



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE BIOCÊNCIAS
DEPARTAMENTO DE GENÉTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA

ANA CAROLINA DE OLIVEIRA LUZ

**ANÁLISE ESTRUTURAL E GENÔMICA DO SISTEMA CRISPR/CAS EM
ISOLADOS CLÍNICOS BRASILEIROS DE *PSEUDOMONAS AERUGINOSA***

Recife

2019

ANA CAROLINA DE OLIVEIRA LUZ

**ANÁLISE ESTRUTURAL E GENÔMICA DO SISTEMA CRISPR/CAS EM
ISOLADOS CLÍNICOS BRASILEIROS DE *PSEUDOMONAS AERUGINOSA***

Tese apresentada ao Programa de Pós-Graduação em Genética da Universidade Federal de Pernambuco como parte dos requisitos exigidos para obtenção do título de Doutor em Genética.

Área de concentração: Genética

Orientadora: Dra. Tereza Cristina Leal Balbino

Coorientadora: Dra. Maria Paloma Silva de Barros

Recife

2019

Catálogo na fonte:
Bibliotecária Claudina Queiroz, CRB4/1752

Luz, Ana Carolina de Oliveira

Análise estrutural e genômica do sistema CRISPR/Cas em isolados clínicos brasileiros de *Pseudomonas aeruginosa* / Ana Carolina de Oliveira Luz - 2019.

134 folhas: il., fig., tab.

Orientadora: Tereza Cristina Leal Balbino

Coorientadora: Maria Paloma Silva de Barros

Tese (doutorado) – Universidade Federal de Pernambuco. Centro de Biociências. Programa de Pós-Graduação em Genética. Recife, 2019.

Inclui referências e anexos.

1. *Pseudomonas aeruginosa* 2. Bacteriófagos 3. Imunidade

I. Balbino, Tereza Cristina Leal (orient.) II. Barros, Maria Paloma Silva de (coorient.) III. Título

579.332

CDD (22.ed.)

UFPE/CB-2019-180

ANA CAROLINA DE OLIVEIRA LUZ

**ANÁLISE ESTRUTURAL E GENÔMICA DO SISTEMA CRISPR/CAS EM
ISOLADOS CLÍNICOS BRASILEIROS DE *PSEUDOMONAS AERUGINOSA***

Tese apresentada ao Programa de Pós-Graduação em
Genética da Universidade Federal de Pernambuco como
parte dos requisitos exigidos para obtenção do título de
Doutor em Genética.

Aprovada em: 27/02/2019

BANCA EXAMINADORA

Dra. Tereza Cristina Leal Balbino (Orientadora)
Centro de Pesquisas Aggeu Magalhães

Profº. Dr. Tercilio Calsa Junior (Examinador Interno)
Universidade Federal de Pernambuco

Dr. Gabriel da Luz Wallau (Examinador externo)
Centro de Pesquisas Aggeu Magalhães

Dra. Isabelle Freire Tabosa Viana (Examinadora externa)
Centro de Pesquisas Aggeu Magalhães

Profª. Dra. Marcia Maria de Camargo Moraes (Examinadora externa)
Universidade de Pernambuco

Dedico este trabalho a todos que trabalham com ciência no Brasil. Nunca é fácil, mas é sempre necessário.

AGRADECIMENTOS

Agradeço aos órgãos de fomento CAPES, Facepe e CNPq, que permitiram a realização deste trabalho através da concessão de bolsa e auxílios financeiros.

Agradeço à minha orientadora e minha coorientadora por todos os anos de aprendizado e amizade. Os conhecimentos adquiridos nestes anos moldaram a pessoa que sou hoje, tanto cientificamente quanto pessoalmente. À minha orientadora, faço um agradecimento especial. Sua amizade, apoio, confiança e companhia foram (e são) extremamente importantes para mim.

Agradeço à Julia, como aluna de Iniciação Científica e, agora, como mestrandia, pela amizade, companhia, conversas e pelo trabalho que desenvolvemos juntas. Este estudo não seria possível sem sua ajuda.

Agradeço ao Programa de Pós-Graduação em Genética pela oportunidade e pelas experiências obtidas.

Agradeço aos técnicos e demais pesquisadores do Departamento de Microbiologia pelo auxílio nas atividades ao longo dos anos.

Agradeço às técnicas do Núcleo de Plataformas Tecnológicas (NPT) do Instituto de Pesquisas Aggeu Magalhães (IAM), pelo serviço e pelo uso de sua infraestrutura.

Agradeço ao Prof. Dr. Valdir, Wilson, Bandeira, e demais integrantes do LABBE, pela parceria e pela ajuda.

Por fim, porém não menos importantes, agradeço ao meu namorado, meus amigos, e meus familiares, pelo apoio e incentivo. Pela comemoração nos momentos alegres e pelo consolo nos momentos difíceis.

“Estou ficando ligeiramente louco. Finalmente aconteceu.” (QUEEN, 1991)

RESUMO

Pseudomonas aeruginosa é uma bactéria oportunista, causa frequente de infecções hospitalares, e frequentemente resistente a diferentes classes de antimicrobianos. O sistema CRISPR/Cas é uma maquinaria de imunidade adaptativa contra elementos genéticos móveis (MGEs). É composto por dois elementos: loco CRISPR, regiões de DNA semi-repetitivo, onde as repetições são separadas por espaçadores (sequências derivadas de MGEs), o que confere a característica adaptativa ao sistema; e por genes *cas*, codificantes dos elementos efetores desta maquinaria. Com o intuito de aplicar os conhecimentos obtidos através do estudo do sistema CRISPR/Cas de *P. aeruginosa* para auxiliar no desenvolvimento de novas estratégias terapêuticas contra este microrganismo, o presente trabalho teve por objetivo identificar as características deste sistema em isolados clínicos desta bactéria. A identificação dos tipos I-F e I-E, e a análise dos locos CRISPR do sistema tipo I-F, foram realizadas através de Reação em Cadeia da Polimerase e sequenciamento. Isolados selecionados tiveram seus genomas sequenciados e analisados para determinar características relacionadas ao sistema em questão. O subtipo I-F é o mais frequente entre os isolados estudados. Foram encontrados dois isolados comportando dois tipos de sistema concomitantemente, característica incomum. Os espaçadores encontrados nos locos CRISPR do subtipo I-F são relacionados a bacteriófagos, plasmídeos e ilhas genômicas, corroborando com a relação deste sistema com MGEs. A organização dos espaçadores demonstrou padrões de incorporação iguais entre isolados diferentes, porém, o sistema não parece ser eficaz como método de tipagem molecular para esta espécie bacteriana. A análise genômica evidenciou a presença de genes anti-CRISPR, de fagos, e de espaçadores contra estes fagos em um mesmo genoma. Foi estabelecido o contexto genético dos tipos de sistemas e de um loco CRISPR órfão. Nenhuma relação genética direta entre os isolados CRISPR/Cas positivos foi encontrada. Os resultados sugerem inatividade deste sistema em relação à imunidade, podendo sua função estar ligada à regulação da expressão gênica de produtos bacterianos e virais. Os resultados também nos permitiram conhecer o panorama do sistema CRISPR/Cas em isolados clínicos brasileiros de *P. aeruginosa*.

Palavras-chave: Bacteriófagos. Imunidade. Anti-CRISPR. CRISPR/Cas.

ABSTRACT

Pseudomonas aeruginosa is an opportunistic bacterium, frequent cause of hospital infections, and frequently resistant to different classes of antibiotics. The CRISPR/Cas system is an adaptive immunity machinery against mobile genetic elements (MGEs). It is composed by two elements: CRISPR locus, regions of semi-repetitive DNA, where the repeats are interspaced by spacers (MGE derived sequences), which confer the adaptive feature to the system; and by *cas* genes, coding for the effector elements of this machinery. In order to apply the knowledge obtained from studying the *P. aeruginosa*'s CRISPR/Cas system to developing new therapeutic strategies against this microorganism, the present work aimed to identify the characteristics of this system in this bacterium's clinical isolates. The identification of I-F and I-E types and the analysis of system's I-F CRISPR loci, were performed through Polymerase Chain Reaction and sequencing. Selected isolates had their genomes sequenced and analyzed to determined characteristics related to the system in question. The subtype I-F was the most frequent between the studied isolates. Two isolates were found carrying two types of the system simultaneously, a rare feature. The spacers found in subtype I-F CRISPR loci are related to phages, plasmids and genomic islands, supporting this system's relationship with MGEs. Spacer arrays showed equal incorporation patterns between different isolates, however, the system does not appear to be effective as a molecular typing method in this bacterial species. Genomic analysis pointed out the presence of anti-CRISPR genes, phages, and spacers against these phages in the same genome. Genetic backgrounds were established for system's different types and an orphan CRISPR locus. No direct relationship between CRISPR/Cas positive isolates was found. Results suggest this system's inactivity regarding immunity, and its function could be related to regulation of gene expression of bacterial and viral products. Results also allowed us to know the scenery of CRISPR/Cas system in Brazilian clinical isolates of *P. aeruginosa*.

Key words: Bacteriophages, immunity, anti-CRISPR, CRISPR/Cas

LISTA DE ILUSTRAÇÕES

Revisão da Literatura

Figura 1 –	Características, atividades metabólicas, e aplicações biotecnológicas de espécies do gênero <i>Pseudomonas</i> spp.....	19
Figura 2 –	Listas de patógenos das quais <i>P. aeruginosa</i> faz parte.....	23
Figura 3 –	Ciclo lítico do bacteriófago e, em vermelho, momentos do ciclo em que diferentes sistemas bacterianos de combate à infecção viral atuam.....	27
Quadro 1 –	Número de genomas analisados e contendo locos CRISPR.....	28
Figura 4 –	Representação esquemática da estrutura genética do sistema CRISPR/Cas.....	31
Figura 5 –	Mecanismo geral de ação do sistema CRISPR/Cas.....	33
Quadro 2 –	Sequências consenso de PAM em alguns organismos com tipos diferentes de sistema CRISPR/Cas.....	35
Figura 6 –	Pareamento de bases entre espaçador e sua molécula alvo, destacando o PAM e a sequência seed, regiões de importância para o reconhecimento de moléculas exógenas.....	36
Figura 7 –	Sistemas CRISPR/Cas de classe 1.....	38
Figura 8 –	Sistemas CRISPR/Cas de classe 2.....	39
Quadro 3 –	Funções variantes do sistema CRISPR/Cas ou de apenas um dos seus elementos em espécies diversas.....	43
Figura 9 –	Técnica empregada para tipagem molecular através da análise do loco CRISPR.....	45
Quadro 4 –	Espécies bacterianas cuja tipagem através do loco CRISPR está estabelecida.....	46
Figura 10 –	Locos anti-CRISPR de fagos da família Mu-like de <i>P. aeruginosa</i>	47
Quadro 5 –	Famílias de proteína anti-CRISPR.....	48

Figura 11 –	Processos do mecanismo do sistema CRISPR/Cas em que proteínas anti-CRISPR específicas interferem.....	49
-------------	---	----

Artigo I

Fig 1 –	Graphic representations of both CRISPR loci from isolate Pae_12.....	63
Fig 2 –	Spacer arrays of I-F CRISPR loci in Brazilian strains of <i>P. aeruginosa</i>	68

Artigo II

Figure 1 –	Genetic context for CRISPR/Cas in <i>P. aeruginosa</i>	92
Figure 2 –	Phylogenetic tree of 174 genomes of <i>P. aeruginosa</i>	95

LISTA DE TABELAS

Artigo I

Table 1 – Primers utilized for amplifications in the present study.....	58
Table 2 – Characteristics of repetitive elements in CRISPR loci.....	62
Table 3 – New spacers found in CRISPR loci related to type I-F CRISPR/Cas system in brazilian strains of <i>P. aeruginosa</i>	64

Artigo II

Table 1 – Characteristics of <i>Pseudomonas aeruginosa</i> Brazilian isolates.....	87
Table 2 – Prophages in <i>P. aeruginosa</i> Brazilian isolates.....	93
Table 3 – Anti-CRISPR genes search.....	94

LISTA DE ABREVIATURAS E SIGLAS

AMP	Peptídeos antimicrobianos (<u>A</u> ntimicrobial <u>p</u> eptides)
Cascade	Complexo de defesa antiviral associado ao CRISPR (<u>C</u> RISPR <u>a</u> ssociated <u>c</u> omplex for <u>a</u> ntiviral <u>d</u> efense)
CDC	Centro de Controle e Prevenção de Doenças (<u>C</u> enter for <u>D</u> isease <u>C</u> ontrol and <u>P</u> revention)
CRISPR	Repetições palindrômicas curtas agrupadas e inter espaçadas (<u>C</u> lustered <u>R</u> egularly <u>I</u> nterspaced <u>S</u> hort <u>P</u> alindromic <u>R</u> epeats)
crRNA	RNA CRISPR
crRNP	Complexo proteico-RNA CRISPR
DR	Repetição direta
DG	Repetição degenerada
ESBL	β -lactamase de espectro estendido (<u>E</u> xtended- <u>S</u> pectrum β - <u>L</u> actamase)
ESKAPE	Acrônimo englobando o nome de seis patógenos associados à resistência antimicrobiana: <u>E</u> nterococcus <i>faecium</i> , <u>S</u> taphylococcus <i>aureus</i> , <u>K</u> lebsiella <i>pneumoniae</i> , <u>P</u> seudomonas <i>aeruginosa</i> , <u>E</u> nterobacter <i>cloacae</i> .
ICE	Elementos integrativos e conjugativos (<u>I</u> ntegrative/ <u>c</u> onjugative <u>E</u> lements)
Mb	Megapares de bases (<u>M</u> egabase <u>p</u> airs)
MGE	Elementos genéticos móveis (<u>M</u> obile <u>G</u> enetic <u>E</u> lements)
MLST	Tipagem de sequências multilocos (<u>M</u> ulti <u>l</u> ocus <u>S</u> equences <u>T</u> yping)

MLVA	Análise de múltiplos locos VNTR (<i><u>M</u>ultiple-<u>l</u>ocus <u>V</u>NTR <u>A</u>nalysis</i>)
NUC	Lóbulo de Nuclease
OMS	<u>O</u> rganização <u>M</u> undial de <u>S</u> aúde
pb	Pares de bases
piRNA	RNA <i>piwi</i>
PFGE	Eletroforese em gel de campo pulsátil (<i><u>P</u>ulsed-<u>F</u>ield <u>G</u>el <u>E</u>lectrophoresis</i>)
pré-crRNA	Precursor de RNA CRISPR
R-M	Restrição e Modificação
REC	Lóbulo de Reconhecimento
RFLP	Polimorfismo de tamanho de fragmentos de restrição (<i><u>R</u>estriction <u>F</u>ragment <u>L</u>enght <u>P</u>olymorphism</i>)
ST	Tipo de sequência (<i><u>S</u>equence <u>T</u>ype</i>)
tracrRNA	Transativador de RNA CRISPR
VNTR	Repetições em tandem de número variável (<i><u>V</u>ariable <u>N</u>umber <u>T</u>andem <u>R</u>epeats</i>)

SUMÁRIO

1	INTRODUÇÃO.....	15
2	REVISÃO DA LITERATURA.....	18
2.1	<i>PSEUDOMONAS AERUGINOSA</i>	18
2.1.1	<i>Pseudomonas aeruginosa</i> e seu sucesso como patógeno.....	21
2.1.2	Métodos alternativos de tratamento.....	23
2.2	MECANISMOS DE PROTEÇÃO BACTERIANA CONTRA FAGOS E OUTROS MGES.....	25
2.3	SISTEMA CRISPR/CAS.....	27
2.3.1	Mecanismo de ação do sistema CRISPR/Cas.....	32
2.3.2	Classificação do sistema CRISPR/Cas.....	37
2.3.3	Outras funções do sistema CRISPR/Cas.....	43
2.3.4	Tipagem molecular através do sistema CRISPR/Cas.....	44
2.4	A RESPOSTA DOS FAGOS: PROTEÍNAS ANTI-CRISPR.....	46
2.5	<i>PSEUDOMONAS AERUGINOSA</i> E O SISTEMA CRISPR/CAS.....	50
3	ARTIGO I - INVESTIGATION OF FREQUENCY, LOCI'S ARCHITECTURE AND SPACER DIVERSITY OF CRISPR/CAS IN BRAZILIAN CLINICAL STRAINS OF <i>PSEUDOMONAS AERUGINOSA</i>.....	52
4	ARTIGO II - GENOMIC ANALYSIS OF <i>PSEUDOMONAS AERUGINOSA</i> HARBORING CRISPR/CAS SYSTEMS SUGGESTS NO IMMUNITY FUNCTION.....	82
5	DISCUSSÃO GERAL.....	105
6	CONCLUSÕES.....	111
	REFERÊNCIAS.....	112
	ANEXO A - NORMAS PARA SUBMISSÃO – ARTIGO I.....	118
	ANEXO B - NORMAS PARA SUBMISSÃO – ARTIGO II.....	130

1 INTRODUÇÃO

Pseudomonas aeruginosa é um bacilo Gram negativo, não fermentador, conhecido por ser um microrganismo versátil, presente em hospedeiros de diferentes reinos. Capaz de causar infecções em plantas, humanos e demais animais, é considerado um patógeno oportunista e uma das principais causas de infecções hospitalares ou relacionadas a ambientes de assistência à saúde.

Infecções por *P. aeruginosa*, em geral, são de difícil tratamento devido à alguns fatores, como sua resistência intrínseca à determinados medicamentos antimicrobianos, sua capacidade de adquirir novos mecanismos de resistência, e sua notável habilidade em formar biofilmes.

Com os objetivos de promover e guiar o desenvolvimento de novos antimicrobianos ou novos métodos terapêuticos, a Organização Mundial de Saúde (OMS), em fevereiro de 2017, elaborou uma lista contendo famílias bacterianas que, atualmente, representam as maiores ameaças à saúde humana. *Pseudomonas aeruginosa* configura entre os microrganismos considerados “Prioridade Crítica”, devido à resistência aos antimicrobianos da classe dos carbapenêmicos. A busca por alternativas terapêuticas para debelar infecções por microrganismos deste tipo tem crescido vertiginosamente. Uma estratégia em desenvolvimento nos últimos anos é a utilização de bacteriófagos, que atacariam especificamente as células bacterianas e as destruiriam, numa técnica conhecida como fagoterapia (*phage therapy*). No entanto, para desenvolver e estabelecer tal estratégia com segurança, ainda é necessário compreender melhor como as bactérias interagem com os bacteriófagos.

Um dos mecanismos de defesa de procariotos e arqueias contra elementos genéticos móveis (MGEs), tais como bacteriófagos, plasmídeos e transposons, é um sistema de imunidade adaptativa chamado CRISPR/Cas.

O sistema CRISPR/Cas é composto por dois elementos principais: o loco CRISPR (acrônimo para *Clustered Regularly Interspaced Short Palindromic Repeats*), formado por uma família de DNA repetitivo que confere a característica adaptativa ao sistema; e pelas proteínas Cas (*CRISPR-associated*), os elementos efetores desta maquinaria. Este sistema e seus elementos são, em geral, altamente variáveis entre os diferentes genomas em termos de ocorrência, composição gênica, sequências, número e tamanho.

O estudo do sistema CRISPR/Cas e seus elementos, além de promover conhecimento acerca da defesa bacteriana contra MGEs, também proporciona outras aplicações, como tipagem molecular baseada na hipervariabilidade de sequências espaçadoras; controle da disseminação de elementos genéticos móveis; e estabelecimento do sistema de defesa natural da bactéria como uma ferramenta para edição de genoma e controle transcricional.

Até o momento, poucos estudos foram feitos com o intuito de esclarecer as nuances do sistema CRISPR/Cas em isolados clínicos de *P. aeruginosa*, e apenas um deles foi realizado com isolados brasileiros. Os resultados obtidos poderão ser promissores para, posteriormente, compreender os diversos mecanismos utilizados por este microrganismo no controle da aquisição de MGEs, seu impacto na obtenção de novos genes de resistência presentes em plasmídeos, e a real função do sistema neste microrganismo específico.

O objetivo geral do presente trabalho foi investigar a presença e sequência do sistema CRISPR/Cas em isolados clínicos brasileiros de *Pseudomonas aeruginosa*,

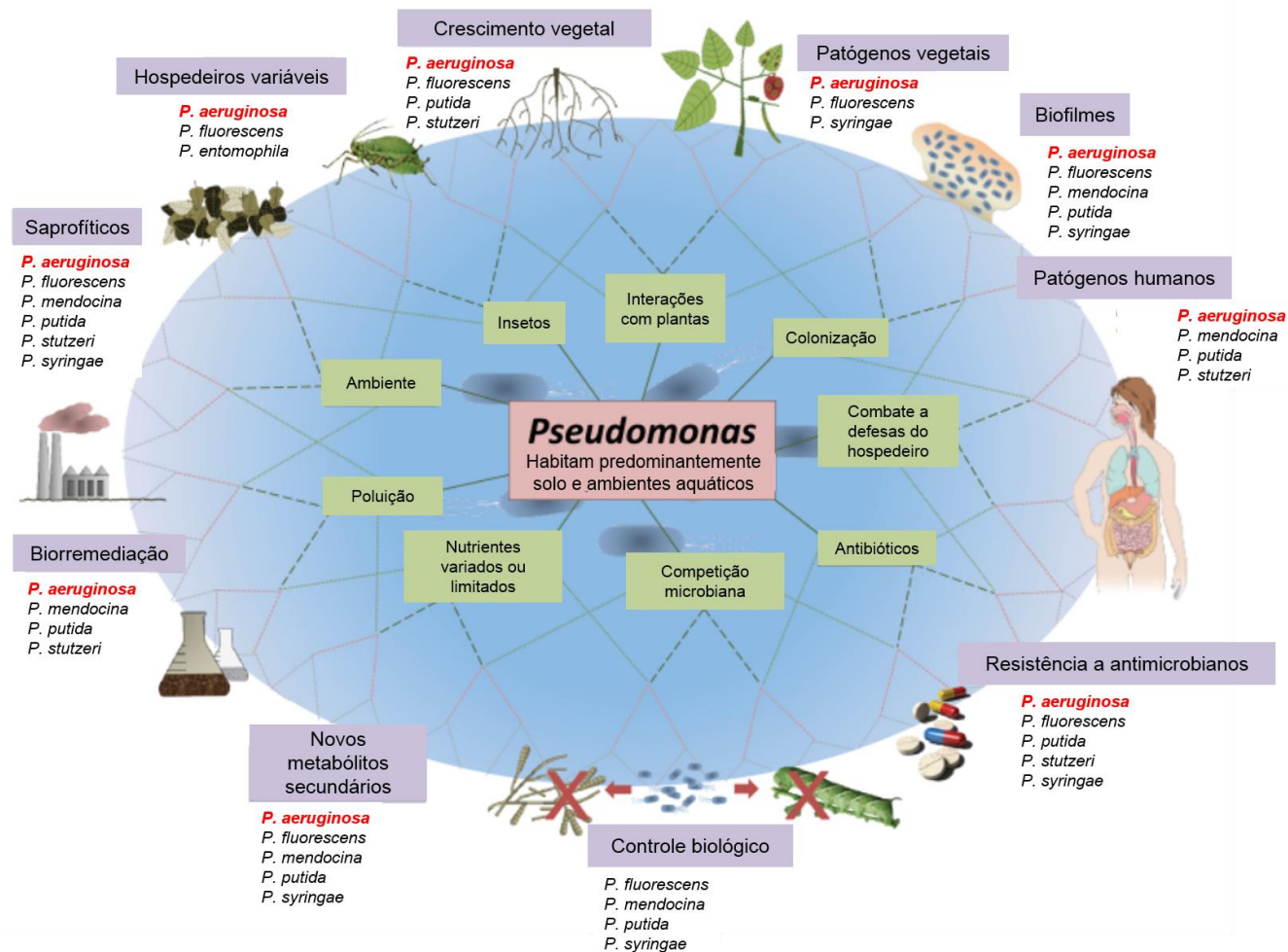
além de analisar os genomas de alguns isolados buscando relações evolutivas e características acerca o funcionamento do sistema. Os objetivos específicos compreendidos no Artigo I foram: investigar a presença dos sistemas tipo I-F e I-E nos isolados clínicos de *P. aeruginosa* brasileiros; identificar os elementos estruturais (DRs, DG e espaçadores) no sistema tipo I-F; investigar a possível origem dos espaçadores; e estabelecer a organização estrutural do loco CRISPR nos isolados clínicos positivos e realizar uma análise comparativa entre eles. Por sua vez, o Artigo II compreendeu os seguintes objetivos: realizar tipagem molecular dos isolados em estudo; investigar os profagos presentes nos genomas de isolados positivos para o sistema CRISPR/Cas; comparar os profagos encontrados aos espaçadores presentes nos locos CRISPR destes isolados; e realizar a filogenia de isolados positivos e negativos para o sistema CRISPR/Cas, para melhor compreender como os isolados brasileiros positivos para o sistema se relacionam com os demais.

2 REVISÃO DA LITERATURA

2.1 *PSEUDOMONAS AERUGINOSA*

Bactérias do gênero *Pseudomonas* demonstram-se versáteis fisiológica e metabolicamente, o que permite que espécies deste gênero sejam encontradas em habitats variados, como solo e ambientes aquáticos (SILBY et al., 2011). São microrganismos Gram negativos, bacilares, não-esporulados, com flagelo polar, aeróbios estritos, não fermentadores, e oxidase-positivos (MADIGAN et al., 2010). Devido a sua versatilidade e sua ampla distribuição na natureza, os microrganismos que compõem o gênero *Pseudomonas* são frequentemente patógenos de plantas e animais, podem estar envolvidos em processos metabólicos naturais, e podem possuir habilidades singulares, como produção de biofilme e resistência intrínseca a antimicrobianos (Figura 1).

Figura 1. Características, atividades metabólicas, e aplicações biotecnológicas de espécies do gênero *Pseudomonas* spp.



Fonte: adaptado de Silby et al. (2011)

Nota: a espécie *P. aeruginosa* encontra-se em vermelho, para destacar sua distribuição em diversas funções.

Pseudomonas aeruginosa encontra-se entre os microrganismos mais importantes de seu gênero, devido à sua multifuncionalidade (Figura 1), e à sua presença frequente em ambientes hospitalares e de assistência à saúde, causando infecções com alta morbidade e mortalidade em pacientes com câncer, fibrose cística, ou demais quadros imunodepressores (GELLATLY; HANCOCK, 2013; LE et al., 2014).

Como patógeno oportunista, *P. aeruginosa* pode levar a quadros infecciosos do trato urinário, trato respiratório, feridas e queimaduras, entre outros sítios. Devido a sua capacidade de sobreviver em superfícies majoritariamente inorgânicas, pode ser isolado de equipamentos médicos, pias e vasos sanitários, que facilitam a dispersão deste microrganismo no ambiente hospitalar (AZAM; KHAN, 2019).

A cepa de referência é a PAO1, cuja sequência genômica foi publicada em 2000 e, desde então, outros isolados clínicos, ambientais e industriais têm sido sequenciados e depositados publicamente. Estes genomas apresentam entre 5,5 e 7 megapares de bases (Mb), 65 a 76% de conteúdo GC, e cerca de 79 regiões de plasticidade genômica (van BELKUM et al., 2015). Estas regiões de plasticidade do genoma permitem à *P. aeruginosa* inserir sequências acessórias em bloco, de forma que seu genoma é composto por um *core* relativamente grande e um genoma acessório passível de constante modificação (VALOT et al., 2015). Em comparação com outros patógenos humanos, *P. aeruginosa* apresenta um dos maiores genomas, com mais de 6 Mb, em comparação com *Escherichia coli* (aproximadamente 4 Mb), *Mycobacterium tuberculosis* (aproximadamente 4 Mb), *Staphylococcus aureus* (Entre 2 e 3 Mb) e *Haemophilus influenzae* (aproximadamente 2 Mb) (AZAM; KHAN, 2019).

Tem-se observado que o genoma de *P. aeruginosa* é composto por um número menor de genes constitutivos altamente conservados, um número um pouco maior de genes e sequências relativamente conservados, e um grande número de genes raros, pouco conservados (PARKINS et al., 2018). Os genes constitutivos de *P. aeruginosa* permitem a elaboração de esquemas de tipagem molecular para identificação de clones ou linhagens, dispersos locais ou internacionalmente. A técnica mais comumente empregada é a de MLST (Tipagem de Sequências Multilocus, ou *Multilocus Sequencing Typing*), na qual são avaliados os polimorfismos de sete genes constitutivos específicos (no caso de *P. aeruginosa*, são os genes *acsA*, *aroE*, *guaA*, *mutL*, *nuoD*, *ppsA* e *trpE*), e ao final, é determinado um perfil único (ST) de acordo com estes polimorfismos (JOLLEY; MAIDEN, 2014). Estudos utilizando esta técnica demonstram uma dispersão não clonal de *P. aeruginosa*, com poucos clones mundialmente distribuídos e de alto risco. Como exemplo, três linhagens amplamente distribuídas e consideradas de alto risco são aquelas relacionadas aos STs 111, 175 e 235 (MULET et al., 2013; van BELKUM et al., 2015; PARKINS et al., 2018).

2.1.1 *P. aeruginosa* e seu sucesso como patógeno

O sucesso das infecções por *P. aeruginosa* se dá, principalmente, pela resistência a várias classes de antimicrobianos e pela formação de biofilme. Ao passo que a resistência a antimicrobianos atua diretamente na evasão da célula bacteriana aos medicamentos utilizados para debelar a infecção, a formação de biofilme, um conjunto de células em uma matriz polimérica extracelular, aumenta a

permanência do patógeno no organismo por dificultar o contato dos antimicrobianos, ou de células do sistema imune do hospedeiro, com as colônias bacterianas (SILBY et al., 2011; AZAM; KHAN, 2019).

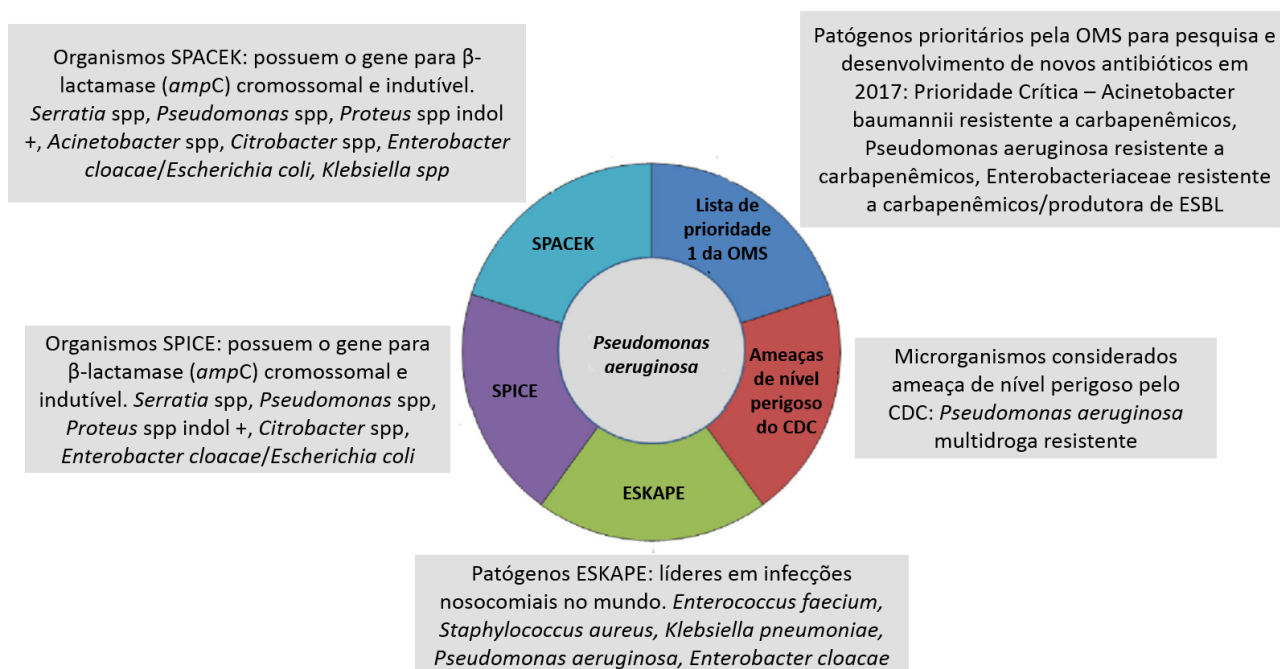
Estes mecanismos de resistência podem ser intrínsecos (diminuição na permeabilidade da membrana externa, expressão de sistemas ativos de efluxo), adquiridos (mutações em genes cromossomais e transferência horizontal de genes) ou adaptativos (fenótipos temporários causados pela regulação gênica diferencial em situações de estresse, tais como a presença de antibióticos, deficiência nutricional, mudança de pH do meio, mudanças de temperatura ou na disponibilidade de oxigênio). Sua facilidade em adquirir novos mecanismos de resistência confere o potencial para tornar-se resistente a todos os antimicrobianos utilizados na clínica atual (MULET et al., 2013; AZAM; KHAN, 2019).

Pseudomonas aeruginosa resistente a um extenso número de drogas antimicrobianas representam um problema grave e crescente em ambientes hospitalares, nos últimos anos (van DER ZEE et al., 2018). Carbapenêmicos são as principais drogas utilizadas para combater infecções por *P. aeruginosa*, porém estima-se que o percentual de cepas resistentes varie entre 10 e 50% na maioria dos países (FENG et al., 2017). Além dos carbapenêmicos, *P. aeruginosa* pode ser resistente aos demais antimicrobianos de escolha, como aminoglicosídeos, fluoroquinolonas e polimixinas, sendo estas últimas utilizadas apenas como último recurso (POTRON et al., 2015).

Em fevereiro de 2017, a OMS estabeleceu uma lista de 12 famílias bacterianas que representam a maior ameaça à saúde humana (OMS, 2017). Esta lista, cujo intuito é de guiar as pesquisas para o desenvolvimento de novos métodos para debelar essas infecções, compreende *P. aeruginosa* resistente a

antimicrobianos carbapenêmicos entre os agentes patogênicos que se encontram em prioridade crítica. Este microrganismo também configura em outras listas de patógenos importantes, como demonstra a Figura 2.

Figura 2. Listas de patógenos das quais *P. aeruginosa* faz parte.



Fonte: adaptado de Azam e Khan (2019)

Legenda: ESBL corresponde à sigla para β -lactamase de espectro estendido (*Extended-spectrum β -lactamase*), CDC corresponde à Centro para Prevenção e Controle de Doenças (*Center for Disease Control and Prevention*).

2.1.2 Métodos alternativos de tratamento

Diante do preocupante aumento da frequência de bactérias patogênicas resistentes a múltiplos antimicrobianos, tem-se buscado, nos últimos anos, o desenvolvimento de estratégias alternativas para debelar, ou amenizar os efeitos de infecções por microrganismos multidroga resistentes. Algumas destas estratégias

são: o uso de anticorpos monoclonais que se ligariam a fatores de virulência dos agentes patogênicos para neutralizá-los; a utilização de probióticos para aumentar a competição por nutrientes entre os microrganismos em questão; produção de nanopartículas lipídicas (lipossomos) que seriam capazes de sequestrar e neutralizar produtos bacterianos; o uso de peptídeos antimicrobianos (AMPs), moléculas integrantes do sistema imune de alguns eucariotos e procariotos; e, por fim, a utilização de bacteriófagos (fagos) específicos, capazes de destruir seletivamente a célula bacteriana (ASIF et al., 2018).

A fagoterapia (*phage therapy*) representa um dos métodos alternativos mais promissores. Antes do advento dos antibióticos, fagos já eram utilizados para o combate de infecções em sociedades da Europa Oriental. Durante a Primeira Guerra Mundial, soldados com disenteria eram frequentemente tratados com um preparado de fagos (DY et al., 2014).

Vantagens da fagoterapia são, em geral, devido às características dos fagos, como o fato de serem agentes exclusivamente bactericidas; sua capacidade de aumentar de número de acordo com o número de células bacterianas, sem necessitar de interferências externas; sua relativa baixa toxicidade para o organismo do hospedeiro; seu impacto mínimo na microbiota do hospedeiro, devido à relação restrita entre fagos e gêneros bacterianos; ausência de interferência em uma antibioticoterapia já existente; e sua melhor penetração em biofilmes bacterianos do que moléculas antimicrobianas. No entanto, esta técnica apresenta desvantagens que ainda necessitam ser melhor estudadas e minimizadas. O fago a ser utilizado deve, obrigatoriamente, ser lítico, ou seja, não pode ser capaz de estabelecer um ciclo lisogênico na bactéria, além de precisar ser efetivo (fagos com baixo potencial de destruir a célula bacteriana não devem ser utilizados). A especificidade de

determinados fagos com alguns gêneros bacterianos, ao passo que confere uma vantagem, também dificulta a utilização desta técnica por diminuir as opções terapêuticas (LOC-CARRILLO; ABEDON, 2011).

A utilização da fagoterapia tem sido bastante estudada para microrganismos que compõem o grupo ESKAPE (Figura 2), já tendo sido reportado algum sucesso no combate a *P. aeruginosa*, *Acinetobacter baumannii* e *Enterococcus faecalis* (KHALIFA et al., 2015; WATERS et al., 2017; ASIF et al., 2018).

2.2 MECANISMOS DE PROTEÇÃO BACTERIANA CONTRA FAGOS E OUTROS MGES

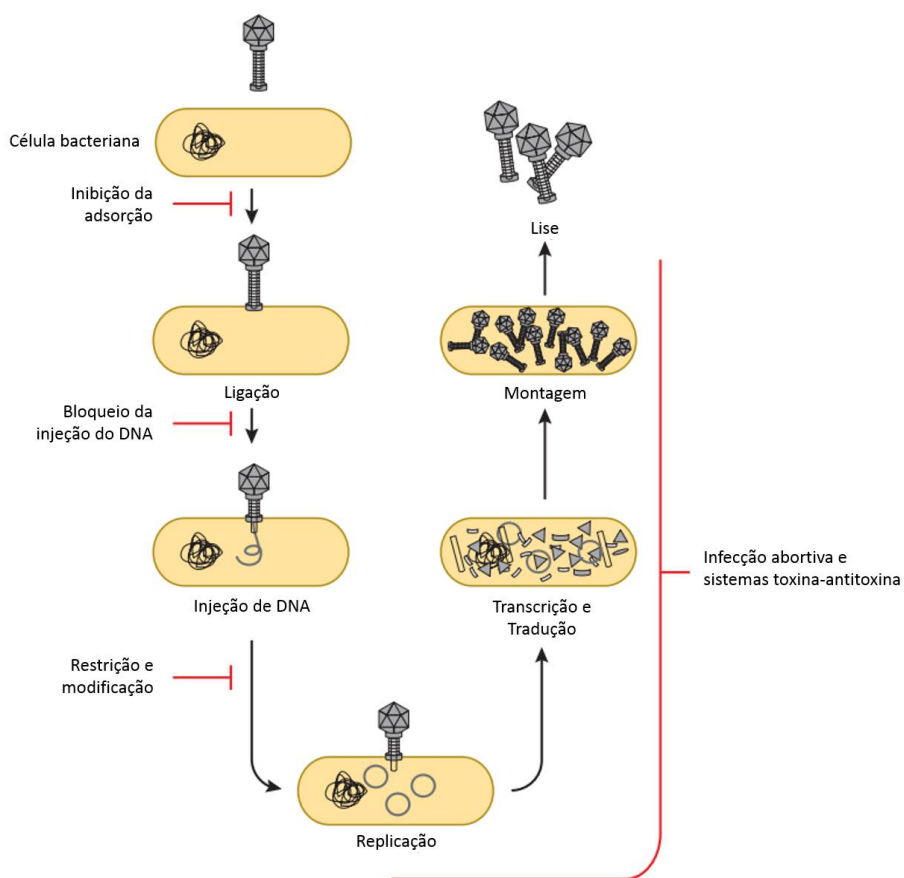
A hipótese da Rainha Vermelha afirma que toda e qualquer adaptação positiva em um organismo causa um impacto negativo no *fitness* no organismo com o qual há uma interação, o que leva a uma série de adaptações e contra adaptações, em uma espécie de “corrida armamentista”. De maneira coerente com esta hipótese, bactérias e seus fagos estão em constante co-evolução. Devido ao grande número de fagos na natureza (cerca de 10 vezes o número de células bacterianas), a pressão seletiva que eles exercem sobre as bactérias, e a capacidade de ambos os organismos desenvolverem mutações rapidamente, o conhecimento de como as bactérias interagem e se protegem de fagos é passível de rápidas mudanças (DY et al., 2014; BORGES et al., 2017).

O contato com bacteriófagos e outros MGEs, como plasmídeos, transposons e elementos integrativos e conjugativos (ICEs), promove a transferência horizontal de genes que podem trazer benefícios para a célula bacteriana hospedeira,

capacitando-a com genes de resistência a antimicrobianos, fatores de virulência, e favorecendo a adaptação bacteriana, principalmente, em patógenos humanos (BARRANGOU, 2015). Porém, na ausência de vantagens seletivas, a aquisição de elementos genéticos custa energia e recursos bacterianos desnecessariamente (RICHTER et al., 2012).

Mecanismos de defesa contra fagos garantem à bactéria sua sobrevivência e multiplicação, impedindo a proliferação viral em vários estágios distintos (BARRANGOU et al., 2007; SOREK et al., 2008). A maioria dos mecanismos de defesa bacteriana acontecem de maneira inespecífica, e incluem modificação e restrição de moléculas de DNA exógenas (R-M); inibição da ligação do bacteriófago à superfície bacteriana; bloqueio da injeção do material genético na célula; sistemas toxina-antitoxina e infecção abortiva (Figura 3) (DY et al., 2014; BURMISTRZ; PYRC, 2015).

Figura 3. Ciclo lítico do bacteriófago e, em vermelho, momentos do ciclo em que diferentes sistemas bacterianos de combate à infecção viral atuam.



Fonte: adaptado de Dy et al. (2014)

2.3 SISTEMA CRISPR/CAS

O sistema CRISPR/Cas consiste em um sistema adaptativo de defesa contra MGEs presente nos genomas da maioria das arqueas e cerca de metade das bactérias (Quadro 1) (BURMISTRZ; PYRC, 2015).

Quadro 1. Número de genomas analisados e contendo locos CRISPR.

	Número de genomas analisados	Número de genomas positivos	%
Arqueas	232	202	87
Bactérias	6.782	3.059	45

Fonte: CRISPRs web server, 2018.

Seu mecanismo assemelha-se ao do RNA piwi (piRNA) em eucariotos, onde uma molécula de RNA é usada como guia para o silenciamento de elementos transponíveis. O sistema CRISPR/Cas, por sua vez, utiliza moléculas de RNA como guia, e em conjunto com um complexo proteico, impede a aquisição de fagos e plasmídeos com os quais a bactéria já interagiu previamente (HIROSE et al., 2014; BARRANGOU, 2015). Este sistema é formado por dois elementos principais: o loco CRISPR e as proteínas Cas (BONDY-DENOMY; DAVIDSON, 2014).

O loco CRISPR (Repetições Palindrômicas Curtas Agrupadas e Regularmente Interespaçadas, do inglês *Clustered Regularly Interspaced Short Palindromic Repeats*) é um loco genético composto por sequências repetitivas separadas entre si por sequências espaçadoras variáveis, únicas, e curtas, provenientes dos MGEs invasores (BURMISTRZ; PYRC, 2015). Foi primeiro descoberto por Ishino e colaboradores em 1987, onde foi descrito como uma região parcialmente repetitiva próxima ao gene *iap* na cepa K12 de *Escherichia coli*, composto por cinco repetições de 29 pares de bases (pb) espaçadas por quatro sequências de 32 pb (ISHINO et al., 1987; RICHTER et al., 2012).

Nos anos posteriores, outros trabalhos reportaram a presença de locos semelhantes em outras espécies, como por exemplo, *Mycobacterium tuberculosis* (GROENEN et al., 1993). Mojica e colaboradores, em 2000 reconheceram que tais

sequências seriam parte de uma mesma família de DNA repetitivo, e Jansen e colaboradores, em 2002, estabeleceram as bases para o reconhecimento destas sequências em diversas espécies e a nomenclatura utilizada atualmente (MOJICA et al., 2000; JANSEN et al., 2002).

As repetições, chamadas Repetições Diretas (DR), são sequências altamente conservadas de 21 a 48 pb. Na maioria das espécies estudadas, é possível observar uma complementariedade de bases nas extremidades das DRs, comumente envolvendo o complemento GTT e AAC, bem como trechos de três ou quatro bases idênticas dentro da sequência, em geral A ou T (JANSEN et al., 2002). Estas estruturas contidas nas DRs conferem ao RNA uma estrutura secundária (BONDY-DENOMY; DAVIDSON, 2014). Estas DRs são idênticas ou quase idênticas dentro de locos do mesmo tipo de sistema CRISPR/Cas na mesma espécie (BURMISTRZ; PYRC, 2015).

As regiões não repetitivas do loco CRISPR são chamadas de Espaçadores e possuem entre 26 e 72 pb. Estes espaçadores, em geral, possuem alta similaridade com sequências virais, plasmidiais ou mesmo bacterianas. No entanto, um grande número de espaçadores não possui similaridade com outros organismos, de forma que sua origem ainda é desconhecida (BURMISTRZ; PYRC, 2015). Em geral, os espaçadores encontrados em uma espécie não são encontrados em outra espécie (JANSEN et al. 2002; BONDY-DENOMY; DAVIDSON, 2014).

Em uma das extremidades do loco CRISPR observa-se uma sequência não codificante, cujo tamanho varia de acordo com a espécie e o tipo de sistema, mas de modo geral, apresenta 200 a 500 pb. Esta sequência é nomeada Sequência Líder, e funciona como promotor para a transcrição do loco e como guia para a incorporação de novos espaçadores. Esta sequência possui uma alta porcentagem de bases AT,

e é encontrada apenas adjacente ao loco CRISPR, estando ausentes em outros locais dos genomas bacterianos. Na extremidade oposta à sequência líder, é encontrada uma DR com mutações pontuais, chamada de Repetição Degenerada (DG), marcando o fim do loco (JANSEN et al., 2002; RICHTER et al., 2012).

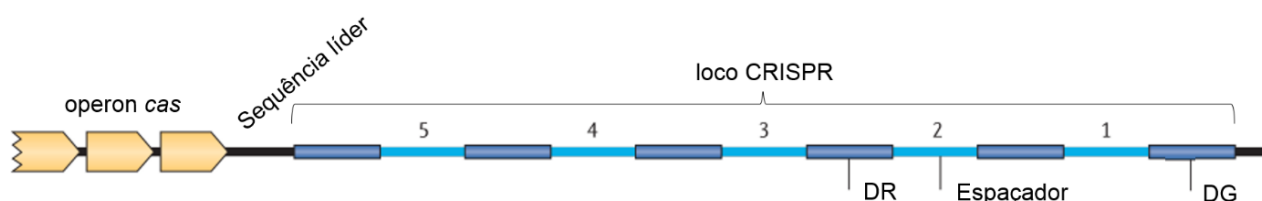
As proteínas Cas (*CRISPR-associated*) são os elementos efetores do sistema CRISPR/Cas (van DER OOST et al., 2014). O conjunto dessas proteínas pode variar entre diferentes espécies, porém as proteínas Cas1 (a qual codifica uma nuclease/integrase) e Cas2 (endorribonuclease), essenciais para o processo de reconhecimento e incorporação de novas sequências espaçadoras ao loco CRISPR, são conservadas e encontradas em quase todos os tipos de sistema (van DER OOST et al., 2009; MAKAROVA et al., 2011b; KOONIN et al., 2017).

Apesar do grande número de proteínas Cas conhecidas, elas podem ser distribuídas em quatro categorias de acordo com sua função: nucleases/recombinases envolvidas na aquisição de novos espaçadores (por exemplo, Cas1); ribonucleases envolvidas no processamento de RNAs (Cas6); proteínas ligantes de RNA que fazem parte dos complexos efetores (Cas7); e nucleases responsáveis pela degradação das moléculas alvo (por exemplo, Cas3) (van DER OOST et al., 2014).

A estrutura e função de algumas proteínas Cas assemelham-se a transposons auto sintetizáveis conhecidos como casposons, e análises filogenéticas inferem que *cas1*, entre outros genes *cas*, evoluíram de casposons para formar o sistema CRISPR/Cas. Neste cenário evolutivo, a repetição dos locos CRISPR teriam surgido a partir das repetições invertidas terminais que tipicamente flanqueiam transposons (KRUPOVIC et al., 2016).

Os genes *cas* encontram-se adjacentes ao loco CRISPR nos genomas. Geralmente encontram-se localizados à esquerda ou à direita do loco, a poucas bases de distância. Em cepas que possuem mais de um loco CRISPR no mesmo genoma, apenas um conjunto de genes *cas* é necessário, estando este próximo a algum dos locos. Entretanto, caso estes múltiplos locos possuam DRs com sequências distintas, ou seja, representem dois tipos diferentes de sistema numa mesma bactéria, cada tipo possuirá seu próprio conjunto de genes *cas* (JANSEN et al., 2002). A estrutura genética do sistema CRISPR/Cas está representada na Figura 4.

Figura 4. Representação esquemática da estrutura genética do sistema CRISPR/Cas.



Fonte: adaptado de Makarova et al. (2011)

Nota: números de um a cinco representam a ordem cronológica de incorporação dos espaçadores.

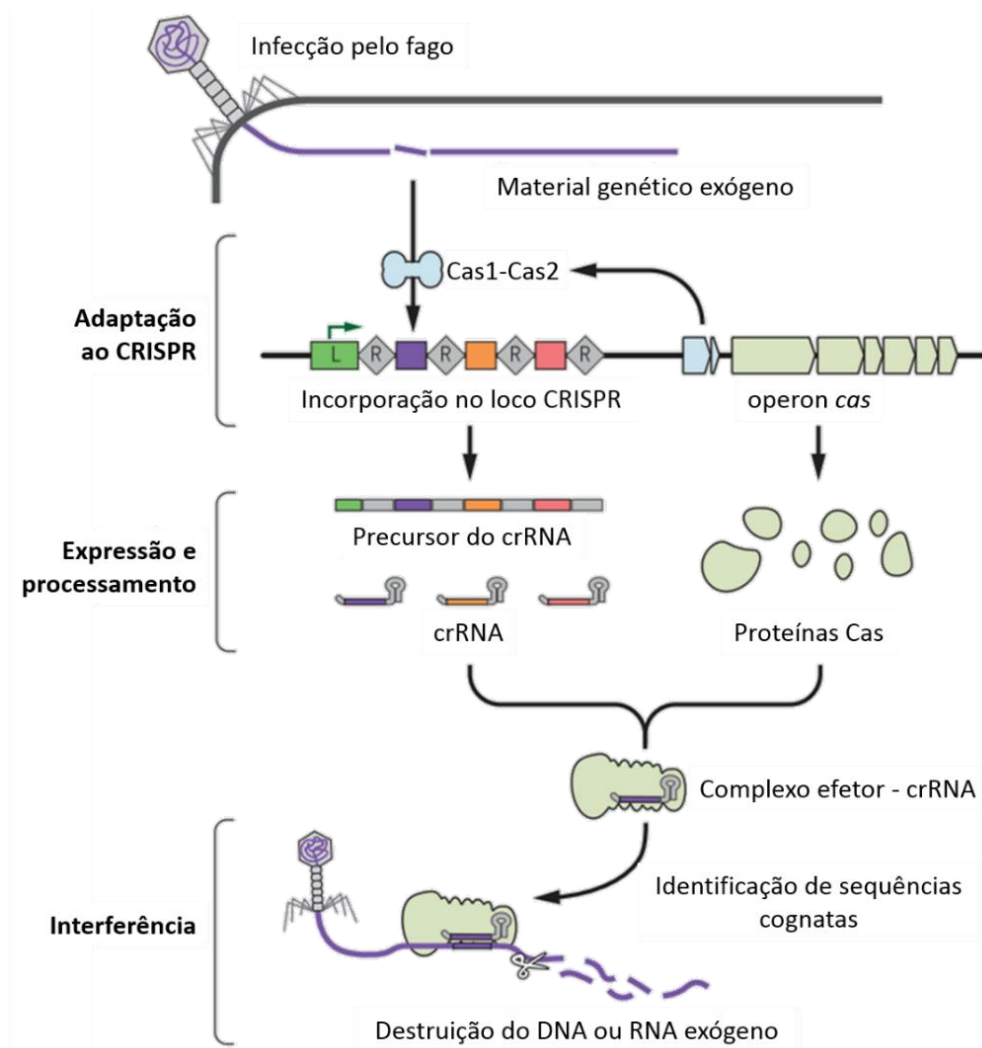
Após o descobrimento destas estruturas, a função do loco CRISPR permanecia desconhecida, enquanto acreditava-se que as proteínas Cas estariam envolvidas no reparo do DNA de arqueas e bactérias (MAKAROVA et al., 2002). O conhecimento de que os espaçadores dos locos CRISPR correspondiam a sequências de DNA provenientes de MGEs foi o motivo para que diversos autores sugerissem que estes locos, de alguma maneira, pudessem prevenir a bactéria contra invasões futuras (BOLOTIN et al., 2005; MOJICA et al., 2005; POURCEL et al., 2005). A proximidade genética entre os genes *cas* e o locos CRISPR levou à

hipótese de que essas proteínas estariam, na realidade, envolvidas com o loco de alguma forma em um sistema de imunidade adaptativa (MAKAROVA et al., 2006). Barrangou e colaboradores confirmaram esta hipótese em estudos com *Streptococcus thermophilus* (BARRANGOU et al., 2007).

2.3.1 Mecanismo de ação do sistema CRISPR/Cas

O mecanismo de ação deste sistema apresenta, em geral, três etapas principais: 1) adaptação, aquisição ou imunização; 2) expressão ou biogênese do RNA-CRISPR (crRNA); 3) interferência ou ação contra uma molécula alvo (Figura 5) (DY et al., 2014).

Figura 5. Mecanismo geral de ação do sistema CRISPR/Cas.



Fonte: adaptado de Jackson et al. (2017)

Legenda: L corresponde à Sequência líder; R, repetição direta; quadrados coloridos, aos espaçadores diferentes.

A etapa de adaptação ocorre quando a bactéria entra em contato com o bacteriófago ou plasmídeo pela primeira vez. Nesta fase inicial do mecanismo de ação do sistema, acontece a fragmentação do DNA exógeno por sistemas R-M naturais da célula bacteriana; o reconhecimento de protoespaçadores através da identificação de uma sequência específica; e incorporação do novo espaçador no

início do loco CRISPR, sempre próximo à sequência líder (van DER OOST et al., 2014; BURMISTRZ; PYRC, 2015).

Um protoespaçador é definido como uma região do DNA exógeno adjacente à uma sequência conhecida como PAM (*Protospacer Adjacent Motif*, ou: motivo adjacente ao protoespaçador). A sequência deste PAM varia de acordo com particularidades do sistema e da espécie em questão (Quadro 2), mas em geral corresponde à 2 a 5 pb adjacentes ao protoespaçador, à montante ou à jusante. A existência do PAM não apenas permite que uma sequência de DNA seja reconhecida como protoespaçadora e incorporada ao loco CRISPR como um novo espaçador, mas na etapa de interferência, permite à bactéria diferenciar sequências exógenas de seus próprios espaçadores, visto que este motivo não é incorporado ao loco CRISPR, evitando, desta forma, autoimunidade e clivagem de seu próprio genoma (LEENAY; BEISEL, 2017; GLEDITZSCH et al., 2019). Algumas espécies não possuem uma sequência PAM específica, mas possuem outros métodos de identificação das sequências exógenas, como por exemplo, degradação de moléculas de DNA que possuam complementariedade de bases com o espaçador, apenas no momento da sua transcrição (LEENAY; BEISEL, 2017).

Quadro 2. Sequências consenso de PAM em diferentes organismos, comportando tipos diferentes de sistema CRISPR/Cas.

Tipo-Subtipo	Organismo	Localização do PAM	Sequência
I-A	<i>Pyrococcus furiosus</i>	5'	YCN
I-B	<i>Clostridium difficile</i>	5'	CCW
I-C	<i>Bacillus halodurans</i>	5'	YYC
I-E	<i>Cyanothece</i> sp.	5'	AWG
I-F	<i>Pseudomonas aeruginosa</i>	5'	CC
III-B	<i>Pyrococcus furiosus</i>	3'	MMA
II-A	<i>Streptococcus pyogenes</i>	3'	NGG
II-A	<i>Streptococcus thermophilus</i>	3'	NNAGAA
II-A	<i>Staphylococcus aureus</i>	3'	NNGRRT
II-C	<i>Neisseria meningitidis</i>	3'	NNNNGWWT

Fonte: adaptado de Leenay e Beisel (2017)

Legenda: na coluna Sequência, onde lê-se M, corresponde à uma das bases A ou C; onde lê-se N, corresponde a uma das quatro bases do DNA; onde lê-se W, corresponde a uma das bases A ou T; e onde lê-se Y, corresponde a uma das bases C ou T.

Apesar de a adaptação ser a etapa mais desconhecida de todo o processo de funcionamento do sistema, sabe-se que as proteínas Cas1 e Cas2 estão intimamente envolvidas e são fundamentais neste momento (DY et al., 2014). Proteínas que não fazem parte do conjunto das Cas também podem estar implicadas nesta fase, como RuvC e RecBCD, envolvidas no reparo e recombinação de DNA (van DER OOST et al., 2014).

A etapa de expressão, ou biogênese do crRNA, consiste na transcrição do loco CRISPR em uma grande molécula precursora, conhecida como pré-RNA-CRISPR (pré-crRNA), e posterior clivagem desta molécula em diversos pequenos

crRNAs, cada um contendo partes da DR, formando uma estrutura secundária, e um espaçador (van DER OOST et al., 2014; BURMISTRZ; PYRC, 2015).

É na etapa de interferência que o complexo proteico efetor associado ao crRNA (crRNP) reconhece a molécula de DNA exógeno através do pareamento de bases entre o fragmento invasor e o complexo efetor. Uma vez feito o pareamento, ocorre uma mudança conformacional no crRNP, cuja estrutura recruta uma nuclease para que ocorra a clivagem do material genético invasor (van DER OOST et al., 2014).

Além do reconhecimento do PAM, para que ocorra a degradação do MGE, é necessário que ocorra o pareamento de bases com precisão em uma determinada região, dentro do espaçador, conhecida como *seed*. A sequência *seed* está envolvida no reconhecimento e na ligação do crRNP à molécula alvo. Em *Haloferax volcanii* (Figura 6), por exemplo, a sequência *seed* corresponde às 10 bases do espaçador, adjacentes ao PAM, porém a sexta base não necessita apresentar um pareamento perfeito (MAIER et al., 2015; GORSKI et al., 2017).

Figura 6. Pareamento de bases entre espaçador e sua molécula alvo, destacando o PAM e a sequência *seed*, regiões de importância para o reconhecimento de moléculas exógenas.



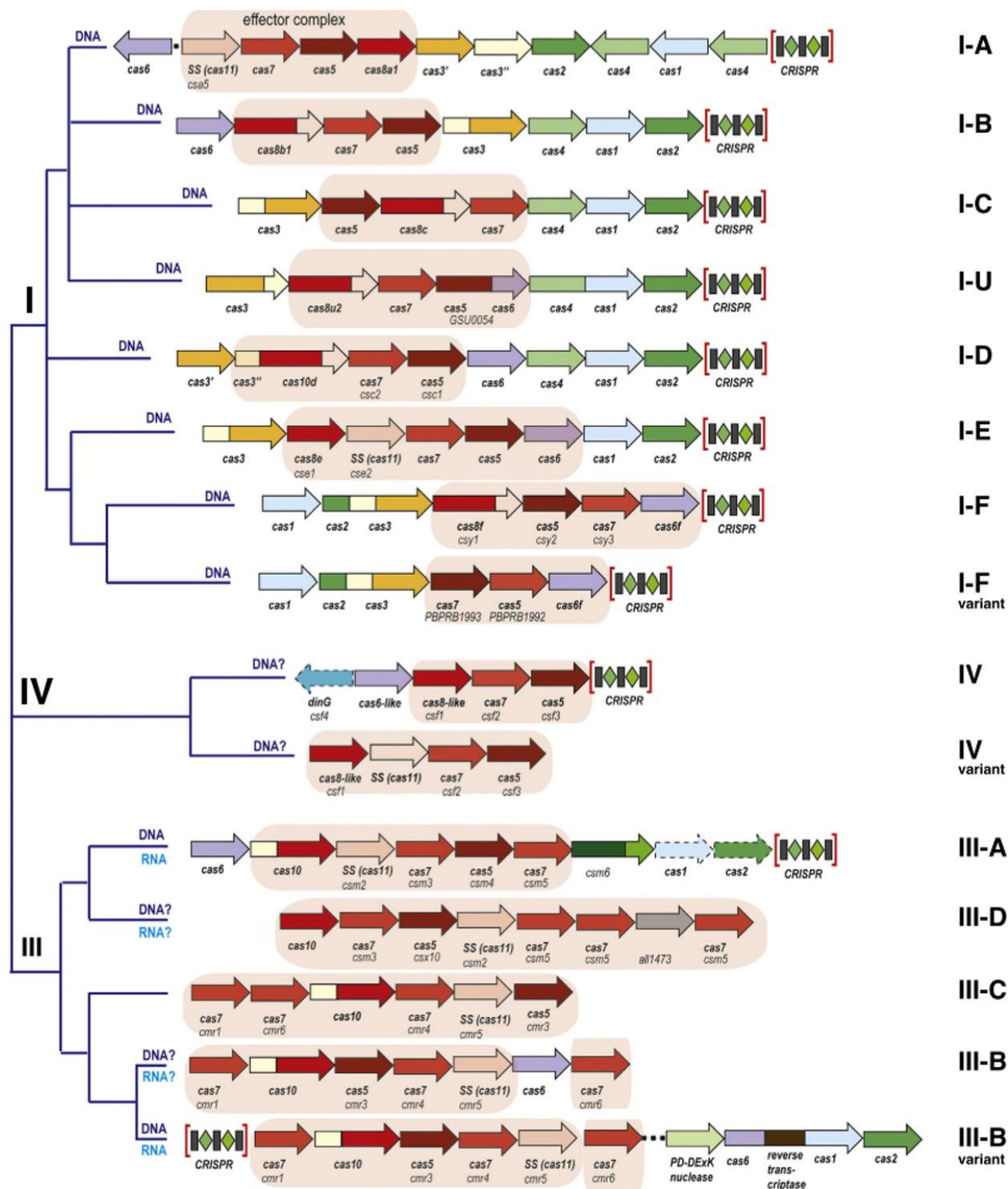
Fonte: adaptado de Maier et al. (2015)

2.3.2 Classificação do sistema CRISPR/Cas

A análise comparativa do sistema CRISPR/Cas em genomas de arqueas e procariotos denota um sistema estruturalmente diverso. Já foram descritas mais de 45 famílias de proteínas relacionadas ao sistema, onde há evidência de homologia distante entre algumas destas famílias (MAKAROVA et al., 2011a). Diante desta diversidade estrutural e funcional, foi necessário criar um sistema de classificação.

De acordo com a última classificação, realizada por Koonin et al (2017), este sistema foi classificado em duas classes, seis tipos e cerca de 30 subtipos diferentes (Figuras 7 e 8). A divisão em classes é baseada na quantidade de proteínas que fazem parte do complexo efetor na etapa final do sistema, onde sistemas de classe 1 possuem várias proteínas compondo seu complexo efetor, enquanto sistemas de classe 2 possuem apenas uma proteína grande e multifuncional atuando nesta etapa. A divisão em tipos, por sua vez, baseia-se em análises filogenéticas, de organização estrutural e de genômica comparativa. Em geral, os tipos compreendem ao menos uma proteína exclusiva, considerada uma “assinatura”. Os critérios para a distinção em subtipos são a presença de determinados alelos e arranjos gênicos específicos (HATOUM-ASLAN; MARRAFFINI, 2014; MAKAROVA et al., 2015; KOONIN et al., 2017).

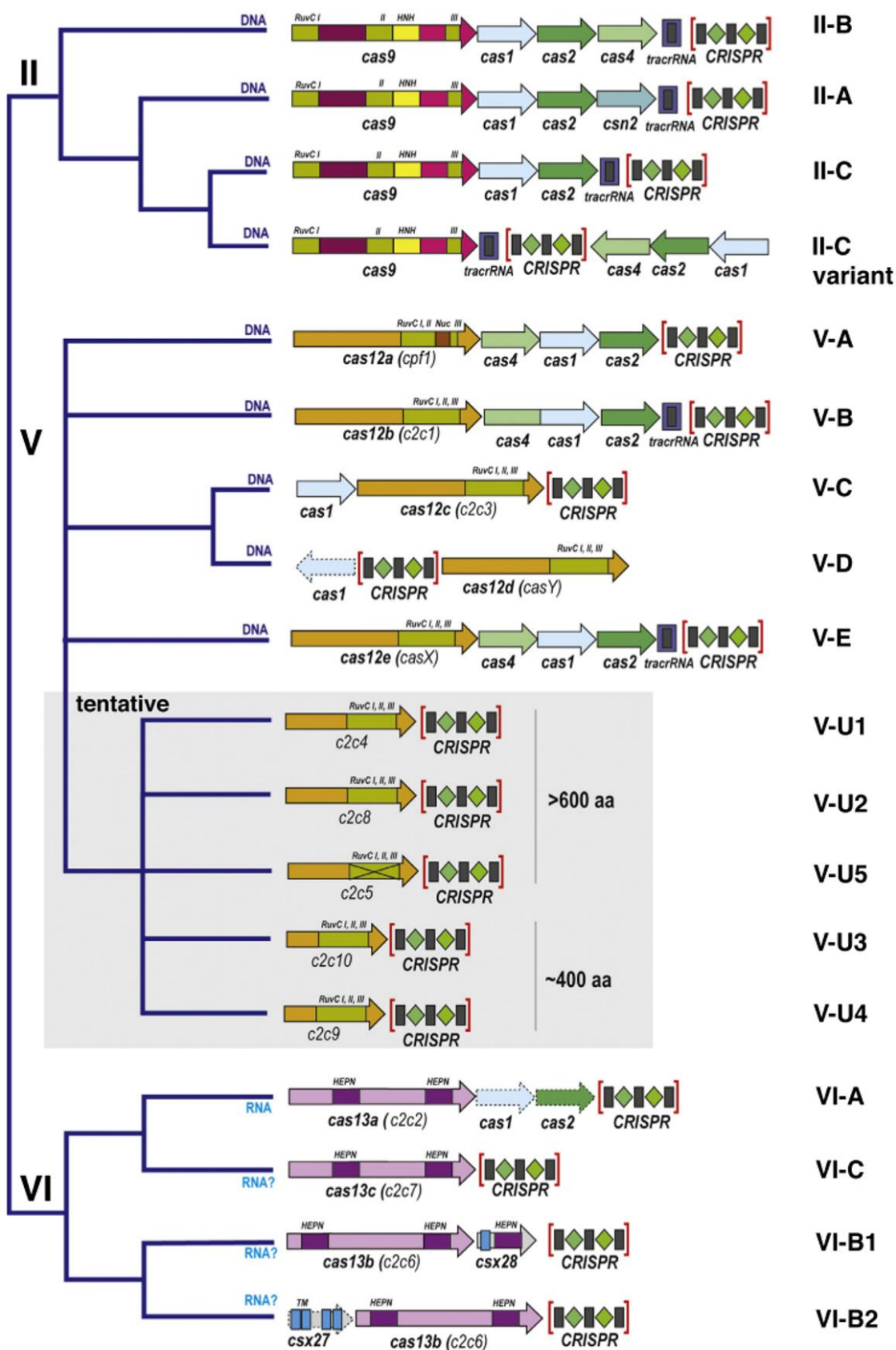
Figura 7. Sistemas CRISPR/Cas de classe 1.



Fonte: Koonin et al. (2017)

Nota: Proteínas marcadas em vermelho correspondem àquelas que fazem parte do complexo efetor.

Figura 8. Sistemas CRISPR/Cas de classe 2.



Fonte: Koonin et al. (2017)

A classe 1 compreende o tipo I, o tipo mais comum e diverso; o tipo III, mais encontrado em arqueas do que em bactérias; e o tipo IV, um tipo raro, rudimentar e ainda pouco estudado (KOONIN et al., 2017).

Os sistemas tipo I possuem com exclusividade o gene *cas3*, o qual codifica uma helicase com capacidade de resolver DNA dupla fita ou complexos DNA-RNA. Frequentemente, o domínio helicase da Cas3 está fusionado com um domínio de endonuclease. As proteínas Cas deste tipo estão codificadas em um ou dois operons, e possuem ortólogos dos genes *cas3*, *cas5*, *cas7* e *cas8*, e formam um complexo efetor conhecido como Cascade (*CRISPR-associated complex of antiviral defence*) (van DER OOST et al., 2014; MAKAROVA et al., 2015).

Os sistemas tipo III são os únicos capazes de interagir e degradar moléculas-alvo tanto de DNA quanto de RNA. A proteína Cas específica deste tipo é a Cas10, que possui um domínio de reconhecimento de RNA semelhante ao domínio observado em polimerases e constitui a maior subunidade do complexo efetor. Todos os sistemas tipo III também possuem os genes para *cas5* e *cas7* (Hatoum-Aslan and Marraffini 2014; Makarova et al. 2015). Sistemas tipo III não necessitam de um PAM para reconhecimento de moléculas exógenas, empregando outros mecanismos devido a particularidades de suas proteínas Cas (BISWAS et al., 2013; LEENAY; BEISEL, 2017).

As proteínas Cas1 e Cas2 são encontradas quase universalmente em todos os tipos de sistema, e eram consideradas marcadores para a presença do sistema CRISPR/Cas em um genoma. No entanto, no subtipo III-B estão ausentes os genes *cas1*, *cas2* e *cas6*, sendo este subtipo dependente de outros sistemas CRISPR/Cas para a realização de etapas específicas do mecanismo do sistema (MAKAROVA et al., 2011b; 2015).

O tipo IV é identificado pela presença das proteínas Csf1 (uma proteína *cas8-like*, considerada a assinatura do tipo IV), e ortólogos de Cas5 e Cas7. Assim como o subtipo III-B já mencionado, o tipo IV também não possui os genes *cas1* e *cas2* (KOONIN et al., 2017). Este sistema já foi encontrado em plasmídeos, porém sistemas CRISPR/Cas carregados por plasmídeos já foram descritos em *Acidithiobacillus ferrooxidans* (considerado o organismo modelo para o sistema tipo IV), na cianobactéria *Synechocystis*, em *Streptomyces rochei*, e em *Lactococcus lactis*. Esses plasmídeos sugerem que o sistema possa ser transferido horizontalmente, o que poderia explicar a presença do sistema CRISPR/Cas em isolados não filogeneticamente relacionados (BONDY-DENOMY; DAVIDSON, 2014; MAKAROVA et al., 2015).

Os tipos II, V e VI são considerados de classe 2, devido à presença de uma única proteína grande e com múltiplos domínios no complexo efetor. Devido à sua simplicidade, sistemas de classe 2 são menos diversos que os de classe 1 (KOONIN et al., 2017).

O tipo mais bem estudado entre aqueles que compõem a classe 2 é o tipo II, cuja proteína efetora é a Cas9. Devido à sua simplicidade, este tipo foi o modelo para o desenvolvimento do sistema como uma ferramenta de edição de genoma e controle da transcrição, vendida comercialmente e empregada na engenharia genética (KOONIN et al., 2017). A Cas9 é uma proteína bilobulada, apresentando um lóbulo de reconhecimento (REC) e um lóbulo de atividade de nuclease (NUC) com dois domínios catalíticos (*RuvC-like* e HNH), mais uma característica que impulsionou a padronização do CRISPR/Cas9 como um método de edição gênica, pois aumenta a versatilidade e as possibilidades na sua utilização (MAKAROVA et al., 2011b; MA et al., 2014; VIEIRA et al., 2016). A clivagem do pré-crRNA em vários

crRNAs é realizada pela RNase III, uma proteína convencional da célula bacteriana. O processamento do crRNA e a clivagem da molécula alvo ocorrem com o auxílio de uma classe de RNA conhecida como tracrRNA (transativador crRNA) (Figura 8), o qual possui complementariedade de bases com a DR. A fita dupla resultante é, então, clivado pela RNase III de maneira dependente da Cas9 (van DER OOST et al., 2014; WESTRA et al., 2014).

O tipo V foi observado em genomas bacterianos e um genoma de arquea. É caracterizado por genes da família da *cas12* (ou *cpf1*), encontrado adjacente a um loco CRISPR e aos genes *cas1* e *cas2*. O organismo modelo para o sistema tipo V foi uma cepa de *Francisella cf. novicida*. O gene *cas12a* codifica uma proteína grande que contém um domínio RuvC-like, como a Cas9, mas não contém um domínio HNH. Também diferente da Cas9, as proteínas da família Cas12 não necessitam de um tracrRNA para clivagem. Estas proteínas possuem uma alta similaridade com a proteína TnpB, da família de transposons IS605, sugerindo que foram recrutadas de elementos transponíveis (MAKAROVA et al., 2015; KOONIN et al., 2017).

O tipo VI foi o último tipo de sistema a ser descoberto (SHMAKOV et al., 2015), e entre os tipos classificados como 2, é o mais distinto dos demais, em função e em sequência (MURUGAN et al., 2017). A Cas13 difere das demais proteínas efetoras de classe 2 (Cas9 e família da Cas12) por não possuir domínios HNH ou RuvC-like, mas domínios HEPN com atividade de RNase, sendo este tipo, então, o único que degrada exclusivamente moléculas de RNA (KOONIN; ZHANG, 2017; YAN et al., 2018). A estrutura genética da Cas13 também é bastante variável entre os subtipos que a possuem, sendo a similaridade nucleotídica praticamente restrita apenas aos motivos HEPN (O'CONNELL, 2019).

2.3.3 Outras funções do sistema CRISPR/Cas

A extensão da função do sistema CRISPR/Cas ainda não é completamente conhecida. Sua aparente baixa taxa evolutiva em *Escherichia coli* (dado inconsistente com a pressão seletiva exercida por fagos sobre as células bacterianas), além da sua similaridade com o mecanismo de RNA de interferência em eucariotos, o qual apresenta outras funções além da imunidade, levaram à questão de que este sistema pode interferir em outros processos biológicos nas arqueas e bactérias que o possuem (WESTRA et al., 2014). Funções variantes do sistema CRISPR/Cas já descritas encontra-se no Quadro 3. Estas funções podem ser promovidas tanto pelas proteínas Cas e o loco CRISPR, quanto por apenas um destes elementos.

Quadro 3. Funções variantes do sistema CRISPR/Cas ou de apenas um dos seus elementos em espécies diversas.

Função	Tipo-subtipo	Mecanismo	Espécie
Regulação gênica	III-B	Degradação de RNA mensageiro complementar	<i>Pyrococcus furiosus</i>
Regulação de comportamento de colônia	I-F	Baseado em complementariedade parcial	<i>Pseudomonas aeruginosa</i>
Regulação de comportamento de colônia	I-C	Desconhecido	<i>Myxococcus xanthus</i>
Regulação de genes de virulência	II-C	Modificação da superfície celular dependente de Cas9	<i>Campylobacter jejuni</i>
Regulação de genes de virulência	II-B	Supressão da produção de lipoproteínas bacterianas mediada por Cas9	<i>Francisella novicida</i>
Regulação de genes de virulência	II-B	Desconhecido	<i>Legionella pneumophila</i>
Regulação de genes de virulência	Loco CRISPR órfão	Regulação do operon <i>feoAB</i> por complementariedade parcial	<i>Listeria monocytogenes</i>
Reparo de DNA	I-E	Reparo mediado por Cas1	<i>Escherichia coli</i>

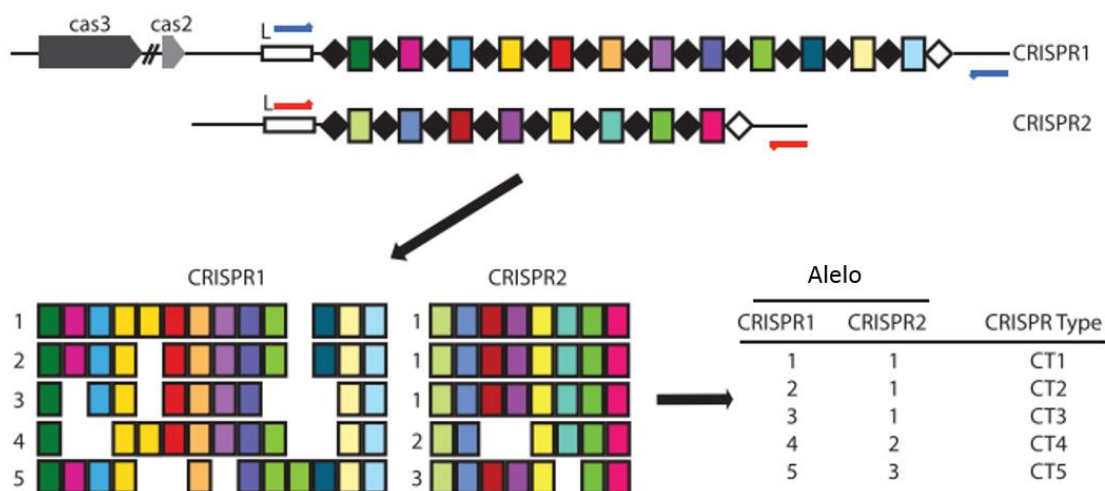
Fonte: adaptado de Westra et al. (2014)

2.3.4 Tipagem molecular através do sistema CRISPR/Cas

A diferenciação de bactérias além do nível de espécie é uma importante ferramenta epidemiológica para reconhecer causas de surtos infecciosos ou origens de infecções (SHARIAT; DUDLEY, 2014). As técnicas empregadas com este fim podem ser divididas em duas categorias: os métodos tradicionais, tais como sorotipagem, fagotipagem, e tipagem pelo perfil de resistência, os quais podem apresentar resultados variantes que não correspondem com a realidade; e os métodos moleculares, que incluem técnicas como análise de múltiplos locos VNTR (repetições em tandem de número variável) (MLVA), análise do polimorfismo do tamanho dos fragmentos de restrição (RFLP), eletroforese em gel de campo pulsátil (PFGE), ribotipagem, MLST, e mais recentemente, técnicas baseadas no sequenciamento de genoma completo (SHARIAT; DUDLEY, 2014; KARIMI et al., 2018).

O loco CRISPR do sistema CRISPR/Cas adquire novos espaçadores, cronologicamente, de maneira ordenada (novos espaçadores são sempre incorporados próximos à sequência líder), o que torna este loco uma região hipervariável única, cuja sequência de incorporação dos espaçadores pode ser utilizada para genotipagem (Figura 9) (BARRANGOU; DUDLEY, 2016).

Figura 9. Técnica empregada para tipagem molecular através da análise do loco CRISPR.



Fonte: adaptado de Shariat e Dudley (2014)

Nota: cada retângulo colorido corresponde a um espaçador diferente. De acordo com o conteúdo de espaçadores e a sua ordem de incorporação, é possível estabelecer perfis para cada um, e, ao final, estabelecer um tipo CRISPR único.

No entanto, para que o loco CRISPR de uma espécie bacteriana possa ser estabelecido como uma ferramenta válida de tipagem molecular, ele não pode incorporar espaçadores mais rapidamente do que o intervalo de tempo do surto em estudo, por exemplo. De maneira oposta, espécies cujo loco CRISPR é homogêneo devido à inatividade do sistema, também não são suficientemente informativos (SHARIAT; DUDLEY, 2014).

Inicialmente restrito à aplicação em espécies de *Yersinia* ou *Mycobacterium*, a tipagem baseada no CRISPR expandiu para isolados industriais e clínicos (BARRANGOU; DUDLEY, 2016). Atualmente, esta técnica já é bem estabelecida para algumas espécies, como demonstra o Quadro 4.

Quadro 4. Espécies bacterianas cuja tipagem através do loco CRISPR está estabelecida.

Organismo	Tipo
<i>Bifidobacterium animalis</i>	Probiótico
<i>Campylobacter jejuni</i>	Patógeno
<i>Clostridium difficile</i>	Patógeno
<i>Corynebacterium diphtheriae</i>	Patógeno
<i>Enterococcus faecalis</i>	Patógeno
<i>Enterococcus faecium</i>	Patógeno
<i>Erwinia amylovora</i>	Patógeno
<i>Escherichia coli</i>	Patógeno
<i>Lactobacillus buchneri</i>	Refugo
<i>Lactobacillus casei</i>	Probiótico
<i>Legionella pneumophila</i>	Patógeno
<i>Microcystis aeruginosa</i>	Patógeno
<i>Mycobacterium tuberculosis</i>	Patógeno
<i>Propionibacterium acnes</i>	Patógeno
<i>Salmonella enterica</i>	Patógeno
<i>Streptococcus agalactiae</i>	Patógeno
<i>Streptococcus thermophilus</i>	Cultura inicial
<i>Staphylococcus aureus</i>	Patógeno
<i>Vibrio parahaemolyticus</i>	Patógeno
<i>Xanthomonas aeruginosa</i>	Patógeno
<i>Yersinia pestis</i>	Patógeno

Fonte: adaptado de Barrangou e Dudley (2016)

2.4 A RESPOSTA DOS FAGOS: PROTEÍNAS ANTI-CRISPR

Numa tentativa de descobrir os mecanismos pelos quais profagos de *P. aeruginosa* modulam os mecanismos da célula bacteriana hospedeira, Bondydenomy e colaboradores, em 2016, se depararam com profagos que permitiam a infecção por outros fagos na célula bacteriana, apesar deste novo fago ser alvo de espaçadores do sistema CRISPR/Cas tipo I-F. Esta observação permitiu a hipótese

de que estes profagos inativariam a função do sistema CRISPR/Cas, que depois se mostrou verdadeira pela descoberta de um loco contendo genes anti-CRISPR, principalmente em fagos da família Mu-like (Figura 10) (BONDY-DENOMY et al., 2016; BORGES et al., 2017).

Figura 10. Locos anti-CRISPR de fagos da família Mu-like de *P. aeruginosa*.



Fonte: adaptado de Borges et al. (2017)

Nota: o loco anti-CRISPR está localizado entre as regiões bem conservadas dos genes G e genes da protease I/scaffold Z. As letras E e F, seguidas por um número, representam proteínas contra os tipos I-E e I-F, respectivamente, e encontram-se sempre adjacentes ao gene *aca1* (anti-CRISPR-associated 1).

Após a descoberta das proteínas anti-CRISPR em *P. aeruginosa*, outras proteínas foram descritas em outras espécies, através da similaridade de sequência e proximidade com o gene *aca1* (PAWLUK et al., 2016). Atualmente, existem cerca de 21 proteínas anti-CRISPR descritas (Quadro 5) (PAWLUK et al., 2018).

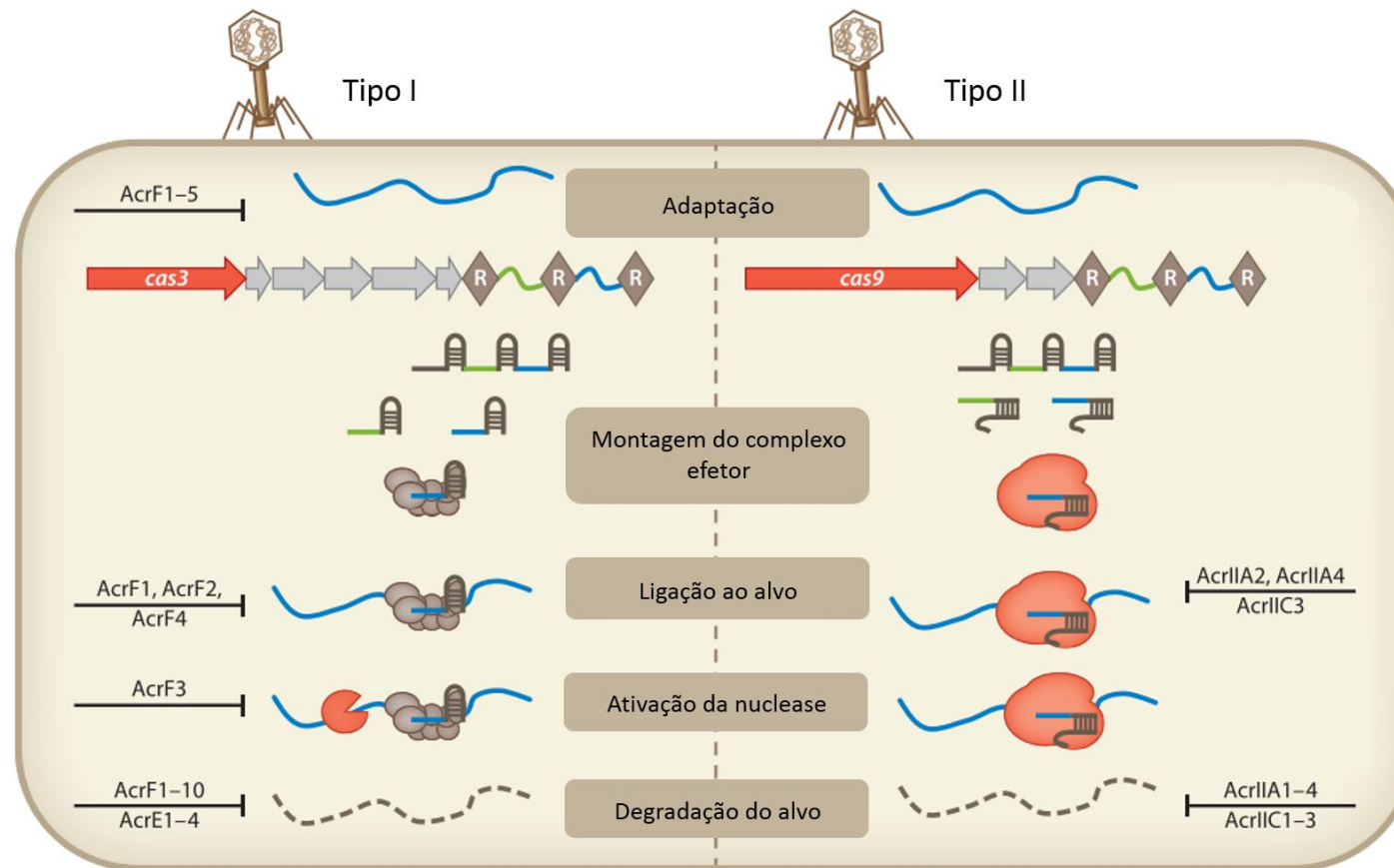
Quadro 5. Famílias de proteína anti-CRISPR.

Família de proteínas anti-CRISPR	Fago	Espécie relacionada	Tipo de sistema inibido	Número de aminoácidos
AcrE1	JBD5	<i>Pseudomonas aeruginosa</i>	I-E	100
AcrE2	JBD88a	<i>Pseudomonas aeruginosa</i>	I-E	84
AcrE3	DMS3	<i>Pseudomonas aeruginosa</i>	I-E	68
AcrE4	D3112	<i>Pseudomonas aeruginosa</i>	I-E	52
AcrF1	JBD30	<i>Pseudomonas aeruginosa</i>	I-F	78
AcrF2	D3112	<i>Pseudomonas aeruginosa</i>	I-F	90
AcrF3	JBD5	<i>Pseudomonas aeruginosa</i>	I-F	139
AcrF4	JBD26	<i>Pseudomonas aeruginosa</i>	I-F	100
AcrF5	JBD5	<i>Pseudomonas aeruginosa</i>	I-F	79
AcrF6	Profago	<i>Pseudomonas aeruginosa</i>	I-E e I-F	100
AcrF7	LPB1	<i>Pseudomonas aeruginosa</i>	I-F	67
AcrF8	ZF40	<i>Pectobacterium atrosepticum</i> .	I-F	92
AcrF9	Profago	<i>Vibrio parahaemolyticus</i>	I-F	68
AcrF10	Profago	<i>Shewanella xiamenensis</i>	I-F	97
AcrIIA1	J0161a	<i>Listeria monocytogenes</i>	II-A	149
AcrIIA2	J0161a	<i>Listeria monocytogenes</i>	II-A	123
AcrIIA3	SLCC2482	<i>Listeria monocytogenes</i>	II-A	125
AcrIIA4	J0161b	<i>Listeria monocytogenes</i>	II-A	87
AcrIIC1	MGE	<i>Neisseria meningitidis</i>	II-C	85
AcrIIC2	Profago	<i>Neisseria meningitidis</i>	II-C	123
AcrIIC3	Profago	<i>Neisseria meningitidis</i>	II-C	116

Fonte: adaptado de Borges et al. (2017) e Pawluk et al. (2018)

Os mecanismos pelos quais estas proteínas atuam para inativar o sistema é variado, e estão representados na Figura 11.

Figura 11. Processos do mecanismo do sistema CRISPR/Cas em que proteínas anti-CRISPR específicas interferem.



Fonte: adaptado de Borges et al. (2017)

Nota: a letra R representada no loco CRISPR corresponde às repetições. A sigla Acr, seguida por letras e números, correspondem às diferentes proteínas anti-CRISPR, e em que etapa do mecanismo do sistema elas interferem.

2.5 *PSEUDOMONAS AERUGINOSA* E O SISTEMA CRISPR/CAS

Já foi reportada a presença de sistemas tipo I-F, I-E e I-C em *P. aeruginosa*, em frequências variáveis. Em um estudo realizado por Van Belkum e colaboradores, em 2015, o sistema do tipo I-F foi observado em 30% dos isolados, o tipo I-E em 12%, e o tipo I-C em 3% (van BELKUM et al., 2015). Mais de um sistema pode ser encontrado no mesmo isolado bacteriano, no entanto, esta é uma característica rara: Essoh et al. (2013) observaram em 6% (3/50) dos isolados, e Van Belkum et al. (2015), em 2,52% (23/672) (ESSOH et al., 2013; van BELKUM et al., 2015).

A cepa UCBPP-PA14 (ou simplesmente PA14) apresenta um conjunto de genes *cas* específicos do subtipo I-F e dois locos CRISPR, um à montante e outro à jusante dos genes *cas*. Esta cepa é a atual referência para estudos do sistema CRISPR/Cas em *P. aeruginosa* (SONNLEITNER et al., 2012).

Uma função alternativa conhecida deste sistema em *P. aeruginosa*, descoberto na cepa PA14, é a sua função na modulação de genes do profago (um fago cujo genoma está integrado no genoma bacteriano) DMS3 (ZEGANS et al., 2009). Cepas que contêm este profago perdem sua capacidade de motilidade em grupo e produção de biofilme, o que ocorre pela similaridade parcial entre um espaçador do loco CRISPR com o gene 42 do fago DMS3. Uma vez que ocorra a mutação ou deleção dos genes *cas* envolvidos, ou deleção do espaçador do loco CRISPR, as funções de motilidade e produção de biofilme são restauradas (CADY; O'TOOLE, 2011).

Cady e colaboradores, em 2011, reportaram que o subtipo I-F de *P. aeruginosa* não parecia conferir imunidade à bactéria contra o ataque de bacteriófagos, apesar de crRNAs serem produzidos. Esta descoberta parecia

contradizer os achados de Zegans e colaboradores, em 2009, que reportaram a interação entre crRNA e a expressão de genes do fago DMS3 (ZEGANS et al., 2009; CADY et al., 2011). A partir de então, passou-se a acreditar que o papel do sistema CRISPR/Cas no genoma de *P. aeruginosa* seria apenas modular a expressão gênica, e não de proteger a célula bacteriana contra a invasão de novos MGEs (SONNLEITNER et al., 2012). No entanto, Cady e colaboradores, em 2012, descreveram a proteção contra fagos pelo subtipo I-F do sistema CRISPR/Cas em *P. aeruginosa*, obtido através da mudança de parâmetros experimentais.

Westra e colaboradores, em 2015, demonstraram que, o fato da *P. aeruginosa* tornar-se resistente aos fagos não seria apenas pela expressão do sistema CRISPR/Cas, mas também pela modificação dos receptores do fago pela célula, e que o sistema CRISPR/Cas parece ser o mecanismo de defesa de escolha sob condições nutritivas pobres, enquanto outros mecanismos estariam ativos em meios ricos em nutrientes (WESTRA et al., 2015). Diante das descrições controversas, a função deste sistema em *P. aeruginosa* ainda é considerado não completamente elucidado (ESSOH et al., 2013; van BELKUM et al., 2015).

3 ARTIGO I - INVESTIGATION OF FREQUENCY, LOCI'S ARCHITECTURE AND SPACER DIVERSITY OF CRISPR/CAS IN BRAZILIAN CLINICAL STRAINS OF *PSEUDOMONAS AERUGINOSA*

Manuscrito submetido à revista *Molecular Genetics and Genomics* (Qualis B1) em 20 de novembro de 2018 sob o código MGAG-D-18-00723. Manuscrito atualmente em correção, de acordo com as recomendações dos revisores, recebidas em 04 de fevereiro de 2019.

O material suplementar eletrônico do manuscrito encontra-se em mídia digital, que acompanha este documento.

**Investigation of frequency, loci's architecture and spacer diversity of
CRISPR/Cas system in Brazilian clinical strains of *Pseudomonas aeruginosa***

Ana Carolina de Oliveira Luz^{1,2}, Julia Mariana Assis da Silva², Antonio Mauro
Rezende², Maria Paloma Silva de Barros², Tereza Cristina Leal-Balbino^{2*}

¹Departamento de Genética, Universidade Federal de Pernambuco, Recife,
Pernambuco, Brazil.

²Departamento de Microbiologia, Instituto Aggeu Magalhães, Recife, Pernambuco,
Brazil.

* Corresponding author

E-mail: cristina@cpqam.fiocruz.br (TCL)

Telephone number: +558121012696

ABSTRACT

CRISPR/Cas is an adaptive immune system found in prokaryotes, with the main function to protect these cells from invasion and possible death by mobile genetic elements. *Pseudomonas aeruginosa* is considered a model for type I-F CRISPR/Cas system studies. However, its CRISPR loci characteristics have not yet been thoroughly described, and its function has not yet been fully unraveled. Here, we aimed to find the system's frequency in Brazilian clinical isolates; to identify the loci's sequence and its spacer diversity and their origins, as well as to propose a unified spacer library, to aid in future structural studies of *P. aeruginosa*'s CRISPR loci. We investigated types I-F and I-E gene markers to establish CRISPR/Cas typing, and observed two strains harboring both systems simultaneously, a very rare feature. Through amplification and sequencing of CRISPR loci related to type I-F system, we describe polymorphisms in DRs and 350 spacers, of which 97 are new. Spacers identified had their possible organisms or proteins of origin identified. Spacer arrays were grouped in five different CRISPR patterns, and the loci's plasticity was inferred by rearrangements in spacer arrays. Here, we perform the first detailed and focused description of CRISPR/Cas elements in Brazilian clinical strains of *P. aeruginosa*. Our findings reflect active and highly diverse CRISPR loci, and we suggest that CRISPR/Cas may also pose as a transcriptional regulatory mechanism. The structural and diversity features here described can provide insights on CRISPR/Cas function in this pathogen, and help guide the development of new therapeutic strategies.

Keywords: bacteriophages, plasmids, I-F, I-E, *Pseudomonas aeruginosa*, CRISPR/Cas

42

43 **INTRODUCTION**

44 Understanding the mechanisms through how bacteria react to the invasion of
45 mobile genetic elements (MGE) is important for a great number of applications, such
46 as comprehending the co-evolution between bacteria and their viruses
47 (bacteriophages), or aiding in the development of new strategies to treat infections by
48 multidrug resistant bacteria, like phage therapy, where bacteriophages are used to
49 specifically target and destroy bacterial cells in a patient (Dy et al. 2014; Kakasis and
50 Panitsa 2019). Among many mechanisms of resistance against phages, the
51 CRISPR/Cas system appears to be the most complex in prokaryotes, comprising two
52 key components to function: The Cas (CRISPR-associated) proteins and the
53 CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) locus (Bondy-
54 Denomy and Davidson 2014). CRISPR loci harbor key elements to learn about
55 CRISPR/Cas functions, the parasitic genetic elements present in a determined
56 environment, and may serve as a target for molecular typing (Barrangou 2015).

57 The CRISPR locus is composed of Direct Repeats (DR) separated by spacers,
58 which are partial sequences of MGE. They are acquired when the bacterial cell is first
59 invaded by a given MGE, and are incorporated in CRISPR locus between DRs.
60 These spacers can later serve as guides for the Cas proteins to block invasions by
61 the MGE from which they originated (van der Oost et al. 2014). This locus is typically
62 flanked by a leader sequence, a 5' AT-rich sequence that functions as the locus'
63 promoter and as a guide for the incorporation of new spacers (Jansen et al. 2002).
64 The last DR, on the opposite side from the leader sequence, usually shows some
65 level of degeneracy, harboring several polymorphisms, and is considered a
66 Degenerate Repeat (DG) (Biswas et al. 2014; Karah et al. 2015).

The CRISPR/Cas system and its components are well studied in several bacterial species, including many human pathogens, e.g. *Staphylococcus epidermidis*, *Streptococcus pneumoniae*, *Yersinia pestis*, *Klebsiella pneumoniae*, *Salmonella sp*, and many others (Marraffini and Sontheimer 2008; Bikard et al. 2012; Barros et al. 2014; Ostria-Hernández et al. 2015; Fu et al. 2017; Karimi et al. 2018).

Pseudomonas aeruginosa, an opportunistic pathogen that causes severe infections in immunocompromised patients within healthcare-associated environments, usually carries a great arsenal of mechanisms to resist several antibiotic treatments, like resistance genes and biofilm formation ability (Silby et al. 2011). Three CRISPR/Cas system types (I-F, I-E and I-C) have already been reported in *P. aeruginosa* strains, with varying frequencies for each type, with type I-F being the most most commonly observed, and type I-C being very rarely found (van Belkum et al. 2015).

Although *P. aeruginosa* is considered a model for CRISPR/Cas system type I-F studies (van Belkum et al. 2015), there are very few detailed descriptions of its CRISPR locus' characteristics, and the natural function of *P. aeruginosa*'s CRISPR/Cas still seems to be unclear, with contradicting reports (Cady et al. 2011, 2012; Westra et al. 2015).

Here, we aim to establish grounds for possible descriptive or structural CRISPR investigations in *P. aeruginosa*, by describing the sequence's characteristics, providing a unified spacer library, and analyzing spacer's incorporation order and rearrangements. Even though there are reports in literature of CRISPR/Cas system type I-F in *P. aeruginosa* isolates from Brazil (van Belkum et al. 2015), this is the first detailed description of some of this locus' genetic architecture and spacer diversity in Brazilian clinical isolates.

MATERIALS AND METHODS

Bacterial strains

A total of 130 *P. aeruginosa* clinical strains were provided by two public hospitals in Recife (State of Pernambuco, Brazil) by spontaneous demand, already isolated and purified by the hospital's bacteriology sectors. The isolates were derived from clinical specimens of patients from outpatient clinics, inpatient wards and Intensive Care Units (ICU). The hospitals performed the specimens' collection, bacterial isolation and identification of bacterial isolates through automated methods. We had no contact with patients, nor have we received any patient information. The bacteria were stored in BHI-glycerol at -80°C in Department of Microbiology of Aggeu Magalhães Institute. Cells were reactivated by plating in BHI-agar supplemented with ampicillin for 24 hours.

DNA extraction, CRISPR/Cas typing, and CRISPR amplification

Genomic DNA was obtained using the InstaGene Matrix (Bio-Rad, California) following manufacturer's instructions. The DNA from strains UCBPP-PA14 and PcyII-10 were used as positive control, and were provided by Dr. Cristine Pourcel (Paris-Sud University, France).

In order to perform CRISPR/Cas system typing, we used the system I-F's specific *cas1* gene with primers to investigate type I-F, and *cse1* with primers designed for this study, to investigate type I-E. CRISPR amplification of I-F's loci 1 and 2 (named CRISPR1 and CRISPR2), already defined for *P. aeruginosa*, was performed in 30 isolates through PCR. Another set of primers for CRISPR2 was

designed for our study, with the sequences annealing even closer to the CRISPR locus to better amplify internal sequences. Primers sequences and references are shown in Table 1.

Table 1 Primers used for amplifications in the present study

Primers	Sequence 5'-3'	Reference
<i>cas1</i> -F	GACATTTCTCCCAGCGAACTGA	(Essoh et al. 2013)
<i>cas1</i> -R	CTTCTTCGGTCAGTAGATGCTC	(Essoh et al. 2013)
<i>cse1</i> -F	ATGACTGAACACTACAACCTGC	This study
<i>cse1</i> -R	GGTCATAGGAATGGTCGAACAC	This study
CRISPR1-F	CTTGACGACCTATGTGGCAG	(Essoh et al. 2013)
CRISPR1-R	GTCGCTGCAAAAAGAGGGTT	(Essoh et al. 2013)
CRISPR2-F	TTTTCGTCTGTGTGAGGAGC	(Essoh et al. 2013)
CRISPR2-R	AGCAAGTTACGAGACCTCGA	(Essoh et al. 2013)
CRISPR2P-F	AGGGTGGTTCTGCCTAGACG	This study
CRISPR2P-R	GGCGGGAAAACTCGGTAT	This study

125 **Amplicon purification and sequencing**

126 A representative group of 30 isolate's PCR products was chosen for
127 sequencing, in order to comprehend the architecture of CRISPR1 and CRISPR2. The
128 PCR products were purified with ExoSAP-IT (Thermo Fisher Scientific,
129 Massachusetts) or Wizard SV Gel and PCR Clean-Up System (Promega,
130 Wisconsin), depending on the presence of spurious bands in agarose gels. Purified
131 samples were sequenced on Applied Biosystems 3100 using BigDye Kit (Applied
132 Biosystems, California).

133

134 **CRISPR loci sequence analysis**

135 Sequence assembly was performed using Phred, Phrap and Consed algorithms
136 (Gordon et al. 1998; Ewing et al. 2005). CRISPRFinder (Grissa et al. 2007a) was
137 chosen to confirm the CRISPR sequences obtained. Degenerate Repeat was
138 analyzed manually, since CRISPRFinder was often not able to identify DGs.

139

140 **Unified spacer library for *P. aeruginosa***

141 A non-redundant spacer library was designed for identifying spacers in *P.*
142 *aeruginosa*, where each spacer was assigned an identification (ID) composed of the
143 letter "S" and a specific number (data in Online Resource 1). This library was built
144 gathering information from two previous works(Cady et al. 2011; Essoh et al. 2013),
145 where spacers were explicitly described. We chose to build the library first
146 chronologically and then alphabetically. Spacers described by more than one author
147 are grouped and named with spacers from the first description. To avoid
148 misclassification due to sequencing errors, spacers differing one to three bases were
149 considered the same spacer and spacers differing four to seven bases were

considered variants, as suggested by a previous study (Cui et al. 2008). These variants were named with the same ID of the original spacer followed by the letter “v”.

Spacer analysis

Spacers found in our study were named after the unified spacer library we built. These spacers were compared with the CRISPR database using the BLAST CRISPRdb tool (Grissa et al. 2007b) to determine if they already are described as spacers in this specific CRISPR database. We determined the possible origins of our spacers using: the CRISPRTarget online software (Biswas et al. 2013) with default configurations; BLASTn tool against microbial genomes in GenBank, where only the highest scores were taken into consideration; and BLAST tool against Uniprot, where only e-values lower than one were considered. Observing these results, we classified spacers as related to Phages, Plasmids, Integrative or Conjugative Elements (ICE), *P. aeruginosa* or Unknown, taking into account only results of *Pseudomonas* genus. The listed results can be found in Online Resource 2.

CRISPR array establishment

Spacers and DRs were assigned to a position according to their proximity to a DG, where the DR and spacer closer to a DG were assigned as position one, as this spacer was the first one to be acquired. Each Spacer ID established in the unified spacer library were arrayed for each locus and each strain studied, where we aimed to identify any patterns of spacer incorporation and similar patterns between strains. All spacers found in our strains and their positions in CRISPR loci can be verified in Online Resource 3.

RESULTS

Analysis of CRISPR loci of *P. aeruginosa*

CRISPR/Cas types I-F and I-E were determined for 130 clinical isolates of *Pseudomonas aeruginosa*, and we observed the following frequencies: 73 isolates negative for any type (56,2%), 50 positive for I-F (38,5%), 5 positive for I-E (3,8%), and 2 positive for both system's types (1,5%).

We attempted to amplify CRISPR1 and CRISPR2 loci, related to I-F system type, in 30 isolates, to better understand the structural basis in these loci and their spacer content. CRISPR1 locus was amplified in 18 isolates, and CRISPR2 locus, in 27 isolates. Fifteen strains amplified both loci. Aside from two CRISPR1 (isolates Pae_8 and Pae_39) and three CRISPR2 products (isolates Pae_96, Pae_99 and Pae_101), all PCR amplicons were successfully sequenced.

To identify if CRISPR loci were sequenced in their entirety in each strain, we investigated the DG sequence and the ending marker of the locus in both extremities of the sequence obtained. Spacers closer to DG are the first ones to be incorporated, therefore, we considered DG as the CRISPR's beginning, the adjacent spacer as the first one, and the next DR as the first repeat. A small and specific sequence was sought in the locus' another extremity, closer to the Leader end, where new spacers are incorporated. This sequence was defined according to UCBPP-PA14 genome sequence, and it assigns the ending of the locus (ending marker). These markers are slightly different in the PcyII-10 genome, in accordance with our findings, but all our clinical strains possessed the UCBPP-PA14 markers.

DG could not be observed in all sequences, because the forward primer for CRISPR1 and CRISPR2 were designed disregarding this feature and favoring

internal sequences. Even so, the beginning of CRISPR2 was reliably recognized in all sequences, due to first spacer being the same for all loci. As for CRISPR1, since the beginning (from DG to second spacer) is identical in UCBPP-PA14 and other genomes, we looked for the first DR as the beginning marker. Markers and point mutations are described in Table 2, and a graphic representation of CRISPR1 and CRISPR2 are provided in Fig 1.

Table 2 Genetic markers, direct repeats and degenerate repeats sequence's characteristics found in Brazilian clinical isolate's CRISPR loci

Element	Sequence	Sequence in Reverse-complement
CRISPR1 ending marker	GGTCTA	TAGACC
CRISPR1 ordinary DR	GTTC <u>ACTGCCGT</u> ATAG GCAGCTAAGAAA ^a	TTTCTTAGCTGCCTA <u>T</u> ACGGCAGTGAAC ^a
CRISPR1 first DR (beginning marker)	<u>I</u> TTCA <u>CTGCC</u> <u>AC</u> ATAG G <u>T</u> <u>C</u> G <u>T</u> CAAGAAA ^b	TTTCTT <u>GACGAC</u> CCTAT <u>G</u> TGGCAGTGA <u>AA</u> ^b
CRISPR2 ending marker	GAGCTA	TAGCTC
CRISPR2 ordinary DR	GTTC <u>ACTGCCGT</u> GTAG GCAGCTAAGAAA ^a	TTTCTTAGCTGCCTA <u>C</u> ACGGCAGTGAAC ^a
CRISPR2 DG (beginning marker)	GTTC <u>G</u> CTGCCGT <u>C</u> TAG GCAGA <u>ACCACCC</u> ^b	<u>GGGTGGTTCTGCCTA</u> <u>G</u> ACGGCAG <u>C</u> GAAC ^b

^abases underlined correspond to point mutations between ordinary DRs in both loci

^bbases underlined correspond to point mutations between the ending markers and their respective's loci ordinary DR.

A. Leader end

GGTCTA GTTCACTGCCGTATAGGCAGCTAAGAAA TGTCAGAAGGGATTACCGCGACCTGGTCGCA
GTTCACTGCCGTATAGGCAGCTAAGAAA TCAGGCTCATTTTCGTTGTCTCGATGCCCCGG
GTTCACTGCCGTATAGGCAGCTAAGAAA TGCTCGCTGATGACCAGCCGCAGCGCATGGTT
GTTCACTGCCGTATAGGCAGCTAAGAAA CGCAAAGCCCCGAGGACAATGACTTGATATC
GTTCACTGCCGTATAGGCAGCTAAGAAA AATCCGAGGCGGAGTTAGCCACTTGGCATAG
GTTCACTGCCGTATAGGCAGCTAAGAAA GCAGTGATCGAGCGCGCACGGTCGCGCAAGAC
GTTCACTGCCGTATAGGCAGCTAAGAAA AGAAGCGCATCCAGCGGATACGAAGATGCACTC
GTTCACTGCCGTATAGGCAGCTAAGAAA AACACAGCGTGTCGATGCCGAAGCGAAGGCCCG
GTTCACTGCCGTATAGGCAGCTAAGAAA AGTGAGCACGATCATCATGTGCGCCCTGCTGGC
GTTCACTGCCGTATAGGCAGCTAAGAAA ATCACCGGAGGCCACGGATTTCGTGTAGCTCA
GTTCACTGCCGTATAGGCAGCTAAGAAA ACCAGCTGCTGCGAGTGCTGGTTTCGCGCTGGC
GTTCACTGCCGTATAGGCAGCTAAGAAA TGTCCCGAAGTTCATAAGCGGGCTTCGGGCGA
TTTCACTTGCCCCCTAGGTCGTCAAGAA

DG end

Analysis of CRISPR direct repeats

exclusive to CRISPR1 and may also pose as a marker for this specific locus in *P. aeruginosa*.

Analysis of spacer sequences

The unified spacer library we propose (Online Resource 1) combines the spacer findings in two previous works (Cady et al. 2011; Essoh et al. 2013). Although other spacer libraries were generated (van Belkum et al. 2015), spacer sequences were not explicitly described, so we were not able to include these data. The proposed library lists 1054 spacers, 350 spacers found in our strains, of which 62 were previously described (Online Resource 1), and 288 of which are first described in this study.

We designed this unified library to approach our findings in a more realistic manner, in a way that the spacers, reported here as new, have not yet been described in those two previous works, have not been found in CRISPR database (Grissa et al. 2007b) and have not been annotated as CRISPR related or adjacent to any Cas protein in NCBI database. Following these parameters, we describe 97 truly new spacers of *P. aeruginosa* (Table 3). These spacers were compared to those in type I-F CRISPR/Cas loci in previously described Brazilian strains (van Belkum et al. 2015), to confirm their novelty.

Table 3 New spacers found in CRISPR loci related to type I-F CRISPR/Cas system in Brazilian strains of *P. aeruginosa*

Spacer ID	Sequence ^a	Related to
S786	AAAGTGTCCGGCGCCGATCCGAAAGCCAGCGC	Unknown
S787	AAATCCTGAAGGACGTTTCGGCAGTGGCAAAAC	Unknown
S789	AACAGGTCGAGTGGTATGGCCGGAAGCTGGAC	Phage
S793	AACGAGCTGAACAGCTCGACCAACGTGGTGCC	Phage
S795	AACGGGTTGCAAGGCTAACTGTACGTATTTGG	Unknown

S797	AACTCGCAGAAACGGGGCCACTGCACGCCAGC	Phage
S801	AAGAGAAAACGAAGAGGTGCTTTCATCTCAGG	Unknown
S804	AAGCTCTGTTTCGACGGCCAGTATTTTCATCGA	Phage
S805	AAGGACTACGTGTTTCAGCAGTCTCGGCGGCGT	Unknown
S813	AATTCCTCTATGGTTCCGGTCCGGTCGCGCAC	Phage
S814	ACAAAGAGGGGCAGCGCAGACGTTTCATCACGC	Unknown
S817	ACAGGTCACAATGCGAACTGAACATCACACTC	Unknown
S818	ACATTCGGTACAGCGGCGGCGGCAGCGTGCTG	Phage
S820	ACCCTTGCGCGCAGTGCTGCGACTTCCTCCGT	Phage
S824	ACGAAGCTGAGCGTCGACGAATTCGGCGTTTA	Phage
S825	ACGATGATCAGCTTCCACGGCGGCCGCGGCTT	Unknown
S826	ACTAAAAGACTCAGAAATCGTGCACGCCTTCC	Unknown
S830	AGAAGGGCGGCCAGGGTGTGTTGGATCGTACT	Unknown
S832	AGACCAGGCGGCACCCAGCCGATCAGGGCATC	Phage
S834	AGCAGCAGGCAGGACAGCGCGATGGTTGTCCA	Plasmid
S835	AGCAGCGCACGGACCTTGGTCTGTAGGTCGTC	Phage
S839	AGCCGCAGGCGGCGCAGCCCGTTAAGCAGGCT	Unknown
S851	AGGCGATGGCTGAAGTTCCGGCTGATGCTACT	Unknown
S854	AGGGGGACGGCCTGTCGGCCAGTCAGTACCGG	Unknown
S855	AGGTACAGGTACTGCTCATCCGCTGCGTACAG	Phage
S856	AGGTGCCGAACGCCTCATCGACCGACTGATGCT	Phage
S858	AGTAACAGTCATGATGTCGGCACTGGCCTGGC	Phage
S862	AGTCGATGAACATGAAAATCCCTGTCAATATC	Unknown
S864	AGTGCCGGGCGGCTCCGGTTCAAGCAGACGGC	Unknown
S866	AGTTGCTGCTCGAACAGTGTGAGCCAGACGTG	Phage
S867	ATAGACCTGAAAGAGGCCGCCCAAGTCCTGAT	Phage
S870	ATCACCATCGACCTGACCAAGGCCGCCCAAGT	Phage
S878	ATCGAGCGGGAAGATGCAGACGGCATCACGCT	Phage
S883	ATGGCGTGCAGATCGTCGAGGAACGGCTGCAT	Phage
S884	ATGTAGGCCGACTGCATATTGGTCTGGTAGAT	Phage and Plasmid
S884v	ATGTAGGCCGACT TTAAGG GGTCTGGTAGAT	Unknown
S898	CGACAATGACCAGGCGGTGATCTGGCCCTGGC	Phage and Plasmid
S899	CGCAGTAGTACCCCGATCAGTAACGGCATCAG	Unknown
S906	CTCGAACCCACCTCGGCCACAACAGCCGCCGG	Unknown
S906	CTCGAACCCACCTCGGCCACAACAN N CCGCCGG	
S908	GAACGGATCATCGGGTACCTCTACGAAACCTG	Phage
S909	GACGCAGGGTGGCCGCCTTCCTCGACGAGAGC	Unknown
S910	GACGGAAAAGAAGGGCCTGGAGCGACTGGTAA	Unknown
S912	GATCTGTCCTCAGCACCCCTCGATAGGTATCT	Unknown
S918	GCGCAAGGGTACGACAACGGTTTTCCGTACTT	Phage
S920	GCTACGGGGTGGAGTTCTATACCGCCCTGGTG	Phage
S923	GGCACCTTCAACAAGCAGCTCGCC CC CGCTGAA	Unknown
S923	GGCACCTTCAACAAGCAGCTCGCCGCGCTGAA	
S925	GGCGAGCCTGGTCGTCGTGGCGGAGTGATGGT	Phage

S926	GGGACGAACATGCGCCGGAGATCGCAGCCACT	Phage
S929	GGTGATGGCTATGCCACACACGTGGATTTCAG	Phage
S931	GTAAGGCATCGGCTGGTTCCATGCCAGCAGTT	Phage
S932	GTAGTGTCTGATAGCAGCCGTTGCCCCGCAC	Unknown
S938	TACATGATGCGACCCGAATTTTCGCATTGAAA	Phage
S941	TACGAGAAAACGTTTCGACAACACCATTCCCGC	Phage
S943	TACTCGGTGGCGGTGCTGCTGGCGGCACTGGC	Phage
S945	TAGCCAGATTGCTGTGCGGCGACGGTTCATTC	Phage
S946	TATCAGTGATGAGCAGTTGGCGGAGGTTGAGG	Phage
S950	TATTACGACATCCAGTGGCGAGACATGGCCATC	Phage
S951	TCAAAGAAAAGGTAGGAGTAGAGAATAGCGCG	Unknown
S953	TCCAGGCTGCTGATGGCGTGTTTCGAGGAAGGC	Unknown
S954	TCCAGGTGCTGGCAAACGTCAGCCCCTTTGCA	Unknown
S956	TCCATGTCCATGAACAACCTCGGCCTGAGCCTG	Phage
S961	TCGATGGAAGCAGGAAGTCCAACGCAGCCGGT	Phage
S963	TCGGGCCATATCGTCGACGCCGACACCCGCAT	Unknown
S963	<u>T</u> TCGGGCCATATCGTCGACGCCGACACCCG <u>C</u> CAT	
S969	TCGTGAATACCGGCAACGTCGCGACGTCGTA	Phage
S970	TCTCCGAGAAAATCATGACCAGCAGACGACGA	Unknown
S971	TCTGGGAAACTGGTAATTCACATAGAGTGAAA	Unknown
S973	TGAAGGCCGGGCACCTCGGCGCGGATGCGCTC	Phage
S981	TGAT <u>A</u> GGTACCCTGATTACTGGTTACCTGATT	Phage
S981	TGATTGGTACCCTGATTACTGGTTACCTGATT	
S982	TGCAGAACCAGCGGCAGTTCCGGGGCTTTCGG	Phage
S983	TGCAGACGATCAAGCCGGGTGCATCGAGTGAG	Phage
S986	TGCAGTTGCCAGACGATGCGAGCGTCCACTAC	Phage
S987	TGCCGCAGCGCCTGGCCCAGCCCGCTATGGTG	Phage
S990	TGCGAGTCCAGAACCAGGTTCGATGCCGACGGT	Phage
S994	TGGAGCGAACCAGATCAGCTACCGCGACTTCGC	Phage
S995	TGGATGATTGGCCAGACCGACAAGGGGAGCAA	Phage
S996	TGGATGTTGGTGGCGCCGTCCAGGGTGTAGGC	Phage
S998	TGGCAGTGGTGCCTACGGTGACAACCTTGATGA	Phage
S1006	TGGCGGGCCATCTGGGACGGCCGGGTCTACAA	Phage
S1007	TGGCGTAGGCGACGAGACTGTCCAATGGCCGC	Unknown
S1007	TGGCGTAGGCGACGAGACTGT <u>I</u> CAATGGCCGC	
S1009	TGGGAGTTCGTTCATGCTCTTCAGGGCTTCGAT	Unknown
S1010	TGGGCGAGATCGGCGCGTAGCCGTAAGGCGAG	Unknown
S1011	TGGGCGATCGAGCACGGCATCACGCGCGAGAC	Phage
S1015	TGGTCGATATTCACGACCGGAGACCGGTGGTG	Phage and Plasmid
S1016	TGGTGGCGCTGCTGGTAGAAGTTCGCACCTTG	Unknown
S1017	TGTACAAAGGCGTTTACCGCAGAGTTGATTCC	Phage
S1023	TTATGCGCCTGGCCGGGTGCTGGTGTGTGGTG	Phage
S1024	TTCAAGAATTCGATCACATCCCGGCGCTTTT	Phage
S1029	TTCCGGAAGAACCTGGCGTACTACCCGGAAGC	Unknown

S1031	TTCGCCGGTGGCTGGTGC GCGACCAACTGGCC	Phage
S1034	TTCGGAATAAAGACGGCGCGCATCGAGCACTAC	Phage
S1046	TTGTCGCTGATATCGAAGGTGTCCCACTTCCG	Unknown
S1048	TTGTCGTAGTAGGTGCTGACGCCACGGACCTT	Unknown
S1048	TTGTCGTAGTAGGTGCTGAC <u>CG</u> CCACGGACCTT	Unknown
S1048v	TTGTCGTATTA <u>AG</u> TGCT <u>GG</u> AC <u>GCC</u> CGGACCTT	Unknown
S1049	TTGTGGGCGGCACATCGCTGTTGGCTCCTGGTA	Phage
S1050	TTGTTGGATGACTGCAACTTGTGACATATTCG	Phage
S1053	TTTGCCTGTGCTACGCCCAGGCTGCGAGTCAG	Unknown
S1054	TTTTCCCAGTAGCGGCGGGACCGGTCATCACG	Unknown

^aBold and underlined bases represent the differences between spacer sequences and variant spacers.

We tried to determine the origins for our spacer's sequences, or at least to what those spacers may protect the isolate against, using three different online tools. Results on origins' hypothesis are completely described in Online Resource 2. We observed that 53% of these spacers are related to phages; 6% to plasmids; 3% to phages and plasmids; 0.86% to plasmid and ICE; 0.57% to phages and ICE; and 36% remain unknown. Also, we found one spacer related to a protein product of *P. aeruginosa* itself (S874 - Oxygen-independent coproporphyrinogen III oxidase). The five most frequent results obtained from CRISPRTarget among our spacers, aside from no significant matches, were from *Pseudomonas* phages phi1 and JBD58a (both described 36 times), JBD26 (described 35 times), Bacteriophage D3112 (34 times) and *Pseudomonas* phage H70 (33 times).

The organization of CRISPR arrays were obtained by allocating spacers per isolate according to their position in the locus (Fig 2). We observed some isolates with similar patterns of spacer incorporation, allowing us to determine CRISPR patterns and infer CRISPR rearrangements. Isolates were grouped in the same pattern if they shared at least three spacers in the same incorporation order.

Fig 2 Spacer arrays of I-F CRISPR loci in Brazilian strains of *P. aeruginosa*. Spacer position is shown in ordinal numbers. The first depicted spacer, marked in grey, is the closest to DG, being the first one incorporated in CRISPR locus. C1 and C2 correspond to CRISPR1 and CRISPR2 locus. Color schemes represent identical spacers incorporated in the CRISPR locus in the same order, when in comparison with other isolates in the same pattern. A) Patterns A, B, C and D. B) Patterns E, B/E, E/D. C) Patterns A, B, C and D. D) Patterns E, B/E, E/D.

Isolate	Locus	1°	2°	3°	4°	5°	6°	7°	8°	9°	10°	11°	12°	13°	14°	15°	16°	17°	18°	19°	20°	21°	22°
Pae_66	C1	S1019	S551	S113	S507	S727	S663	S744	S748	S761	S752	S703	S738	S720	S697	S711	S994	S878	S1010	S943	S941		
Pae_113	C1	S1019	S551	S113	S507	S744	S774	S714	S525	S764	S758	S749	S748	S761	S752	S703	S738	S720	S697	S711	S832		
Pae_113	C2	S906	S913	S859	S1047	S816	S1055																
PA14	C1	S1019	S675	S450	S180	S101	S131	S533	S619	S156	S95	S625	S600	S161									
PA14	C2	S906	S1043	S1027	S848	S877	S806	S815	S792	S849	S934	S881	S993	S803	S975	S935	S823	S893	S841	S903			
Pae_29	C2	S906	S894	S1000	S882	S807	S1008	S949	S1043	S1027	S848	S877	S1032	S806	S815	S849	S934	S978	S1004	S850	S1036		
Pae_114	C1	S1019	S60	S890	S860	S998	S1041	S814	S982	S852	S871V	S794	S888	S910	S871								
Pae_114	C2	S906	S1012	S343	S789	S988	S835	S1023	S986	S969	S858	S1003	S912	S857	S876	S951	S1050	S1044	S932	S839			
Pae_115	C1	S1019	S60	S890	S860	S998	S1041	S814	S982	S852	S871V	S794	S825	S888	S910	S871							
Pae_115	C2	S906	S1012	S343	S789	S988	S835	S1023	S986	S969	S858	S1003	S912	S857	S876	S951	S1050	S1044	S932	S839	S856		
Pae_28	C2	S906	S884	S896	S1054	S840	S1042	S942	S981	S788	S1011	S908	S926	S817	S987	S956	S1007	S963	S1049	S923			
Pae_39	C2	S906	S884	S896	S1054	S840	S1042	S942	S981	S788	S1011	S908	S926	S817	S987	S956	S1007	S990	S996	S953	S965	S971	
Pae_61	C2	S906	S884V	S896	S1054	S840	S1042	S942	S981	S788	S1011	S908	S926	S817	S987	S956	S1007	S963	S1049	S923	S970	S990	∴
Pae_67	C2	S906	S884	S896	S1054	S840	S1042	S942	S981	S788	S1011	S908	S926	S817	S987	S956	S1007	S963	S1049V	S923	S970	S990	∴
Pae_83	C2	S906	S884	S896	S1054	S840	S1042	S942	S981	S788	S1011	S908	S926	S817	S987	S956	S1007	S963	S1049	S923	S970	S990	∴

273 B.

Isolate	Locus	1°	2°	3°	4°	5°	6°	7°	8°	9°	10°	11°	12°	13°	14°	15°	16°	17°	18°	19°	20°	21°	22°
Pae_12	C1	S1019	S70	S160	S146	S14	S103	S716	S49	S251	S558V	S484	S590										
Pae_12	C2	S906	S968	S933	S977	S799	S1039	S1021	S911	S891	S924	S901	S829	S1020	S1022	S861	S420	S872	S1012	S879			
Pae_106	C1	S1019	S70	S160	S146	S14	S103	S716	S49	S251	S558	S484	S590	S804									
Pae_106	C2	S906	S968	S933	S977	S799	S1039	S1021	S911	S891	S843	S924	S901	S829	S1020	S1022	S861	S420	S872	S1012	S879		
Pae_127	C1	S1019	S70	S160	S146	S14	S103	S716	S49	S251	S558	S484	S590	S804									
Pae_127	C2	S906	S968	S933	S977	S799	S1039	S1021	S911	S891	S843	S924	S901	S829	S1020	S1022	S861	S420	S872	S1012	S879		
Pae_129	C1	S1019	S70	S160	S146	S14	S103	S716	S49	S251	S558	S484	S590	S804									
Pae_129	C2	S906	S968	S933	S977	S799	S1039	S1021	S911	S891	S843	S924	S901	S829	S1020	S1022	S861	S420	S872	S1012	S879		
Pae_74	C1	S1019	S70	S160	S146	S14	S103	S716	S49	S251	S558	S11	S256	S44	S591	S592	S410	S406	S484	S590	S950	S1051	S925
Pae_74	C2	S906	S968V	S933	S977	S799	S1039	S1021	S911	S891	S843	S924	S901	S831	S961	S914	S800	S905	S790	S784	S811	S940	:
Pae_104	C1	S1019	S1055	S160	S146	S14	S716	S49	S251	S558	S11	S256	S938	S946	S793	S931							
Pae_105	C1	S1019	S70	S160	S146	S14	S716	S49	S251	S558	S11	S256	S938	S946	S793	S931							
Pae_105	C2	S906	S827	S940	S980	S957	S1012	S862	S965	S867	S875	S818											
Pae_116	C1	S1019	S597	S104	S34	S491	S338	S574	S199	S614	S864	S995											
Pae_116	C2	S906	S968	S933	S977	S799	S1039	S1021	S911	S891	S843	S924	S901	S831	S900	S847	S914	S833	S922	S873	S809	S829	:
Pae_110	C2	S906	S968	S933	S1039	S1021	S911	S802	S891	S843	S924	S901	S831	S900	S1026	S930	S927	S892					
Pae_42	C2	S906	S894	S1000	S882	S807	S1008	S949	S1043	S1027	S848	S877	S1032	S806	S815	S792	S849	S934	S978	S1004	S850	S937	S1035
Pae_42	C1	S1019	S816	S160	S146	S987	S1033	S863	S14	S103	S49	S251	S558	S11	S874	S836							
Pae_107	C2	S906	S968	S933	S1039	S1021	S911	S802	S891	S924	S901	S831	S900	S1026	S930	S812	S927	S824	S870	S787	S973		
Pae_107	C1	S1019	S60	S890	S860	S998	S852																

F

B/E

E/D

274

275

As shown (Fig 2), S906 was the first spacer for CRISPR2 loci, and S1019 was the second spacer for CRISPR1 (the first one observed, our beginning marker being the first DR). Even if they were only explicitly depicted in our study, these sequences are found in the control strains we used, whose genomes are fully sequenced.

CRISPR1 sequences of six pattern E isolates held two sets of mutations in DRs: T to C in 22th base of a DR and A to G in 24th base in another one, always spaced apart by 117 base pairs. Strains with pattern C of spacer incorporation showed point mutations in adjacent DRs of CRISPR1 (A to C in 15th position in one DR, and C to T in 10th position in the other). A deletion of a T in position 22 on the first DR of CRISPR1 was also found in five strains from other patterns, including control PcyII-10.

Most of our isolates belong to CRISPR pattern E, and as so, we observe the most diversity between their loci, particularly in isolates Pae_74, Pae_104, Pae_105, Pae_110 and Pae_116, evidencing the differential incorporation of spacers closer to leader sequence. This pattern also has the most noticeable possible spacer rearrangements and deletions, like the S977, S799, S843, S802 and S831 inconsistency in CRISPR2 loci (Fig 2B). Similar rearrangements can be seen in pattern A (Fig 2A).

Isolates Pae_42 and Pae_107 showed two different patterns. Their CRISPR1 arrays clearly resemble one pattern's signature, but CRISPR2 clearly relates to another's. Therefore, we name them B/E and E/D, respectively (Fig 2B).

All pattern D strains harbored a deletion of supposedly 20 base pairs between DR three and spacer three (which is S896, with 21 base pairs only)

DISCUSSION

The full extent of CRISPR/Cas function in *P. aeruginosa* is rather unclear, with reports of successfully impeding MGE invasion in laboratory environment (Cady et al. 2012); having a role in regulating virulence (Li et al. 2016); and possibly only being active under given nutritional circumstances and virus density (Westra et al. 2015). A previous study with *P. aeruginosa* strains from India and United States reported the system present in 33% of isolates (Cady et al. 2011). Among 17 Brazilian strains used in Van Belkum et al (van Belkum et al. 2015), only 4 (23,5%) carried any elements of type I-F CRISPR/Cas system, and no isolates carried I-E type.

The absence in most strains and the apparent inertness under certain nutritional conditions suggest that CRISPR/Cas elements in *P. aeruginosa* are under constant selective pressure due to its high fitness costs, and therefore, are excluded from some *P. aeruginosa*'s lineages. Nevertheless, their persistence in some strains may indicate they somehow confer advantages, such as regulating the transcription of prophage or bacterial products. The absence of CRISPR/Cas system elements in *P. aeruginosa*'s genomes, or even the presence of *cas* genes and absence of CRISPR loci, seem to suggest the importance of MGEs in shaping the bacterial genome and increasing virulent or resistant phenotypes (Barrangou 2015; van Belkum et al. 2015).

Furthermore, observing types I-F and I-E simultaneously in the same isolate is rather uncommon. This feature has been described only once (Essouh et al. 2013), in three isolates. The findings in Van Belkum et al (van Belkum et al. 2015), also show this feature in 17 isolates inside a universe of 672 other isolates, accounting for 2,5%. However, this information is not described in their main article, and none of these 17 isolates are Brazilian. This is the first report on I-F and I-E types

simultaneously in Brazilian strains. Further studies are needed to determine if both CRISPR/Cas types are active in these isolates.

Concerning the spacers found in our strains, results obtained from CRISPRTarget show spacers matching both bacteriophages and plasmids, although, in most cases, these plasmids belong to other bacterial genera. Even though plasmids are mobile elements and would be able to be transferred between different species, these similarities could have happened occasionally. Therefore, these cases do not offer a proper origin hypothesis for those spacers, and were considered as of unknown origin.

The spacers found in our strains are mainly against bacteriophages (53%). Most spacers match more than one MGE, which may suggest the protospacer region is conserved between different bacteriophages or these bacteriophages are related, which is the case with MP22 and DMS3, both D3112-like phages (Ceyssens and Lavigne 2010). As so, exact origin cannot be established, but the spacer may confer protection against all matched bacteriophages.

Different spacers against the same MGE reported in this study could be explained by the rapid rate of mutation of these MGEs, especially bacteriophages (Dy et al. 2014). The polymorphisms described between spacers or the partial matches found between spacers and databases might be because when the bacteriophage had its contact with our strain, it had already evolved and accumulated the mutations in protospacer, when in comparison with its deposited genome.

Multiple spacers targeting the same bacteriophage or other MGEs in the same strain suggest primed acquisition, event already described for type I CRISPR/Cas systems where the presence of a spacer targeting a specific MGE leads to further acquisition of more spacers against the same MGE, which can be a handful tool due

to bacteriophage rate of mutation and possible mismatches generated by these polymorphisms (Burmistrz and Pyrc 2015).

We observed spacers matching plasmids and ICEs (Online Resource 2). As happens with spacers against bacteriophages, these spacers usually match more than one plasmid or ICE simultaneously. Targeting plasmids and ICEs clearly impacts on possible resistance and virulence genes acquisition. Genomic islands to which we found spacers against are PAGI-5, PAGI-6 and PAPI-1. PAPI-1 is a pathogenicity island first described in the UCBPP-PA14 genome, which is a highly virulent strain (Morales-Espinosa et al. 2012). PAGI-5 was found to be a hybrid genomic island of new sequences and PAPI-1 sequences, rendering the strain a virulent phenotype (Battle et al. 2008). PAGI-6, interestingly, has many regions homologous to phiCTX, a *P. aeruginosa* phage (Battle et al. 2009). This trait can be observed in our findings (Online Resource 2), where all spacers matched for PAGI-6 also matched for phiCTX, probably conferring the isolate protection against both.

One spacer (S874) appears to be related to the oxygen-independent coproporphyrinogen III oxidase protein in *P. aeruginosa*. This protein has been reported in *E. coli* as part of tetrapyrrole rings of heme biosynthesis, and is present in other bacterial species (Layer et al. 2002). Li et al (Li et al. 2016) have already reported that type I-F CRISPR/Cas system in *P. aeruginosa* targets *lasR*, and endogenous gene, to alter its expression. It could be possible for spacer S874 to be acting in the same manner. Further studies are needed to address this hypothesis.

A total of 126 sequences (36%) returned no significant results, which was expected, as a high number of CRISPR spacers are still of unknown origin (Burmistrz and Pyrc 2015). This might happen because of the number of MGEs not yet sequenced or even known. A practical example is the S70 spacer: it was previously

described (Cady et al. 2011) as a spacer with no matches, however its sequence was later deposited as plasmid pPA7790 in a *P. aeruginosa* strain.

Even though CRISPR array analysis is performed in other bacterial species as a typing method (Barrangou and Dudley 2016), it has not been established for *P. aeruginosa* as a valid technique, mainly because it is not present in all isolates. Here, we aimed only to observe similarities between CRISPR loci in different strains in order to observe possible spacer recombination or unusual incorporation of spacers, and attempted no typing inferences. However, the CRISPR similarities found in our strains could be indicative of genetic relationships, but further investigations are needed to confirm this hypothesis. Analysis of CRISPR array showed that CRISPR loci in *P. aeruginosa* displays a high plasticity and spacer diversity, but even so, similarity between strains were undeniable, and spacer rearrangement can be inferred.

Homologous recombination may occur within a CRISPR locus or between two CRISPR loci in the same genome due to sequence similarity in DRs, promoting spacer rearrangements and deletions. This phenomenon has already been described for *Streptococcus mutans* and *Salmonella* spp, among other bacterial species (Kupczok et al. 2015). The deletion in isolates harbouring CRISPR pattern C may be due to a previous conjugative event, a homologous recombination deletion, or a DNA polymerase slippage during replication.

Despite the ever increasing number of publications regarding CRISPR/Cas system mechanistic and their applicability in genetic engineering, there are still some aspects of this systems' structure that are mysterious, such as its evolutionary origins and spacers unrelated to known sequenced elements (Nature Editorial 2018).

Although many characteristics observed here are also present in control strains, this is the first detailed description of CRISPR/Cas types I-F and I-E frequencies, point mutations, DR and DG sequences, and the first focused description of subtype I-F CRISPR loci in Brazilian strains of *P. aeruginosa*. Also, we report for the first time the presence of I-F and I-E types simultaneously, and 97 new spacers, in Brazilian isolates. We can conclude that CRISPR loci in *P. aeruginosa* appear to be diverse, adaptable, and possibly active due to incorporation of different spacers at the beginning of the loci observed in several grouped strains. The new spacers and the polymorphic sequences described may represent a growing adaptation to bacteriophage threats specific to the hospital environment from where these isolates were obtained, another indicative of CRISPR/Cas activity. Understanding the nuances of CRISPR/Cas components in *P. aeruginosa* and other human pathogens may provide insights about their relationship with bacteriophages and other MGEs, in addition to infer possible dynamics on plasmid and resistance genes uptake. Even though some issues may still be addressed in the future, as CRISPR/Cas' possible function of transcriptional regulator in this species, the data gathered here is important to aid in future genetic studies of this bacterium, and may also aid in the successful development of phage therapy strategies for *P. aeruginosa*.

ACKNOWLEDGMENTS

The authors thank the Núcleo de Plataformas Tecnológicas (NPT) from Instituto Aggeu Magalhães (Fiocruz/IAM/PE) for use of the sequencing platform. Also, we thank Dr. Christine Pourcel for donating the control strains' genetic material.

COMPLIANCE WITH ETHICAL STANDARDS

Funding

This study was funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico (Process Number: 404016/2016-2) and by Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (Process Number: APQ-0836-2.02/16).

Conflict of Interest

Author Ana C. O. Luz declares that she has no conflict of interest.

Author Julia M. A. da Silva declares that she has no conflict of interest.

Author Antonio M. Rezende declares that he has no conflict of interest.

Author Maria P. S. Barros declares that she has no conflict of interest.

Author Tereza C. Leal-Balbino declares that she has no conflict of interest.

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

DATA AVAILABILITY

CRISPR, *cas1* and *cse1* sequences were submitted in GenBank database. The accession numbers are: CRISPR1 (MH665277 to MH665294); CRISPR2 (MH665295 to MH665320); *cas1* (MH665321 to MH665348); and *cse1* (MK184280 to MK184286).

REFERENCES

- 450 Barrangou R (2015) The roles of CRISPR-Cas systems in adaptive immunity and
451 beyond. *Curr Opin Immunol* 32:36–41.
- 452 Barrangou R, Dudley EG (2016) CRISPR-Based Typing and Next-Generation
453 Tracking Technologies. *Annu Rev Food Sci Technol*. doi: 10.1146/annurev-food-
454 022814-015729
- 455 Barros MPS, França CT, Lins RHFB, Santos MDV, Silva EJ, Oliveira MBM, Silveira-
456 Filho VM, Rezende AM, Balbino VQ, Leal-Balbino TC (2014) Dynamics of
457 CRISPR loci in microevolutionary process of *Yersinia pestis* strains. *PLoS One*.
458 doi: 10.1371/journal.pone.0108353
- 459 Battle SE, Meyer F, Rello J, Kung VL, Hauser AR (2008) Hybrid pathogenicity island
460 PAGI-5 contributes to the highly virulent phenotype of a *Pseudomonas*
461 *aeruginosa* isolate in mammals. *J Bacteriol* 190:7130–7140.
- 462 Battle SE, Rello J, Hauser AR (2009) Genomic islands of *Pseudomonas aeruginosa*.
463 *FEMS Microbiol Lett* 290:70–78.
- 464 Bikard D, Hatoum-Aslan A, Mucida D, Marraffini LA (2012) CRISPR interference can
465 prevent natural transformation and virulence acquisition during in vivo bacterial
466 infection. *Cell Host Microbe* 12:177–186.
- 467 Biswas A, Fineran PC, Brown CM (2014) Accurate computational prediction of the
468 transcribed strand of CRISPR non-coding RNAs. *Bioinformatics* 30:1805–1813.
- 469 Biswas A, Gagnon JN, Brouns SJJ, Fineran PC, Brown CM (2013) CRISPRTarget -
470 Bioinformatic prediction and analysis of crRNA targets. *RNA Biol* 10:817–827.
- 471 Bondy-Denomy J, Davidson AR (2014) To acquire or resist: The complex biological
472 effects of CRISPR-Cas systems. *Trends Microbiol* 22:218–225.
- 473 Burmistrz M, Pirc K (2015) CRISPR-Cas systems in prokaryotes. *Polish J Microbiol*
474 64:193–202

- 475 Cady KC, Bondy-Denomy J, Heussler GE, Davidson AR, O'Toole GA (2012) The
 476 CRISPR/Cas adaptive immune system of *Pseudomonas aeruginosa* mediates
 477 resistance to naturally occurring and engineered phages. J Bacteriol 194:5728–
 478 5738.
- 479 Cady KC, White AS, Hammond JH, Abendroth MD, Karthikeyan RSG, Lalitha P,
 480 Zegans ME, O'Toole GA (2011) Prevalence, conservation and functional
 481 analysis of *Yersinia* and *Escherichia* CRISPR regions in clinical *Pseudomonas*
 482 *aeruginosa* isolates. Microbiology 157:430–437.
- 483 Ceysens P-J, Lavigne R (2010) Bacteriophages of *Pseudomonas*. Future Microbiol
 484 5:1041–1055
- 485 Cui Y, Li Y, Gorgé O, Platonov ME, Yan Y, Guo Z, Pourcel C, Dentovskaya SV,
 486 Balakhonov SV, Wang X, Song Y, Anisimov AP, Vergnaud G, Yang R (2008)
 487 Insight into microevolution of *Yersinia pestis* by clustered regularly interspaced
 488 short palindromic repeats. PLoS One. doi: 10.1371/journal.pone.0002652
- 489 Dy RL, Richter C, Salmond GPC, Fineran PC (2014) Remarkable Mechanisms in
 490 Microbes to Resist Phage Infections. Annu Rev Virol 1:307–331.
- 491 Nature Editorial (2018) Crispr still needs microbiologists. Nat Microbiol. doi:
 492 10.1038/s41564-018-0175-x
- 493 Essoh C, Blouin Y, Loukou G, Cablanmian A, Lathro S, Kutter E, Thien HV,
 494 Vergnaud G, Pourcel C (2013) The Susceptibility of *Pseudomonas aeruginosa*
 495 Strains from Cystic Fibrosis Patients to Bacteriophages. PLoS One. doi:
 496 10.1371/journal.pone.0060575
- 497 Ewing B, Hillier L, Wendl MC, Green P (2005) Base-Calling of Automated Sequencer
 498 Traces Using Phred. I. Accuracy Assessment. Genome Res 8:175–185.
- 499 Fu S, Hiley L, Octavia S, Tanaka MM, Sintchenko V, Lan R (2017) Comparative

- 500 genomics of Australian and international isolates of *Salmonella Typhimurium*:
 501 Correlation of core genome evolution with CRISPR and prophage profiles. Sci
 502 Rep 7:9733. doi: 10.1038/s41598-017-06079-1
- 503 Gordon D, Abajian C, Green P (1998) Consed: a graphical tool for sequence
 504 finishing. Genome Res 8:195–202.
- 505 Grissa I, Vergnaud G, Pourcel C (2007a) CRISPRFinder: a web tool to identify
 506 clustered regularly interspaced short palindromic repeats. Nucleic Acids Res
 507 35:52–57.
- 508 Grissa I, Vergnaud G, Pourcel C (2007b) The CRISPRdb database and tools to
 509 display CRISPRs and to generate dictionaries of spacers and repeats. BMC
 510 Bioinformatics. doi: 10.1186/1471-2105-8-172
- 511 Jansen R, Van Embden JDA, Gaastra W, Schouls LM (2002) Identification of genes
 512 that are associated with DNA repeats in prokaryotes. Mol Microbiol 43:1565–
 513 1575.
- 514 Kakasis A, Panitsa G (2019) Bacteriophage therapy as an alternative treatment for
 515 human infections. A comprehensive review. Int J Antimicrob Agents 53:16–21.
- 516 Karah N, Samuelsen Ø, Zarrilli R, Sahl JW, Wai SN, Uhlin BE (2015) CRISPR-cas
 517 subtype I-Fb in *Acinetobacter baumannii*: Evolution and utilization for strain
 518 subtyping. PLoS One. doi: 10.1371/journal.pone.0118205
- 519 Karimi Z, Ahmadi A, Najafi A, Ranjbar R (2018) Bacterial CRISPR Regions: General
 520 Features and their Potential for Epidemiological Molecular Typing Studies. Open
 521 Microbiol J 12:59–70.
- 522 Kupczok A, Landan G, Dagan T (2015) The contribution of genetic recombination to
 523 CRISPR array evolution. Genome Biol Evol 7:1925–1939.
- 524 Layer G, Verfürth K, Mahlitz E, Jahn D (2002) Oxygen-independent

- 525 coproporphyrinogen-III oxidase hemN from *Escherichia coli*. J Biol Chem
526 277:34136–34142.
- 527 Li R, Fang L, Tan S, Yu M, Li X, He S, Wei Y, Li G, Jiang J, Wu M (2016) Type I
528 CRISPR-Cas targets endogenous genes and regulates virulence to evade
529 mammalian host immunity. Cell Res 26:1273–1287.
- 530 Marraffini LA, Sontheimer EJ (2008) CRISPR interference limits horizontal gene
531 transfer in Staphylococci by targeting DNA. Science (80-) 322:1843–1845.
- 532 Morales-Espinosa R, Soberón-Chávez G, Delgado-Sapién G, Sandner-Miranda L,
533 Méndez JL, González-Valencia G, Cravioto A (2012) Genetic and phenotypic
534 characterization of a *Pseudomonas aeruginosa* population with high frequency
535 of genomic islands. PLoS One. doi: 10.1371/journal.pone.0037459
- 536 Ostria-Hernández ML, Sánchez-Vallejo CJ, Ibarra JA, Castro-Escarpulli G (2015)
537 Survey of clustered regularly interspaced short palindromic repeats and their
538 associated Cas proteins (CRISPR/Cas) systems in multiple sequenced strains of
539 *Klebsiella pneumoniae*. BMC Res Notes 8:332. doi: 10.1186/s13104-015-1285-7
- 540 Silby MW, Winstanley C, Godfrey SAC, Levy SB, Jackson RW (2011) *Pseudomonas*
541 genomes: Diverse and adaptable. FEMS Microbiol Rev 35:652–680.
- 542 van Belkum A, Soriaga LB, LaFave MC, Akella S, Veyrieras J, Barbu EM, Shortridge
543 D, Blanc B, Hannum G, Zambardi G, Miller K, Enright MC, Mugnier N, Bami D,
544 Schicklin S, Felderman M, Schwartz AS, Richardson TH, Peterson TC, Hubby B,
545 Cady KC (2015) Phylogenetic Distribution of CRISPR-Cas Systems in Antibiotic-
546 Resistant *Pseudomonas aeruginosa*. MBio. doi: 10.1128/mBio.01796-15
- 547 van der Oost J, Westra ER, Jackson RN, Wiedenheft B (2014) Unravelling the
548 structural and mechanistic basis of CRISPR-Cas systems. Nat Rev Microbiol
549 12:479–92.

550 Westra ER, van Houte S, Oyesiku-Blakemore S, Makin B, Broniewski JM, Best A,
551 Bondy-Denomy J, Davidson A, Boots M, Buckling A (2015) Parasite Exposure
552 Drives Selective Evolution of Constitutive versus Inducible Defense. Curr Biol
553 25:1043–1049.

554

555 **SUPPLEMENTARY MATERIAL**

556 **Online Resource 1. Unified spacer library.** Library of spacers built by gathering
557 data from two previous studies and our results. Comprises 1054 spacers.

558 **Online Resource 2. Table of spacer origin.** Spacers found in our strains and their
559 similarity with sequences in databases, stating their possible origin or to what they
560 might confer protection against.

561 **Online Resource 3. Spacer results in our strains.** Presents all spacers found in all
562 strains studied, their sequence, their spacer ID, their locus of origin and their position
563 in the CRISPR locus.

564

4 ARTIGO II - GENOMIC ANALYSIS OF *Pseudomonas aeruginosa* HARBORING CRISPR/CAS SYSTEMS SUGGESTS NO IMMUNITY FUNCTION

Manuscrito a ser submetido à revista *Genome Biology and Evolution* (Qualis A2). O material suplementar eletrônico deste manuscrito encontra-se em mídia digital, que acompanha este documento.

**Genomic Analysis of *Pseudomonas aeruginosa* Harboring CRISPR/Cas
Systems Suggests No Immunity Function**

Ana Carolina de Oliveira Luz^{1,2}, Julia Mariana Assis da Silva², Wilson José da Silva Junior³, José Bandeira do Nascimento Junior³, Valdir de Queiroz Balbino³, Tereza Cristina Leal-Balbino^{2*}

¹Department of Genetics, Federal University of Pernambuco, Recife, Pernambuco, Brazil

²Department of Microbiology, Aggeu Magalhães Institute, Oswaldo Cruz Foundation, Recife, Pernambuco, Brazil

³Laboratory of Bioinformatics and Evolutionary Biology, Department of Genetics, Federal University of Pernambuco, Recife, Pernambuco, Brazil

Author for Correspondence: Tereza Cristina Leal-Balbino, Department of Microbiology, Aggeu Magalhães Institute, Oswaldo Cruz Foundation, Recife, Pernambuco, Brazil, +558121012696, cristina@cpqam.fiocruz.br

Data deposition: The BioProject has been deposited at the GenBank database under the accession PRJNA514718.

ABSTRACT

Pseudomonas aeruginosa is a versatile and adaptable species, acting as an opportunistic pathogen, commonly infecting burned and immunocompromised patients, or those with other conditions of susceptibility. Strains of *P. aeruginosa* may carry CRISPR/Cas systems types I-F, I-E and I-C, however, this system's function in this species is not yet fully understood. Here, we report 13 genomes of clinical *P. aeruginosa* Brazilian strains positive for CRISPR/Cas systems, two of which harbor types I-F and I-E simultaneously, a rare feature between this species' representatives. The phylogenetic tree, constructed with other 161 publicly available genomes, suggests no direct relationship between positive strains, with the various types of CRISPR/Cas systems spread through the tree. We established the common genetic background for types I-F, I-E and a specific orphan CRISPR (without cas genes in the same genome), named LES-orphan CRISPR, showing this orphan locus does not appear to be directly related to type I-F due to its different backgrounds. We also report the presence of prophages to which there are spacers in the genome's CRISPR locus, and the presence of anti-CRISPR genes. Observing anti-CRISPR-free genomes carrying prophages and spacers against these prophages is a finding that contradicts the very basis of CRISPR/Cas function, suggesting this system's main function in *P. aeruginosa* may not be immunity against bacteriophages, but actually regulation of gene expression of phage products.

Key words: bacteriophage; I-F; I-E; anti-CRISPR; prophage; Brazilian strains.

INTRODUCTION

The *Pseudomonas* genus includes a great number of ubiquitous and

opportunistic species, responsible for infecting plants and humans (Silby et al. 2011). *Pseudomonas aeruginosa*, one the most prominent pathogens in hospitals and clinics, infects burnt or immunocompromised patients, causing healthcare-associated pneumonia, cystic fibrosis (CF) associated infections of respiratory system, bacteremia, and severe urinary tract, gastrointestinal and skin infections (Azam & Khan 2018). This pathogen's success is mainly due to its remarkable ability to resist antibiotic treatment by acquiring new resistance genes and forming biofilm, leading to high morbidity, mortality, and healthcare costs (Pereira et al. 2018).

Pseudomonas aeruginosa typically presents a large and plastic genome (higher than 6 Mbp) when in comparison with other bacterial human pathogens. Aside from metabolism-associated genes, the core genome of *P. aeruginosa* also comprises resistance encoding genes, such as cephalosporinases (AmpC) and efflux pumps (MexAB-OprM), as well as a number of virulence-related genes. This specie's accessory genome is a major determinant to each strain success and evolution, since it harbours components obtained by horizontal gene transfer, conferring different phenotypes. This extensive pangenome is responsible for *P. aeruginosa*'s high adaptability to different niches and stress situations (Valot et al. 2015; Azam & Khan 2018).

Some housekeeping genes have been studied and typified, and their allelic combination make for the Multilocus Sequence Typing (MLST) of *P. aeruginosa* strains. Observing the Sequence Types (STs) of several *P. aeruginosa* isolates distributed worldwide, this microorganism usually shows a nonclonal dispersion, with very few Clonal Complexes (CCs) being able to be defined (van Belkum et al. 2015; Parkins et al. 2018).

The CRISPR/Cas system in *P. aeruginosa* is considered a part of its accessory genome, as this system is not observed in all isolates. It is an adaptive immunity mechanism against bacteriophages and plasmids, preventing recurring invasions by mobile genetic elements (MGE). CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) is a locus composed of several repeats interspaced by different and unique sequences called spacers. These spacers are the key to the adaptive characteristic of the system: they are segments of MGE's DNA incorporated in the CRISPR locus during the first contact with the given MGE. In a second contact, the bacterial cell is able to cleave the invading DNA molecule, guided by the spacer sequence. The effectors proteins in this entire process are the Cas (CRISPR-associated) proteins (van der Oost et al. 2014; Hille et al. 2018). According to the latest classification, CRISPR/Cas systems are divided in two classes, six types and more than 30 subtypes. Subtypes I-F, I-E and I-C have already been found in *P. aeruginosa* isolates in different frequencies, being I-F the most frequent, and I-C identified in very few strains (Cady et al. 2011; van Belkum et al. 2015; Koonin et al. 2017).

As well as bacteria have evolved to evade bacteriophage infections, bacteriophages have evolved to better infect bacteria. This constant arms race has led to the discovery of anti-CRISPR proteins, bacteriophage products capable of inactivate the CRISPR/Cas system mechanism. *Pseudomonas aeruginosa* is the bacterial species with the greatest number of anti-CRISPR proteins known, with at least 11 proteins already described against system's types I-F and I-E (Pawluk et al. 2018).

Here, we report 13 Brazilian isolates harboring the CRISPR/Cas system, their phage and anti-CRISPR protein content, and their genetic relationship with other

publicly available genomes.

MATERIALS AND METHODS

Bacterial strains

The 13 *P. aeruginosa* isolates were provided purified by the bacteriology sector from two public hospitals in Recife (State of Pernambuco, northeast Brazil) by spontaneous demand, derived from clinical routine, already isolated and identified. We had no contact with patients, nor received any patient information. The strains are stored in BHI-glycerol at -80°C in the Department of Microbiology of Aggeu Magalhães Institute, and their characteristics are found in Table 1. All isolates were previously screened for CRISPR/Cas system through PCR (data not shown).

Table 1. Characteristics of *Pseudomonas aeruginosa* Brazilian isolates.

Isolate	Genome Coverage	ST	CRISPR/Cas type	Year of collection	Hospital unit	Clinical specimens
Pae12	176	244	I-F	2015	Ward (Hospital 1)	Bone fragment and tissue
Pae21	161	1394	I-F	2015	Ward (Hospital 1)	Urine
Pae22	123	252	I-F and I-E	2015	ICU ^a (Hospital 1)	Sputum
Pae28	151	3079	I-F	2015	Ward (Hospital 1)	Surgery material

Pae29	124	357	I-F	2015	ICU (Hospital 1)	Sputum
Pae39	122	3079	I-F	2015	ICU (Hospital 1)	Sputum
Pae42	348	357	I-F	2016	ICU (Hospital 1)	Sputum
Pae66	142	3078	I-F	2016	Ward (Hospital 1)	Bone fragment
Pae70	182	275	I-F	2016	Ward (Hospital 1)	Urine
Pae74	156	244	I-F	2016	ICU (Hospital 1)	Sputum
Pae81	186	3080	I-F	2016	Ward (Hospital 1)	Urine
Pae83	123	3079	I-E	2016	ICU (Hospital 1)	Sputum
Pae113	19	3137	I-F and I-E	2010	Ward (Hospital 2)	Tissue fragment

^aICU: intensive care unit

Genome extraction, sequencing, *de novo* assembly and annotation

Genomic DNA was extracted with DNeasy Blood and Tissue Kit (Qiagen) following manufacturer's instructions. Genomic libraries were obtained with Nextera XT Library Preparation Kit (Illumina) and quantified with NEBNext Library Quant Kit

for Illumina (New England Biolabs). Libraries were sequenced in MiSeq Platform (Illumina) using MiSeq Reagent Kit V3 600 cycles (Illumina). The output reads were visualized and checked quality using FastQC (Andrews, 2010). We removed the adapters using CUTADAPT (Martin 2013) and then filtered the reads by size (17 base pairs) and quality (20) with SICKLE (Joshi & Fass, 2011) software. The remaining reads were then used to assemble the genomes using the Unicycler (Wick et al. 2017) which has the SPAdes 3.12.0 (Bankevich et al. 2012) implemented. To view and check the quality of both Unicycler and SPAdes assemblies, we use Quast (Gurevich et al. 2013) in order to decide the best assembly, where metrics were evaluated, such as largest contig, GC content, N50, number of contigs larger than 500 bp, genome size and total contigs. The assembled genomes were then annotated by PROKKA (Seemann 2014).

MLST and phylogenetic analysis

MLST was performed online, in the specific database for *P. aeruginosa* (<https://pubmlst.org/paeruginosa/>), where new alleles and STs were submitted as well (Table 1). E-burst analysis was performed with eBurst V3 (<http://eburst.mlst.net/>). Phylogenetic analysis was performed comparing the genomes of our 13 isolates to 161 genomes available in GenBank as complete genomes and chromosomes (Supplementary file 1). We created a local database with all genomes. To retrieve MLST sequences, we performed a local blast using the alleles from pubMLST as query sequences, with the following parameters: 70% identity and coverage, and 1e-25 e-value. After blast, sequences were aligned with MAFFT algorithm, and in order to choose the best possible alignment, we treated the

sequences with both GUIDANCE 2 (Sela et al. 2015) using default configurations, and GBLOCKS (Castresana 2000) using a less restrictive configuration. The evolutionary model was calculated for each individual gene, using PAUP* (Swofford 2003) and MrModelTest (Nylander 2004). Phylogeny was obtained using MrBayes software (Ronquist et al. 2012), with 5 million generations.

CRISPR/Cas confirmation and spacer analysis

CRISPR/Cas system type was confirmed for the 13 genomes, and investigated for the 174 publicly available genomes, through analysis in CRISPRCasFinder (<https://crisprcas.i2bc.paris-saclay.fr/CrisprCasFinder/index>). Spacers in CRISPR loci were submitted to analysis in CRISPRtarget (http://bioanalysis.otago.ac.nz/CRISPRTarget/crispr_analysis.html) to identify their possible origin (Supplementary file 2). CRISPR/Cas system overall structure was observed with Artemis (<https://www.sanger.ac.uk/science/tools/artemis>), and proteins of interest were confirmed through BlastP in NCBI database (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

Bacteriophage and anti-CRISPR proteins investigation

Prophages in the studied genomes were investigated using the online software PHASTER (<http://phaster.ca/>). Phages found were compared to phages related to each genome's spacer content. Anti-CRISPR proteins were investigated using the blast tool available at NCBI database (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>), using the specific DNA sequences of 14 different proteins (Supplementary file 3), previously defined (van Belkum et al. 2015;

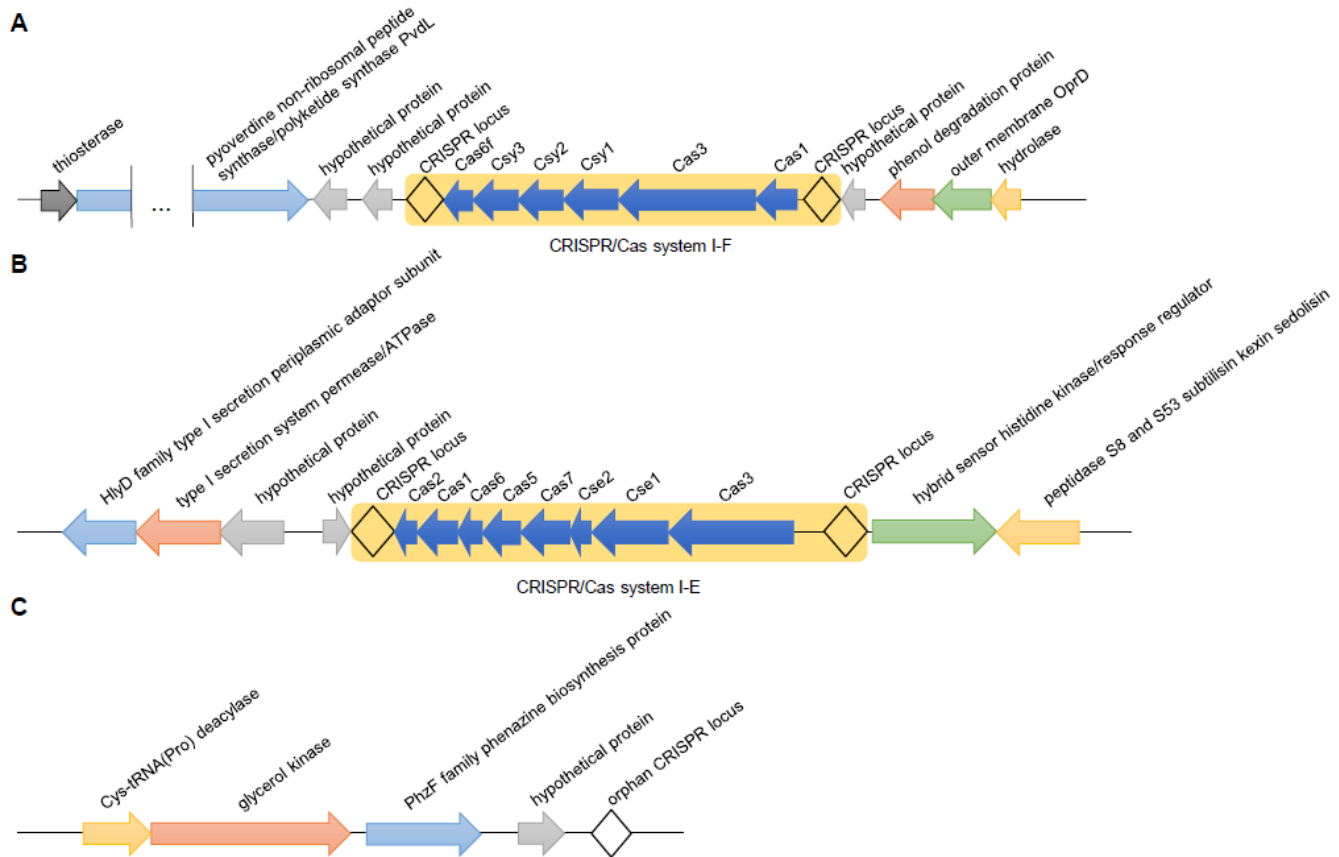
Pawluk et al. 2016, 2018).

RESULTS

Our genomes present nine different STs (Table 1), showing a non-clonal dispersion of CRISPR/Cas positive isolates in Brazil. The Brazilian isolates studied comprehended the already described STs 244, 1394, 252, 357 and 275, and the newly established STs 3078, 3079, 3080, and 3137 (Table 1). Using e-burst algorithm for MLST data, no single-locus variants or double-locus variants were found (data not shown). Therefore, no direct genetic relationship between different isolates was determined through this technique.

CRISPR/Cas confirmation and investigation has classified each genome accordingly to their CRISPR/Cas content. Aside from I-F, I-E and I-C containing genomes, some isolates were classified as containing an orphan CRISPR locus (one or more CRISPR loci without any *cas* genes associated), or absent (without CRISPR loci and *cas* genes) (Supplementary file 2). CRISPR/Cas system types I-F and I-E usually carry two CRISPR loci each, and I-C, only one locus. However, some genomes harbor an uneven number of these loci. Analyzing these orphan CRISPR loci, either if they coexist with CRISPR/Cas types or not, we noticed a common genetic background in 47 isolates. This background was named LES background, because they are present in all Liverpool Epidemic Strains (LES) of *Pseudomonas aeruginosa* studied, and these loci are here on described as LES-orphan CRISPR. The I-F, I-E and LES backgrounds are depicted in Figure 1.

183 **Figure 1. Genetic context for CRISPR/Cas in *P. aeruginosa*.** A) type I-F; B) type I-E; C) LES-
184 orphan CRISPR.



185

186

187 Regarding the 13 Brazilian isolates here described, we attempted to determine
188 the prophages in these genomes. Comparing their phage content with their
189 CRISPR's spacer content, we were able to observe spacers against phages already
190 incorporated (Table 2), contradicting the principle of CRISPR/Cas' primary function,
191 immunity. In order to find if our strains' systems are somehow inactive, we performed
192 a search for anti-CRISPR genes (Table 3), but we found that only five of our 13
193 genomes may carry inactive CRISPR/Cas systems.

194

195 **Table 2. Prophages in *P. aeruginosa* Brazilian isolates.** Prophages without a genus specified are
 196 all related to the *Pseudomonas* genus. Genomes that harbor a prophage, and also show a spacer in
 197 their CRISPR arrays against this prophage, are denoted by the color grey.

Phage	12	21	22	28	29	39	42	66	70	74	81	83	113	Genome count
D3	+							+						2
F10	+			+		+	+		+	+		+		7
F116				+		+								2
JBD18							+							1
JBD25	+													1
JBD44												+		1
MP38					+									1
Pf1	+	+		+			+		+			+	+	7
phi2											+			1
phi297							+			+				2
phi3	+			+		+						+		4
phiCTX				+	+	+	+	+	+	+		+		8
phiPSA1									+	+				2
PMG1							+							1
vB_PaeP_Tr60_Ab31				+		+	+					+		4
vB_PaeS_PM105					+									1
YMC11/02/R656	+	+	+		+		+		+	+	+		+	9
YMC11/07/P54_PAE_BP									+	+				2
<i>Sphingomonas</i> phage PAU	+													1
<i>Burkholderia</i> phage KL3	+													1
<i>Salmonella</i> phage SJ46				+										1
<i>Vibrio</i> phage VP882							+							1
<i>Mannheimia</i> phage vB_MhM_3927AP2									+					1
<i>Synechococcus</i> phage S-CBS1												+		1

198

199 **Table 3. Anti-CRISPR genes search.**

Isolates	anti-CRISPR genes														CRISPR/Cas possible function status
	AcrE1	AcrE2	AcrE3	AcrE4	AcrF1	AcrF2	AcrF3	AcrF4	AcrF5	AcrF6	AcrF7	JBD24-35	JBD24-36	MP29-29	
12															Active
21	+						+		+						Inactive I-F I-E
22															Active
28	+						+		+						Inactive I-F I-E
29															Active
39	+						+		+						Inactive I-F I-E
42											+				Inactive I-F
66															Active
70															Active
74															Active
81															Active
83	+						+		+						Inactive I-F I-E
113															Active

200

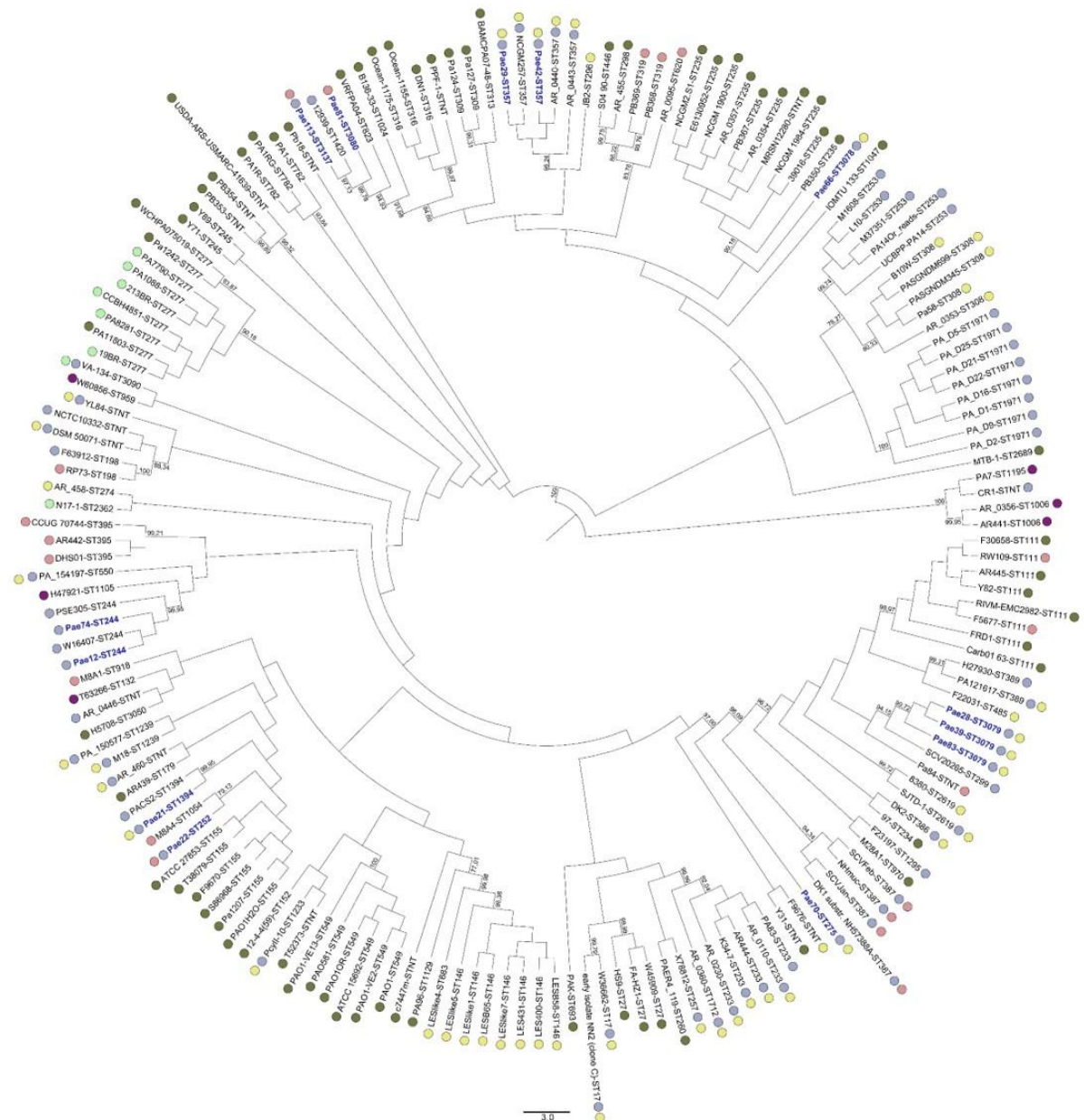
201 The phylogenetic tree, comprehending 174 genomes, show that CRISPR/Cas systems presence varies between strains and

202 does not appear to follow a clear distribution (Figure 2). Therefore, we chose to evaluate CRISPR/Cas presence by lineage, according

203 to their STs.

204

Figure 2. Phylogenetic tree of 174 genomes of *P. aeruginosa*. Strains have their ST stated next to their name. Bold blue names correspond to the 13 Brazilian isolates reported in this study. Colored circles next to the names represent each strain's CRISPR/Cas status: Olive – absent CRISPR/Cas; blue – type I-F; pink – type I-E; green – type I-C; purple – orphan CRISPR; yellow – LES-orphan CRISPR.



Strains with the same ST usually present the same CRISPR/Cas status (presence, absence or orphan CRISPR). The most noticeable difference is in ST111, where two isolates harbor type I-E, while the other six isolates are CRISPR/Cas negative. Inside ST2619, strain SJTD-1 possess type I-F and a LES-orphan CRISPR, while the other strain, 8380, shows only the LES-orphan CRISPR. A similar event has occurred in ST1394, where Pae21 presents I-F and LES-orphan CRISPR, but PACS2 appears to have lost the LES-orphan CRISPR locus.

Among our 13 isolates, two (Pae22 and Pae113) harbors CRISPR/Cas system types I-F and I-E simultaneously. This feature is rather uncommon, as it appears in only four other publicly available complete genomes: DK1, NHmuc, SCVFeb and SCVJan. However, SCVFeb and SCVJan are small colony variants recovered from murine models infected with NHmuc, and are genetically identical to NHmuc (Irvine et al. 2018). Thus, we can consider I-F/I-E to be a feature observed in four lineages between the genomes here considered: Pae22, Pae113, DK1 and NHmuc. Interestingly, the Brazilian lineages are not closely related with the other I-F/I-E lineages, and are not closely related between them. Regarding to the system's general structure in these lineages, Pae113 is the only one that presents a mutation in any of these CRISPR/Cas systems. In this case, *cas3* gene of I-E type is interrupted, close to protein's C-terminus, probably rendering interference by I-E type inactive.

DISCUSSION

The non-clonal dispersion of *Pseudomonas aeruginosa* has been established previously in globally distributed isolates (Wiehlmann et al. 2007). We attempted to

238 comprehend where the Brazilian strains of *P. aeruginosa* fit in this background, and
239 we observed our isolates scattered through the phylogenetic tree, most likely being
240 not closely related.

241 In the phylogenetic tree, aside from our strains, seven isolates belonging to
242 ST277 are Brazilian as well. They carry CRISPR/Cas system type I-C, which is a
243 very rare feature in *P. aeruginosa*. In this bacterial species, type I-C is associated
244 with a pKLC102-like integrative and conjugative element (van Belkum et al. 2015),
245 and even though none of our isolates carry this CRISPR/Cas type, it is interesting to
246 observe type I-C circulating in Brazilian strains, with the possibility to spread among
247 isolates.

248 Orphan CRISPRs have already been known to exist in *P. aeruginosa* (van
249 Belkum et al. 2015). Also, a type I-F orphan CRISPR in LES and related strains have
250 already been described (England et al. 2018). In these LES-orphan CRISPR, we do
251 observe a direct repeat similar to those in type I-F, but their genetic background is
252 completely different, and, therefore, we feel we cannot assume this CRISPR locus is
253 related to this specific CRISPR/Cas type. However, they might be functional when
254 present with a complete CRISPR/Cas system in the same genome, and only I-F
255 system type was observed in this situation. We observed a few strains with orphan
256 CRISPR loci with completely different genetic backgrounds, but they were rare, and
257 LES-orphan CRISPR was the third most frequent background in all CRISPR loci we
258 analyzed. LES are a group of well-studied epidemic strains infecting CF patients,
259 spread in Europe and North America. Considered a highly resistant and adaptable
260 epidemic clone, this adaptability is believed to be related to virulence genes in
261 prophage -rich genomic islands (Parkins et al. 2018). This intimate relationship with
262 prophages may be the reason why all LES-related strains retain no *cas* genes, even

263 though they still carry an orphan CRISPR.

264 In previous works, types I-F and I-E were explicitly described once (Essoh et
265 al. 2013) in three CF related isolates, where CRISPR/Cas typing was performed
266 through PCR and, therefore, we could not compare their results to ours. The findings
267 of one other previous study (van Belkum et al. 2015) also show 17 draft genomes
268 (therefore, not considered here for phylogeny) containing I-F and I-E CRISPR/Cas
269 types, accounting for 2,5% of their analyzed strains. Gathering these data, we can
270 notice all strains carrying this feature are derived from clinical sources, with one
271 exception of a strain from well water (Supplementary file 4).

272 DK1 and NHmuc present the same ST and are placed adjacent in the
273 phylogenetic tree, but Pae22 and Pae113 are distant, showing the I-F/I-E feature
274 distant in the phylogenetic tree. Also, these isolates were derived from two different
275 hospitals, with a five-year period between them. Aside from Pae113, we have no
276 reason to believe that only one CRISPR/Cas type is being expressed or is fully active
277 in these lineages, but we are approaching this subject only in structural terms.
278 Functional studies are needed to better comprehend what this feature can mean to
279 the bacterial fitness or to the bacterium-bacteriophage relationship.

280 Finding lysogenic phages in our genomes and spacers against these same
281 phages in the same genomes was not expected, and suggests no immunity function.
282 We have no reason to infer these systems are inactive, since their system's structure
283 appears to be intact and not all of our isolates harbor anti-CRISPR genes. This
284 observation contradicts the foundation of CRISPR/Cas immunity, the primary
285 objective of this system.

286 In *P. aeruginosa*, CRISPR/Cas has been shown to be active in a controlled

environment, against temperate phages (Cady et al. 2012). Studies, also in *P. aeruginosa*, have already suggested this system is the preferred mechanism of defense only in nutrient-poor media (Westra et al. 2015). It has been proved that CRISPR/Cas interferes with the expression of phage DMS3 products, altering the biofilm production (Zegans et al. 2009). Also, CRISPR/Cas appears to regulate the expression of *lasR*, a gene related to the control of quorum-sensing systems (Li et al. 2016). Observing this possible inactivity of interference in our genomes is another evidence that raises the question if the CRISPR/Cas system, in *P. aeruginosa*, would mainly be a machinery to better regulate the expression of phage and internal proteins.

SUPPLEMENTARY DATA

Supplementary file 1. Information on the 161 publicly available genomes. Information comprehend access numbers, ST, CRISPR/Cas status, and epidemiologic data.

Supplementary file 2. Non-redundant list of targets for spacers in CRISPR loci in the 13 brazilian clinical strains. Results obtained from CRISPRtarget online software.

Supplementary file 3. Anti-CRISPR gene sequences investigated.

Supplementary file 4. Information on all discussed I-F/I-E genomes.

ACKNOWLEDGMENTS

The authors thank the Núcleo de Plataformas Tecnológicas (NPT) from Aggeu Magalhães Institute for use of the sequencing platform. This work was supported by

Conselho Nacional Conselho Nacional de Desenvolvimento Científico e Tecnológico
[Grant Number 404016/2016-2]; and Fundação de Amparo à Ciência e Tecnologia
do Estado de Pernambuco [Grant Number: APQ-0836-2.02/16].

LITERATURE CITED

Andrews S. 2010. FastQC: A quality control tool for high throughput sequence
data. Babraham Bioinformatics.

Azam MW, Khan AU. 2018. Updates on the pathogenicity status of
Pseudomonas aeruginosa. Drug Discov. Today. 00:1–10. doi:
10.1016/j.drudis.2018.07.003.

Bankevich A et al. 2012. SPAdes: a new genome assembly algorithm and its
applications to single-cell sequencing. J. Comput. Biol. 19:455–477. doi:
10.1089/cmb.2012.0021.

van Belkum A et al. 2015. Phylogenetic Distribution of CRISPR-Cas Systems
in Antibiotic-Resistant *Pseudomonas aeruginosa*. MBio. 6:1–13. doi:
10.1128/mBio.01796-15.

Cady KC et al. 2011. Prevalence, conservation and functional analysis of
Yersinia and Escherichia CRISPR regions in clinical *Pseudomonas aeruginosa*
isolates. Microbiology. 157:430–437. doi: 10.1099/mic.0.045732-0.

Cady KC, Bondy-Denomy J, Heussler GE, Davidson AR, O'Toole GA. 2012.
The CRISPR/Cas adaptive immune system of *Pseudomonas aeruginosa* mediates
resistance to naturally occurring and engineered phages. J. Bacteriol. 194:5728–
5738. doi: 10.1128/JB.01184-12.

- 333 Castresana J. 2000. Selection of conserved blocks from multiple alignments
334 for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17:540–552.
- 335 England WE, Kim T, Whitaker RJ. 2018. Metapopulation Structure of CRISPR-
336 Cas Immunity in *Pseudomonas aeruginosa* and Its Viruses. *mSystems*. 3:e00075-18.
- 337 Essoh C et al. 2013. The Susceptibility of *Pseudomonas aeruginosa* Strains
338 from Cystic Fibrosis Patients to Bacteriophages. *PLoS One*. 8. doi:
339 10.1371/journal.pone.0060575.
- 340 Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: Quality
341 assessment tool for genome assemblies. *Bioinformatics*. 29:1072–1075. doi:
342 10.1093/bioinformatics/btt086.
- 343 Hille F et al. 2018. The Biology of CRISPR-Cas: Backward and Forward. *Cell*.
344 172:1239–1259. doi: 10.1016/j.cell.2017.11.032.
- 345 Irvine S et al. 2018. Genomic inversion drives small colony variant formation
346 and increased virulence in *P. aeruginosa*. *bioRxiv*.
347 <http://biorxiv.org/content/early/2018/07/09/356386.abstract>.
- 348 Joshi N, Fass J. 2011. Sickel - A windowed adaptive trimming tool for FASTQ
349 files using quality. (Version 1.33).
- 350 Koonin E V., Makarova KS, Zhang F. 2017. Diversity, classification and
351 evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol.* 37:67–78. doi:
352 10.1016/j.mib.2017.05.008.
- 353 Li R et al. 2016. Type I CRISPR-Cas targets endogenous genes and regulates
354 virulence to evade mammalian host immunity. *Cell Res.* 26:1273–1287. doi:
355 10.3389/fmicb.2017.00319.

- 356 Martin M. 2013. Cutadapt removes adapter sequences from high-throughput
357 sequencing reads kenkyuhojokin gan rinsho kenkyu jigyo. EMBnet.journal. 17:10–
358 12. doi: 10.14806/ej.17.1.200.
- 359 Nylander JAA. 2004. MrModeltest v2. program distributed by the author.
360 Evolutionary biology centre, Uppsala University.
- 361 van der Oost J, Westra ER, Jackson RN, Wiedenheft B. 2014. Unravelling the
362 structural and mechanistic basis of CRISPR-Cas systems. Nat. Rev. Microbiol.
363 12:479–92. doi: 10.1038/nrmicro3279.
- 364 Parkins MD, Somayaji R, Waters J. 2018. Epidemiology , Biology , and Impact
365 of Clonal *Pseudomonas aeruginosa* Infections in Cystic Fibrosis. Clin. Microbiol. Rev.
366 31:1–38.
- 367 Pawluk A et al. 2016. Inactivation of CRISPR-Cas systems by anti-CRISPR
368 proteins in diverse bacterial species. Nat. Microbiol. 1:1–6. doi:
369 10.1038/nmicrobiol.2016.85.
- 370 Pawluk A, Davidson AR, Maxwell KL. 2018. Anti-CRISPR: Discovery,
371 mechanism and function. Nat. Rev. Microbiol. 16:12–17. doi:
372 10.1038/nrmicro.2017.120.
- 373 Pereira RPA, Peplies J, Mushi D, Brettar I, Höfle MG. 2018. Pseudomonas-
374 Specific NGS Assay Provides Insight Into Abundance and Dynamics of
375 Pseudomonas Species Including *P. aeruginosa* in a Cooling Tower. Front. Microbiol.
376 9:1–15. doi: 10.3389/fmicb.2018.01958.
- 377 Ronquist F et al. 2012. MrBayes 3.2: efficient bayesian phylogenetic inference
378 and model choice across a large model space. Syst. Biol. 61:539–542. doi:
379 10.1093/sysbio/sys029.

- 380 Seemann T. 2014. Prokka: Rapid prokaryotic genome annotation.
381 Bioinformatics. 30:2068–2069. doi: 10.1093/bioinformatics/btu153.
- 382 Sela I, Ashkenazy H, Katoh K, Pupko T. 2015. GUIDANCE2: Accurate
383 detection of unreliable alignment regions accounting for the uncertainty of multiple
384 parameters. Nucleic Acids Res. 43:W7–W14. doi: 10.1093/nar/gkv318.
- 385 Silby MW, Winstanley C, Godfrey SAC, Levy SB, Jackson RW. 2011.
386 Pseudomonas genomes: Diverse and adaptable. FEMS Microbiol. Rev. 35:652–680.
387 doi: 10.1111/j.1574-6976.2011.00269.x.
- 388 Swofford DL. 2003. PAUP*: phylogenetic analysis using parsimony, version
389 4.0b10.
- 390 Valot B et al. 2015. What it takes to be a *Pseudomonas aeruginosa*? The core
391 genome of the opportunistic pathogen updated. PLoS One. 10:1–15. doi:
392 10.1371/journal.pone.0126468.
- 393 Westra ER et al. 2015. Parasite Exposure Drives Selective Evolution of
394 Constitutive versus Inducible Defense. Curr. Biol. 25:1043–1049. doi:
395 10.1016/j.cub.2015.01.065.
- 396 Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: Resolving bacterial
397 genome assemblies from short and long sequencing reads. PLoS Comput. Biol.
398 13:1–22. doi: 10.1371/journal.pcbi.1005595.
- 399 Wiehlmann L et al. 2007. Population structure of *Pseudomonas aeruginosa*.
400 Proc Natl Acad Sci U S A. 104:8101–8106. doi: 10.1073/pnas.0609213104.
- 401 Zegans ME et al. 2009. Interaction between bacteriophage DMS3 and host
402 CRISPR region inhibits group behaviors of *Pseudomonas aeruginosa*. J. Bacteriol.

403 91:210–219. doi: 10.1128/JB.00797-08.

404

405 TABLES AND FIGURES CAPTIONS

406 **Table 1. Characteristics of *Pseudomonas aeruginosa* Brazilian isolates.**

407 **Fig 1. Genetic context for CRISPR/Cas in *P. aeruginosa*.** A) type I-F; B) type I-E;
408 C) LES-orphan CRISPR.

409 **Table 2. Prophages in *P. aeruginosa* Brazilian isolates.** Prophages without a
410 genus specified are all related to the *Pseudomonas* genus. Genomes that harbor a
411 prophage and also show a spacer in their CRISPR arrays against this prophage, are
412 denoted by the color grey.

413 **Table 3. Anti-CRISPR genes search.**

414 **Fig 2. Phylogenetic tree of 174 genomes of *P. aeruginosa*.** Strains have their ST
415 stated next to their name. Bold blue names correspond to the 13 Brazilian isolates
416 reported in this study. Colored circles next to the names represent each strain's
417 CRISPR/Cas status: Olive – absent CRISPR/Cas; blue – type I-F; pink – type I-E;
418 green – type I-C; purple – orphan CRISPR; yellow – LES-orphan CRISPR.

419

5 DISCUSSÃO GERAL

O primeiro estudo direcionado ao sistema CRISPR/Cas de *P. aeruginosa* foi realizado em 2009 (ZEGANS et al., 2009), e desde então, informação acerca deste microrganismo tem crescido bastante. Nessa espécie, o sistema do tipo I-F apresenta algumas particularidades, as quais tornaram *P. aeruginosa*, em especial a cepa UCBPP-PA14, um dos mais bem estudados microrganismos para o estudo do mecanismo e funcionamento do sistema. No entanto, apesar do grande número de trabalhos com a cepa UCBPP-PA14, considerada a referência da espécie, poucos estudos têm utilizado isolados clínicos como objeto de estudo (CADY et al., 2011; ESSOH et al., 2013; van BELKUM et al., 2015; ENGLAND et al., 2018).

Estudos têm utilizado genomas brasileiros, porém as informações sobre estes isolados acabam perdidas diante do volume de informações discutido nestes trabalhos. Van Belkum e colaboradores, em 2015, descreveram o sistema CRISPR/Cas tipo I-C em *P. aeruginosa* pela primeira vez, e dos 20 isolados que possuem este sistema, três são brasileiros (van BELKUM et al., 2015). Em nossas análises, dos 161 genomas disponíveis no banco de dados do NCBI, sete correspondem a isolados brasileiros, e seis possuem o sistema do tipo I-C. Como descrito por Van Belkum e colaboradores, este tipo de sistema está associado a um ICE, e, portanto, é passível de ser compartilhado entre isolados diferentes. A circulação deste MGE em isolados brasileiros sugere que, futuramente, poderemos observar uma maior frequência do sistema tipo I-C em isolados brasileiros de *P. aeruginosa*.

A transferência horizontal é o principal mecanismo pelo qual bactérias evoluem. Enquanto, por um lado, o sistema CRISPR/Cas provê bactérias com um

mecanismo de proteção capaz de se adaptar às ameaças do meio, por outro, ele impede a aquisição de plasmídeos conjugativos, elementos integrativos/conjugativos, ou DNA livre no meio (JIANG et al., 2013). A perda de elementos do sistema CRISPR/Cas, ou dele completo, permite a incorporação de novos elementos genéticos, os quais podem trazer benefícios para a célula bacteriana. Bactérias patogênicas, em particular, são beneficiadas pela aquisição de plasmídeos contendo genes de resistência, ou fagos que carregam genes de virulência, pois aumentam as chances de uma infecção bem-sucedida. A ausência deste sistema em espécies de *Campylobacter*, *Enterococcus*, *Streptococcus* e *Staphylococcus* pode ser um retrato da importância que MGEs possuem para a evolução destes patógenos (JIANG et al., 2013; BARRANGOU, 2015). Apesar da sua ausência em 73 (56%) dos nossos 130 isolados clínicos, o que corrobora a hipótese de que agentes patogênicos que possuem o sistema CRISPR/Cas apresentariam uma desvantagem adaptativa, Gophna e colaboradores demonstraram que, em escalas evolutivas, o impacto do sistema CRISPR/Cas não é detectável (GOPHNA et al., 2015).

A diversidade dos espaçadores encontrados e a incorporação diferencial observada ao final dos locos CRISPR permite sugerir que estes locos são ativos e potencialmente de evolução rápida. A relação dos espaçadores com fagos, plasmídeos e ilhas genômicas de patogenicidade evidenciam a importância que este sistema pode ter na aquisição de novos determinantes de virulência e resistência (JIANG et al., 2013).

A presença dos tipos I-F e I-E concomitantemente nos genomas não é uma característica observada com frequência. De fato, esta característica só foi descrita uma vez, por Esoh e colaboradores em 2013, em apenas três isolados (ESSOH et

al., 2013). Apesar de não ser mencionado no trabalho de Van Belkum e colaboradores, em 2015, estudando o material suplementar fornecido, são observados 23 genomas com esta mesma característica (van BELKUM et al., 2015). Neste trabalho, que compreendeu 672 genomas, 17 isolados correspondem a 2,5% de frequência. Entre os nossos isolados, esta característica foi presente em dois (Pae22 e Pae113), o que corresponde a 1,5% dos 130 isolados analisados.

Espécies bacterianas que não possuem o sistema CRISPR/Cas distribuído uniformemente entre seus isolados, não são boas candidatas para a aplicação de tipagem molecular através do conteúdo espaçador do loco CRISPR. Desta maneira, esta técnica nunca foi padronizada para *P. aeruginosa*. Comparando os locos CRISPR dos nossos isolados com o intuito de observar padrões de incorporação, seria possível assumir que estes isolados fossem geneticamente mais próximos. Esta hipótese se manteve para quase todos os isolados cujos genomas foram sequenciados, onde isolados correspondentes aos padrões D e E apresentam os mesmos STs e localizam-se próximos na árvore filogenética. O mesmo foi verdadeiro para isolados dos padrões B (Pae29) e B/E (Pae42). A exceção, porém, foram os isolados que apresentaram o padrão de incorporação A Pae66 e Pae113, que apresentam STs diferentes e estão localizados distantes na árvore filogenética. Diante destes dados, a inferência de relação genética entre isolados com padrões semelhantes de incorporação de espaçadores no loco CRISPR não parece ser uma ferramenta completamente confiável.

Com relação à função do sistema, esta permanece relativamente desconhecida em *P. aeruginosa*. Embora não se conheça a expressão e atividade do sistema do tipo I-E, o sistema tipo I-F é bem estudado e possui relatos controversos quanto à sua eficiência. Zegans e colaboradores, em 2009, ao

estudarem o mecanismo da interação do fago DMS3 com o sistema CRISPR/Cas para inibição da produção de biofilme, observaram que o fago em questão se encontrava como profago no genoma da UCBPP-PA14, apesar desta cepa possuir ao menos um espaçador contra este fago, sugerindo que a interferência do sistema na aquisição de MGEs não ocorre em *P. aeruginosa* (ZEGANS et al., 2009). No entanto, Cady e colaboradores, em 2011, descreveram a produção de crRNA por isolados de *P. aeruginosa*, e em 2012, a eficiência do sistema contra fagos temperados em condições laboratoriais (CADY et al., 2011; CADY et al., 2012). Essoh e colaboradores, em 2013, descreveram que o sistema não parece ser efetivo contra fagos líticos (ESSOH et al., 2013), e Westra e colaboradores, em 2015, identificaram que este sistema pode apenas proteger contra MGEs em situações nas quais o meio é pobre em nutrientes, enquanto em meios ricos, outros mecanismos de defesa são preferíveis (WESTRA et al., 2015). Em 2016, Li e colaboradores detectaram a regulação da expressão do gene *lasR* através de um espaçador específico contra este gene constitutivo (LI et al., 2016).

Além do observado por Li e colaboradores, funções relacionadas à regulação da expressão de genes já foi descrita em *Pyrococcus furiosus*, *Campylobacter jejuni*, *Francisella novicida*, e *Listeria monocytogenes*. Porém, nenhuma destas espécies possui sistemas do tipo I (WESTRA et al., 2014).

A observação de fagos incorporados aos genomas dos nossos isolados, onde também há contra estes mesmos fagos, representa mais uma evidência para a ineficácia do sistema CRISPR/Cas como um mecanismo de imunidade adaptativa em *P. aeruginosa*, e sua possível função na regulação da expressão de proteínas fágicas. Outra evidência que suporta esta hipótese é o espaçador encontrado em um dos nossos isolados, contra o gene de uma proteína do metabolismo normal de *P.*

aeruginosa, a coproporfirinogênio oxidase III independente de oxigênio, cuja expressão poderia estar sendo regulada de maneira semelhante à regulação do gene *lasR*. A presença de genes codificantes de proteínas anti-CRISPR em apenas cinco dos nossos 13 isolados não são o suficiente para que possamos definir que estes 13 sistemas estão inativos, e, portanto, justificar a presença dos fagos lisogênicos nestes genomas.

As três proteínas anti-CRISPR encontradas em quatro genomas dos nossos isolados (Pae21, Pae28, Pae29 e Pae83) correspondem ao fago JBD5, o que indica que este estaria lisogênico nestes genomas. Avaliando-se a organização dos espaçadores nestes isolados, observa-se um espaçador contra este mesmo fago. Assim como descrito por Zegans e colaboradores, em 2009, há a incorporação bem sucedida do fago JBD5, mesmo havendo espaçadores específicos contra o mesmo (ZEGANS et al., 2009). No entanto, nestes casos, a incorporação seria justificada pela presença das proteínas anti-CRISPR, que estariam inativando o sistema.

Segundo Borges e colaboradores, as proteínas AcrE1, AcrF3 e AcrF5 atuam na etapa de interferência do sistema, o que é consistente com nossos dados, visto que apesar do sistema supostamente inativo, há a incorporação diferencial e polarizada nos locos CRISPR destes isolados. A proteína AcrF3, porém, também parece estar implicada na inativação dos mecanismos de aquisição de novos espaçadores, o que impediria a incorporação diferencial observada (BORGES et al., 2017). Uma hipótese que podemos sugerir para esta observação é a incorporação do fago JBD5 apenas após a aquisição diferencial de espaçadores no final do loco, que não seria impedida pela existência de um espaçador contra o próprio fago devido à suposta função do sistema CRISPR/Cas em mecanismos não relacionados à imunidade adaptativa em *P. aeruginosa*.

O presente estudo teve por objetivo conhecer o panorama do sistema CRISPR/Cas em isolados clínicos brasileiros de *P. aeruginosa*. Os resultados obtidos permitiram: compreender a frequência deste sistema; a diversidade de espaçadores nos isolados; a relação destes com fagos, plasmídeos e ilhas genômicas; inferir acerca da funcionalidade deste sistema de acordo com achados estruturais; e observar a característica não-clonal da distribuição do sistema CRISPR/Cas nessa espécie bacteriana.

6 CONCLUSÕES

- a) Como hipótese para a função em *P. aeruginosa*, levando-se em consideração o espaçador encontrado contra uma proteína do metabolismo da própria bactéria, bem como os espaçadores e profagos observados nos genomas, é que este sistema possivelmente atue apenas como um regulador transcricional de produtos bacterianos e virais.
- b) Levando-se em conta a quantidade de espaçadores diferentes, a organização destes nos locos CRISPR, e os rearranjos observados, este sistema em isolados brasileiros é diverso, adaptável e possivelmente está ativamente incorporando novos espaçadores;
- c) A presença de dois sistemas diferentes (I-F e I-E) em um mesmo isolado, apesar de ser encontrado em outros poucos genomas, ainda não havia sido encontrado em isolados brasileiros, e o estudo das consequências desta característica ainda se fazem necessários;
- d) A ausência do sistema CRISPR/Cas em parte dos isolados sugere uma possível desvantagem evolutiva na manutenção deste sistema em algumas linhagens;
- e) A análise da organização dos espaçadores pode corresponder ao perfil ST e à filogenia, mas não parece ser o suficiente para estabelecer relações genéticas com sucesso;

REFERÊNCIAS

- ASIF, M.; ALVI, I. A.; REHMAN, S. U. Infection and Drug Resistance Dovepress Insight into *Acinetobacter baumannii*: pathogenesis, global resistance, mechanisms of resistance, treatment options, and alternative modalities. **Infect. Drug Resist.**, Londres, v. 11, p. 1249–1260, 2018.
- AZAM, M. W.; KHAN, A. U. Updates on the pathogenicity status of *Pseudomonas aeruginosa*. **Drug Discov. Today**, Oxford, v. 24, n. 1, p. 350-359, 2019.
- BARRANGOU, R. The roles of CRISPR-Cas systems in adaptive immunity and beyond. **Curr. Opin. in Immunol**, Filadélfia, v. 32, p. 36–41, 2015.
- BARRANGOU, R.; DUDLEY, E. G. CRISPR-Based Typing and Next-Generation Tracking Technologies. **Annu. Rev. Food Sci. Technol.**, Palo Alto, v. 7, p. 395-411, 2016.
- BARRANGOU, R. et al. CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. **Science**, Washington, v. 315, n. 5819, p. 1709–1712, 2007.
- BISWAS, A. et al. CRISPRTarget - Bioinformatic prediction and analysis of crRNA targets. **RNA Biol.**, Oxford, v. 10, n. 5, p. 817–827, 2013.
- BOLOTIN, A. et al. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. **Microbiology**, Reading, v. 151, p. 2551–2561, 2005.
- BONDY-DENOMY, J.; DAVIDSON, A.R. To acquire or resist: The complex biological effects of CRISPR-Cas systems. **Trends Microbiol.**, Cambridge, v. 22, n. 4, p. 218–225, 2014.
- BONDY-DENOMY, J. et al. Prophages mediate defense against phage infection through diverse mechanisms. **ISME J.**, Wageningen, v. 10, n. 12, p. 2854–2866, 2016.
- BORGES, A. L.; DAVIDSON, A. R.; BONDY-DENOMY, J. The Discovery, Mechanisms, and Evolutionary Impact of Anti-CRISPRs. **Annu. Rev. Virol.**, Palo Alto, v. 4, p. 37-59, 2017.
- BURMISTRZ, M.; PYRC, K. CRISPR-Cas systems in prokaryotes. **Polish J. Microbiol.**, Warsaw, v. 64, p. 193–202, 2015.
- CADY, K. C. et al. The CRISPR/Cas adaptive immune system of *Pseudomonas aeruginosa* mediates resistance to naturally occurring and engineered phages. **J. Bacteriol.**, Washington, v. 194, p. 5728–5738, 2012.
- CADY, K. C.; O'TOOLE, G. A. Non-identity-mediated CRISPR-bacteriophage

interaction mediated via the Csy and Cas3 proteins. **J. Bacteriol.**, Washington, v. 193, p. 3433–3445, 2011.

CADY, K. C. et al. Prevalence, conservation and functional analysis of *Yersinia* and *Escherichia* CRISPR regions in clinical *Pseudomonas aeruginosa* isolates. **Microbiology**, Reading, v. 157, p. 430–437, 2011.

CRISPR *web server*. Disponível em: <<http://crispr.i2bc.paris-saclay.fr/>>. Acesso em 21 nov 2018.

DY, R. L. et al. Remarkable Mechanisms in Microbes to Resist Phage Infections. **Annu. Rev. Virol.**, Palo Alto, v. 1, p. 307–331, 2014.

ENGLAND, W. E.; KIM, T.; WHITAKER, R. J. Metapopulation Structure of CRISPR-Cas Immunity in *Pseudomonas aeruginosa* and Its Viruses. **mSystems**, Washington, v. 3, n. 5, p. e00075-18, 2018.

ESSO, H. C. et al. The Susceptibility of *Pseudomonas aeruginosa* Strains from Cystic Fibrosis Patients to Bacteriophages. **PLoS One**, São Francisco, v. 8, n. 4, 2013. Disponível em <<https://doi.org/10.1371/journal.pone.0060575>>. Acesso em 22 mai 2019.

FENG, W. et al. Epidemiology and resistance characteristics of *Pseudomonas aeruginosa* isolates from the respiratory department of a hospital in China. **J. Glob. Antimicrob. Resist.**, Londres, v. 8, p. 142–147, 2017.

GELLATLY, S. L.; HANCOCK, R. E. W. *Pseudomonas aeruginosa*: New insights into pathogenesis and host defenses. **Pathog. Dis.**, Oxford, v. 67, n. 3, p. 159–173, 2013.

GLEITZSCH, D. et al. PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. **RNA Biol.**, Oxford, v. 16, n. 4, p. 504–517, 2019.

GOPHNA, U. et al. No evidence of inhibition of horizontal gene transfer by CRISPR-Cas on evolutionary timescales. **ISME J.**, Wageningen, v. 9, p. 2021–2027, 2015.

GORSKI, S. A.; VOGEL, J.; DOUDNA, J. A. RNA-based recognition and targeting: Sowing the seeds of specificity. **Nat. Rev. Mol. Cell. Biol.**, Londres, v. 18, n. 4, p. 215–228, 2017.

GROENEN, P. M. A. et al. Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method. **Mol. Microbiol.**, Oxford, v. 10, p. 1057–1065, 1993.

HATOUM-ASLAN, A.; MARRAFFINI, L. A. Impact of CRISPR immunity on the emergence and virulence of bacterial pathogens. **Curr. Opin. Microbiol.**, Filadélfia, v. 17, p. 82–90, 2014.

HIROSE, T.; MISHIMA, Y.; TOMARI, Y. Elements and machinery of non-coding

RNAs: Toward their taxonomy. **EMBO Rep.**, Oxford, v. 15, p. 489–507, 2014.

ISHINO, Y. et al (1987) Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. **J. Bacteriol.**, Washington, v. 169, p. 5429–5433, 1987.

JACKSON, S. A. et al CRISPR-Cas: Adapting to change. **Science**, Washington, v. 356, n. 6333, 2017. Disponível em <<https://science.sciencemag.org/content/356/6333/eaal5056>>. Acesso em 22 mai 2019.

JANSEN, R. et al. Identification of genes that are associated with DNA repeats in prokaryotes. **Mol. Microbiol.**, Oxford, v. 43, p. 1565–1575, 2002.

JIANG, W. et al. Dealing with the Evolutionary Downside of CRISPR Immunity: Bacteria and Beneficial Plasmids. **PLoS Genet.**, São Francisco, v. 9, n. 9. Disponível em <<https://doi.org/10.1371/journal.pgen.1003844>>. Acesso em 22 mai 2019.

JOLLEY, K. A.; MAIDEN, M. C. Using MLST to study bacterial variation: prospects in the genomic era. **Future Microbiol.**, Londres, v. 9, p. 623–630, 2014.

KARIMI, Z. et al. Bacterial CRISPR Regions: General Features and their Potential for Epidemiological Molecular Typing Studies. **Open Microbiol. J.**, Hilversum, v. 12, p. 59–70, 2018.

KHALIFA, L. et al Targeting *Enterococcus faecalis* biofilms with phage therapy. **Appl. Environ. Microbiol.**, Washington, v. 81, p. 2696–2705, 2015.

KOONIN, E. V.; MAKAROVA, K. S.; ZHANG, F. Diversity, classification and evolution of CRISPR-Cas systems. **Curr. Opin. Microbiol.**, Filadélfia, 37:67–78, 2017.

KOONIN, E. V.; ZHANG, F. Coupling immunity and programmed cell suicide in prokaryotes: Life-or-death choices. **BioEssays**, Cambridge, v. 39, p. 1–9, 2017.

LE, S. et al Chromosomal DNA deletion confers phage resistance to *Pseudomonas aeruginosa*. **Sci. Rep.**, Heidelberg, v. 4, 2014. Disponível em <<https://www.nature.com/articles/srep04738>>. Acesso em 22 mai 2019.

KRUPOVIC, M. et al. Recent mobility of casposons, self-synthesizing transposons at the origin of the CRISPR-Cas immunity. **Gen. Biol. Evol.**, Oxford, v. 8, n. 2, p. 375–386, 2016.

LEENAY, R. T.; BEISEL, C. L. Deciphering, Communicating, and Engineering the CRISPR PAM. **J. Mol. Biol.**, Londres, v. 429, p. 177–191, 2017.

LI, R. et al. Type I CRISPR-Cas targets endogenous genes and regulates virulence to evade mammalian host immunity. **Cell Res.**, Pequim, v. 26, p. 1273–1287, 2016.

LOC-CARRILLO, C.; ABEDON, S. T. Pros and cons of phage therapy. **Bacteriophage**, Oxford, v. 1, p. 111–114, 2011.

MA, Y.; ZHANG, L.; HUANG, X. Genome modification by CRISPR/Cas9. **FEBS J.**, Oxford, v. 281, p. 5186–5193, 2014.

Madigan, M. T. et al. **Microbiologia de Brock**. 12ª edição. Porto Alegre, Artmed, 2010.

MAIER, L. K.; DYALL-SMITH, M.; MARCHFELDER, A. The Adaptive Immune System of *Haloferax volcanii*. **Life**, Chicago, v. 5, p. 521–537, 2015.

MAKAROVA, K. S. et al. A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. **Nucleic Acids Res.**, Londres, v. 30, p. 482–496, 2002.

MAKAROVA, K. S. et al. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. **Biol. Direct**, Heidelberg, v. 1, p. 7, 2006.

MAKAROVA, K. S. et al. Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. **Biol. Direct**, Heidelberg, v. 6, p. 1–27, 2011a.

MAKAROVA, K. S. et al. Evolution and classification of the CRISPR–Cas systems. **Nat. Rev. Microbiol.**, v. 9, p. 467–477, 2011.b.

MAKAROVA, K. S. et al. An updated evolutionary classification of CRISPR-Cas systems. **Nat. Rev. Microbiol.**, Londres, v. 13, p. 722–736, 2015.

MOJICA, F. J. M. et al. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.*, Nova Iorque, v. 60, p.174–182, 2005.

MOJICA, F. J. M. et al. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. **Mol. Microbiol.**, Oxford, v. 36, p. 244–246, 2000.

MULET, X. et al. Biological markers of *Pseudomonas aeruginosa* epidemic high-risk clones. **Antimicrob. Agents Chemother.**, Washington, v. 57, p. 5527–5535, 2013.

MURUGAN, K. et al. The Revolution Continues: Newly Discovered Systems Expand the CRISPR-Cas Toolkit. **Mol. Cell**, Cambridge, v. 68, p. 15–25, 2017.

O'CONNELL, M. R. Molecular Mechanisms of RNA Targeting by Cas13-containing Type VI CRISPR–Cas Systems. **J. Mol. Biol.**, Londres, v. 431, n. 1, p. 66-87, 2019.

OMS – Organização Mundial de Saúde. WHO publishes list of bacteria for which new antibiotics are urgently needed. Disponível em <<http://www.who.int/news-room/detail/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed>>. Acesso em 21 nov 2018.

PARKINS, M. D.; SOMAYAJI, R.; WATERS, J. Epidemiology , Biology , and Impact of Clonal *Pseudomonas aeruginosa* Infections in Cystic Fibrosis. **Clin. Microbiol. Rev.**, Washington, v. 31, p. 1–38, 2018.

PAWLUK, A.; DAVIDSON, A. R.; MAXWELL, K. L. Anti-CRISPR: Discovery, mechanism and function. **Nat. Rev. Microbiol.**, Londres, v. 16, p. 12–17, 2018.

PAWLUK, A. et al. Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species. **Nat. Microbiol.**, Londres, v. 1, p. 1–6, 2016.

POTRON, A.; POIREL, L.; NORDMANN, P. Emerging broad-spectrum resistance in *Pseudomonas aeruginosa* and *Acinetobacter baumannii*: Mechanisms and epidemiology. *Int. J. Antimicrob. Agents.*, Amsterdam, v. 45, p. 568–585, 2015.

POURCEL, C.; SALVIGNOL, G.; VERGNAUD, G. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. **Microbiology**, Reading, v. 151, p. 653–663, 2005.

RICHTER, C.; CHANG, J. T.; FINERAN, P. C. Function and regulation of clustered regularly interspaced short palindromic repeats (CRISPR) / CRISPR associated (Cas) systems. **Viruses**, Basel, v. 4, p.2291–2311, 2012.

SHARIAT, N.; DUDLEY, E. G. CRISPRs: Molecular Signatures Used for Pathogen Subtyping. **Appl. Environ. Microbiol.**, Washington, v. 80, p. 430–439, 2014.

SHMAKOV, S. et al. Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. **Mol. Cell**, Cambridge, v. 60, p. 385–397, 2015.

SILBY, M. W. et al. *Pseudomonas* genomes: Diverse and adaptable. **FEMS Microbiol. Rev.**, Amsterdam, v. 35, p. 652–680, 2011.

SONNLEITNER, E.; ROMEO, A.; BLÄSI, U. Small regulatory RNAs in *Pseudomonas aeruginosa*. **RNA Biol.**, Oxford, v. 9, p. 364–371, 2012.

SOREK, R.; KUNIN, V.; HUGENHOLTZ, P. CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. **Nat. Rev. Microbiol.**, Londres, v. 6, p. 181–186, 2008.

VALOT, B. et al. What it takes to be a *Pseudomonas aeruginosa*? The core genome of the opportunistic pathogen updated. **PLoS One**, São Francisco, v. 10, 2005. Disponível em <<https://doi.org/10.1371/journal.pone.0126468>>. Acesso em 22 mai 2019.

van BELKUM, A. et al. Phylogenetic Distribution of CRISPR-Cas Systems in Antibiotic-Resistant *Pseudomonas aeruginosa*. **MBio**, Washington, v. 6, p. 1–13, 2015.

van DER OOST, J. et al. CRISPR-based adaptive and heritable immunity in

prokaryotes. **Trends Biochem. Sci.**, Cambridge, v. 34, p. 401–407, 2009.

van DER OOST, J. et al. Unravelling the structural and mechanistic basis of CRISPR-Cas systems. **Nat. Rev. Microbiol.**, Londres, v. 12, p. 479–492, 2014.

van DER ZEE, A. et al. Spread of Carbapenem Resistance by Transposition and Conjugation Among *Pseudomonas aeruginosa*. **Front. Microbiol.**, Lausanne, v. 9, p. 2057, 2018.

VIEIRA, G. V. et al. Visão Geral do Mecanismo Básico de Ação. In: PEREIRA, T. C. (Org.). **Introdução à técnica de CRISPR**. 1ª edição. Ribeirão Preto, Sociedade Brasileira de Genética, 2016. p. 39-50.

WATERS, E. M. et al. Phage therapy is highly effective against chronic lung infections with *Pseudomonas aeruginosa*. **Thorax**, Londres, v. 72, p. 666–667, 2017.

WESTRA, E. R.; BUCKLING, A.; FINERAN, P. C. CRISPR-Cas systems: beyond adaptive immunity. **Nat. Rev. Microbiol.**, Londres, v. 12, p. 317–326, 2014.

WESTRA, E. R. et al. Parasite Exposure Drives Selective Evolution of Constitutive versus Inducible Defense. **Curr. Biol.**, Londres, v. 25, p. 1043–1049, 2015.

YAN, W. X. et al. Cas13d Is a Compact RNA-Targeting Type VI CRISPR Effector Positively Modulated by a WYL-Domain-Containing Accessory Protein. **Mol. Cell**, Cambridge, v. 70, p. 327–339, 2018.

ZEGANS, M. E. et al. Interaction between bacteriophage DMS3 and host CRISPR region inhibits group behaviors of *Pseudomonas aeruginosa*. **J. Bacteriol.**, Washington, v. 91, p. 210–219, 2009.

ANEXO A - NORMAS PARA SUBMISSÃO – ARTIGO I

Revista: Molecular Genetics And Genomics

Disponível em: https://academic.oup.com/gbe/pages/General_Instructions

Acesso em: 21/11/18

Original Papers and Methodology Papers should be written using the following structure

Introduction

The introduction should guide the reader and hence contain the following:

- A short introduction to the field of research
- The scope and incentives of the work
- The background, i.e. previous work by your and other research groups strictly on the subject
- A conclusion summarizing the novelty and the significance of this paper.

Make sure that the introduction focuses on the background that is relevant for the data presented and the conclusions drawn.

Materials and Methods

This section should provide sufficient information to precisely evaluate how the data have been obtained, including the number of repetitions, and should permit a different research group to repeat the experimental work.

Results

Data should be presented as precisely and concisely as possible, if appropriate in the form of tables or figures.

Make sure not to repeat in the text values given in figures or tables.

The data presentation should follow a logical flow starting with a short description of the incentive of the experiment, followed by a description of the results. Do not repeat information on the experimental details; those belong to Materials and Methods. Describe and discuss the data, but interpretations and conclusions should be given in the Discussion.

Avoid combining Results and Discussion: separate Results and Discussion sections make for easier reading and a better structure of your article.

Discussion

The discussion should cover the following aspects:

- An interpretation of the results and their significance with reference to previous work.
- A description how the data advance the state-of-the-art
- The importance of the findings for the research field
- Any models realistically emerging from the data presented
- Implications for future research

Avoid repeating Introduction or Results.

Present your work so that it attracts the interest of referees and readers. This will increase the chances of acceptance and subsequent citation of the published work. Interest is attracted early on in the manuscript by highlighting the research questions addressed, the main results and conclusions derived from the work and their broader scientific impact.

For Reviews authors should use a suitable structure of their choice. The manuscript should be broken down using informative headings. The Abstract and Introduction should be written so as to attract the reader and point out the broader impact and importance of the field covered.

Title page

The title page should include:

The name(s) of the author(s)

A concise and informative title

The affiliation(s) and address(es) of the author(s)

The e-mail address, and telephone number(s) of the corresponding author

If available, the 16-digit ORCID of the author(s)

Abstract

Please provide an abstract of 150 to 250 words. The abstract should not contain any undefined abbreviations or unspecified references.

Keywords

Please provide 4 to 6 keywords which can be used for indexing purposes.

Text Formatting

Manuscripts should be submitted in Word.

Use a normal, plain font (e.g., 10-point Times Roman) for text.

Use italics for emphasis.

Use the automatic page numbering function to number the pages.

Do not use field functions.

Use tab stops or other commands for indents, not the space bar.

Use the table function, not spreadsheets, to make tables.

Use the equation editor or MathType for equations.

Save your file in docx format (Word 2007 or higher) or doc format (older Word versions).

Manuscripts with mathematical content can also be submitted in LaTeX.

LaTeX macro package (zip, 182 kB)

Headings

Please use no more than three levels of displayed headings.

Abbreviations

Abbreviations should be defined at first mention and used consistently thereafter.

Footnotes

Footnotes can be used to give additional information, which may include the citation of a reference included in the reference list. They should not consist solely of a reference citation, and they should never include the bibliographic details of a reference. They should also not contain any figures or tables.

Footnotes to the text are numbered consecutively; those to tables should be indicated by superscript lower-case letters (or asterisks for significance values and other statistical data). Footnotes to the title or the authors of the article are not given reference symbols.

Always use footnotes instead of endnotes.

Acknowledgments

Acknowledgments of people, grants, funds, etc. should be placed in a separate section on the title page. The names of funding organizations should be written in full.

Scientific Style

Please follow the download below:

Scientific Style guidelines (pdf, 118 kB)

References

Cite references in the text by name and year in parentheses. Some examples:

Negotiation research spans many disciplines (Thompson 1990).

This result was later contradicted by Becker and Seligman (1996).

This effect has been widely studied (Abbott 1991; Barakat et al. 1995a, b; Kelso and Smith 1998; Medvec et al. 1999, 2000).

Reference list

The list of references should only include works that are cited in the text and that have been published or accepted for publication. Personal communications and unpublished works should only be mentioned in the text. Do not use footnotes or endnotes as a substitute for a reference list.

Reference list entries should be alphabetized by the last names of the first author of each work. Order multi-author publications of the same first author alphabetically with respect to second, third, etc. author. Publications of exactly the same author(s) must be ordered chronologically.

Journal article

Gamelin FX, Baquet G, Berthoin S, Thevenet D, Nourry C, Nottin S, Bosquet L (2009) Effect of high intensity intermittent training on heart rate variability in prepubescent children. *Eur J Appl Physiol* 105:731-738.
<https://doi.org/10.1007/s00421-008-0955-8>

Ideally, the names of all authors should be provided, but the usage of “et al” in long author lists will also be accepted:

Smith J, Jones M Jr, Houghton L et al (1999) Future of health insurance. *N Engl J Med* 341:325–329

Article by DOI

Slifka MK, Whitton JL (2000) Clinical implications of dysregulated cytokine production. *J Mol Med*. <https://doi.org/10.1007/s001090000086>

Book

South J, Blass B (2001) The future of modern genomics. Blackwell, London

Book chapter

Brown B, Aaron M (2001) The politics of nature. In: Smith J (ed) The rise of modern genomics, 3rd edn. Wiley, New York, pp 230-257

Online document

Cartwright J (2007) Big stars have weather too. IOP Publishing PhysicsWeb. <http://physicsweb.org/articles/news/11/6/16/1>. Accessed 26 June 2007

Dissertation

Trent JW (1975) Experimental acute renal failure. Dissertation, University of California

Always use the standard abbreviation of a journal's name according to the ISSN List of Title Word Abbreviations, see ISSN LTWA

If you are unsure, please use the full journal title.

For authors using EndNote, Springer provides an output style that supports the formatting of in-text citations and reference list.

EndNote style (zip, 2 kB)

EndNote style for Molecular Genetics and Genomics

Instead of the general file in the above download, please use the Endnote style given below

EndNote style for Molecular Genetics and Genomics (zip, 1 kB)

Tables

All tables are to be numbered using Arabic numerals.

Tables should always be cited in text in consecutive numerical order.

For each table, please supply a table caption (title) explaining the components of the table.

Identify any previously published material by giving the original source in the form of a reference at the end of the table caption.

Footnotes to tables should be indicated by superscript lower-case letters (or asterisks for significance values and other statistical data) and included beneath the table body.

Artwork and Illustrations Guidelines

Electronic Figure Submission

Supply all figures electronically.

Indicate what graphics program was used to create the artwork.

For vector graphics, the preferred format is EPS; for halftones, please use TIFF format. MSOffice files are also acceptable.

Vector graphics containing fonts must have the fonts embedded in the files.

Name your figure files with "Fig" and the figure number, e.g., Fig1.eps.

Line Art

Definition: Black and white graphic with no shading.

Do not use faint lines and/or lettering and check that all lines and lettering within the figures are legible at final size.

All lines should be at least 0.1 mm (0.3 pt) wide.

Scanned line drawings and line drawings in bitmap format should have a minimum resolution of 1200 dpi.

Vector graphics containing fonts must have the fonts embedded in the files.

Halftone Art

Definition: Photographs, drawings, or paintings with fine shading, etc.

If any magnification is used in the photographs, indicate this by using scale bars within the figures themselves.

Halftones should have a minimum resolution of 300 dpi.

Combination Art

Definition: a combination of halftone and line art, e.g., halftones containing line drawing, extensive lettering, color diagrams, etc.

Combination artwork should have a minimum resolution of 600 dpi.

Color Art

Color art is free of charge for online publication.

If black and white will be shown in the print version, make sure that the main information will still be visible. Many colors are not distinguishable from one another when converted to black and white. A simple way to check this is to make a xerographic copy to see if the necessary distinctions between the different colors are still apparent.

If the figures will be printed in black and white, do not refer to color in the captions.

Color illustrations should be submitted as RGB (8 bits per channel).

Figure Lettering

To add lettering, it is best to use Helvetica or Arial (sans serif fonts).

Keep lettering consistently sized throughout your final-sized artwork, usually about 2–3 mm (8–12 pt).

Variance of type size within an illustration should be minimal, e.g., do not use 8-pt type on an axis and 20-pt type for the axis label.

Avoid effects such as shading, outline letters, etc.

Do not include titles or captions within your illustrations.

Figure Numbering

All figures are to be numbered using Arabic numerals.

Figures should always be cited in text in consecutive numerical order.

Figure parts should be denoted by lowercase letters (a, b, c, etc.).

If an appendix appears in your article and it contains one or more figures, continue the consecutive numbering of the main text. Do not number the appendix figures, "A1, A2, A3, etc." Figures in online appendices (Electronic Supplementary Material) should, however, be numbered separately.

Figure Captions

Each figure should have a concise caption describing accurately what the figure depicts. Include the captions in the text file of the manuscript, not in the figure file.

Figure captions begin with the term Fig. in bold type, followed by the figure number, also in bold type.

No punctuation is to be included after the number, nor is any punctuation to be placed at the end of the caption.

Identify all elements found in the figure in the figure caption; and use boxes, circles, etc., as coordinate points in graphs.

Identify previously published material by giving the original source in the form of a reference citation at the end of the figure caption.

Figure Placement and Size

Figures should be submitted separately from the text, if possible.

When preparing your figures, size figures to fit in the column width.

For most journals the figures should be 39 mm, 84 mm, 129 mm, or 174 mm wide and not higher than 234 mm.

For books and book-sized journals, the figures should be 80 mm or 122 mm wide and not higher than 198 mm.

Permissions

If you include figures that have already been published elsewhere, you must obtain permission from the copyright owner(s) for both the print and online format. Please be aware that some publishers do not grant electronic rights for free and that Springer will not be able to refund any costs that may have occurred to receive these permissions. In such cases, material from other sources should be used.

Accessibility

In order to give people of all abilities and disabilities access to the content of your figures, please make sure that

All figures have descriptive captions (blind users could then use a text-to-speech software or a text-to-Braille hardware)

Patterns are used instead of or in addition to colors for conveying information (colorblind users would then be able to distinguish the visual elements)

Any figure lettering has a contrast ratio of at least 4.5:1

Electronic Supplementary Material

Springer accepts electronic multimedia files (animations, movies, audio, etc.) and other supplementary files to be published online along with an article or a book chapter. This feature can add dimension to the author's article, as certain information cannot be printed or is more convenient in electronic form.

Before submitting research datasets as electronic supplementary material, authors should read the journal's Research data policy. We encourage research data to be archived in data repositories wherever possible.

Submission

Supply all supplementary material in standard file formats.

Please include in each file the following information: article title, journal name, author names; affiliation and e-mail address of the corresponding author.

To accommodate user downloads, please keep in mind that larger-sized files may require very long download times and that some users may experience other problems during downloading.

Audio, Video, and Animations

Aspect ratio: 16:9 or 4:3

Maximum file size: 25 GB

Minimum video duration: 1 sec

Supported file formats: avi, wmv, mp4, mov, m2p, mp2, mpg, mpeg, flv, mxf, mts, m4v, 3gp

Text and Presentations

Submit your material in PDF format; .doc or .ppt files are not suitable for long-term viability.

A collection of figures may also be combined in a PDF file.

Spreadsheets

Spreadsheets should be submitted as .csv or .xlsx files (MS Excel).

Specialized Formats

Specialized format such as .pdb (chemical), .wrl (VRML), .nb (Mathematica notebook), and .tex can also be supplied.

Collecting Multiple Files

It is possible to collect multiple files in a .zip or .gz file.

Numbering

If supplying any supplementary material, the text must make specific mention of the material as a citation, similar to that of figures and tables.

Refer to the supplementary files as “Online Resource”, e.g., “... as shown in the animation (Online Resource 3)”, “... additional data are given in Online Resource 4”.

Name the files consecutively, e.g. "ESM_3.mpg", "ESM_4.pdf".

Captions

For each supplementary material, please supply a concise caption describing the content of the file.

Processing of supplementary files

Electronic supplementary material will be published as received from the author without any conversion, editing, or reformatting.

Accessibility

In order to give people of all abilities and disabilities access to the content of your supplementary files, please make sure that

The manuscript contains a descriptive caption for each supplementary material

Video files do not contain anything that flashes more than three times per second (so that users prone to seizures caused by such effects are not put at risk)

ANEXO B - NORMAS PARA SUBMISSÃO – ARTIGO II

Revista: Genome Biology and Evolution

Disponível em: https://academic.oup.com/gbe/pages/General_Instructions

Acesso em: 21/11/18

PREPARATION OF MANUSCRIPTS

General Format

Prepare your manuscript text using a Word processing package (save in .doc or .rtf format). Remember to number each page. Use double spacing (space between lines of type not less than 6 mm) throughout the manuscript and leave margins of 25 mm (1 inch) at the top, bottom and sides of each page. Please avoid footnotes. Type references in the correct order and in *GBE* style (see below). Type without hyphenation, except for compound words. Type headings in the style of the journal. Use the TAB key once for paragraph indents. Where possible use Times for the text font and Symbol for the Greek and special characters.

Use the word processing formatting features to indicate strong **Bold**, *Italic*, Greek, Maths, ^{Superscript}, and _{Subscript} characters. Clearly identify unusual symbols and Greek letters. Differentiate between the letter O and zero, and the letters l and I and the number 1. Mark the approximate position of each figure and table.

Check the final copy of your paper carefully, as any spelling mistakes and errors may be translated into the typeset version.

Sections of the Manuscript

- **Title:** The title should accurately advertise the paper's content and contain 150 characters or less including spaces.
- **Authors and affiliations:** Provide the name and institutional address of all authors, match addresses to names using superscript numbers.
- **Corresponding author:** The name of the author to whom all correspondence is to be addressed should be indicated with an asterisk in the author line and specified as follows:
 - *Author for Correspondence: John Smith, Department of Science, University of Somewhere, Anytown, USA, telephone number, fax number, email address
- **Data deposition:** Supply all accession numbers for the relevant databases. New sequence data must be deposited in GenBank/DDBJ/EMBL. Any sequence alignments used must be made available as supplemental information or by the corresponding author upon request.
- **Abstract:** The first page of the manuscript should begin with the abstract, which should be a concise summary of the paper. Avoided reference citations in the Abstract; if mentioned, the full reference must be given. The Abstract should contain 250 words or less.
- **Key words:** Up to six key words should be given below the abstract. Key words facilitate retrieval of articles by search engines, web directories and indexes; therefore, terms that are too general should be avoided. The selected key words should not repeat words given in the title. The aim is to assist potential readers to find the article by clearly and specifically describing its subject matter, including aspects of methodology or the theoretical framework.
- **References:** Published articles and those in press (state the journal that has accepted them, provide a doi where possible) may be included. Do not include any reference cited *only* in Supplementary files.

In the text citation, a reference should be cited by author and date. Do not place text other than the author and date within the parentheses. No more than two authors may be cited per text citation; if there are more than two authors, use et al. in the text (unless more are necessary to distinguish between references).

In the reference list, list all authors if the author total is five authors or fewer; with more than five list the first author (only) followed by et al.

At the end of the manuscript, the references should be typed in alphabetical order, with the authors' names, year, paper title, journal, volume number, inclusive page numbers, and name and address of publisher (for books only). The name of the journal should be abbreviated according to the World List of Scientific Periodicals. References should therefore be listed as follows:

- Cagan RH, Rhein LD. 1980. Biochemical basis of recognition of taste and olfactory stimuli. In: van der Starre H, editor. *Olfaction and Taste VII*. Oxford: IRL Press. p. 35-44.
- Marshall DA, Moulton DG. 1981. Olfactory sensitivity to alpha-ionone in humans and dogs. *Chem Senses*. 6:53-61.
- van der Starre H, editor. 1980. *Olfaction and Taste VII*. Oxford: IRL Press.
- Avoid personal communications and mention of unpublished data.
- References to websites should be avoided, but if they are given, the references should give authors (if known), title of cited page, URL in full, and year of posting in parentheses.
- **Tables:** Tables should be prepared on separate sheets and numbered consecutively with Arabic numerals. They must be supplied in an editable format (.xlsx or .docx, for example) rather than an image. They should be self-explanatory and include a brief descriptive title. They should be of such a size that they fit easily onto a journal page, the type area of which is 234 (height) x 185 mm (double column width) or 89 mm (single column width). Footnotes to tables indicated by lower case letters are acceptable, but they should not include extensive experimental details.
- **Illustrations:** All illustrations (line drawings and photographs) must be referred to in the text (as Figure 1 etc.) and should be abbreviated to 'Fig. 1.' only in the figure legend. At online submission, you will be required to submit images electronically in one of the following formats: .jpg, .gif, .tif, .pdf or .eps.
- Each figure should be on a separate page and should be submitted at roughly final magnification. Use sans serif fonts such as Arial or Helvetica in figures. Use uniform font size in each figure whenever possible and recall that labels should never smaller than 6 pt at final magnification.
- **Electronic submission of figures:** Save figures at a resolution of at least 300 pixels per inch at the final printed size for color figures and photographs, and 600

pixels per inch for black and white line drawings. Color art **must** be submitted in CMYK rather than RGB format. Authors should be satisfied with the colors in CMYK (both on screen and when printed) before submission. Please also keep in mind that colors can appear differently on different screens and printers. Failure to follow these guides could result in complications and delays. For useful information on preparing your figures for publication, click [here](#).

- **Figure legends:** These should be included at the end of the manuscript text. Define all symbols and abbreviations used in the figure. Common abbreviations and others in the preceding text need not be redefined in the legend.
- **Color Figures:** All figures submitted to the journal in color will be published in color online at no cost to authors.
- **Permissions for Illustrations and Figures:** Permission to reproduce copyright material, for print and online publication in perpetuity, must be cleared and if necessary paid for by the author; this includes applications and payments to DACS, ARS, and similar licensing agencies where appropriate. Evidence in writing that such permissions have been secured from the rights-holder must be made available to the editors. It is also the author's responsibility to include acknowledgements as stipulated by the particular institutions. Oxford Journals can offer information and documentation to assist authors in securing print and online permissions: please see the [Guidelines for Authors](#) section. Information on permissions contacts for a number of main galleries and museums can also be provided. Should you require copies of this, please contact the editorial office of the journal in question or the [Oxford Journals Rights](#) department.
- **Funding:** Authors who are NIH-funded will have their paper automatically deposited in PubMed Central. Details of all funding sources for the work should be given in the 'Acknowledgements' section. A full list of RIN-approved UK funding agencies may be found [here](#).
- The following convention should be followed:
- The sentence should begin: 'This work was supported by ...'
- The full official funding agency name should be given, i.e. 'the National Cancer Institute at the National Institutes of Health' or simply 'National Institutes of Health' not 'NCI' (one of the 27 subinstitutions) or 'NCI at NIH' ([full RIN-approved list](#))

of UK funding agencies) Grant numbers should be complete and accurate and provided in brackets as follows: '[grant number ABX CDXXXXXX]'

- Multiple grant numbers should be separated by a comma as follows: '[grant numbers ABX CDXXXXXX, EFX GHXXXXXX]'
- Agencies should be separated by a semi-colon (plus 'and' before the last funding agency)
- Where individuals need to be specified for certain sources of funding the following text should be added after the relevant agency or grant number 'to [author initials]'.
 - Example: This work was supported by the National Institutes of Health [AP50 CA098252 and CA118790 to R.B.S.R.]; and the Education Research Council [hfygr667789].
- Oxford Journals will deposit all NIH-funded articles in PubMed Central. See [this page](#) for details. Authors must ensure that manuscripts are clearly indicated as NIH-funded using the guidelines above.