Research paper

CRISPR-based subtyping to track the evolutionary history of a global clone of *Acinetobacter baumannii*

Nabil Karah [*], Sun Nyunt Wai, Bernt Eric Uhlin [*]

Department of Molecular Biology, the Laboratory for Molecular Infection Medicine Sweden (MIMS), Umeå Centre for Microbial Research (UCMR), Umeå University, Umeå, Sweden

ABSTRACT

Acinetobacter baumannii global clone 1 (GC1) is the second most common clone in the global population of *A. baumannii* isolates and a key cause of hospital-acquired infections. In this study, comparative analysis of the clustered regularly interspaced short palindromic repeats (CRISPR)-based sequence types (CST) was performed to determine the genetic relatedness and track patterns of descent among 187 GC1 isolates, as a complement to the evolutionary inferences from their multilocus sequence types and genome-wide single nucleotide polymorphism (SNP)-based phylogeny.

The CST2 cluster, CST2 and all the CSTs descending from CST2, corresponded to GC1 lineage 1. This cluster included 143 of the 187 isolates showing a prevalent geographical distribution worldwide. A well-demarcated group of 13 CSTs, accounting for 33 of the 187 isolates, corresponded to GC1 lineage 2. All the CSTs of this group were characterized by the absence of spacer Ab-18. Many of the GC1 lineage 2 isolates had an epidemiological link to the Middle East and/or were obtained in military healthcare facilities. GC1 lineage 3 was a novel lineage that has so far been limited to Afghanistan, Pakistan and India. Diversification of *A. baumannii* GC1 into lineages and clades has probably been related to a dynamic expansion after passing a migration bottleneck to enter the hospital environment.

We conclude that CRISPR-based subtyping is a convenient method to trace the evolutionary history of particular bacterial clones, such as *A. baumannii* GC1.

1. Introduction

Acinetobacter baumannii is a glucose-non-fermenting Gram-negative bacterium that has widely been implicated in hospital-acquired infections (Nowak and Paluchowska, 2016). Although it has been found in a variety of environmental samples, the natural habitat of *A. baumannii* is still not known. This opportunistic pathogen has remarkable abilities to endure desiccation and starvation, acquire resistance to different classes of antibiotics, and disseminate in and between medical facilities (Chapartegui-González et al., 2018; Hamidian and Nigro, 2019). Worryingly, carbapenem-resistant *A. baumannii* has recently been labelled as "Priority 1: CRITICAL" in the world health organization list of antimicrobial resistant bacteria to guide research and development of new effective antibiotic treatments (Tacconelli et al., 2018). Extensive utilization of ventilators, due to the recent COVID-19 pandemic, has increased the risk of hospital-acquired ventilator-associated secondary infections, for which *A. baumannii* has been a main cause (Lescure et al.,

2020; Chen et al., 2020).

The exposed global population of *A. baumannii*, largely biased by clinical isolates, is predominated by few highly successful clones, including global clones (GC) 1 and 2, also known as international clones I and II (Hamidian and Nigro, 2019; Higgins et al., 2017). GC1 is currently the second most common *A. baumannii* clone with a widespread geographical distribution in more than 30 countries (Karah et al., 2012). Strains belonging to GC1 have showed a steady increase in the rates and ranges of their antimicrobial resistance over the past five decades (Holt et al., 2016). The oldest known GC1 isolate, HK302, was collected in 1977 in Switzerland (Krizova and Nemec, 2010). HK302 was multidrug-resistant, but susceptible to imipenem, and was associated with an outbreak of nosocomial infections (Devaud et al., 1982). Concurrently, time-stamped phylogenetic analysis of the whole-genomes of 44 GC1 strains estimated that the most recent common ancestor of GC1 emerged around 1960 and then diverged into two phylogenetically distinct lineages (Holt et al., 2016). The two GC1

lineages, L1 and L2, have then diversified into multiple clades accumulating different resistance determinants through the acquisition of plasmids and transposons and/or chromosomal mutations (Douraghi et al., 2020; Hamidian et al., 2019).

According to the Pasteur scheme for multilocus sequence typing (MLST), GC1 can be interrelated with clonal complex (CC) 1 (Diancourt et al., 2010). CC1 was initially composed of only five sequence types (STs), namely ST1, ST7, ST8, ST19 and ST20. The neat demarcation of these five STs from other non-CC1 STs endorsed the potential of this scheme to identify isolates belonging to CC1, and subsequently to GC1 (Diancourt et al., 2010). In addition, several programs to extract the MLST allelic profiles from assembled bacterial genomes are openly accessible online, providing a valuable tool for rapid assignments of STs to isolates subjected to whole genome sequencing (Larsen et al., 2012).

We have previously reported that CRISPR-*cas* subtype I-Fb, standing for clustered regularly interspaced short palindromic repeats (CRISPR) - CRISPR associated (*cas*) genes of the subtype I-Fb, is a conserved genetic element in the genome of GC1 (Karah et al., 2015). The occurrence of CRISPR-*cas* subtype I-Fb at the same locus on the chromosome of 106 GC1 isolates confirmed our results and indicated that this element was most likely acquired by a common ancestor of GC1 before diversification into intra-clonal lineages (Alvarez et al., 2020). The CRISPR-*cas* subtype I-Fb locus, located at position 1,057,691 to 1,069,768 of the genome of *A. baumannii* strain AYE (GenBank accession number: CU459141), included six genes, encoding for the Cas machinery, and an array of spacers (Karah et al., 2015). Upon the entry of a foreign element, the Cas machinery takes up a short sequence, called proto-spacer, from the invasive DNA and integrates it into the leading end of the array, where the adjacent direct repeat is duplicated and the integrated sequence becomes a new spacer flanked by two direct repeats (Barrangou and Marraffini, 2014). Interestingly, the spacer denoted Ab-1 was present at the trailer end of the CRISPR arrays in all the GC1 isolates, proposing it as a genomic indicator of GC1 (Karah et al., 2015). However, Ab-1 was also present in few non-GC1 isolates, such as NIPH 201 (APQV00000000.1) and Naval-82 (AMSW00000000.1) belonging to ST38 and ST428, respectively.

CRISPR-Cas systems were suggested to have an important role in the genomic changes of the *Acinetobacter* genus, particularly in controlling the transfer of conjugative elements (Touchon et al., 2014; Mangas et al., 2019). Due to their dynamic nature, comparative analysis of the CRISPR arrays of spacers provided a valuable secondary technique for subtyping isolates belonging to particular clones of *A. baumannii*, including GC1 (Karah et al., 2015). For instance, analysis of few spacers at the leading end of the CRISPR arrays was very informative to reveal diversity among local GC1 isolates (Hauck et al., 2012). Close clonal relationship was detected between twelve ST409 isolates collected in Greece between 2018 and 2019 and two ST1 isolates collected in Norway in 2011 and 2013 (Galani et al., 2020). Interestingly, one of these two Norwegian isolates had a history of import from Greece (Karah et al., 2015). In this study, the CRISPR-based subtyping approach was used to investigate the genetic relatedness among a collection of 187 clinical isolates belonging to *A. baumannii* GC1, and to provide new insights into the evolutionary path of this eminent clone.

2. Material and methods

2.1. In-silico plot to search for A. baumannii GC1 genomes

The Nucleotide Basic Local Alignment Search Tool (BLASTn) algorithm (https://blast.ncbi.nlm.nih.gov/Blast.cgi; Altschul et al., 1990) was used to screen for *A. baumannii* genomes carrying Ab-1 at the leading end of their CRISPR arrays of spacers. A query of 88 bp, including the nucleotide sequence of spacer Ab-1 and the two surrounding direct repeats, was used to search for similarities against the "RefSeq Genome Database (refseq_genomes)" database. "*Acinetobacter baumannii* (taxid:470)" was used as the target organism. The BLASTn

search was run under default parameters except for using 500, instead of 100, as the "max target sequences". Epidemiological data (year of isolation, country of isolation, and type of sample) were retrieved, from the online records or relevant literature, for all the hits identified by BLASTn. The nucleotide sequences of the corresponding whole genomes (complete or contigs) were downloaded on a local drive as FASTA files.

The online service "Sequence query - *Acinetobacter baumannii* MLST (Pasteur)", hosted at the *Acinetobacter baumannii* MLST website (https://pubmlst.org/abaumannii/), was used to determine the ST of the isolates according to the Institute Pasteur's MLST scheme (Diancourt et al., 2010). goeBURST and PHYLOViZ were used to generate and visualize a minimum spanning tree based on the allelic profiles of the whole *A. baumannii* MLST dataset (Francisco et al., 2009; Francisco et al., 2012). The n Locus Variant (nLV) Graph service was applied to display all possible links. We used one LV as the maximum number of differences between nodes (Ribeiro-Gonçalves et al., 2016). Accordingly, STs were grouped into CCs if they shared 6/7 of their MLST alleles with at least one other ST in the group (Feil et al., 2004).

2.2. Single-nucleotide polymorphism (SNP)-based phylogenetic analysis

CSI Phylogeny 1.4 (Call SNPs and Infer Phylogeny) was used to generate a genome-wide SNP-based phylogenetic tree for all the GC1 isolates (Kaas et al., 2014). The SNPs were called, filtered, site validated, concatenated, and aligned following the default parameters of the web-based service (https://cge.cbs.dtu.dk/services/CSIPhylogeny/). The genomic sequence of strain DSM30011, isolated before 1944, was used as a reference genome (GenBank accession no. JJOC02000000; Repizo et al., 2017). The reference genome was excluded from the phylogenetic tree. FigTree (http://tree.bio.ed.ac.uk/software/figtree/) was used to provide a high-quality graphical view of the generated phylogenetic tree.

2.3. CRISPR-based subtyping

The CRISPRCasfinder platform was used to detect and retrieve the nucleotide sequence of the CRISPR arrays of spacers (https://crisprcas.i2bc.paris-saclay.fr/CrisprCasFinder/Index; Couvin et al., 2018). Each spacer with a newly defined sequence was assigned a new consecutive number, and each array with a newly defined assortment of spacers was given a new CRISPR-based sequence type (CST), as previously described (Karah et al., 2015). Then, a binary file was manually created to visualize and compare the presence (red rectangle) or absence (empty) of spacers in each CST and to trace the ancestry of CRISPR arrays in GC1.

3. Results and discussion

3.1. Identification of A. baumannii GC1 genomes

Our BLASTn search (as of June 11, 2020) yielded a total of 260 hits with query coverage of 100% and nucleotide identity of 94.62% to 100% (Supplementary Table S1). Two hits represented a duplicate of the same CRISPR array on two different contigs (NZ_ASFN01000020.1 and NZ_ASFN01000054.1) in the genomic record of isolate TG22148 (Supplementary File S1). One of these two hits was omitted. Additional seven hits were excluded since they represented a repetitive genomic record of the same isolate (for instance, both JABU00000000.1 and JPHW00000000.1 corresponded to isolate MRSN 58). On the other hand, isolates recovered from the same patient or obtained during an outbreak caused by one strain were not excluded.

Among the 252 non-repetitive hits, 209 hits were carried by isolates belonging to CC1, as inferred by their MLST results (Table 1). The remaining 43 hits corresponded to non-CC1 isolates, among which 23 isolates were members of 4 different CCs, 13 isolates belonged to singleton STs, and 7 isolates showed incomplete or new STs (Supplementary File S2). CC1 was well detached from the non-CC1 query-

positive STs, as shown by the minimum spanning tree generated for the whole dataset of *A. baumannii* STs (Fig. 1A). The only exception could be ST106 (3, 3, 16, 1, 13, 1, 1) since it was a double locus variant of ST1217 (3, 3, 1, 1, 5, 1, 1) and had a subsequent linear linkage to ST736 (1, 3, 1, 1, 5, 1, 1) and ST1 (1, 1, 1, 1, 5, 1, 1). Nonetheless, a comprehensive genomic analysis is needed in order to determine the exact mode of acquisition, vertical or horizontal, of Ab-1 at the trailer end of the CRISPR arrays both in GC1 and among several non-GC1 isolates.

3.2. Updates on CC1

CC1 is currently part of a large complex consisting of several other CCs and hundreds of STs (Fig. 1B). Although we could not spot a strong founder for the large complex, clonal expansion of an overlooked ST is still the rational theory behind the formation of this complex (Feil et al., 2004). Alternatively, the occurrence of STs that might accidentally link different CCs is probable due to the large size of the MLST dataset. In order to make an outline for CC1, we had to make artificial borders and break the ties between ST94/ST325, ST94/ST495, and ST174/ST325 (Fig. 1B). Accordingly, ST325 and ST495 and their subsequent connections were subjectively considered to be outside CC1. Based on this outline, CC1 has so far included a total of 70 STs (Supplementary Table S2). As expected, ST1 has retained a robust central position, supporting the proposal that CC1 has emerged as a clonal expansion of ST1 (Diancourt et al., 2010). In other words, the most recent common ancestor of CC1, corresponding to GC1, most likely belonged to ST1 (Feil et al., 2004). Gradually, allelic changes have accumulated and descendants with new STs have emerged. Then, some of the descendent STs, such as ST20, ST81, ST19, ST493, and ST594, have become the founders of additional expansions and diversifications in CC1 (Fig. 1B).

3.3. Genome-wide SNP-based phylogenetic analysis

The phylogenetic tree of GC1 demonstrated the emergence of around ten clades and few subclades of isolates (Supplementary Fig. S1). Isolates belonging to lineage 1, according to previous studies (Alvarez et al., 2020; Hamidian et al., 2019; Holt et al., 2016), were overdistributed among several clades, making it difficult to delineate this lineage. Most of these isolates were supported by short internal branches. On the other hand, isolates from lineage 2, according to Douraghi et al., 2020 and Hamidian et al., 2019, were all assembled on a well delineated bush characterized by relatively long branches. Although the branch length in our tree was not proportional to time and only reflected to the amount of evolutionary divergence (the number of nucleotide substitutions) that has occurred along that branch, we could infer that lineage 1 was older than lineage 2, as reported by other studies (Holt et al., 2016). Our analysis did not exclude SNPs found in the accessory genomes nor those probably acquired via a homologous recombination event. Yet, our results were largely consistent with previous phylogenetic studies where strict parameters were used (Hamidian et al., 2019; Alvarez et al., 2020).

3.4. CRISPR-based subtyping of GC1

Full sequences were available for the CRISPR arrays of only 187 of the 209 GC1 isolates. Different compositions of the arrays enabled us to assign these 187 GC1 isolates into 45 CSTs, including 35 novel CSTs (Supplementary Tables S1, S3, and S4). The novel CSTs, CST41 to CST75, were designated according to the current numbering system for CSTs in *A. baumannii* (Karah et al., 2015). The arrays mainly differentiated by the acquisition of new spacers at the leading ends and/or due to internal duplications or deletions of vertically inherited spacers, as described in other bacterial species (Kupczok et al., 2015). The arrays ranged in size between 15 spacers (CST66) and 110 spacers (CST63). CST1 was the most common subtype, with 65 isolates recovered between 2002 and 2016 from 13 countries in North America, Europe, Asia, Africa and Australia (Supplementary Table S1). The year and/or country

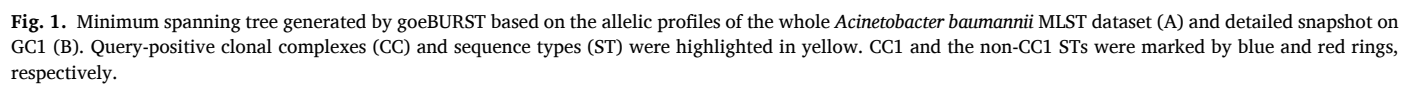of isolation were not available for 4 isolates.

Twenty-one isolates belonged to CST2, the second main subtype. CST2 was collected between 1984 and 2019 from different parts of the world, with the exception of Australia. Importantly, deletion of spacer Ab-40 was the only difference between CST2 and CST1. Each of the remaining CSTs were composed of 1 to 15 isolates. However, the size of some subtypes (for example, CST8 and CST13) was augmented by the presence of epidemiologically related isolates (Higgins et al., 2016; Lesho et al., 2013). We could not make a definite assignment for the remaining 22/209 GC1 isolates because the sequence of their arrays of spacers was divided on ≥2 contigs and some parts were missing or overlapping (Supplementary Table S1). Nonetheless, potential relationships were inferred from the available sequences. For example, isolates MRSN960 (NZ_VHDR01000016.1) and MRSN489678 (NZ_VHEN01000006.1) would perfectly fit into CST2 and CST65, respectively. Isolates ACB5 (NZ_OEON01000001.1) and aba_5m (NZ_CABEFJ010000030.1) belonged to a novel CST, which would be distinguished from CST61 by the deletion of 3 consecutive spacers (Supplementary File S3).

3.5. Post-migration diversification

The common ancestor of CRISPR arrays in GC1, designated $CST0_{GC1}$, consisted of 54 spacers (Ab-1 to Ab-54) as shown in Table S3. We propose that the most recent common ancestor of GC1 has entered the hospital ecosystem at this point. It is tempting to propose that this entry has happened on only one occasion and that GC1 has since then been persisting in the hospital environment. Earlier to this point, this ancestor was challenged by foreign DNA (plasmids or viruses) coexisting in the same unknown pre-hospital niche. Spacers Ab-55 to Ab-106 and Ab-877 to Ab-1024 were acquired after this point and could accordingly provide insights into the pool(s) of DNA inhabiting the hospital environment. $CST0_{GC1}$, and correspondingly the most recent common ancestor of GC1, has been evolving through two major pathways, designated 1 and 2 (Fig. 2). Deletion of Ab-15 and Ab-16 and deletion of Ab-18 are the genetic markers of pathways 1 and 2, respectively. *A. baumannii* isolate AB307–0294 (CP001172.2) has been standing alone with a secluded subtype (CST10) characterized by the deletion of Ab-23 to Ab-31. The exceptional position of AB307–0294 was also reported by other studies (Holt et al., 2016).

In order to survive in and between hospitals and succeed to infect and spread among patients, bacteria need to overcome a number of stressful conditions caused by the regular use of disinfectants, human immune mechanisms, and antibiotic treatments. Going through such harsh selective pressures, which can be described as a migration bottleneck, has shaped the clinical population of *A. baumannii*. The *A. baumannii* clinical isolates appear to have a constrained diversity in comparison with the environmental counterpart *e.g.* isolates from soil (Furlan et al., 2018). To our knowledge, only one isolate from GC1 has so far been reported from a hospital-unrelated environmental sample (Rafei et al., 2015). The whole genome of this isolate is under sequencing (according to personal communication with Dr. Rayane Rafei).

Once the bottleneck is passed, the newcomers gain a growth advantage, which could explain the predominance of few clones, such as GC1, in the current population of healthcare-associated *A. baumannii* (Karah et al., 2012). Further adaptability to the new environment has probably been a key factor behind the persistence, expansion, and widespread dissemination of GC1, as reported for other bacterial species (Martínez and Baquero, 2002). The post-migration GC1 lineages and clades are expected to be very homogenous in their core genomes (Antunes et al., 2014). Intra-clonal genomic deviations will be mainly related to genetic determinants providing better adoption to the hospital environment, including the accumulation of horizontally acquired antimicrobial resistance genes or having distinct outer surface molecules (Holt et al., 2016).

Fig. 1. Minimum spanning tree generated by goeBURST based on the allelic profiles of the whole *Acinetobacter baumannii* MLST dataset (A) and detailed snapshot on GC1 (B). Query-positive clonal complexes (CC) and sequence types (ST) were highlighted in yellow. CC1 and the non-CC1 STs were marked by blue and red rings, respectively.
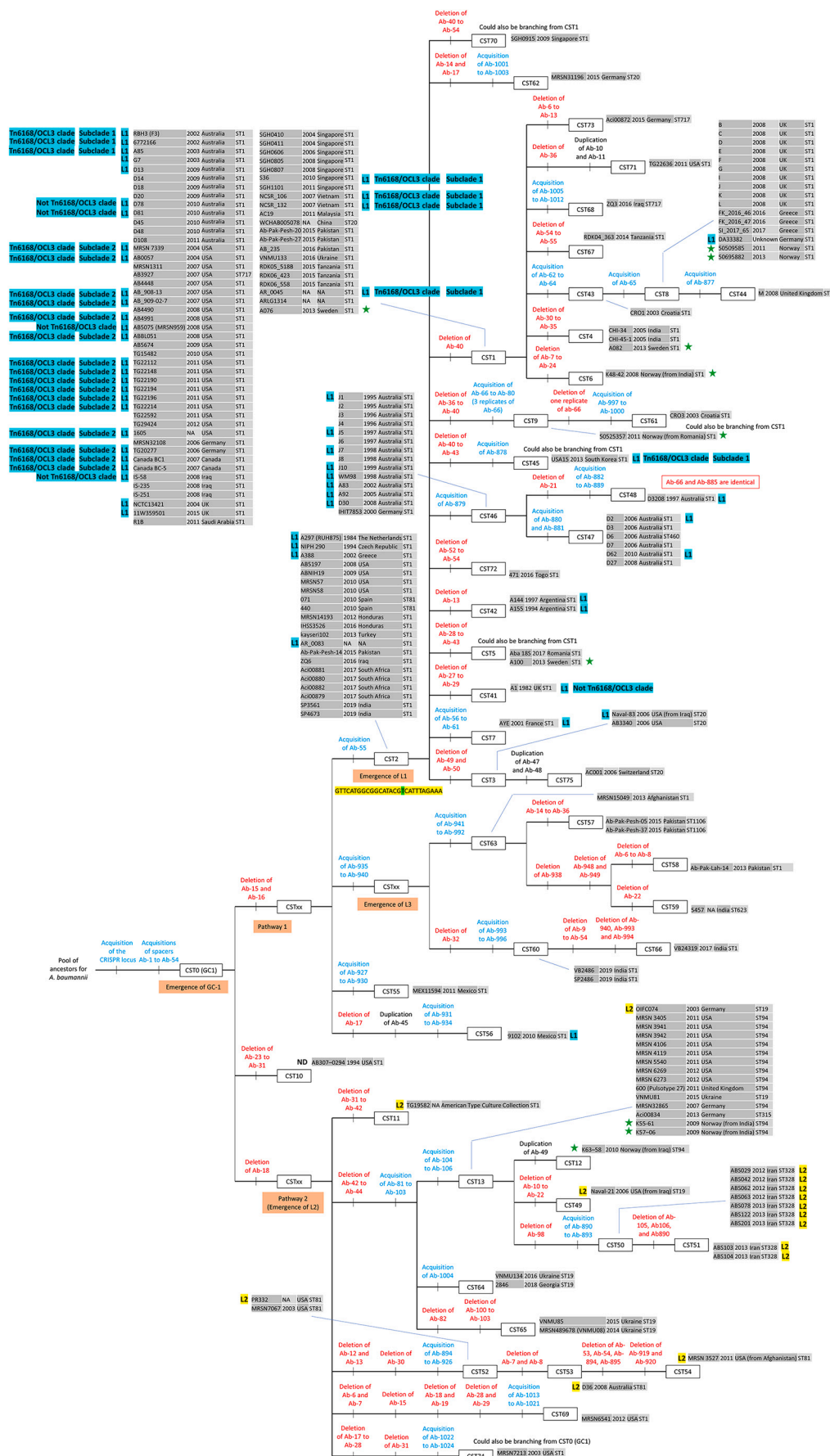
Fig. 2. Graphic tree of the evolutionary history of *Acinetobacter baumannii* GC1, inferred from the composition of their CRISPR arrays. Patterns of decent were established for 48 CRISPR-based sequence types (CSTs) including CST6, CST9 and CST12, retrieved from a previous study (Karah et al., 2015, the green stars referred to isolates from this study). L1, highlighted in blue, and L2, in yellow, were used to label isolates assigned to GC1 lineage 1 and lineage 2, respectively, as described by previous studies (Alvarez et al., 2020; Douraghi et al., 2020; Hamidian et al., 2019; Holt et al., 2016). ND, standing for not determined, was used to label isolate AB307–0294.

Table 1
Multilocus sequence types of 209 query-positive isolates belonging to clonal complex 1 and their geographical distribution.

Sequence type (ST)	Number of isolates	Geographical distribution
ST1 (1, 1, 1, 1, 5, 1, 1)	156	Vietnam, Australia, the United States of America (USA), Germany, Canada, Iraq, Pakistan, Ukraine, Singapore, Malaysia, Tanzania, Saudi Arabia, the United Kingdom (UK), India, South Africa, Honduras, Turkey, the Netherlands, Czech Republic, Greece, Romania, France, Argentina, Croatia, Bosnia and Hercegovina, South Korea, Mexico, Afghanistan, Togo
ST94 (1, 2, 2, 1, 5, 1, 1)	10	Germany, USA, UK
ST328 (1, 1, 1, 25, 5, 1, 2)	9	Iran
ST20 (3, 1, 1, 1, 5, 1, 1)	8	China, USA, Germany, Switzerland
ST81 (1, 1, 1, 1, 5, 1, 2)	8	Spain, USA, Australia
ST19 (1, 2, 1, 1, 5, 1, 1)	7	Ukraine, Germany, USA, Georgia
ST717 (1, 1, 1, 1, 5, 1, 30)	3	USA, Iraq, Germany
ST460 (1, 1, 1, 1, 5, 1, 64)	1	Australia
ST623 (1, 1, 2, 1, 5, 1, 1)	1	India
ST315 (1, 56, 1, 1, 5, 1, 1)	1	Germany
ST1106 (2, 1, 2, 1, 5, 1, 1)	1	Pakistan
Incomplete STs (1, 1, 1, 1, x, 1, 1), (1, 1, 1, x, 5, 1, 1), or (1, x, 1, 1, 5, 1, 1)	4	Togo, USA, Singapore

3.6. GC1 lineage 1

The main branch in pathway 1 corresponded to GC1 lineage 1 (Holt et al., 2016). The most recent common ancestor of this lineage belonged to CST2 (Fig. 2), descending from CST0_GC1 by the acquisition of spacer Ab-55 and the occurrence of a single nucleotide polymorphism (C to A substitution) in the direct repeat separating spacers Ab-22 and Ab-23 (Supplementary File S4). CST2 has so far been identified in the Netherlands 1984, Czech Republic 1994, Greece 2002, USA 2008–2010, Spain 2010, Honduras 2012 and 2016, Turkey 2013, Pakistan 2015, Iraq 2016, South Africa 2017, and India 2019, indicating that this antecedent subtype is still alive. Comparison of the year of isolation among the earliest isolates in GC1 lineage 1 was consistent with our proposal that CST2 was born first. In agreement, most of the CST2 isolates had very short internal branches and were mainly positioned at the base of the SNP-based phylogenetic tree (Supplementary Fig. S1).

Although CST2 has subsequently evolved toward a variety of CSTs, the emergence of CST1 was probably the most significant step in the dissemination of GC1 lineage 1. CST1 was also the founding node of several following subtypes (CST4, CST6, CST67, CST68, CST71, CST73, CST43, CST8, and CST44). We have previously reported that several isolates from CST1 were obtained or had a history of import from Iraq (Karah et al., 2015). Importantly, CST1, CST4 and CST67 have been involved in the dissemination of carbapenem-resistant isolates in India and Tanzania (Jones et al., 2014; Kumburu et al., 2019). The dissemination of carbapenem-resistant isolates belonging to CST8 has recently been reported in Greek hospitals (Galani et al., 2020).

CST46/CST47/CST48 was the second largest cluster branching from CST2. CST46/CST47/CST48 showed a local spread in Australia between 1995 and 2010, in addition to one sporadic isolate obtained in Germany in 2000. Since none of the latest Australian isolates belonged to this cluster, CST46/CST47/CST48 might have died out. The CST9/CST61 cluster consisted of two isolates, one obtained in Norway with a history of import from Romania (Karah et al., 2015) and one from Croatia

(Sousa et al., 2014). Together with isolate ACB5 from Bosnia and Hercegovina (Petrović et al., 2018), representing a novel CST related to CST61 as mentioned above, this cluster has shown a probable epidemiological linkage to Southeastern Europe. The vast majority of the GC1 lineage 1 isolates (CST2 and all CSTs branching from CST2) belonged to ST1. In this lineage, ST1 (1, 1, 1, 1, 5, 1, 1) has so far diversified to ST81 (1, 1, 1, 1, 5, 1, 2), ST20 (3, 1, 1, 1, 5, 1, 1), ST717 (1, 1, 1, 1, 5, 1, 30), and ST460 (1, 1, 1, 1, 5, 1, 64). Interestingly, the ST20 isolates in CST3/CST75 ($n = 3$) and CST62 ($n = 1$) have emerged directly from CST2 and were not related to the ST20 isolate of CST1 (Fig. 2 and Fig. S1).

3.7. GC1 lineage 3

The second main branch in pathway 1 was characterized by the acquisition of Ab-935 to Ab-940. Subtypes CST57, CST58, CST59, CST63, CST60, and CST66 have descended from this branch, hereby designated as GC1 lineage 3. GC1 lineage 3 included a total of eight isolates obtained between 2013 and 2019 from Afghanistan, Pakistan (Karah et al., 2020), and India. Isolates belonging to this lineage formed a demarcated cluster with long internal branches and an outstanding position in the genome-wide SNP-based phylogenetic tree (Fig. S1). Further analysis is needed to confirm the identity and characterize the phenotypic and genotypic features of this lineage. A number of phage DNA elements were detected as a potential source of the proto-spacers for Ab-935 to Ab-940 (Supplementary Table S5). However, we were not able to infer weather these spacers were acquired due to multiple independent interactions with several phage or plasmid DNA molecules or following a single contact with some yet unknown phage/plamsid.

3.8. GC1 Lineage 2

Pathway 2, corresponding to GC1 lineage 2 (Holt et al., 2016), was characterized by the lack of spacer Ab-18. Pathway 2 has so far split into five branches. The main branch included CST64, CST65 and the CST13 cluster (CST13 and all CSTs descending from CST13). This branch had 28 isolates belonging to ST19 (1, 2, 1, 1, 5, 1, 1), ST94 (1, 2, 2, 1, 5, 1, 1), ST315 (1, 56, 1, 1, 5, 1, 1), or ST328 (1, 1, 1, 25, 5, 1, 2). The second main branch included CST52, CST53, and CST54. Only 4 isolates were carried on this branch and were all assigned to ST81 (1, 1, 1, 1, 5, 1, 2). The remaining three branches corresponded to CST11, CST69, and CST74. They collectively had 3 isolates that belonged to ST1.

GC1 lineage 2 had a total of 35 isolates showing a widespread geographical dissemination between the United states (2003−2012), Germany (2003−2013), Australia (2008), Norway (2009–2010), the United Kingdom (2011), Iran 2012–2013, Ukraine 2014–2016, and Georgia 2018. Many of these isolates were obtained from military healthcare facilities and others had a history of import from Iraq or Afghanistan (Chan et al., 2015; Douraghi et al., 2020; Farlow et al., 2019; Huang et al., 2010; Kovalchuk et al., 2018; Lesho et al., 2013). In accord, three isolates obtained from one military hospital in France in 2009 were also linked to CST13 (Hauck et al., 2012; Karah et al., 2015). However, we could not make a precise assignment for the French isolates since only the leader end of the CRISPR arrays was available for comparison. Two of the CST13 and ST94 isolates were collected in Norway with a probable import from India (Karah et al., 2015). There was no epidemiological data to link the last two isolates with military hospitals or war zones (Karah et al., 2011).

The isolates from Iran belonged to CST50 ($n = 7$) and CST51 ($n = 2$) and were all assigned to ST328 (Douraghi et al., 2020). Both CST50 and CST51 descended from CST13 according to our CST-based analysis (Fig. 2). However, the SNP-based phylogenetic tree demonstrated that the Iranian isolates were more related to the isolates of CST52, CST53, and CST54 in comparison to the CST13 isolates (Fig. S1). The MLST results also demonstrated that ST328 was more related to ST81 (1 locus variant) than to ST19, ST94, or ST315 ($\geq$ 2 locus variants). In fact, the minimum spanning tree proposed that ST238 evolved from ST81

(Fig. 1B).

Searching for other isolates that might belong to GC1 lineage 2, we have found few articles demonstrating the occurrence of ST19 in Bulgaria in 2002 (from a military medical center), Saudi Arabia in 2006, and Croatia in 2009 (Aly et al., 2016; Diancourt et al., 2010; Dobrewski et al., 2006; Vranić-Ladavac et al., 2014). ST315 was reported in Belgium in 2010 (De Vos et al., 2016). The ST315 isolate had a history of import from Tunisia and was collected at a military hospital in Brussels. The occurrence of ST94 was also reported in the Kurdistan region of Iraq in 2012 (Ganjo et al., 2016). The *A. baumannii* MLST database included another isolate (Kh_4) that belonged to ST94 and was collected in Iraq in 2019 (https://pubmlst.org/bigsdb?page=infoanddb=pubmlst_ab aumannii_isolatesandid=4149).

4. Conclusion

CRISPR-based subtyping is suggested as a powerful and practical method to detect and track the patterns of descent among isolates of particular bacterial clones, such as *A. baumannii* GC1. Our study demonstrated that the most recent common ancestor of the current known population of GC1 carried an array of 54 spacers. Passing through a migration bottleneck, this ancestor has managed to establish itself as a permanent resident of the hospital environment. Since then, it has been following two main pathways of evolution, through which several lineages and clades have emerged. We found that the most recent common ancestor of GC1 lineage 1 belonged to CST2, an early but still active subtype. GC1 lineage 2 included a well demarcated group of isolates, which was mostly linked to military hospitals and war zones. We identified a novel lineage, designated GC1 lineage 3, which was distributed in the region of Pakistan, Afghanistan, and India. The occurrence of some spacers can be used to trace particular lineages, clades, or single CSTs. For instance, spacer Ab-55 is a hallmark of GC1 lineage 1 while the absence of spacer Ab-18 points toward GC1 lineage 2.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2021.104774.

Credit author statement

Nabil Karah: Conceptualization, Methodology, Writing - Original Draft, Visualization; Sun Nyunt Wai: Conceptualization, Writing - Review & Editing, Supervision; Bernt Eric Uhlin: Conceptualization, Writing - Reviewing and Editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare no competing interests.

Acknowledgments

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Alvarez, V.E., Quiroga, M.P., Galan, A.V., Vilacoba, E., Quiroga, C., Ramírez, M.S., Centron, D., 2020. Crucial role of the accessory genome in the evolutionary trajectory of *Acinetobacter baumannii* global clone 1. Front. Microbiol. 11, 342.

Aly, M.M., Abu Alsoud, N.M., Elrobh, M.S., Al Johani, S.M., Balkhy, H.H., 2016. High prevalence of the PER-1 gene among carbapenem-resistant *Acinetobacter baumannii* in Riyadh, Saudi Arabia. Eur. J. Clin. Microbiol. Infect. Dis. 35, 1759–1766.

Antunes, L.C., Visca, P., Towner, K.J., 2014. *Acinetobacter baumannii*: evolution of a global pathogen. Pathog. Dis. 71, 292–301.

Barrangou, R., Marraffini, L.A., 2014. CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. Mol. Cell 54, 234–244.

Chan, A.P., Sutton, G., DePew, J., Krishnakumar, R., Choi, Y., Huang, X.Z., Beck, E., Harkins, D.M., Kim, M., Lesho, E.P., Nikolich, M.P., Fouts, D.E., 2015. A novel method of consensus pan-chromosome assembly and large-scale comparative analysis reveal the highly flexible pan-genome of *Acinetobacter baumannii*. Genome Biol. 16, 143.

Chapartegui-González, I., Lázaro-Díez, M., Bravo, Z., Navas, J., Icardo, J.M., Ramos-Vivas, J., 2018. *Acinetobacter baumannii* maintains its virulence after long-time starvation. PLoS One 13, e0201961.

Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., Qiu, Y., Wang, J., Liu, Y., Wei, Y., Xia, J., Yu, T., Zhang, X., Zhang, L., 2020. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet 395, 507–513.

Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B., Rocha, E.P.C., Vergnaud, G., Gautheret, D., Pourcel, C., 2018. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Res. 46, W246–W251.

De Vos, D., Pirnay, J.P., Bilocq, F., Jennes, S., Verbeken, G., Rose, T., Keersebilck, E., Bosmans, P., Pieters, T., Hing, M., Heuninckx, W., De Pauw, F., Soentjens, P., Merabishvili, M., Deschaght, P., Vaneechoutte, M., Bogaerts, P., Glupczynski, Y., Pot, B., van der Reijden, T.J., Dijkshoorn, L., 2016. Molecular epidemiology and clinical impact of *Acinetobacter calcoaceticus-baumannii* complex in a Belgian burn wound center. PLoS One 11, e0156237.

Devaud, M., Kayser, F.H., Bachi, B., 1982. Transposon-mediated multiple antibiotic resistance in Acinetobacter strains. Antimicrob. Agents Chemother. 22, 323–329.

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

Dobrewski, R., Savov, E., Bernards, A.T., van den Barselaar, M., Nordmann, P., van den Broek, P.J., Dijkshoorn, L., 2006. Genotypic diversity and antibiotic susceptibility of *Acinetobacter baumannii* isolates in a Bulgarian hospital. Clin. Microbiol. Infect. 12, 1135–1137.

Douraghi, M., Kenyon, J.J., Aris, P., Asadian, M., Ghourchian, S., Hamidian, M., 2020. Accumulation of antibiotic resistance genes in carbapenem-resistant *Acinetobacter baumannii* isolates belonging to lineage 2, global clone 1, from outbreaks in 2012-2013 at a Tehran burns hospital. mSphere 5 e00164-20.

Farlow, J., Nozadze, M., Mitaishvili, N., Kotorashvili, A., Kotoria, N., Arobelidze, K., Tavadze, V., Simsive, T., Imnadze, P., Latif, N., Nikolich, M.P., Washington, M., Hinkle, M.K., Kwon, P., Trapaidze, N., 2019. Comparative genomic analysis of four multidrug-resistant isolates of *Acinetobacter baumannii* from Georgia. J. Glob. Antimicrob. Resist. 21, 363–368.

Feil, E.J., Li, B.C., Aanensen, D.M., Hanage, W.P., Spratt, B.G., 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J. Bacteriol. 186, 1518–1530.

Francisco, A.P., Bugalho, M., Ramirez, M., Carriço, J.A., 2009. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. BMC Bioinformatics 10, 152.

Francisco, A.P., Vaz, C., Monteiro, P.T., Melo-Cristino, J., Ramirez, M., Carriço, J.A., 2012. PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. BMC Bioinformatics 13, 87.

Furlan, J.P.R., Pitondo-Silva, A., Stehling, E.G., 2018. New STs in multidrug-resistant *Acinetobacter baumannii* harbouring β-lactamases encoding genes isolated from Brazilian soils. J. Appl. Microbiol. 125, 506–512.

Galani, V., Papagiannitsis, C.C., Petinaki, E., 2020. First description of ST409 OXA-23-producing *Acinetobacter baumannii*, carrying a CST8 CRISPR/Cas system, in Central Greece. J. Glob. Antimicrob. Resist. 22, 137–138.

Ganjo, A.R., Maghdid, D.M., Mansoor, I.Y., Kok, D.J., Severin, J.A., Verbrugh, H.A., Kreft, D., Fatah, M.H., Alnakshabandi, A.A., Dlnya, A., Hammerum, A.M., Ng, K., Goessens, W., 2016. OXA-carbapenemases present in clinical *Acinetobacter baumannii-calcoaceticus* complex isolates from patients in Kurdistan region. Iraq. Microb. Drug Resist. 22, 627–637.

Hamidian, M., Nigro, S.J., 2019. Emergence, molecular mechanisms and global spread of carbapenem-resistant *Acinetobacter baumannii*. Microb. Genom. 5, e000306.

Hamidian, M., Hawkey, J., Wick, R., Holt, K.E., Hall, R.M., 2019. Evolution of a clade of *Acinetobacter baumannii* global clone 1, lineage 1 via acquisition of carbapenem- and aminoglycoside-resistance genes and dispersion of ISAba1. Microb. Genom. 5, e000242.

Hauck, Y., Soler, C., Jault, P., Merens, A., Gerome, P., Nab, C.M., Trueba, F., Bargues, L., Thien, H.V., Vergnaud, G., Pourcel, C., 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.

Higgins, P.G., Chan, J.Z., Seifert, H., Pallen, M.J., Millard, A.D., 2016. Draft genome sequences of 11 clinical isolates of *Acinetobacter baumannii*. Genome Announc. 4 e00269–16.

Higgins, P.G., Prior, K., Harmsen, D., Seifert, H., 2017. Development and evaluation of a core genome multilocus typing scheme for whole-genome sequence-based typing of *Acinetobacter baumannii*. PLoS One 12, e0179228.

Holt, K., Kenyon, J.J., Hamidian, M., Schultz, M.B., Pickard, D.J., Dougan, G., Hall, R., 2016. Five decades of genome evolution in the globally distributed, extensively antibiotic-resistant *Acinetobacter baumannii* global clone 1. Microb. Genom. 2, e000052.

Huang, X.Z., Frye, J.G., Chahine, M.A., Cash, D.M., Barber, M.G., Babel, B.S., Kasper, M. R., Whitman, T.J., Lindler, L.E., Bowden, R.A., Nikolich, M.P., 2010. Genotypic and phenotypic correlations of multidrug-resistant *Acinetobacter baumannii*-

A. calcoaceticus complex strains isolated from patients at the National Naval Medical Center. J. Clin. Microbiol. 48, 4333–4336.

Jones, L.S., Toleman, M.A., Weeks, J.L., Howe, R.A., Walsh, T.R., Kumarasamy, K.K., 2014. Plasmid carriage of *bla*$_{NDM-1}$ in clinical *Acinetobacter baumannii* isolates from India. Antimicrob. Agents Chemother. 58, 4211–4213.

Kaas, R.S., Leekitcharoenphon, P., Aarestrup, F.M., Lund, O., 2014. Solving the problem of comparing whole bacterial genomes across different sequencing platforms. PLoS One 9, e104984.

Karah, N., Haldorsen, B., Hermansen, N.O., Tveten, Y., Ragnhildstveit, E., Skutlaberg, D. H., Tofteland, S., Sundsfjord, A., Samuelsen, Ø., 2011. Emergence of OXA-carbapenemase- and 16S rRNA methylase-producing international clones of *Acinetobacter baumannii* in Norway. J. Med. Microbiol. 60, 515–521.

Karah, N., Sundsfjord, A., Towner, K., Samuelsen, Ø., 2012. Insights into the global molecular epidemiology of carbapenem non-susceptible clones of *Acinetobacter baumannii*. Drug Resist. Updat. 15, 237–247.

Karah, N., Khalid, F., Wai, S.N., Uhlin, B.E., Ahmad, I., 2020. Molecular epidemiology and antimicrobial resistance features of *Acinetobacter baumannii* clinical isolates from Pakistan. Ann. Clin. Microbiol. Antimicrob. 19 (2).

Karah, N., Samuelsen, Ø., Zarrilli, R., Sahl, J.W., Wai, S.N., Uhlin, B.E., 2015. CRISPR-*cas* subtype I-Fb in *Acinetobacter baumannii*: evolution and utilization for strain subtyping. PLoS One 10, e0118205.

Kovalchuk, V., Kondratiuk, V., McGann, P.T., Kovalenko, I., Snesrud, E., 2018. The epidemiology and antibiotic resistance in *A. baumannii* isolates from military health care facilities. Reports of Vinnytsia National Medical University 22, 248–252. https://doi.org/10.31393/reports-vnmedical-2018-22(2)-02.

Krizova, L., Nemec, A., 2010. A 63 kb genomic resistance island found in a multidrug-resistant *Acinetobacter baumannii* isolate of European clone I from 1977. J. Antimicrob. Chemother. 65, 1915–1918.

Kumburu, H.H., Sonda, T., van Zwetselaar, M., Leekitcharoenphon, P., Lukjancenko, O., Mmbaga, B.T., Alifrangis, M., Lund, O., Aarestrup, F.M., Kibiki, G.S., 2019. Using WGS to identify antibiotic resistance genes and predict antimicrobial resistance phenotypes in MDR *Acinetobacter baumannii* in Tanzania. J. Antimicrob. Chemother. 74, 1484–1493.

Kupczok, A., Landan, G., Dagan, T., 2015. The contribution of genetic recombination to CRISPR array evolution. Genome Biol. Evol. 7, 1925–1939.

Larsen, M.V., Cosentino, S., Rasmussen, S., Friis, C., Hasman, H., Marvig, R.L., Jelsbak, L., Sicheritz-Pontén, T., Ussery, D.W., Aarestrup, F.M., Lund, O., 2012. Multilocus sequence typing of total-genome-sequenced bacteria. J. Clin. Microbiol. 50, 1355–1361.

Lescure, F.X., Bouadma, L., Nguyen, D., Parisey, M., Wicky, P.H., Behillil, S., Gaymard, A., Bouscambert-Duchamp, M., Donati, F., Le Hingrat, Q., Enouf, V., Houhou-Fidouh, N., Valette, M., Mailles, A., Lucet, J.C., Mentre, F., Duval, X., Descamps, D., Malvy, D., Timsit, J.F., Lina, B., van der Werf, S., Yazdanpanah, Y., 2020. Clinical and virological data of the first cases of COVID-19 in Europe: a case series. Lancet Infect. Dis. 20, 697–706.

Lesho, E., Yoon, E.J., McGann, P., Snesrud, E., Kwak, Y., Milillo, M., Onmus-Leone, F., Preston, L., St Clair, K., Nikolich, M., Viscount, H., Wortmann, G., Zapor, M., Grillot-Courvalin, C., Courvalin, P., Clifford, R., Waterman, P.E., 2013. Emergence of

colistin-resistance in extremely drug-resistant *Acinetobacter baumannii* containing a novel *pmrCAB* operon during colistin therapy of wound infections. J. Infect. Dis. 208, 1142–1151.

Mangas, E.L., Rubio, A., Álvarez-Marín, R., Labrador-Herrera, G., Pachón, J., Pachón-Ibáñez, M.E., Divina, F., Pérez-Pulido, A.J., 2019. Pangenome of *Acinetobacter baumannii* uncovers two groups of genomes, one of them with genes involved in CRISPR/Cas defence systems associated with the absence of plasmids and exclusive genes for biofilm formation. Microb Genom. 5 (11), e000309.

Martínez, J.L., Baquero, F., 2002. Interactions among strategies associated with bacterial infection: pathogenicity, epidemicity, and antibiotic resistance. Clin. Microbiol. Rev. 15, 647–679.

Nowak, P., Paluchowska, P., 2016. *Acinetobacter baumannii*: biology and drug resistance - role of carbapenemases. Folia Histochem. Cytobiol. 54, 61–74.

Petrović, T., Uzunović, S., Barišić, I., Luxner, J., Grisold, A., Zarfel, G., Ibrahimagić, A., Jakovac, S., Slaćanac, D., Bedenić, B., 2018. Arrival of carbapenem-hydrolyzing-oxacillinases in *Acinetobacter baumannii* in Bosnia and Herzegovina. Infect. Genet. Evol. 58, 192–198.

Rafei, R., Hamze, M., Pailhoriès, H., Eveillard, M., Marsollier, L., Joly-Guillou, M.L., Dabboussi, F., Kempf, M., 2015. Extrahuman epidemiology of *Acinetobacter baumannii* in Lebanon. Appl. Environ. Microbiol. 81, 2359–2367.

Repizo, G.D., Viale, A.M., Borges, V., Cameranesi, M.M., Taib, N., Espariz, M., Brochier-Armanet, C., Gomes, J.P., Salcedo, S.P., 2017. The environmental *Acinetobacter baumannii* isolate DSM30011 reveals clues into the preantibiotic era genome diversity, virulence potential, and niche range of a predominant nosocomial pathogen. Genome Biol. Evol. 9, 2292–2307.

Ribeiro-Gonçalves, B., Francisco, A.P., Vaz, C., Ramirez, M., Carriço, J.A., 2016. PHYLOViZ online: web-based tool for visualization, phylogenetic inference, analysis and sharing of minimum spanning trees. Nucleic Acids Res. 44, W246–W251.

Sousa, C., Silva, L., Grosso, F., Lopes, J., Peixe, L., 2014. Development of a FTIR-ATR based model for typing clinically relevant *Acinetobacter baumannii* clones belonging to ST98, ST103, ST208 and ST218. J. Photochem. Photobiol. B 133, 108–114.

Tacconelli, E., Carrara, E., Savoldi, A., Harbarth, S., Mendelson, M., Monnet, D.L., Pulcini, C., Kahlmeter, G., Kluytmans, J., Carmeli, Y., Ouellette, M., Outterson, K., Patel, J., Cavaleri, M., Cox, E.M., Houchens, C.R., Grayson, M.L., Hansen, P., Singh, N., Theuretzbacher, U., Magrini, N., WHO Pathogens Priority List Working Group, 2018. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. Lancet Infect. Dis. 18, 318–327.

Touchon, M., Cury, J., Yoon, E.J., Krizova, L., Cerqueira, G.C., Murphy, C., Feldgarden, M., Wortman, J., Clermont, D., Lambert, T., Grillot-Courvalin, C., Nemec, A., Courvalin, P., Rocha, E.P., 2014. The genomic diversification of the whole Acinetobacter genus: origins, mechanisms, and consequences. Genome Biol. Evol. 6, 2866–2882.

Vranić-Ladavac, M., Bedenić, B., Minandri, F., Ištok, M., Bošnjak, Z., Frančula-Zaninović, S., Ladavac, R., Visca, P., 2014. Carbapenem resistance and acquired class D beta-lactamases in *Acinetobacter baumannii* from Croatia 2009-2010. Eur. J. Clin. Microbiol. Infect. Dis. 33, 471–478.