https://doi.org/10.48048/tis.2024.7603

Identification of CRISPR/Cas12a (Cpf1) guideRNA Sequence Targeting the Mitochondrial DNA D-loop Region in Wild Pig *(Sus scrofa)* Through Homology Difference and Mismatch Analysis

Muhammad Ihda Hamlu Liwaissunati Zein, Ari Hardianto, Irkham and Yeni Wahyuni Hartati*

Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Jawa Barat 45363, Indonesia

(*Corresponding author's e-mail: yeni.w.hartati@unpad.ac.id)

Received: 12 October 2023, Revised: 6 December 2023, Accepted: 12 December 2023, Published: 20 March 2024

Abstract

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeat) or CRISPR-associated (Cas) System has become a major gene editing tool. Gene editing with CRISPR requires the Cas protein and the corresponding guide RNA (gRNA). However, low cleavage efficiency and off-target effects can impede the application of the CRISPR/Cas system. Therefore, the determination of specific gRNAs is essential. In biosensor applications, CRISPR/Cas12a can enhance specificity and sensitivity in identifying target genes due to the trans-cleavage activity of Cas12a (Cpf1). The mtDNA D-loop sequence is the most variable part of mtDNA, making it suitable for distinguishing between species. Consequently, the objective of this study was to determine the gRNA sequence of the D-loop of wild pig mtDNA *in silico*. Candidate gRNAs were predicted using the Benchling application with the assistance of the GenBank database. The gRNA candidates were subsequently subjected to a homology difference analysis using BLAST nucleotide and a mismatch test using Jalview. Among several candidates, candidate 1 was selected as the best option, with an Off-target value of 99.8. The homology difference analysis against competitors and the mismatch test against the Sus genus resulted in high E-values and high percentage values, respectively. This suggests that the candidate will not recognize other species but can detect members of the Sus scrofa species. These gRNA candidates can be selectively and sensitively applied to biosensors for the detection of meat adulteration.

Keywords: Benchling, gRNA CRISPR/Cas12a, Biosensor, mtDNA D-loop, Wild pig (Sus scrofa), *In silico*

Introduction

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeat)/CRISPR-associated system (Cas) now is a major genome editing tool. Compared to previous genome editing methods such as zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs), which use protein as a target and cut specific genomic loci, CRISPR has the advantage of being cheaper both in terms of time and price [1-3].

CRISPR/Cas encompasses several proteins, including Cas9 and Cas12a (often referred to as Cpf1). It serves as a versatile genome editing tool capable of manipulating, detecting, and describing DNA and RNA sequences within cells. The CRISPR/Cas system was initially harnessed for genome editing in 2012 [1], building upon earlier research that merged CRISPR RNA (crRNA) with trans-activating RNA (tracrRNA) to create a single guide RNA (sgRNA) in Cas9. In this Cas12a, guide RNA (gRNA) interacts with Cas proteins and target DNA sequences. Upon recognizing the protospacer adjacent motif (PAM) sequence, the Cas protein-gRNA complex binds to the target genomic locus, activating the RuVc nuclease domain, which cleaves the target double-stranded DNA (dsDNA) and nearby single-stranded DNA (ssDNA). This distinctive mechanism of CRISPR/Cas12a finds application in biosensor detection [4].

CRISPR/Cas12a belongs to the type 5 CRISPR nuclease family and recognizes the 5'-TTTV-3' PAM sequence, demonstrating high efficiency in the recognition of animal and plant organisms. In addition to Cas12a, several other Cas proteins such as Cas9, Cas12b, Cas13a, and Cas14 have been identified and employed for genome editing, each with unique characteristics [5]. However, as highlighted in previous studies, there are 2 primary challenges associated with the development and application of CRISPR/Cas systems: The risk of potential off-target effects and the need to enhance on-target efficiency [3]. Therefore, the design of specific and sensitive CRISPR guide RNAs (gRNAs) is essential to address these challenges.

In biosensor applications, the utilization of the CRISPR/Cas system can enhance both specificity and sensitivity owing to the trans-cleavage activity exhibited by the Cas protein against the target molecule. This approach is particularly effective when targeting specific DNA sequences, such as wild pig DNA. For instance, researchers have explored the use of mitochondrial DNA as a target due to its high variability and several advantages [6-8].

Mitochondrial DNA (mtDNA) is commonly used to distinguish animal species due to several reasons. Firstly, mtDNA has a high degree of molecular diversity, making it a valuable marker for species identification [9]. Furthermore, mtDNA has a higher mutation rate compared to nuclear DNA, which contributes to its utility as a molecular marker [10]. Forensic genetics also relies on mtDNA analysis for species identification [6]. In addition, mtDNA's clonal mode of inheritance, primarily through the maternal line, makes it a reliable marker for species identification [11].

The diversity apparent in mitochondrial DNA sequences found within both the cytochrome b gene (*cyt b*) and the D-loop region renders them valuable assets for distinguishing between species [12]. Additionally, the fact that mitochondrial DNA is present in thousands of copies per cell has made it a target in numerous mitochondrial gene databases, where researchers study its variability, including intraspecific variation.

The wild pig, scientifically known as *Sus scrofa*, serves as a crucial source of meat, particularly pork, in human diets globally. However, wild pigs are sometimes implicated in food adulteration cases due to their versatile nature. Instances of fraud may involve the substitution of pork for other, less expensive or culturally sensitive meats, especially in processed products like sausages and hams [13,14]. Such adulteration can deceive consumers and compromise religious or cultural dietary restrictions. While not universal, these cases underscore the importance of stringent food safety regulations and monitoring systems to prevent deceptive practices in the food industry and ensure accurate product labeling [15].

Furthermore, the design of guide RNA (gRNA) sequences, which is typically accomplished computationally (*in silico*), is facilitated by these characteristics [16]. In this study, D-loop mtDNA was employed as a database to design a guide RNA (gRNA) to provide strategies for using CRISPR/Cas12a with biosensors and determine gRNA that can optimally detect wild pig mtDNA.

Materials and methods

Wild pig (*Sus scrofa*) mitochondrial D-loop DNA database

The overall animal DNA database was obtained from the GenBank database with the website https://www.ncbi.nlm.nih.gov/genbank/. The wild pig mitochondrial DNA database was obtained with the code *Sus scrofa* mitochondrion, complete genome (NC-00845.1).

CRISPR/Cas12a gRNA prediction

gRNA is a combination of Cas12a sequences that have been fixed depending on the origin of the Cas protein obtained. In our study, we used a recombinant Cas12a derived from *Acidaminococcus sp. BV3L6* purchased from GenScript (https://www.genscript.com/enzyme/Z03502-GenCRISPR_Cas12a_Cpf1_Nuclease.html).

The sequence is 5'-UAA UUU CUA CUC UUG UAG AU-3'. In CRISPR/Cas12a gRNA prediction, further sequences that can hybridize specifically to the target are required. These sequences can be predicted using the open-source application *Benchling* with the domain https://www.benchling.com.

Homology difference analysis and Mismatch analysis

Homology difference analysis was conducted using a web-based open-source application, BLAST nucleotide: Basic Local Alignment Search Tool with the domain https://blast.ncbi.nlm.nih.gov/Blast.cgi. Homology difference analyses were conducted on several species of mammals, rodents, poultry, and fish. Meanwhile, the mismatch analysis was carried out by using the open-source application, JalView with the domain https://www.jalview.org. to see its similarity with the *Sus* subspecies.

Results and discussion

CRISPR/Cas12a gRNA candidate determination

Using prediction tools to determine gRNA is crucial for efficient and specific genome editing with CRISPR/Cas systems [17]. These tools analyze target DNA sequences and identify potential sites for gRNA binding based on criteria like the presence of a PAM. By predicting binding efficiency and off-target effects, these tools reduce the need for extensive experimental testing, allowing researchers to focus on the most promising gRNA candidates. The computational analysis provided by prediction tools not only streamlines

the selection process but also guides the optimization of experimental design, considering factors such as gRNA sequence composition and target site accessibility. This strategic use of prediction tools ensures the customization of gRNA selection for specific genome editing applications, ultimately enhancing the precision and success of the editing process [18].

Based on the prediction results of CRISPR/Cas12a gRNA candidates for wild pig mtDNA detection, a total of 76 gRNA candidates were obtained with varying off-score values. the off-target score value refers to the likelihood of the guide RNA (gRNA) binding to unintended genomic sites. The off-target score is used to assess the specificity of the gRNA and predict potential off-target effects. A higher off-score value indicates a smaller probability that the competitor will hybridize with the gRNA [19]. The results of the gRNA candidate prediction can be seen in **Table 1**.

Table 1 Prediction results of gRNA candidates using Benchling.

Position	Type	Sequences	PAM	Off-score target
643	−	GATTGTCGTGCCGGATCATG	TTTG	99.79302
1163	+	AGTATACGCCCACAATCTGA	TTTA	99.01922
838	−	CCTGCTTATATGCGCGTGTA	TTTA	98.93631
585	+	GGGATGCTTAGACTCAGCCA	TTTG	97.69425
1074	−	TTTATGGTAGATTGGCGTAA	TTTA	97.54287
1093	+	TACGCCAATCTACCATAAAT	TTTT	96.82005
1094	+	ACGCCAATCTACCATAAATA	TTTT	96.12975
1095	+	CGCCAATCTACCATAAATAA	TTTA	94.99655
963	−	GTGTAAGTTAGGCTTATTGT	TTTA	94.30631
1000	−	ATAAGATATGGTGGTTTGTG	TTTT	94.07952
1014	+	ACAACACAAACCACCATATC	TTTA	94.06573
1070	−	TGGTAGATTGGCGTAAAAAT	TTTA	93.91279
74	−	TGCGCAATTTTTATGGGTTT	TTTG	93.4048
253	−	ACATGATATGTGCTATGTAC	TTTG	92.10233
1026	−	GTAGCACGTATTTAAGTAAG	TTTC	92.0227
583	+	TGGGGATGCTTAGACTCAGC	TTTT	91.86147
97	−	GGGGTCACATATTTGTATGT	TTTT	91.83355
10	−	CATACGAATGGAATGCTTGT	TTTG	90.28348
98	−	TGGGGTCACATATTTGTATG	TTTT	89.70734
999	−	TAAGATATGGTGGTTTGTGT	TTTA	89.44644
578	+	ATTTTTGGGGATGCTTAGAC	TTTA	88.47047
576	+	TAATTTTTGGGGATGCTTAG	TTTT	88.23871
96	−	GGGTCACATATTTGTATGTT	TTTG	86.95539
577	+	AATTTTTGGGGATGCTTAGA	TTTT	85.95712
575	+	TTAATTTTTGGGGATGCTTA	TTTT	84.36448
113	−	CAATGGTAAAATTTTTGGGG	TTTT	84.29109
112	−	AATGGTAAAATTTTTGGGGT	TTTC	83.76505
82	−	TATGTTTGTGCGCAATTTTT	TTTG	81.50934
1120	+	AAATTACTACACAATAACCT	TTTA	81.42173

Position	Type	Sequences	PAM	Off-score target
119	−	GGTTTTCAATGGTAAAATTT	TTTT	78.56816
982	−	TGTTGTTAAATGTTGTTTAG	TTTG	77.81008
44	−	AAGATAGTAATTAAGTACTT	TTTA	76.99493
120	−	TGGTTTTCAATGGTAAAATT	TTTT	76.98143
145	+	CCATTGAAAACCAAAAAATC	TTTA	76.87959
144	+	ACCATTGAAAACCAAAAAAT	TTTT	76.19705
121	−	TTGGTTTTCAATGGTAAAAT	TTTT	75.63295
1012	−	AGTAAGTGTTTTATAAGATA	TTTA	75.52645
45	−	AAAGATAGTAATTAAGTACT	TTTT	75.51119
49	−	TTTTAAAGATAGTAATTAAG	TTTG	74.05303
1080	−	AATTTATTTATGGTAGATTG	TTTA	73.39937
63	−	TATGGGTTTTTTTGTTTTAA	TTTT	71.99859
118	−	GTTTTCAATGGTAAAATTTT	TTTG	71.69421
52	−	TTGTTTTAAAGATAGTAATT	TTTT	71.44159
51	−	TGTTTTAAAGATAGTAATTA	TTTT	71.22887
84	+	AAACAAAAAAACCCATAAAA	TTTA	70.07588
1081	−	AAATTTATTTATGGTAGATT	TTTT	67.677
61	−	TGGGTTTTTTTGTTTTAAAG	TTTA	66.85714
574	−	ACGGCCATGGCTGAGTCTAA	TTTG	65.81544
918	−	GGGTTTGGCAAGGCGTTATA	TTTG	65.63826
584	+	GGGGATGCTTAGACTCAGCC	TTTT	65.48243
911	−	GCAAGGCGTTATAGGGTGTG	TTTG	65.3896
15	−	GTTTGCATACGAATGGAATG	TTTG	64.76549
62	−	ATGGGTTTTTTTGTTTTAAA	TTTT	64.63995
16	−	GGTTTGCATACGAATGGAAT	TTTT	64.54209
53	−	TTTGTTTTAAAGATAGTAAT	TTTT	62.91194
938	−	TACACTCTGCTTTGTTTTTG	TTTG	57.0696
855	−	AATGAGCTAATAATTTACCT	TTTG	56.17003
50	−	GTTTTAAAGATAGTAATTAA	TTTT	52.00342
919	−	GGGGTTTGGCAAGGCGTTAT	TTTT	49.75732
367	−	TGGGCCCGGAGCGAGAAGAG	TTTA	49.53062
318	−	ACGCGGCATGGTAGTTAAGC	TTTC	49.39631
876	−	ATGGGGGGTAAGGGGGGTTT	TTTA	48.6617
592	−	ACTGTGTTAGGGCCTTTGAC	TTTG	48.64616
920	−	TGGGGTTTGGCAAGGCGTTA	TTTT	39.47614
422	+	TATTGATGAACTTTAACAGG	TTTC	38.21115
568	+	GTATTTTTTAATTTTTGGGG	TTTG	29.63709
924	−	TTTTTGGGGTTTGGCAAGGC	TTTG	28.05501

Position	Type	Sequences	PAM	Off-score target
437	+	ACAGGCATCTGGTTCTTACT	TTTA	27.80165
458	−	AGGGGAAAGAGTGGGCGATT	TTTA	27.47914
559	+	ATACATTTGGTATTTTTTAA	TTTC	21.96633
435	−	GGTGAGATGGTCCTGAAGTA	TTTA	21.72432
492	+	CCCTTAAATAAGACATCTCG	TTTC	21.20244
436	−	AGGTGAGATGGTCCTGAAGT	TTTT	19.58801

Referring to **Table 1**, the initial phase of our analysis focused on the selection of the top 10 CRISPR/Cas12a guide RNA candidates. These candidates were then subjected to a more detailed examination using the BLAST nucleotide method, specifically against *Bos taurus*, commonly known as cattle. This step was significant as beef, originating from cattle, is frequently susceptible to adulteration with pork, often driven by the aim of reducing costs. The selection of candidates was prioritized based on their off-score values, with the highest off-score positioned at the top, followed by progressively lower values. To delve deeper into the specificity and potential cross-reactivity with cattle DNA, the BLAST results for these first 10 candidates are outlined in **Table 2**, providing a comprehensive insight into their compatibility with the primary competitor, *Bos taurus*. This meticulous analysis is crucial for ensuring the effectiveness of the chosen CRISPR/Cas12a guide RNA candidates in differentiating wild pig DNA from cattle DNA, particularly in scenarios where food authenticity, such as in beef products, is of paramount importance.

In theory, the CRISPR/Cas system identifies specific PAM sequences in target locations and triggers the endonuclease activity to cut DNA at precise sites. However, the efficiency of this cutting process is significantly influenced by various factors, including the design of the gRNA, the Cas protein, the method of delivery, differences in the target sites, and the potential for off-target effects [3,20-22]. This complexity implies that multiple elements can impact how well the gRNA-Cas complex binds to and cuts the DNA. As highlighted by Doench & Root (2016), the selection of an effective gRNA involves considering several key factors: The gRNA sequence itself (including sequence composition, nucleotide position, and GC content), genetic and ewild pigenetic factors (such as accessible chromatin and gene expression), and energy-related factors (including RNA secondary structure, melting point, and free energy). All these considerations collectively influence the overall effectiveness of the gRNA in the CRISPR/Cas system. Understanding and optimizing these factors are crucial for enhancing the precision and efficiency of genome editing processes.

Table 2 BLAST results of gRNA candidates against *Bos taurus*.

No.	Max score	Total score	Query coverage (%)	E-value	Percent identity (%)
1	28.2	153	100	38	94.44
2	30.2	590	100	9.7	100
3	28.2	305	75	38	100
4	32.2	32.2	100	2.4	95
5	32.2	1281	100	2.4	95
6	32.2	679	85	2.4	100
7	32.2	784	90	2.4	100
8	32.2	1239	95	2.4	100
9	32.2	946	100	2.4	100
10	32.2	2344	100	2.4	100

Referring to **Table 2**, we chose candidate number 1 for further investigation due to its high number of E-values. The E-value, or expectation value, is a statistical measure employed in BLAST to gauge the likelihood of a random match between a target sequence and a competitor sequence. This parameter provides an estimate of the expected number of matches with a similar or better score that could occur

purely by chance. The calculation of the E-value takes into an account the alignment score, the size of the database, and other statistical parameters. Essentially, it helps us assess how likely a sequence similarity match is, filtering out matches that could occur randomly. In this context, opting for candidate number 1 based on a higher number of E-values indicates a greater statistical significance and raises confidence in its potential as a suitable guide RNA for the next stage of the investigation [24].

Query coverage is like a measure of how much of our original DNA sequence (the query) matches up with another sequence in the database. It's calculated by looking at the length of the part that aligns (matches) and dividing it by the total length of our original DNA. So, if a large portion of our DNA aligns with another sequence, we get a higher query coverage percentage. This helps us understand how complete the alignment is and allows us to filter out matches that only cover a small part of our DNA [25].

Percent identity is about figuring out how similar two sequences are. It represents the percentage of identical building blocks (residues) between the aligned sections of the sequences. We calculate this by dividing the number of identical building blocks by the length of the alignment and then multiplying by 100. This gives us a percentage that tells us how much the sequences resemble each other. Percent identity is often used alongside other parameters like query coverage and E-value to really understand the significance and quality of the matches we find in our sequence analysis. Essentially, it helps us evaluate how much the sequences have in common and how reliable the match is [24]. In a nutshell, query coverage tells us if the matching part is big, and percent identity tells us how similar those matched parts are. Together, they help us gauge the quality and significance of the matches we find in our sequence analysis.

Candidate number 1 underwent additional testing through BLAST, focusing on variations in the number of nucleotides. This involved testing the candidate with different lengths of nucleotides; specifically, using 20, 24, 25, and 30 nucleotides. The results of this testing are detailed in **Table 3**, providing insights into the selection process for determining the optimal number of nucleotides for candidate 1's gRNA. This step is crucial because the length of the gRNA sequence can impact its effectiveness and specificity in targeting the desired DNA region. Analyzing the outcomes in **Table 3** helps fine-tune the choice of nucleotide length, optimizing the gRNA for subsequent stages of the study.

Table 3 Selection result of nucleotide number of candidate 1 gRNA.

Nucleotide count (nt)	Max score	Total score	Query coverage (%)	E-value	Percent identity (%)
20	28.2	153	100	38	94.44
24	28.2	28.2	75	76	94.44
25	28.2	80.8	96	89	94.44
30	32.2	644	80	9.8	100

Reviewing **Table 3**, we observe that the candidate with 25 nucleotides stands out with a higher E-Value. Now, the E-Value serves as a measure of how different or homologous a sequence is compared to a competitor. In this context, a higher E-Value suggests greater dissimilarity from the competitor's sequence. Surprisingly, the second-best performer is the sequence with 24 nucleotides, demonstrating a somewhat less pronounced difference. Yet, when considering the Query Coverage values between the 25 and 24 nucleotide sequences, a distinct preference emerges for the 24 nucleotides sequence. Query Coverage measures how much of our original DNA aligns with the competitor's sequence, and a higher value indicates a more comprehensive match. In this instance, the 24 nucleotides sequence exhibits a notably greater Query Coverage value compared to the 25 nucleotides sequence.

Therefore, even though the 25 nucleotides sequence shows a more pronounced dissimilarity in terms of E-Value, the higher Query Coverage of the 24 nucleotides sequence makes it the preferred choice. This intricate analysis underscores the importance of considering multiple factors, such as E-Value and Query Coverage, to make informed decisions in optimizing the guide RNA sequence for subsequent stages of the study.

Homology difference analysis of the selected gRNA

The homology difference analysis was conducted because the cutting efficiency is highly dependent on the difference between the target sequences and the competitors. Competitor sequences include mammals, rodents, fish, and poultry. The selection of these animals is considered the main competitor to

the target. The results of the homology difference analysis against several competitors can be seen in **Table 4**.

Table 4 Selected gRNA Homology difference analysis Results.

Scientific name	Max score	Total score	Query cover (%)	E-value	Percent identity (%)
Bos taurus (Cattle)	28.2	28.2	58	62	100
Bubalus buballis (Buffalo)	28.2	28.2	75	7.1	100
Ovis aries (Sheep)	28.2	28.2	58	7.1	100
Capra hircus (Goat)	28.2	28.2	58	4.9	100
Camelus dromedarus (Camel)	26.3	26.3	54	23	100
Equus caballus (Horse)	28.2	28.2	58	8.5	100
Canis familiaris (Dog)	30.2	111	70	23	100
Cervus unicolor (Sambar Deer)	18.3	260	100	39	100
Elephas maximus (Elephant)	28.2	28.2	75	6.3	94.44
Panthera tigris (Tiger)	28.2	28.2	58	5.5	100
Panthera pardus (Leopard)	28.2	28.2	58	5.6	100
Muntiacus muntjak (Deer)	20.3	284	100	2.7	100
Cervus nippon (Sika Deer)	20.3	36.7	54	13	100
Naemorhedus goral (Goral)	18.3	239	100	3.8	100
Homo sapiens (Human)	30.2	30.2	62	44	100
Orytolagus cuniculus (Rabbit)	30.2	30.2	62	1.0	100
Tor putitora (Golden Mahseer)	22.3	51.0	51	1.1	100
Schizothoraxrichardsonii (Snow trout)	24.3	280	100	0.22	100
Raiamas bola (Trout barb)	16.4	30.7	62	1.4	100
Hypophthalmichthysmolitrix (Silver carp)	26.3	26.3	54	0.48	100
Garragotylagotyla (Sucker head)	22.3	22.3	45	0.058	100
Gallus gallus (Chicken)	30.2	103	100	9.2	100
Anas patyrhynchos (Duck)	28.2	28.2	58	22	100

Scientific name	Max score	Total score	Query cover (%)	E-value	Percent identity (%)
Meleagris gallopavo (Turkey)	30.2	323	100	12	94.74
Numida meleagris (Geunia Bird)	28.2	28.2	58	5.9	100
Coturnix japonica (Quail)	28.2	28.2	58	5.7	100
Milvus migrans (Eagle)	18.3	272	100	7.0	100
Psittacula krameri (Parakeet)	16.4	16.4	33	13	100

According to the results of the selected gRNA homology difference analysis, various E-values were obtained with values that were not close to 0, indicating the level of homology with these species. However, some animals that are likely to hybridize with the designed gRNA such as *Garragotylagotyla, Hypophthalmichthysmolitrix, and Schizothoraxrichardsonii*.

Mismatch analysis of selected gRNAs

Mismatch analysis was conducted to see what subspecies can hybridize with the selected gRNA. Subspecies tested included *verrucosus, barbatus, domesticus, celebnesis, bucculentus, phillippensis, cebifrons, affinis, andamensis, coreanus, cristasus, leucomystax, lybicus, majori, maridionalis, papuensis, ussuricus, riukiuanus, vittatus,* and *taivanus*. The selection of these animals was considered to be a subspecies of *Sus scrofa*. The results of the mismatch analysis on several subspecies of *Sus scrofa* can be seen in **Table 5**.

Table 5 Mismatch analysis results for several *Sus scrofa* subspecies.

Scientific name	Sequences (5'-3')	Mismatch	Percent (%)
Sus verrucosus (Java wild pig)	TATTCATGACCTGGCACGACAATC	4/24	83.33
	AACTCATGATCCGGCA	0/16	100
Sus barbatus (Bearded wild pig)	ATTCATGACCTGGCACGACAATC	4/24	83.33
Sus domesticus (Livestock wild pig)	ACTCATGATCCGGCACGACAATCC	1/24	95.83
Sus celebensis (Sulawesi wild pig)	TGATCTGGCGCGACAACCTAAACA	9/24	62.5
	AACTCATGATCCGGCA	0/16	100
	CATTCATGATCTGGCGCGACAATC	4/24	83.33
Sus bucculentus (Heude's wild pig) extinct	No mitochondrial database found		
Sus phillippensis (Philippine warty wild pig)	ATAGTACATATTATTATTGGTCGT	12/24	50
Sus cebifrons (Visayan warty wild pig)	ATTCATGACCTGGCACGGCAATCT	5/24	79.17
Sus scrofa affinis (Sri lankan wild boar)	ATTACAACACAATAACCTCTCAAA	14/24	41.66

Scientific name	Sequences (5'-3')	Mismatch	Percent (%)
Sus scrofa andamensis	No d-loop database		
Sus scrofa coreanus	ACTCATGATCCGGCACGACAATCC	2/24	91.67
Sus scrofa cristatus	TATTCATGATCCGGCACGACAACC	3/24	87.5
	TATTCATGATCCGGCACGACAATC	2/24	91.67
Sus scrofa leucomystax (Japanese wild boar)	ACTCATGATCCGGCACGATAATCC	2/24	91.67
Sus celebensis (Sulawesi wild pig)	TGATCTGGCGCGACAACCTAAACA	9/24	62.5
Sus scrofa lybicus. *Sus scrofa majori.* *Sus scrofa maridionalis.* *Sus scrofa papuensis* *Sus scrofa ussuricus*	No d-loop database		
Sus scrofa riukiuanus	ACTCATGATCCGGCACGACAATCC	1/24	95.83
Sus scrofa vittatus.	AGCTCGTGATCTAATGGTAGTGAT	14/24	41.67
Sus scrofa taivanus (Taiwanese wild boar)	ACTCATGATCCGGCACGACAATCC	1/24	95.83

Retrieved from the mismatch analysis results, various percentage values were obtained with values close to 100 %, indicating the compatibility of the sub-species with the selected gRNA. There are several species with presentation values above 90 %, indicating that there are several possible species such as *domesticus, coreanus, cristatus, leucomystax, riukiuanus,* and *taivanus* that can hybridize with the selected gRNA. The proposed CRISPR/Cas12a model with optimal gRNA is shown in **Figure 1**.

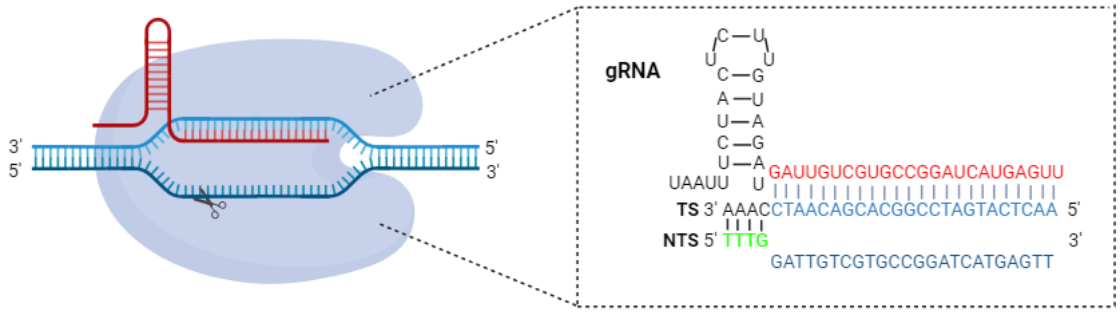

Figure 1 The CRISPR/Cas12a model of optimal gRNA. TS and NTS are dsDNA Wild pig, TS is the target strand that will be hybridized with gRNA, and NTS is a non-target strand that will have PAM (green) as a recognition for trigger nuclease activity.

Based on **Figure 1**, CRISPR/Cas12a is a CRISPR-associated protein that possesses 2 distinct nuclease activities, Cas12a uses only a single catalytic site to both cleave target dsDNA (cis-activity) and indiscriminately degrade ssDNA (trans-activity). The mechanism of action involves the binding of a guide RNA to the Cas12a protein, leading to a conformational change that activates its nuclease activity. In the first step, Cas12a makes a ssDNA cut in the target DNA, resulting in a dsDNA with a single-stranded

overhang. This initial cut is referred to as the "cis-activity" of Cas12a. In the second step, Cas12a indiscriminately degrades the ssDNA in a process known as "trans-activity" This trans-activity allows Cas12a to degrade not only the target DNA but also any other single-stranded DNA present in the system [26].

Conclusions

Several candidates have the potential as a wild pig mtDNA D-loop gRNA, but candidate 1 was selected with the sequence 5'-GAT TGT CGT GCC GGA TCA TGA GTT-3' with an off score of 99.8 and a cut position at the 643rd base on mtDNA with PAM TTTG and the number of base pairs of 24 nt.

Based on the specificity and mismatch analysis results of candidate 1, a high E value was obtained from several species of mammals, poultry, and fish tested. This indicates that Candidate 1 will not recognize these species and can detect the genus *Sus*, especially *domesticus*, *coreanus, cristatus, leucomystax, riukiuanus,* and *taivanus*. Optimal gRNA can be used to distinguish a wild pig from another animal even in the same subspecies. This optimal gRNA can be selectively and sensitively applied to biosensors for the detection of meat adulteration.

Acknowledgements

We would like to thank the Indonesian Ministry of Education and Culture, Directorate of Higher Education, Fundamental Research scheme, 3018/UN6.3.1/PT.00/2023 and Universitas Padjadjaran for supporting this work with scheme Academic Leadership Grant of Padjadjaran University No. 1549/UN6.3.1/PT.00/2023 for funding the research articles

References

[1] JA Doudna and E Charpentier. The new frontier of genome engineering with CRISPR-Cas9. *Science* 2014; **346**, 1258096.
[2] T Gaj, CA Gersbach and CF Barbas. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* 2013; **31**, 397-405.
[3] PD Hsu, ES Lander and F Zhang. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 2014; **157**, 1262-78.
[4] JS Chen, E Ma, LB Harrington, MD Costa, X Tian, JM Palefsky and JA Doudna. CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* 2018; **360**, 436-9.
[5] SN Zakiyyah, AU Ibrahim, MS Babiker, S Gaffar, M Ozsoz, MIHL Zein and YW Hartati. Detection of tropical diseases caused by mosquitoes using CRISPR-Based biosensors. *Trop. Med. Infect. Dis.* 2022; **7**, 309.
[6] M Roccaro, C Bini, P Fais, G Merialdi, S Pelotti and A Peli. Who killed my dog? Use of forensic genetics to investigate an enigmatic case. *Int. J. Legal Med.* 2021; **135**, 387-92.
[7] YW Hartati, TA Setiawati, T Sofyatin, F Fitrilawati, A Anggraeni and S Gaffar. Electrochemical DNA biosensor for detection of pork (Sus scrofa) using screen printed carbon-reduced graphene oxide electrode. *ScienceAsia* 2020; **46**, 72-9.
[8] YW Hartati, AA Suryani, M Agustina, S Gaffar and A Anggraeni. A gold Nanoparticle–DNA Bioconjugate–Based electrochemical biosensor for detection of sus scrofa mtDNA in raw and processed meat. *Food Anal. Meth.* 2019; **12**, 2591-600.
[9] N Galtier, B Nabholz, S GlÉmin and GDD Hurst. Mitochondrial DNA as a marker of molecular diversity: A reappraisal. *Mol. Ecol.* 2009; **18**, 4541-50.
[10] R Allio, S Donega, N Galtier and B Nabholz. Large variation in the ratio of mitochondrial to nuclear mutation rate across animals: Implications for genetic diversity and the use of mitochondrial DNA as a molecular marker. *Mol. Biol. Evol.* 2017; **34**, 2762-72.
[11] Q Yang, CMM Franco, SJ Sorokin and W Zhang. Development of a multilocus-based approach for sponge (phylum Porifera) identification: Refinement and limitations. *Sci. Rep.* 2017; **7**, 41422.
[12] HD Nguyen, TA Bui, PT Nguyen, OTP Kim and TTB Vo. The complete mitochondrial genome sequence of the indigenous i pig (Sus scrofa) in Vietnam. *Asian Australas. J. Anim. Sci.* 2017; **30**, 930-7.
[13] S Hermanto, T Rudiana, MIHL Zein and AW Wisudawati. Methods validation of pork authentication in processed meat products (Sausages) through densitometry analysis. *Indonesian J. Halal Res.* 2022; **4**, 35-44.
[14] YW Hartati, I Irkham, I Sumiati, S Wyantuti, S Gaffar, SN Zakiyyah, MIHL Zein and M Ozsoz. The

optimization of a label-free electrochemical DNA biosensor for detection of sus scrofa mtDNA as food adulterations. *Biosensors* 2023; **13**, 657.
[15] TA Okuma and RS Hellberg. Identification of meat species in pet foods using a real-time polymerase chain reaction (PCR) assay. *Food Contr.* 2015; **50**, 9-17.
[16] AP Karlapudi, TC Venkateswarulu, J Tammineedi, K Srirama, L Kanumuri and VP Kodali. *In silico* sgRNA tool design for CRISPR control of quorum sensing in acinetobacter species. *Gene. Dis.* 2018; **5**, 123-9.
[17] A Sgro and P Blancafort. Epigenome engineering: New technologies for precision medicine. *Nucleic Acids Res.* 2020; **48**, 12453-82.
[18] L Hernandez-Huertas, G Kushawah, A Diaz-Moscoso, L Tomas-Gallardo, I Moreno-Sanchez, GDS Pescador, AA Bazzini and MA Moreno-Mateos. Optimized CRISPR-RfxCas13d system for RNA targeting in zebrafish embryos. *STAR Protocol.* 2022; **3**, 101058.
[19] R Torres-Ruiz and S Rodriguez-Perales. CRISPR-Cas9: A revolutionary tool for cancer modelling. *Int. J. Mol. Sci.* 2015; **16**, 22151-68.
[20] P Mali, L Yang, KM Esvelt, J Aach, M Guell, JE DiCarlo, JE Norville and GM Church. RNA-guided human genome engineering via Cas9. *Science* 2013; **339**, 823-6.
[21] Y Fu, JA Foden, C Khayter, ML Maeder, D Reyon, JK Joung and JD Sander. High frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* 2013; **31**, 822-6.
[22] CY Kharismasari, Irkham, MIHL Zein, A Hardianto, SN Zakiyyah, AU Ibrahim, M Ozsoz, and YW Hartati. CRISPR/Cas12-based electrochemical biosensors for clinical diagnostic and food monitoring. *Bioelectrochemistry* 2024; **155**, 108600.
[23] JG Doench, N Fusi, M Sullender, M Hegde, EW Vaimberg, KF Donovan, I Smith, Z Tothova, C Wilen, R Orchard, HW Virgin, J Listgarten and DE Root. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9 Synthesis of an arrayed sgRNA library targeting the human genome. *Nat. Biotechnol.* 2016; **34**, 184-91.
[24] S Dash, S Rahman, H Hines and W Feng. Incremental BLAST: Incremental addition of new sequence databases through e-value correction. *PLos One* 2021, https://doi.org/10.1371/journal.pone.0249410.
[25] C Camacho, G Coulouris, V Avagyan, N Ma, J Papadopoulos, K Bealer and TL Madden. BLAST+: Architecture and applications. *BMC Bioinformatics* 2009; **10**, 421.
[26] Muflihah, A Hardianto, P Kusumaningtyas, S Prabowo and YW Hartati. DNA-based detection of pork content in food. *Heliyon* 2023; **9**, e14418.