Supporting Information

Zika virus genomic RNA possesses conserved G-quadruplexes characteristic of the *Flaviviridae* family

Aaron M. Fleming,* Yun Ding, Anton Alenko, and Cynthia J. Burrows*

Department of Chemistry, University of Utah, 315 S 1400 East,

Salt Lake City, UT USA 84112-0850

*To whom correspondence should be addressed.

Phone: (801) 585-7290 E-mail: burrows@chem.utah.edu; afleming@chem.utah.edu

Item	Page
Methods	1
Figure S1. Predicted G-quadruplexes using QGRS mapper, Quadparser, and G4Hunter	7
Figure S2. Complete sequence alignments from the *Flaviviridae* family for the PQS in the prM, E, NS1-A, NS1-B, NS3, NS5-A, and NS5-B gene regions	17
Figure S3. Alignment statistics	24
Figure S4. Complete sequence alignment for the Zika specific PQS in the 3`-UTR	26
Figure S5. Sequence logos generated from the alignment data	27
Figure S6. Data for the randomized genomes with the same nucleotide frequency as the Zika virus	30
Figure S7. Representative T_m curves	31
Figure S8. Representative thioflavin T emission spectra	32
Figure S9. Denaturing gel electrophoresis for polymerase bypass of the NS5-A and 3`-UTR PQS	33
References	34

Methods

Identification of PQS sites in the Zika viral genome:

The Zika viral genome was obtained from the NCBI database (reference sequence: NC_012532.1). The genome was analyzed for PQS sites using three different published algorithms (QGRS mapper, Quadparser, and G4Hunter).[1-3] The G-quadruplexes predicted by these algorithms are provided in Figure S1.

Sequence alignment and analysis:

The 66 viral genomes from the *Flaviviridae* family were downloaded from the NCBI database (each genome's reference number is provided below). The genomes were globally aligned in R studio using the DECIPHER package.[4] We did not conduct any local alignment on the sequence population, meaning the loop regions between G runs remained constant and were not allowed to float from sequence to sequence. The global alignment allowed >56 of the viral genomes to be aligned at each of the 7 PQSs identified. Next, the G-quadruplex sequences were extracted in R for further data analysis using the session information provided below. The aligned sequences were loaded into MEGA 7.0.16 software for phylogenetic analysis.[5] The Zika viral strain genomes were downloaded in R studio by rentrez_1.0.2 from NCBI. A similar process was performed to extract G-quadruplex sequences from the Zika virus strains and to conduct the phylogenetic analysis.

```
> sessionInfo()
R version 3.2.2 (2015-08-14)
Platform: x86_64-apple-darwin13.4.0 (64-bit)
Running under: OS X 10.11.4 (El Capitan)

locale:
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

attached base packages:
[1] stats4    parallel  stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
 [1] rentrez_1.0.2                        TxDb.Hsapiens.UCSC.hg19.knownGene_3.2.2
 [3] GenomicFeatures_1.22.13              AnnotationDbi_1.32.3
 [5] Biobase_2.30.0                       BSgenome.Mmusculus.UCSC.mm9_1.4.0
 [7] BiocInstaller_1.20.3                 DECIPHER_1.16.1
 [9] RSQLite_1.0.0                        DBI_0.4-1
[11] ape_3.4                              BSgenome_1.38.0
[13] rtracklayer_1.30.4                   Biostrings_2.38.4
[15] XVector_0.10.0                       GenomicRanges_1.22.4
[17] GenomeInfoDb_1.6.3                   IRanges_2.4.8
[19] S4Vectors_0.8.11                     BiocGenerics_0.16.1

loaded via a namespace (and not attached):
 [1] zlibbioc_1.16.0          GenomicAlignments_1.6.3    BiocParallel_1.4.3
 [4] lattice_0.20-33          R6_2.1.2                   httr_1.1.0
 [7] tools_3.2.2              grid_3.2.2                 SummarizedExperiment_1.0.2
[10] nlme_3.1-128             lambda.r_1.1.7             futile.logger_1.4.1
[13] futile.options_1.0.0     bitops_1.0-6               curl_0.9.7
[16] RCurl_1.95-4.8           biomaRt_2.26.1             Rsamtools_1.22.0
[19] XML_3.98-1.4             jsonlite_0.9.20
>
```

Reference numbers for the genomes studied:

		NCBI Reference Number	GenBank Number
1	Zika virus	NC_012532.1	
2	West Nile virus		AY646354.1
3	Dengue virus		EF025110.1
4	Tick-borne encephalitis virus	NC_001672.1	
5	Yellow fever virus		AF094612.1
6	Kadam virus from Uganda		DQ235146.1
7	Absettarov virus		KJ000002.1
8	Alkhurma virus	NC_004355.1	
9	Gadgets Gully virus		DQ235145.1
10	Karshi virus		DQ462443.1
11	Langat virus	NC_003690.1	
12	Louping ill virus		KF056331.1
13	Powassan virus		KU886216.1
14	Royal Farm virus		DQ235149.1
15	Sokoluk virus	NC_026624.1	
16	Kama virus	NC_023439.1	
17	Meaban virus		DQ235144.1
18	Saumarez Reef virus		DQ235150.1
19	Tyuleniy virus		KT224356.1
20	Aedes flavivirus		KC181923.1
21	Barkedji virus		KC496020.1
22	Chaoyang virus	NC_017086.1	
23	Culex theileri flavivirus		HE574574.1
24	Culiseta flavivirus		KT599442.1
25	Donggang virus		JQ086551.1
26	Hanko virus		JQ268258.1
27	Kamiti River virus	NC_005064.1	
28	Lammi virus		FJ606789.2
29	Nakiwogo virus		GQ165809.2
30	Nounane virus		EU159426.2
31	Nhumirim virus	NC_024017.1	
32	Nienokoue virus	NC_024299.1	
33	Palm Creek virus		KC505248.1
34	Quang Binh virus	NC_012671.1	
35	Xishuangbanna aedes flavivirus		KU201526.1
36	Aroa virus		KF917535.1
37	Bussuquara virus		AY632536.4
38	Kedougou virus	NC_012533.1	
39	Cacipacore virus	NC_026623.1	

40	Koutango virus		EU082200.2
41	Ilheus virus		KC481679.1
42	Japanese encephalitis virus (strain Beijing-1)		L48961.1
43	Murray Valley encephalitis virus		KM259934.1
44	Rocio virus		AY632542.4
45	St. Louis encephalitis virus	NC_007580.2	
46	Usutu virus		KF573410.1
47	Yaounde virus		EU082199.2
48	Kokobera virus	NC_009029.2	
49	New Mapoon virus		KC788512.1
50	Stratford virus		KF917540.1
51	Bagaza virus		KR108246.1
52	Baiyangdian Virus		JF312912.1
53	Duck egg-drop syndrome virus		JQ920426.1
54	Israel turkey meningoencephalomyelitis virus		KC734553.1
55	Ntaya virus	NC_018705.3	
56	Tembusu virus		KF192951.1
57	Spondweni virus	NC_029055.1	
58	Banzi virus		DQ859056.1
59	Bamaga virus		KU308380.1
60	Bouboui virus		DQ859057.1
61	Edge Hill virus		DQ859060.1
62	Jugra virus		DQ859066.1
63	Saboya virus		DQ859062.1
64	Sepik virus	NC_008719.1	
65	Uganda S virus		DQ859065.1
66	Wesselsbron virus	NC_012735.1	

Oligomer preparation. The RNA oligomers were synthesized by the DNA/Peptide core facility at the University of Utah using commercially available phosphoramidites and standard protocols for solid-phase synthesis. After the oligomers were cleaved from the solid support and deprotected via standard protocols, they were HPLC purified. Purification was achieved using a semi-preparative, ion-exchange HPLC column running line A = 1:9 ddH_2O:MeCN, and line B = 20 mM Tris (pH 8) and 1 M LiCl in 1:9 ddH_2O:MeCN with a flow rate = 3 mL/min while monitoring the elution via the absorbance at 260 nm. After purification, the oligomers were dialyzed against ddH_2O for 36 h while changing the water three times daily to remove the purification salts. The dialyzed samples were lyophilized to dryness and resuspended in ddH_2O. The concentrations were determined by the absorbance at 260 nm using the primary sequence to determine the extinction coefficients. All RNA samples were stored at -20 °C when not being studied. The G-quadruplex strands were annealed in the desired salt and buffer by heating the samples to 90 °C for 5 min and then slowly cooling the samples to room temperature over ~4 h. After reaching room temperature, the samples were stored at 4 °C for 24 h prior to their study.

Native gel electrophoresis. The G-quadruplex samples were 5` labeled with γ-^{32}P-ATP using T4-polynucleotide kinase following the manufacturer's protocol (New England Biolabs). The labeled

oligomers were passed through a G-25 column following the manufacturer's protocol (GE Healthsciences) to remove unreacted γ-^{32}P-ATP. Next, 20,000 cpm of radiolabeled oligomer was mixed with 10 µM unlabeled strands in the analysis buffer (20 mM lithium cacodylate buffer (pH 7.4) with 140 mM KCl and 12 mM NaCl), and the mixtures were annealed by heating to 90 °C for 5 min followed by slowly cooling to room temperature. The annealed samples were placed at 4 °C overnight prior to analysis. The samples were then applied to a 20% acrylamide native gel and electrophoresed for 8 h at 4 °C. The thrombin binding aptamer was used as a positive control because it adopts a two-tetrad G-quadruplex and single-stranded controls of 20, 25, and 30 nucleotide lengths were provided by homopolymers of 2`-deoxycytidine. The gel was then placed in a phosphor screen and left to expose for 18 h before imaging via phosphorimager autoradiography. The bands were visualized using ImageQuant software.

^{1}H-NMR analysis. The G-quadruplex samples were annealed in 300 µL at a 300 µM concentration in 20 mM KP_i (pH 7.0) and 50 mM KCl in 9:1 H_2O:D_2O. The annealed samples were placed in a D_2O-matched Shigemi NMR tube. The samples were analyzed on an 800-MHz NMR spectrometer with the temperature set to 24 °C. Each sample was scanned 1024 times using the Watergate solvent suppression pulse sequence.

CD analysis. The G-quadruplex samples were annealed at 10 µM concentrations in 20 mM lithium cacodylate buffer (pH 7.4) with 140 mM KCl and 12 mM NaCl. The samples were placed in a 0.2 cm quartz cuvette for CD analysis at 20 °C. The recorded data were background subtracted and then normalized on the y-axis to units of molar ellipticity for plotting and comparative purposes.

***T_m* analysis**. The annealed G-quadruplex samples at 5 µM concentration in the same buffer used in the CD studies were used for T_m analysis. The samples were placed in a UV-vis cuvette and thermally equilibrated at 20 °C for 3 min prior to initiation of the heating experiment. The samples were heated at 1 °C/min and held at each degree for 1 min prior to making an absorbance reading at 260 and 295 nm. The samples were heated to 90 °C. Plots of the temperature vs. absorbance were constructed and evaluated by two-point analysis using the manufacture's software (Shimadzu).

Thioflavin T fluorescence analysis. The G4 samples were annealed at 4 µM concentrations in 20 mM lithium cacodylate buffer (pH 7.4) with 140 mM KCl and 12 mM NaCl. The fluorescence assay was conducted by diluting the G-quadruplex stocks to a final concentration of 1 µM with 0.5 µM thioflavin T in the same buffer in which they were annealed. The samples were placed in a 0.2 cm quartz cuvette with the spectrometer set to excite thioflavin T at 425 nm. The emission spectra were collected over the range of 440 to 700 nm at 2 nm intervals. The experimental spectra had the buffer background spectrum subtracted from them prior to data analysis.

Polymerase stop assay (reverse transcription). Stock solutions of RNA template, 5'-^{32}P labeled DNA primer, reverse transcription buffer, and dNTPs were combined to give final concentrations of 20 mM Tris-HCl pH 7.5, 3 mM $MgCl_2$, 150 mM LiCl or KCl, 5 mM DTT, 500 µM dNTPs, and 200 nM (200 nM template, 212 nM primer), 500 nM (500 nM template, 530 nM primer) or 1 µM (1.0 µM template, 1.1 µM primer) template-primer duplex in 20 µL. The mixture was incubated at room temperature for 15 minutes, and then 1 µL of Superscript III (200 U/µL) was added. Reverse transcription was performed at 37 °C for 1 h. The reactions were quenched by addition of an equal volume of gel loading buffer (8 M urea, 20 mM EDTA, 20 mM Tris-HCl pH 7.8, 0.25% bromophenol blue, 0.25% xylene cyanol FF) followed by a 10 min incubation at 93 °C. The resulting cDNAs were resolved on a 20% denaturing polyacrylamide gel. To produce a G-lane standard, 10 mM dNTP mixture was substituted with the one containing 20 mM ddCTP, 1 mM dCTP, and an unchanged amount (10 mM) of other dNTPs.

Polymerase Extension Sequences:

3`-UTR RNA Template 5'-AACUUC**GG**C**GG**CC**GG**UGU**GGGG**AAAUCCAUGGUUUCUCACGGA-3'

3`-UTR Primer (16nt) 5'-TCCGTGAGAAACCATG-3'

NS5-A RNA Template (41nt) 5'-AGACGU**GG**A**GG**U**GGG**AC**GGG**AGAGACUCUGGGAGAGAAGUG-3'

NS5-A RT Primer (16nt) 5'-CACTTCTCTCCCAGAG-3'

Figure S1. Predicted G-quadruplexes using QGRS mapper, Quadparser, and G4Hunter

The three algorithms used to inspect the Zika viral genome for G-quadruplexes did identify the conserved PQS sites found in the sequence alignment studies.

PQS Sequences Identified with QGRS Mapper

QGRS sequences found (overlaps not included)

Position	Length	QGRS	G-Score
187	27	**GGGAGGTTTGAAGAGGTTGCCAGCCGG**	14
224	28	**GGTCATGGACCCATCAGAATGGTTTTGG**	13
314	28	**GGTTCCGTGGGGAAAAAAGAGGCTATGG**	17
514	23	**GGATAGGAGCGATGCCGGGAAGG**	15
844	30	**GGTTGAAAACTGGATATTCAGGAACCCCGG**	18
1016	21	**GGCATGTCAGGTGGGACCTGG**	16
1058	28	**GGAGGCTGCGTTACCGTGATGGCACAGG**	7
1162	26	**GGCATCGATATCGGACATGGCTTCGG**	15
1274	20	**GGTTGGGGAAACGGTTGTGG**	21
1504	21	**GGGAGGCTTTGGAAGCTTAGG**	16
1638	22	**GGCATGCTGGGGCAGACACCGG**	13
1738	15	**GGGGAGCCAGGAAGG**	16
1795	19	**GGATGGTGCAAAGGGAAGG**	17
1948	27	**GGAGGTGCAGTATGCAGGGACAGATGG**	11
2111	29	**GGGGATTCTTACATTGTCATAGGAGTTGG**	4
2163	30	**GGCATAGGAGTGGTAGCACCATCGGAAAGG**	14
2239	30	**GGATACAGCCTGGGACTTCGGATCAGTCGG**	18

2509	23	**GGAAACGAGATGTGGCACGGGGGG**	11
2620	24	**GGCCTGGGAAGAGGGGATCTGTGG**	19
2816	18	**GGCTGGAAAGCCTGGGGGG**	15
2941	18	**GGAGGATCACGGGTTTGG**	17
3029	25	**GGAACAGCTGTTAAGGGAAGGGAGG**	11
3310	29	**GGAATGTCCAGGCACCAAGGTTTACGTGG**	19
3389	21	**GGAAGGGTCATTGAGGAATGG**	17
3458	19	**GGCTGCTGGTATGGAATGG**	19
3516	20	**GGTCAATGGTGACAGCGGGG**	14
3589	13	**GGTGCAGGAGGGG**	17
3646	30	**GGCAGTGCTGGTAGTCATGATCTTGGGAGG**	10
3743	28	**GGAGGAGATGTAGCTCACTTGGCATTGG**	7
3923	22	**GGATTTGCTTTGGCCTGGTTGG**	14
4036	20	**GGCATGGAGAGCGGGCCTGG**	19
4197	20	**GGAGTGGGAAGCGGAGCTGG**	21
4261	20	**GGCCGGAGGGTTTGCCAAGG**	15
4294	24	**GGCTGGACCCATGGCTGCAGTAGG**	15
4393	24	**GGAAAAGGACGCGGAAGTCACTGG**	17
4428	28	**GGCTTGACGTGGCACTGGATGAGAGTGG**	17
4510	24	**GGTGGTCCTGATGGCCATCTGTGG**	14
4593	23	**GGAAAAGGAGTGGCGCCCTCTGG**	16
4694	22	**GGTTCAACACAGGTTGGAGTGG**	14
4763	26	**GGAGCCGCACTGAGGAGCGGTGAGGG**	13
4806	24	**GGGGGGATGTCAAGCAGGACTTGG**	13
4841	27	**GGGCCTTGGAAGTTGGATGCAGCTTGG**	17
4957	15	**GGACGGGGACATCGG**	16
5030	30	**GGAAGAGTGATAGGACTCTATGGCAATGGG**	16

5102	25	**GGAAAGAGGGAGGAGGAGACTCCGG**	19
5296	26	**GGAGGAGGCCTTGAGAGGACTTCCGG**	17
5512	20	**GGTTGAAATGGGCGAGGCGG**	15
5967	20	**GGAGAGGACGTATAGGCAGG**	16
6002	27	**GGAGATGAGTACATGTATGGAGGTGGG**	6
6457	26	**GGCTTTGGGAGTAATGGAGGCCCTGG**	19
6560	25	**GGAAGCAGGCCTTATAAGGCAGCGG**	17
6626	27	**GGTTTGCTGGGAACAGTTTCACTGGGG**	9
6680	23	**GGCATCGGGAAGATGGGCTTTGG**	21
6718	29	**GGCCAGTGCATGGCTCATGTGGCTTTCGG**	17
6868	21	**GGTGGCAGTGGGCCTTCTAGG**	15
6952	29	**GGGAAGGAGAGAAGAAGGAGCAACCATGG**	15
7124	27	**GGAGTGCTGTTTGGCATGGGCAAAGGG**	15
7476	13	**GGGGGAGGCTGG**	20
7655	13	**GGAGGTGGGACGG**	20
7789	18	**GGATGGAGTGGCCACAGG**	18
7845	16	**GGTTGGAGGAGAGAGG**	17
7895	21	**GGATGTGGCAGAGGGGGCTGG**	20
7951	30	**GGTGAGAGGATACACAAAGGGAGGTCCCGG**	17
8039	25	**GGAGTGGACGTCTTCCACATGGCGG**	9
8155	13	**GGTGGGGGACTGG**	19
8256	13	**GGCATGGGGGAGG**	19
8386	16	**GGATGGCCCCAGGAGG**	17
8416	24	**GGAGGATGTGAACCTCGGCTCGGG**	11
8652	25	**GGGACGTGGTGACTGGAGTTACAGG**	19
8796	18	**GGCTGTGGAAGGAGCTGG**	19

8902	26	**GGAAAAAGAATGGAAGACGGCTGTGG**	16
8946	24	**GGTTTTGGGCCCTAGTGGATAGGG**	18
9081	30	**GGTACATGTGGTTGGGAGCCAGATTCTTGG**	12
9125	29	**GGATTCTTGAACGAGGACCATTGGATGGG**	12
9167	23	**GGAGGTGGAGTCGAAGGGTTAGG**	18
9328	29	**GGAGGAAGGGCACAGAACTCTGGCGTTGG**	14
9388	30	**GGTGAAGGTTCTCAGACCAGCTGAAGGAGG**	5
9505	26	**GGTGGTGCAGCTTATCCGGAACATGG**	10
9611	28	**GGATGGGATAGACTCAAACGAATGGCGG**	6
9713	30	**GGAAAAGTTAGGAAAGACACACAGGAGTGG**	13
9887	12	**GGGGCAGGATGG**	19
10044	26	**GGTCAATCCATGGAAAGGGAGAATGG**	16
10093	26	**GGTGTGGAATAGAGTGTGGATTGAGG**	14
10178	21	**GGAAAAAGGGAGGACTTATGG**	17
10334	23	**GGTGAGGAAGGGTCCACACCCGG**	15
10555	25	**GGAAGCGCAGGATGGGAAAAGAAGG**	17
10646	28	**GGACTAGTGGTTAGAGGAGACCCCCCGG**	17
10764	14	**GGCGGCCGGTGTGG**	19

PQS Sequences Identified with Quadparser

Source name	Location		Source Name + Location + PQS Length		Strand Polarity	Sequence
Zika virus	186	231	Zika virus_186_231_for	45	+	GGGAGGTTTGAAGAGGTTGCCAGCCGGACTTCTGCTGGGTCATGG
Zika virus	311	341	Zika virus_311_341_for	30	+	GGGGTTCCGTGGGGAAAAAAGAGGCTATGG
Zika virus	513	536	Zika virus_513_536_for	23	+	GGATAGGAGCGATGCCGGGAAGG
Zika virus	843	874	Zika virus_843_874_for	31	+	GGTTGAAAACTGGATATTCAGGAACCCCGGG
Zika virus	1011	1037	Zika virus_1011_1037_for	26	+	GGAGGGCATGTCAGGTGGGACCTGGG
Zika virus	1110	1142	Zika virus_1110_1142_for	32	+	GGTCACGACGACGGTTAGTAACATGGCCGAGG
Zika virus	1161	1187	Zika virus_1161_1187_for	26	+	GGCATCGATATCGGACATGGCTTCGG
Zika virus	1266	1316	Zika virus_1266_1316_for	50	+	GGACAGAGGTTGGGGAAACGGTTGTGGACTTTTTGGCAAAGGGAGCTTGG
Zika virus	1491	1524	Zika virus_1491_1524_for	33	+	GGAAGCAACCTTGGGAGGCTTTGGAAGCTTAGG
Zika virus	1764	1824	Zika virus_1764_1824_for	60	+	GGCTCTCGCTGGAGCTCTAGAGGCTGAGATGGATGGTGCAAAGGGAAGGCTGTTCTCTGG
Zika virus	2162	2201	Zika virus_2162_2201_for	39	+	GGCATAGGAGTGGTAGCACCATCGGAAAGGCATTTGAGG
Zika virus	2226	2271	Zika virus_2226_2271_for	45	+	GGCAGTCCTGGGGGATACAGCCTGGGACTTCGGATCAGTCGGGGG
Zika virus	2619	2644	Zika virus_2619_2644_for	25	+	GGCCTGGGAAGAGGGGATCTGTGGG
Zika virus	2815	2833	Zika virus_2815_2833_for	18	+	GGCTGGAAAGCCTGGGGG
Zika virus	2940	2958	Zika virus_2940_2958_for	18	+	GGAGGATCACGGGTTTGG
Zika virus	3302	3360	Zika virus_3302_3360_for	58	+	GGTTTGAGGAATGTCCAGGCACCAAGGTTTACGTGGAGGAGACATGCGGAACTAGAGG

Zika virus 3388 3419 Zika virus_3388_3419_for 31 +
GGAAGGGTCATTGAGGAATGGTGCTGTAGGG

Zika virus 3457 3490 Zika virus_3457_3490_for 33 +
GGCTGCTGGTATGGAATGGAGATAAGGCCCAGG

Zika virus 3922 3944 Zika virus_3922_3944_for 22 +
GGATTTGCTTTGGCCTGGTTGG

Zika virus 4035 4069 Zika virus_4035_4069_for 34 +
GGCATGGAGAGCGGGCCTGGCTACTTGTGGAGGG

Zika virus 4196 4216 Zika virus_4196_4216_for 20 +
GGAGTGGGAAGCGGAGCTGG

Zika virus 4260 4280 Zika virus_4260_4280_for 20 +
GGCCGGAGGGTTTGCCAAGG

Zika virus 4293 4317 Zika virus_4293_4317_for 24 +
GGCTGGACCCATGGCTGCAGTAGG

Zika virus 4391 4416 Zika virus_4391_4416_for 25 +
GGGAAAAGGACGCGGAAGTCACTGG

Zika virus 4427 4455 Zika virus_4427_4455_for 28 +
GGCTTGACGTGGCACTGGATGAGAGTGG

Zika virus 4509 4533 Zika virus_4509_4533_for 24 +
GGTGGTCCTGATGGCCATCTGTGG

Zika virus 4591 4616 Zika virus_4591_4616_for 25 +
GGGAAAAGGAGTGGCGCCCTCTGGG

Zika virus 4693 4716 Zika virus_4693_4716_for 23 +
GGTTCAACACAGGTTGGAGTGGG

Zika virus 4805 4829 Zika virus_4805_4829_for 24 +
GGGGGGATGTCAAGCAGGACTTGG

Zika virus 4840 4895 Zika virus_4840_4895_for 55 +
GGGCCTTGGAAGTTGGATGCAGCTTGGGATGGACTCAGCGAGGTACAGCTTTTGG

Zika virus 5029 5060 Zika virus_5029_5060_for 31 +
GGAAGAGTGATAGGACTCTATGGCAATGGGG

Zika virus 5100 5126 Zika virus_5100_5126_for 26 +
GGGAAAGAGGGAGGAGGAGACTCCGG

Zika virus 5295 5321 Zika virus_5295_5321_for 26 +
GGAGGAGGCCTTGAGAGGACTTCCGG

Zika virus 5510 5531 Zika virus_5510_5531_for 21 +
GGGTTGAAATGGGCGAGGCGG

Zika virus 5966 5986 Zika virus_5966_5986_for 20 + GGAGAGGACGTATAGGCAGG

Zika virus 6442 6495 Zika virus_6442_6495_for 53 + GGAAAAAGAGGAGCGGCTTTGGGAGTAATGGAGGCCCTGGGAACACTGCCAGG

Zika virus 6559 6584 Zika virus_6559_6584_for 25 + GGAAGCAGGCCTTATAAGGCAGCGG

Zika virus 6671 6746 Zika virus_6671_6746_for 75 + GGAATAAGGGCATCGGGAAGATGGGCTTTGGAATGGTAACCCTTGGGGCCAGTGCATGGCT CATGTGGCTTTCGG

Zika virus 6867 6888 Zika virus_6867_6888_for 21 + GGTGGCAGTGGGCCTTCTAGG

Zika virus 6951 6992 Zika virus_6951_6992_for 41 + GGGAAGGAGAGAAGAAGGAGCAACCATGGGATTCTCAATGG

Zika virus 7123 7150 Zika virus_7123_7150_for 27 + GGAGTGCTGTTTGGCATGGGCAAAGGG

Zika virus 7460 7488 Zika virus_7460_7488_for 28 + GGACCGCCTGGGGATGGGGGGAGGCTGG

Zika virus 7654 7690 Zika virus_7654_7690_for 36 + GGAGGTGGGACGGGAGAGACTCTGGGAGAGAAGTGG

Zika virus 7788 7809 Zika virus_7788_7809_for 21 + GGATGGAGTGGCCACAGGAGG

Zika virus 7844 7860 Zika virus_7844_7860_for 16 + GGTTGGAGGAGAGAGG

Zika virus 7894 7915 Zika virus_7894_7915_for 21 + GGATGTGGCAGAGGGGGCTGG

Zika virus 7947 7980 Zika virus_7947_7980_for 33 + GGAGGTGAGAGGATACACAAAGGGAGGTCCCGG

Zika virus 8154 8167 Zika virus_8154_8167_for 13 + GGTGGGGGACTGG

Zika virus 8255 8268 Zika virus_8255_8268_for 13 + GGCATGGGGGAGG

Zika virus 8376 8401 Zika virus_8376_8401_for 25 + GGGACGCATGGATGGCCCCAGGAGG

Zika virus 8651 8676 Zika virus_8651_8676_for 25 + GGGACGTGGTGACTGGAGTTACAGG

Zika virus 8795 8827 Zika virus_8795_8827_for 32 + GGCTGTGGAAGGAGCTGGGGAAACGCAAGCGG

Zika virus	8901	8927	Zika virus_8901_8927_for	26	+
GGAAAAAGAATGGAAGACGGCTGTGG					
Zika virus	8945	8969	Zika virus_8945_8969_for	24	+
GGTTTTGGGCCCTAGTGGATAGGG					
Zika virus	9166	9189	Zika virus_9166_9189_for	23	+
GGAGGTGGAGTCGAAGGGTTAGG					
Zika virus	9886	9908	Zika virus_9886_9908_for	22	+
GGGGCAGGATGGAGCATCCGGG					
Zika virus	10019	10069	Zika virus_10019_10069_for	50	+
GGGTACCAACTGGGAGAACCACCTGGTCAATCCATGGAAAGGGAGAATGG					
Zika virus	10080	10118	Zika virus_10080_10118_for	38	+
GGACATGCTCATGGTGTGGAATAGAGTGTGGATTGAGG					
Zika virus	10177	10216	Zika virus_10177_10216_for	39	+
GGAAAAAGGGAGGACTTATGGTGTGGATCCCTTATAGGG					
Zika virus	10332	10356	Zika virus_10332_10356_for	24	+
GGGTGAGGAAGGGTCCACACCCGG					
Zika virus	10554	10582	Zika virus_10554_10582_for	28	+
GGAAGCGCAGGATGGGAAAAGAAGGTGG					
Zika virus	10644	10673	Zika virus_10644_10673_for	29	+
GGGACTAGTGGTTAGAGGAGACCCCCCGG					
Zika virus	10763	10789	Zika virus_10763_10789_for	26	+
GGCGGCCGGTGTGGGGAAATCCATGG					

PQS Sequences Identified with G4Hunter

Sequences are written 5` to 3`

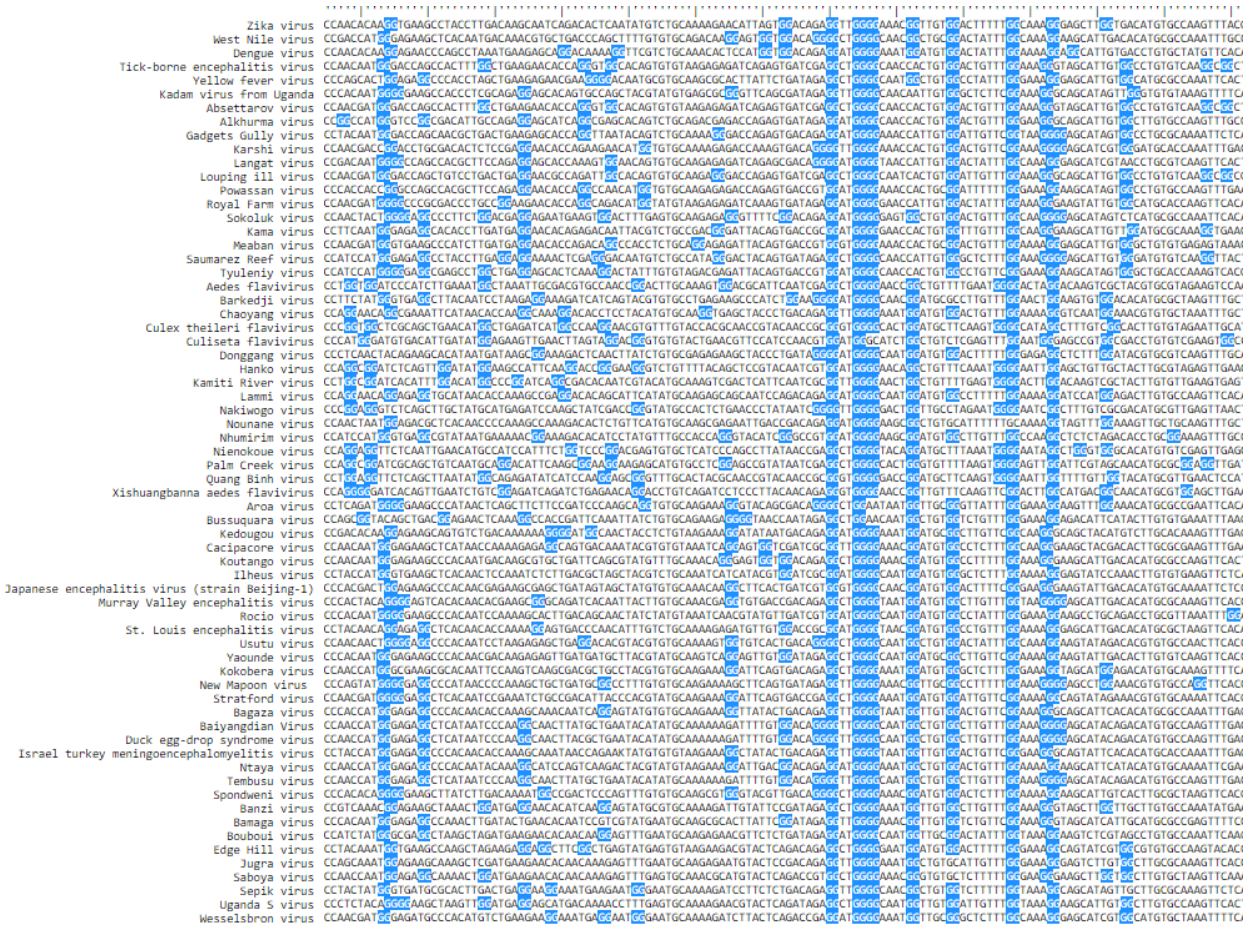
TTGGGAGGTTTGAAGAGG
TATCAACAGATGGGGTTCCGTGGGGAAAAAAGAGGCT
TGCTGGATGAGGGAGTGGAA
GGAGGGCATGTCAGGTGGGACCTGGGTTGAT
GTGGACAGAGGTTGGGGAAACGGTTGTGGA
TTGGGAGGCTTTGGAAG
CTTGGCATGCTGGGGCAGACACCGGA
GTCGTCGTTCTGGGGAGCCAGGAAGGA
GGATGGTGCAAAGGGAAGG
ATTTGGGGATTCTTA
GTCATAGGAGTTGGGGACAAGAAAATC
ATGGCAGTCCTGGGGGATACAGCCTGGG
TCGGATCAGTCGGGGGTGTGTTCAACTCACTGGGTAAGGGCATTCA
GCTTGGCCCTGGGGGGAGTGATGATCT
TCTGCTGACGTGGGGTGCTCAGTGGAC
GAGATGTGGCACGGGGGTATTCATCTA
AGGCCTGGGAAGAGGGGATCTGTGGGATCT
AATCAGTAGAAGGGGAGCTCAATGCT
CTGGAAAGCCTGGGGGAAATCGTATTT
GTGGAGGATCACGGGTTTGG
GCTGTTAAGGGAAGGGAGGCCGCG
TGGAGGCTGAAGAGGG
GGAAGGGTCATTGAGG
AATGGTGACAGCGGGGTCAA
TCATGGTGCAGGAGGGGTTGAAGAAGAGA
GGCATGGAGAGCGGG
GAAAGGGAAAGGTAGTGTG
AAGGAGTGGGAAGCGGAGCTGG
CGGGAAAGAGTGTGGA
GGGAAAAGGACGCGG
CTGGGAAAAGGAGTGGCG
ACAGGTTGGAGTGGGAGTCATG
TGAGGAGCGGTGAGGGAAGA
GATCCATACTGGGGGGATGTCAAGCAGG
AAGACAAAGGACGGGGACATCGGAGCAG
TCTATGGCAATGGGGTTGTGATCAAG
CAGGGAAAGAGGGAGGAGGAGACTC
CAAGGGTTGAAATGGGCGAGGCGGCTG
AGGCTGGAAAGCGGGT
ATGTATGGAGGTGGGTGTGCAGAG

GGATGGATGCTAGGGT
AGGAGCGGCTTTGGGAGTAATGGAGGC
AGTTTCACTGGGGAT
GAATAAGGGCATCGGGAAGATGGGCTTTGGAATGGT
TGGGGCCAGTGCATGG
CATGGTGGCAGTGGGC
TAATGGGAAGGAGAGAAGA
TGGCATGGGCAAAGGGATGC
ATTTATGCATGGGGACCTTGGAGT
GGAGAAGAAGATGGG
GACCGCCTGGGGATGGGGGGAGGCTGGAGCTC
GAGACGTGGAGGTGGGACGGGAGAGACTCTGGGAGAGAAGTGGAA
CGGGGAAGTGCAAAGA
GGTTGGAGGAGAGAGG
ATGGGAAGGTTGTTGA
CGGATGTGGCAGAGGGGGCTGGAGCTATTA
CTCTCTATGGTGGGGGACTGGCTTGAAA
AACGTAGGCATGGGGGAGGATTAGTCAG
GTACTGGGTCTCTGGGGCAAAGAGCAAC
GGGACGCATGGATGG
TCGTGAACGGGGTTGTTAGACTC
GGGACGTGGTGACTG
TGTGGAAGGAGCTGGGGAAACGCAAGCGG
TAGTGGATAGGGAGAGAGA
GTACATGTGGTTGGG
ATTGGATGGGAAGAGAAA
AGGTGGAGTCGAAGGGTTAGGATTG
CAAATGGAGGAAGGGCACAG
AGAGAGGGAGTGGACAAG
GATGGAGCAATTGGG
ACCAGGGGCAGGATGGAGCAT
TGGAAAGGGAGAATGGATGA
TAGGAAAAAGGGAGGACTTATGGT
ACTTGGGTGAGGAAGGGTCC
ACAGTTTGGGGAAAGCTGTGCA
GAAGCGCAGGATGGGAAAAGAAGGTGG
TGGGGCCTGAACTGGA
CGGCGGCCGGTGTGGGGAAATCCATGGT

Figure S2. Complete sequence alignments from the *Flaviviridae* family for the PQS in the prM, E, NS1-A, NS1-B, NS3, NS5-A, and NS5-B gene regions

Each GG site is labeled blue.

Regions with the conserved prM sequence.

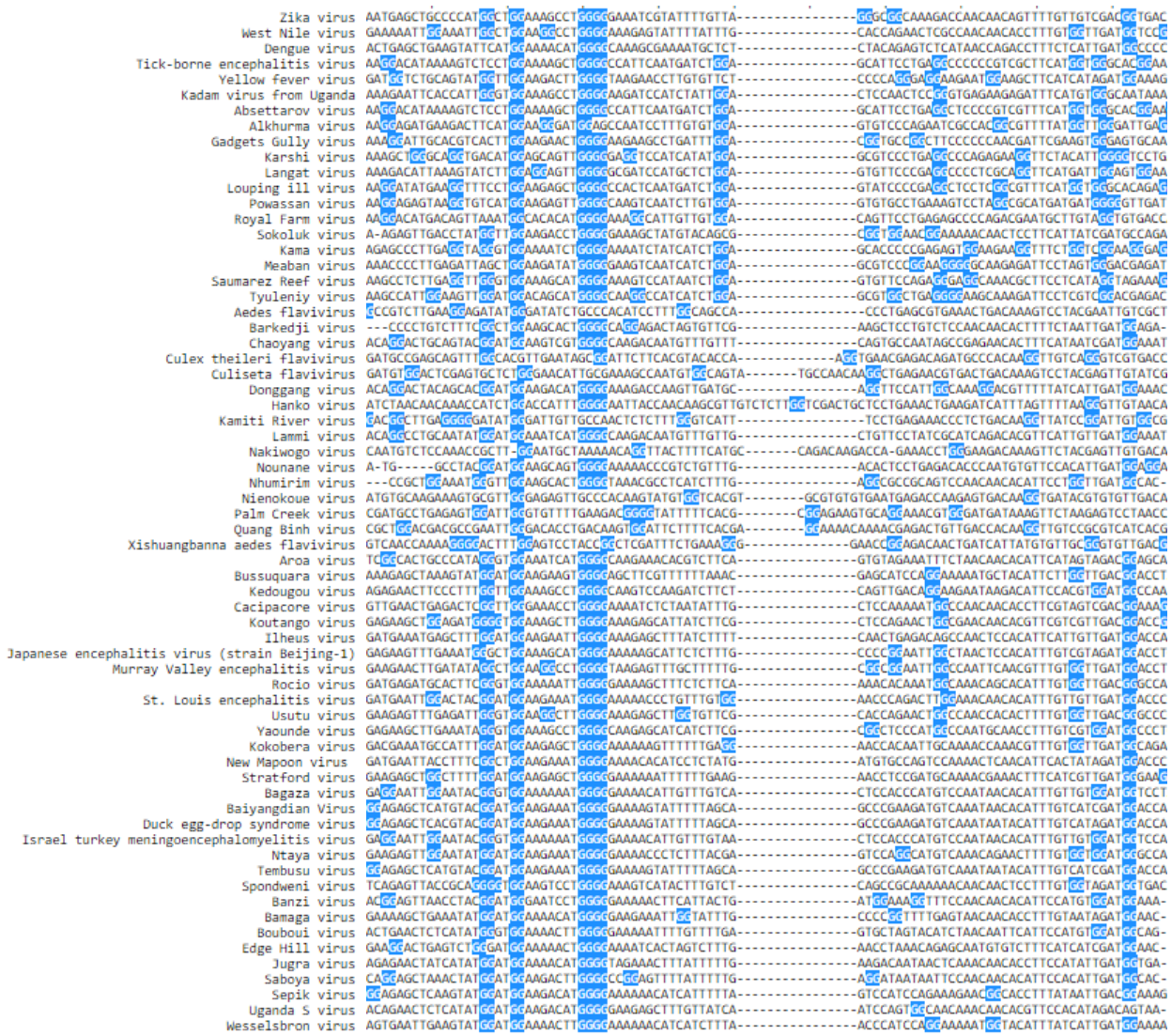


PQS site above

Regions with the conserved E sequence

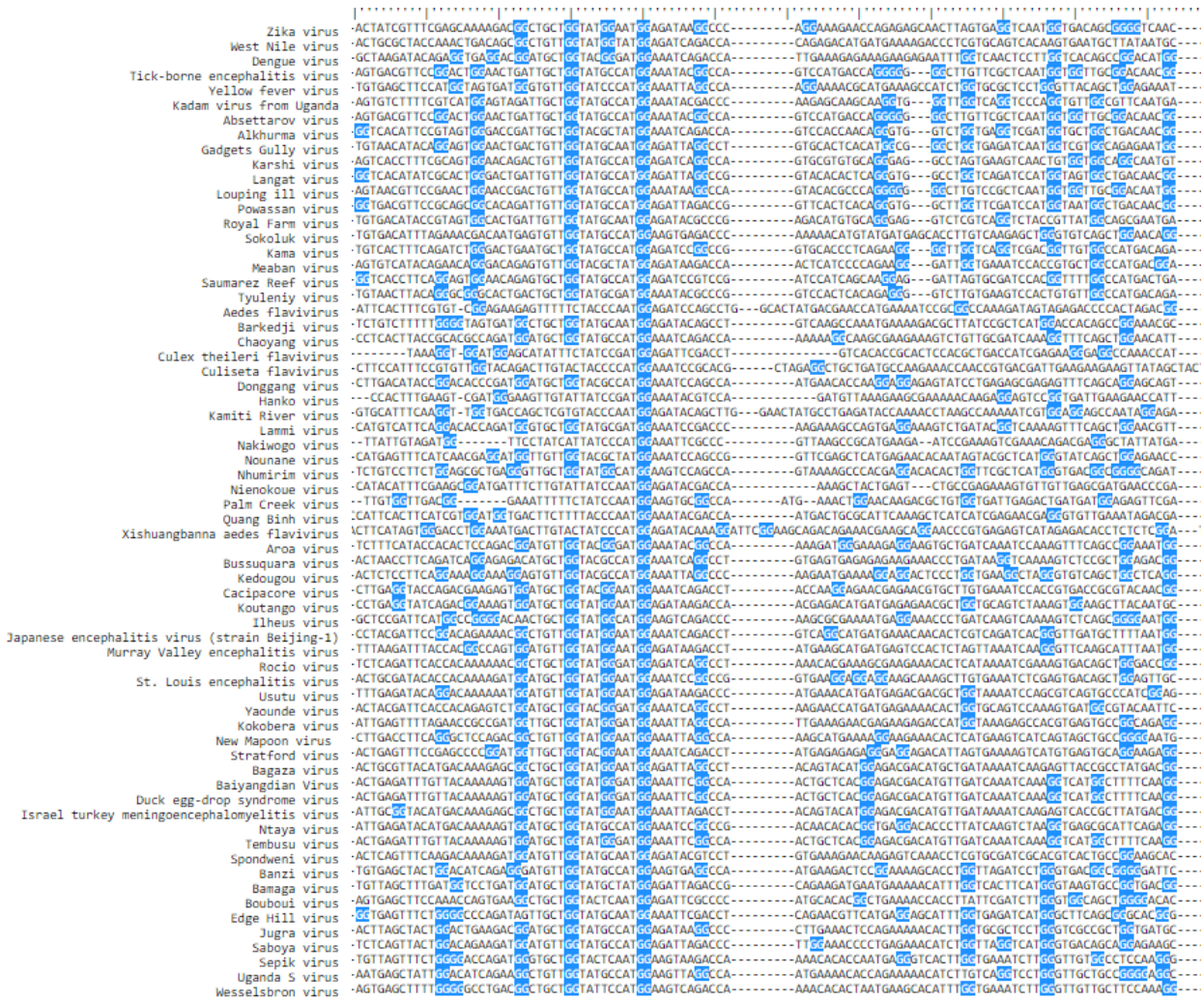
PQS Sequence Above

Regions with the conserved NS1-A sequence



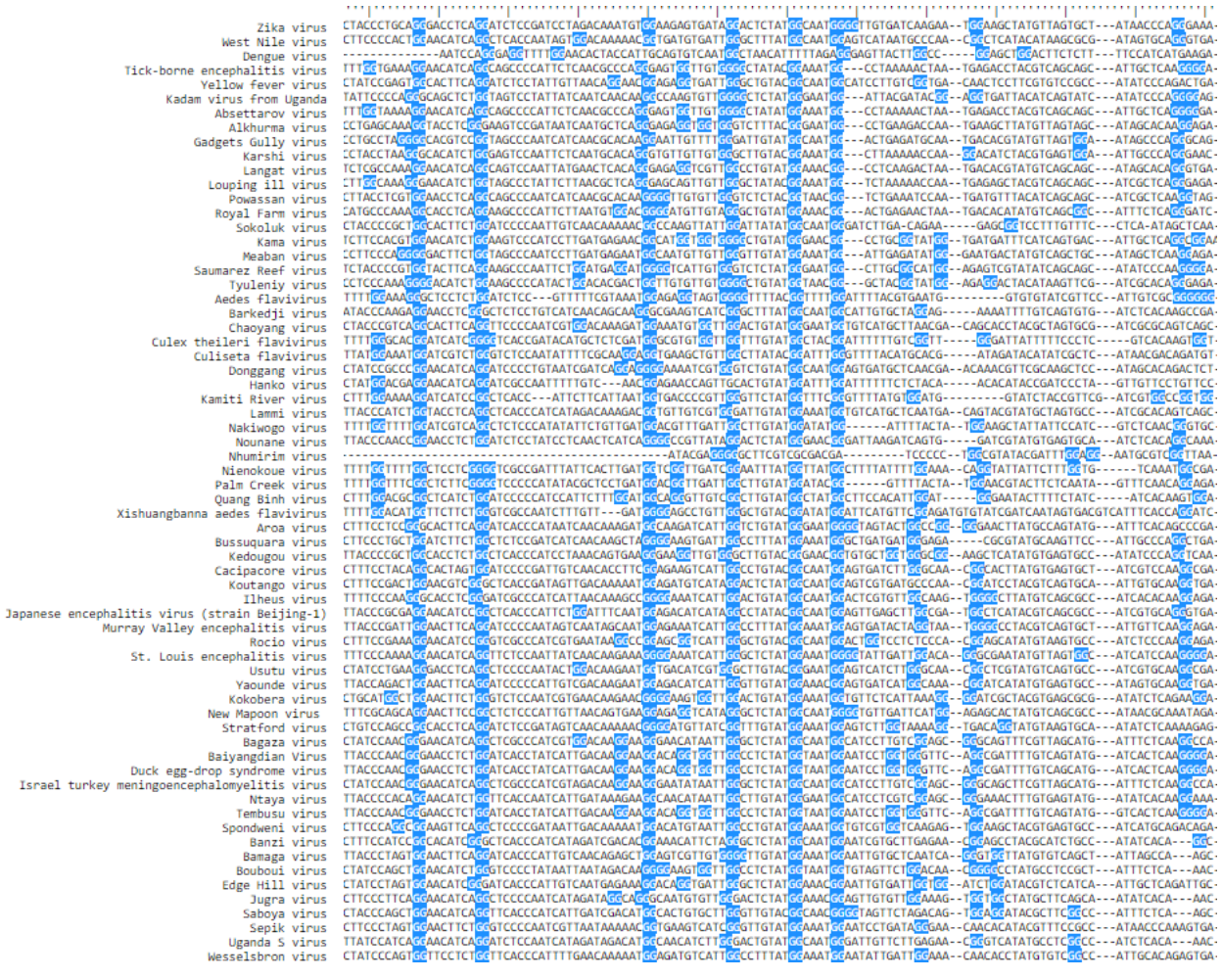
PQS Site

Regions with the conserved NS1-B sequence



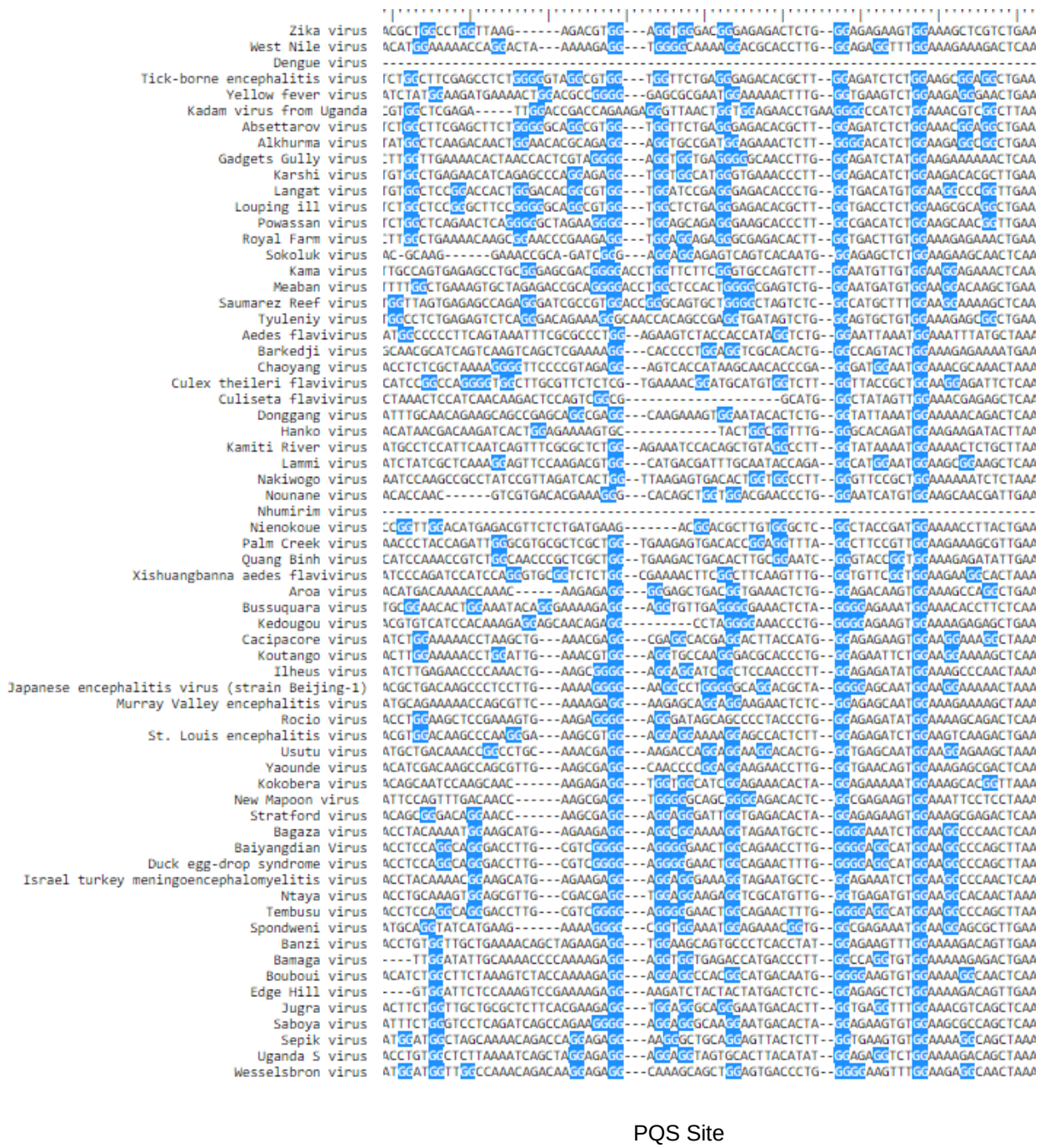
PQS Site

Regions with the conserved NS3 sequence



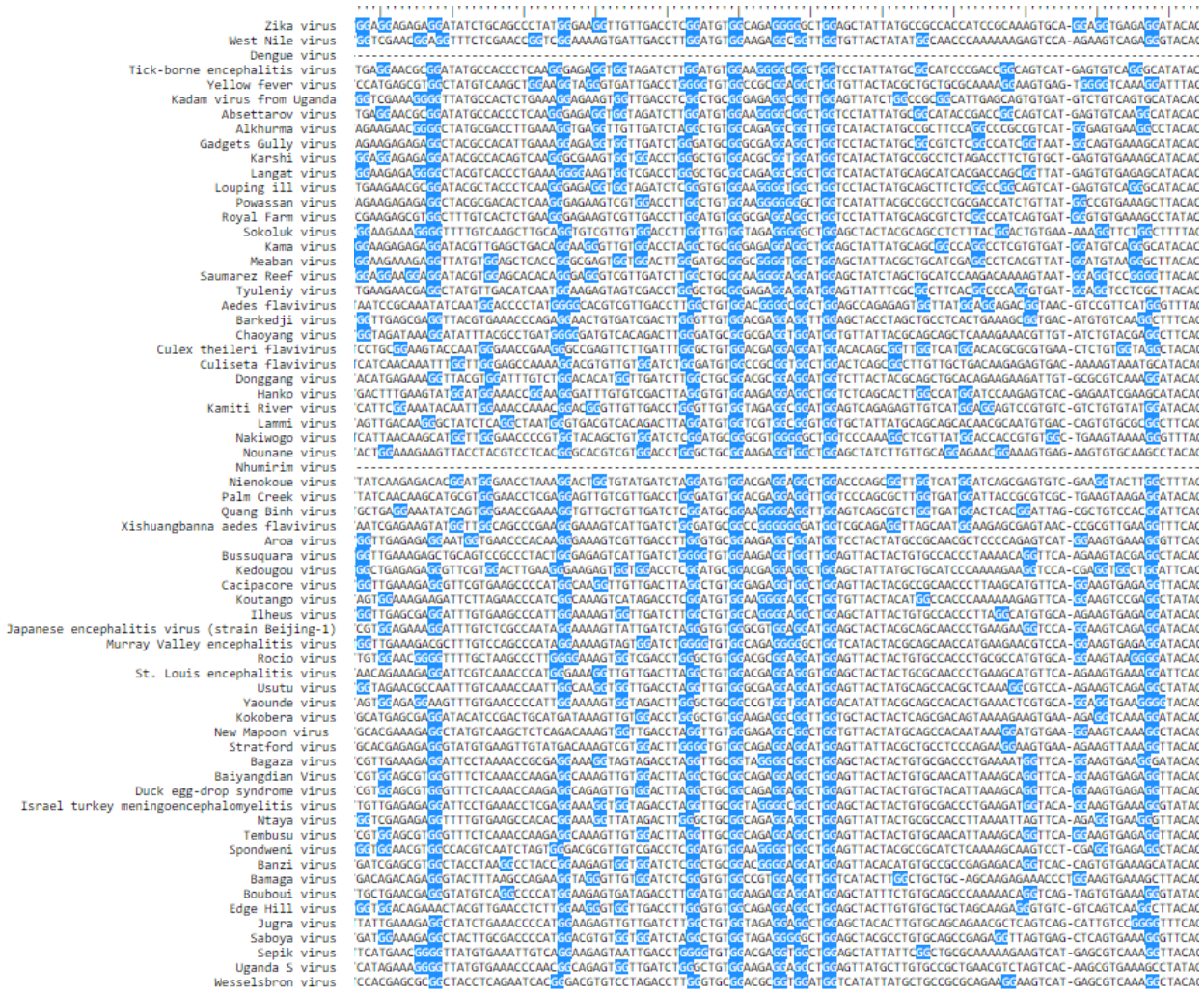
PQS Site

Regions with the conserved NS5-A sequence



PQS Site

Regions with the conserved NS5-B sequence



PQS Site

Figure S3. Alignment statistics

	Fraction Similarity with Zika	Fraction Difference with Zika
Zika_virus	1.00	0.00
West_Nile_virus	0.62	0.38
Dengue_virus	0.571	0.429
Tick-borne_encephalitis_virus	0.526	0.474
Yellow_fever_virus	0.565	0.435
Kadam_virus_from_Uganda	0.528	0.472
Absettarov_virus	0.525	0.475
Alkhurma_virus	0.519	0.481
Gadgets_Gully_virus	0.531	0.469
Karshi_virus	0.528	0.472
Langat_virus	0.529	0.471
Louping_ill_virus	0.529	0.471
Powassan_virus	0.529	0.471
Royal_Farm_virus	0.537	0.463
Sokoluk_virus	0.545	0.455
Kama_virus	0.532	0.468
Meaban_virus	0.534	0.466
Saumarez_Reef_virus	0.532	0.468
Tyuleniy_virus	0.527	0.473
Aedes_flavivirus	0.423	0.577
Barkedji_virus	0.582	0.418
Chaoyang_virus	0.579	0.421
Culex_theileri_flavivirus	0.432	0.568
Culiseta_flavivirus	0.431	0.569
Donggang_virus	0.583	0.417
Hanko_virus	0.448	0.552
Kamiti_River_virus	0.427	0.573
Lammi_virus	0.581	0.419
Nakiwogo_virus	0.428	0.572
Nounane_virus	0.57	0.43
Nhumirim_virus	0.557	0.443
Nienokoue_virus	0.441	0.559
Palm_Creek_virus	0.43	0.57
Quang_Binh_virus	0.433	0.567
Xishuangbanna_aedes_flavivirus	0.432	0.568
Aroa_virus	0.613	0.387

Bussuquara_virus	0.618	0.382
Kedougou_virus	0.621	0.379
Cacipacore_virus	0.62	0.38
Koutango_virus	0.621	0.379
Ilheus_virus	0.626	0.374
Japanese_encephalitis_virus_(strain_Beijing-1)	0.609	0.391
Murray_Valley_encephalitis_virus	0.615	0.385
Rocio_virus	0.62	0.38
St._Louis_encephalitis_virus	0.629	0.371
Usutu_virus	0.618	0.382
Yaounde_virus	0.623	0.377
Kokobera_virus	0.617	0.383
New_Mapoon_virus	0.606	0.394
Stratford_virus	0.613	0.387
Bagaza_virus	0.611	0.389
Baiyangdian_Virus	0.615	0.385
Duck_egg-drop_syndrome_virus	0.614	0.386
Israel_turkey_meningoencephalomyelitis_virus	0.614	0.386
Ntaya_virus	0.613	0.387
Tembusu_virus	0.614	0.386
Spondweni_virus	0.703	0.297
Banzi_virus	0.559	0.441
Bamaga_virus	0.553	0.447
Bouboui_virus	0.564	0.436
Edge_Hill_virus	0.554	0.446
Jugra_virus	0.561	0.439
Saboya_virus	0.567	0.433
Sepik_virus	0.563	0.437
Uganda_S_virus	0.559	0.441
Wesselsbron_virus	0.566	0.434

Figure S4. Complete sequence alignment for the Zika specific PQS in the 3`-UTR

The PQS site specific to the Zika virus in the 3`-UTR (G runs are highlighted in yellow)

```
gi|1008913395|gb|KU955593.1| Zika virus isolate Zika virus/H.sapiens-tc/KHM/2010/FSS13025, complete genome   :ATGAGTTTCCACCACGCTGGCCGCCAGGCACAGATCGCCGAATAGCGGCGGCCGGTGTGGGGAAATCCATGGGTCT-
gi|1000381325|gb|KU681082.3| Zika virus isolate Zika virus/H.sapiens-tc/PHL/2012/CPC-0740, complete genome   :ATGAGTTTCCACCACGCTGGCCGCCAGGCACAGATCGCCGAATAGCGGCGGCCGGTGTGGGGAAATCCATGGGTCT-
gi|1000381324|gb|KU681081.3| Zika virus isolate Zika virus/H.sapiens-tc/THA/2014/SV0127- 14, complete genome   :ATGAGTTTCCACCACGCTGGCCGCCAGGCACAGATCGCCGAATAGCGGCGGCCGGTGTGGGGAAATCCATGGGTCT-
gi|1031621529|gb|KX262887.1| Zika virus isolate 103451, complete genome   :ATGAGTTTCCACCACGCTGGCCGCCAGGCACAGATCGCCGAATAGCGGCGGCCGGTGTGGGGAAATCCATGGGTCT-
gi|1029032430|gb|KX253996.1| Zika virus isolate ZKC2/2016, complete genome   :ATGAGTTTCCACCACGCTGGCCGCCAGGCACAGATCGCCGAATAGCGGCGGCCGGTGTGGGGAAATCCATGGGTCT-
gi|1028727781|gb|KX247646.1| Zika virus isolate Zika virus/Homo sapiens/COL/UF-1/2016, complete genome   :ATGAGTTTCCACCACGCTGGCCGCCAGGCACAGATCGCCGAATAGCGGCGGCCGGTGTGGGGAAATCCATGGGTCTT
gi|1027485270|gb|KX087101.2| Zika virus strain ZIKV/Homo sapiens/PRI/PRVABC59/2015, complete genome   :ATGAGTTTCCACCACGCTGGCCGCCAGGCACAGATCGCCGAATAGCGGCGGCCGGTGTGGGGAAATCC---------
gi|1026288143|gb|KX198135.1| Zika virus strain ZIKV/Homo sapiens/PAN/BEI-259634_V4/2016, complete genome   :ATGAGTT-----------------------------------------------------------------------
gi|1026288139|gb|KX197192.1| Zika virus isolate ZIKV/H.sapiens/Brazil/PE243/2015, complete genome   :ATGAGTTTCCACCACGCTGGCCGCCAGGCACAGATCGCCGAATAGCGGCGGCCGGTGTGGGGAAATCCATGGGTCT-
gi|1024286109|gb|KX156776.1| Zika virus strain ZIKV/Homo sapiens/PAN/CDC-259364_V1-V2/2015, complete genome   ----------
gi|1024286104|gb|KX156775.1| Zika virus strain ZIKV/Homo sapiens/PAN/CDC-259249_V1-V3/2015, complete genome   ----------
gi|1024286099|gb|KX156774.1| Zika virus strain ZIKV/Homo sapiens/PAN/CDC-259359_V1-V3/2015, complete genome   ----------
gi|998961613|gb|KU744693.1| Zika virus isolate VE_Ganxian, complete genome   ----------
gi|1021395933|gb|KX117076.1| Zika virus isolate Zhejiang04, complete genome   :ATGAGTTTCCACCACGCTGGCCGCCAGGCACAGATCGCCGAATAGCGGCGGCCGGTGTGGGGAAATCCATGGGTCTT
gi|1016563098|gb|KU509998.3| Zika virus strain Haiti/1225/2014, complete genome   :ATGAGTTTCCACCACGCTGGCCGCCAGGCACAGATCGCCGAATAGCGGCGGCCGGTGTGGGGAAATCCATGGGTCTT
gi|1020267437|gb|KX087102.1| Zika virus strain ZIKV/Homo sapiens/COL/FLR/2015, complete genome   ----------
gi|1016884664|gb|KX051563.1| Zika virus isolate Haiti/1/2016, complete genome   :ATGAGTTTCCACCACGCTGGCCGCCAGGCACAGATCGCCGAATAGCGGCGGCCGGTGTGGGGAAATCCATGGGTCTT
gi|969945756|gb|KU321639.1| Zika virus strain ZikaSPH2015, complete genome   ----------
gi|1006593136|gb|KU870645.1| Zika virus isolate FB-GWUH-2016, complete genome   :ATGAGTTTCCACCACGCTGGCCGCCAGGCACAGATCGCCGAATAGCGGCGGCCGGTGTGGGGAAATCCATGCAGC--
gi|1006593054|gb|KU926310.1| Zika virus isolate Rio-S1, complete genome   :ATGAGTTTCCACCACGCTGGCCGCCAGGCACAGATCGCCGAATAGCGGCGGCCGGTGTGGGGAAATCCATGGGCC--
gi|1006593051|gb|KU926309.1| Zika virus isolate Rio-U1, complete genome   :ATGAGTTTCCACCACGCTGGCCGCCAGGCACAGATCGCCGAATAGCGGCGGCCGGTGTGGGGAAATCCATG------
gi|1005883181|gb|KU922960.1| Zika virus isolate MEX/InDRE/Sm/2016, complete genome   ----------
gi|1005883179|gb|KU922923.1| Zika virus isolate MEX/InDRE/Lm/2016, complete genome   ----------
gi|1001904408|gb|KU820898.1| Zika virus isolate GZ01 polyprotein gene, complete cds   ----------
gi|1005500895|gb|KU740184.2| Zika virus isolate GD01 polyprotein gene, complete cds   ----------
gi|1003109675|gb|KU820899.2| Zika virus isolate ZJ03, complete genome   :ATGAGTTTCCACCACGCTGGCCGCCAGGCACAGATCGCCGAATAGCGGCGGCCGGTGTGGGGAAATCCATGGGTCTT
gi|1002634691|gb|KU853013.1| Zika virus isolate Dominican Republic/2016/PD2, complete genome   ----------
gi|1002634689|gb|KU853012.1| Zika virus isolate Dominican Republic/2016/PD1, complete genome   ----------
gi|985578255|gb|KU497555.1| Zika virus isolate Brazil-ZKV2015, complete genome   :ATGAGTTTCCACCACGCTGGCCGCCAGGCACAGATCGCCGAATAGCGGCGGCCGGTGTGGGGAAATCCA--------
gi|992324757|gb|KU707826.1| Zika virus isolate SSABR1, complete genome   ----------
gi|982894594|gb|KU527068.1| Zika virus strain Natal RGN, complete genome   :ATGAGTTTCCACCACGCTGGCCGCCAGGCACAGATCGCCGAATAGCGGCGGCCGGTGTGGGGAAATCCATGGGTCTT
gi|226377833|ref|NC_012532.1| Zika virus, complete genome   :ATGAGTTTCCACCACGCTGGCCGCCAGGCACAGATCGCCGAACTTCGGCGGCCGGTGTGGGGAAATCCATGGTTTCT
gi|984874581|gb|KU501215.1| Zika virus strain PRVABC59, complete genome   ----------
```

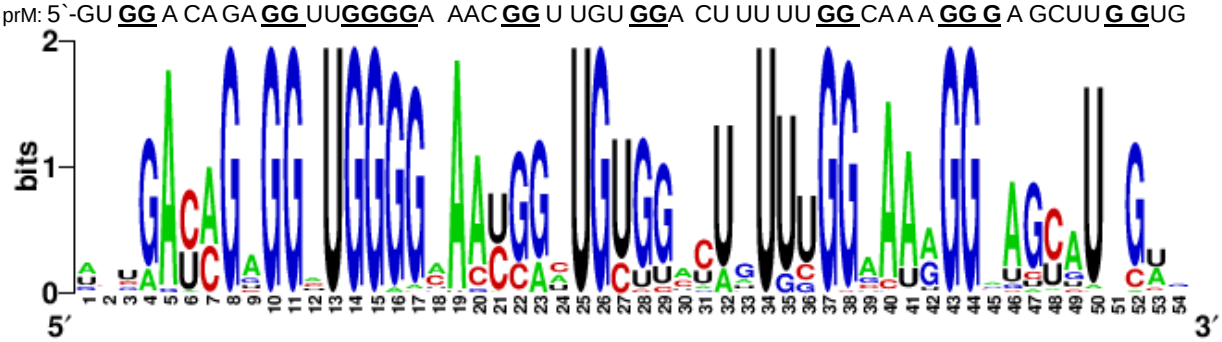
PQS Site

Figure S5. Sequence logos generated from the alignment data

prM: Family Consensus RNA Sequence

Zika Virus Sequence

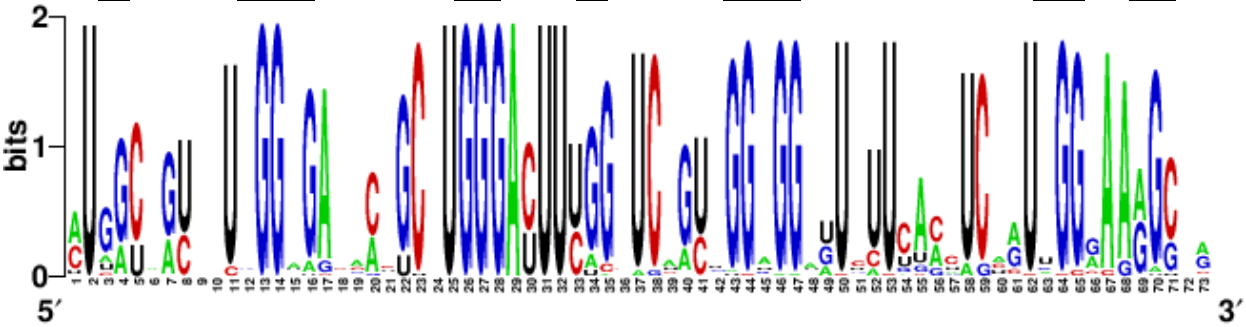
prM: 5`-GU **GG** A CA GA **GG** UU**GGGG**A AAC **GG** U UGU **GG**A CU UU UU **GG** CA A A **GG G** A GCUU **G G**UG



E: Family Consensus RNA Sequence

Zika Virus Sequence

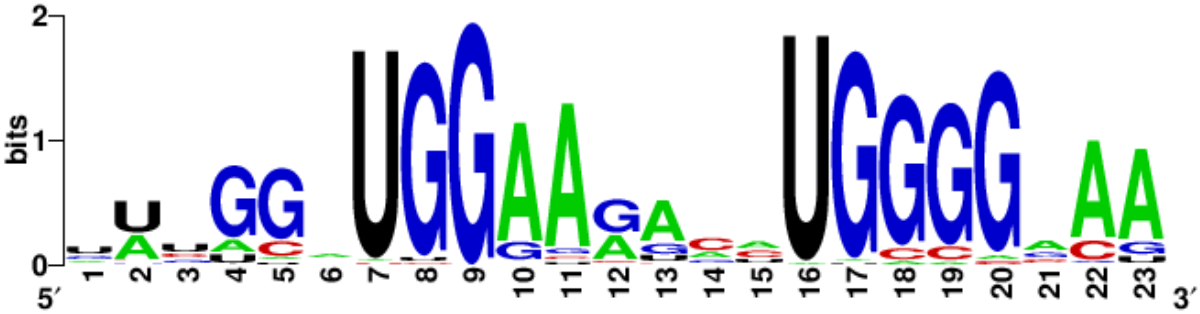
E: 5`-AU**GG**CAGUCCU**GGGGG**AUACAGCCU **GGG**ACUUC**GG**AUCAGUC **GGGGG**UGUGUUCAACUCACU **GGG** UAA**GGG**CA



NS1-A: Family Consensus RNA Sequence

Zika Virus Sequence NS1 1st:

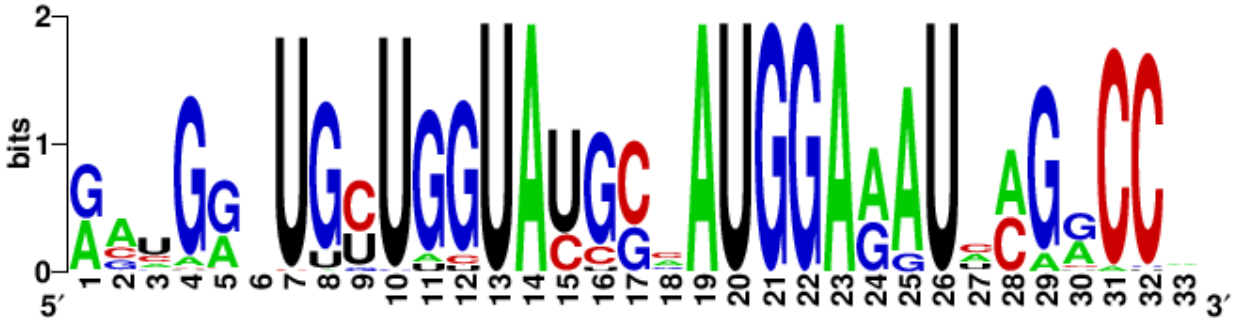
5`- C AU **G G** C U **G G** A A A G C C U **G G G G G** A A



NS1-B: Family Consensus RNA Sequence

Zika Virus NS1 2nd:

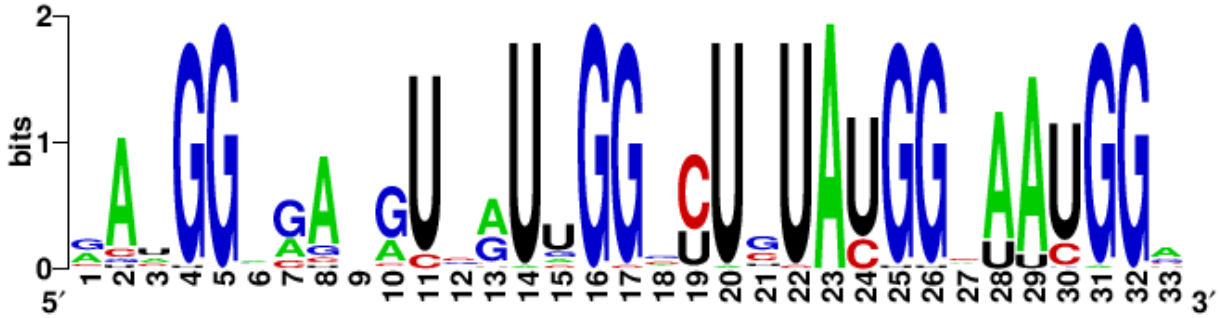
5`-G AC **G G** C U G C U **GG** U A U **G G** A A U **G G** A G AU A A **G G** C CCA



NS3: Family Consensus RNA Sequence

Zika Virus Sequence NS3:

5`- GU **G G**A A G A G U G A U A **G G** A C U C U A U **G G** C A A U **G G GG**

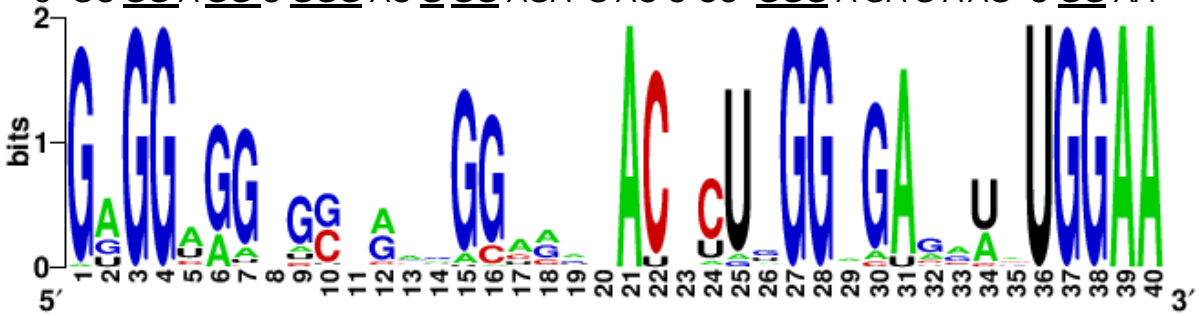


NS5-A: Family Consensus RNA Sequence

Zika Virus Sequence

NS5 1st:

5`-GU **GG** A **GG** U **GGG** AC **G** **GG** AGA G AC U CU **GGG** A GA G A AG U **GG** AA



NS5-B: Family Consensus RNA Sequence

Zika Virus Sequence NS5 2nd:

5`-**GGG** AA **GG**U U GU U GA C CU C**GG** A UG U **GG**C A GA **GG G GG**C U **GG** AG

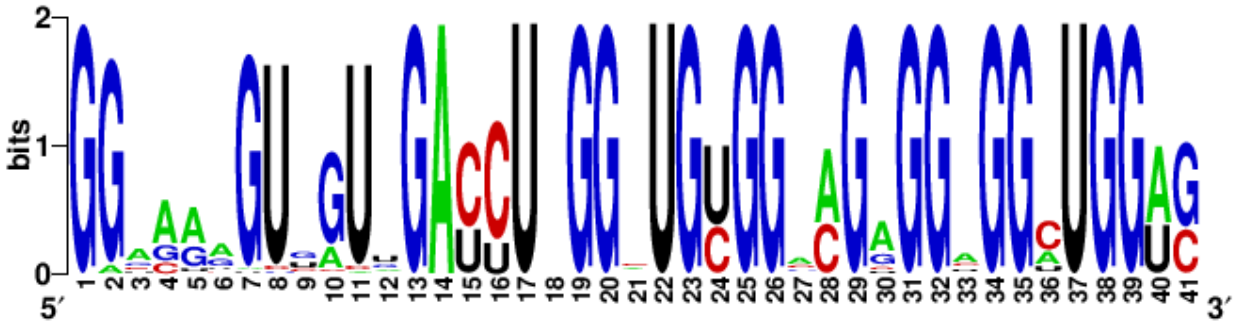
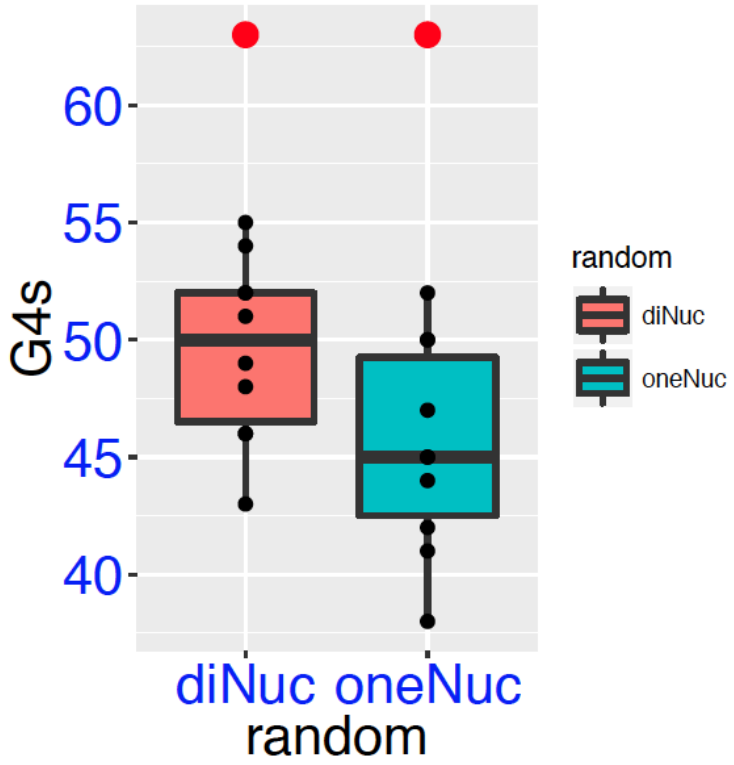


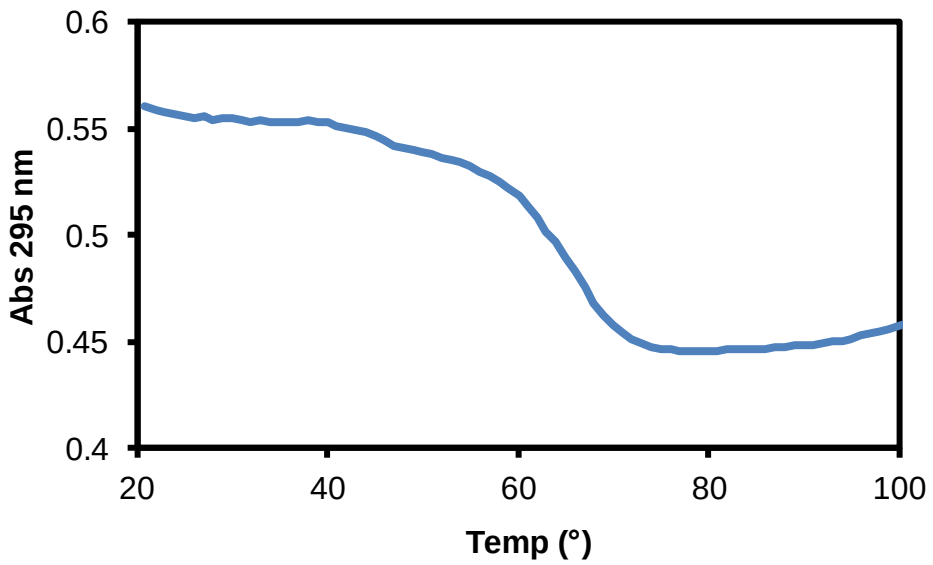
Figure S6. Data for the randomized genomes with the same nucleotide frequency as the Zika virus



The randomized sequences were generated using the GenRGenS package running a Markovian model.[6] Two types of random sequences were generated based on the Zika genome sequence. The first type of random sequences (oneNuc) has the same single nucleotide content as Zika, while the second type of random sequences (diNuc) has the same dinucleotide content as Zika. Ten random sequences were generated for each type. Quadparser[1] was utilized to search for PQSs in Zika viral and randomized genome sequences to search for PQSs with >2 Gs per tetrad and up to 10 nucleotides loops. The number of PQSs found in the random sequences was plotted as a box plot. The black dots represent the number of observed PQSs in the random genomes and the red dot represents the number of PQSs found in the Zika viral genome. The *P* values for the number of PQSs in the Zika viral genome when compared to the randomized genomes were oneNuc = 5.3×10^{-4} and diNuc = 7.0×10^{-5}. The *P* values were calculated using the pnorm function in R.

Figure S7. Representative T_m curves

T_m curve for NS5-A



T_m curves for NS1-B

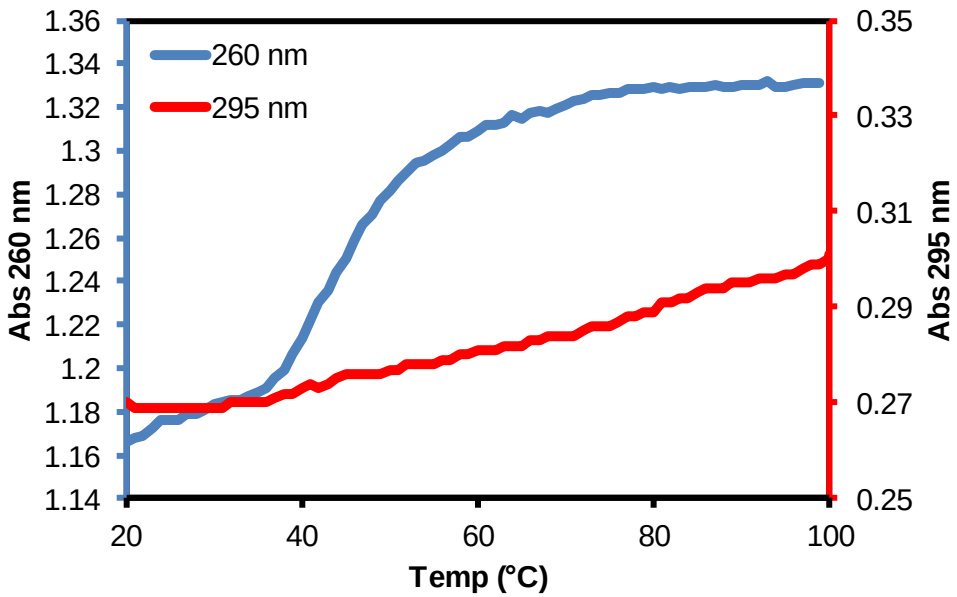


Figure S8. Representative thioflavin T emission spectra

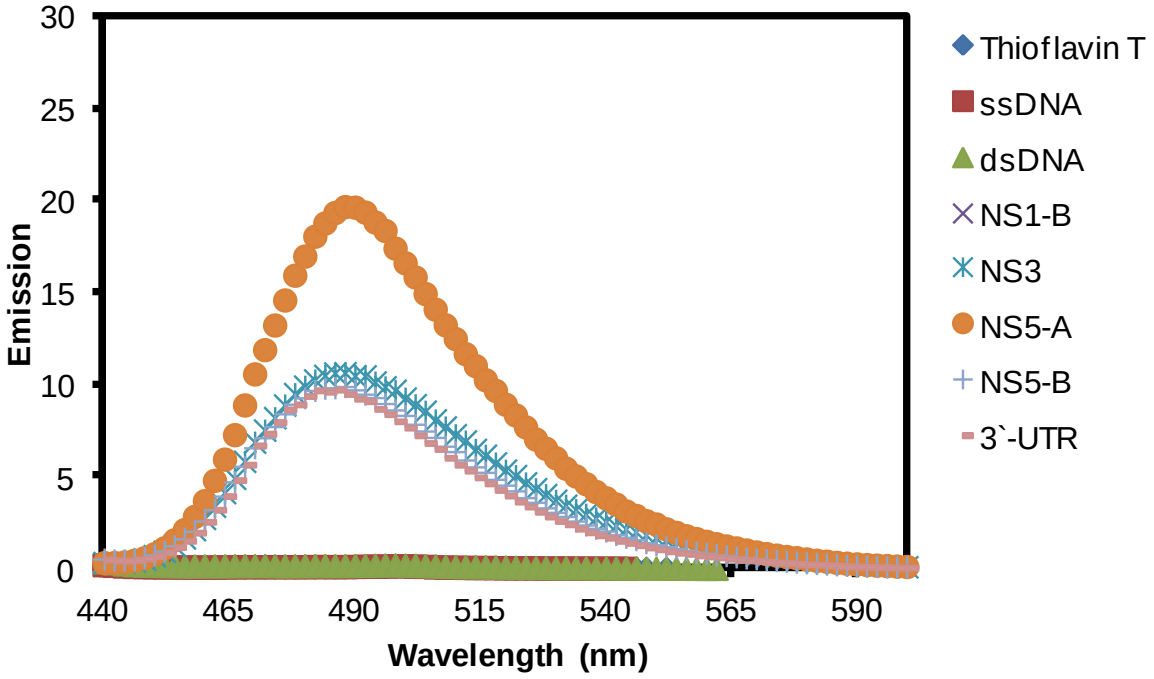
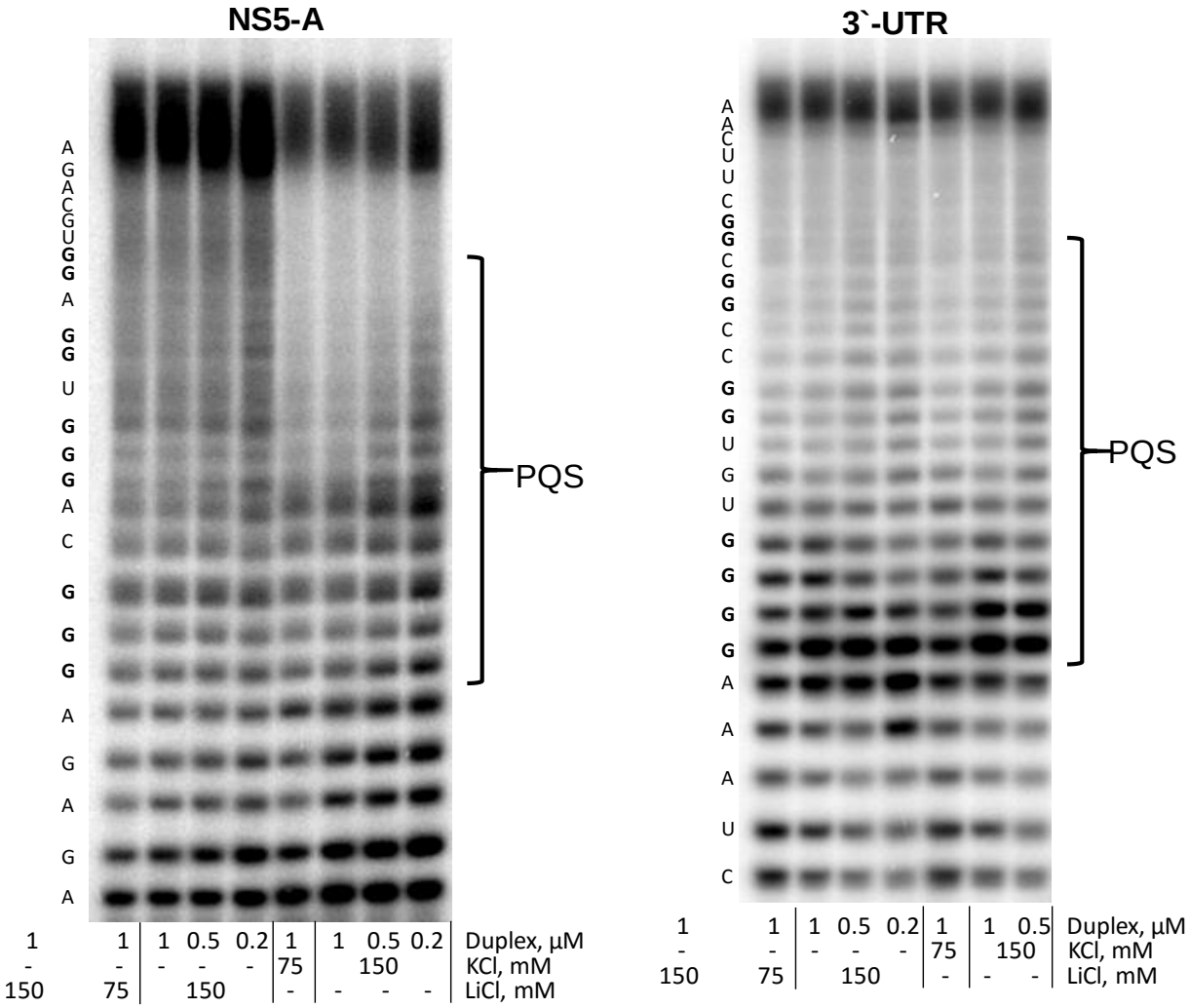


Figure S9. Denaturing gel electrophoresis for polymerase bypass of the NS5-A and 3`-UTR PQS



References

1. Huppert, J. L., and Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res. 33*, 2908-2916.

2. Kikin, O., D'Antonio, L., and Bagga, P. S. (2006) QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res. 34*, W676-W682.

3. Bedrat, A., Lacroix, L., and Mergny, J. L. (2016) Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res. 44*, 1746-1759.

4. Wright, E. S. (2015) DECIPHER: harnessing local sequence context to improve protein multiple sequence alignment. *BMC Bioinformatics 16*, 1-14.

5. Kumar, S., Stecher, G., and Tamura, K. (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol. 33*, 1870-1874.

6. Ponty, Y., Termier, M., and Denise, A. (2006) GenRGenS: software for generating random genomic sequences and structures. *Bioinformatics 22*, 1534-1535.