Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Palindromic target site identification in SARS-CoV-2, MERS-CoV and SARS-CoV-1 by adopting CRISPR-Cas technique

Nimisha Ghosh\textsuperscript{a,b,1}, Indrajit Saha\textsuperscript{c,*1}, Nikhil Sharma\textsuperscript{d}

\textsuperscript{a} Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, Poland
\textsuperscript{b} Department of Computer Science and Information Technology, Institute of Technical Education and Research, Siksha ‘O’ Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India
\textsuperscript{c} Department of Computer Science and Engineering, National Institute of Technical Teachers’ Training and Research, Kolkata, West Bengal, India
\textsuperscript{d} Department of Electronics and Communication Engineering, Jaypee Institute of Information Technology, Noida, Uttar Pradesh, India

ARTICLE INFO

Edited by Xavier Carette

Keywords:
- CRISPR-Cas
- Primer
- SARS-CoV-2
- Palindrome
- Protospacer Adjacent Motif
- Target site identification

ABSTRACT

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) associated Cas protein (CRISPR-Cas) has turned out to be a very important tool for the rapid detection of viruses. This can be used for the identification of the target site in a virus by identifying a 3–6 nt length Protospacer Adjacent Motif (PAM) adjacent to the potential target site, thus motivating us to adopt CRISPR-Cas technique to identify SARS-CoV-2 as well as other members of Coronaviridae family. In this regard, we have developed a fast and effective method using k-mer technique in order to identify the PAM by scanning the whole genome of the respective virus. Subsequently, palindromic sequences adjacent to the PAM locations are identified as the potential target sites. Palindromes are considered in this work as they are known to identify viruses. Once all the palindrome-PAM combinations are identified, PAMs specific for the RNA-guided DNA Cas9/Cas12 endonuclease are identified to bind and cut the target sites. In this regard, PAMs such as 5’-TGG-3’ and 5’-TTTA-3’ in NSP3 and Exon for SARS-CoV-2, 5’-GGG-3’ and 5’-TTTG-3’ in Helicase and NSP3 respectively for SARS-CoV-1 are identified corresponding to SpCas9 and FnCas12a endonucleases. Finally, to recognise the target sites of Coronaviridae family as cleaved by SpCas9 and FnCas12a, complements of the palindromic target regions are designed as primers or guide RNA (gRNA). Therefore, such complementary gRNAs along with respective Cas proteins can be considered in assays for the identification of SARS-CoV-2, MERS-CoV and SARS-CoV-1.

1. Introduction

COVID-19, the disease caused by Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) has affected a lot of people around the globe and has claimed more than 5.3 million lives as of 15th December 2021.\textsuperscript{2} SARS-CoV-2 belongs to the family of Coronaviridae which also accommodates MERS-CoV and SARS-CoV-1 viruses (Zhou et al., 2020). The symptoms of COVID-19 include cough, fever, dyspnoea, diarrhoea, myalgia (Hosseini et al., 2020) and in some extreme cases may also lead to severe respiratory distress leading to eventual death. Moreover, comorbidity issue in COVID-19 is relatively high and targets different organs like kidney, liver, heart, brain, etc. (Dey et al., 2020; Qi et al., 2020).

Since its spread, symptom-based diagnosis of COVID-19 is being performed which includes chest X-ray and CT scan, quantitative reverse transcription polymerase chain reaction (qPCR) and antibody test. Most recently, another rapid detection method based on CRISPR-Cas has been proposed by the researchers (Broughton et al., 2020b). Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) associated Cas protein (CRISPR-Cas) is an adaptive immune system in prokaryotic organism that can provide resistance to foreign elements. This system has been exploited in recent times as a powerful gene editing tool and for the diagnosis and inactivation of viruses (Jia et al., 2020). The efficiency of CRISPR-Cas technique is dependent on the design of guide RNA

Abbreviations: CRISPR, Clustered Regularly Interspaced Short Palindromic Repeats; PAM, Protospacer Adjacent Motif; gRNA, guide RNA.

* Corresponding author.
E-mail address: indrajit@nitttrkol.ac.in (I. Saha).
\textsuperscript{1} Equally contributed.
\textsuperscript{2} https://www.worldometers.info/coronavirus/.

https://doi.org/10.1016/j.gene.2021.146136

Received 1 October 2021; Received in revised form 16 December 2021; Accepted 20 December 2021
Available online 6 January 2022
0378-1119/© 2022 Elsevier B.V. All rights reserved.
For the identification of SARS-CoV-2, MERS-CoV and SARS-CoV-1 viruses, initially the PAM and their corresponding genomic locations are considered for the RNA-guided DNA Cas9/Cas12 endonuclease. In this regard, PAMs such as 5′-TG3′ and 5′-TTTA-3′ in NSP3 and Enox for SARS-CoV-2, 5′-GGG-3′ and 5′-TGG-3′ in Enox and NSP2 for MERS-CoV and 5′-AGG-3′ and 5′-TTT-3′ in Helicase and NSP3 respectively for SARS-CoV-1 are identified corresponding to SpCas9 and Fncas12a. It is worth mentioning that studies performed by Cain et al. (2001), Dirac et al. (2002), Chew et al. (2004) have suggested that palindromic sequences can be considered to be involved in target identification, viral packaging and defence mechanisms. A palindromic sequence is a symmetrical sequence so that when read from the reverse direction, it is the exact complement of itself. For example, TGCA is a palindrome of length 4. It is to be noted that a palindrome is always even in length. Thereafter, to recognise these target sites in a virus genome as cleaved by SpCas9 and Fncas12a, primers are designed as complementary to the target site sequences. Thus, these complementary palindromic primers can be considered in assays for the rapid identification of SARS-CoV-2, MERS-CoV and SARS-CoV-1. These primers are akin to guide RNA (gRNA) in CRISPR-Cas technology.

2. Materials and methods

To find PAM and the corresponding palindromic sequences, initially the three reference genomic sequences of SARS-CoV-2 (NC_045512.2),3 MERS-CoV (NC_019843.3)4 and SARS-CoV-1 (NC_004718.3)5 are collected from NCBI (National Center for Biotechnology Information). Once the palindromic and PAM combinations for each virus are identified on their respective reference sequence, their presence or coverage is verified for 108246, 291 and 340 virus sequences of SARS-CoV-2, MERS-CoV and SARS-CoV-1 respectively. SARS-CoV-2 virus sequences are collected from Global Initiative on Sharing All Influenza Data (GISAID)6 in fasta format while 291 and 340 virus sequences of MERS-CoV7 and SARS-CoV-18 are downloaded from NCBI. It is to be noted that 108246 SARS-CoV-2 sequences are considered up to September 2021 after performing filtering for considering only complete SARS-CoV-2 genomes. For docking purposes, P3DOCK Server9 is used and PuYMol10 is considered for visualization. For such docking, the PDB ID of PAMs like Fncas12a is 5NG6 and that of SpCas9 is 4OO8 while the corresponding palindromes of SARS-CoV are designed with the help of Chimera v.1.15 RNA/DNA build tool.

Algorithm 1 presents the Palindrome-PAM finding method in details. To identify PAMs (P), the reference sequence r of each of the viruses is initially divided into patterns of sequences of length k using the popular k-mer technique. Thereafter, these k-mers are searched for in the reference sequence to learn about their corresponding genomic locations in the FindPAMpositions step. These sequences of length k are considered as PAMs in this work. Subsequently, all these k-mer PAMs are used to identify the palindromic sequences in the reference sequence. Based on the intended palindromic sequence length η, some calculations are performed to find the starting and ending genomic locations of the target site of a virus. Since, the palindromic sequence and PAM are adjacent, the ending (end_loc_pal) and the starting locations of the palindrome

\[ \text{end_loc_pal} = \text{start_loc_pal} + \eta \]

\[ \text{start_loc_pal} = \text{start_loc_PAM} - \eta \]

3 https://www.ncbi.nlm.nih.gov/nuccore/1798174254.
4 https://www.ncbi.nlm.nih.gov/nuccore/667498388.
5 https://www.ncbi.nlm.nih.gov/nuccore/302771926.
6 gisaid.org/CoV2020.
7 https://www.ncbi.nlm.nih.gov/gene?term=Virus%3AVoL%3A%20Severe%20%5BRespiratory%20%5D%20%5BCoronavirus%20%5D
8 https://www.ncbi.nlm.nih.gov/labs/virus/vski/virus?SeqType=Nucleotide&VirusLineageSeq=Severe%20%3B%20respiratory%20%3B%20Coronavirus
9 http://www.mabinding.com/P3DOCK/P3DOCK.html.
10 https://pymol.org/2/.
(strt_loc_pal) are calculated as end_loc_pal = strt_loc_PAM – 1 and strt_loc_pal = end_loc_pal – (η – 1) respectively. If these starting and ending locations in the reference sequence of a virus genome return a palindromic sequence in the PalindromeCheck step, then they are chosen as the target sites (ℳ) for the identification of the virus. Subsequently, all the virus genomic sequences are checked for the presence of the set of identified palindrome and PAM combinations (S_{PalPAM}) to check for the presence of these combinations (Population Coverage) in the Coverage step. This method for identification of the target site for the subsequent identification of virus is applied for SARS-CoV-2, MERS-CoV and SARS-CoV-1 viruses. The CRISPR-Cas9 technology for gene editing is shown in Fig. 1 (a) while the proposed method for virus identification are shown in Fig. 1 (b).

Algorithm 1. Pseudo-code for palindromic target site identification

\begin{verbatim}
Input : ρ (reference sequence of virus genome), k (k-mer), η (length of palindrome), χ (all virus genomic sequences)

Output: S_{PalPAM} (set of Palindrome and PAM combination), \theta (Population Coverage)

begin
      Initialize a NULL Set, S_{PalPAM}
      \mathcal{P} \leftarrow \kmercount(\rho, k)
      \lambda \leftarrow \text{FindPAMpositions(}\mathcal{P})
      \alpha \leftarrow \text{numrows}()\lambda)
      \beta \leftarrow \text{numcols}()\lambda)

      for i = 1 to \alpha do
         for j = 1 to \beta do
            end_loc_pal = strt_loc_PAM – 1
            strt_loc_pal = end_loc_pal – (η – 1)
            \mathcal{M} \leftarrow \text{PalindromeCheck(strt_loc_pal, end_loc_pal)}
            \text{PalPAM} \leftarrow \mathcal{M}\mathcal{P}
            S_{PalPAM} \leftarrow S_{PalPAM} \cup \text{PalPAM}

      end

      \theta \leftarrow \text{Coverage(\chi, S_{PalPAM})}
      return S_{PalPAM}, \theta

end
\end{verbatim}

Once all the palindrome-PAM combinations are identified, PAMs like 5′-NGG-3′ and 5′-TTTN-3′ (N is any nucleotide base) specific for the RNA-guided DNA SpCas9 and FnCas12a endonucleases are identified out of all the combinations to bind and cut the target sites. Thereafter, SpCas9 and FnCas12a cleave the target regions for the final virus identification. Similarly, target sites are identified by the gRNAs which are the complementary primers of the palindromic sequences. For example, if SARS-CoV-2 is present in a sample collected from a person, the corresponding gRNA will attach itself to the fragment of the virus as cleaved by Cas9/Cas12 endonuclease. Such fragments can be then amplified thereby identifying the corresponding virus. Similar experiments can be performed for MERS-CoV and SARS-CoV-1 as well.

The time complexity of the proposed method for finding the palindromic-PAM combinations can be calculated as: let the length of the reference sequence be \mathcal{N}. Thus, the time complexity of finding PAM using \kmercount is \mathcal{O}(\mathcal{N}). To find the corresponding genomic locations of these PAM in \text{FindPAMpositions} step, the time taken is \mathcal{O}(\log\mathcal{N}). Computing the palindromic sequences takes \mathcal{O}(\alpha\beta) time, where \alpha and \beta are the number of rows and columns respectively of the matrix returned by \text{FindPAMpositions} step. Finally, the computation of population coverage takes \mathcal{O}(\chi\mathcal{D}) time, where \chi and \mathcal{D} are the dimensions of the set which holds the palindrome-PAM combinations and \mathcal{D} is the number of virus genomic sequences. Thus, the overall complexity of the proposed target site identification method is \mathcal{O}(\mathcal{N} + \log\mathcal{N} + \alpha\beta + \chi\mathcal{D}).

3. Results

The results for the total number of PAM in the reference genomic sequence of each of SARS-CoV-2, MERS-CoV and SARS-CoV-1 are shown in Fig. 1(c), where the value of \k is varied from 3 to 6. From the figure, we can see that with increasing \k, the number of PAM is also increasing. For example, for SARS-CoV-2, with \k = 3, 4, 5 and 6, number of PAM are 64, 256, 1023 and 3756 respectively. With these PAM, the corresponding palindromic sequences are identified in the reference genomic sequence following Algorithm 1. The results for the number of identified palindrome-PAM combinations in the reference sequences of SARS-CoV-2, MERS-CoV and SARS-CoV-1 viruses are depicted in Fig. 1 respectively for varying lengths of \k and where the palindrome of even length is taken from 10 to 20. For example, 38 such combinations exist in the reference sequence of SARS-CoV-2 for PAM lengths of 3 to 6 and palindrome length of 10. For MERS-CoV, the identified palindrome-PAM combination sequences exist for lengths of 10 to 18 but not for 20 while for SARS-CoV-1 such combinations exist for all the palindromic lengths as considered in this work. It is to be noted that beyond the length of 20, no palindrome-PAM combinations exist for any of the virus sequences. Table 1 reports the percentage of coverage of the identified palindrome-PAM combinations in 108246, 291 and 340 SARS-CoV-2, MERS-CoV
and SARS-CoV-1 virus sequences respectively for palindromic lengths of 14 to 20. For all the palindromic sequences, the results are provided in the supplementary as an excel file. For example, for SARS-CoV-2 with $k = 6$ and target palindromic sequence length of 20, palindrome-PAM combination of ACACTGGTAATTACCAGTGT and GGTCAC is present in 108048 out of 108246 virus sequences, i.e. this combination is present in 99.81% of the sequences. This coverage percentage shows the validity of the identified palindrome-PAM combinations for the respective virus genomic sequences. The results are similarly reported for MERS-CoV and SARS-CoV-1. Table 1 also reports the corresponding GC content of the identified target sites. According to (Reynolds et al., 2004; Haeussler et al., 2016), it is difficult to target GC-rich genes and thus it can be said that sequences with moderate GC-content are good candidates for being target sites of a virus. As can be observed from Table 1, the GC content of the identified target sites for each of SARS-CoV-2, MERS-CoV and SARS-CoV-1 are quite moderate. For example, for $k = 6$ and target palindromic sequence length of 20, the target palindromic site of SARS-CoV-2 has a GC content of 40%.

Once all the palindrome-PAM combinations are identified for all lengths of palindrome and PAM, specific PAMs which are recognised by either SpCas9 or FnCas12a endonuclease for cleaving the target sites are further identified. These palindrome-PAM combinations are specifically important for the identification of the viruses. Out of the total 248 unique palindrome-PAM combinations for SARS-CoV-2, 2 such combinations are identified corresponding to SpCas9 and FnCas12a, for MERS-CoV, out of 248 combinations, 3 such combinations are identified while for SARS-CoV-1 2 such combinations out of 224 are identified. Table 2 reports such PAMs along with the corresponding palindromes and their

![Fig. 1.](image-url)
associated details along with the gRNA, coding gene, percentage of coverage and GC content. For example, for identifying SARS-CoV-2, TTTA is a PAM in Exon needed to guide FnCas12a endonuclease for binding to and cleaving TTGGTACCAA with the help of gRNA AACCATGGTT as a complementary primer. Such palindrome-PAM is present in 99.71% of the total number of sequences while the GC content of the palindrome is 40. On the other hand, TGG is a PAM in NSP3 for SpCas9 for cleaving CACTGGTAATTACCAGTG with GTGACCATTAATGGTCAC and this combination is present in 99.82% of the sequences with the GC content of the palindrome being 44.44%. Similarly, PAMs corresponding to SpCas9 and FnCas12a are also identified for MERS-CoV and SARS-CoV-1 and reported in Table 2. The corresponding target sequences as DNA in SpCas9 and FnCas12a along with the Cas Binding with PAM in order to cleave the target region are shown in Figs. 2–4. In each figure, (I) shows the palindrome-PAM combinations with Cas Protein and (II) shows the binding of PAM with the Cas protein for cleaving the target site. For example, Fig. 2(a)(I) shows the palindromic sequence TTGGTACCAA in purple and PAM TTTA in red and (II) shows the binding of TTTA with FnCas12a. This binding is a proof that indeed FnCas12a can recognise TTTA and cleave the target region TTGGTACCAA.

Please note that all the palindrome-PAM combinations are unique to each virus, thereby confirming the fact that they can indeed be used for virus identification. Furthermore, we have also checked for the combinations from Table 2 in the reference sequence of Ebola, Dengue, Influenza and Zika viruses. Also, nucleotide BLAST\textsuperscript{11} is used to check the specificity of the same and it has been observed that such palindrome-PAM combinations are not present in any of the other viruses. Apart from aforementioned results, all the palindrome-PAM combinations are provided in the supplementary as an excel file. It is to be further noted that though this work specifically focuses on PAMs as recognised by Cas9 or Cas12 endonuclease, we have reported other palindrome-PAM combinations as well in the hope that if any new endonuclease is engineered, our work can serve as a way for further virus identification.

Table 1
Population Coverage and GC Content of the target sites for SARS-CoV-2, MERS-CoV and SARS-CoV-1.

| Virus     | Length of Palindrome | Population Coverage (%) | GC Content (%) | Population Coverage (%) | GC Content (%) | Population Coverage (%) | GC Content (%) | Population Coverage (%) | GC Content (%) |
|-----------|---------------------|-------------------------|---------------|-------------------------|---------------|-------------------------|---------------|-------------------------|---------------|
| SARS-CoV-2| 14                  | 99.89                   | 42.85         | 99.88                   | 42.85         | 99.79                   | 42.85         | 90.81                   | 42.85         |
|           | 70.37               | 70.37                   | 14.28         | 91.63                   | 0             | 90.34                   | 14.28         | 70.33                   | 14.28         |
|           | 16                  | 99.85                   | 37.5          | 99.79                   | 37.5          | 90.34                   | 37.5          | 70.32                   | 37.5          |
|           | 18                  | 99.82                   | 44.44         | 99.82                   | 44.44         | 90.34                   | 44.44         | 90.32                   | 44.44         |
| MERS-CoV  | 14                  | 99.65                   | 57.14         | 99.65                   | 57.14         | 90.78                   | 57.14         | 90.78                   | 57.14         |
|           | 16                  | 98.62                   | 14.28         | 98.62                   | 14.28         | 89.78                   | 14.28         | 80.32                   | 14.28         |
|           | 18                  | 97.62                   | 33.33         | 97.62                   | 33.33         | 88.78                   | 33.33         | 83.82                   | 33.33         |
| SARS-CoV-1| 14                  | 91.17                   | 28.57         | 90.58                   | 28.57         | 90.05                   | 28.57         | 90.05                   | 28.57         |
|           | 16                  | 87.05                   | 22.22         | 87.05                   | 22.22         | 87.05                   | 22.22         | 87.05                   | 22.22         |
|           | 20                  | 87.05                   | 30            | 87.05                   | 30            | 87.05                   | 30            | 87.05                   | 30            |

Table 2
Combination of Palindrome and PAM in SARS-CoV-2, MERS-CoV and SARS-CoV-1 viruses along with the complementary primer.

| Virus     | Palindrome as Target | Length of Palindrome | Start Coordinate of Target | PAM | Start Coordinate of PAM | Complementary Primer (gRNA) | Coding Gene | Population Coverage (%) | GC Content (%) |
|-----------|----------------------|----------------------|-----------------------------|-----|------------------------|-----------------------------|-------------|-------------------------|---------------|
| SARS-CoV-2| TTGGTACCAA           | 10                   | 18341                       | TTTA| 18351                  | AACCATGGTT                  | Exon        | 99.71                   | 40            |
|           | CACTGGTAATTACAGTG     | 18                   | 5746                        | TGG | 5764                   | GTGACCATTAGGTCAC            | NSP3        | 99.82                   | 44.44         |
| MERS-CoV  | TTATGCCATAA          | 10                   | 18456                       | GGG | 18466                  | AATAGTATT                   | Exon        | 100                     | 20            |
|           | ATCTATAGATAG         | 12                   | 1018                        | TGG | 1030                   | ATCTATAGATAG               | NSP2        | 98.62                   | 16.66         |
|           | CTTATGCATAG         | 12                   | 18455                       | GGG | 18467                  | GAATAGTATTGTC              | Exon        | 99.31                   | 33.33         |
| SARS-CoV-1| AACTAGTGCTGT        | 10                   | 16450                       | AGG | 16460                  | TTGGTACCAA                 | Helicase     | 90.81                   | 16.67         |
|           | TAAACTTGGTAA         | 12                   | 5208                        | TTG | 5220                   | ATTTACCTAAT               | NSP3        | 93.64                   | 16.67         |

\textsuperscript{11}https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn\&PAGE_TYPE=BlastSearch\&LINK_LOC=blasthome.
4. Conclusion

This study adopts the idea of CRISPR-Cas technology in order to identify palindrome-PAM combinations as target sites for virus identification. To achieve this, initially PAMs are identified using k-mer technique. Thereafter, palindromic sequences which are adjacent to the PAM locations are identified as the potential target sites. Next, PAMs specific for the RNA-guided DNA Cas9/Cas12 endonuclease to bind and cut the target sites are detected. In this regard, corresponding to SpCas9 and FnCas12a endonuclease, PAMs such as 5′-TGG-3′ and 5′-TTTA-3′ in NSP3 and Exon for SARS-CoV-2, 5′-GGG-3′ and 5′-TGG-3′ in Exon and NSP2 for MERS-CoV and 5′-AGG-3′ and 5′-TTTG-3′ in Helicase and NSP3 respectively for SARS-CoV-1 are identified. Finally, gRNAs as primers complementary to the identified target sites are designed to recognise these target sites in a virus genome. The palindrome-PAM combinations are initially identified in the corresponding reference sequences of SARS-CoV-2, MERS-CoV and SARS-CoV-1. By varying the length of k in k-mer technique and by considering different length of palindromes, the highest number of palindrome-PAM combinations for each of SARS-CoV-2, MERS-CoV and SARS-CoV-1 are found to be 38, 43 and 36 respectively for k = 3 to 6 and palindrome length = 10. The palindrome-PAM combinations suitable for CAS binding in SARS-CoV-2 are TTGGTACCAATTTA in Exon and CACTGGTAATTACAGTGTGG in NSP3. For MERS-CoV such combinations are TTATGCTAAGGG in Exon, ATCTATAGATTGG in NSP2 and CTATGCTAAGGGG in Exon while for SARS-CoV-1 such combinations are ACACATGTGTAGG in Helicase and TAAAAATGTTAGGG in NSP3. These palindrome-PAM combinations when searched for in 108246, 291 and 340 SARS-CoV-2, MERS-CoV and SARS-CoV-1 virus sequences respectively, they are present in 99.71%, 99.82%, 100%, 98.62%, 99.31%, 94.41% and 92.64% of the virus sequences. Furthermore, the GC content of these identified target sites are evaluated to judge their candidature. The method proposed in this work can be deemed to be a very efficient and quick way to detect SARS-CoV-2, MERS-CoV and SARS-CoV-1 where the gRNAs as primers complementary to the target sites along with the respective Cas proteins can be considered in assays for the identification method.

Fig. 2. (I) Palindrome-PAM combinations with Cas Proteins and (II) Cas binding with PAM for cleaving the target site in SARS-CoV-2 for (a) FnCas12a (b) SpCas9. In the figures, purple represents the palindromic sequence and red represents the PAM sequence.
Fig. 3. (I) Palindrome-PAM combinations with Cas Proteins and (II) Cas binding with PAM for cleaving the target site in MERS-CoV for (a) (b) and (c) SpCas9. In the figures, purple represents the palindromic sequence and red represents the PAM sequence.
Furthermore, apart from the Cas recognised PAMs we have provided a list of other palindrome-PAM combinations as well for all the three viruses in the hope that if in the future some other endonucleases are engineered, such combinations may be further used for successful virus identification. Moreover, the in vitro verification of the identified palindrome-PAM combinations may also be further investigated.

Ethics approval and consent to participate

The ethical approval or individual consent was not applicable.

Availability of data and materials

All the SARS-CoV-2, MERS-CoV and SARS-CoV-1 virus genomes with their corresponding reference sequences and the final results of this work are available at “http://www.nitttrkol.ac.in/indrajit/projects/COVID-CRISPR-Cas/”.

Funding

This work has been partially supported by CRG short term research grant on COVID-19 (CVD/2020/000991) from Science and Engineering Research Board (SERB), Department of Science and Technology, Govt. of India.

Fig. 4. (I) Palindrome-PAM combinations with Cas Proteins and (II) Cas binding with PAM for cleaving the target site in SARS-CoV-1 for (a) SpCas9 and (b) FnCas12a. In the figures, purple represents the palindromic sequence and red represents the PAM sequence.

Author contributions

Nimisha Ghosh: Conceptualization; Methodology; Visualization; Writing – original draft, Indrajit Saha: Conceptualization; Data curation; Supervision; Funding acquisition; Formal analysis; Investigation; Methodology; Project administration; Resources; Validation; Visualization; Writing – review & editing, Nikhil Sharma: Methodology; Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was partially carried out during the tenure of an ERCIM ‘Alain Bensoussan’ Fellowship Program awarded to Nimisha Ghosh. Also, the authors would like to thank all those who have contributed sequences to GISAID and NCBI databases.

References

Bhat, M.A., Bhat, M.A., Kumar, V., Wani, I.A., Bashir, H., Shah, A.A., Rahman, S., Jan, A. T., 2020. The era of editing plant genomes using crisps/cas: A critical appraisal. J. Biotechnol. 324, 34–60. https://doi.org/10.1016/j.jbiotec.2020.09.013.
Chew, D., Choi, K.P., Heidner, H., Leung, M.-Y., 2004. Palindromes in sars and other
Freije, C.A., Myhrvold, C., Boehm, C.K., Lin, A.E., Welch, N.L., Carter, A., Metsky, H.C.,
Dey, A., Sen, S., Maulik, U., 2020. Unveiling COVID-19-associated organ-specific cell
Chen, S.J., Ma, E., Harrington, B.L., Costa, M.D., Tian, X., Palefsky, J.M., Doudna, A.J.,
Haeussler, M., Schnig, K., Eckert, H., Eschstruth, A., Miann, J., Renaud, J.-B., Schneider-
Gootenberg, S.J., Abudayyeh, O.O., Lee, J.W., Essletzbichler, P., Dy, A.J., Joung, J.,
Broughton, J.P., Deng, X., Yu, G., Fasching, C.L., Servellita, V., Singh, J., Miao, X.,
Streithorst, J., Granados, A., Sotomayer-Gonzalez, A., Zorn, K., Gopez, A., Hsu, E.,
Gu, W., Miller, S., Pan, C., Guevara, H., Wadford, D.A., Chen, J.S., Chiu, C.Y., 2020a. Crispr cas12-based detection of sars-cov-2. Nat. Biotechnol. 38, 870–874. https://doi.org/10.1101/2020.03.06.20025334.
Broughton, J.P., Deng, X., Yu, G., Fasching, C.L., Singh, J., Streithorst, J., Granados, A.,
Sotomayer-Gonzalez, A., Zorn, K., Gopez, A., Hsu, E., Gu, W., Miller, S., Pan, C.,
Guevara, H., Wadford, D.A., Chen, J.S., Chiu, C.Y., 2020b. Rapid detection of 2019
novel coronavirus sars-cov-2 using a crispr-based detectr lateral flow assay.
medRxiv: the preprint server for health sciences, doi: https://doi.org/10.1101/2020.03.06.20025334.
Cain, D., Erlewine, O., Grigg, A., Russell, R., Mcclure, M., 2001. Palindromic sequence
plays a critical role in human foamy virus dimerization. J. Virol. 75, 3731–3739.
https://doi.org/10.1128/JVI.75.10.3731-3739.2001.
Chen, S.J., Ma, E., Harrington, B.L., Costa, M.D., Tian, X., Palefsky, J.M., Doudna, A.,
2018. Crispr-cas12a target binding unleashes indiscriminate single-stranded dna
activity. Science 360, 436–439. https://doi.org/10.1126/science.aan6245.
Chew, D., Choi, K.P., Heidner, H., Leung, M.Y., 2004. Palindromes in sars and other
coronaviruses. INFORMS J. Comput. 16, 331–340. https://doi.org/10.1287/ijoc.1040.0087.
Dey, A., Sen, S., Maulik, U., 2020. Unveiling COVID-19-associated organ-specific cell
types and cell-specific pathway cascade. Briefings Bioinformat. https://doi.org/10.1093/bib/bbaa214.
Dirac, A.M.G., Huthoff, H., Kjems, J., Berkshire, B., 2002. Requirements for rna
heterodimerization of the human immunodeficiency virus type 1 (hiv-1) and hiv-2
genomes. J. Gen. Virol. 83, 3731–3739. https://doi.org/10.1099/0022-1317-83-10-2533.
Freije, C.A., Myhrvold, C., Boehm, C.K., Lin, A.E., Welch, N.L., Carter, A., Metsky, H.C.,
Luo, C.Y., Abudayyeh, O.O., Gootenberg, J.S., Youzzaik, N.L., Zhang, F., Saberi, P.C.,
2019. Programmable inhibition and detection of rna viruses using cas13. Mol. Cell 76, 826–837. https://doi.org/10.1016/j.molcel.2019.09.013 e11.
Gootenberg, S.J., Abudayyeh, O.O., Lee, J.W., Esletzbichler, P., Dy, A.J., Joung, J.,
Verdine, V., Donghia, N., Daringer, N.M., Freije, C.A., Myhrvold, C.,
Bhattacharyya, R.P., Livny, J., Regev, A., Koonin, E.V., Hung, D.T., Saberi, P.C.,
Collins, J.J., Zhang, F., 2017. Nucelic acid detection with crispr-cas13a/c2c2.
Science 356, 438–442. https://doi.org/10.1126/science.aam9321.
Haeussler, M., Schnig, K., Eckert, H., Echsttruch, A., Mzaoua, J., Renaud, J.-B., Schneider-
Maunoury, S., Shikumata, A., Teboul, J., Kent, J., Joly, J.-S., Concordet, J.-P.,
2016. Evaluation of off-target and on-target scoring algorithms and integration into
the guide rna selection tool crispor. Nat. Biotechnol. 17 https://doi.org/10.1186/s13059-016-1012-2.
Hosseini, E.S., Kashani, N.R., Nikzad, H., Azadbakht, J., Bafrani, H.H., Kashani, H.H.,
2020. The novel coronavirus disease-2019 (covid-19): Mechanism of action,
detection and recent therapeutic strategies. Virology 551, 1–9. https://doi.org/10.1016/j.virology.2020.08.011.
Jia, F., Li, X., Zhang, C., Tang, X., 2020. The expanded development and application of
crisp-cas12a system for sensitive nucleotide detection. Protein Cell 11, 624–629. https://
doi.org/10.1007/s13238-020-0709-8.
Kayesh, M.E.H., Amako, Y., Hashem, M.A., Murakami, S., Ogawa, S., Yamamoto, N.,
Hifumi, T., Miyoshi, N., Sugiyama, M., Tanaka, Y., Mizokami, M., Kohara, M.,
Tsukiyama-Kohara, K., 2020. Development of an in vivo delivery system for crispr/
cas9-mediated targeting of hepatitis b virus cccdna. Virus Res. 290, 198191. https://doi.org/10.1016/j.virome.2019.10.021.
King, M.W., Mungen, J., 2019. Editing the human cytomegalovirus genome with the
crisp-cas9 system. Virology 529, 186–194. https://doi.org/10.1016/j.virology.2019.01.021.
Li, L., Li, S., Wu, N., Wu, J., Wang, G., Zhao, G., Wang, J., 2019. Holmesv2: A crispr-
cas12b-assisted platform for nucleic acid detection and dna methylation
quantitation. ACS Synthetic Biol. 8, 2228–2237. https://doi.org/10.1021/acssynbio.9b00209.
Lyu, C., Shi, H., Cui, Y., Li, M., Yan, Z., Yan, L., Jiang, Y., 2020. Crispr-based biosensing is
prospective for rapid and sensitive diagnosis of pediatric tuberculosis. Int. J.
Infectious Dis. 101, 183–187. https://doi.org/10.1016/j.ijid.2020.09.1428.
Qi, F., Qian, S., Zhang, S., Zhang, Z., 2020. Single cell rna sequencing of 13 human tissues
identify cell types and receptors of human coronaviruses. Biochim. Biophys. Res.
Commun. 526, 135–140. https://doi.org/10.1016/j.bbrc.2020.03.044.
Rahman, M.R., Hossain, M.A., Mozibullah, M., et al., 2021. CRISPR is a useful biological
tool for detecting nucleic acid of SARS-CoV-2 in human clinical samples. Biomed.
Pharmacotherapy 140, 111772. https://doi.org/10.1016/j.biopsych.2021.111772.
Reynolds, A., Leake, D., Boese, Q., Scarring, S., Marshall, W.S., Khvorova, A., 2004.
Rational sirna design for rna interference. Nat. Biotechnol. 22, 326–330. https://doi.org/10.1038/nbt938.
Zhang, F., Abudayyeh, O.O., Gootenberg, J.S., 2020. A protocol for detection of covid-19
using crispr diagnostics. In: A protocol for detection of COVID-19 using CRISPR
diagnostics, p. 8.
Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B.,
Huang, C., Chen, H., Chen, J., Luo, Y., Guo, H., Jiang, R., Liu, M., Chen, Y., Shen, X.,
Wang, X., Zheng, X., Zhao, K., Chen, Q., Deng, F., Liu, L., Yan, B., Zhan, F., Wang, Y.,
Xiao, G., Shi, Z., 2020. A pneumonia outbreak associated with a new coronavirus of
probable bat origin. Nature 579, 270–273. https://doi.org/10.1038/s41586-020-1202-7.