Genomics 113 (2021) 881–888

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Original Article

A computational approach to identify CRISPR-Cas loci in the complete genomes of the lichen-associated *Burkholderia* sp. PAMC28687 and PAMC26561

Prasansah Shrestha [a], So-Ra Han [a], Jun Hyuck Lee [b,c], Hyun Park [d,*], Tae-Jin Oh [a,e,f,**]

[a] Department of Life Science and Biochemical Engineering, Graduate School, SunMoon University, Asan 31460, South Korea
[b] Unit of Research for Practical Application, Korea Polar Research Institute, Incheon 21990, South Korea
[c] Department of Polar Sciences, University of Science and Technology, Incheon 21990, South Korea
[d] Division of Biotechnology, College of Life Sciences and Biotechnology, Korea University, Seoul 02841, South Korea
[e] Genome-based BioIT Convergence Institute, Asan 31460, South Korea
[f] Department of Pharmaceutical Engineering and Biotechnology, SunMoon University, Asan 31460, South Korea

ARTICLE INFO

Keywords:
CRISPR-Cas classification
Burkholderia PAMC26561
Burkholderia PAMC28687
Lichen

ABSTRACT

The genus *Burkholderia* and its strains PAMC28687 and PAMC26561 are lichen-associated bacteria isolated from the Antarctic region. Our study is the first to provide the genome sequence of the *Burkholderia* sp. PAMC26561 strain. The genus *Burkholderia* includes bacteria that are pathogenic to plants, animals, and humans. Computational analysis of complete genomes of strains from the uncategorized *Burkholderia* group was performed using the NCBI databank and PATRIC (for genome sequence information) and CRISPRCasFinder (online and offline versions) software in order to predict the CRISPR loci and Cas genes. The RNAfold Webserver online software was used to predict RNA secondary structures. Our study showed that strain MSMB0852 (plasmid) possesses CRISPR-Cas system Class 2, and two lichen-associated strains, PAMC28687 (chromosome I) and PAMC26561 (chromosome I), possess CRISPR-Cas system Class 1. Additionally, only the two lichen-associated strains possess a variety of Cas genes.

1. Introduction

The genus *Burkholderia* belongs to the phylum Proteobacteria and family Burkholderiaceae [1]. The members of this genus include bacteria that are pathogenic to animals and plants, cause nodulation in legumes, inhabit soil and water, or are endophytes [2]. The pathogenic members include the *Burkholderia cepacia* complex that attacks humans, and *B. mallei*, responsible for glanders, a disease that occurs mostly in horses and related animals. *B. pseudomallei* is a causative agent of melioidosis. The members of the genus are gram-negative, obligate aerobic, and rod-shaped bacteria that are motile by means of a single flagellum or multiple polar flagella, apart from *B. mallei*, which is non-motile [3].

We have sequenced two species of *Burkholderia* from Antarctica (*B.* sp. PAMC28687 and *B.* sp. PAMC26561), which are both cold-adapted lichen-associated organisms. Their sequence information was deposited within the NCBI (National Center for Biotechnology Information) in 2016 as an uncategorized group. As these two strains were isolated from the polar region, they could be classified as psychrophilic bacteria. Numerous organisms, in particular, bacteria, yeasts, unicellular algae, and fungi, have successfully colonized cold environments, which are the most abundant environments on the surface of our planet. As these organisms do not have any thermal regulatory mechanisms, their internal temperature is close, if not identical, to that of the surrounding environment. Despite the strong negative effect of low temperatures on biochemical reactions, the rates of reproduction, growth, and locomotion in these organisms are similar to those of species that inhabit temperate environments [4]. The strains PAMC28687 and PAMC26561 have three and two chromosomes, respectively. In addition, these lichen-associated bacteria also have CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) loci and Cas-associated (Cas) genes governing their immune systems.

* Corresponding author.
** Corresponding author at: Department of Pharmaceutical Engineering and Biotechnology, SunMoon University, Asan 31460, South Korea.
E-mail addresses: hpark@korea.co.kr (H. Park), tjoh3782@sunmoon.ac.kr (T.-J. Oh).

https://doi.org/10.1016/j.ygeno.2021.01.019
Received 10 September 2020; Received in revised form 31 October 2020; Accepted 26 January 2021
Available online 29 January 2021

Table 1
Genome features of *Burkholderia* sp. PAMC26561 and PAMC28687.

Name	PAMC28687			PAMC26561		
	Chromosome 1	Chromosome 2	Chromosome 3	Chromosome 1	Chromosome 2	Chromosome 3
Genome size (bp)	3,397,021	1,611,937	1,623,425	3,260,450	1,696,029	1,495,500
GC%	60.1	60.2	60.3	60.2	60.2	60.4
Protein	2926	1468	1365	2990	1426	1415
rRNA	12	N	–N	12	–N	–N
tRNA	48	1	1	53	1	–N
Other RNA	4	N	–N	4	–N	–N
Gene	3060	1514	1392	3106	1459	1465
Pseudogene	70	45	26	47	32	50
NCBI No.	NZ_CP014505.1	NZ_CP014506.1	NZ_CP014507.1	NZ_CP014306.1	NZ_CP014307.1	NZ_CP014315.1

Note; "N" referes to not available.

CRISPR and Cas genes are present in many bacterial and archaeal genomes [5]. Cas genes are present only in CRISPR-containing prokaryotes, and are always located adjacent to CRISPR loci [6–8]. Studies suggest that CRISPR loci and Cas genes have related functions, especially in terms of DNA metabolism and/or gene expression [9]. The typical genomic architecture of a CRISPR-Cas system consists of a CRISPR locus, a series of Cas genes, and a leader region. The genomic component of the CRISPR-Cas system is formed by a series of tandem repeats separated by a unique spacer sequence, which may share sequence similarity with viruses, plasmids, or bacteria [7,10,12]. Interestingly, several pathogenic bacteria possess a CRISPR-associated ribonucleoprotein complex; this complex is thought to play a dual role in defense as well as virulence [11]. Some bacterial species harbor more than one CRISPR locus within their genomes [11]. Comparative genomic analyses have revealed that CRISPRs and their associated Cas genes are present in diverse bacterial phylogenetic groups, which has resulted in the classification of these genes into several protein families [11].

In the present study, we predicted CRISPR loci and Cas genes in 21 uncategorized complete genomes of *Burkholderia* species, along with the lichen-associated *B.* sp. PAMC28687 and PAMC26561 isolated from the Antarctic region. CRISPRCasFinder (online and offline) was used to study the CRISPR loci and Cas genes present in the respective strains. Lichen-associated *Burkholderia* strains have not been studied as widely as *Burkholderia* strains associated with plants and mammals. The prediction and classification of CRISPR loci and Cas genes in these strains may improve our understanding of pathogenic bacteria, as the genus *Burkholderia* also contains members that are opportunistic human pathogens. The CRISPR-Cas systems of these strains might provide a new direction to identify pathogens more accurately and rapidly at the genetic level. Furthermore, identifying various types of CRISPR and Cas genes in psychrophilic bacteria may help us to understand the nature of Cas genes and their underlying mechanisms. Use of bioinformatic tools such as CRISPRCasFinder can allow us to study CRISPR loci and Cas genes in respective strains without performing laborious laboratory experiments.

2. Materials and methods

2.1. Data sources

The complete genome sequences of the uncategorized group of *Burkholderia* species were obtained from the NCBI nucleotide database (https://www.ncbi.nlm.nih.gov/). Genomes deposited up to January 2020 were included in this study. A total of 21 strains, including our two lichen-associated strains (*B.* sp. PAMC28687 and PAMC26561), were included in the study [13].

2.2. Comparative genome analysis

The PATRIC database (https://patricbrc.org/) was used to obtain genomic information and the number of virulence factors involved in the respective strains. The PATRIC database uses the virulence factor

Table 2
Genomic information for 21 uncategorized *Burkholderia* strains.

Organism/name	Strain	GC (%)	Isolate info.			Virulence factors		
			Isolation source	Isolation country	Geographic location	Victors	VFDB	PATRIC_VF
Burkholderia sp.	DHOD12	63.06	Forest soil	China	China: Guangdong Province	19	14	1
Burkholderia sp.	OLGA172	60.85	Soil	Russia	Russia	NA	NA	NA
Burkholderia sp.	BDU8	66.46	Soil	Australia	Australia: Badu Island, Torres Strait	108	36	1
Burkholderia sp.	MSMB0266	67.01	Soil	Australia	Australia: Northern Territory	81	33	1
Burkholderia sp.	MSMB617WGS	67.50	Soil	Australia	Australia: Northern Territory	80	32	1
Burkholderia sp.	MSMB0852	67.31	Soil	Australia	Australia: Northern Territory	58	25	1
Burkholderia sp.	CCGE1003	63.23	N/A	N/A	N/A	18	15	1
Burkholderia sp.	MSMB43	67.14	Water	Australia	Australia: Northern Territory	2	1	
Burkholderia sp.	2,002,721,687	67.14	Water	Australia	Australia	95	37	1
Burkholderia sp.	CCGE1001	63.62	N/A	N/A	N/A	17	15	1
Burkholderia sp.	BDU6	66.31	Soil	Australia	Australia: Badu Island, Torres Strait	55	28	1
Burkholderia sp.	MSMB0175	64.38	Soil	Australia	Australia: Northern Territory	42	25	1
Burkholderia sp.	JP2-270	65.95	Rice rhizosphere	China	China: Hangzhou	28	20	1
Burkholderia sp.	IDO3	66.37	Reactor sludge	China	China: Dalian	25	19	1
Burkholderia sp.	NRF60-BP8	67.26	Soil	Thailand	Thailand: Ubon, Pibul	31	20	1
Burkholderia sp.	KBS0801	66.64	Soil	USA	USA: Hickory Corners, MI	27	20	NA
Burkholderia sp.	LA-2-3-30-S1-D2	66.35	Soil	United States	USA: New Orleans, Plaquemines, LA	28	21	1
Burkholderia sp.	MSMB0856	66.49	Soil	Australia	Australia: Northern Territory	28	20	1
Burkholderia sp.	KJ006	67.19	Rice root	South Korea	Korea	27	20	1
Burkholderia sp.	PAMC28687*	59.98	Lichen	Antarctica	Antarctica	15	12	1
Burkholderia sp.	PAMC26561*	59.24	Lichen	Antarctica	Antarctica: King George Island,	16	12	1

Note that "*" refers to lichen-associated strains isolated from Antarctica.

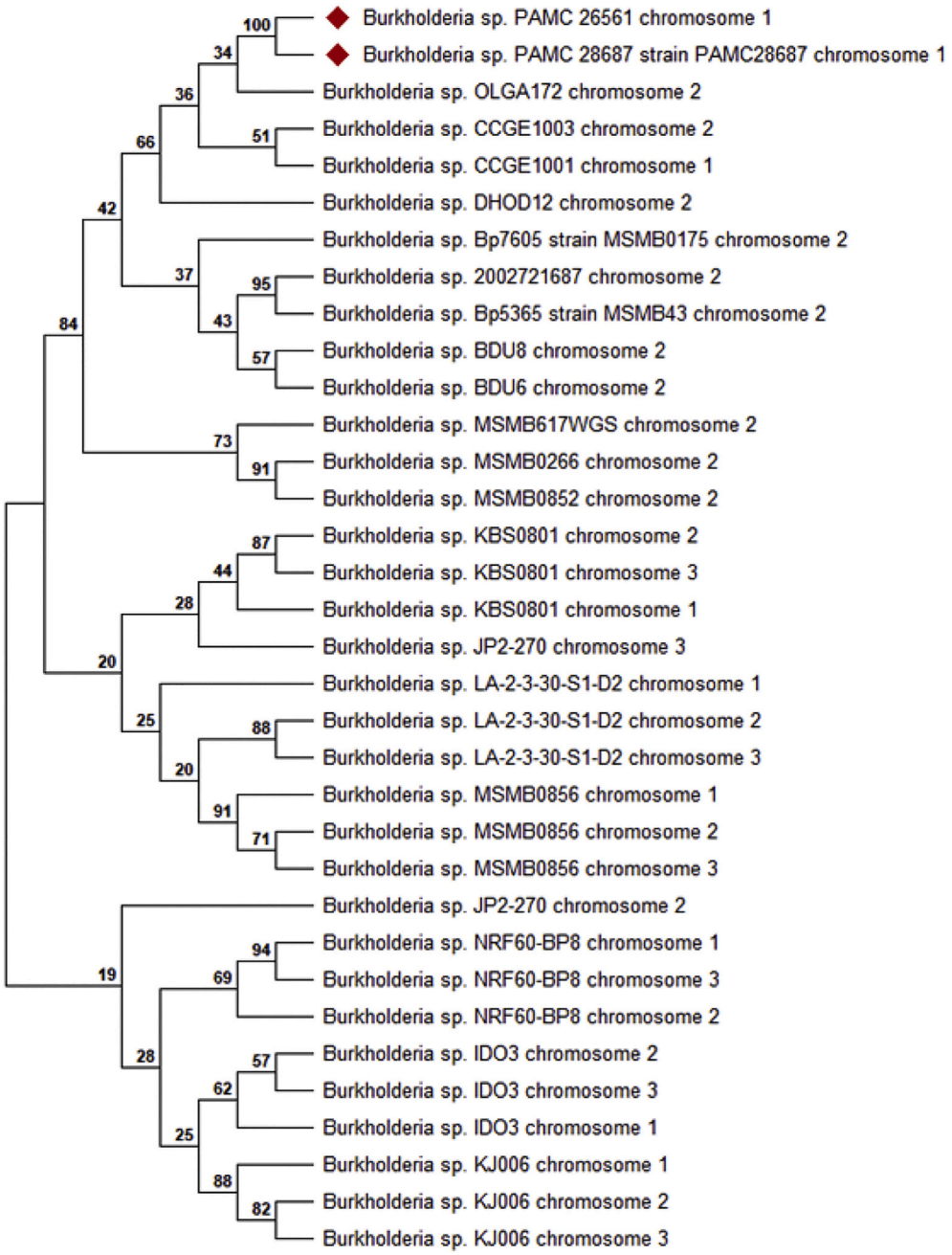


Fig. 1. A phylogenetic tree was constructed using the neighbor-joining method for the uncategorized microorganisms using the complete genomes of *Burkholderia* strains. The strains marked with a red marker represent strains isolated from an Antarctic region. The bootstrap value was 1000. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

database (VFDB), Victors, and PATRIC_VF. The VFDB is an integrated and comprehensive online resource for curating information regarding the virulence factors of bacterial pathogens [14]. CRISPR loci and Cas genes were predicted using the online tool (https://crisprcas.i2bc.paris-saclay.fr/), 2020 and standalone version 1.4 June 2019 of CRISPRCasFinder (https://crisprcas.i2bc.parissaclay.fr/CrisprCasFinder/Index); the outcomes were compared for the uncategorized group of *Burkholderia*. In order to identify CRISPR arrays, CRISPRCasFinder uses CRISPRFinder v.4.2, which is based on Vmatch v.2.3 (http://www.vmatch.de/) to identify CRISPR repeats [15]. The RNA secondary structures and Minimum Free Energy (MFE) of each direct repeat (DR) sequence were predicted by using the RNAfold Web server (https://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi). The CRISPRminer (http://www.microbiome-bigdata.com/CRISPRminer/) online software [16] was also used to predict CRISPR loci and Cas genes, and to identify gene clusters.

2.3. Phylogenomic classification

A phylogenetic tree was constructed using 16S rRNA sequences of the complete genomes of uncategorized *Burkholderia* strains as well as for the Cas genes of *Burkholderia* sp. MSMB0852 (plasmid), *Burkholderia* sp. PAMC28687 (chromosome I), and PAMC26561 (chromosome I) obtained from the NCBI database. MEGA X software (https://www.megasoftware.net/) was used to construct the phylogenetic tree. Phylogenetic studies performed on the CAS protein suggest that CRISPRs can be acquired through horizontal gene transfer [17]. The average nucleotide identity of closely related species was determined using the

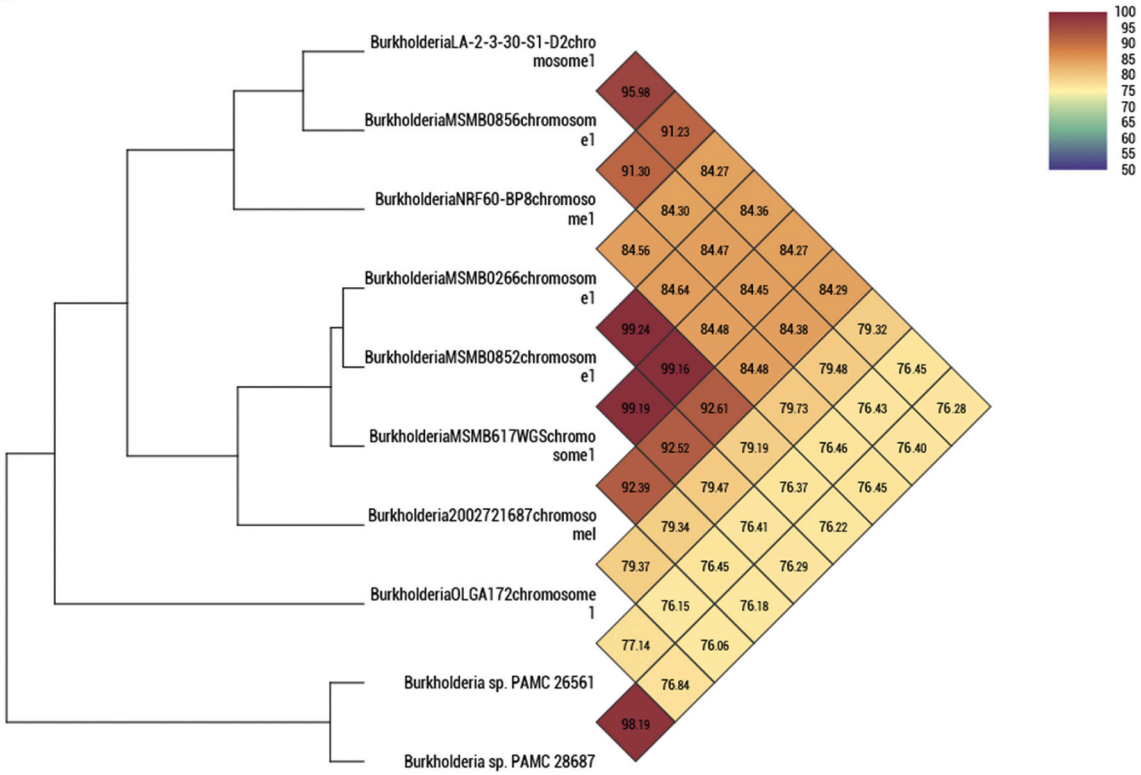


Fig. 2. Heatmap generated using the OrthoANI values, which were calculated using the OAT software for *Burkholderia* sp. PAMC26561 and PAMC28687, and other closely related *Burkholderia* species.

Orthologous Average Nucleotide Identity Software Tool (OAT) [18].

3. Results and discussion

3.1. General genomic features of the uncategorized Burkholderia group

Among the 21 strains from the uncategorized group of *B.* sp., MSMB617WGS had the highest GC content (67.50%), whereas the lichen-associated strains PAMC28687 and PAMC26561 had the lowest GC content (59.98% and 59.24%, respectively). The general genome features and genomic information for all 21 *Burkholderia* strains are summarized in Table 1 and Table 2 respectively. According to the genomic information obtained from the PATRIC database (https://patricbrc.org/), 19 strains of *Burkholderia* species were isolated from soil, water, and rice root, with the exception of PAMC28687 and PAMC26561. The strains PAMC28687 and PAMC26561 were isolated from the Antarctic region, and can survive low temperatures. Psychrophilic enzymes produced by cold-adapted microorganisms display high catalytic efficiency, and are typically associated with high thermosensitivity [4].

3.2. Phylogenetic comparison within the genus Burkholderia

A phylogenetic tree was constructed from the 16S rRNA sequences of the *Burkholderia* strains. PAMC28687 (chromosome I) and PAMC26561 (chromosome I) showed close relationships to strain OLGA172 (chromosome I) (Fig. 1). A comparative study of all 21 complete genomes was performed to understand the genomic distances based on the 16S rRNA genes and Average Nucleotide Identity (ANI). The strains PAMC28687 and PAMC26561 exhibited similar identities, with an ANI value of 98.19% (Fig. 2). Phylogenetic studies using ANI values reflect functional relationships involving strains more effectively than 16S rRNA sequence studies [19].

Table 3
Number of chromosomes and plasmids in strains of the genus *Burkholderia*.

Strain	Length (bp)	Chromosome	Plasmid
Burkholderia sp. 2002721687	7,285,824	2	1
Burkholderia sp. BDU6	6,590,914	2	0
Burkholderia sp. BDU8	7,357,530	2	0
Burkholderia sp. Bp5365 (MSMB43)	7,287,809	2	1
Burkholderia sp. Bp7605 (MSMB0175)	5,705,688	2	1
Burkholderia sp. CCGE1001	6,833,751	2	1
Burkholderia sp. CCGE1003	7,043,595	2	0
Burkholderia sp. DHOD12	8,576,517	3	0
Burkholderia sp. IDO3	8,003,806	3	1
Burkholderia sp. JP2-270	8,925,310	3	2
Burkholderia sp. KBS0801	7,381,868	3	0
Burkholderia sp. KJ006	6,629,912	3	1
Burkholderia sp. LA-2-3-30-S1-D2	7,129,526	3	0
Burkholderia sp. MSMB0266	7,427,555	2	1
Burkholderia sp. MSMB0852	7,081,173	2	1
Burkholderia sp. MSMB0856	7,249,502	3	1
Burkholderia sp. MSMB617WGS	6,913,798	2	0
Burkholderia sp. NRF60-BP8	7,079,943	3	0
Burkholderia sp. OLGA172	8,574,890		3
Burkholderia sp. PAMC26561*	9,055,344	3	5
Burkholderia sp. PAMC28687*	6,881,273	3	5

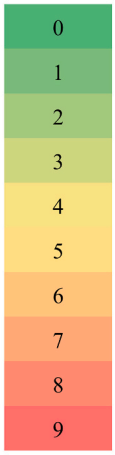
Note that "*" refers to lichen-associated strains isolated from Antarctica.

3.3. Comparative genomics

In this study, only 21 uncategorized groups of *Burkholderia* species were used to predict and classify CRISPR-Cas systems, since two lichen-associated strains were deposited with NCBI as an uncategorized group. Table 3 summarizes the number of chromosomes and plasmids present in the respective strains. Strains originating from Antarctic regions possessed high numbers of plasmids (5), followed by strain OLGA172, which possessed 3. Studies have shown that approximately 40% of bacterial genomes contain CRISPR loci [20]. In addition, Sorek et al.

Table 4
Number of CRISPR loci and Cas genes in uncategorized *Burkholderia* sp.

Strain	CRISPR (offline)	CRISPR (online)	Cas (offline)	Cas (online)
2002721687 (Chromosome I)	0	0	0	0
2002721687 (Chromosome II)	2	2	0	1
BDU6 (Chromosome I)	2	1	0	1
BDU6 (Chromosome II)	1	3	0	1
BDU8 (Chromosome I)	1	1	0	1
BDU8 (Chromosome II)	0	3	0	1
Bp5365 MSMB43 (Chromosome I)	0	0	0	0
Bp5365 MSMB43 (Chromosome II)	0	0	0	0
Bp5365 MSMB43 (plasmid)	0	0	0	0
Bp7605 MSMB0175 (Chromosome I)	1	7	0	1
Bp7605 MSMB0175 (Chromosome II)	0	0	0	0
CCGE1001 (Chromosome I)	0	2	0	1
CCGE1001 (Chromosome II)	0	1	0	1
CCGE1003 (Chromosome I)	1	2	0	0
DHOD12 (Chromosome I)	2	5	0	1
DHOD12 (Chromosome II)	1	3	0	1
DHOD12 (Chromosome III)	0	0	0	0
IDO3 (Chromosome I)	0	1	0	1
IDO3 (Chromosome II)	0	0	0	0
IDO3 (Chromosome III)	3	0	0	0
IDO3 (plasmid 1)	1	3	0	1
JP2-270 (Chromosome III)	0	0	0	0
JP2-270 (Chromosome I)	6	5	0	1
JP2-270 (Chromosome II)	0	0	0	1
JP2-270 (plasmid 1)	1	1	0	1
JP2-270 (plasmid 2)	0	0	0	0
JP2-270 (plasmid 3)	2	0	0	0
KBS0801 (Chromosome I)	7	0	0	1
KJ006 (Chromosome I)	1	2	0	1
KJ006 (Chromosome II)	1	2	0	1
KJ006 (Chromosome III)	1	1	0	1
KJ006 (plasmid)	1	1	0	1
LA-2-3-30-S1-D2 (Chromosome I)	0	3	0	1
LA-2-3-30-S1-D2 (Chromosome II)	0	1	0	1
LA-2-3-30-S1-D2 (Chromosome III)	0	0	0	1
MSMB0266 (Chromosome I)	2	1	0	1
MSMB0266 (Chromosome II)	1	1	0	1
MSMB0266 (plasmid)	0	0	0	0
MSMB0852 (Chromosome I)	2	2	0	1
MSMB0852 (Chromosome II)	1	2	0	1
MSMB0852 (plasmid)	0	0	1	0
MSMB0856 (Chromosome I)	0	0	0	0
MSMB0856 (Chromosome II)	0	0	0	0
MSMB0856 (Chromosome III)	0	1	0	1
MSMB617WGS (Chromosome I)	3	0	0	0
MSMB617WGS (Chromosome II)	2	0	0	0
NRF60- BP8 (Chromosome III)	3	1	0	1
NRF60-BP8 (Chromosome I)	1	1	0	1
NRF60-BP8 (Chromosome II)	0	1	0	1
OLGA172 (Chromosome I)	9	7	0	1
OLGA172 (Chromosome II)	3	3	0	0
OLGA172 (plasmid 1)	0	0	0	0
OLGA172 (plasmid 2)	0	0	0	0
OLGA172 (plasmid 3)	0	0	0	0
PAMC26561 (Chromosome I)*	2	2	7	2
PAMC26561 (plasmid)*	1	0	0	0
PAMC28687 (Chromosome I)*	1	1	0	2



Note that "*" refers to lichen-associated strains isolated from Antarctica.

[21] have shown that most of the CRISPR loci in prokaryotes are located chromosomally, and are rarely located on plasmids. The existence of CRISPR loci on plasmids would be detrimental to their heritability. Among the 21 strains, strain OLGA172 (chromosome I) showed 9 CRISPR loci in the offline version of CRISPRCasFinder, but only 7 CRISPR loci in the online version. OLGA172 showed the highest number of CRISPR loci in our study (Table 4). The strain PAMC28687 (chromosome I) showed only one CRISPR locus in the offline and online versions, but showed two Cas genes in the online version of CRISPRCasFinder (Table 4). The CRISPRCasFinder web server currently accepts (multi-) Fasta DNA sequence files up to 50 Mb in size, including files containing up to 100 sequences. The standalone application has no

Table 5
Cas genes and their putative functions in CRISPR-Cas types (standalone program).

Strain	Sequence ID	System	Cas-type/subtype		Functions	References
			Cas protein	Cas Type		
Burkholderia sp. MSMB0852 (plasmid)	NZ_CP013423.1_44	General-Class2	Cas4	I-II-V	Cas4 plays a role in acquiring new viral DNA sequences and incorporating these into the host genome for further crRNA production	[26]
Burkholderia sp. PAMC28687 (Chromosome I)	NZ_CP014505.1_1707	General-Class1	Cas6	IF	An endoribonuclease, which cleaves the pre-crRNA within the CRISPR repeat sequence during the crRNA maturation process	[27]
	NZ_CP014505.1_1708		Csy3	IF	Csy RNP backbone stabilizing protein	[28]
	NZ_CP014505.1_1709		Csy2	IF	Csy RNP stabilization, possibly target recognition protein	[28]
	NZ_CP014505.1_1710		Csy1	IF	Csy RNP large subunit, possibly a polymerase	[22,28]
	NZ_CP014505.1_1711		Cas3_cas2	IF	RNase, specific to U-rich regions, DNase, spacer integration	[27]
	NZ_CP014505.1_1712		Cas1	IF	Specially exhibits nuclease activity against single stranded and branched DNA, replication forks and any be implicated in addition of novel repeats and/or spacers	[27]
Burkholderia sp. PAMC26561 (Chromosome I)	NZ_CP014306.1_1235	General-Class1	Cas2	I-II-III-V	Involved in novel spacer acquisition, novel repeat synthesis and repeat-spacer insertion at the leader end	[27]
	NZ_CP014306.1_1236		Cas1	IC	Specially exhibits nuclease activity against single-stranded and branched DNA, replication forks, and any be implicated in addition of novel repeats and/or spacers	[27]
	NZ_CP014306.1_1237		Cas4	I-II	Cas4 plays a role in acquiring of new viral DNA sequences and incorporating these into the host genome for further crRNA production	[26]
	NZ_CP014306.1_1238		Cas7	IC	CASCADE stabilization protein	[29,30]
	NZ_CP014306.1_1239		Cas8c	IC	This is experimental evidence that Cas8 is important for targeting Cascade to invader DNA	[31]
	NZ_CP014306.1_1240		Cas5	IC	Nuclease, CASCADE complex, crRNA maturation	[30]
	NZ_CP014306.1_1241		Cas3	I	Encodes a nuclease involved in the cleavage of the target DNA, and is thus, responsible for the procession of crRNA and involved in the recognition of target DNA	[27]

pre-defined input size limit, and is only limited by the computer memory available. Likewise, the CRISPRCasFinder database depends on the evidence level in the database that helps to discriminate spurious CRISPR-like elements from authentic CRISPR loci. The results obtained from the offline version of the database showed that strains PAMC26561 and PAMC28687 each had an evidence level of 4, which confirms that these

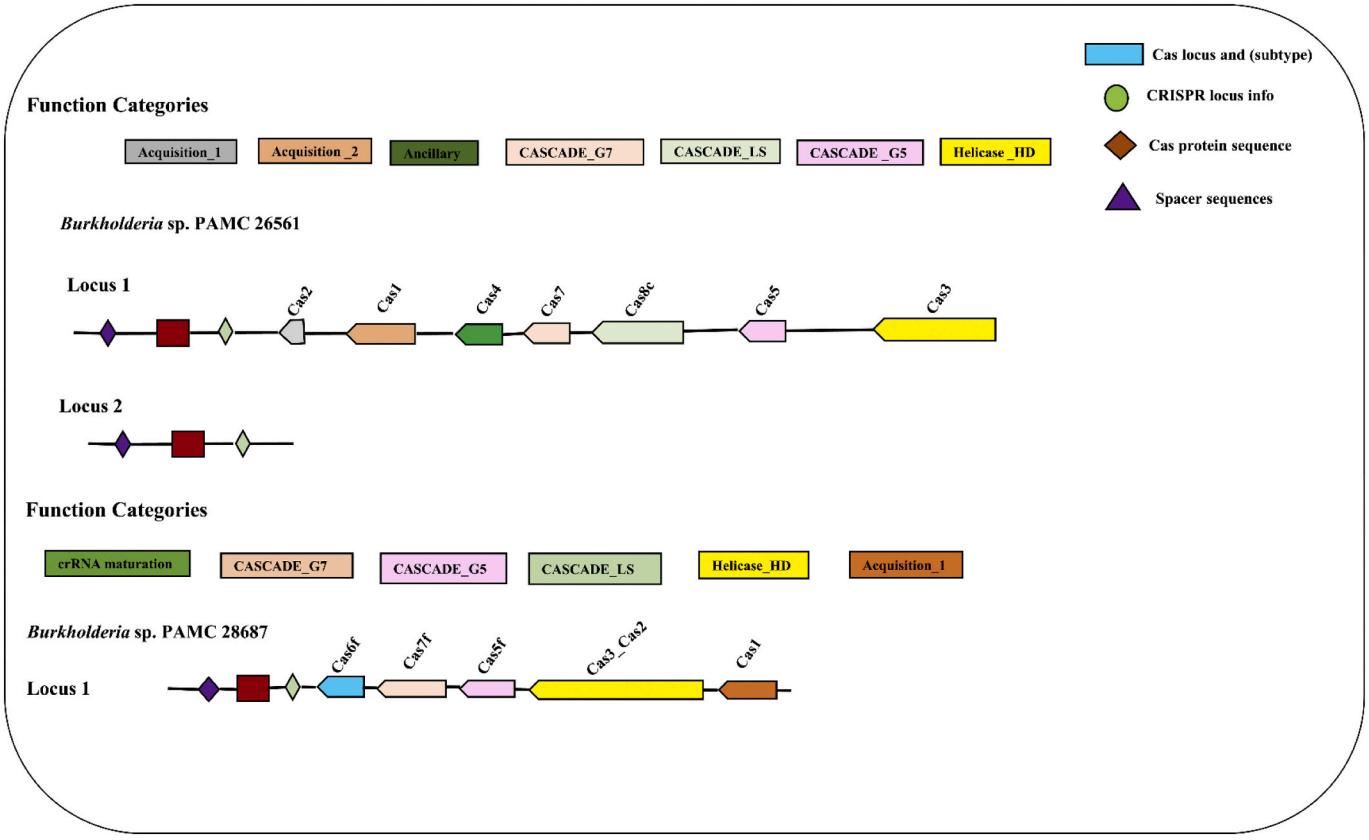


Fig. 3. Schematic representation of the CRISPR (Clustered Interspaced Short Palindromic Repeats) array and Cas genes of the strains PAMC26561 and PAMC28687 generated using CRISPRminer.

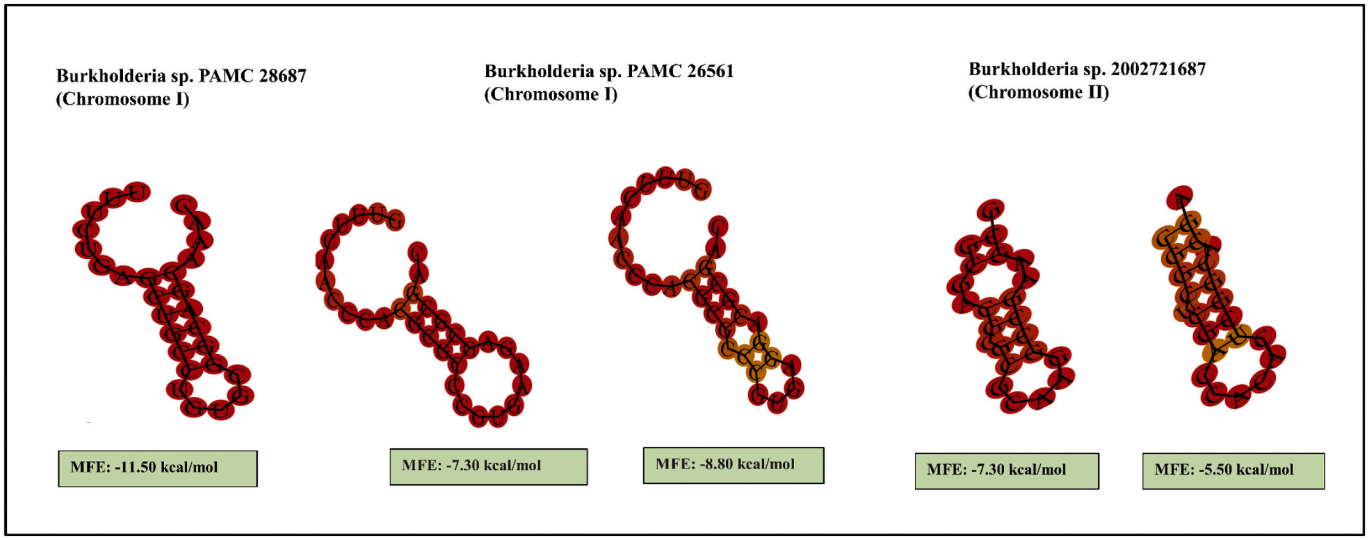


Fig. 4. Prediction of RNA secondary structures were generated using DR sequences and their MFE values

two lichen-associated strains have true CRISPR loci. The evidence levels of 2 to 4 are assigned based on combined degrees of similarity of repeats and spacers [17]. Another strain, PAMC26561 (chromosome I), showed one CRISPR locus in its chromosomes, and two CRISPR loci in its plasmid, along with seven Cas genes in the offline version, but two CRISPR loci and two Cas genes in the online version of CRISPRCasFinder (Table 4).

3.4. Classification of CRISPR loci and Cas genes in Burkholderia sp. PAMC28687 and PAMC26561

The current classification of CRISPR-Cas systems is based on the sequence of the Cas genes, the sequences of the repeats within the CRISPR arrays, and the organization of the Cas operons [22]. CRISPR is classified into two classes; Class 1 (Type I, Type III, and Type IV) and Class 2 (Type II, Type V, and Type VI). Each type is also divided into subtypes, which include subtypes I-A to I-F, subtypes II-A to II-C, and subtypes III-A and III-B. The *Cas1* and *Cas2* genes are common to all three CRISPR-Cas types. The major criterion for classification is the presence or absence of certain type of Cas proteins. For example, the Cas3, Cas9, and Cas10 proteins are hallmarks of CRISPR-Cas types I, II, and III, respectively. Systems that do not have the specific hallmarks of CRISPR-Cas system types I–III are termed as unclassified (type U) [12]. According to the classification of CRISPR-Cas systems, the complete genome in our study belongs to the Class 1 CRISPR system, with the exception of strain MSMB0852 (plasmid), which belongs to Class 2 CRISPR-Cas systems. The CRISPR-Cas system is divided into three subtypes: CRISPR-Cas system type I, II, and III. This classification is based on signature genes that are present in each subtype. However, it is important to note that all types and subtypes of CRISPR systems contain Cas1 and Cas2. Notably, these two Cas proteins play a key role as spacers [23]. According to Makarova et al. [24], Class 1 CRISPR systems are divided into three types: I, III, and IV, and 12 subtypes. Class I CRISPR systems represent approximately 90% of all CRISPR-Cas loci discovered in bacteria and archaea. For example, type I has *Cas3*, type II has *Cas9*, and type III has *Cas10*. Class 1 encompasses the most common and diversified types I and III, and includes diverse variants. It is commonly found in numerous archaea, but is less commonly seen in bacteria [25]. Class 1 CRISPR systems utilize multi-protein complexes.

3.5. Cas genes in Burkholderia sp. PAMC28687 and PAMC26561

CRISPR-associated Cas genes are present only in CRISPR-containing prokaryotes and are always located adjacent to CRISPR loci. Of the 21 strains in our study, only MSMB0852 (plasmid), PAMC28687 (chromosome I), and PAMC26561 (chromosome I) contain Cas genes in their respective genomes. The strain MSMB0852 (plasmid) possesses one *Cas4* gene whereas PAMC28687 (chromosome I) possesses *Cas6*, *Csy3*, *Csy2*, *Csy1*, and a combination of *Cas3_Cas2* and *Cas1*. The strain PAMC26561 possesses *Cas2*, *Cas1*, *Cas4*, *Cas7*, *Cas8c*, *Cas5*, and *Cas3*. The putative functions of Cas genes in each strain are summarized in Table 5 and Fig. 3.

3.6. RNA secondary structure

RNA secondary structures for the 2 lichen-associated strains were predicted using their DR sequences and their MFE values was recorded using the RNA fold web server [32]. Strain PAMC28687 (chromosome I) showed a minimum MFE value of −11.50 kcal/mol and formed the longest stem, compared with strain PAMC26561 (chromosome I), with MFE values of −7.30 kcal/mol and −8.80 kcal/mol (Fig. 4). This indicated a certain correlation between stem length and secondary structure [33]. Although the stability of secondary structures depends on other factors such as GC content, DR sequences with lower MFE values were more stable than those with high MFE values.

According to Kunin et al. [34], the presence of short palindrome sequences in DRs leads to the formation of RNA secondary structures during transcription. These structures may serve to mediate contact between the spacer-target foreign RNA or DNA and Cas-encoded proteins [35]. Furthermore, we observed compensatory base changes in the stems of structured repeats, including G:U base pairs, indicating that this CRISPR system likely functions through an RNA intermediate [36]. Overall, the stem-loop structure of some repeats may contribute to recognition-mediated contact between a gap-targeted foreign RNA or DNA and a Cas-encoded protein, which suggests that the stability of RNA secondary structures may affect CRISPR function.

4. Conclusions

In this study, we predicted CRISPR loci and Cas genes present in the complete genomes of uncategorized groups of *Burkholderia*. The results obtained from online and offline versions of CRISPRCasFinder were not identical in certain instances. The online version showed that all the strains from our study possessed Cas genes, whereas the offline version indicated that only PAMC28687 (chromosome I), PAMC26561 (chromosome I), and MSMB0852 (plasmid) possessed Cas genes. It was found

that most strains have CRISPR loci, but lack Cas genes. These results suggest that bioinformatics tools can only provide preliminary information, which requires further experimental confirmation. Despite having a relatively small genome, PAMC28687 and PAMC26561 (isolated from an Antarctic region and associated with lichens) each possess three chromosomes and five plasmids, unlike the other members from the uncategorized group of *Burkholderia*. Apart from their small-sized genome, they also possess more Cas types and subtypes than those seen in other strains. Surprisingly, these strains also show a historically high record of phage infection, despite having a habitat that is exposed to relatively lower human contamination. Such studies concerning psychrophilic strains could be instrumental for understanding the functions of Cas genes, apart from genome editing, and could improve our understanding of psychrophilic strains of pathogens and their CRISPR-Cas systems.

Funding

This research was a part of the project titled “Development of potential antibiotic compounds using polar organism resources (15250103, KOPRI Grant PM20030)”, funded by the Ministry of Oceans and Fisheries, Korea.

Declaration of Competing Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to thank Mr. Ermal Elbasani for his help in downloading an offline version of the CRISPRCasFinder program.

References

[1] D.E. Woods, P.A. Sokol, The genus Burkholderia, in: M. Dworkin, S. Falkow, E. Rosenberg, K.H. Schleifer, E. Stackebrandt (Eds.), The Prokaryotes-A Handbook on the Biology of Bacteria, 3rd, Springer-Verlag, New York, NY, 2006, pp. 848–860.
[2] A.A. Angus, C.M. Agapakis, S. Fong, S. Yerrapragada, P.E. Santos, P. Yang, N. Song, S. Kano, J. Caballero-Mellado, S.M. de Faria, F.D. Dakora, G. Weinstock, A. M. Hirsch, Plant-associated symbiotic *Burkholderia* species lack hallmark strategies required in mammalian pathogenesis, PLoS One 9 (2014), e83779.
[3] Z. Weinberg, J.E. Barrick, Z. Yao, A. Roth, J.N. Kim, J. Gore, J.X. Wang, E.R. Lee, K. F. Block, N. Sudarsan, S. Neph, M. Tompa, W.L. Ruzzo, R.R. Breaker, Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline, Nucleic Acids Res. 35 (2007) 4809–4819.
[4] C. Gerday, M. Aittaleb, M. Bentahir, J.P. Chessa, P. Claverie, T. Collins, S. D’Amico, J. Dumont, G. Garsoux, D. Georlette, A. Hoyoux, T. Lonhienne, M.A. Meuwis, G. Feller, Cold-adapted enzymes: from fundamentals to biotechnology, Trends Biotechnol. 18 (2000) 103–107.
[5] R.K. Lillestøl, P. Redder, R.A. Garrett, K. Brügger, A putative viral defence mechanism in archaeal cells, Archaea 2 (2006) 59–72.
[6] D.H. Haft, J. Selengut, E.F. Mongodin, K.E. Nelson, A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes, PLoS Comput. Biol. 1 (2005), e60.
[7] P. Horvath, R. Barrangou, CRISPR/Cas, the immune system of bacteria and archaea, Science 327 (2010) 167–170.
[8] K.S. Makarova, N.V. Grishin, S.A. Shabalina, Y.I. Wolf, E.V. Koonin, 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 1 (2006) 7.
[9] R. Barrangou, CRISPR-Cas systems and RNA-guided interference, Wiley Interdiscip. Rev. RNA 4 (2013) 267–278.
[10] E.V. Koonin, K.S. Makarova, CRISPR-Cas: evolution of an RNA-based adaptive immunity system in prokaryotes, RNA Biol. 10 (2013) 679–686.
[11] R. Louwen, R.H.J. Staals, H.P. Endtz, P. van Baarlen, J. van der Oost, The role of CRISPR-Cas systems in virulence of pathogenic bacteria, Microbiol. Mol. Biol. Rev. 78 (2014) 74–88.
[12] M. Burmistrz, K. Pyrć, CRISPR-Cas systems in prokaryotes, Pol. J. Microbiol. 64 (2015) 193–202.
[13] S.R. Han, S.C. Yu, D.H. Ahn, H. Park, T.-J. Oh, Complete genome sequence of *Burkholderia* sp. strain PAMC28687, a potential octopine-utilizing bacterium isolated from Antarctica lichen, J. Biotechnol. 226 (2016) 16–17.
[14] A.R. Wattam, J.J. Davis, R. Assaf, S. Boisvert, T. Brettin, C. Bun, N. Conrad, E. M. Dietrich, T. Disz, J.L. Gabbard, S. Gerdes, C.S. Henry, R.W. Kenyon, D. Machi, C. Mao, E.K. Nordberg, G.J. Olsen, D.E. Murphy-Olson, R. Olson, R. Overbeek, B. Parrello, G.D. Pusch, M. Shukla, V. Vonstein, A. Warren, F. Xia, H. Yoo, R. L. Stevens, Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center, Nucleic Acids Res. 45 (2017) D535–D542.
[15] M.I. Abouelhoda, S. Kurtz, E. Ohlebusch, Replacing suffix trees with enhanced suffix arrays, J. Discrete Algorithms 2 (2004) 53–86.
[16] F. Zhang, S. Zhao, C. Ren, Y. Zhu, H. Zhou, Y. Lai, F. Zhou, Y. Jia, K. Zheng, Z. Huang, CRISPRminer is a knowledge base for exploring CRISPR-Cas systems in microbe and phage interactions, Commun. Biol. 1 (2018) 180.
[17] D. Couvin, A. Bernheim, C. Toffano-Nioche, M. Touchon, J. Michalik, B. Néron, E. P.C. Rocha, G. Vergnaud, D. Gautheret, C. Pourcel, CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins, Nucleic Acids Res. 46 (2018) W246–W251.
[18] I. Lee, Y. Ouk Kim, S.C. Park, J. Chun, OrthoANI: an improved algorithm and software for calculating average nucleotide identity, Int. J. Syst. Evol. Microbiol. 66 (2016) 1100–1103.
[19] W.H. Chung, J. Kang, M.Y. Lim, T.-J. Lim, S. Lim, S.W. Roh, Y.-D. Nam, Complete genome sequence and genomic characterization of *Lactobacillus acidophilus* LA1 (11869BP), Front. Pharmacol. 9 (2018) 83.
[20] J.A. Hakim, H. Koo, J.D. van Elsas, J.T. Trevors, A.K. Bej, CRISPR-Cas system: a new paradigm for bacterial stress response through genome rearrangement, in: F. J. de Bruijn (Ed.), Stress and Environmental Regulation of Gene Expression and Adaptation in Bacteria (Ch.2.11), John Wiley & Sons, Inc, 2016, pp. 146–160.
[21] R. Sorek, V. Kunin, P. Hugenholtz, CRISPR - a widespread system that provides acquired resistance against phages in bacteria and archaea, Nat. Rev. Microbiol. 6 (2008) 181–186.
[22] K.S. Makarova, D.H. Haft, R. Barrangou, S.J. Brouns, E. Charpentier, P. Horvath, S. Moineau, F.J. Mojica, Y.I. Wolf, A.F. Yakunin, J. van der Oost, E.V. Koonin, Evolution and classification of the CRISPR-Cas systems, Nat. Rev. Microbiol. 9 (2011) 467–477.
[23] K.S. Makarova, F. Zhang, E.V. Koonin, SnapShot: class 1 CRISPR-Cas systems, Cell 168 (2017) (946–946.e1).
[24] K.S. Makarova, Y.I. Wolf, E.V. Koonin, Classification and nomenclature of CRISPR-Cas systems: where from here? CRISPR J. 1 (2018) 325–336.
[25] M.A. Shabbir, H. Hao, M.Z. Shabbir, H.I. Hussain, Z. Iqbal, S. Ahmed, A. Sattar, M. Iqbal, J. Li, Z. Yuan, Survival and evolution of CRISPR-Cas system in prokaryotes and its applications, Front. Immunol. 7 (2016) 375.
[26] H. Lee, Y. Zhou, D.W. Taylor, D.G. Sashital, Cas4-dependent prespacer processing ensures high-fidelity programming of CRISPR arrays, Mol. Cell 70 (2018) 48–59 (e5).
[27] Z. Karimi, A. Ahmadi, A. Najafi, R. Ranjbar, Bacterial CRISPR regions: general features and their potential for epidemiological molecular typing studies, Open Microbiol. J. 12 (2018) 59–70.
[28] C. Richter, P.C. Fineran, The subtype I-F CRISPR-Cas system influences pathogenicity island retention in *Pectobacterium atrosepticum* via crRNA generation and Csy complex formation, Biochem. Soc. Trans. 41 (2013) 1468–1474.
[29] S. Mulepati, A. Héroux, S. Bailey, Structural biology. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target, Science 345 (2014) 1479–1484.
[30] K.S. Makarova, S. Karamycheva, S.A. Shah, G. Vestergaard, R.A. Garrett, E. V. Koonin, Predicted highly derived class 1 CRISPR-Cas system in Haloarchaea containing diverged Cas5 and Cas7 homologs but no CRISPR array, FEMS Microbiol. Lett. 366 (2019) (pii: fnz079).
[31] S.D.B. Cass, K.A. Haas, B. Stoll, O.S. Alkhnbashi, K. Sharma, H. Urlaub, R. Backofen, A. Marchfelder, E.L. Bolt, The role of Cas8 in type I CRISPR interference, Biosci. Rep. 35 (2015), e00197.
[32] I. Grissa, G. Vergnaud, C. Pourcel, CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats, Nucleic Acids Res. 35 (2007) W52–W57.
[33] M.H. David, Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization, RNA 10 (2004) 1178–1190.
[34] V. Kunin, R. Sorek, P. Hugenholtz, Evolutionary conservation of sequence and secondary structures in CRISPR repeats, Genome Biol. 8 (2007) R61.
[35] G.S. James, B. Amanda, The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes, J. Mol. Evol. 62 (2006) 718–729.
[36] C. Stephen, RNA-protein complexes, Curr. Opin. Struct. Biol. 9 (1999) 66–73.