISAba1 in expression of OXA carbapenemase genes in Acinetobacter baumannii. FEMS Microbiol Lett 258:72-77.

Vallenet, D, P Nordmann, V Barbe, et al. 2008. Comparative analysis of Acinetobacters: three genomes for three lifestyles. PLoS ONE 3:e1805.

Vaneechoutte, M, T De Baere, A Nemec, M Musilek, TJ van der Reijden, L Dijkshoorn. 2008. Reclassification of Acinetobacter grimontii Carr et al. 2003 as a later synonym of Acinetobacter junii Bouvet and Grimont 1986. IJSEM 58:937-940.

Vaz-Moreira, I, A Novo, E Hantsis-Zacharov, AR Lopes, M Gomila, OC Nunes, CM Manaia, M Halpern. 2011. Acinetobacter rudis sp. nov., isolated from raw milk and raw wastewater. IJSEM 61:2837-2843.

Wilharm, G, J Piesker, M Laue, E Skiebe. 2013. DNA Uptake by the Nosocomial Pathogen Acinetobacter baumannii Occurs during Movement along Wet Surfaces. J Bacteriol 195:4146-4153.

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

Williams, KP, JJ Gillespie, BW Sobral, EK Nordberg, EE Snyder, JM Shallom, AW Dickerman. 2010. Phylogeny of gammaproteobacteria. J Bacteriol 192:2305-2314.

Wright, MS, DH Haft, DM Harkins, F Perez, KM Hujer, S Bajaksouzian, MF Benard, MR Jacobs, RA Bonomo, MD Adams. 2014. New insights into dissemination and variation of the health care-associated pathogen Acinetobacter baumannii from genomic analysis. MBio 5:e00963-00913.

Yamahira, K, K Hirota, K Nakajima, N Morita, Y Nodasaka, I Yumoto. 2008. Acinetobacter sp. strain Ths, a novel psychrotolerant and alkalitolerant bacterium that utilizes hydrocarbon. Extremophiles 12:729-734.

Yamamoto, S, PJ Bouvet, S Harayama. 1999. Phylogenetic structures of the genus Acinetobacter based on gyrB sequences: comparison with the grouping by DNA-DNA hybridization. International journal of systematic bacteriology 49 Pt 1:87-95.

Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586-1591.

Yoon, EJ, P Courvalin, C Grillot-Courvalin. 2013. RND-type efflux pumps in multidrug-resistant clinical isolates of Acinetobacter baumannii: major role for AdeABC overexpression and AdeRS mutations. Antimicrob Agents Chemother 57:2989-2995.

Figure legends

Figure 1. Core- and pan-genomes of the genus and of *A. baumannii* (left) and spectrum of frequencies for *A. baumannii* gene repertoires (right). The pan- and core-genomes were used to perform gene accumulation curves using the statistical software R (Team 2014). These curves describe the number of new genes (pan-genome) and genes in common (core-genome) obtained by adding a new genome to a previous set. The procedure was repeated 1000 times by randomly modifying the order of integration of genomes in the analysis. The spectrum of frequencies (right) represents the number of genomes where the families of the pan-genome can be found, from 1 for strain-specific genes to 34 for core genes. Red indicates accessory genes and green the genes that are highly persistent in *A. baumannii*.

Figure 2. Comparisons of the pan-genome of *A. baumannii* computed with random samples of different size (boxplots) with those of the other species or genomic species. Each species, except *A. baumannii*, is only represented once, in the graph corresponding to the full number of available genomes for the taxa (*e.g.*, nine genomes for *A. lwoffii*). The boxplots show the distribution of the size of the pan-genome of *A. baumannii* using random samples of K *A. baumannii* genomes (K={2, 3, 4, 8, 9} genomes). Green dots correspond to pan-genomes of other species that are within the 25-75 percentiles of the distribution of the pan-genomes of *A. baumannii*, *i.e.*, these are pan-genomes approximately the size of *A. baumannii* given the same number of genomes. Red dots correspond to species with pan-genomes smaller than 75% of the *A. baumannii* pan-genomes (see Table S5 for full data).

Figure 3. Phylogeny of the *Acinetobacter* genus based on the alignment of the protein families of the core genome (see Methods). Triangles mark groups of taxa that are from the same species or have more than 95% average nucleotide identity (ANI) values and therefore might be regarded as coming from the same species. The nodes in red have bootstrap supports higher than 95%. The tree was rooted using two outgroup genomes (see main text).

Figure 4. Analysis of the association between average nucleotide identity (ANI) and gene repertoire relatedness (GRR) (see Methods). The points in black correspond to the clades in triangles in Figure 3. The points in gray correspond to comparisons between genomes that are closely related but not of the same species. We highlight three clades where some strains are closely related to the genomic species 13, *A. pittii*, and *A. calcoaceticus*.

Figure 5. Analysis of the association between gene repertoire relatedness (GRR) and the phylogenetic distance. Points in black indicate comparisons between pairs of genomes of the

same species/genomic species (triangles in Figure 3) and points in gray indicate the other pairs. The red line is a spline fit of the data. The insert shows the relation between the evolutionary distance and the probability that comparing two genomes will result in a GRR value higher than the average within-species GRR (red) and higher than the minimal within-species GRR (green).

Figure 6. Distribution of elements potentially related with genetic diversification in the genus. White indicates absence of the trait and black its presence. Genomes with many elements of a given type are indicated in red and those with few elements are indicated in yellow. Intermediate values are indicated in shades of orange. Black asterisks indicate complete genomes from GenBank.

Figure 7. Distribution of elements potentially related with genetic diversification in *A. baumannii*. White indicates absence of the trait and black its presence. Genomes with many elements of a given type are indicated in red and those with few elements are indicated in yellow. Intermediate values are indicated in shades of orange. Black asterisks indicate complete genomes from GenBank, blue asterisks indicate genomes sequenced at Pasteur Institute and at Walter Read.

Figure 8. Molecular phylogeny of the Cas1 protein across the genus. Phylogenetic tree for the Cas1 proteins was performed using PhyML with the WAG model and a Gamma correction. Cluster of *cas* genes organization, the most common repeat sequence, and the number of repeat sequences in each genome are indicated on the right part of the figure. Black circles indicate incomplete CRISPR-Cas systems. The left insert shows the genomes sharing spacers, each edge corresponds to the SRR (see Methods). Each color corresponds to a given species.
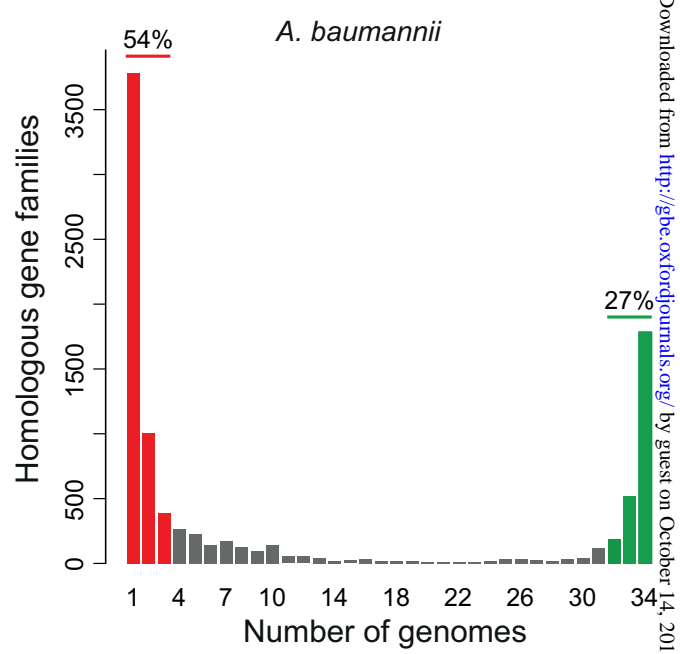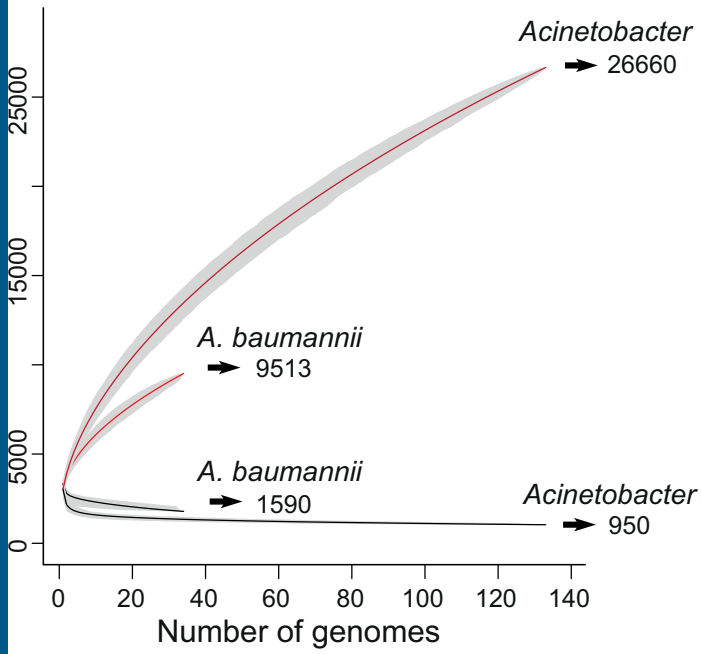
Figure 9. Distribution of genes of the core-genome of *A. baumannii* presenting significant evidence of recombination using Phi (p<0.05 after sequential Bonferroni correction) computed in sliding windows of 50 core genes. The dashed line indicates the average.
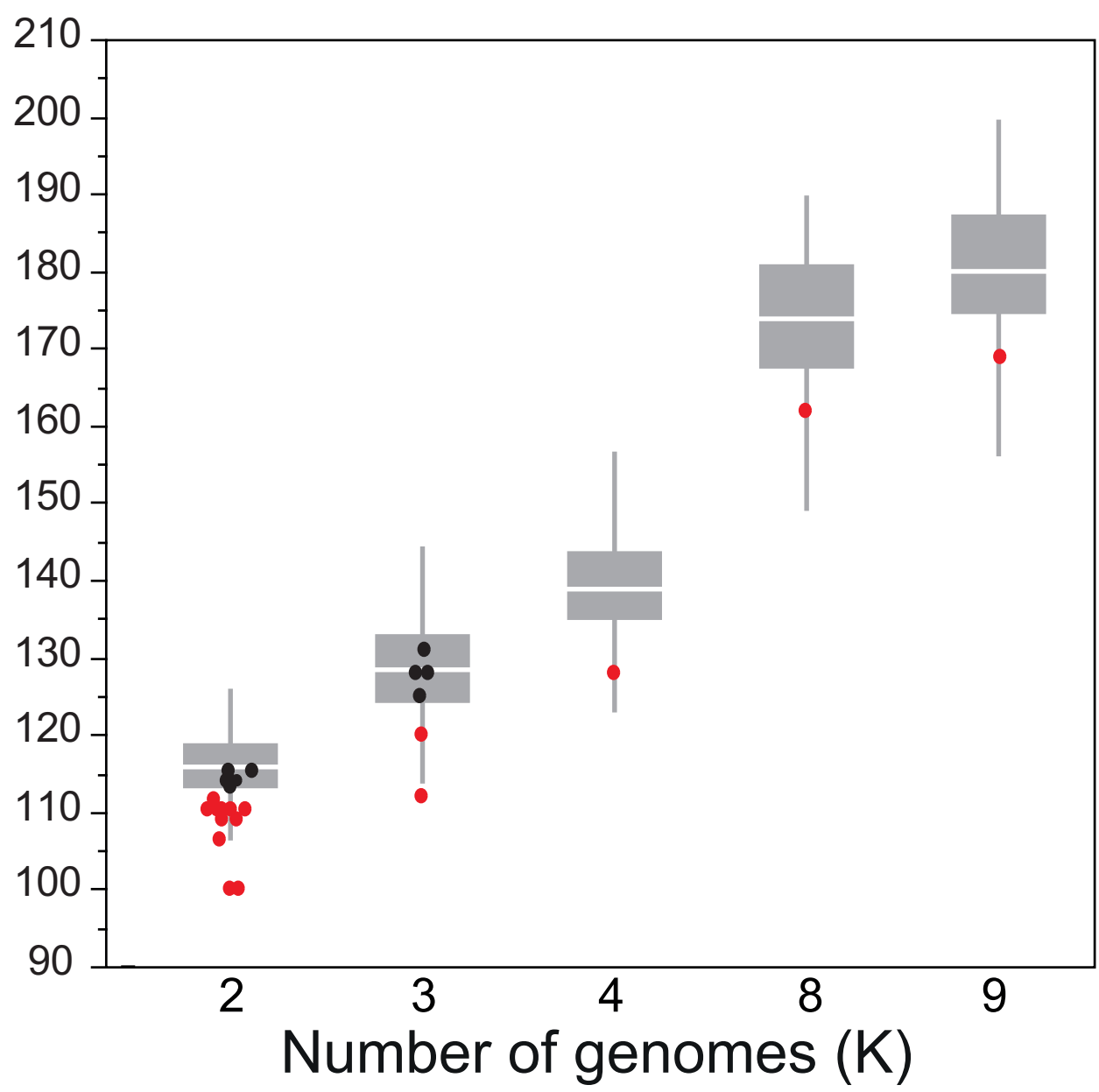
Figure 10. Distribution of integration/deletion hotspots along the core-genome of *A. baumannii* using gene orders of *A. baumannii* AYE strain as a reference (see Methods). The bars represent the number of different gene families in all the genomes found between two consecutive genes of the core genome. The colors represent the diversity of these gene families, i.e., the number of gene families divided by the number of genes found between two consecutive genes of the core genome. If the number of genes is identical to the number of gene families (1, maximal diversity), then every genome has a different set of genes in the
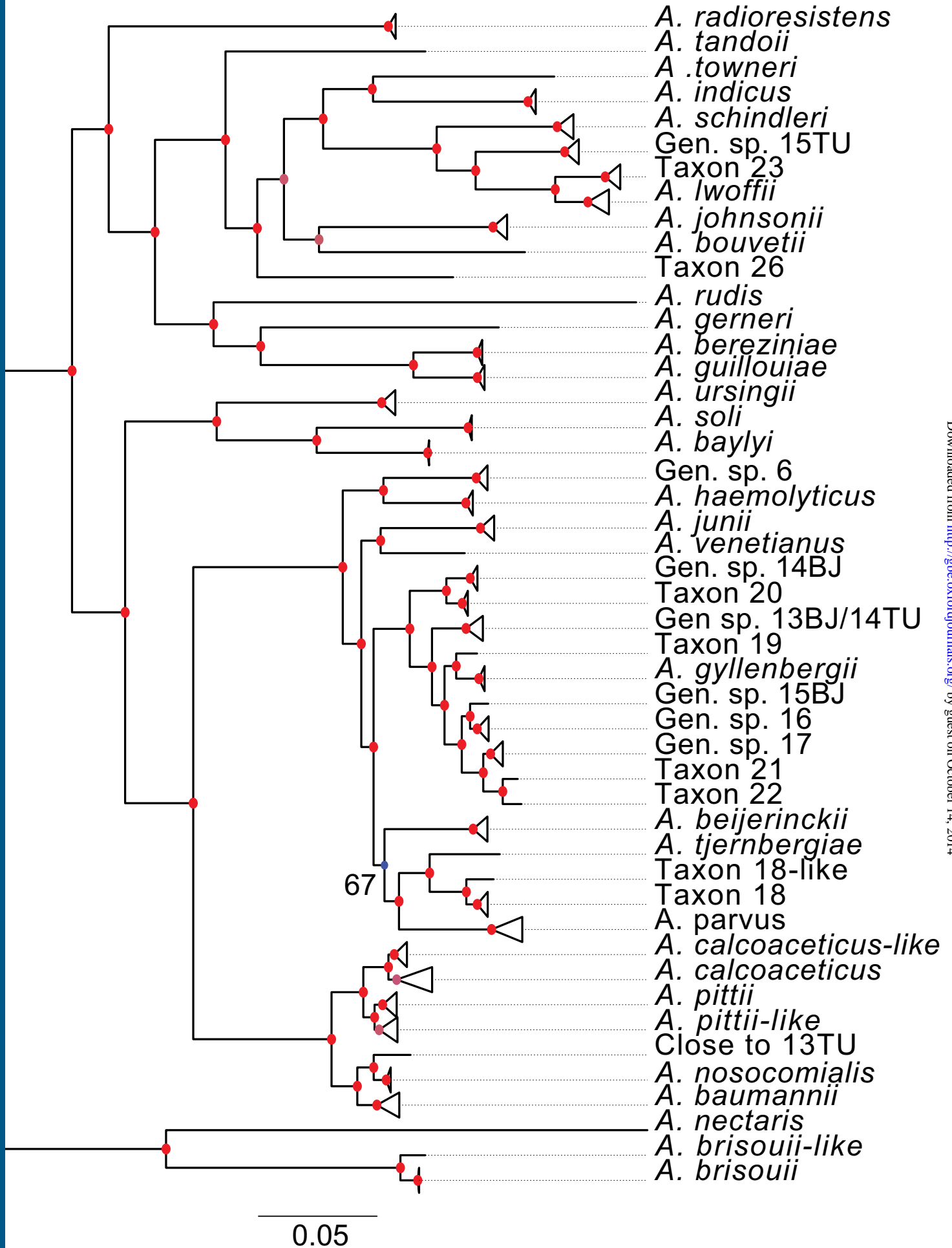
hotspot indicating many different insertions in the region. If the number of families equals the number of genes per genome (close to 1/33, minimal diversity), then most genomes have the same genes in the hotspot. This last scenario typically corresponds to strain-specific large deletions.
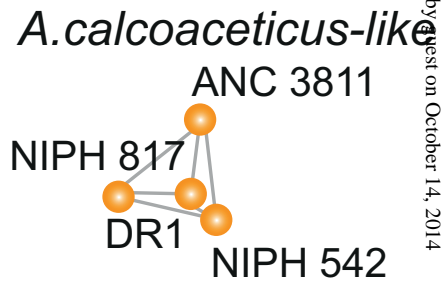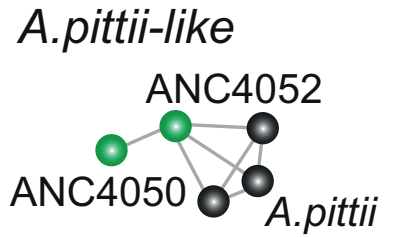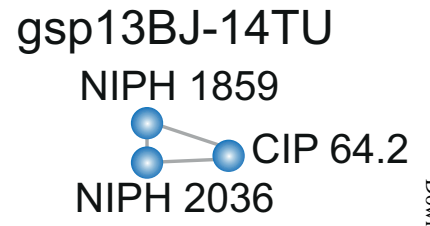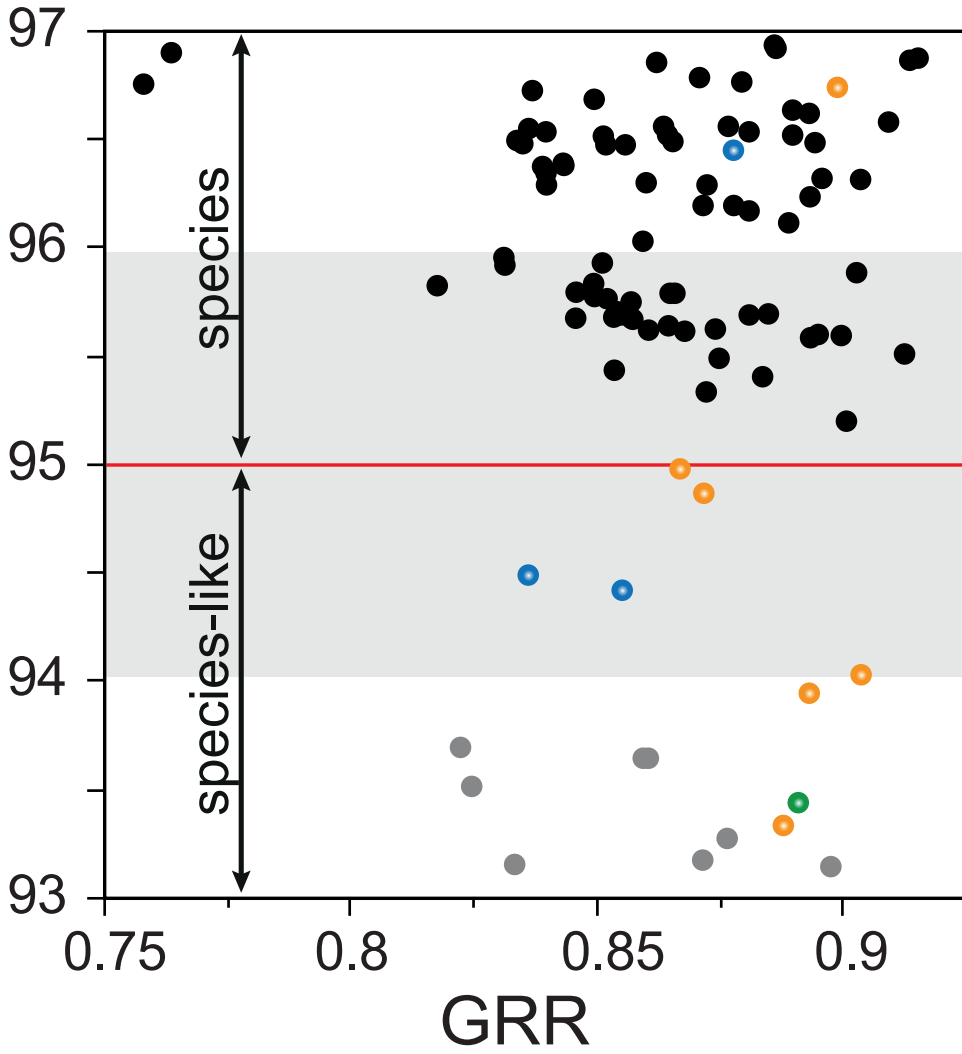
Table 1 – Core and pan-genomes for species of the *Acinetobacter* genus with at least three sequenced genomes (see Table S5 for all species). For each species the number of genomes (#), size of the core- and pan-genome, and percentage of the two relative to the size of the average genome in the clade are indicated.

Species	#	Core-genome		Pan-genome	
		size	%	size	%
A. calcoaceticus	4	2951	81%	4677	128%
A. indicus	3	2340	79%	3309	112%
A. lwoffii	9	2161	66%	5557	169%
A. parvus	8	1810	64%	4576	162%
A. pittii	3	2926	81%	4282	120%
A. schindleri	3	2391	76%	3929	125%
A. ursingii	3	2458	72%	4353	128%
A. junii	3	2293	70%	4292	131%
Genomic sp. 13BJ/14TU	3	2782	73%	4839	128%
Genomic sp. 16	3	3222	78%	5210	125%
A. baumannii	34	1590	42%	10849	288%

A. radioresistens
A. tandoii
A .towneri
A. indicus
A. schindleri
Gen. sp. 15TU
Taxon 23
A. lwoffii
A. johnsonii
A. bouvetii
Taxon 26
A. rudis
A. gerneri
A. bereziniae
A. guillouiae
A. ursingii
A. soli
A. baylyi
Gen. sp. 6
A. haemolyticus
A. junii
A. venetianus
Gen. sp. 14BJ
Taxon 20
Gen sp. 13BJ/14TU
Taxon 19
A. gyllenbergii
Gen. sp. 15BJ
Gen. sp. 16
Gen. sp. 17
Taxon 21
Taxon 22
A. beijerinckii
A. tjernbergiae
Taxon 18-like
Taxon 18
A. parvus
A. calcoaceticus-like
A. calcoaceticus
A. pittii
A. pittii-like
Close to 13TU
A. nosocomialis
A. baumannii
A. nectaris
A. brisouii-like
A. brisouii

67

0.05

Public genome
Bootstrap values
Type strain

Column headers: Genome Size | G+C Content | Competence | Flagellum | Conjugation | Mobilization | Prophage | Integron | Transposase | CRISPR-Cas | PolV | YpdA
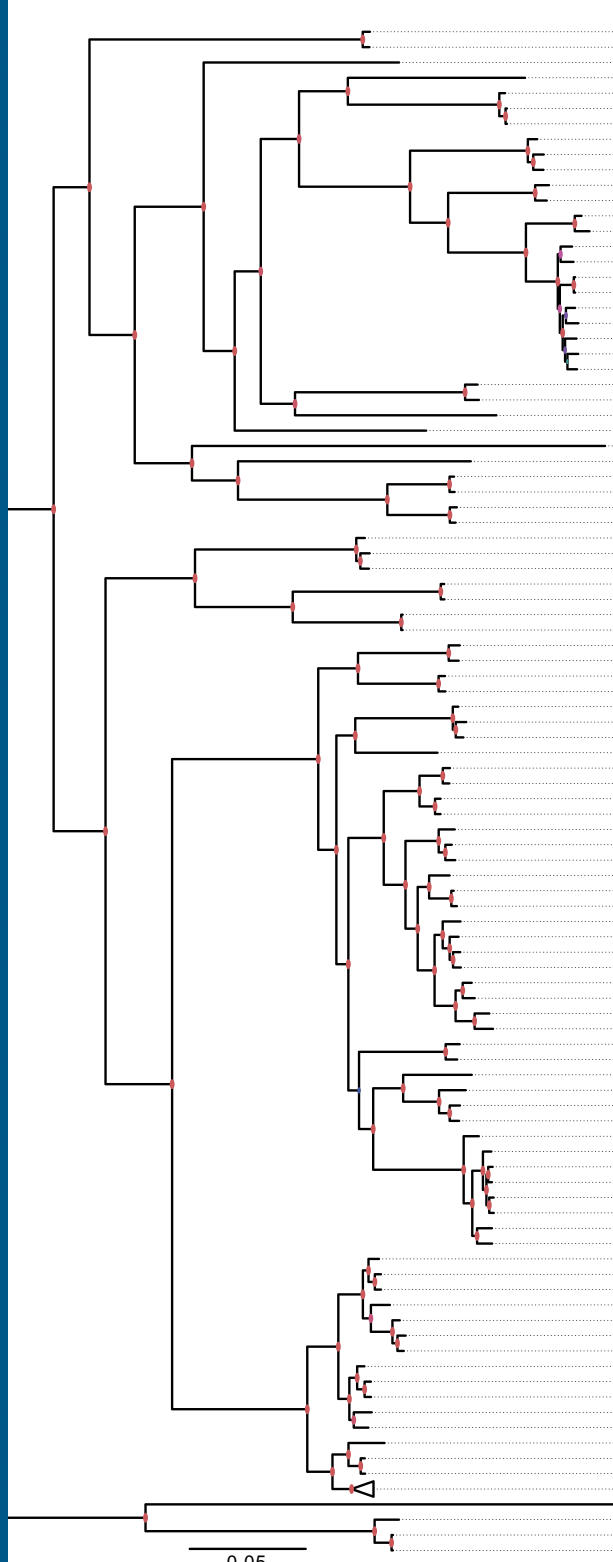
Taxon
A. radioresistens CIP 103788T
A. radioresistens NIPH 2130
A. tandoii CIP 107469T
A .towneri CIP 107472T
A. indicus CIP 53.82
A. indicus ANC 4215T
A. indicus CIP 110367T
A. schindleri NIPH 900
A. schindleri CIP 107287T
A. schindleri CIP 101934
Gen. sp. 15TU NIPH 2171
Gen. sp. 15TU NIPH 899
Taxon 23 CIP 64.10
Taxon 23 NIPH 713
A. lwoffii NIPH 478
A. lwoffii CIP 51.11
A. lwoffii NIPH 512
A. lwoffii CIP A162
A. lwoffii NIPH 715
A. lwoffii CIP 64.7
A. lwoffii CIP 102136
A. lwoffii CIP 70.31
A. lwoffii CIP101966
A. johnsonii ANC 3681
A. johnsonii CIP 64.6T
A. bouvetii CIP 107468T
Taxon 26 ANC 3994
A. rudis CIP 110305T
A. gerneri CIP 107464T
A. bereziniae CIP 70.12T
A. bereziniae NIPH 3
A. guillouae CIP 63.46T
A. guillouae NIPH 991
A. ursingii NIPH 706
A. ursingii CIP 107286T
A. ursingii ANC 3649
A. soli CIP 110264T
A. soli NIPH 2899
A. baylyi CIP 107474T
A. baylyi ADP1
Gen. sp. 6 CIP A165
Gen. sp. 6 NIPH 298
A. haemolyticus CIP 64.3T
A. haemolyticus NIPH 261
A. junii NIPH 182
A. junii CIP 107470
A. junii CIP 64.5T
A. venetianus CIP 110063T
Gen. sp. 14BJ NIPH 1847
Gen. sp. 14BJ NIPH 3623
Taxon 20 NIPH 2168
Taxon 20 NIPH 758
Gen. sp. 13BJ/14TU CIP 64.2
Gen.sp. 13BJ/14TU NIPH 1859
Gen.sp. 13BJ/14TU NIPH 2036
Taxon 19 NIPH 809
A. gyllenbergii CIP 110306T
A. gyllenbergii NIPH 230
Gen. sp. 15BJ CIP 110321
Gen. sp. 16 CIP 56.2
Gen. sp. 16 ANC 3880
Gen. sp. 16 CIP70.18
Gen. sp. 17 ANC 4105
Gen. sp. 17 NIPH 1867
Taxon 21 ANC 3929
Taxon 22 NIPH 2100
A. beijerinckii ANC 3835
A. beijerinckii CIP 110307T
A. tjernbergiae CIP 107465T
Taxon 18-like NIPH 284
Taxon 18 ANC 3862
Taxon 18 NIPH 236
A. parvus NIPH 1103
A. parvus CIP 108168T
A. parvus CIP 102143
A. parvus CIP 102529
A. parvus CIP 102082
A. parvus CIP 102159
A. parvus CIP 102129
A. parvus CIP 102637
A. calcoaceticus-like NIPH 817
A. calcoaceticus-like DR1
A. calcoaceticus-like NIPH 542
A. calcoaceticus-like ANC 3811
A. calcoaceticus ANC 3680
A. calcoaceticus CIP 81.8T
A. calcoaceticus NIPH 13
A. pittii PHEA-2
A. pittii ANC 3678
A. pittii CIP 70.29T
A. pittii-like ANC 4050
A. pittii-like ANC 4052
Close to 13TU NIPH 973
A. nosocomialis NIPH 2119T
A. nosocomialis NIPH 386
A. baumannii NIPH 527
A. nectaris CIP 110549T
A. brisouii-like ANC 3789
A. brisouii ANC 4119T
A. brisouii CIP 110357T

0.05

ISPR-CAS Type III

Subtype III-A

GTCTGAACAGACTTCCCTGTAGATAAGGGATTAAGAC

Gen. sp. 15TU NIPH 899	3+6
A. junii NIPH 182	5+4

cas2 cas1 cas2   cas10 csm2 csm3 csm4 csm5 cas6 csx1

Subtype III-A

A. venetianus CIP 110063	120
Gen. sp. 15BJ CIP 110321	68
A. baumannii CIP 70.34	61
A. calcoaceticus CIP 81.8	88
A. radioresistens CIP103788	23
Taxon 20 NIPH 758	74
Taxon 18 NIPH 236	78
A. baylyi ADP1	91
A. baylyi CIP 107474	100
A. soli NIPH 2899	31
A. bereziniae NIPH 3	64
A. gerneri CIP 107464	41

Subtype I-Fa

GTTCTTCATCGCATAGATGATTTAGAAA

cas1 cas6 csy3 csy2 csy1   cas3-cas2

Subtype I-Fa

ISPR-CAS Type I

A. brisouii-like ANC 3789	5
A. baumannii NIPH 527	29
A. baumannii NIPH 290	54
A. baumannii NIPH 201	98
A. baumannii ANC 4097	46
A. baumannii AB0057	53
A. baumannii AB307-0294	46
A. baumannii AYE	60
A. baumannii MRSN 3405	77
Gen. sp.16 ANC 3880	53 ●
Gen. sp. 16 CIP 56.2	40
A. gyllenbergii NIPH 230	55
Taxon 19 NIPH 809	90
A. haemolyticus CIP 64.3	26 ●
A. baumannii NIPH 190	55
A. baumannii NIPH 146	19
A. junii CIP 64.5	148
A.baumannii NIPH 615	111
A. parvus CIP 102529	119
A. parvus CIP 102143	249
A. parvus CIP 108168	135
A. parvus CIP 102159	83
A. parvus CIP 102082	65
A. parvus NIPH 1103	5 ●
A. parvus CIP 102637	204
A. parvus CIP 102129	118
A. ursingii ANC 3649	86
A. ursingii CIP 107286	9 ●
A. ursingii NIPH 706	304
A. tjernbergiae CIP 107465	6
Taxon 18 ANC 3862	140
Gen. sp. 6 CIP A165	83 ●
Taxon 18-like NIPH 284	6
A. junii CIP 107470	3
A. lwoffii NIPH 715	71
A. schindleri CIP 107287	3
Taxon 26 ANC 3994	22

Subtype I-Fb

0.8

GTTCATGGCGGCATACGCCATTTAGAAA

cas6 csy3 csy2 csy1   cas3-cas2   cas1

ABAYE1010 ABAYE0994

Subtype I-Fb

NIPH 201

A. parvus

A. baumannii

NIPH 146 NIPH 615

A. baylyi

Ori

Ter

Position along the core genome (# genes)