Jurnal Kejuruteraan SI 1(7) 2018: 1-9
https://doi.org/10.17576/jkukm-2018-si1(7)-01

Endogenous CRISPR/Cas Systems Prediction: A Glimpse towards Harnessing CRISPR/Cas Machineries for Genetic Engineering

(Ramalan Sistem Endogen CRISPR/Cas: Menuju ke Arah Memanfaatkan CRISPR/Cas Untuk Kejuruteraan Genetik)

Rozieffa Roslan[a], Peer Mohamed Abdul[a,b*], Jamaliah Md Jahim[a,b]
[a]Research Centre for Sustainable Process Technology (CESPRO),
[b]Chemical Engineering Programme
Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia

ABSTRACT

Genetic engineering field has become an imperative approach for enhancement of various bioproducts yield and productivity; and found extended applications in gene therapy, nanotechnology, as well as industrial microbiology. Modern genetic engineering tool CRISPR/Cas system, specifically the Type II system from Streptococcus pyogenes, is gaining traction in recent years and being utilized to engineer novel strains to overproduce primary fermentation product of interest. Employing this technology for non-model microorganism such as Clostridium spp is still restricted due to several limitations such as inadequate genome information, resistance against transformation, low plasmid replication, and the ability for gene expression. The prediction of CRISPR/Cas systems in microbial genomes is fundamentally the initial step towards exploitation of this technology to engineer Clostridium spp. In this study, we demonstrate a simple yet effective method to predict component of endogenous CRISPR/Cas systems, using Clostridium spp genomes as a proof-of-concept. We identified the "real" CRISPR array together with the cas gene operon consist of Type I-B signature proteins in Clostridium pasteurianum which is in agreement with the previous report, implying that this strategy generates reliable CRISPR/Cas systems prediction. Thus, this provides a glimpse on how bioinformatics and biocomputational tools can be utilized to overcome barriers in genetic engineering.

Keyword: Bioinformatics; Non-Model Organism; Bacteria; Archaea; Clostridium; CRISPR Array; Cas Gene

ABSTRAK

Bidang kejuruteraan genetik telah menjadi satu pendekatan yang penting untuk penghasilan dan peningkatan produktiviti pelbagai bioproduk dan aplikasinya ditemui dalam pelbagai bidang termasuk terapi gen, nanoteknologi, serta mikrobiologi dalam industri. Peralatan molekular kejuruteraan genetik sistem CRISPR/Cas, khususnya Jenis II dari Streptococcus pyogenes mendapat perhatian sejak beberapa tahun kebelakangan ini, dan telah digunakan untuk membangunkan strain bakteria novel yang berpotensi menghasilkan produk fermentasi yang dikehendaki. Penggunaan teknologi CRISPR/Cas dalam mikroorganisma bukan-model seperti spesis Clostridium adalah terhad, kerana beberapa batasan seperti maklumat genom yang tidak mencukupi, rintang terhadap transformasi, replikasi plasmid yang rendah, serta keupayaan untuk ekspresi gen. Ramalan sistem CRISPR/Cas yang terdapat di dalam genom mikrob adalah langkah awal ke arah pengeksploitasian teknologi ini untuk mengubahsuai spesis Clostridium. Dalam kajian ini, kami menunjukkan kaedah yang mudah tetapi berkesan untuk meramal komponen sistem CRISPR/Cas, menggunakan genom spesis Clostridium sebagai model. Kami telah mengesan CRISPR array "sebenar" yang disertai dengan operon gen cas dalam Clostridium pasteurianum yang terdiri daripada protein petanda Jenis I-B. Hasil peramalan ini setanding dengan kajian terdahulu, menyifatkan strategi ini mampu menghasilkan ramalan yang tepat. Ketepatan ramalan ini memberi gambaran tentang bagaimana perkakas bioinformatik dan biokomputer dapat digunakan untuk mengatasi permasalahan dalam kejuruteraan genetik.

Kata kunci: Bioinformatik; Organisma Bukan-Model; Bakteria, Arkea, Clostridium, CRISPR Array; Gen Cas

INTRODUCTION

In recent years, the world has turned towards the application of "green chemistry" or "sustainable development", aiming of utilizing microbes for industrial-scale chemical compound production. The genus *Clostridium* spp has long been recognised for its biotechnological potential, and been extensively used in bioconversion of various fermentable carbon sources to value-added products such as 1,3-propanediol (Tee et al. 2017), hydrogen (Alalayah et al. 2008; Abdul et al. 2013), acetone-butanol-ethanol (ABE) (Ibrahim et al. 2012), and etc. *Clostridium* spp is an endospore-forming, Gram-positive anaerobe that is commonly found in natural environment, and a commensal gut microbiota in human and animal alike. Apart from the economically-important strains, the genus *Clostridium* also includes disease-causing,

pathogenic strains such as *C. difficile*, *C. botulinum*, and *C. perfringens* which are associated with infectious diarrhoea, botulism, and food poisoning, respectively (M. Num & Useh 2014).

As more living microorganisms are employed to manufacture biotechnological and chemical products, fermentation technology became essential for numerous industries. In order for industrial fermentation to be economically-viable, high product yield and productivity are critical and must be maintained throughout the fermentation process. Naturally, microorganism has a limited ceiling of maximum product yield and productivity. Conventionally, yield and productivity enhanced through the optimization of biotic and abiotic factors such as oxygenation, temperature, pH, carbon sources, bioreactor design and operation strategy (Kumar & Prasad 2011). However, the microorganisms will produce only in a certain amount of targeted products depending on the its growth and needs. These microbes tend not to overproduce the targeted product of fermentation, hence decreasing the potential of commercial application. To break away from maximum ceiling of product yield and productivity limits, there is a need for microbes to be genetically-engineered prior to commercialisation. The development of genetic engineering technology has imparted new tools over the past few years to improve the overall microbial performance.

Genetic engineering typically leads to genome editing which is basically the insertion (Chung et al. 2016), replacement (Li et al. 2016) and/or deletion (Wang et al. 2016) of DNA in the genome of an organism in order to enhance, eliminate or stimulate metabolic pathways. Genetic engineering has been applied in *Clostridium* spp to enhance the production of major fermentation product via different improvement strategies such as: 1) knocking out genes responsible for competing product of butanol production in *Clostridium thermocellum* using yeast recombineering tool (Rydzak, Lynd & Guss 2015); 2) eliminating butyrate formation pathway in ethanol production using Targetron system and Clostron technology in *Clostridium butyricum* (Cai et al. 2011); 3) knocking down a gene employing antisense RNA for high-yield ethanol production in *Clostridium pasteurianum* (Pyne et al. 2015).

A number of studies show that the manipulation of bacterial genetic content can influence the pathogenicity of disease-causing strain. Microbial genetic manipulation is becoming a new breakthrough that could be used to completely cure diseases. The application of gene editing technologies to potentially abolish disease in pathogenic *Clostridium* spp has long been reported. For example, in *C. perfringens*, the ability to affect rabbit ileal loops has been eliminated by the inactivation of the gene encoding enterotoxin (Sarker, Carman & McClane 1999). Genetic engineering is advancing as the revolutionary genetic engineering tool CRISPR/Cas9 has been introduced, exploiting the adaptive microbial immune system to provide faster targeted genome editing. The concept of editing the genome using CRISPR/Cas9 machineries is practically the same as other genetic engineering tools. The major difference over the previous technologies is that several genes can be edited efficiently in just one attempt. This latest breakthrough has empowered scientific community to successfully modify microbial genomes at quicker rate, thus making it promising for industrial,engineering and medical applications (Hsu, Lander & Zhang 2014; Rath et al. 2015).

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) are the hallmark of host defence mechanisms that are present in bacteria and most archaea. The exposure to foreign genetic element by the way of transformation, conjugation and transduction resulting in the development of CRISPR/Cas systems, an adaptive microbial immune system that provides acquired immunity against viruses and plasmids (Horvath & Barrangou 2010). CRISPR loci typically consist of several repeats separated by spacer, adjacent to *cas* genes encode for polynucleotide binding protein, polymerases, nucleases, and helicases. CRISPR/Cas systems rely on RNA guides (gRNA) which lead the system's Cas protein to degrade target DNA sequence. The targeted sequence must have a 2-6 base pair Protospacer Adjacent Motif (PAM) sequence in order for Cas nuclease protein to function (Barrangou et al. 2007; Horvath & Barrangou 2010; Shah et al. 2013). This system is divided into three stages: 1) adaptation or spacer acquisition; 3) crRNA biogenesis; 3) target interference (Hille & Charpentier 2016).

There are three main groups of CRISPR/Cas system, namely Type I, II, and III, which are characterized by the presence their signature protein Cas3, Cas9 and Cas10, respectively (Makarova et al. 2011; Rath et al. 2015). Nevertheless, Type IV system has been recently proposed and characterized by the absence of CRISPR, *cas1* or *cas2* gene (Koonin & Krupovic 2015). Cas1 and Cas2 are the two universally conserved proteins usually found in all prokaryotic CRISPR (Nuñez et al. 2014). The differences between Type I, Type II and Type III are presented in Figure 1. In Type I system (A), Cas5 or Cas6 is needed for the maturation of precursor crRNA, and target sequence degradation requires Cas3 along with Cascade and crRNA. Type II system (B) requires trans-activating RNA (tracrRNA), RNAse III, along with an unknown factor for crRNA processing. crRNA guides Cas9 to cleave target DNA. In Type III system (C), Cas6 along with unknown factor process the precursor crRNA. Csm (DNA targeting) or Cmr (RNA targeting) complex cleaves DNA or RNA. The type I system is the key type presents in the bacterial genomes, comprising of six subtypes (I-A to I-F) (Pyne et al. 2016; Zhang & Ye 2017). Type II system was only found in bacteria so far, and Type III system is typical in archea (Makarova et al. 2011). Among all, Type II system is the most simple, whereby it only requires a single Cas protein (Cas9) to function (Li et al. 2016). Owing to its simplicity, CRISPR/Cas Type II system from *Streptococcus pyogenes* has become a revolution in genome editing technology. The use of Cas9 is popular in many research laboratories around the world, and it is exploited mainly for gene silencing.

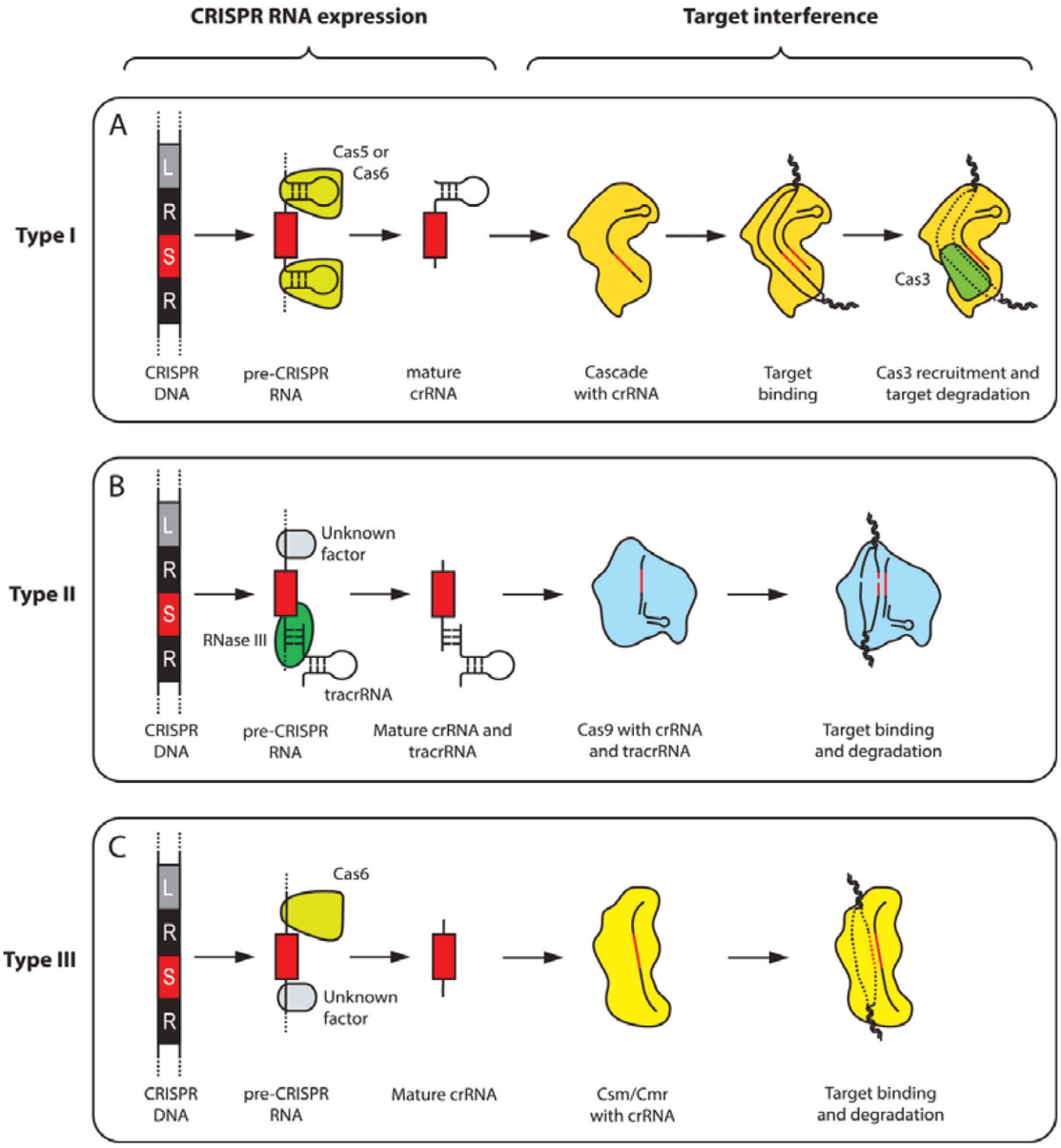

FIGURE 1. Three stages of CRISPR-Cas: adaptation, expression and interference (adapted from Rath et al. 2015)

CRISPR/Cas9 gene editing systems has been successfully applied to Clostridia such as *Clostridium cellulolyticum* (Xu et al. 2015), *Clostridium ljungdahlii* (Huang et al. 2016), *Clostridium autoethanogenum* (Nagaraju et al. 2016), *Clostridium pasteurianum* (Pyne et al. 2016) and *Clostridium acetobutylicum* (Wasels et al. 2017). However, in some *Clostridium* species, the application of Type II CRISPR/Cas9 system has been shown to reduce transformation efficiency by 25% (Pyne et al. 2016), and the overexpression of Cas9 endonuclease is proved to be fairly toxic in some species such as *Schizosaccharomyces pombe* (Jacobs et al. 2014) and *Clostridium pasteurianum* (Pyne et al. 2016), even without the presence of gRNA. CRISPR/Cas9-mediated cell killing is highly unfavourable and raised many concerns among geneticists. For this reason, researchers are finding the right solution to counter this challenging problem. One of the most applicable solutions is to exploit endogenous CRISPR-Cas system in the respective bacterial and archaeal (e.g. *Clostridium pasteurianum* – Type I; *Sulfolobus islandicus* – Type I) (Li et al. 2016; Pyne et al. 2016). However, not all bacteria and archaea possess CRISPR/Cas system, approximately 60% of bacteria and 10% of archaea are lacking of CRISPR/Cas system (Horvath & Barrangou 2010). In order to aim endogenous CRISPR/Cas as a genome editing tool, the foremost crucial stage is to determine if the endogenous CRISPR/Cas system is present or absent in the microorganism of interest. CRISPR/Cas prediction is inevitably one of the key step to find and locate CRISPR/Cas system in a species once it has been sequenced. Currently, there are numerous bacterial strains in which their CRISPR-Cas

systems remain uncharacterized. Hence, a reliable prediction of CRISPR-Cas system in a strain is crucial and fundamental procedure in understanding uncharacterized endogenous CRISPR-Cas system of an organism.

Few prediction tools such as CRISPRFinder (Grissa, Vergnaud & Pourcel 2007), CRT (Bland et al. 2007), and CRISPRone (Zhang & Ye 2017) have been develop to predict endogenous CRISPR/Cas systems in a genome. Each prediction tool implements different methods with diverse prediction algorithm. Hence, the accuracy of the prediction is fairly unknown. In most existing CRISPR identification program such as CRISPRFinder and CRT, the predictions are primarily based on detecting regions with repeat-and-spacer like structures. This method will likely produces false-CRISPRs during prediction, as this structure is quite similar with other elements such as simple repeats, tandom repeats, and STAR-like element (Zhang & Ye 2017). Alternately, CRISPRone program delivers CRISPR/Cas system prediction integrated with false-CRISPR checking pipeline, allowing the identification of real CRISPR with a lower risk of false-positive. Despite of increasing effort devoted to CRISPR/Cas identification, a methodical approach to guide researcher is still lacking. Here, using CRISPRone as a predictor program, a detailed workflow from data collection to CRISPR/Cas systems validation will be explored in this study. This workflow may serve as a guide that assists scientific community toward bioinformatics and biocomputing implementation in their research. We demonstrate the prediction of CRISPR/Cas systems in two Clostridial strains (*Clostridium pasteurianum* and *Clostridium acetobutylinum*) where its endogenous CRISPR/Cas have been readily outlined in laboratory as a proof-of-concept and to validate this guidelines:

METHODOLOGY

Accurate prediction of CRISPR array and *cas* operon is the initial step for finding endogenous CRISPR system in a genome. Hence, choosing a precise prediction program followed by verification process is extremely important to generate reliable data. Here we present our prediction approach which consists of 3 stages: 1) Data collection; 2) CRISPR/Cas prediction; 3) Validation. Detail strategy workflow is illustrated in Figure 2.

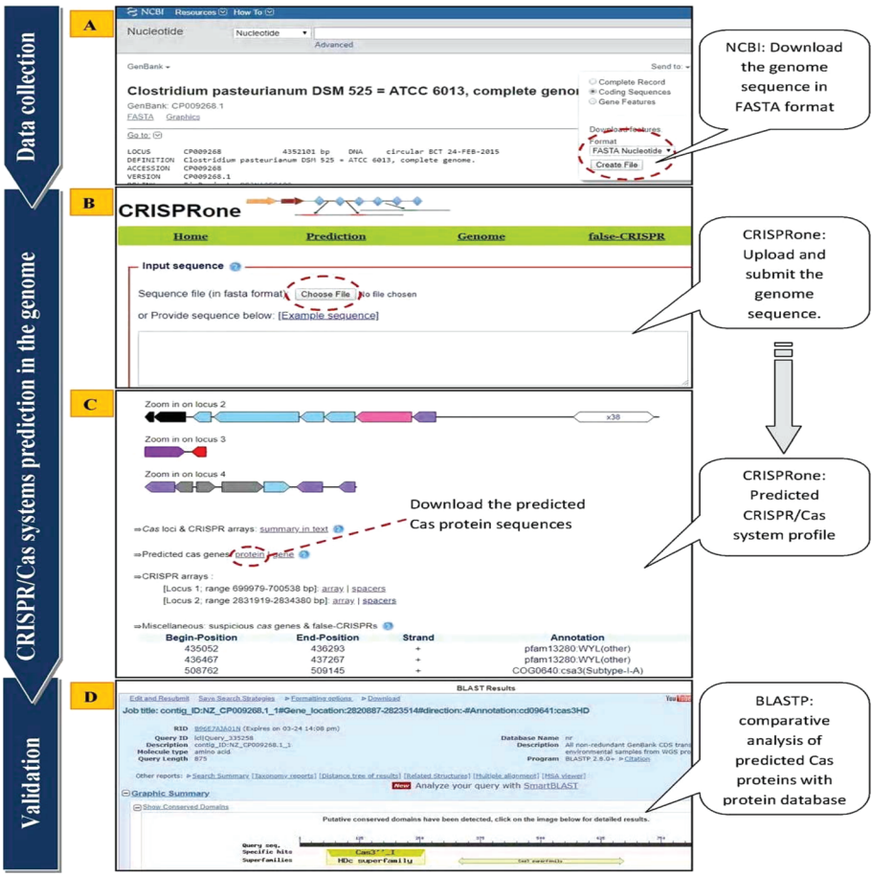

FIGURE 2. Workflow of CRISPR/Cas system prediction in a genome. There are three stages in CRISPR/Cas prediction which are data collection stage, CRISPR/Cas system prediction and validation stage. The prediction procedure is as follow: A) Download raw sequence data from NCBI nucleotide collection. B) Upload a genomic sequence and submit to CRISPRone program. C) Visualization result after analysis complete D) Perform protein comparative analysis via BLAST NCBI to verity the prediction

DATA COLLECTION / DATA MINING

The genome sequence of *Clostridium pasteurianum* DSM 525 (Poehlein et al. 2015) under the accession number CP009268 was obtained from NCBI Reference Sequence (RefSeq). For *Clostridium acetobutylicum* ATCC 824 strain (Nolling et al. 2001), the genome sequence under the accession number AE001437 and the megaplasmid pSOL1 under accession number AE001438 were retrieved from NCBI database (http://www.ncbi.nlm.nih.gov)

CRISPR/CAS SYSTEMS PREDICTION USING CRISPRONE

CRISPR/Cas systems in the genome were predicted using CRISPRone program (omics.informatics.indiana.edu/CRISPRone/). CRISPRone is a bioinformatics tool developed and maintained by Yuzhen Ye, School of Informatics and Computing, Indiana University (Zhang & Ye 2017). It provides online prediction of genomic sequences with integrated false-CRISPR detection.

IDENTIFICATION AND VALIDATION OF PREDICTED CRISPR/CAS SYSTEMS

To develop more accurate result, integrated strategy consists of prediction and validation must be employed. Predicted Cas proteins were validated using comparative analysis (sequence alignment) via NCBI BLASTP program (https://blast.ncbi.nlm.nih.gov/). BLASTP program was used to identify protein that matches to amino acid sequence of predicted Cas proteins. The database was set to non-redundant protein which comprises of SwissProt, PIR (Protein Identification Resource), and PDB (Protein Data Bank). Algorithm parameter was set to default.

RESULTS AND DISCUSSION

DATA COLLECTION

Three sequences (two chromosomal, one plasmid) were used to determine the efficiency CRISPR/Cas system prediction using CRISPRone program (Zhang & Ye 2017). In general, there are two types of genome reference which are complete genomes and draft genomes. This study concentrated on complete genomes, as draft genomes consist of separate independent contigs. The usage of draft genomes may cause a long CRISPR array split into multiple ones, thus, CRISPR/Cas system may be found in separate contigs.

PREDICTION CRISPR/CAS SYSTEM IN BACTERIAL STRAIN MODEL: CLOSTRIDIUM PASTEURIANUM

In this study, CRISPRone program was used to re-analyse the genome in attempt to evaluate the accuracy of this online prediction tool. Detailed prediction analysis data is presented in Table 1. A total of four CRISPR loci (Locus 1-4) and two CRISPR arrays (located at Locus 1 and Locus 2) were identified. Locus 1 consist of CRISPR array containing nine repeats (R1-R9) and eight spacers (S1-S8) with no *cas* gene operon, while Locus 2 consist of CRISPR array containing 37 repeats (R1-R37) and 38 spacers (S9-S45) with the presence of a *cas* gene operon, *predicted as Cas6-Cas8b1-Cas7b-Cas5-Cas3-Cas4-Cas1-Cas2 (Figure 3). Each repeats consist of the same sequence of nucleotides (Table 2) and each repetition was followed by a spacer, as shown in Figure 4. The existent of Cas3 (Type I signature protein),* Cas7b and Cas 8b indicate that this locus is within Type I-B subtype. The nucleases Cas1 (metal-dependent DNase) and Cas2 (metal-dependent endoribonuclease) are two universally conserved proteins, crucial for adaptation. While Cas3 contains HD nuclease domain, catalyzing DNA target cleavage (Makarova et al. 2011). Cas4 assist Cas1-Cas2 complex during adaptation (Kieper et al. 2018). The predicted CRISPR-Cas system also encodes for RAMPs (Repeat-Associated Mysterious Protein) which consist subunits of Cascade complex, Cas5, Cas6 and Cas7. Cas5 and Cas7 groups and were located adjacent to each other. Cas8 is a large subunit of Cascade complex which will interacts with HD domain and a RAMP carrying crRNA (Makarova et al. 2011).

Typically, the CRISPR-Cas locus consists of CRISPR array(s) together with its nearby (within 10,000 bps) *cas* genes (Zhang & Ye 2017). CRISPRs that lack of *cas* genes such as Locus 1 was not classified as real CRISPR. They were called false-CRISPR element, which was probably a genomic element (such as tandem repeats). Thus, Locus 2 was considered to be "real" CRISPR/Cas system. Considering that not all CRISPR/Cas predictions are perfect, the predicted Cas proteins sequences were then validated and confirmed using Protein BLAST program. The comparative analysis revealed that the predicted proteins in Locus 2 have been shown to be CRISPR-associated protein, validated CRISPRone prediction. Whereas, the analysis of Locus 3 and Locus 4 show only the *cas* genes without the companion of CRISPR array within 10,000 bps. A locus which contains at least three *cas* genes without CRISPR array is known as isolated *cas* locus (Zhang & Ye 2017). In this case, Locus 4 was possibly an isolated *cas*locus as it has more than three *cas* genes. However, BLAST comparative analysis of Locus 3 and Locus 4 predicted proteins were completely false-positive, as they turned out to be non-CRISPR-related proteins such as RNA-directed DNA polymerases.

Previous study by Pyne et al. has identified that *Clostridium pasteurianum* possess a Type I-B CRISPR/Cas system containing a 37-spacer CRISPR array upstream of a *cas* gene operon (cas6-cas8b-cas7-cas5-cas3-cas4-cas1-cas2), and an additional 8 spacers that were not associated with *cas* gene operon. Pyne et al. has successfully harness endogenous Type I-B CRISPR/Cas system for gene editing using a plasmid containing synthetic CRISPR expression cassette mimicking the sequence and arrangement of endogenous Type I-B CRISPR array identified in *C. pasteurianum*is in agreement with *C. pasteurianum*'s CRISPR loci reported by Pyne et al., our CRISPR/Cas prediction and validation have proved to be almost identical, and the result can be considered comparable (Figure 5). Here we demonstrate that: (1) choosing the right

TABLE 1. CRISPRone analysis for Clostridium pasteurianum DSM 525

Locus	Location	Predicted	Protein BLAST analysis	BLAST Similarity (%)
1	699979 700538		CRISPR array (Table 2)	
2	2818776 2819063	Cas2	CRISPR-associated endonuclease Cas2 [Clostridium pasteurianum]	100%
2	2819065 2820042	Cas1	type I-B CRISPR-associated endonuclease Cas1 [Clostridium pasteurianum]	100%
2	2820256 2820813	Cas4	CRISPR-associated protein Cas4 [Clostridium pasteurianum]	100%
2	2820887 2823514	Cas3HD	CRISPR-associated helicase/endonuclease Cas3 [Clostridium pasteurianum]	100%
2	2823538 2824275	Cas5	type I-B CRISPR-associated protein Cas5 [Clostridium pasteurianum]	100%
2	2824279 2825223	Cas7b	hypothetical protein [Clostridium pasteurianum]	100%
			* CRISPR-associated protein [Clostridium autoethanogenum]	73%
2	2825245 2826960	Cas8b1	hypothetical protein [Clostridium pasteurianum]	100%
			*type I-B CRISPR-associated protein Cas8b/Csh1 [Clostridium tetani]	47%
2	2827001 2827708	Cas6	hypothetical protein [Clostridium pasteurianum]	100%
			*CRISPR-associated protein [Clostridium autoethanogenum]	64%
2	2831919 2834380		CRISPR array (data not shown)	100%
3	4122464 4123708	Cas5u	IS110 family transposase [Clostridium pasteurianum]	100%
3	4123898-4124344	Cas3	transcriptional regulator [Clostridium pasteurianum]	100%
4	4136582 4137508	RT	RNA-directed DNA polymerase [Clostridium pasteurianum	100%
4	4137495 4138049	unknown	hypothetical protein [Clostridium pasteurianum]	100%
			*group II intron reverse transcriptase/maturase [Clostridium lundense]	88%
4	4138181 4138753	unknown	hypothetical protein [Clostridium pasteurianum]	100%
4	4138953 4140215	unknown	transposase of IS1604-like element [Clostridium pasteurianum	100%
4	4140231 4141037	Cas3	DUF2075 domain-containing protein [Clostridium pasteurianum]	100%
4	4141239 4142042	RT	RNA-directed DNA polymerase [Clostridium pasteurianum]	100%
4	4142526 4143023	RT	RNA-directed DNA polymerase [Clostridium pasteurianum]	100%

TABLE 2. Locus 1 CRISPR array predicted by CRISPRone program

Position	Repeat Sequence	Spacer Sequence
699979	R1: GTTGAACCTTAACATAGGATGTATTTAAAT	S1: CAGATAATGCTACATGGAGAGGGGCTATAACACAAG
700045	R2: GTTGAACCTTAACATAGGATGTATTTAAAT	S2: TAATCATTATTTTCTCCTAACCAAATTCCATATTTC
700111	R3: GTTGAACCTTAACATAGGATGTATTTAAAT	S3: ATATCAGGACAGTCATCTAATCTGCTAGTGTTTAG
700176	R4: GTTGAACCTTAACATAGGATGTATTTAAAT	S4: TGGAATAGTTAATACACAACTTACTTTTGAAGATTT
700242	R5: GTTGAACCTTAACATAGGATGTATTTAAAT	S5: TAATAATACCTTCCTGCAAATTCATAATTTTTTTGA
700308	R6: GTTGAACCTTAACATAGGATGTATTTAAAT	S6: TAATTTGGAAAACCTATATGAAAAGAGGATAGTGCA
700374	R7: GTTGAACCTTAACATAGGATGTATTTAAAT	S7: ACAGCAGTTGCAAATAATGCAACTGCTACAGTAC
700438	R8: GTTGAACCTTAACATAGGATGTATTTAAAT	S8: CTACTTGGTTAAATGAATTAAGGAGAGAATAAT
700509	R9: GTTGAACCTTAACATAGGATGTATTTAAAC	AAGATATG

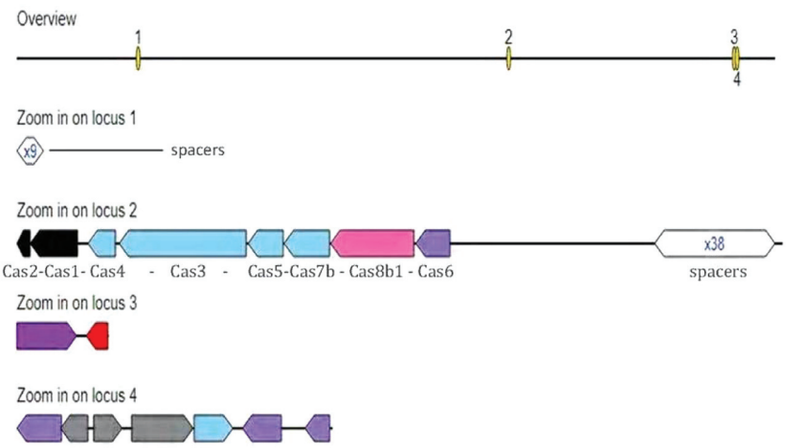

FIGURE 3. Visualization of predicted CRISPR/Cas systems in *Clostridium pasteurianum* DSM 525 by CRISPRone. *Cas* genes are colour-coded according to types or subtypes

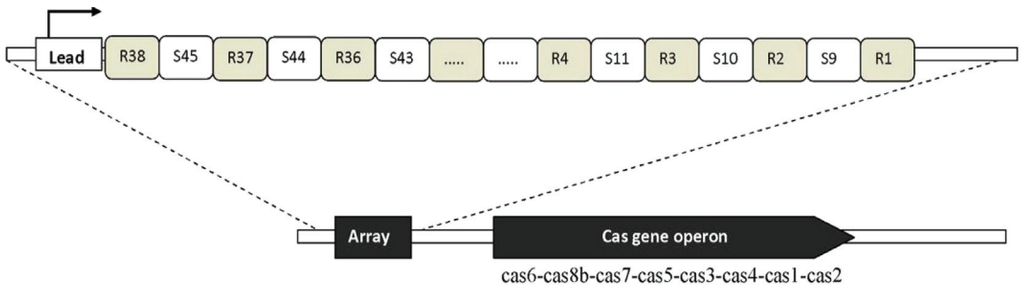

FIGURE 4. Schematic representation based on CRISPR prediction in this study to illustrate a true CRISPR array (Locus 2) in Clostridium pasteurianum DSM 525 genome based on prediction data in this study. CRISPR array usually starting with a leader sequence, followed by an array of repeated sequences (R1-R38) separated by spacers (S9-S45), and a set of cas genes encoding the Cas proteins

CRISPR/Cas identification tool with low-risk of false positive is definitely the most vital; (2) validating stage is important to ensure the complete accuracy of an analysis. This 3-steps strategy can be readily implemented as initial step to study endogenous CRISPR/Cas in bacterial and archeal genomes.

PREDICTION CRISPR/CAS SYSTEM IN CLOSTRIDIUM ACETOBUTYLICUM

CRISPRone analysis runs on *Clostridium acetobutylicum* ATCC 824 genome produce no result, as it cannot detect any CRISPR element in the genome. For its megaplasmid pSOL sequence, one false-CRISPR element consist of five *cas* genes were predicted. Nevertheless, protein comparative analysis reveals that those predicted proteins do not belong to CRISPR-associated protein (data not shown). Based on prediction and validation result, we can assure that *Clostridium acetobutylicum* lacks of CRISPR/Cas system. *Clostridium acetobutylicum* documented to be one of the microbes which devoid of endogenous CRISPR/Cas loci (Pyne et al. 2016). This prediction was coherent with previously reported result, ruling out the possibility of false prediction. Previous analysis reveals that only 10% of 1724 sample microorganisms possess endogenous CRISPR/Cas system (Burstein et al. 2016). The absent of CRISPR/Cas system was hypothesized to be correlated with a symbiotic lifestyle of an organism. Burstein et al. reported that organisms which lack of CRISPR/Cas such as Chlamydiae shared a general feature of symbionts. Chlamydiae members are obligate bacteria and known as symbionts in eukaryotes (Horn 2008). To further examine this hypothesis, we searched for previous studies conducting on symbionts feature in *Clostridium acetobutylicum*. A study from (Wang et al. 2015) discloses that *Clostridium acetobutylicum* and *Bacillus cereus* are found to be coexisting in a symbiotic system TSH06 isolated from corn powder. Whereas, *Clostridium pasteurianum* has long been known as asymbiotic (non-symbiotic) nitrogen fixing bacteria (Silver & Postgate 1973). Based on this observation, the hypothesis from Burstein et al. (2016) on correlation between the absent of CRISPR/Cas with symbionts could possibly be accepted.

Since not all organisms possess CRISPR, hijacking endogenous CRISPR/Cas system for gene editing is therefore might not possible in certain organisms. For organisms devoid of endogenous CRISPR loci, the usage of current exogenous Type II CRISPR/Cas machinery is unavoidable. Hence, inducible CRISPR genome editing tool is highly recommended to deal with the toxicity of constitutively-expressed Cas9 in the first generation of CRISPR/Cas9 gene editing technology (Pyne et al. 2016; Dai et al. 2018). By using inducible approach, the enzyme Cas9 is expressed only when needed. Inducible CRISPR systems consist of inducible promoters which are regulated chemically or physically. Chemically-induced promoters are regulated by chemicals such as antibiotics, IPTG, steroids, alcohol, etc while physically-induced promoters are regulated by environmental alterations and stresses such as temperature, salt stress, light, etc (Dai et al. 2018). For the fact that harnessing endogenous CRISPR is impossible in certain microbes, the employment of inducible CRISPR genome editing tool is definitely an alternative to constitutively-expressed CRISPR/Cas9 system.

CONCLUSION

Essentially, we have outlined here a strategy to predict the endogenous CRISPR/Cas in Clostridia that has been proved to be reliable and coherent with the previous data, in which endogenous CRISPR identification leads to the hijacking of CRISPR machineries for genome editing. Thereby, this study highlights the application of integrated strategy in generating precise analysis data, providing an insight into utilization online bioinformatics tools to predict microbial CRISPR/Cas systems. Given the situation that only CRISPR/Cas system type II is currently available for gene editing, the other CRISPR/Cas systems need to be explored and characterized in order to overcome the CRISPR/Cas9-mediated cell killing in certain bacterial species. Thus, the detection of endogenous CRISPR array and *cas* gene operon in bacterial genome essentially will aid scientific community to develop next generation CRISPR/Cas gene editing tools. Moreover, the CRISPR prediction will inspire researchers to aim for a suitable and cost-effective gene editing strategy in the microbial of interest.

ACKNOWLEDGEMENT

We gratefully acknowledge the financial and technical support provided by Universiti Kebangsaan Malaysia for funding this work through Geran Galakan Penyelidik Muda (GGPM-2017-041) and Skim Zamalah Penyelidikan UKM PPI/244/43/24(P92856).

REFERENCES

Abdul, P. M., Md. Jahim, J., Harun, S., Markom, M., Hassan, O., Mohammad, A. W. & Asis, A. J. 2013. Biohydrogen production from pentose-rich oil palm empty fruit bunch molasses. *International Journal of Hydrogen Energy* 38(35): 15693-15699.

Alalayah, W. M., Kalil, M. S., Kadhum, A. A. H., Jahim, J. M. & Alauj, N. M. 2008. Hydrogen production using *Clostridium saccharoperbutylacetonicum* N1-4 (ATCC 13564). *International Journal of Hydrogen Energy* 33(24): 7392-7396.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A. & Horvath, P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315(5819): 1709-1712.

Bland, C., Ramsey, T. L., Sabree, F., Lowe, M., Brown, K., Kyrpides, N. C. & Hugenholtz, P. 2007. CRISPR Recognition Tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* 8(1): 209.

Burstein, D., Sun, C. L., Brown, C. T., Sharon, I., Anantharaman, K., Probst, A. J., Thomas, B. C. & Banfield, J. F. 2016. Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. *Nature Communications* 7: 10613.

Cai, G., Jin, B., Saint, C. & Monis, P. 2011. Genetic manipulation of butyrate formation pathways in Clostridium butyricum. *Journal of Biotechnology* 155(3): 269-274.

Chung, M. E., Yeh, I. H., Sung, L. Y., Wu, M. Y., Chao, Y. P., Ng, I. S. & Hu, Y. C. 2016. Enhanced integration of large DNA into E. coli chromosome by CRISPR/Cas9. *Biotechnology and Bioengineering* 114(1): 172-183.

Dai, X., Chen, X., Fang, Q., Li, J. & Bai, Z. 2018. Inducible CRISPR genome-editing tool: classifications and future trends. *Critical Reviews in Biotechnology* 38(4): 573-586.

Grissa, I., Vergnaud, G. & Pourcel, C. 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Research* 35: W52-W57.

Hille, F. & Charpentier, E. 2016. CRISPR-Cas: biology, mechanisms and relevance. *Philosophical Transactions of the Royal Society B: Biological Sciences* 371(1707): 20150499.

Horn, M. 2008. Chlamydiae as symbionts in eukaryotes. *Annu Rev Microbiol* 62: 113-131.

Horvath, P. & Barrangou, R. 2010. CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327(5962): 167-170.

Hsu, P. D., Lander, E. S. & Zhang, F. 2014. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157(6): 1262-1278.

Huang, H., Chai, C., Li, N., Rowe, P., Minton, N. P., Yang, S., Jiang, W. & Gu, Y. 2016. CRISPR/Cas9-Based Efficient Genome Editing in Clostridium ljungdahlii, an Autotrophic Gas-Fermenting Bacterium. *ACS Synthetic Biology* 5(12): 1355-1361.

Ibrahim, M. F., Abd-Aziz, S., Razak, M. N. A., Phang, L. Y. & Hassan, M. A. 2012. Oil palm empty fruit bunch as alternative substrate for acetone–butanol–ethanol production by Clostridium butyricum EB6. *Applied Biochemistry and Biotechnology* 166(7): 1615-1625.

Jacobs, J. Z., Ciccaglione, K. M., Tournier, V. & Zaratiegui, M. 2014. Implementation of the CRISPR-Cas9 system in fission yeast. *Nature Communications* 5: 5344-5344.

Kieper, S. N., Almendros, C., Behler, J., McKenzie, R. E., Nobrega, F. L., Haagsma, A. C., Vink, J. N. A., Hess, W. R. & Brouns, S. J. J. 2018. Cas4 Facilitates PAM-Compatible Spacer Selection during CRISPR Adaptation. *Cell Reports* 22(13): 3377-3384.

Koonin, E. V. & Krupovic, M. 2015. Evolution of adaptive immunity from transposable elements combined with innate immune systems. *Nature Reviews Genetics* 16(3): 184-192.

Kumar, R. R. & Prasad, S. 2011. Metabolic engineering of bacteria. *Indian Journal of Microbiology* 51(3): 403-409.

Li, J., Meng, X., Zong, Y., Chen, K., Zhang, H., Liu, J., Li, J. & Gao, C. 2016. Gene replacements and insertions in rice by intron targeting using CRISPR–Cas9. *Nature Plants* 2: 16139.

Li, Y., Pan, S., Zhang, Y., Ren, M., Feng, M., Peng, N., Chen, L., Liang, Y. X. & She, Q. 2016. Harnessing Type I and Type III CRISPR-Cas systems for genome editing. *Nucleic Acids Research* 44(4): e34-e34.

Makarova, K. S., Aravind, L., Wolf, Y. I. & Koonin, E. V. 2011. Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biology Direct* 6: 38-38.

Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F. J., Wolf, Y. I., Yakunin, A. F., van der Oost, J. & Koonin, E. V. 2011. Evolution and classification of the CRISPR-Cas systems. *Nature Reviews Microbiology* 9(6): 467-477.

Nagaraju, S., Davies, N. K., Walker, D. J. F., Köpke, M. & Simpson, S. D. 2016. Genome editing of Clostridium autoethanogenum using CRISPR/Cas9. *Biotechnology for Biofuels* 9(1): 219.

Nolling, J., Breton, G., Omelchenko, M. V., Makarova, K. S., Zeng, Q., Gibson, R., Lee, H. M., Dubois, J., Qiu, D., Hitti, J., Wolf, Y. I., Tatusov, R. L., Sabathe, F., Doucette-Stamm, L., Soucaille, P., Daly, M. J., Bennett, G. N., Koonin, E. V. & Smith, D. R. 2001. Genome sequence

and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum. *Journal of Bacteriology* 183(16): 4823-4838.

Num, S. M. & Useh, N. M. 2014. Clostridium: pathogenic roles, industrial uses and medicinal prospects of natural products as ameliorative agents against pathogenic species. *Jordan Journal of Biological Sciences* 7(2): 0008220.

Nuñez, J. K., Kranzusch, P. J., Noeske, J., Wright, A. V., Davies, C. W. & Doudna, J. A. 2014. Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. *Nature Structural & Molecular Biology* 21(6): 528-534.

Poehlein, A., Grosse-Honebrink, A., Zhang, Y., Minton, N. P. & Daniel, R. 2015. Complete genome sequence of the nitrogen-fixing and solvent-producing Clostridium pasteurianum DSM 525. *Genome Announc* 3(1): 01591-14

Pyne, M., Moo-Young, M., Chung, D. & Chou, C. 2015. Antisense-RNA-Mediated Gene Downregulation in Clostridium pasteurianum. *Fermentation* 1(1): 113-116.

Pyne, M. E., Bruder, M. R., Moo-Young, M., Chung, D. A. & Chou, C. P. 2016. Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in Clostridium. *Scientific Reports* 6: 25666.

Rath, D., Amlinger, L., Rath, A. & Lundgren, M. 2015. The CRISPR-Cas immune system: Biology, mechanisms and applications. *Biochimie* 117: 119-128.

Rydzak, T., Lynd, L. R. & Guss, A. M. 2015. Elimination of formate production in *Clostridium thermocellum*. *Journal of Industrial Microbiology and Biotechnology* 42(9): 1263-1272.

Sarker, M. R., Carman, R. J. & McClane, B. A. 1999. Inactivation of the gene (cpe) encoding Clostridium perfringens enterotoxin eliminates the ability of two cpe-positive C. perfringens type A human gastrointestinal disease isolates to affect rabbit ileal loops. *Molecular Microbiology* 33(5): 946-958.

Shah, S. A., Erdmann, S., Mojica, F. J. M. & Garrett, R. A. 2013. Protospacer recognition motifs. *RNA Biology* 10(5): 891-899.

Silver, W. S. & Postgate, J. R. 1973. Evolution of asymbiotic nitrogen fixation. *Journal of Theoretical Biology* 40(1): 1-10.

Tee, Z. K., Jahim, J. M., Tan, J. P. & Kim, B. H. 2017. Preeminent productivity of 1,3-propanediol by Clostridium butyricum JKT37 and the role of using calcium carbonate as pH neutraliser in glycerol fermentation. *Bioresource Technology* 233: 296-304.

Wang, G., Wu, P., Liu, Y., Mi, S., Mai, S., Gu, C., Wang, G., Liu, H., Zhang, J., Borresen, B. T., Mellemsaether, E. & Kotlar, H. K. 2015. Isolation and characterisation of non-anaerobic butanol-producing symbiotic system TSH06. *Applied Microbiology and Biotechnology* 99(20): 8803-8813.

Wang, Q., Lu, Y., Xin, Y., Wei, L., Huang, S. & Xu, J. 2016. Genome editing of model oleaginous microalgae Nannochloropsis spp. by CRISPR/Cas9. *The Plant Journal* 88(6): 1071-1081.

Wasels, F., Jean-Marie, J., Collas, F., Lopez-Contreras, A. M. & Lopes Ferreira, N. 2017. A two-plasmid inducible CRISPR/Cas9 genome editing tool for Clostridium acetobutylicum. *Journal of Microbiological Methods* 140: 5-11.

Xu, T., Li, Y., Shi, Z., Hemme, C. L., Li, Y., Zhu, Y., Van Nostrand, J. D., He, Z. & Zhou, J. 2015. Efficient Genome Editing in Clostridium cellulolyticum via CRISPR-Cas9 Nickase. *Applied and Environmental Microbiology* 81(13): 4423-4431.

Zhang, Q. & Ye, Y. 2017. Not all predicted CRISPR–Cas systems are equal: isolated cas genes and classes of CRISPR like elements. *BMC Bioinformatics* 18: 92.

*Peer Mohamed Abdul, Rozieffa Roslan,
Jamaliah Md Jahim
Research Centre for Sustainable Process Technology (CESPRO),
Faculty of Engineering & Built Environment,
Universiti Kebangsaan Malaysia, Bangi, Malaysia.

*Corresponding author; email:
peer@ukm.edu.my

Received date: 6th April 2018
Accepted date: 23rd July 2018
Online First date: 1st October 2018
Published date: 30th November 2018