©Biomedical Informatics (2022)

OPEN ACCESS GOLD



Declaration on Publication Ethics:

The author's state that they adhere with COPE guidelines on publishing ethics as described elsewhere at https://publicationethics.org/. The authors also undertake that they are not associated with any other third party (governmental or non-governmental agencies) linking with any form of unethical issues connecting to this publication. The authors also declare that they are not withholding any information that is misleading to the publisher in regard to this article.

Declaration on official E-mail:

The corresponding author declares that lifetime official e-mail from their institution is not available for all authors

License statement:

This is an Open Access article which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. This is distributed under the terms of the Creative Commons Attribution License

Comments from readers:

Articles published in BIOINFORMATION are open for relevant post publication comments and criticisms, which will be published immediately linking to the original article without open access charges. Comments should be concise, coherent and critical in less than 1000 words.

Edited by P Kangueane Citation: Prajapati *et al.* Bioinformation 18(8): 661-668 (2022)

Towards the diagnosis of dengue virus and its serotypes using designed CRISPR/Cas13 gRNAs

Archana Prajapati*, AshmitaTandon* & Vikrant Nain#

School of Biotechnology, Gautam Buddha University, Greater Noida-201312, Uttar Pradesh, India; #Corresponding author,*Equal contribution

Institution URL: https://www.gbu.ac.in/

Author contacts:

Archana Prajapati - E-mail: btphd2018003@gbu.ac.in Vikrant Nain - E-mail: vikrant@gbu.ac.in

Abstract:

Dengue Virus (DENV) is a mosquito-borne virus that is prevalent in the world's tropical and subtropical regions. Therefore, early detection and surveillance can help in the management of this disease. Current diagnostic methods rely primarily on ELISA, PCR, and RT-PCR, among others, which can only be performed in specialized laboratories and require sophisticated instruments and technical expertise.

DOI: 10.6026/97320630018661

CRISPR-based technologies on the other hand have field-deployable viral diagnostics capabilities that could be used in the development of point-of-care molecular diagnostics. The first step in the field of CRISPR-based virus diagnosis is to design and screen gRNAs for high efficiency and specificity. In the present study, we employed a bioinformatics approach to design and screen DENV CRISPR/Cas13 gRNAs for conserved and serotype-specific variable genomic regions in the DENV genome. We identified one gRNA sequence specific for each of the lncRNA and NS5 regions and identified one gRNA against each of DENV1, DENV2, DENV3, and DENV4 to distinguish the four DENV serotypes. These CRISPR/Cas13 gRNA sequences will be useful in diagnosing the dengue virus and its serotypes for *in vitro* validation and diagnostics.

Keywords: Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR), Cas13, Dengue virus, serotypes, diagnosis, gRNA pool, secondary structure, free energy.

Background:

Dengue is a major arthropod-borne viral disease caused by the dengue virus (DENV), which is widespread in tropical and subtropical regions of the world and is spread by the Aedesaegyptis and Aedesalbopictus mosquitos [1]. Its clinical manifestations range from mild Dengue Fever (DF) to severe Dengue Haemorrhagic Fever (DHF) and Dengue shock syndrome (DSS) [1]. Dengue virus (DENV) is an enveloped, positive-sense, ss-RNA virus of the Flaviviridae family and the genus Flavivirus [2], [3]. DENV has four well-known serotypes (DENV1, DENV2, DENV3, DENV4) based on antigen cross-reactivity, with each serotype having distinct genotypes [2][3]. The DENV genome (~11 kb) comprises one open reading frame (ORF) flanked by 5' and 3' UTR regions [2][4]. Its ORF encodes a single polyprotein, which is cleaved posttransactionally to generate three structural proteins: capsid (C), premembrane/membrane (prM/M), and envelope (E), as well as seven non-structural (NS) proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5)(Figure 1) [1]. There are currently no particular antiviral medicines or vaccinations available to treat or prevent dengue infection[2],[5],[6]. The sole alternative that relies on an early and accurate dengue diagnosis is symptomatic and supportive therapy of the disease. Dengue infection is diagnosed through the isolation and characterization of the DENV virus, serological tests that detect DENV-related antigens and/or antibodies in patient plasma/serum, and molecular diagnosis [4], [5], [7]. The DENV nucleic acid (RNA) is detected using an isothermal nucleic acid sequence-based amplification assay (NASBA), a reverse transcriptase-polymerase chain reaction (RT-PCR) (one-step or nested RT-PCR assay), and a real-time RT-PCR assay; these molecular methods require skilled technicians and specialized laboratories [4] [5] [8]. As a result, more precise and point-of-care (POC) molecular approaches that enable improved clinical treatment, proper disease management, surveillance, and prevention of dengue outbreaks are required. Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR), is a highly variable locus consisting of short and identical repeated palindromic sequences that are naturally found in the genome of bacteria, separated by large spacer sequences[9]. The CRISPR locus is transcribed to generate pre-crRNA, which is then processed by CRISPR-associated (Cas) proteins to generate mature CRISPR guide RNA (gRNAs) [10]. CRISPR-associated (Cas) proteins specifically act as RNA-guided endonucleases which together with gRNA induce indiscriminate cleavage of target nucleic acid thereby protecting the bacterial cell against any invading bacteriophage [9], [10]. CRISPR-Cas-based diagnosis relies on the identification of viral nucleic acid sequences by gRNAs followed by non-specific cleavage of viral nucleic acid by Cas [11], [12]. This reactionis usually coupled with probe molecules that generate a fluorescent/visual signal which is detected [11], [12]. SHERLOCK (Specific High Sensitivity Enzymatic Reporter Unlocking), DETECTR endonuclease-targeted (DNA CRISPR transReporter), CARVER (Cas13-assisted restriction of viral expression and readout), PAC-MAN (Prophylactic Antiviral CRISPR in human cells), SHINE (Streamlined Highlighting of Infections to Navigate Epidemics), All-In-One Dual CRISPR-Cas12a (AIOD-CRISPR)are examples of CRISPR-Cas molecular methods used in disease diagnosis **[13][15]**. In this study, we have designed DENV-specific CRISPR-Cas13 gRNAs for the conserved and variable/hyper variable genomic regions among four DENV serotypes using bioinformatics tools. These CRISPR-Cas13 gRNA sequences serve as molecular detectors that provide point-of-care molecular diagnosis useful in the detection and clinical evaluation of the dengue virus and its serotypes.

Methodology:

The dataset of complete genome sequences of four Dengue virusserotypesDENV1 (NC_001477), DENV2 (NC_001474), DENV3 (NC_001475), DENV4 (NC_002640) were compiled from the Dengue Virus Variation Resource Centre at National Centre for Biotechnology Information (NCBI) [16]. The Multiple Sequence Alignment (MSA) of these sequences was performed using Bio Edit v 7.2 [17]. The CRISPR-Cas13 gRNAs were designed the using CHOPCHOP v2 CRISPR designing web tool [18]. These gRNA sequences were visually inspected for the number of mismatches and number of bases conserved to obtain potentialCRISPR-Cas13 gRNA sequences showing high specificity (maximum target annealing) and low off-target selectivity (minimum off-target binding) for the conserved and variable portions in the DENV genome. The secondary structures and free energies of gRNA sequences were predicted using the default parameters in RNA fold web-server [19].

. .

©Biomedical Informatics (2022)

Self-

Free

Mismatch in seed region(11-

Table 1	Free energies of	DENV specific gRNA sequences		
S.	Genomic		Start	
No.	region		position	p

INO.	region	CRISPR/Cas13 gRNA	position	position	No. of mismatch es	18 69)	ity	(ΔG in kcal/mo
								1)
1	lncRNA	TAGAGGAGACCCCCCCGAAACAAA AAAC	293	321	1	no	0	-0.7
2	lncRNA	AGAGGAGACCCCCCCGAAACAAAA AACA	294	322	1	no	0	-0.7
3	lncRNA	CTGGGAAAGACCAGAGATCCTGCTG TCT	333	361	1	no	1	-6.1
4	lncRNA	AGATCCTGCTGTCTCCTCAGCATCAT TC	347	375	1	no	2	-6.2
5	lncRNA	CAGAGATCCTGCTGTCTCCTCAGCAT CA	344	372	1	no	3	-6.2
6	lncRNA	AGAGATCCTGCTGTCTCCTCAGCATC AT	345	373	1	no	3	-6.2
s7	lncRNA	TTAGAGGAGACCCCCCCGAAACAAA AAA	292	320	2	no	0	-0.7
8	lncRNA	AGGAGACCCCCCCGAAACAAAAAA CAGC	296	324	2	no	0	-0.7
9	lncRNA	ACGCTGGGAAAGACCAGAGATCCTG CTG	330	358	2	no	1	-5.9
10	lncRNA	CGCTGGGAAAGACCAGAGATCCTGC TGT	331	359	2	no	1	-5.9
11	lncRNA	ATCCTGCTGTCTCCTCAGCATCATTCC A	349	377	2	no	2	-6.2
12	NS5	ATGTATGCCGATGACACCGCAGGA TGGG	1588	1616	4	Yes, 1	0	-0.9
13	NS5	TGTATGCCGATGACACCGCAGGATG GGA	1589	1617	5	Yes, 1	0	-2
14	NS5	ATGACACCGCAGGATGGGATACAAG AAT	1598	1626	5	Yes, 1	0	-1.3
15	NS5	TGACACCGCAGGATGGGATACAAGA ATC	1599	1627	5	Yes, 1	0	-1.3

End



Figure 1: Genome organization of dengue virus

Results and Discussion:

We identified the conserved and variable sequences across the genome of four common DENV serotypes. Subsequently, these selected genomic regions were used to design CRISPR-Cas13gRNAsequences.

Sequence retrieval and multiple sequence alignment:

The extensive whole-genome Multiple sequence alignment (MSA) of four DENV reference serotypes DENV1 (Gen Bank:

NC_001477.1), DENV2 (Gen Bank: NC_001474.2), DENV3 (Gen Bank: NC_001475.2), and DENV4 (Gen Bank: NC_002640.1) revealed lncRNA and NS5 genomic regions that are conserved across all DENV serotypes, while variable sequences found in the NS2A and NS2B gene. The NS2A gene is involved in coordination during RNA packaging and replication whereas the NS2B gene serves as a co-factor in the structural activation of DENV serine protease NS3 [1]. NS2B gene also assists in viral replication and blocks IFN- induced signal transduction [1]. Gene NS5 codes for

©Biomedical Informatics (2022)

methyltransferase domain (located at residues 1-269 amino acids) and RNA-dependent RNA polymerase (located at residue 270-900

amino acids) [1].

DENV1 DENV2 DENV3 DENV4	10 20 30 40 50 60 70 80 90 AAGAAGTCAGGCCGGATTAAGCCATAGCACGGTAAGAGCTATGCTGCTGTGAGCCCCGTCCAAGGACGT AAGAAGTCAGGCCGGATTAAGCCATAGCACGGTAAGAGCTATGCTGCTGTGAGCCCCGTCCAAGGACGT T
DENV1 DENV2	110 120 130 140 150 160 170 180 190 CACCGCTTCGAGCCAGCCGTGCTGCCAGCCTGTAGCTCCATC-GTGGGGATGTAAAAA-CCCGGGAGGCTGCAAACCATGGAAGCCTGTACGCAT . TA. C. T T. A. TA A C. T. A AA. G AT
DENV3 DENV4	$\mathbf{T} \qquad \mathbf{A} \qquad \qquad \mathbf{G} = \mathbf{C} = \mathbf{C} = \mathbf{C} = \mathbf{C} = \mathbf{T} = \mathbf{C} $
DENV1 DENV2 DENV3 DENV4	210 220 230 240 250 260 270 280 290 AGCAGACTAGTGGTTAGAGGAGACCCCTCCCAAGACACAACGCAGCAGCGGGGCCCCAA-CACCAGGGGAAGCTGTACCCTGGTGGTAA .TG. .C .TTA.T. .A.AAT. .G.C.AGG.G.AT. .GT.C.CT.G.
DENV1 DENV2 DENV3 DENV4	310 320 330 340 350 360 370 380 390 AGAGGTTAGAGGAGACCCCCCCCCCCACAACAACAACAACAGCATATTGACGCTGGGAGAGACCAGAGATCCTGCTGTCTCTACAGGATCATTCCAGGCA
DENV1 DENV2 DENV3 DENV4	410 420 430 440 CGCCAAAAAATGGAATGGTGCTGTTGAATCAACAGGTTCT G

Figure 2: Multiple sequence alignment of lncRNA in four dengue serotypes

Designing of CRISPR/Cas13 gRNAs against conserved regions in DENV genome:

MSA of reference DENV serotypes revealed that the most conserved sequences are in the NS5 and long non-coding RNA (lncRNA) regions (Figures 2 and 3). As a result, genomic sequences of the DENV2 (Gen Bank: NC_001474.2) lncRNA (0.425 kb) and NS5 (2.7 kb) regions were used as a query in the CHOPCHOP tool to design CRISPR/Cas13 gRNAs for Dengue diagnosis. The CHOPCHOP online tool returned 293 gRNA sequences for DENV IncRNA and 1973 gRNAs for NS5. These gRNA sequences were visually inspected for the number of conserved nucleotides in the MSA, and the sequences with the highest number of conserved nucleotides among four DENV serotypes were selected, yielding a total of 251 gRNA sequences for lncRNA and 268 gRNA sequences for NS5. To narrow the search even further and to improve gRNA specificity against the target, gRNA sequences were re-examined for the number of sequence mismatches in the lncRNA and NS5 regions, and the gRNA sequences with the least mismatches were selected. We selected a total of 17 lncRNA-specific sequences by selecting gRNA sequences with a mismatch of less than or equal to two nucleotides and a total of 10 NS5 specific gRNA sequences by selecting gRNA sequences with a mismatch of less than or equal to five nucleotides because lncRNA was found to be more conserved among DENV serotypes than NS5, as indicated by MSA. Furthermore, gRNA sequences with no mismatches in the core region, i.e. within a stretch of 11-18nts, were highly target-specific and anneal quickly with complementary sequences in the target genome **[20]**. As a result, we visually inspected the lncRNA and NS5 specific gRNA sequences for mismatches and eliminated those with mismatches within the core region. Our findings showed that 11 of the 17 lncRNA gRNA sequences have no mismatches within an 8-base sequence stretch spanning 11-18 nucleotides, while 4 of 10 NS5 gRNA sequences have only one mismatch within the central seed region (**Table 1**).

Prediction of gRNA secondary structure:

There is a link between the secondary structure of gRNA and gene editing efficiency. The formation of secondary structures in gRNA has a significant impact on cleavage efficiency. More specifically, when the Gibbs free energy (ΔG) for the formation of the most

ISSN 0973-2063 (online) 0973-8894 (print)

Bioinformation 18(8): 661-668 (2022)

stable RNA structure is between -2 and 0 kJmol⁻¹, cleavage efficiencies are highest **[21]**. As a result, the gRNA can invade the target strand without the aid of the gRNA scaffold. RNA fold Web Server was used to evaluate the secondary structures and minimum free energy of lncRNA and NS5 specific gRNA sequences (**Table 1** and Figure 4). We selected two DENV-specific gRNAs based on the minimum free energy, least number of mismatches among DENV serotypes, zero self-complementarity of gRNA, GC content, and no

mismatch within the central region and gRNA secondary structure. One sequence (TAGAGGAGACCCCCCCGAAACAAAAAC) is specific for targeting the lncRNA and the other (ATGTATGCCGATGACACCGCAGGATGGG) is specific for targeting the NS5 gene (Table 1 and Figure 4). These two CRISPR-Cas13 gRNA sequences have a low possibility of self-annealing; allowing them to recognize and bind to the target site within the DENV genome's conserved regions.

	10	20	30	40	50	60	70		90	100	110	120
DENVI	TGACATCTAGAATGC	TGCTAAATCG	TTCACAATGG	TCACAGGAAG	CCAACATAT	GAAAGAGACG	TGGACTTAG	GCGCTGGAAC	AAGACATGT	GGCAGTAGAA	CAGAGGTGGC	CAACC
DENV2	TTAGT		TA O	AT A A	G.C.T.C	GCCGT.	.TC.C.	. AAGC	CC.TA.CA.	C.GGA.T	AGTA.AC.	A
DENV3		.AGCA.C	G G A	A	CCATA	G.AT.	T	.AA	TC	TAAT.CG	AACAC.	A
DENV4	CAA.AGT						.ATC.T.	.GA	GAG	CT.CACT	AAAAAC.	AG. A
	1.3.0	140	150	160	170	180	1 90	200	21.0	22.0	23.0	240
						.	· · · · · · ·			1 • • • • 1 • • • • 1		
DENVI	TAGATATCATTGGCC	AGAGGATAGAG	CAATATAAAAA	TGAACACAAA	TCAACATGG	CATTATGATC	AGGACAATC	CATACAAAAA	ATGGGCCTA	TCATGGATCAT	TATGAGGTCAA	CCAT
DENV2 DENV3	.GG	A. A. A.	A.GG. C. GG	G. G. T. T	A T		.C. A. C.	.c.	GT.	C	A	AG. CA
DENV4	.GACAGA	GAC.TC.	CGAT.GCG	AG	GAC		AC.	G	CG.	AGC	A.CTCC	TT GA
	250	260	270	200	290	300	310	320	330	340	350	360
DENVI	CAGGATCAGCCTCAT	CCATGGTCAAT	CGTGTGTGGTGA	ACTOCTAACO	ABACCATCC	GATGTCATTC	CCATGGTCA	CACAAATAG	CATGACTGA	CACCACACCC	FTTGGACAACA	CACCC
DENV2	.T			G	T	c	G.	GG.	A A	G. T. A		C.C.
DENV3	c	A.A		cc			G.	GG.	A A	A. T. A.	G	A .
DENV4	CTAC.	G	GA.J		••••••••••••••••••••••••••••••••••••••	G	.AG.	.TGT	A	TACT.		AA.
	37.0	38.0	30.0	400	410	4.30	4.30	4.40	45.0	460	47.0	400
DENV1	TGTTTAAAGAGAAAG	TTGACACGCG1	ACACCAAAAG	GAAACGAGGG	ACAGCACAA	ATTATGGAGG	TGACAGCCA	GETCETTAT	CCCTTTTCT	CTCTAGAAAC	AAAAAA <mark>CCC</mark> AG	ATCT
DENV2	. <u>T</u>	.G	AC.A.GC.	GA	GAAGA	C.A A. AA	A AG	AC.T	. AAAGAAT .	AGGG.AG.A.		3G.
DENVA	C G	G T CA Z		A CC 7	CG ATG	G ACCA	C	AT CG	CCC C	TGGA AG G		CG
	490	500	51.0	520	530	540	550	5 60	570	58.0	590	600
		.				.						
DENV1	C A A	TCACAAGAAA	C AGGTCAA	T CT C		AC	G CA G	G G	CGT	T T Z	AACGGTTCTG	G
DENV3	GA	.T		c			.GC	G. AGT.	GCT.	TTG	GAAT	A.A.
DENV4	GA	TCTC	TA		cc	TCAGA.	C.GGG	C A	C.GTA	TA.TC	AGCT	A .
		.								1		
DENVI	TTGTGCACAGAGAGA	GGGAGCTTCAT	AAACAAGGAAJ	ATGTGCCACG	TGTGTCTAC	AACATGATGG	GAAAGAGAG	AGAAAAATT	AGGAGAGTT	CGGAAAGGCAJ	AAAGGAAGTCG	GCAA
DENV2	.GTGAGA.	A.TC	CTTG	GAAA			A	GG <mark>C</mark> .	GA	c	CCA.	AC.
DENV3	AC	.T. A. C	TTGC.	GA.GC			.C	· · · · · · · · · · · · · · · · · · ·	T	TC.A	· · · · · · C · · · A · ·	A T.
DENVA	. (s 11(s										-	
		CCAC								TC.GAC.		
	730	740	750	760	770	780	7.90	800	81.0	TC.GAC.	GC	840
	730	740	750	760	770	780	790	noo	810	n20		n40
DENV1	730 TATGGTACATGTGGT	740 	TTTTAGAGT	TGAAGCCCT	770 GGTTTCATG	780	7 90 	GCAGAGAGA	TTCACTCAG	n20 TGGAGTGGAAC	nan GAGAAGGACT	
DENV1 DENV2 DENV3	730 TATGGTACATGTGGT	740 TEGGAGCGCG	750	760	770 	AATGAAGATC	790 	GCAGAGAGA	810 TTCACTCAG CCG	T.C.GAC.	GAGAAGGACT	840 CACA
DENV1 DENV2 DENV3 DENV4	730 TATGGTACATGTGGT 	740 TGGGAGCGCGG	750 TTTTTAGAGTT C. 3.ACC.T. 3C.G.A	760 TGAAGCCCTI	770 	780	790 ACTGGTTCA	GCAGAGAGAZ C.CGC.T.A.	n10 TTCACTCAG CC.G. CTTA. TGG.	T. C. GAC.	GAGAAGGACT	040 CACA
DENV1 DENV2 DENV3 DENV4	730 TATGGTACATGTGGT	740 TGGGAQCGCG T.A.CA.C	750 	760 TGAAGCCCTT	770 	780	790 ACTGGTTCA T	GCAGAGAGA C.CGC.T.A.	TTCACTCAG C. C. G. C. C. TTA	T. C. GAC.	GAGAAGACT	040 CACA
DENV1 DENV2 DENV3 DENV4	730 TATGGTACATGTGGT 	740 740 TGGGACCCCC T	750 	760 TGAAGCCCTT 	770 GGTTTCATG 	780 1	790 ACTGGTTCA T T 910	GCAGAGAGA C	810 TTCACTCAG CC.G. CTTA TGG.	T.C.GA.C.	GAGAAGGACT GAGAAGGACT G. G. T. G. T. 950	840 CACA 3 3 960
DENV1 DENV2 DENV3 DENV4	730 TATGGTACATGGT 	740 TGGGAGCGCGC T	750 		770 	780 	790 ACTGGTTCA T T 910 CCGGATGGG		810 TTCACTCAG CC.G. C.TTA 930 930	940		840
DENV1 DENV2 DENV3 DENV4 DENV1 DENV2	730 TATGGTACATGTGGT C C C C C C C C C C	740 TGGGAGCGCGG .TA CA.C CA.C CA.C 	750 TTTTTAGAGT 3. C. G. A 	760 TGAAGCCCTT 	770 	700 AATGAAGATC C. 900 GATGACACAG	790 	000 GCAGAGAGAGA C.C.T.A. 920 920 ACACAAGAA	010 TTCACTCAG C. C. G. TTG. TGG. 930 3ACAGAGGA C CTA.	940 TGATCTTCAGI A. C. AA.A	BOD BOD BOD BOD BOD BOD BOD BOD BOD BOD	960 960
DENV1 DENV2 DENV3 DENV4 DENV1 DENV2 DENV3	730 TATGGTACATGTGGT 	740 740 TGGGACCCCCC T. A. A. C. A. C.	750 	760 TGAAGCCCT C	770 	780 	790 	000 GCAGAGAGAGA CGC.T.A. 920 1 ACACAAGAA T.	010 TTCACTCAG C.C.G. C.TTA 	T. C. GAC. 120 TGGAGTGGAAC A 940 940 TGATCTTCAGJ A. C. AA.A.	GGAGAAGGACT GGAGAAGGACT GT 950 950 AATGAGGCCAA A.A.T. A.A.A.T.	960
DENV1 DENV3 DENV4 DENV4 DENV1 DENV2 DENV3 DENV4	730 TATGGTACATGTGGT .CC .CC .CC .A.CTTGGATACATAC .G. A. CT. G. T. G. T. C.	740 740 TCAGAACCACA 	750 		770 CGTTTCATG A.A.T.A A.T.A .T.A .T.A .T.A .T.A .T	700 	7 90 	800 GCAGAGAGAG C.C.C.T.A. 920 I.C.C.C.G.T.A. 920 I.C.C.C.A.GAAA T.	010 TTCACTCAG C. C. G. C. TTA. TGG. 930 930 930 C. CTA. C. T.	T.C.GAC. 120 1700 1000 1		940 CCACA 3 960 4ATCA 3 3 3 3 3
DENV1 DENV2 DENV3 DENV4 DENV4 DENV2 DENV3 DENV4	730 TATGGTACATGTGGT 	740 	750 TTTTAGAGT 3. ACC 3. ACC 4. C. G. A. 1. C. G. A. 1. C. A. 1. C 1. C	760 TGAAGCCCTT C	770 	700 100 000 1000	790 	100 GCAGAGAGAG C. CGC.T.A. 920 1040	010 TTCACTCAG C. C. G. C. TTA. 	T. C. GA C. 1 10 1 10 TCGACTCGAAC 	азо саладаасаасасаа с с с с с с с с с с с с с с с	960 960 960 971 960 960 960 960 960 960 960 960 960 960
DENV1 DENV2 DENV3 DENV4 DENV1 DENV2 DENV3 DENV4	713 TATGGTACATGGT 	740 740 T. A. C.	750 TTTTAGAGT C C C C C C C C	760 TGAAGCCCTT A C 	770 GGTTCATG A.T.A noo il argTATGCA T T 1010	7 80 	790 	920 0 40 40	810 TTCACTCAG C. TTA . •30 •30 •30 •30 C. CTA . C. CTA . C. CTA . •30 •30 •30 •30 •30 •30 •30 •30	T.C.GAC. 120 110 TCGACTGGAAC 940 TCGACTTCACJ A.C.AA.A. 1060 1060	6	960 960 960 960 960 960 960 960 960 960
DENV1 DENV2 DENV3 DENV4 DENV1 DENV3 DENV3 DENV4	730 TATGGTACATGGT 	740 740 TGGAACGGGG TGAACACGGGG 860 TCAAGAGACATG 600 100 100 100 100 100 100 100	750 TTTTAGAGT C.C.T. J.ACC.T. J.ACC.T. MTCAAAGATTCC GAC.AAG 990 CTATTGGCCAO	760 TCAAAGCCT C	770 GGTTCATG A.T.A B.T.A I	700 ATCAAAAC	790 790 700 70 70 910 910 70 70 70 70 70 70 70 70 70 70 70 70 70	900 CCACAGAGAA CCC.T.A. 920 1040 1040 CCCCCAGAG	010 TTACTCAG C. C. G. G. TTA 	10.00 10		840 840 82 82 82 83 84 84 84 84 84 84 84 84 84 84 84 84 84
DENV1 DENV2 DENV3 DENV4 DENV4 DENV2 DENV3 DENV4 DENV1 DENV1 DENV2	713 TATGGTACATGGT 	740 TGGGAGCGCGC TA CA.C 	750 TTTTAGAGT ACC. AC	760 TGAAGCCCT C	770 NGTTTCATG A. A. T. A A. T. A . T. A . T. A . T. A . T. A . T. A 	700 AATGAAGATC C. 900 GATGACACAC C. 1020	790 790 910 910 010 010 010 010 010 010 010 0	000 CCACAGAGAGA CGC. T. A. 920 ACACAAGAA .T. 1040 .T. A.	010 TTCACTCAG C C TTA. C. TTA. 010 AACAGAGGA C CTA. 1050 110 1050 110 1050 110 1050 1050 1050 1050	ПССАД. (1) 1000		960
DENV1 DENV2 DENV3 DENV4 DENV1 DENV3 DENV3 DENV4 DENV1 DENV2 DENV4	730 TARGGTACARGTGGT 	740 TGGGAGCGCG TGGGAGCGCG TGGGAGCGCG TGGGAGCGCG 000 000 000 000 000 000 000	750 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	760 TGAAGCCCTT 	770 	700 AATGAARGATC 000 000 GATGACACAG 1020 1020	290 	000 GCAGAGAGA C	a10 TTCACTCAG C. C. G. 	020 020 020 020 020 020 020 020	G C C C C C C C C C C C C C C C C C C C	840 840 840 840 840 840 840 840 840 840
DENV1 DENV2 DENV3 DENV4 DENV2 DENV2 DENV4 DENV4 DENV1 DENV3 DENV4	730 TATGGTACATGTGGT 	740 TGGGAGCGCG TGGGAGCGCGCG T ACA.(CA.(750 	760 760 760ACCCTT 760ACCCTT 800 800 100 1000 1	770 GGTTTCATG A. T. A T. A. T. A 	700 AATCAAGATC C. 000 GATCACACAC C. 1020 TACCAAAACA T.	790 	800 GCAGAGAGA C. C. C. C. ACACAAGAA 1040 1040 1040 1040 1040 1040 1040 10	010 TTCACTCAG C. C. G. TTGA TGG. 930 TGG. 0.10 TGC. 0.10 TG	820 1	азо Саларана Салара	960 960 960 960 960 960 960 960 960 960
DENV1 DENV2 DENV3 DENV4 DENV3 DENV3 DENV4 DENV2 DENV3 DENV4	730 	740 740 TGGGAGCGGCGCG T A. C.	750 	760 TGAAGCCCT 	770 	700 ATCAAGATC 000 000 1020 TACCAAAAC TACCAAAAAC 1020 TACCAAAAAC	910 910 910 910 910 000 000 000 000 000	000 GCAGAGAGA C. 	110 TTCACTCAG C. C. G. - TTA - TGG 930 - 10 - C. CTA - C. CTA - C. CTA - C. CTA - C. CA - CA	820 176GACTGGAAG 940 17GATCTTCAG A. C. AA. 1060 1. AATTGGAACC G. C. G	азо Сасаласаст Сасаласаст Сасаласаст Сасаста Аласассала Аласассала Аласассала Аласассала Сасасассала Аласассала Сасасассала Аласассала Аласассала Аласассала Аласассала Аласассала Аласассала Аласассала Аласассала Аласассала Аласассала Аласассала Аласассала Аласассала Аласассала Аласасассала Аласасассала Аласасассала Аласасассала Аласасасала Аласасасала Аласасасала Аласасаса Аласасаса Аласасаса Аласасасаса Аласасасаса Аласасасаса Аласасасасасаса Аласасасасасасасасасасасасасасасасасасас	960 960 960 960 960 960 960 960 960 960
DENV1 DENV2 DENV3 DENV4 DENV2 DENV2 DENV3 DENV4 DENV3 DENV3 DENV3 DENV4	730 TATGGTACATGTGGT 	740 740 7603A676C6 T. A. C. C. 760 760 760 760 760 760 760 760	750 TTTTTAGGT . C.G.A. . C.G.A. . C.G.A. . C.G.A. . C.G.A. . C.A.G. . C.A.G. . C.A.G. . C.A.G. . A.G. . A.G. . A.G. J. . C.A.G. . A.G. . A.G. J. . C.A.G. J. . A.G. J.	760 760 760 760 760 760 760 760	770 GGTTTCATG A. T. A T. T. T. 90 1010 1000 1000 1000 1000 1000 100	700 AATCAACAGATC 000 GATCAACACA C 1020 T T CCAAAACA T 1140	790 	800 GCAGAGAGA CCC.T.A. 920 ACACARGAN T. 1040 GGGGCCAGAG AA.T.AC AA.C.TC. 1160	110 TTCACTCAG C. C. G. 910 930 3ACAGAGGA C. CTA. 1010	820 176GAGTGGAAG 176GAGTGGAAG 940 1	азо Саларана ала аларана аларана ала ала ала ала ала ала ала ала ала	040 040 200 200 200 200 200 200 200 200
DENV1 DENV2 DENV3 DENV4 DENV4 DENV3 DENV4 DENV4 DENV4 DENV4 DENV4 DENV4	730 	740 740 TGGGAGCGGCGCGC T A. CA. C A. C. CA. C A. C. C. CA. C GA. C. C. CA. C 900 CTGAACATGCC GGA. CAAC 000 1100 1100 100 100 100 100	750 C C C C C C C C C C C C C C C C C C C	760 TGAAGCCCT 	770 CGTTTCATG A. T. A 	700 AATGAAGATC 900 GATGACACAC 1020 TACCAAAACA 1140 ACCAAAACA	700 	CCC. T. A 920 ACAAGAA 1040 GOTTCAGAA A. C. TC. 1100 1140 1160	810 TTCACTCAG C. C. G. 930 3ACAGAGGA C. TTA 1050	20 17GCACTCGCAAC 940 17GCACTCGCAAC 940 17GCACTCACC A. C. AA. 1060 100 100 100 100 100 100 10	азо Сасаласаста Сасаласаста Сасаласаста Сасаласаста А.А.А.Т Золо ТССАТОСАСТСА А.А.А. Забото Сасаласаста А.А.А. Забото Сасаласаста А.А.А. Эво Сасаласаста А.А.А. Эво Сасаласаста Сасаласаста Сасаласаста Сасаласаста Сасаласаста А.А.А. Эво Сасаласаста А.А.А. Эво Сасаласаста Сасаласаста Сасаласаста А.А.А. Эво Сасаласаста А.А.А. Эво Сасаласаста Сасаласаста А.А.А. Эво Сасаласаста Сасаласаста Сасаласаста Сасаласаста Сасаласаста Сасаласаста Сасаласаста Сасаласаста Сасаласаста Сасаласаста Сасаласаста Сасаласта Сасаста Сасаста Сасаста Сасаста Сасаста Сасаста Сасаста Сасаста Сасаста Сасаста Сасаста Сасаста Сасаста Сасаста Сасаста Саста Сасаста Саста Састаста Саста Саста Саста Сас	040 040 040 040 040 040 040 040
DENV1 DENV2 DENV3 DENV4 DENV2 DENV2 DENV3 DENV4 DENV3 DENV4 DENV4 DENV4	730 TATGGTACATGTGGT 	740 740 740 7603A05400 7706AACATA A. CA. C 700 700 700 700 700 700 700 70	750 	760 760 760 760 760 760 760 760	770 770 770 770 770 770 770 770	700 AATCAARGATC 000 GATCAACACA C 1020 TACCAARACA 1140 ACCAACACACA T	700 	000 GCAGAGAGAS CCC.T.A. 920 ACACAAGAAT .T. 10.40 10.40 	110 TTCACTCAG C. C. G. 	820 1	азо азадаладаст ададла ададаст ададла ададла ададла ададаст ададла адада ададла ад	000 000 000 000 000 000 000 000 000 00
DENV1 DENV2 DENV3 DENV4 DENV3 DENV3 DENV4 DENV3 DENV4 DENV3 DENV4	730 	740 740 TGGGAGCGGC T A. C.	750 C C C C C C C C C C C C C C C C C C C	760 TGAAGCCCTT 	770 CGTTTCATG A. A. T. A 90 1010 100	700 AATGAAGATC 900 GATGACACAC 1020 TACCAAAACA T 1140 ACCAACAACA	700 	CCC.T.A. OCC.T.A. OCC.T.A. OCC.T.A. OCC.T.A. OCC.TC. OCC.TC. OCC.TC. OCC.TC. OCC.C.C.	810 TTCACTCAG C. C. G. 930 7ACAGAGGA C. TTA 930 7ACAGAGGA C. T. 930 7ACAGAGGA 1050 1	20 17 17 17 17 17 17 17 17 17 17	азо азо ададаладаст ададаладаст ададаладаст ададаладаст ададаладаст ададалада ададала ададалада ададалада адада адада адада адада адада ада адада адада адада адада адада адада адада адада адада адада адада адада адада адада ада адада а	040 040 040 040 040 040 040 040

Figure 3: Multiple sequence alignment of NS5 (700-1200 nt positions) in four dengue serotypes

Designing of serotype-specificCRISPR/Cas13 gRNAs:

Using reference genomic sequences of the NS2A and NS2B genes from four DENV serotypes, the CHOPCHOP online tool identified a total of 3,072 CRISPR-Cas13 gRNA sequences, with 776, 772, 763, and 761 gRNA sequences for the NS2A and NS2B regions of DENV1, DENV2, DENV3, and DENV4 respectively. The number of conserved bases in the NS2A and NS2B regions of these gRNA was visually inspected in MSA. The gRNA sequences showing the lowest number of conserved bases among the serotypes was selected to provide the most variable serotype-specific targets. It narrowed down serotype-specific gRNAs to 143, 173, 136, and 275 for DENV1, DENV2, DENV3, and DENV4 respectively. To further reduce the number of gRNA sequences we identified gRNA sequences with the highest specificity and selected only those gRNAs that show the least conserved bases not more than 3 bases among serotypes. This screening further reduced the gRNA sequences to 5, 19, 14, and 11for DENV1, DENV2, DENV3, and DENV4 respectively (**Supplementary Table 1**, selected gRNA sequences).

Prediction of secondary structure in gRNA:

The secondary structures and minimum free energy of DENV serotype-specific gRNAs were predicted to aid in the selection of gRNAs that are more efficient and specific to target RNA. For the targets DENV1, DENV2, DENV3, and DENV4, we obtained 1, 5, 5, and 2 potential serotype-specific gRNA sequences with zero or nearly zero free energy, respectively (**Figure 5**). These gRNAs were then screened for minimum free energy, secondary structure, and the fewest number of conserved bases across DENV serotypes. We got one gRNA for each of the DENV serotypes, which were TCAAAACAACTTTTTCATTGCACTATGC,

AACCCTCTCAAGAACCAGCAAGAAAAGG,

TTAGCTTGAAAGACACACTCAAAAGGAG, and

AACAGCACTCATCCTAGGAGCCCAAGCT for DENV1, DENV2,

DENV3, and DENV4 respectively (**Figure 5**, gRNA in bold). These gRNA sequences have the potential for efficient binding with the genome of selected DENV serotypes. CRISPR-based diagnostics are next generation biosensing techniques for detection of viral and bacterial infections. SHERLOCK has been used for detection of Dengue, Zika, bacterial infections, and tumour changes in cell-free DNA using CRISPR-Cas13a [13], [22]. These *In-silico* designed28nt long gRNAs sequences can detect complementary sequences in the DENV genome. These DENV-specific gRNA sequences coupled with Cas13 enzyme can be used in sequence-specific cleavage of the target genome and reporter molecules which can be developed into molecular test useful in identification of dengue virus and its serotypes. Molecular tests like these provide rapid detection and quick identification of virus in clinical samples from dengue infected patients providing an early and accurate disease diagnosis.

Conclusions:

Early and serotype specific diagnosis is important in management of spread of the any diseases and treatment of patients. CRISPR-Cas13 based diagnostics are highly sensitive and specific as well as cost-effective, fast, and can be used as point of care diagnostics. The present study gRNA sequences specific for DENV were designed to provide molecular diagnosis. These gRNA will be valuable for *in vitro* evaluation the sensitivity and efficiency of CRISPR-Cas13-based POC diagnostics.

List of Abbreviations:

DENV: Dengue virus; ZIKV: Zika virus; ELISA: Enzyme-linked Immunosorbent Assay; PCR: Polymerase Chain Reaction; RT-PCR: reverse transcriptase-polymerase chain reaction; CRISPR: Clustered Regularly Interspaced Short Palindromic Repeat; Cas: CRISPR associated proteins; gRNA: guide RNA; lncRNA: Long non-coding RNA; DF: Dengue Fever; DHF: Dengue Haemorrhagic Fever; DSS: Dengue shock syndrome; ORF: Open Reading Frame; UTR: Untranslated region; Capsid; prM/M: C: premembrane/membrane; NS: Non-structural; E: envelope; NASBA: Nucleic acid sequence-based amplification assay; SHERLOCK: Specific High Sensitivity Enzymatic Reporter Unlocking; DETECTR: DNA endonuclease-targeted CRISPR transReporter; CARVER: Cas13-assisted restriction of viral expression and readout; PAC-MAN: Prophylactic Antiviral CRISPR in human cells; SHINE : Streamlined Highlighting of Infections to Navigate AIOD-CRISPR:All-In-One Dual CRISPR-Cas12a; Epidemics; FELUDA: FNCas9 Editor Limited Uniform Detection Assay; MSA: Multiple sequence alignment; IFN: Interferons; ΔG : Gibbs free energy; RAA: recombinase aided amplification; STOP: SHERLOCK testing in one pot; CREST: Cas13-based, rugged, equitable, scalable testing; HPV: Human papillomavirus.



Figure 4: Secondary structure of selected gRNA designed against conserved regions of DENV lncRNA and NS5 regions

©Biomedical Informatics (2022)



Figure 5: DENV Serotype specific gRNAs after designing and screening, Abbreviation: MNCB= minimum number of conserved bases found among reference serotypes in alignment

Acknowledgments:

This work was supported by the Senior Research Fellowship awarded to Archana Prajapati by the Council of Scientific and Industrial Research (CSIR), Human Resource Development group Government of India is highly acknowledged.

References:

- [1] Harapan H et al. Viruses 2020 12:829. [PMID: 32751561]
- [2] Waman VP et al. Peer [2016 4:e2326. [PMID: 27635316]
- [3] Laiton-Donato K *et al. Virol. J.* 2019 **16**:62. [PMID: 31068191]
- [4] Cecchetto J *et al. Biosens. Bioelectron.* 2020 **151**:111972. [PMID: 31999580]
- [5] Mardekian SK and Roberts AL, *Biomed Res. Int.* 2015 2015.[PMID: 26509163]
- [6] Wong PF *et al.J. Infect. Public Health* 2020 **13**:198. [PMID: 31405788]
- [7] Shu P and Huang J, Clin. Diagn. Lab. Immunol. 2004 11:650.[PMID: 15242935]
- [8] Muller DA, J. Infect. Dis. 2017 215:S89. [PMID: 28403441]
- [9] Prajapati A and Nain V, *Bioinformation* 2021 17:645. [PMID: 35173386]

- [10] Jolany Vangah S *et al. Biol. Proced. Online* 2020 22:14. [PMID: 32939188]
- [11] Kumar P et al. Front. Cell. Infect. Microbiol. 2020 10:576875.[PMID: 33251158]
- [12] Rahimi H et al. ACS Sensors 2021 6:1430. [PMID: 33502175]
- [13] Gootenberg JS et al. Science 2017 356: 442. [PMID: 28408723]
- [14] Ding X et al. Nat. Commun. 2020 11:10. [PMID: 32948757]
- [15] Arizti-Sanz J et al. Nat. Commun. 2020 11. [PMID: 33219225]
- [16] Benson DA *et al. Nucleic Acids Res.* 2013 **41**:42. [PMID: 23193287]
- [17] Hall T GERF Bull. Biosci. 2011 2:61.
- [18] Labun K et al. Nucleic Acids Res. 2016 44:W276. [PMID: 27185894]
- [19] https://rna.tbi.univie.ac.at/cgibin/RNAWebSuite/RNAfold.cgi.
- [20] Bandaru S et al. Sci. Rep. 2020 10:12. [PMID: 32665590]
- [21] Jensen KT et al. FEBS Lett. 2017 591:1901. [PMID: 28580607]
- [22] Vatankhah M et al. Crit. Rev. Clin. Lab. Sci. 2021 58:241. [PMID: 33245685]

ISSN 0973-2063 (online) 0973-8894 (print) Bioinformation 18(8): 661-668 (2022)

©Biomedical Informatics (2022)

