IRAQI Academic Scientific Journals
العراقية المجلات الاكاديمية العلمية
University of Kut Journal
ISSN (E): 2616 - 7808 II ISSN (P): 2414 - 7419
www.kutcollegejournal.alkutcollege.edu.iq
k.u.c.j.sci@alkutcollege.edu.iq
Special Issue for the Researches of the 6th Int. Sci. Conf. for Creativity for 16 -17 April 2025

Characterization of Simple Sequence Repeats (SSRs) in genomes of Acinetobacter baumannii and evaluation of their potential role in phase variation

Ali Abdulwahid Al-Rubaiawi [1]

Abstract

The multidrug resistant Acinetobacter baumannii is one of the emerging health threatening -bacteria globally. Phase variation in this bacterium showed to play a role in switching between different phenotypes that displayed different capacities to resist adverse conditions. In this study, computer based analysis was applied to characterize the Simple Sequence Repeats (SSRs) in 32 genomes of A. bumannii, and to evaluate their potential mediating of phase variation of certain genes. SSRs were screened in terms of their types, periodic length, motifs and location within coding or intergenic regions. A total of 2277 SSR loci were detected in the total genomes analyzed, dominated by poly A/T tracts which represented 57.1 % of the whole SSRs detected. Excluding the tracts that not affecting gene expression, the total number of candidate phase variable (PV) genes identified were 1860 genes, ranged from 49 to 79 genes per genome. The majority of these genes have their SSRs within coding sequences. The genes clustered into 402 distinct homologous groups, and 42 of them are found in more than 50 % of the total isolates analyzed. Only 27 genes showed interstrain variability in their SSR numbers, and only five of them switched between different expression states (On/ Off) over genomes. In general, the detected SSRs are mainly short, and have a limited role in variation of surface antigens, but it potentially mediates phase variation in genes that are supposed to participate in drug resistance, reflecting their possible importance in regulating such ability in A. baumannii.

Keywords: Phase Variation; Simple Sequence Repeats; Multi-Drug Resistance; Genome analysis; Acinetobacter baumannii

Affiliation of Author

[1] Department of dormitory affairs, Mustansiriyah University, Iraq, Baghdad, 10052

[1] aliabed.81@uomustansiriyah.edu.iq

[1] **Corresponding Author**

Paper Info.
Published: Oct. 2025

توصيف التكرارات التسلسلية البسيطة في جينومات بكتيريا الراكدة البومانية (*Acinetobacter baumannii*) وتقييم دورها المحتمل في التغاير الطوري

علي عبد الواحد عبد الربيعاوي [1]

المستخلص

تُعد بكتيريا الراكدة البومانية *Acinetobacter baumannii* المقاومة للعديد من المضادات الحيوية واحدة من البكتيريا الصاعدة كتهديد للصحة على مستوى العالم. وقد أظهر التباين الطوري (phase variation) في هذه البكتيريا أنه يلعب دورًا في التبديل بين الانماط الظاهرية المختلفة والتي أظهرت قدرات مختلفة على مقاومة البكتريا للظروف المعاكسة. في هذه الدراسة، تم تطبيق التحليل القائم على الكمبيوتر لتوصيف التكرارات التسلسلية البسيطة (SSRs) في الحمض النووي لـ 32 جينومًا من بكتريا الراكدة البومانية، وتم تقييم قدرتها على التوسط في الاختلاف الطوري لبعض الجينات. تم فحص SSRs من حيث أنواعها واطوالها الدورية وأنماطها وموقعها داخل المناطق المشفرة أو المناطق بين الجينات. تم الكشف عن إجمالي 2277 موضعًا لـ SSR في إجمالي الجينومات التي تم تحليلها، والتي تهيمن عليها تتابعات الادنين/ الثايمين (A/T) المتعددة والتي تمثل 57.1٪ من إجمالي SSRs المكتشفة. باستثناء المناطق التي لا تؤثر على التعبير الجيني، بلغ العدد الإجمالي للجينات المرشحة لحدوث التباين الطوري فيها 1860 جينًا، وبمعدل من 49 إلى 79 جينًا لكل جينوم. تحتوي غالبية هذه الجينات على SSRs ضمن التسلسلات التي تشفر للبروتين. صنفت هذه الجينات الى 402 مجموعة اعتمادا على تماثلها، وكان 42 منها في موجودا في أكثر من 50% من إجمالي العزلات التي تم تحليلها. أظهر 27 جينًا فقط تباينًا في اطوال SSR الخاصة بها بين الجينومات المختلفة، وخمسة منها فقط تغايرت بين حالات

انتساب الباحث
[1] قسم شؤون الاقسام الداخلية، الجامعة المستنصرية، العراق، بغداد، 10052

[1] aliabed.81@uomustansiriyah.edu.iq

[1] **المؤلف المراسل**

معلومات البحث
تأريخ النشر: تشرين الاول 2025

تعبير جيني مختلفة (تشغيل/إيقاف) في الجينومات المختلفة. بشكل عام، كانت (SSRs) المشخصة قصيرة في الغالب، ولها دور محدود في تباين الطور في المستضدات السطحية، ولكنها قد تتوسط تباين الطور في الجينات التي من المفترض أن تشارك في مقاومة الأدوية، مما يعكس أهمية محتملة لها في تنظيم قدرة هذه البكتريا على مقاومة المضادات الحيوية.

الكلمات المفتاحية: التباين الطوري، التكرارات التسلسلية البسيطة، مقاومة الادوية المتعددة، تحليل الجينوم، الراكدة البومانية

Introduction

A. baummannii is a multidrug-resistant, Gram-negative bacterium that is rated as the sixth most important nosocomial pathogen in the world [1]. The bacterium responsible for a wide range of hospital acquired infections in human including urinary tract infection, wound infection, meningitis, septicemia and most importantly ventilator associated pneumonia which are correlated with a high rate of mortality and morbidity [2]. It can also cause community acquired infection mainly in people with risk factors including chronic obstructive lung diseases or diabetic mellitus [3]. A. baummannii is ubiquitous in the nature and also colonize different body cites of human such as skin and digestive tracts asymptomatically [4]. The bacteria are one of the highly prevalent opportunistic pathogens in Iraq, and large number of cases are continuously reported over the different regions of the country with a wide range of nosocomial infections [5–7].

Phase variation (PV) is an adaptive mechanism that has been reported to happen in a wide range of pathogenic bacteria, which enhance their virulence. It involves several genetic mechanisms that lead to generate reversible phenotypic heterogeneity within the clonal population, giving it more flexibility to overcome the adverse conditions in their local environment [8].

Simple sequence repeats (SSRs) are short motifs of DNA sequences (ranging from 1-6 nucleotides) that are tandemly repeated several times within a locus [9]. SSRs showed high rate of length polymorphism due to reversible insertion/deletion mutations happened within it stretch during chromosome replication [10]. These hypermutable loci mediate phase variation in large number of surface antigens in many pathogenic bacteria increasing their adaptability to the environmental stress such the immune response [11]. Alteration in the length of SSR tracts affects the expression state of protein, changing it reversibly from high to low level when it lies within the regulatory region, or switched it from expression (ON state) to no expression (OFF state) when it occur in the coding sequence of a gene [12].

Phase variation has been reported previously in A. baummannii . It mediates switching between two different colony opacity phenotypes, opaque and translucent. The two phenotypes exhibit further variability including surface motility, cell morphology, drug resistance and biofilm formation [13]. Although these studies investigated the phase variation at the genetic level, they focused mainly on understanding the role of some negative regulators on the switching between the two phenotypes and other parameters differ between them such as motility and biofilm formation [14]. However, there was no study documented the characteristics of the simples sequence repeats in A. baumannii and their possible role in phase variation. Due to medical importance of A. baummannii as human pathogen that has emerged as global life threatening bacteria, understanding

the pathogenicity and the main mechanism that play a role in enhancing their virulence is crucial to develop an effective therapy. Therefore, this study was aimed to characterize the SSRs in 32 genomes of A. baumannii (publicly available) isolated from different sources and evaluate their potential role in phase variation and consequently in the pathogenicity of the Bactria.

Methods

Acinetobacter baumannii genomic datasets

The SSRs have been identified within 32 complete genomes of A. baumannii available in GenBank database. These genomes were chosen from literature and belongs to different A. baumannii isolates isolated from different sources, 16 of them from clinical sources and the other 16 were from non-clinical sources. The accession numbers and the details of the isolates are summarized in table 1.**Screening for SSRs loci and detection of candidate PV genes.**

The SSRs of each genome were investigated using phasomeIt software, which is available on github website (https://github.com/JackAidley/PhasomeIt/) [15]. The repeats were analyzed in terms of their types (homo or heteromeric repeats), their motif lengths (mono, di, tri, tetra, penta and hexameric repeats) and their number as array within a locus. Genomic loci harboring SSRs within their genic or intergenic regions were considered as candidate PV genes according to criteria used by [16]. The SSRs must be located within open reading frames (ORFs) or within the 200 bp upstream the ORFs if they occur within the intragenic regions. The length of the monomeric repeats should be more than eight repeats in the case of (A/T) tracts, and more than six repeats in the case of (G/C) motifs. Whereas in the heteromeric repeats, the dinucleotide must consist of more than five repeats, and the tri and tetranucleotides should consist of more than four repeats. All pentameric and hexameric repeats should be consist of more than two repeats as shown in Table (1).

Table (1): Details of A. baumannii genomes involved in this study

Genomes accession number	Isolates	Source	Reference
CP003849	BJAB0868	Ascites	[17]
CP003846	BJAB07104	Blood	[17]
CP003967	D1279779	Blood	[18]
CU459141	AYE	Urine	[19]
CP000521	ATCC 17978	Meningitis	[20]
CP001937	MDR-ZJ06	Blood	[21]
CP002522	TCDC-AB0715	Blood	[22]
CP003856	TYTH-1	Blood	[23]
CP000863	ACICU	CSF	[24]
CP001172	AB307-0294	Blood	[25]
CU468230	SDF	Body louse	[19]

SRX14056989	**354/17**	**Feces**	**[26]**
CP001921.	**1656-2**	**Sputum**	**[26]**
SRX14056995	**5250/17**	**Feces**	**[26]**
SRX14056996	**11453/17**	**Feces**	**[26]**
SRX14056997	**14259/17**	**Feces**	**[26]**
SRX14056999	16644/17	Feces	[26]
SRX14057000	29946/18	Feces	[26]
SRX14056988	37889/18	Feces	[26]
ACYS00000000.2	6014059	Skin	[27]
NZ_RZNI01000000	AB18PR065	Pig feces	[28]
NC_011586	AB0057	Blood	[25]
NZ_AEOX01000000	AB210	Pre-therapy clinical isolate	[29]
NZ_LAPU00000000	AB210M	Clinically derived strain from AB210	[30]
NZ_ABXK00000000	AB900	Perinea	[25]
PRJNA738868	DETAB-P95	Nasogastric tube	[31]
PRJNA738869	DETAB-P96	Rectal swab	[31]
PRJNA738870	DETAB-P99	Rectal swab	[31]
PRJNA738871	DETAB-P100	Oral swab	[31]
NZ_CP076812.1	UC23022	Blood culture	[32]
NZ_AMZT00000000	WC-348	Skin	[33]
NZ_AMGG00000000	WC-692	Skin	[33]

Clustering of homologous PV genes

The PV genes were assigned –by PhasomeIt- to homology groups if they share 50% similarity in their sequences. These groups may also contain non-PV homologous genes which in most cases differ in their repeat numbers or types (reduced to low number or interrupted with other bases), but share high similarity in other sequences with their homologous PV genes. The homology were also confirmed by extracting a number of putative PV loci from different genomes and blasted them against the 32 genomes, so their presence and their similarity within those genomes were confirmed.

Statistical analysis

Student t test was used to compare means of PV genes in clinical and non-clinical isolates, while Z-test was used to compare means of PV genes having SSRs in ORF and intergenic regions.

Results

SSRs have a significant role in the pathogenicity of the majority of pathogenic bacteria [8]. However, phase variation mediated by SSRs is poorly studied within the bacterium A. baumannii. A total of 32 genomes of A. baumannii were investigated for their SSR-loci using the software PhasomIt, 16

genomes were of clinical isolates and other 16 genomes were of isolates from non-clinical sources including faecal, skin and other body sites. The total number of SSR loci detected within the 32 genomes was 2277, and their average was 72.2 loci per genome (the average number of SSRs in clinical and non-clinical isolates were relatively similar which counts 71.12 and 71 .18 loci per genome respectively). This was dominated by homopolymeric repeats which represents 85.9 % of the total detected SSRs , where poly A/T motifs represented 57.1 % of the total SSRs and poly G/C, represented 28.8% of the total SSRs**.** Among the hetero repeats, the hexameric repeats were with the highest number that accounted 5.9 % of the detected SSRs, followed by the penta and tri nucleotide repeats which representing 5.4 % and 4.3 % respectively (figure 1). Whereas the other repeats, (di and tetrameric tracts) were with a few number and did not exceed nine loci within the total 32 genomes. The mean number and the aboundancy of the SSRs motifs within genomes were summarized in figure 2 A&B. Number and distribution pattern of the SSRs is quite similar between the both group of isolates (clinical and non-clinical) and no clear difference was noted (data not shown). A negative correlation was found between the length of SSR motifs and the number of their loci per genome. The highest number (92.3 %) of SSR loci harbor poly A/T tracts with 9 A/T repeats, 7 % with 10 repeats and only 0.06 of them with more than 10 repeats. In the case of poly G/C tracts, the majority of their loci (80%) were with seven repeats, 14 % with eight repeats, and only 5% of them were with more than eight repeats (figure 2, C). Such negative correlation become very clear in the case of shorter motifs especially poly (A/T) tracts (< 9) and the other short heteropolymer tracts (except penta and hexameric tracts). Such SSR loci increased exponentially with the decrease of their SSR'S length and distributed at a high number within both coding and non-coding regions (data not shown). as shown in Figure (1) and figure (2)**.**

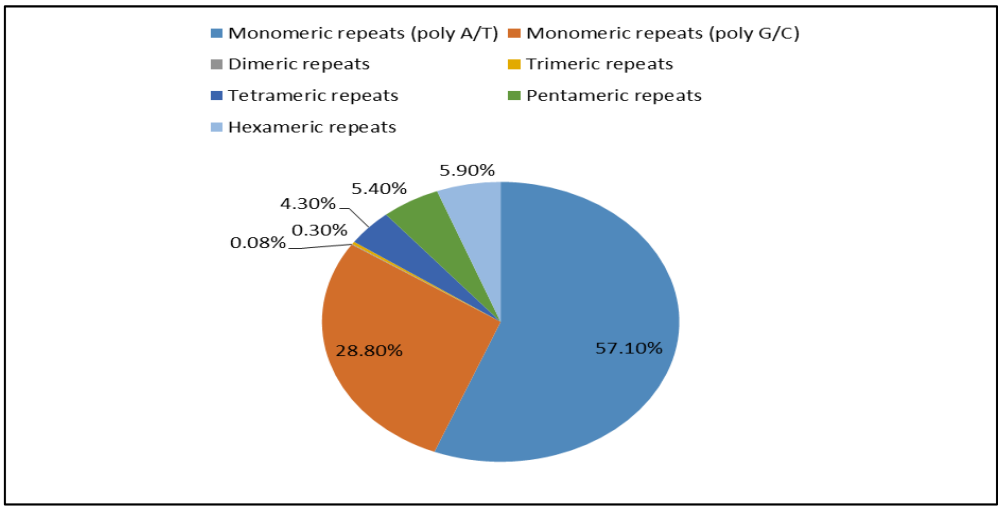

Figure (1): The ratios of the different types of SSRs loci detected within the whole genomes analysed

The figure displays the percentages of each SSRs types among the total 2277 SSR loci detected within the 32 genomes analysed in this study.

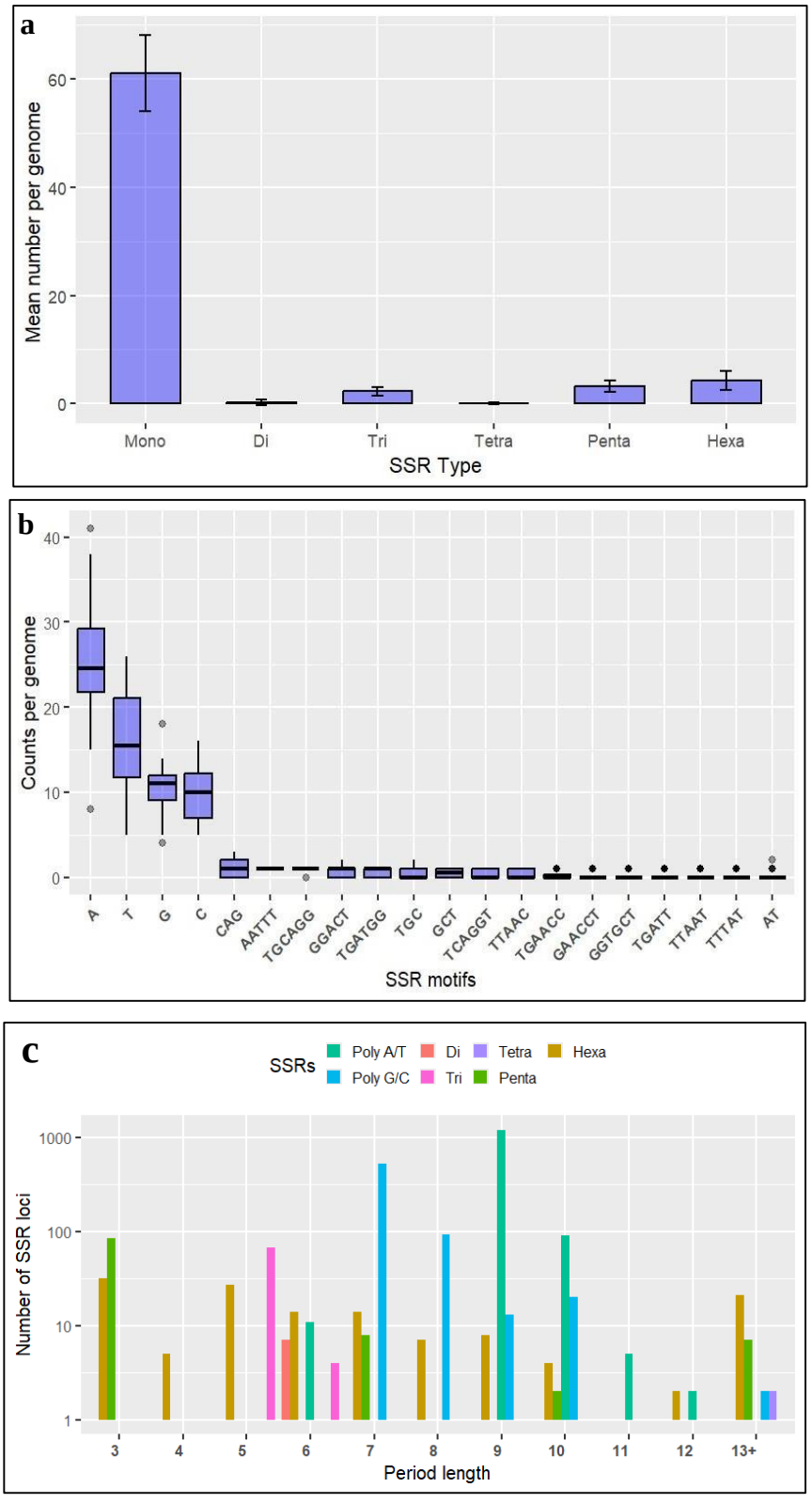

Figure (2): Abundance and distribution of SSRs within analyzed genomes. a: The mean number of different types of SSRs within the different genomes. The error bar represents the standard error. b: counts of the most abundant SSR motifs within genomes. The boxplots diplay the median, minimum, maximum and outlier numbers of motifs within the total genomes. c: the total number of SSRs according to their types and periodic length

Putative PV genes

Putative PV genes were characterized according to the criteria depended by Siena et al. [16] mentioned in previous section. Excluding the SSR loci located at >200 bp upstream genes, or occur directly downstream genes, the total number of candidate PV genes identified were 1860 genes within the 32 genomes analyzed, ranged from 49 to 79 genes per genome (with average of 58.125 genes per genome).

The genes were clustered into 402 distinct homologous gene groups. Regarding the SSRs location, 298 gene groups (74.1 %) harbor SSRs within their coding sequence and 104 groups (25.9 %) containing SSRs within 200 bps upstream their coding sequence. The differences between means of PV genes according to their SSRs location (within ORF or within intergenic regions) were statically significant: $p < 0.00001$using z test. The comparison of numbers of PV genes between clinical and non-clinical isolates showed no statistically significant difference between the mean number of PV gens in the groups of isolates. Among the total gene groups, 42 of these groups were found in more than 50 % of the total isolates analyzed; representing the core PV genes distributed within A. baumanni genomes as shown in (table 2).

Intergenomic variability of SSRs

Intergenomic variability of SSRs has been analyzed in the all candidate PV genes that presented in more than one genome using the software PhasomeIt. DNA sequence of a number of these genes were also extracted from one genome (either the whole ORFs or the SSRs with their flanked 200bps upstream and downstream sequence), and blasted against the 32 genomes to confirm the variability in SSR tract number noticed in homologous genes within different genomes. The data showed that only 27 candidate PV genes showed variability in their SSRs number, fourteen of them have their SSRs within coding regions while the others (13 genes) were containing SSRs within their intergenic sequence (table 3). The variability was mainly in genes with homopolymeric tracts, which represented 78 % (22 genes) of the total variable genes. Whereas, only genes with pentameric and hexameric repeats showed variability among the total genes harboring heteropolymeric SSRs. In general, the vast majority of variation noticed in the monomeric SSR loci was by adding/ removing only one nucleotide. In contrary, the pentameric and hexameric loci showed strong variability ranged from single repeat to more than 30 repeats. However, the insertion/ deletion of SSRs within coding sequences did not affect the expression states of majority of PV genes detected, which kept the same expression state (either On or Off) even they varied in their SSR numbers (as shown in table 3).

There was a high rate of reduction of the number of SSR to short number in homologous genes within different genomes. Such reduction may stabilize the repeat tracts at short number and prevent the slipped stand mechanism to happen. This was noticed along with nucleotide replacement that also happened within a large number of SSRs especially with poly G/C tracts. Such SSRs showed high rate of replacement in one to three G/C with A/T within their stretch that also lead to stabilize them as it breaks the long tracts into many shorter repeats (figure 5). These genes were identified as non-PV genes by PhasomeIt software due to their SSR number lower than the cutoff, and

sometime were not detected by the program even if they are present in all genomes analyzed. as shown in Figure (3), (4) and (5).

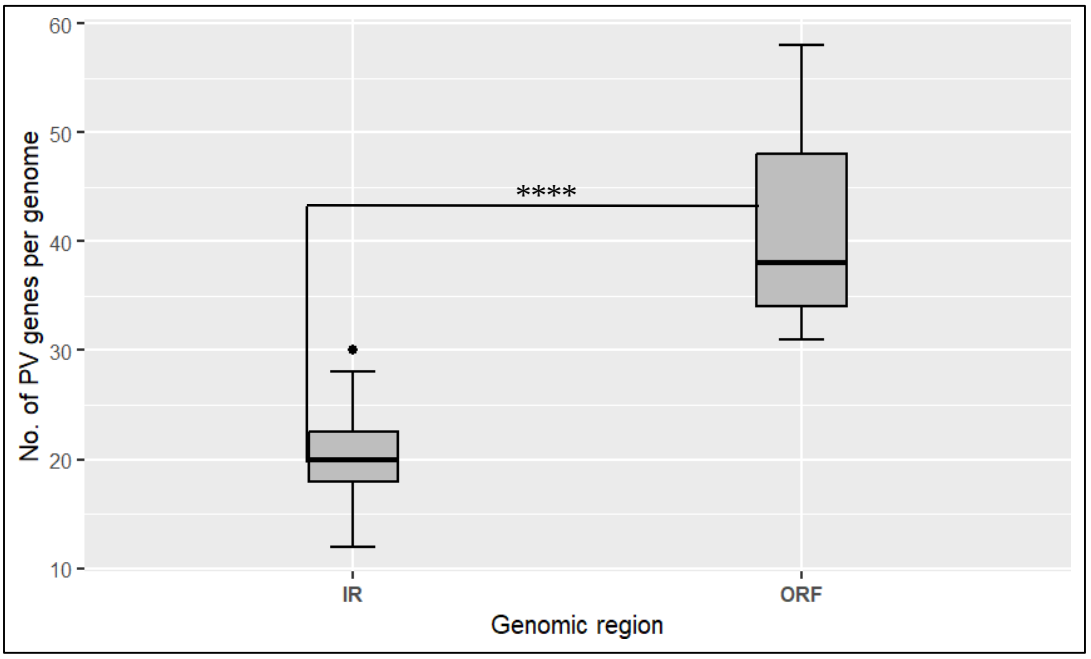

Figure (3): The number of candidate PV genes in A. baumannii isolates according to chromosomal regions of their SSRs. The data compare the number of SSRs per genomes according to their location whether intergenic regions (IR) or open reading frames (ORF).The boxplots display median, minimum, maximum and the outlier of the data. ** highly statically significant difference was found between the two categories (P<0.00001)**

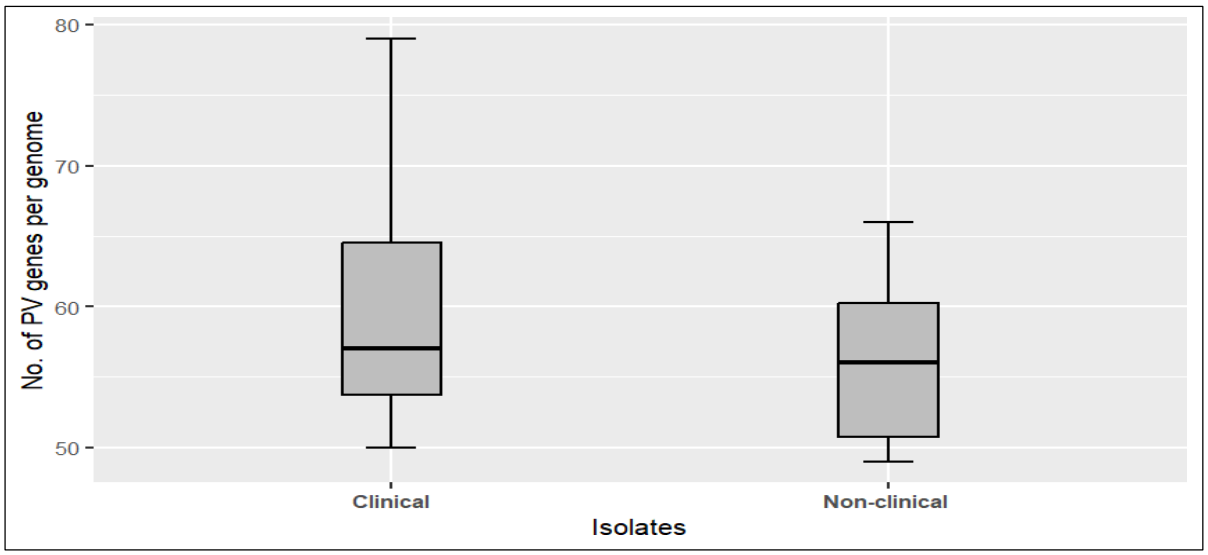

Figure (4): The number of candidate PV genes in A. baumannii isolates according to their source of isolation. The data compare the number of candidate PV genes per genome of the clinical (n=16) and non-clinical (n=16), and the boxplots display the median, minimum, maximum and the outlier of the data. No statically significant difference was found between the two categories of data

Table (2): Core SSR loci detected within 32 A. baumannii isolates analyzed in this study

Percentage of strains	PV Genes	Function	SSR motif
100	MBA24_000193	Tim44 domain-containing protein	A
	folP	dihydropteroate synthase	G
	btuB_2	TonB-dependent receptor	A
	aceE	pyruvate dehydrogenase (acetyl-transferring), homodimeric type	C
	thrC	threonine synthase	CTG
	MBA24_003030	RND transporter	AAATT
	iclR	IclR family transcriptional regulator	G
96.9	gltB	glutamate synthase large subunit	T
	emrA_5	HlyD family secretion protein	CCTGCA
	dmeF	CDF family Co(II)/Ni(II) efflux transporter DmeF	CATCAC
93.8	rplL	50S ribosomal protein L7/L12	CTG5
	ybaL	cation:proton antiporter	T
90.6	dnaX_1	DNA polymerase III subunit gamma/tau	ACCTGA
	ABK1_1601	conserved hypothetical protein; putative exported protein	T
87.5	phnA	alkylphosphonate utilization protein	T
	MBA24_001865	conserved hypothetical protein; putative membrane protein	A

	MBA24_002125	No annotation data	A
81.2	glsA	glutaminase	A
78.1	hfq	RNA chaperone Hfq	T
75	yvdD_1	TIGR00730 family Rossman fold protein	A
71.9	MBA24_003062	alpha/beta hydrolase	A
	MBA24_003076	insulinase family protein	G
68.8	MBA24_000422	GNAT family N-acetyltransferase	A
	MBA24_003482	EAL domain-containing protein	C
65.6	MBA24_000459	glutathione-dependent formaldehyde dehydrogenase	G
	ureC	urease subunit alpha	G
	MBA21_002413	hypothetical protein	A
62.5	MBA24_000894	glutathione S-transferase family protein	T
59.4	MBA24_003153	PQQ-dependent sugar dehydrogenase	G
	MBA24_003421	hypothetical protein	G
56.2	ABK1_1089	TonB-dependent receptor	G
	hemE	uroporphyrinogen decarboxylase	GGACT
53.1	spoT_1	guanosine polyphosphate pyrophosphohydrolases/synthetase	A

Table (3): Candidate PV genes that showed interstrain variability in their repeat tract number

PV Genes	Function	SSR motif	SSR location*	Expression state**	No. of genomes
MBA24_000193	Tim44 domain-containing protein	A (9-11)	ORF	All off	32
MBA24_000422	GNAT family N-acetyltransferase	A (9-10)	IR	-	22
MBA24_000459	glutathione-dependent formaldehyde dehydrogenase	G (7-8)	ORF	All off	21
MBA24_000478	porin Omp33-36	T (9-10)	IR	-	11
MBA24_000751	urease subunit alpha	G (9-13)	ORF	On/off	21
MBA24_000856	DNA polymerase III subunit gamma/tau	ACCTGA (5-31)	ORF	All off	29
MBA24_001560	HlyD family secretion protein	CCTGCA (4-8)	ORF	All on	31
MBA24_001734	ribosomal protein S18-alanine N-acetyltransferase	T (9-10)	IR	-	11
MBA24_001922	pyruvate dehydrogenase (acetyl-transferring), homodimeric type	C (8-9)	IR	-	32
MBA24_002610	cation:proton antiporter	T (9-10)	IR	-	30
MBA24_003016	CDF family Co(II)/Ni(II) efflux transporter DmeF	CATCAC (3-10)	ORF	All off	31
ABK1_0865	Tyrosine--tRNA ligase, putative	AATCA (11-31)	ORF	On/off	7

ACINWC692_1067	hypothetical protein	G (7-8)	IR	-	7
ABK1_3614	chitinase	CTACAC (4-7)	ORF	All on	3
HMPREF0022_01143	Sel1 repeat protein	G (7-9)	IR	-	9
ACICU_01027	hypothetical protein	A (9-10)	ORF	All on	21
HMPREF0022_02355	conserved domain protein	G (7-8)	IR	-	10
HMPREF0022_03607	putative ATP synthase F0, A subunit	G (7-8)	ORF	On/off	13
MBA32_002640	hypothetical protein	T (9-10)	IR	-	4
ACINWC692_0988	D-serine ammonia-lyase	A (9-11)	IR	-	5
KQ253_06715	glucose/quinate/shikimate family membrane-bound PQQ-dependent dehydrogenase	G (7-8)	ORF	On/off	6
KQ253_16005	FKBP-type peptidyl-prolyl cis-trans isomerase	T (9-10)	IR	-	9
AB57_0647	prohead core protein	CAAAG A (7-9)	ORF	All on	2
ACINWC692_1045	endopeptidase La	A (9-10)	IR	-	6
ABBFA_02125	FMN-dependent NADH-azoreductase	G (7-8)	ORF	All off	4
BJAB0868_01195	hypothetical protein	G (7-9)	ORF	On /off	6
ACINWC692_1481	hypothetical protein	G (7-8)	IR	-	3

*ORF: open reading frame, IR: intergenic region..

** Expression state: For genes that have their SSRs in ORF. On/Off: genes with different SSR numbers showed different expression state; all On, all Off: genes with different SSR numbers showed same expression states.

Figure (5): Multiple sequence alignment showing the variability in the poly G tracts within the gene MBA24_000751 (urease subunit alpha) from different genomes of A. baumannii analyzed in this study. The figure showed that variability in poly G tracts ranged from variability in their tracts number, interruption with multiple Ts, or the tracts were deleted completely from the genes. The software classified the two later types of genes as non-PV genes but they are homologous to the PV genes

Discussion

The role of phase variation in enabling the pathogenic bacteria to overcome adverse condition has been reported obviously [34]. Many pathogenic bacteria such as Neisseria meningitidis and Haemophylus influenza have many surface antigens that are subject to phase variation mediated by alteration of their SSRs that lies either in their coding sequences or in their regulatory regions [35,36] . One example is the outer membrane proteins Opa in N. meningitidis which display a clear switching between different variants within the clonal population during the longitudinal colonization of healthy carriers [37]. Such switching was in response to the immune stress imposed on Opa repertoire, and it participates in the immune evasion of the bacteria during the long-term colonization [38]. However, phase variation in A. baumannii has also shown to mediate reversible switching between different phenotypes, involving opaque colony producing phenotype and translucent colonies producing phenotype [13]. The two phenotypes differ in their reactivity with their environmental conditions including their resistance to immune cells, antibiotics, disinfectants, hydrolysis enzymes and producing of thick/thin capsules [13,39,40]. Genetic mechanisms that can affect PV in A. baumannii have been studied in a manner that focusing on the effects of the global regulators genes - such as transcriptional regulators (ABUW_ 1132) and (RND)-type efflux system (arpAB) – on the phase variation process. Mutating these regulators has shown to effect the rate of PV in the two mentioned phenotypes [41,42]. However, studying the genetic bases of mechanisms mediating PV in A. baummanii has a great implication in understanding the role of such mechanisms in the pathogenicity of these bacteria. In spite of the importance of the SSRs in PV, no studies have documented the features of SSRs in A.

baummannii and their potential role in phase variation. In this study, the simple sequence repeats in A. baumannii was investigated in 32 genomes published publicly. The genomes belong to different strains isolated from different sources including clinical specimens and other non-clinical specimens. Our data showed that SSRs are divers and highly represented within A. baumnnii genomes. A total of 2247 SSRs loci were detected within the 32 of A. baumnnii genomes. Repeats were dominated with monomeric tracts that represented about 86 % of total SSRs detected, where poly A/T represent the highest number in both the genic and intergenic regions. The SSRs length cut off used in this study were chosen depending on preliminary data from the same genomes analyzed and also according to previous studies in purpose of detection only SSRs that can mediate PV in their loci. Preliminary data showed that applying cut off lower than the one used here (A/T >8, G/C >6, di >5, tri & tetra >4 and penta & hexa >2 repeats), a very high number of SSRs will be detected that making it so difficult to interpret the results obtained. In the same time, all these tracts with shorter number were very stable and did not show any variability in their length among the different strains (data not shown), which may not reflect any potential mediation of phase variability in these loci. In general, tracts length in A. baumannii tend to be shorter than those in many other pathogenic bacteria characterized in other studies such as N. meningitidis [43], H. influenza [44] and Campylobacter jujuni [15]. This is clearer in monomeric repeats especially the poly G/C where their length peaked around seven and eight repeats and only one locus were with thirteen Gs , while in poly A/T , the length ranged between nine-eleven and peaked at nine As / Ts . Exception of pentameric and hexameric repeats, the other heteropolymeric repeats (di, tri and tetra) displayed similar pattern and did not exceed the threshold length used in the study.

The instability of the SSRs is positively correlated with their length, and shorter repeats tend to be more stable and may not mediate phase variation in the genes they lie within [45]. This suggests that these loci were favored to be more stable rather than highly variable, reflecting their possible essentiality to the cellular processes, especially genes that code for proteins with metabolic functions. Such possible essentiality may also justify the absence of a statically significant difference between clinical and non-clinical isolates of A. baumannii in their numbers of predicted PV genes. An equal importance of those genes to the both types of isolates might lead to maintain them to be expressed in a similar manner within the two groups. Alternatively, these genes may not critically participate in the virulence or in other adaptation mechanisms of clinical isolates, so there is no particular selection for those loci in such group rather than the other- non-clinical-group.

Interstrain variability in SSRs were noticed in 27 genes, fourteen of them carry their SSRs in coding sequences and the other twelve genes have SSRs within their regulatory regions. Such variability in SSR length display the role of the slipped strand mispairing mechanism (SSM) that happens within these loci leading to insert/delete one or more homopolymeric or heterpolyomeric repeat. SSM happening in coding regions lead to generate frame-shift mutations that may affect the expression state of the gene, leading to switching reversibly between on and off state. Whereas insertion/deletion happening in SSRs within regulatory region can affect the expression level of gene products by changing it from high level to

low and vice versa [46] . However, among the fourteen genes showed variability in SSRs within their coding sequence, only five showed changes in their expression state (i.e. switching between ON and OFF state), while others were in the same expression state even they varied in heir SSRs number. The gene MBA24_003016 encodes for heavy metal efflux transporter protein (DmeF) that mediates resistance of bacteria to zinc and heavy metal ions [47]. This gene showed strong variation in their hexameric SSRs length among different genomes, which may effects the function of their protein (when they in on state) and hence reflects a possible role for SSRs variation in regulating tolerance/resistance of the bacteria to the heavy metal ions. However, the majority of the variable SSRs were located in gens that mediate either metabolic function and cellular process, or they with unknown functions (hypothetical proteins), and no variability was detected in genes coding for surface proteins, except in poly Ttracts within regulatory region of the gene MBA24_000478 that encodes for Omp33-36 Porin. This protein is a surface protein that act as a channel for water traffic, and it plays a role in inducing cytotoxicity in immune and connective tissue and its expression participates in antibiotic resistance [48]. Similarly, Variability has also been noticed in intergenic region of mba24_000422 that encodes for n-acetyltransferase belong to GNAT family. Such n-acetyltransferase has been shown to confer antibiotic resistance towards aminoglycosides in bacteria [49]. The variability in the SSR number within these two genes as well as within MBA24_003016 and their effects on the expression levels / state of their proteins need to be more investigated to assess if they have any impact on drug resistance ability of A. baumannii.

Generally, SSRs of A. baumannii tend to be condensed in genes that code mainly for non-surface products where mutability in their SSRs are usually lower than those in genes coding for surface antiges. Such antigens are continuously exposed to environmental stress and display a high rate of phase variation via mutations occur in their SSRs as noticed in many pathogenic bacteria [50]. Lacking of such events may reflect a weak environmental stress imposed on surface proteins of A. baumannii during presence within living hosts, making it unnecessary to phase vary and change their expression states. Alternatively, the bacteria might depend other mechanisms to immune evasion rather than using phase variation, which then lead to maintain their surface protein repertoires in a fixed expression state.

Conclusion

SSRs are highly represented within A. baumannii genomes with mostly similar pattern in both clinical and non- clinical isolates, but they tend to be shorter and hence lower mutable than what is common in many pathogenic bacteria. However, evidences for phase variation mediated by SSRs have been reported in a considerable number of genes within A. baumannii, some of these genes encode for proteins that participate in antibiotic and heavy metal ions resistance, which may reflect a role for phase variation in regulating drug resistance in this bacterium. The effects of SSRs variability on the expression state/ level of proteins need to be confirmed by further investigation in order to assess their impacts on the adaptation capacity of A. baumannii during their persistence within hosts.

Acknowledgments

The author would like to thank Mustansiriyah University (www.uomustansiriyah.edu.iq)

Baghdad-Iraq, for its support in this work.

References

[1] Antunes LCS, Visca P, Towner KJ. Acinetobacter baumannii : evolution of a global pathogen. Pathogens Disease. 2014 Aug;71(3):292–301.

[2] Munoz-Price LS, Weinstein RA. Acinetobacter Infection. N Engl J Med. 2008 Mar 20;358(12):1271– 81.

[3] Dexter C, Murray GL, Paulsen IT, Peleg AY. Community-acquired Acinetobacter baumannii: clinical characteristics, epidemiology and pathogenesis. Expert Review of Anti-infective Therapy. 2015 May 4;13(5):567–73.

[4] Almasaudi SB. Acinetobacter spp. as nosocomial pathogens: Epidemiology and resistance features. Saudi Journal of Biological Sciences. 2018 Mar;25(3):586–96.Sajad N, Al-Saedi F, Israa Burhan, Mojtaba Mohammadzadeh Vazifeh. Virulence Determinants of Acinetobacter baumannii Isolated from Different Infections in Baghdad Hospitals. AJPS. 2024 Jul 9;24(3):264–73.

[5] Hussein NH, Al-Mathkhury HJF. Imipenem-Resistant Acinetobacter baumannii isolated from patients and hospitals environment in Baghdad. 2013;54.

[6] Ahmad NH, Mohammad GA. Identification of Acinetobacter baumannii and Determination of MDR and XDR Strains. Baghdad SciJ. 2020 Sep 1;17(3):0726.

[7] Van Der Woude MW, Bäumler AJ. Phase and Antigenic Variation in Bacteria. Clin Microbiol Rev. 2004 Jul;17(3):581–611.

[8] Brazda V, Fojta M, Bowater RP. Structures and stability of simple DNA repeats from bacteria. Biochemical Journal. 2020 Jan 31;477(2):325–39.

[9] Bayliss CD, Dixon KM, Moxon ER. Simple sequence repeats (microsatellites): mutational mechanisms and contributions to bacterial pathogenesis. A meeting review. FEMS Immunology & Medical Microbiology. 2004 Jan;40(1):11–9.

[10] Mrázek J, Guo X, Shah A. Simple sequence repeats in prokaryotic genomes. Proc Natl Acad Sci USA.

2007 May 15;104(20):8472–7.

[11] Moxon R, Bayliss C, Hood D. Bacterial Contingency Loci: The Role of Simple Sequence DNA Repeats in Bacterial Adaptation. Vol. 40, Annual Review of Genetics. Annual Reviews; 2006. p. 307–33.

[12] Tipton KA, Dimitrova D, Rather PN. Phase-Variable Control of Multiple Phenotypes in Acinetobacter baumannii Strain AB5075. J Bacteriol. 2015 Aug;197(15):2593–9.

[13] Tipton KA, Rather PN. An ompR-envZ Two-Component System Ortholog Regulates Phase Variation, Osmotic Tolerance, Motility, and Virulence in Acinetobacter baumannii Strain AB5075. J Bacteriol [Internet]. 2017 Feb [cited 2024 Nov 4];199(3). Available from: https://journals.asm.org/doi/10.1128/JB.00705-16

[14] Aidley J, Wanford JJ, Green LR, Sheppard SK, Bayliss CD. PhasomeIt: an 'omics' approach to cataloguing the potential breadth of phase variation in the genus Campylobacter. Microbial Genomics [Internet]. 2018 Nov 1 [cited 2024 Nov 1];4(11). Available from: https://www.microbiologyresearch.org/content/journal/mgen/10.1099/mgen.0.000228

[15] Siena E, D'Aurizio R, Riley D, Tettelin H,

Guidotti S, Torricelli G, Moxon ER, Medini D. In-silico prediction and deep-DNA sequencing validation indicate phase variation in 115 Neisseria meningitidis genes. BMC Genomics. 2016 Dec;17(1):843.

[16]Zhu L, Yan Z, Zhang Z, Zhou Q, Zhou J, Wakeland EK, Fang X, Xuan Z, Shen D, Li QZ. Complete Genome Analysis of Three Acinetobacter baumannii Clinical Isolates in China for Insight into the Diversification of Drug Resistance Elements. PLoS ONE. 2013 Jun 24;8(6):e66584.

[17]Farrugia DN, Elbourne LDH, Hassan KA, Eijkelkamp BA, Tetu SG, Brown MH, Shah BS, Peleg AY, Mabbutt BC, Paulsen IT. The Complete Genome and Phenome of a Community-Acquired Acinetobacter baumannii. PLoS ONE. 2013 Mar 19;8(3):e58628.

[18]Vallenet D, Nordmann P, Barbe V, Poirel L, Mangenot S, Bataille E, Dossat C, Gas S, Kreimeyer A, Lenoble P, Oztas S, Poulain J, Segurens B, Robert C, Abergel C, Claverie JM, Raoult D, Médigue C,Weissenbach J, Cruveiller S. Comparative Analysis of Acinetobacters: Three Genomes for Three Lifestyles. PLoS ONE. 2008 Mar 19;3(3):e1805.

[19]Smith MG, Gianoulis TA, Pukatzki S, Mekalanos JJ, Ornston LN, Gerstein M, Snyder M. New insights into Acinetobacter baumannii pathogenesis revealed by high-density pyrosequencing and transposon mutagenesis. Genes Dev. 2007 Mar 1;21(5):601–14.

[20]Zhou Hua, Zhang Tongwu, Yu Dongliang, Pi Borui, Yang Qing, Zhou Jianying, Hu Songnian, Yu Yunsong. Genomic Analysis of the Multidrug-Resistant Acinetobacter baumannii Strain MDR-ZJ06 Widely Spread in China. Antimicrobial Agents and Chemotherapy. 2011 Sep 15;55(10):4506–12.

[21]Chen CC, Lin YC, Sheng WH, Chen YC, Chang SC, Hsia KC, Liao MH, Li SY. Genome Sequence of a Dominant, Multidrug-Resistant Acinetobacter baumannii Strain, TCDC-AB0715. J Bacteriol. 2011 May;193(9):2361–2.

[22]Liou ML, Liu CC, Lu CW, Hsieh MF, Chang KC, Kuo HY, Lee CC, Chang CT, Yang CY, Tang CY. Genome Sequence of Acinetobacter baumannii TYTH-1. J Bacteriol. 2012 Dec 15;194(24):6974–6974.

[23]Iacono M, Villa L, Fortini D, Bordoni R, Imperi F, Bonnal RJP, Sicheritz-Ponten T, De Bellis G, Visca P, Cassone A, Carattoli A. Whole-Genome Pyrosequencing of an Epidemic Multidrug-Resistant Acinetobacter baumannii Strain Belonging to the European Clone II Group. Antimicrob Agents Chemother. 2008 Jul;52(7):2616–25.

[24]Adams MD, Goglin K, Molyneaux N, Hujer KM, Lavender H, Jamison JJ, MacDonald IJ, Martin KM, Russo T, Campagnari AA, Hujer AM, Bonomo RA, Gill SR. Comparative Genome Sequence Analysis of Multidrug-Resistant Acinetobacter baumannii. J Bacteriol. 2008 Dec 15;190(24):8053–64.

[25]Balázs B, Tóth Z, Nagy JB, Majoros L, Tóth Á, Kardos G. Faecal Carriage of Carbapenem-Resistant Acinetobacter baumannii: Comparison to Clinical Isolates from the Same Period (2017–2019). Pathogens. 2022 Sep 2;11(9):1003.

[26]Nucleotide [Internet]). Nucleotide [Internet]. Bethesda, MD: National Library of Medicine (US), National Center for Biotechnology Information; [1988] –. Accession No. ACYS00000000.2, Acinetobacter baumannii

6014059, whole genome shotgun sequencing project; [Accessed 13/9/2023]. Available from: https://www.ncbi.nlm.nih.gov/nuccore/ACYS00000000.2.

[27]Ma F, Shen C, Zheng X, Liu Y, Chen H, Zhong L, Liang Y, Liao K, Xia Y, Tian GB, Yang Y. Identification of a Novel Plasmid Carrying mcr-4.3 in an Acinetobacter baumannii Strain in China. Antimicrob Agents Chemother. 2019 Jun;63(6):e00133-19.

[28]Hornsey M, Loman N, Wareham DW, Ellington MJ, Pallen MJ, Turton JF, Underwood A, Gaulton T, Thomas CP, Doumith M, Livermore DM, Woodford N. Whole-genome comparison of two Acinetobacter baumannii isolates from a single patient, where resistance developed during tigecycline therapy. Journal of Antimicrobial Chemotherapy. 2011 Jul 1;66(7):1499–503.

[29]Hammerstrom TG, Beabout K, Clements TP, Saxer G, Shamoo Y. Acinetobacter baumannii Repeatedly Evolves a Hypermutator Phenotype in Response to Tigecycline That Effectively Surveys Evolutionary Trajectories to Resistance. PLoS ONE. 2015 Oct 21;10(10):e0140489.Doughty EL, Liu H, Moran RA, Hua X, Ba X, Guo F, Chen X, Zhang L, Holmes M, Van Schaik W, McNally A, Yu Y. Endemicity and diversification of carbapenem-resistant Acinetobacter baumannii in an intensive care unit. The Lancet Regional Health - Western Pacific. 2023 Aug;37:100780.

[30]Brito BP, Koong J, Wozniak A, Opazo-Capurro A, To J, Garcia P, Hamidian M. Genomic Analysis of Carbapenem-Resistant Acinetobacter baumannii Strains Recovered from Chilean Hospitals Reveals Lineages Specific to South America and Multiple Routes for Acquisition of Antibiotic Resistance Genes. Microbiol Spectr. 2022 Oct 26;10(5):e02463-22.

[31]Chan AP, Sutton G, DePew J, Krishnakumar R, Choi Y, Huang XZ, Beck E, Harkins DM, Kim M, Lesho EP, Nikolich MP, Fouts DE. A novel method of consensus pan-chromosome assembly and large- scale comparative analysis reveal the highly flexible pan-genome of Acinetobacter baumannii. Genome Biol. 2015 Jul 21;16(1):143.

[32]Phillips ZN, Tram G, Seib KL, Atack JM. Phase-variable bacterial loci: how bacteria gamble to maximise fitness in changing environments. Biochemical Society Transactions. 2019 Aug 30;47(4):1131–41.

[33]van Ham SM, van Alphen L, Mooi FR, van Putten JP. Phase variation of H. influenzae fimbriae: transcriptional control of two divergent genes through a variable combined promoter region. Cell. 1993 Jun 18;73(6):1187–96.

[34]Tauseef I, Ali YM, Bayliss CD. Phase Variation of PorA, a Major Outer Membrane Protein, Mediates Escape of Bactericidal Antibodies by Neisseria meningitidis. Infect Immun. 2013 Apr;81(4):1374–80.

[35]Green LR, Al-Rubaiawi AA, Al-Maeni MARM, Harrison OB, Blades M, Oldfield NJ, Turner DPJ, Maiden MCJ, Bayliss CD. Localized Hypermutation is the Major Driver of Meningococcal Genetic Variability during Persistent Asymptomatic Carriage. mBio. 2020 Apr 28;11(2):e03068-19.

[36]Al-Rubaiawi AA. Contribution of phase variation of Opa proteins to persistent carriage and immune evasion of Neisseria meningitides. [Internet]. 2018 [cited 2024 Dec

29]. Available from: https://hdl.handle.net/2381/43068.

[37] Ahmad I, Karah N, Nadeem A, Wai SN, Uhlin BE. Analysis of colony phase variation switch in Acinetobacter baumannii clinical isolates. PLoS ONE. 2019 Jan 4;14(1):e0210082.

[38] Mushtaq F, Nadeem A, Yabrag A, Bala A, Karah N, Zlatkov N, Nyunt Wai S, Uhlin BE, Ahmad I. Colony phase variation switch modulates antimicrobial tolerance and biofilm formation in Acinetobacter baumannii. Microbiol Spectr. 2024 Jan 11;e02956-23.

[39] Tipton KA, Farokhyfar M, Rather PN. Multiple roles for a novel RND-type efflux system in Acinetobacter baumannii AB5075. MicrobiologyOpen. 2017 Apr;6(2):e00418.

[40] Tierney ARP, Chin CY, Weiss DS, Rather PN. A LysR-Type Transcriptional Regulator Controls Multiple Phenotypes in Acinetobacter baumannii. Front Cell Infect Microbiol. 2021 Nov 4;11:778331.

[43] Klughammer J, Dittrich M, Blom J, Mitesser V, Vogel U, Frosch M, Goesmann A, Müller T, Schoen C. Comparative Genome Sequencing Reveals Within-Host Genetic Changes in Neisseria meningitidis during Invasive Disease. PLoS ONE. 2017 Jan 12;12(1):e0169892.Power PM, Sweetman WA, Gallacher NJ, Woodhall MR, Kumar GA, Moxon ER, Hood DW. Simple sequence repeats in Haemophilus influenzae. Infection, Genetics and Evolution. 2009 Mar;9(2):216– 28.

[44] Bayliss CD, Bidmos FA, Anjum A, Manchev VT, Richards RL., Grossier JP, Wooldridge KG, Ketley JM, Barrow PA, Jones MA, Tretyakov MV. Phase variable genes of Campylobacter jejuni exhibit high mutation rates and specific mutational patterns but mutability is not the major determinant of population structure during host colonization. Nucleic Acids Research. 2012 Jul 1;40(13):5876–89.

[45] Zhou K, Aertsen A, Michiels CW. The role of variable DNA tandem repeats in bacterial adaptation.

FEMS Microbiol Rev. 2014 Jan;38(1):119–41.

[46] Munkelt D, Grass G, Nies DH. The Chromosomally Encoded Cation Diffusion Facilitator Proteins DmeF and FieF from Wautersia metallidurans CH34 Are Transporters of Broad Metal Specificity. J Bacteriol. 2004 Dec;186(23):8036–43.

[47] Rumbo C, Tomás M, Fernández Moreira E, Soares NC, Carvajal M, Santillana E, Beceiro A, Romero A, Bou G. The Acinetobacter baumannii Omp33-36 Porin Is a Virulence Factor That Induces Apoptosis and Modulates Autophagy in Human Cells. Infect Immun. 2014 Nov;82(11):4666–80.

[48] Favrot L, Blanchard JS, Vergnolle O. Bacterial GCN5-Related N -Acetyltransferases: From Resistance to Regulation. Biochemistry. 2016 Feb 23;55(7):989–1002.

[49] Alamro M, Bidmos FA, Chan H, Oldfield NJ, Newton E, Bai X, Aidley J, Care R, Mattick C, Turner DPJ, Neal KR, Ala'Aldeen DAA, Feavers I, Borrow R, Bayliss CD. Phase Variation Mediates Reductions in Expression of Surface Proteins during Persistent Meningococcal Carriage. Infect Immun. 2014 Jun;82(6):2472–84.