Genomic Analysis of Molecular Bacterial Mechanisms of Resistance to Phage Infection

Antón Ambroa

University of A Coruña (UDC), A Coruña

Lucía Blasco

University of A Coruña (UDC), A Coruña

María López

University of A Coruña (UDC), A Coruña

Olga Pacios

University of A Coruña (UDC), A Coruña

Inés Bleriot

University of A Coruña (UDC), A Coruña

Laura Fernández-García

University of A Coruña (UDC), A Coruña

Concha Ortiz-Cartagena

University of A Coruña (UDC), A Coruña

Andrew Millard

University of Leicester

María Tomás (✉ ma.del.mar.tomas.carmona@sergas.es)

University of A Coruña (UDC), A Coruña

Research Article

Keywords: bacterial, resistance, genomic island, phages, *Acinetobacter baumannii*

Posted Date: August 5th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-699717/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background

In order to optimize phage therapy, we need to understand how bacteria evolve against phage attack. One of the main problems of the phage therapy is the appearance of bacterial resistance variants. The use of genomics to track antimicrobial resistance is increasingly developed and used in clinical laboratories. For that reason, it is important to consider, in an emerging future with phage therapy, to detect and avoid phage resistant strains, that can be overcome by the analysis of metadata provided by WGS. Here, we identified genes associated with phage resistance in 18 *Acinetobacter baumannii* clinical strain belonging to the ST-2 clonal complex during a decade (Ab2000 vs 2010): 9 from 2000 and 9 from 2010.

Results

The presence of genes putatively associated to phage resistance were detected. Genes detected were associated with an abortive infection system, restriction-modification system, genes predicted to be associated with defence systems but with unknown function and CRISPR-Cas system. Between 118 and 171 genes were found in the 18 clinical strains. On average, 26% of these genes were detected inside genomic islands (GIs) in the 2000 strains and 32% in 2010 strains. Furthermore, 38 potential CRISPR arrays in 17 of 18 of the strains were found, as well as 705 proteins associated with CRISPR-Cas systems.

Conclusions

A moderately higher presence of these genes in the strains of the 2010 in comparison to those of the 2000 were found, especially those related to the R-M system and CRISPR-Cas system. The presence of these genes in GIs in a higher rate in the strains of the 2010 compared to those of the 2000 was also detected. WGS and bioinformatics could be powerful tools to avoid drawbacks when a personalized therapy is applied. In this study, it allows us to take care of the phage resistance in *A. baumannii* clinical strains to prevent a failure in a possible phage therapy.

Background

As part of the ESKAPE pathogens, *A. baumannii* is frequently isolated from infections in clinical environments, and its resistance against multiple antibiotics is increasingly common (1). For this reason, it is necessary to opt for alternative treatments, such as phage therapy. However, the ability of bacteria to develop resistance mechanisms against phages is possible, even when there is no previous treatment with phage therapy due to the constant coevolutionary interactions (2). The spread of phage resistance presents a significant challenge to the efficacy of the therapy (3), (4).

It is important to know and characterize the phage resistance mechanisms of a certain species, clone or strain, prior to phage treatment in order to minimise treatment failure. Whole-genome sequencing (WGS) has been demonstrated to be a powerful tool in the detection of phage-resistance mechanisms, as well as the evolution of CRISPR-Cas arrays in bacteria subjected to phage pressure (5, 6). WGS is increasingly becoming

a cheaper and faster technology, thus it is implemented progressively in routine hospital diagnostics and research (7).

Recently, new or modified phage resistance mechanisms have been discovered and characterized (8). Although a large part of defence systems against phages are maintained over generations, there is a continuous emergence of resistance mechanisms due to spontaneous mutations as a consequence of the coexistence of phage and bacteria. Most of these mutations occur in the phage receptors proteins, employed by the phages to adhere to the cell (8). In recent years, phage resistance mechanisms are attracting increasing interest due to the rising knowledge in phage interactions with bacteria. This is leading to the discovery and characterization of new phage resistance mechanisms such as Zorya, Druantia or Thoeis (9). Phage resistance mechanisms are typically clustered in genomic “defence islands”: mobile genetic evolutionary elements that contain genes associated with phage defence systems (9, 10).

The main resistance mechanisms are related to the inhibition of the phage adsorption, blocking of phage DNA injection, cutting of the injected DNA, inhibition of the phage DNA replication, interference in the phage assembly, and bacterial suicide (11). In Fig. 1 we summarized all the characterized phage resistance mechanisms.

In this study, we focused in those which could be bioinformatically detected without any experimental process:

- i) Abortive Infection systems, characterized by the fact that the phage enters the cell, but its development is interrupted in any phase (replication, transcription or translation). The mechanism of action is not entirely clear, either because of their complexity or because they are widely varied from one species to another (12).
- ii) Toxin/Antitoxin systems are a specific type of ABI system but they are well-characterized and widespread through diverse species (13). In this system, a toxin is produced by the cell and is neutralized by an antitoxin. The expression of these molecules is highly controlled and varies from one system to another. When the balance between one molecule and the other is disturbed, the toxin is released and the bacteria die (14).
- iii) Restriction-Modification systems consist of a restriction endonuclease and a This type of system distinguishes the DNA of the host from foreign DNA to recognize and destroy phage DNA after its injection into the When unmethylated phage DNA enters a bacteria which possesses the R-M system, it will be cleaved by the restriction endonuclease or methylated by the methyltransferase to escape the restriction (15).
- iv) CRISPR-Cas (clustered regularly interspaced short palindromic repeats – CRISPR-associated system is an adaptative immune system that bacteria develop against phage DNA/RNA and other foreign DNA (16). The typical structure of the CRISPR-Cas locus is a leader sequence, followed by the repeat-spacer array and the *cas* genes operon (17). The adaptation of the CRISPR-Cas system is due to the acquisition of the spacer sequences, which are small fragments of foreign nucleic acids, between the repeats of the CRISPR locus (17). The functioning of the CRISPR-Cas system, usually divided into three steps (adaptation, processing and guidance of the crRNA-CRISPR RNA- and targeting and interference of the foreign DNA/RNA) is carried out by the Cas (CRISPR-associated) proteins (18). CRISPR-Cas systems are classified according their conserved *cas* genes and the architecture of the *cas* operon (19). Until recently, little data existed about CRISPR-Cas systems in *A. baumannii*. The pangenome analysis of *A. baumannii* has shown CRISPR-Cas

systems in the species (20). One of the most characterized systems in *A. baumannii* is the CRISPR I-Fb system (21). However, most of the Cas-related genes and CRISPR arrays are yet not identified and characterized.

In this study, we searched for putative genes associated with phage resistance and we focused on CRISPR-Cas systems by studying the CRISPR arrays and Cas protein presence through bioinformatic approach in 18 genomes of clinical strains of *A. baumannii* isolated in the “II Spanish Study of *A. baumannii* GEIH-REIPI 2000–2010”.

Methodology

Genome database

18 clinical *A. baumannii* genomes previously sequenced and annotated (II Spanish Multicenter Study, GEIH-REIPI *Acinetobacter baumannii* 2000–2010; Umbrella Bioproject PRJNA422585) (22), (23) have been studied. Nine strains were from the 2000 and nine from 2010. All of the strains belong to the ST-2 clone (22).

Search for general genes associated to bacteriophage resistance and their presence in genomic islands

In order to analyse the presence of genes putatively associated with phage resistance systems, a custom database based on genes from the “PADS Arsenal database” (<https://bigd.big.ac.cn/padsarsenal/>) was created (24). The genes were grouped in five systems: ABI systems related (not belonging to toxin/antitoxin system), TA systems related, R-M system related, CRISPR-Cas associated proteins and newly (NEW) characterized systems related genes. In this last category we included those genes which hit against known phage resistance genes but were associated with genes predicted to be associated with phage resistance functions and whose function in *A. baumannii* is not clear yet, such as newly characterized systems (e.g. Zorya, Druantia, Thoeris). A blast search of the complete genomes against this database and filtered out those hits which e-value > 1E-04 was made. The percentage of the genes involved in resistance was calculated by dividing the genes predicted to be associated with phage resistance by the total number of genes in the bacteria genome.

To locate Genomic Islands (GIs) three different approaches were used: IslandViewer with default settings, (25), blast search with default settings and cut-off of e-value < 1e-03 against a previously constructed ICEberg database (26) and checking the Guanine-Cytosine (GC) content of the contigs of each genome (27). The previously detected phage-resistance genes were localised in the GIs detected per genome and average percentage of genes by collection was calculated.

Search and characterization of CRISPR arrays

In a first try, CRISPRCasFinder (28) was employed but no putative CRISPR-Cas system was found. For this reason, CRISPR arrays were found using the CRISPR Recognition Tool (CRT) (29). The modification proposed by Rho et al (29) for whole-metagenomic assembled genomes called metaCRT was used (30) with

the following parameters: minimum number of repeats: 3, minimum repeat length: 12, maximum repeat length: 70, minimum spacer length: 18, maximum spacer length: 80 and with a search window of: 6.

In order to filter and validate the CRISPR arrays, a similar procedure to the first step in the protocol developed by Shmakov et al was followed (31) (Fig. 2). In the first step, CRISPR arrays separated by no more than 6 Open Reading Frames (ORF) to a putative Cas protein identified before, were considered to be part of a putative CRISPR-Cas system. Those which not were part of a putative CRISPR-Cas system were considered to be single possible CRISPR arrays. To validate these arrays, we made a short-blastn search of the spacers of the possible CRISPR arrays against all phage genomes using the INPHARED database (32). Results that were > 95% in identity and those arrays whose query hit was larger than 20 were considered putative CRISPR arrays.

Negative blast hits of the spacers of the CRISPR arrays against bacteriophage follow the procedure described by Shmakov for isolated arrays: first, arrays > 400 bp and with an ORF coverage > 0.95 were filtered out. Second, all arrays with < 850 bp that had domain in the Conserved Domain Database (CDD) search were filtered out (33), (34). Then, pairwise distances between spacers of each array were calculated (number of matches in the longest blastn hit between them, divided by the length of the smaller spacer in the pair). The spacers of each array were clustered using single linkage clustering following the same procedure as Shmakov et al. with a cut-off of 0.3. A spacer similarity index was calculated for each CRISPR array as the number of clusters formed divided by the number of spacers in the array (1 means that all the spacers are different). Those arrays whose spacer similarity index was < 0.85 were filtered out. The rest were considered putative CRISPR arrays.

To complement the procedure made by Shmakov, a search about common special low-complexity sequences that may be confused as CRISPR arrays was made, known as false-CRISPR elements (35). The presence of Tandem repeats, potentially hypermutable regions which enable bacteria to adapt to evolving environments without increasing their mutation rate, was checked with Tandem Repeat Finder (36–38). The presence of short low-complexity repeats was also examined with RepeatMasker (39). To test the results and to complete the search of low-complexity sequences a blast search against an existing false-CRISPR elements database was obtained from the CRISPRone website (35).

The search for possible Cas-related proteins was made, based on the method of Zhang et al. (40), but adding a search in “HMMCAS” website of all of their available HMM models, performed with a cut-off of reported e-values of (41). The 18 genomes were examined using hmmscan with all of the pfam HMM profiles based on NCBI entries of known Cas protein families searching in the pfam database 70 Cas-related protein families and other CRISPR-associated proteins in the pfam database (e. g. DEAD/DEAH box helicase), 93 families which were in the TIGRFAMS resource and the 24 newly characterized families (40, 42–44). For the TIGRFAMS database proteins it was necessary to build an HMM profile with hmmbuild with default settings (42) making a previous alignment for each compound of proteins with Clustal Omega version 1.2.4 (ClustalO) (45).

All of the input and output data for the searches (genes associated to phage resistance, genomic islands, CRISPR arrays and Cas proteins) were processed with Python (www.python.org) and BioPython (19304878

(46), (47).

In order to establish the evolution and to compare the presence of the CRISPR arrays among the same clonal complex of the 18 clinical strains of *A. baumannii*, a phylogenetic tree was made using the CRISPR spacers detected. Trees were built using MEGA7 with CLUSTALW alignment (MEGA version 7.9.26) (48–50).

Results

Genes putatively associated with phage resistance in *A. baumannii* clinical strains and their presence in GIs

Between 118 and 171 genes were detected per genome, those could be putatively associated with bacterial defence against bacteriophages (Table 1, Additional file 1). The frequency (%) of each resistance system was calculated with the number of genes of each group divided by the total genes per genome. It was observed that the genes related to R-M and CRISPR-Cas systems showed a slightly higher prevalence in 2010 strains (Fig. 3, A). The frequency of the genes related to ABI, TA and new systems remained constant in both collections. The presence of putative phage resistance genes in GIs was also predicted (Table 2, Additional file 1), and it was found that in GIs represents in strains of the year 2010 approximately a 24% in average and approximately a 19% in the strains of the year 2000 in average (Fig. 3, B). The observed increase was produced specially in genes related to the RM system, to those related with new phage resistance mechanisms and CRISPR-Cas system.

CRISPR arrays

180 putative predicted arrays were found (without filtering) using metaCRT in the 18 genomes of ST-2 *A. baumannii* clinical strains (Table 3, Additional file 2) Post the complete process of filtering designed by Shmakov et al. (31) and removing the low-complexity sequences (35), only 40 CRISPR arrays were selected (Table 1): 18 CRISPR arrays were present in the 2000 strains and 22 in the 2010 strains. All the strains, excepting in the Ab161_GEIH-2000 strain, presented at least 1 CRISPR array.

Table 1
CRISPR arrays present in the genomes of 18 *A. baumannii* clinical strains.

Strain	Contig	Size	Start	Stop	Repeat	N° of spacers
Ab33_GEIH-2010	MSMK01000003	160	10462	10622	ATTTTGAATTTAAAA	4
Ab33_GEIH-2010	MSMK01000187	198	18280	18478	ACAAAAGAAAAAT	4
Ab49_GEIH-2010	MSMM01000317	96	1114	1210	TCATTTTGCTGTTGTT	2
Ab49_GEIH-2010	MSMM01000323	198	78	276	ACAAAAGAAAAAT	4
Ab49_GEIH-2010	MSMM01000347	122	367	489	TTTTAAATTCAAAA	3
Ab54_GEIH-2010	MSML01000240	198	3914	4112	AATTTTCTTTTCT	4
Ab54_GEIH-2010	MSML01000469	96	1108	1204	TCATTTTGCTGTTGTT	2
Ab54_GEIH-2010	MSML01000525	164	8017	8181	ATATATTTTGA	3
Ab76_GEIH-2010	MSLY01000008	96	835	931	TCATTTTGCTGTTGTT	2
Ab76_GEIH-2010	MSLY01000677	198	3369	3567	AATTTTCTTTTCT	4
Ab76_GEIH-2010	MSLY01000708	164	714	878	ATATATTTTGA	3
Ab103_GEIH-2010	MSLX01000655	160	9148	9308	ATTTTGAATTTAAAA	4
Ab103_GEIH-2010	MSLX01000266	164	2680	2844	ATATATTTTGA	3
Ab103_GEIH-2010	MSLX01000388	96	1108	1204	TCATTTTGCTGTTGTT	2
Ab103_GEIH-2010	MSLX01000506	198	55	253	ACAAAAGAAAAAT	4
Ab104_GEIH-2010	MSMA01000019	96	1450	1546	TCATTTTGCTGTTGTT	2
Ab104_GEIH-2010	MSMA01000107	160	4402	4562	TTTTAAATTCAAAAT	4
Ab104_GEIH-2010	MSMA01000246	164	10815	10979	ATATATTTTGA	3

Strain	Contig	Size	Start	Stop	Repeat	N° of spacers
Ab105_GEIH-2010	LJHB01000001	198	125508	125706	ACAAAAGAAAAAT	4
Ab105_GEIH-2010	LJHB01000010	292	7321	7613	TAAAATAATTTTAA	5
Ab121_GEIH-2010	MSLZ01000141	198	4992	5190	AATTTTCTTTTCT	4
Ab122_GEIH-2010	MSMD01000782	164	711	875	ATATATTTTGA	3
Ab155_GEIH-2000	LJHA01000001	198	125512	125710	ACAAAAGAAAAAT	4
Ab155_GEIH-2000	LJHA01000002	292	7323	7615	TAAAATAATTTTAA	5
Ab158_GEIH-2000	MSMC01000196	198	4027	4225	AATTTTCTTTTCT	4
Ab158_GEIH-2000	MSMC01000525	136	868	1004	ATTTTTTAATATTTA	3
Ab166_GEIH-2000	MSMG01000383	86	859	945	AAATAGCCTAAGC	2
Ab166_GEIH-2000	MSMG01001001	198	293	491	ACAAAAGAAAAAT	4
Ab166_GEIH-2000	MSMG01000974	79	1310	1389	TCTGCTGTCGGAAA	2
Ab166_GEIH-2000	MSMG01001128	194	304	498	ACGACGTGGACGATCTTC	3
Ab169_GEIH-2000	MSMF01000039	96	797	893	TCATTTTGCTGTTGTT	2
Ab169_GEIH-2000	MSMF01000336	198	152	350	ACAAAAGAAAAAT	4
Ab175_GEIH-2000	MSMI01000153	79	8115	8194	TTTCCGACAGCAGA	2
Ab175_GEIH-2000	MSMI01000682	86	2355	2441	AAATAGCCTAAGC	2
Ab177_GEIH-2000	MSME01000459	198	215	413	ACAAAAGAAAAAT	4
Ab183_GEIH-2000	MSMJ01000620	96	1077	1173	TCATTTTGCTGTTGTT	2
Ab183_GEIH-2000	MSMJ01000380	198	78	276	ACAAAAGAAAAAT	4

Strain	Contig	Size	Start	Stop	Repeat	N° of spacers
Ab192_GEIH-2000	MSMH01000263	96	1139	1235	TCATTTTGCTGTTGTT	2
Ab192_GEIH-2000	MSMH01000273	157	0	157	TTGAATTAAAA	4
Ab192_GEIH-2000	MSMH01000395	198	21634	21832	ACAAAAGAAAAAT	4

A phylogenetic tree of the complete CRISPR array sequences was constructed (Fig. 4), and showed an equal distribution of the spacers between the strains of the two years. Some of the spacers were predicted to be the same even in strains different year collections. Few of the arrays were unique respect to the other, such as the present in the 2000 strains Ab158_GEIH-2000_MSMC01000525, Ab166_MSMG01000383, Ab166_MSMG01000974, Ab166_MSMG01001128, Ab175_MSMI01000153 or Ab175_MSMI01000682. However, there were 5 CRISPR arrays grouped that only were represented in the 2010 strains (Fig. 4).

Cas-related proteins

When HMM against Cas-known, Cas-related and CRISPR-associated protein families was employed, 705 Cas-related proteins were identified in the 18 genomes: 341 Cas-related proteins were detected in 2000 strains and 364 in 2010 strains (Table 4, Additional file 3). Most of them were identified as DEAD/DEAH box helicase (207 of the total) and as Type III Restriction Unit Res III (195 of 705). The vast majority of them were located next to proteins whose predicted function does not match with a Cas protein function or to proteins whose function was unknown. Other Cas-related were close in the same contig thereby giving us a clue to help identifying a functional Cas cluster. For example, in the contig MSLX01000260 from the Ab103_GEIH-2010 strain a putative Helicase_C protein (OLV37994.1) and a Cas_St_Csn2 protein (OLV37998.1) were only of 2 ORF distance between them. However, the function of the surrounding proteins was hypothetical, thus hindering the identification process as a Cas cluster.

Discussion

In clinical laboratories, genomics is rapidly being developed and utilized to track antibiotic resistance. As a result, it is critical to explore how to detect and avoid phage resistant strains, if a treatment based on phages was going to be applied, by using WGS metadata analysis. In this study, we looked for genes linked to phage resistance in 18 clinical strains of *A. baumannii*. We constructed a database with genes based in the public PADS database, as it is the most complete database about prokaryotic antiviral defence systems so far, as well as being collecting newly discovered types of defence systems to the BIG Data Center (24, 51). In this case, the high number of genes made us establish groups in order to simplify the results of the blast hits. We also tried to identify the presence of CRISPR-Cas systems by separating the search in CRISPR arrays and Cas proteins.

A difference between the presence of phage resistance genes in 2010 strains and 2000 strains was observed, with a higher presence of genes related to the RM system and CRISPR-Cas system and lower of

TA-related genes. The natural reciprocal selection pressure between host bacteria and phage increases the infectivity of the phage and the phage-resistance in the bacterium side (52). In fact, phage populations are ubiquitous at body surfaces such as lungs, intestines or skin, and they outnumber bacteria at least by 10-fold (53). In this study, the acquisition of phage resistance genes is correlated with a higher presence of complete prophages in the strains of the 2010 in comparison with those of 2000 (54). This could be a result of the development of phage resistance adaptative systems, that could promote the emergence of new phages that can overcome them, such as could happen with Ab105-1 ϕ and Ab105-2 ϕ , two prophages present in the 2010 collection strains but not in the 2000 collection strains (54).

Defence systems are regularly obtained by bacteria and archaea through horizontal gene transfer (HGT) owing to environmental adaptation of the bacterial communities (55, 56). We found a major average of genes acquired by HGT in the 2010 strains rather than in the 2000 ones, especially those genes related to RM system and CRISPR-Cas. It was demonstrated that only ~ 4% of RM systems are in the core genomes of prokaryotic species, suggesting they are commonly transferred (57). CRISPR-Cas systems display weak consistency within the core genome, demonstrating the prevalence of the HGT spreading this system (57, 58). The RM system and the CRISPR-Cas system commonly coexist with an elevated contribution to the bacterial immunity and they rarely operate on their own (57, 59). However, they are far from being perfect in the bacterial resistance, and phage can escape these systems by many different ways, for example the anti-CRISPR proteins (8, 60). We also observed a decreasing number of TA-related genes through the years, even their presence in GIs is higher in the 2000 strains than in the 2010's. This could be because the counteradaptation of the phage may be reached by developing antitoxin in the phage genome that inhibit the cell death and thus promote the infection of the phage (61, 62) or because they could have evolved into Cas proteins of the CRISPR-Cas system, as the TA proteins are considered as ancestors of Cas2 proteins (63).

Furthermore, we found the CRISPR-Cas genes blast hit results incomplete due to the separation in contig assembly of the genomes, which prevented us from identify proteins or arrays related to the CRISPR-Cas proteins identified in small contigs (data not shown), and also due to the high diversity of the Cas proteins and the little knowledge about these proteins in clinical strains of *A. baumannii*, which increases the difficulty identifying these type of proteins (20, 64). As a consequence, we examined the presence of CRISPR arrays and Cas proteins separately. We establish a methodology to discard false-CRISPR elements based on the method of Shmakov et al. (65) and posteriorly completed with a full evaluation of the quality of the CRISPR arrays filtered based on the search of tandem repeats, simple repeats and their presence on phage genomes (35). Secondly, another reason of developing an alternative method is the nature of the multi-resistant pathogens, their constant adaptation to different environments and thus the continuous acquisition of different mobile elements, which provokes the appearance of new CRISPR-Cas yet to be identified (66). This also fosters and extends the variability in the Cas proteins, complicating their characterization.

40 CRISPR arrays were found in the 18 *A. baumannii* clinical strains from the ST-2 clone. All of the strains presented at least one CRISPR array except one, Ab161_GEIH-2000. The vast majority of the arrays are shared between the clone ST-2 in both collections, with some exceptions such as the five arrays only found

in 2010 strains. It has been shown that the distribution of CRISPR-Cas system is MLST dependent and non-random, and thought to be a better discriminating tool than classical MLST in discriminating different *K. pneumoniae* (67, 68). On the other hand, the detection of different unique CRISPR arrays only in the 2000 strains demonstrates the dynamic interaction of these arrays throughout the years.

All of the CRISPR arrays in this study were without any Cas or putative Cas protein near to them. It was described that these “orphan” arrays belong to unknown CRISPR-Cas systems due to be to an extremely evolutionarily remote type of CRISPR-Cas (65). This existence of isolated CRISPR arrays could be explained for four reasons. First, the contig format of the studied genomes could provoke that some arrays are detected in small or incomplete contigs. Secondly, some Cas endonucleases such as Cas1 and/or Cas6 can recognize remote CRISPR arrays (69, 70). Third, it may occur the possibility of some of the unique isolated arrays form part of an undescribed CRISPR-Cas cluster extremely distant to the ones already characterized (65). And fourth, the strains may have lost the *cas* genes thus leaving the isolated arrays (65). The Cas distribution observed in this work would correspond and complete any of the hypothesis about the explanation of “orphan” CRISPR arrays mentioned before as the putative Cas proteins hit through the HMM search could form part of a complete Cas cluster. However, as it was said at the ending of the results section, it was impossible to determine *in silico* if the putative Cas detected form part of a complete and functional Cas loci.

The localization and characterization of defence systems against phages is a necessary step when designing an effective phage therapy. The WGS combined with an effective bioinformatics strategy would allow us to know what mechanisms the clinical strains have. This study shows the wide presence of genes associated with resistance against phages and their acquisition by GIs for 10 years in clinical *A. baumannii* strains from the same clonal complex ST-2 and the CRISPR arrays present on them.

Abbreviations

WGS: Whole Genome Sequencing; ST: Sequence Type; CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats; GI: Genomic Island; ESKAPE: *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa* and *Enterobacter spp.*; ABI: Abortive Infection; TA: Toxin Antitoxin; R-M: Restriction-Modification; GEIH: *Grupo de Estudio de Infección Hospitalaria* (Hospitalary Infection Group of Study); GC: Guanine Cytosine; CRT: CRISPR Recognition Tool; ORF: Open Reading Frames; CDD: Conserved Domain Database; HMM: Hidden Markov Model; PADS: Prokaryotic Antiviral Defence System; HGT: Horizontal Gene Transfer; MLST: Multi-Locus Sequence Typing

Declarations

Funding

This study was funded by grants PI16/01163 and PI19/00878 awarded to M. Tomás within the State Plan for R+D+I 2013-2016 (National Plan for Scientific Research, Technological Development and Innovation 2008–2011) and co-financed by the ISCIII-Deputy General Directorate of evaluation and Promotion of

Research-European Regional Development Fund “A way of Making Europe” and Instituto de Salud Carlos III FEDER. M.Tomás was financially supported by the Miguel Servet Research Programme (SERGAS and ISCIII).

Acknowledgements

We are grateful to the following organizations and researchers who participated in the study GEIH-GEMARA (SEIMC, <http://www.seimc.org>) and the REIPI (Spanish Network for the Research in Infectious Disease-REIPI, RD12/0015/0010 and RD16/0016/0001): A Coruña Hospital (Germán Bou Arévalo) Virgen Macarena (Jesús Rodríguez-Baño, Alvaro Pascual, Felipe Fernández-Cuenca) Virgen Rocío (Jerónimo Pachón, Jose Miguel Cisneros, Younes Smani, José Garnacho, Antonio Gutierrez Pizarra, Juan Antonio Márquez Vácaro), Clinic Hospital (Jordi Vila), Hospital Marqués de Valdecilla (María Eliecer Cano, M. Carmen Fariñas), Hospital SAS La Línea (Antonio Sánchez Porto, Gloria Esteban Meruendano, Luis Barbeyto Vales, Javier Casas Ciria, Luis Vallejo), Complejo hospitalario de Ourense (Begona Fernández Pérez, José Carlos Villar Chao), Hospital Gregorio Marañón (Belén Padilla Ortega, Emilia Cercenado Mansilla), Hospital de Navarra (José Javier García Irure), Hospital Costa del Sol-Marbella (Alfonso del Arco Jiménez), Hospital General de Valencia (Concepción Gimeno Cardona, Juan Carlos Valía, Núria Tormo Palop, Vicente Abril, Josefina Rifa, Maria Jesus, Martinez Garcia), Consorci Hospitalari de Vic (Joseph Vilaró Pujals, Marian Navarro Aguirre, Ana Vilamala), Policlínica Guipúzkoa (José Antonio Jiménez Alfaro, Carlos Reviejo Jaca), Hospital Puerta del Mar (Pilar Marín Casanova, Francisca Guerreiro, Evelyn Shaw, Virginia Plasencia), Complejo Hospitalario de Soria (Teresa Nebreda Mayoral, María José Fernández Calavia, Susana García de Cruz, Carmen Aldea Mansilla), Hospital Universitario de Alicante (Esperanza Merino de Lucas, Alfredo Zorraquino, Sergio Reus Bañuls), Hospital Infanta Cristina (Eugenio Garduno Esevenri, Luis López Sánchez), Hospital Universitario Central de Asturias (Ana Fleites Gutiérrez, Azucena Rodríguez Guardado, Alfonso Moreno), Hospital Donostia (José María García-Arenzana Anguera), Complejo Hospitalario Torrecárdenas (Serafín López Palmero, Manuel Rodríguez Maresca), Complejo Hospitalario Xeral-Calde Lugo (Fernando García Garrote, José Varela Otero, María del Pilar Alonso), Hospital Universitario Reina Sofía de Córdoba (Elisa Vidal Verdú, Fernando Rodríguez López), Hospital Universitario Santiago Compostela (Fernanda Pardo Sánchez, E. Ferrer Vizoso, B.Regueiro Garcia), Hospital Sant Pau (Mercé Gurgui, Roser Pericas, Virginia Pomar), Hospital Galdakao-Usansolo (Pedro María Olaechea Astigarraga, Rafael Ayarza Igartua), Hospital Son Dureta (María Dolores Maciá Romero, Enrique Ruiz de Gopegui Bordes), Hospital Puerta de Hierro (María Isabel Sánchez Romero), Hospital Juan Grande (Jesús García Mata, María José Goyanes, Cristina Morales Mateos), Hospital San Cecilio (José Hernández Quero, Trinidad Escobar Lara), Hospital Sant Joan de Reus (Frederic Ballester Bastardie, Simona Iftimie, Isabel Pujol Bajador), Hospital de Motril (María Isabel Galán Navarro, María Luz Cádiz Gurrea), Hospital San Agustín (Carmen Amores Antequera, Montserrat Gómez, Purificación Cantudo), Hospital de Granollers (Carmina Martí Salas, Jordi Cuquet Peragosa, Antonio Moreno Flores, Luis Anibarro García), Hospital de Segovia (Susana Hernando Real, Pablo A. Carrero González), Complejo Hospitalario de Pontevedra (María Angeles Pallarés González, Sergio Rodríguez Fernández), Hospital de Bellvitge (Miquel Pujol Rojo, Fe Tubau), Hospital Virgen de la Victoria de Málaga (Enrique Nuno Alvarez, María Ortega Torres), Hospital Doctor Moliner (Salvador Giner Almaraz, María Rosa Roca Castelló, Manuela Castillo, Elena Hortelano), Hospital 12 de Octubre (Fernando Chaves Sánchez, Ana García Reyne), Hospital

del Mar (Juan Pablo Horcajada Gallego, Concha Segura), Hospital San Agustín de Avilés (Gema Sierra Dorado, Raquel Yano Escudero), Complejo Hospitalario Materno Insular de Gran Canaria (María Elena Dorta Hung, Cristóbal R.

Author contributions

A.A., L.B., M.L., O.P. L.F-G., I.B., developed the analysis of results and wrote manuscript. C.O.C., A.M., revised manuscript. M.T., financed and directed the experiments as well as supervised the writing of the manuscript.

Transparency declarations

The authors have not actions to declare

Availability of data and materials

The datasets analysed during the current study are genomes of the "II Spanish Multicenter Study. GEIH-REIPI *Acinetobacter baumannii* 2000-2010" available in the BioProject PRJNA422585 repository with the link <https://www.ncbi.nlm.nih.gov/bioproject/422585>

References

1. Boral B, Unaldi Ö, Ergin A, Durmaz R, Eser Ö, Group AS. A prospective multicenter study on the evaluation of antimicrobial resistance and molecular epidemiology of multidrug-resistant *Acinetobacter baumannii* infections in intensive care units with clinical and environmental features. *Ann Clin Microbiol Antimicrob.* 2019;18(1):19.
2. Stern A, Sorek R. The phage-host arms race: shaping the evolution of microbes. *Bioessays.* 2011;33(1):43–51.
3. Schooley RT, Biswas B, Gill JJ, Hernandez-Morales A, Lancaster J, Lessor L, et al. Development and Use of Personalized Bacteriophage-Based Therapeutic Cocktails To Treat a Patient with a Disseminated Resistant *Acinetobacter baumannii* Infection. *Antimicrob Agents Chemother.* 2017;61(10).
4. Oechslin F. Resistance Development to Bacteriophages Occurring during Bacteriophage Therapy. *Viruses.* 2018;10(7).
5. Bishop-Lilly KA, Plaut RD, Chen PE, Akmal A, Willner KM, Butani A, et al. Whole genome sequencing of phage resistant *Bacillus anthracis* mutants reveals an essential role for cell surface anchoring protein CsaB in phage AP50c adsorption. *Viol J.* 2012;9:246.
6. Broniewski JM, Meaden S, Paterson S, Buckling A, Westra ER. The effect of phage genetic diversity on bacterial resistance evolution. *ISME J.* 2020;14(3):828–36.
7. Mintzer V, Moran-Gilad J, Simon-Tuval T. Operational models and criteria for incorporating microbial whole genome sequencing in hospital microbiology - A systematic literature review. *Clin Microbiol*

- Infect. 2019;25(9):1086–95.
8. Labrie SJ, Samson JE, Moineau S. Bacteriophage resistance mechanisms. *Nat Rev Microbiol.* 2010;8(5):317–27.
 9. Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, et al. Systematic discovery of antiphage defence systems in the microbial pangenome. *Science.* 2018;359(6379).
 10. Krupovic M, Makarova KS, Forterre P, Prangishvili D, Koonin EV. Casposons: a new superfamily of self-synthesizing DNA transposons at the origin of prokaryotic CRISPR-Cas immunity. *BMC Biol.* 2014;12:36.
 11. Azam AH, Tanji Y. Bacteriophage-host arm race: an update on the mechanism of phage resistance in bacteria and revenge of the phage with the perspective for phage therapy. *Appl Microbiol Biotechnol.* 2019;103(5):2121–31.
 12. Lopatina A, Tal N, Sorek R. Abortive Infection: Bacterial Suicide as an Antiviral Immune Strategy. *Annu Rev Virol.* 2020;7(1):371–84.
 13. Harms A, Brodersen DE, Mitarai N, Gerdes K. Toxins, Targets, and Triggers: An Overview of Toxin-Antitoxin Biology. *Mol Cell.* 2018;70(5):768–84.
 14. Song S, Wood TK. A Primary Physiological Role of Toxin/Antitoxin Systems Is Phage Inhibition. *Front Microbiol.* 2020;11:1895.
 15. Vasu K, Nagaraja V. Diverse functions of restriction-modification systems in addition to cellular defence. *Microbiol Mol Biol Rev.* 2013;77(1):53–72.
 16. Mojica FJ, Rodriguez-Valera F. The discovery of CRISPR in archaea and bacteria. *FEBS J.* 2016;283(17):3162–9.
 17. Guan L, Han Y, Zhu S, Lin J. Application of CRISPR-Cas system in gene therapy: Pre-clinical progress in animal model. *DNA Repair (Amst).* 2016;46:1–8.
 18. Barrangou R. CRISPR-Cas systems and RNA-guided interference. *Wiley Interdiscip Rev RNA.* 2013;4(3):267–78.
 19. Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, et al. Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol.* 2011;9(6):467–77.
 20. Mangas EL, Rubio A, Álvarez-Marín R, Labrador-Herrera G, Pachón J, Pachón-Ibáñez ME, et al. 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.* 2019;5(11).
 21. Karah N, Samuelsen Ø, Zarrilli R, Sahl JW, Wai SN, Uhlin BE. CRISPR-cas subtype I-Fb in *Acinetobacter baumannii*: evolution and utilization for strain subtyping. *PLoS One.* 2015;10(2):e0118205.
 22. López M, Rueda A, Florido JP, Blasco L, Gato E, Fernández-García L, et al. Genomic Evolution of Two *Acinetobacter baumannii* Clinical Strains from ST-2 Clones Isolated in 2000 and 2010 (ST-2_clon_2000 and ST-2_clon_2010). *Genome Announc.* 2016;4(5).
 23. López M, Mayer C, Fernández-García L, Blasco L, Muras A, Ruiz FM, et al. Quorum sensing network in clinical strains of *A. baumannii*: AidA is a new quorum quenching enzyme. *PLoS One.* 2017;12(3):e0174454.

24. Zhang Y, Zhang Z, Zhang H, Zhao Y, Xiao J. PADS Arsenal: a database of prokaryotic defence systems related genes. *Nucleic Acids Res.* 2020;48(D1):D590-D8.
25. Bertelli C, Laird MR, Williams KP, Lau BY, Hoad G, Winsor GL, et al. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res.* 2017;45(W1):W30-W5.
26. Liu M, Li X, Xie Y, Bi D, Sun J, Li J, et al. ICEberg 2.0: an updated database of bacterial integrative and conjugative elements. *Nucleic Acids Res.* 2019;47(D1):D660-D5.
27. Zhang R, Ou HY, Gao F, Luo H. Identification of Horizontally-transferred Genomic Islands and Genome Segmentation Points by Using the GC Profile Method. *Curr Genomics.* 2014;15(2):113–21.
28. Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Néron B, et al. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* 2018;46(W1):W246-W51.
29. Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyrpides NC, et al. CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics.* 2007;8:209.
30. Rho M, Wu YW, Tang H, Doak TG, Ye Y. Diverse CRISPRs evolving in human microbiomes. *PLoS Genet.* 2012;8(6):e1002441.
31. Shmakov SA, Utkina I, Wolf YI, Makarova KS, Severinov KV, Koonin EV. CRISPR Arrays Away from cas Genes. *Crispr j.* 2020;3(6):535–49.
32. Cook R, Brown N, Redgwell T, Rihtman B, Barnes M, Clokie M, et al. INfrastructure for a PHAge REference Database: Identification of large-scale biases in the current collection of phage genomes. *bioRxiv.* 2021:2021.05.01.442102.
33. Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, et al. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res.* 2020;48(D1):D265-D8.
34. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, et al. CDD: NCBI's conserved domain database. *Nucleic Acids Res.* 2015;43(Database issue):D222-6.
35. Zhang Q, Ye Y. Not all predicted CRISPR-Cas systems are equal: isolated cas genes and classes of CRISPR like elements. *BMC Bioinformatics.* 2017;18(1):92.
36. Zhou K, Aertsen A, Michiels CW. The role of variable DNA tandem repeats in bacterial adaptation. *FEMS Microbiol Rev.* 2014;38(1):119–41.
37. Rando OJ, Verstrepen KJ. Timescales of genetic and epigenetic inheritance. *Cell.* 2007;128(4):655–68.
38. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 1999;27(2):573–80.
39. Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics.* 2004;Chap. 4:Unit 4.10.
40. Zhang Q, Doak TG, Ye Y. Expanding the catalog of cas genes with metagenomes. *Nucleic Acids Res.* 2014;42(4):2448–59.
41. Chai G, Yu M, Jiang L, Duan Y, Huang J. HMMCAS: A Web Tool for the Identification and Domain Annotations of CAS Proteins. *IEEE/ACM Trans Comput Biol Bioinform.* 2019;16(4):1313–5.

42. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 2011;39(Web Server issue):W29-37.
43. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. *Nucleic Acids Res.* 2014;42(Database issue):D222-30.
44. Haft DH, Selengut JD, White O. The TIGRFAMs database of protein families. *Nucleic Acids Res.* 2003;31(1):371-3.
45. Sievers F, Higgins DG. Clustal Omega, accurate alignment of very large numbers of sequences. *Methods Mol Biol.* 2014;1079:105-16.
46. Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics.* 2009;25(11):1422-3.
47. Chapman B, Chang J. Biopython: Python tools for computational biology. *SIGBIO Newsl.* 2000;20(2):15-9.
48. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol.* 2016;33(7):1870-4.
49. Kumar S, Tamura K, Nei M. MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. *Comput Appl Biosci.* 1994;10(2):189-91.
50. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994;22(22):4673-80.
51. Members BDC. Database Resources of the BIG Data Center in 2019. *Nucleic Acids Res.* 2019;47(D1):D8-D14.
52. Hampton HG, Watson BNJ, Fineran PC. The arms race between bacteria and their phage foes. *Nature.* 2020;577(7790):327-36.
53. Batinovic S, Wassef F, Knowler SA, Rice DTF, Stanton CR, Rose J, et al. Bacteriophages in Natural and Artificial Environments. *Pathogens.* 2019;8(3).
54. López M, Rueda A, Florido JP, Blasco L, Fernández-García L, Trastoy R, et al. Evolution of the Quorum network and the mobilome (plasmids and bacteriophages) in clinical strains of *Acinetobacter baumannii* during a decade. *Sci Rep.* 2018;8(1):2523.
55. Koonin EV, Makarova KS, Wolf YI. Evolutionary Genomics of Defense Systems in Archaea and Bacteria. *Annu Rev Microbiol.* 2017;71:233-61.
56. van Houte S, Buckling A, Westra ER. Evolutionary Ecology of Prokaryotic Immune Mechanisms. *Microbiol Mol Biol Rev.* 2016;80(3):745-63.
57. Oliveira PH, Touchon M, Rocha EP. The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Res.* 2014;42(16):10618-31.
58. Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, et al. An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol.* 2015;13(11):722-36.
59. Dupuis M, Villion M, Magadán AH, Moineau S. CRISPR-Cas and restriction-modification systems are compatible and increase phage resistance. *Nat Commun.* 2013;4:2087.

60. Bondy-Denomy J, Pawluk A, Maxwell KL, Davidson AR. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature*. 2013;493(7432):429–32.
61. Otsuka Y, Yonesaki T. Dmd of bacteriophage T4 functions as an antitoxin against *Escherichia coli* LsoA and RnIA toxins. *Mol Microbiol*. 2012;83(4):669–81.
62. Wei Y, Gao Z, Zhang H, Dong Y. Structural characterizations of phage antitoxin Dmd and its interactions with bacterial toxin RnIA. *Biochem Biophys Res Commun*. 2016;472(4):592–7.
63. Makarova KS, Wolf YI, Iranzo J, Shmakov SA, Alkhnbashi OS, Brouns SJJ, et al. Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat Rev Microbiol*. 2020;18(2):67–83.
64. Pinilla-Redondo R, Mayo-Muñoz D, Russel J, Garrett RA, Randau L, Sørensen SJ, et al. Type IV CRISPR-Cas systems are highly diverse and involved in competition between plasmids. *Nucleic Acids Res*. 2020;48(4):2000–12.
65. Shmakov SA, Utkina I, Wolf YI, Makarova KS, Severinov KV, Koonin EV. CRISPR Arrays Away from *cas* genes. *CRISPR J*. 2020;3(6):535–49.
66. Kamruzzaman M, Iredell JR. CRISPR-Cas System in Antibiotic Resistance Plasmids in. *Front Microbiol*. 2019;10:2934.
67. Liao W, Liu Y, Chen C, Li J, Du F, Long D, et al. Distribution of CRISPR-Cas Systems in Clinical Carbapenem-Resistant *Klebsiella pneumoniae* Strains in a Chinese Tertiary Hospital and Its Potential Relationship with Virulence. *Microb Drug Resist*. 2020;26(6):630–6.
68. Shen J, Lv L, Wang X, Xiu Z, Chen G. Comparative analysis of CRISPR-Cas systems in *Klebsiella* genomes. *J Basic Microbiol*. 2017;57(4):325–36.
69. Reimann V, Ziemann M, Li H, Zhu T, Behler J, Lu X, et al. Specificities and functional coordination between the two Cas6 maturation endonucleases in *Anabaena* sp. PCC 7120 assign orphan CRISPR arrays to three groups. *RNA Biol*. 2020;17(10):1442–53.
70. Hoikkala V, Ravantti J, Díez-Villaseñor C, Tirola M, Conrad RA, McBride MJ, et al. Cooperation between Different CRISPR-Cas Types Enables Adaptation in an RNA-Targeting System. *mBio*. 2021;12.

Figures

Figure 1

Representation of the main mechanisms of bacterial resistance against phage infection (from the left in a clockwise sense). 1. a: The phage recognizes the bacterial membrane receptor and can carry out the infection; b: Alterations in the receptors are produced by mutations and prevent the phage from recognizing the receptor, so it will not infect the bacteria; c: The bacteria can block recognition by producing inhibitors that bind to receptors. 2. Production of exopolysaccharide or extracellular matrix. 3. OMVs are composed of membrane lipids, membrane proteins and periplasmic components and they are as a decoy against phages as a defence mechanism. 4. a. Bacteria block the injection of DNA from other phages, acquiring Sie systems through prophages with this type of protein; b. Once the bacterium has the prophage in its genome with the proteins that code for the Sie system, it will be able to block the entry of DNA from other phages. 5. a. The R-M system distinguishes between methylated and unmethylated DNA. Restriction enzymes cannot cut methylated DNA; b. If the phage DNA is not methylated, this system can cut the injected DNA. 6. CRISPR-Cas recognizing phage DNA sequences, incorporating them into the system and producing enzymes that are capable of recognizing these sequences to cut them. 7. The PICI system is found in the bacterial genome and induced by helper prophages to produce mature phage particles that assemble the PICI system itself to kill the infected cell and spreading this system to adjacent cells. 8. a. The most characterized Abi system is the toxin-antitoxin system. Under normal conditions, the bacterium expresses both proteins equally, so cell death does not occur; b. When the organism is subjected to stress situations, such as phage infection, the toxin is highly expressed in comparison with the antitoxin, causing cell death.

Figure 2

Search procedure of CRISPR-Cas systems in 18 genomes of *A. baumannii* clinical strains. Orange rectangle represents bioinformatic programmes used for that task. Green plus marks represent a positive result for each operation. Red minus marks represent a negative result for each operation.

Figure 3

A. Frequency (% , rounded to two decimal numbers) of each group of genes in each genome respect to the total of genes: ABI (Abortive Infection System), TA (Toxin/Antitoxin system), RM (Restriction-Modification system) and NEW (genes associated with newly phage resistance bacterial mechanisms, e. g., Zorya, Hachiman, Druantia. B. Presence and non-presence of the putative phage resistance genes in GIs. The presence (% , rounded without decimal numbers) section is divided into the different groups of genes.

Figure 4

Phylogenetic classification of the CRISPR arrays detected in 18 genomes of *A. baumannii* ST-2 clinical strains through a maximum-likelihood tree with the suggested model by the MEGA analysis Tamura 3-parameter with uniform rates among sites and a bootstrap of 100. Blue rhomboid indicates that the strain belongs to the 2000 collection Red circles to the 2010 collection. Green rectangle indicates 5 CRISPR arrays only detected in the 2010 strains.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.docx](#)
- [Additionalfile2.xls](#)
- [Additionalfile3.xls](#)