

## SUPPLEMENTARY MATERIAL

# Characterization of a novel mitovirus of the sand fly *Lutzomyia longipalpis* using genomic and virus-host interaction signatures

Paula Fonseca<sup>1</sup>, Flavia Ferreira<sup>2</sup>, Felipe da Silva<sup>3</sup>, Liliane Santana Oliveira<sup>4,7</sup>, João Trindade Marques<sup>2,3,5</sup>, Aristóteles Goes-Neto<sup>1,3</sup>, Eric Aguiar<sup>3,6,\*†</sup> and Arthur Gruber<sup>4,7,8\*,†</sup>

<sup>1</sup> Department of Microbiology, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte MG, 30270-901, Brazil; [camargos.paulaluize@gmail.com](mailto:camargos.paulaluize@gmail.com)

<sup>2</sup> Department of Biochemistry and Immunology, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, 30270-901, Brazil; [fvianaferreira@gmail.com](mailto:fvianaferreira@gmail.com)

<sup>3</sup> Bioinformatics Postgraduate Program, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, 30270-901, Brazil; [felselva@gmail.com](mailto:felselva@gmail.com), [arigoesneto@gmail.com](mailto:arigoesneto@gmail.com)

<sup>4</sup> Bioinformatics Postgraduate Program, Universidade de São Paulo, São Paulo, SP, 05508-000, Brazil; [liliane.sntn@gmail.com](mailto:liliane.sntn@gmail.com)

<sup>5</sup> Université de Strasbourg, CNRS UPR9022, Inserm U1257, 67084 Strasbourg, France; [jtmarques2009@gmail.com](mailto:jtmarques2009@gmail.com)

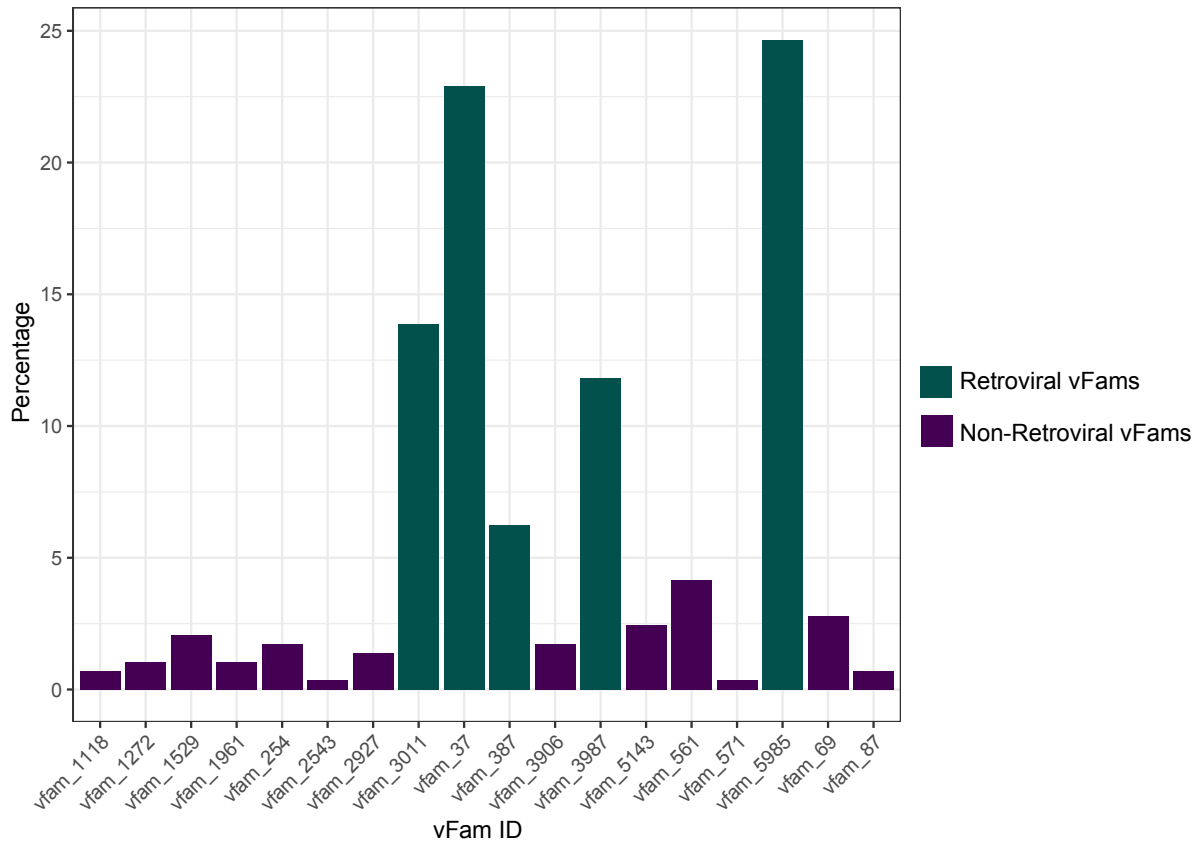
<sup>6</sup> Department of Biological Science (DCB), Center of Biotechnology and Genetics (CBG), State University of Santa Cruz (UESC), Rodovia Ilhéus-Itabuna km 16, Ilhéus, BA, 45652-900, Brazil; [ericgdp@gmail.com](mailto:ericgdp@gmail.com)

<sup>7</sup> Department of Parasitology, Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, SP, 05508-000, Brazil; [argruber@usp.br](mailto:argruber@usp.br)

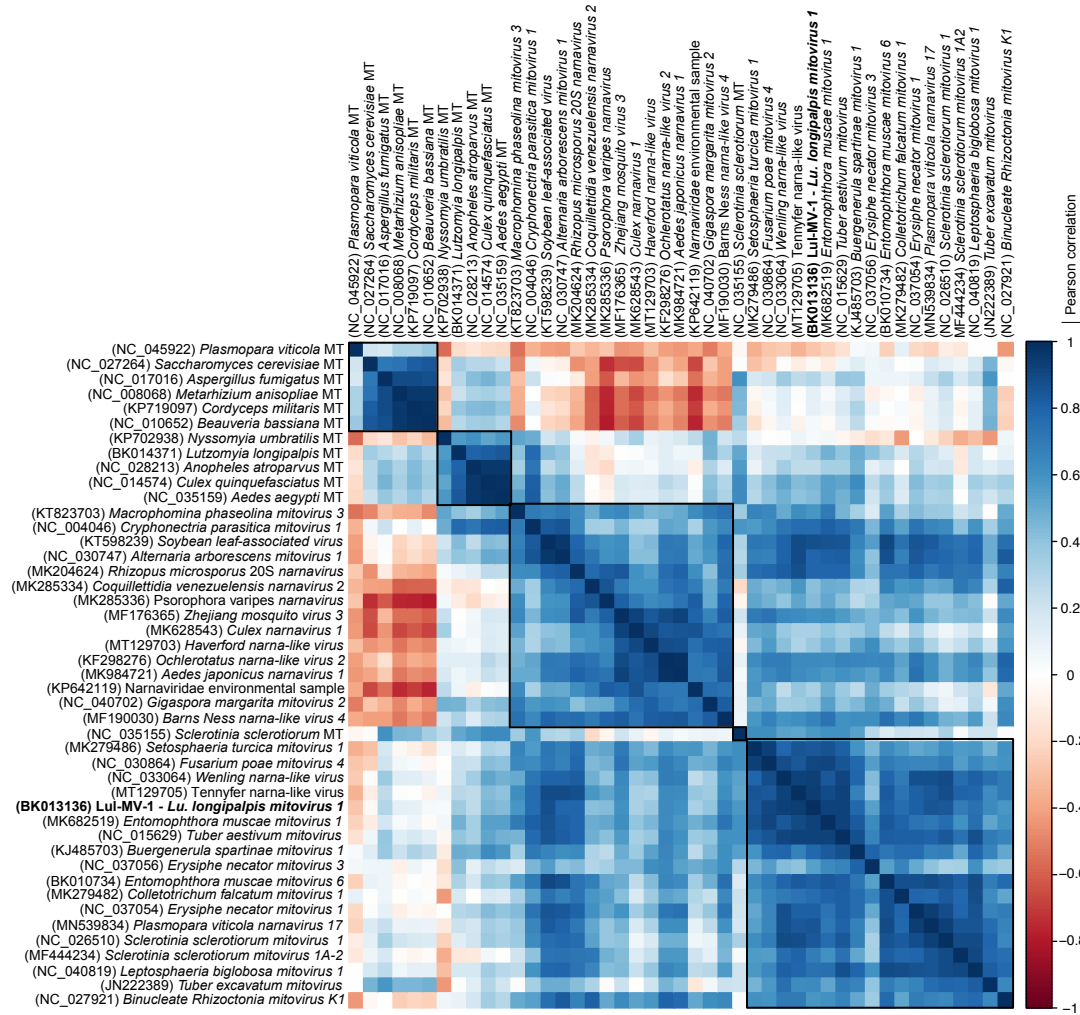
<sup>8</sup> European Virus Bioinformatics Center, Leutragraben 1, Jena, 07743, Germany

\* Correspondence: [ericgdp@gmail.com](mailto:ericgdp@gmail.com) (EA) and [argruber@usp.br](mailto:argruber@usp.br) (AG).

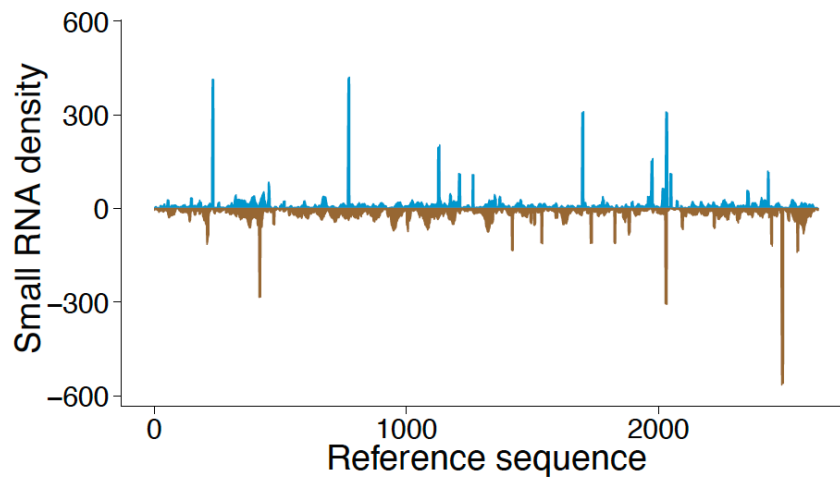
† Both corresponding authors contributed equally to this work.



**Figure S1.** Percentage of vFam domains found in the long RNA library of *Lutzomyia longipalpis*. Functional annotation of the contigs obtained by progressive assembly using vFam models as seeds is listed in the Supplementary Table S5.



**Figure S2. Comparative analysis of dinucleotide frequencies of the novel virus and putative host genomes.** Correlation plot calculated with `corrplot` program and based on dinucleotide odds ratio of the *Lutzomyia longipalpis* mitovirus 1 (Lul-MV-1), other mitoviruses and mitochondrial genomes (MT) of some fungi and insects. Clustering was performed using K-means method.



**Figure S3. Distribution of virus-derived small RNAs mapping on the *Lutzomyia longipalpis mitovirus 1* (Lul-MV-1) genome.** Blue and brown colors indicate density of small RNA reads mapped to the positive and negative strands of the viral genome, respectively.

Sequence ID: Query\_137189 Length: 2634 Number of Matches: 1  
Range 1: 1926 to 2409

Score	Expect	Identities	Gaps	Strand	Frame
767 bits(415)	0.0()	461/484(95%)	0/484(0%)	Plus/Plus	
Features:					
Query 1		AGGTCCAGAGGCAGAATATACACTTCCGGTTGGTAAACCAGGAGTATATATTTTCACCTAA			60
Sbjct 1926		AGGTCCAGAAGCAGAATATACACTTCCGGTTTGTAAACCAGGAGTATATATTTTCGCCTAA			1985
Query 61		GGATTTAATGGAAATTCCATCTTTTTCGAAGATCAGCTACTGATTGGACTAACTTCAAGTC			120
Sbjct 1986		AGATTTAATGGAAATTCCATCTTTTTCGAAGATCTGCTACTGATTGAACTAACTTCAAGTC			2045
Query 121		ATATATTTTAAGAATAATAATTCATTCTTAATAATATGGTAAATGTTGGAAAAGGTTAT			180
Sbjct 2046		ATATATTTTAAGAATAATAATTCATTCTTAATAATATGGTAAATGTTGGTAAAAGATAT			2105
Query 181		TACCTTAACTACAGATACCAAATTAAGAGGAACTGGAGTATTAAACTTTTATATACTCG			240
Sbjct 2106		AACTTTTACTACAGATACCAAATTAAGAGGAACTGGAGTATTAAACTTTTATATACTCG			2165
Query 241		AAGTTGGGTTAGGATAATTATCTTATCATAATCTGGAATTCATACAGTGAAAGAGGATA			300
Sbjct 2166		AAGTTGGGTTAGGATAATCATTTTATCATAATCTGGAATTCATACAGTGAAAGAGGATA			2225
Query 301		AAGATATGATGGAAAACCATCATTGACCTTTCGTCATCAGACAGATTGGTCAATCTTTAC			360
Sbjct 2226		AAGATATGATGGAAAACCATCATTGACCTTTCGTCATCAGACAGATTGGTCAATCTTTAC			2285
Query 361		CTTTTCACCTGTAAGAAAATACAGTTTGATAAAATTCATATCCTCCTTAAAGCGAGCAAC			420
Sbjct 2286		CTTTTCACCTGTAAGAAAATACAGTTTGATAAAATTTATATCCTCCTTAAATCGAGCAAC			2345
Query 421		AATCATTTCTGGTTGTTGCTTCTTTAAAGATTCTTCTGAACATGATAGAGTCTATTTAG			480
Sbjct 2346		AATCATTTCTGGTTGTTGTTTCTTTAAAGATTCTTCTGAACATGATAGAGTCTATTTAG			2405
Query 481		GACT			484
Sbjct 2406		GACT			2409

**Figure S4. Sequence alignment of a fragment of the RdRp gene.** Primers based on the assembled sequence of Lul-MV-1 were designed and used to amplify a 509-bp target of the RdRp gene using total RNA from pools of *Lu. longipalpis* individuals maintained in a laboratory colony. PCR fragments were submitted to Sanger sequencing and aligned to the Lul-MV-1 sequence. The query corresponds to the nucleotide sequence obtained from one of the pools and the subject to the Lul-MV-1 sequence.

**Table S1.** Public RNA libraries of *Lu. longipalpis* analyzed in this study.

SRA ID	Strategy	Number of reads
SRR535766	long RNAs	24,496,036
SRR535767	long RNAs	18,577,781
SRR1803384	small RNAs	10,679,412
SRR1803385	small RNAs	9,463,241
SRR1803386	small RNAs	8,109,613
SRR5224248	small RNAs	8,587,131
SRR5224249	small RNAs	8,500,448
SRR6429752	small RNAs	8,766,757

**Table S2.** Public protein sequences used in this work.

Organism	Accession number
<i>Aedes japonicus narnavirus 1</i>	QIP67846
<i>Alternaria arborescens mitovirus 1</i>	YP_009270635
<i>Aspergillus fumigatus mitovirus 1</i>	AXE72932
<i>Barns Ness breadcrumb sponge narna-like virus 4</i>	ASM94069
<i>Beihai narna-like virus 9</i>	YP_009333316
<i>Beihai narna-like virus 18</i>	YP_009333315
<i>Binucleate Rhizoctonia mitovirus K1</i>	YP_009165597
<i>Buergenerula spartinae mitovirus 1</i>	AHY03257
<i>Cassava virus C</i>	YP_003104770
<i>Coquillettidia venezuelensis narnavirus 2</i>	QBA55487
<i>Cryphonectria parasitica mitovirus 1</i>	NP_660174
<i>Culex narnavirus 1</i>	QBR53296
<i>Enterobacteria phage MS2</i>	YP_009640127
<i>Epirus cherry virus</i>	YP_002019754
<i>Escherichia virus BZ13</i>	NP_040755
<i>Fusarium poae mitovirus 4</i>	YP_009272901
<i>Grapevine-associated narnavirus 1</i>	CEZ26304
<i>Haverford narna-like virus</i>	QIJ70061
<i>Hubei narna-like virus 3</i>	YP_009337787
<i>Hubei narna-like virus 13</i>	YP_009337805
<i>Lutzomyia longipalpis mitovirus 1 (Lul-MV-1)</i>	BK013136
<i>Mitovirus AEF-2013</i>	AGW51760
<i>Mitovirus JS3</i>	QHA33929.1
<i>Narnaviridae environmental sample</i>	AJT39596
<i>Ochlerotatus-associated narna-like virus 1</i>	AGW51766
<i>Ochlerotatus-associated narna-like virus 2</i>	AGW51768
<i>Ophiostoma mitovirus 4</i>	NP_660179
<i>Ourmia melon virus</i>	YP_002019757
<i>Plasmopara viticola lesion-associated mitovirus 56</i>	QIR30279
<i>Plasmopara viticola lesion-associated narnavirus 17</i>	QIR30296
<i>Plasmopara viticola lesion-associated mitovirus 23</i>	QIR30246
<i>Plasmopara viticola lesion-associated mitovirus 39</i>	QIR30262
<i>Psorophora varipes narnavirus</i>	QBA55486
<i>Rhizoctonia solani mitovirus 28</i>	QDW65418
<i>Rhizoctonia solani mitovirus 30</i>	QDW65420
<i>Rhizopus microsporus 20S narnavirus</i>	QBC65280
<i>Rhizopus microsporus 23S narnavirus</i>	QBC65281
<i>Saccharomyces 20S RNA narnavirus</i>	NP_660178
<i>Saccharomyces 23S RNA narnavirus</i>	AAC98708
<i>Sclerotinia sclerotiorum mitovirus 1 HC025</i>	YP_009121785
<i>Sclerotinia sclerotiorum mitovirus 1-A2</i>	AWY10962
<i>Setosphaeria turcica mitovirus 1</i>	AZT88625
<i>Shahe narna-like virus 6</i>	APG77166
<i>Soybean leaf-associated mitovirus 2</i>	ALM62242
<i>Tuber excavatum mitovirus</i>	AEP83726
<i>Tynnyfer narna-like virus</i>	QIJ70064
<i>Wenling narna-like virus 9</i>	YP_009337200
<i>Wuhan insect virus 18</i>	YP_009342440
<i>Zhejiang mosquito virus 3</i>	YP_009333331.1

**Table S3.** Public genome sequences used in this work.

Sample source/organism	Accession number
<b>Mitochondrial genome</b>	
<i>Aedes aegypti</i>	NC_035159
<i>Anopheles atroparvus</i>	NC_028213
<i>Aspergillus fumigatus</i>	NC_017016
<i>Beauveria bassiana</i>	NC_010652
<i>Cordyceps militaris</i>	KP719097
<i>Culex quinquefasciatus</i>	NC_014574
<i>Cyberlindnera jadinii</i>	NC_022163
<i>Lutzomyia longipalpis</i>	BK014371
<i>Metarhizium anisopliae</i>	NC_008068
<i>Nyssomyia umbratilis</i>	KP702938
<i>Plasmopara viticola</i>	NC_045922
<i>Saccharomyces cerevisiae</i>	NC_027264
<i>Sclerotinia sclerotiorum</i>	NC_035155
<b>Viral genome</b>	
<i>Aedes japonicus narnavirus 1</i>	MK984721
<i>Alternaria arborescens mitovirus 1</i>	NC_030747
<i>Barns Ness breadcrumb sponge narna-like virus 4</i>	MF190030
<i>Binucleate Rhizoctonia mitovirus K1</i>	NC_027921
<i>Buergenerula spartinae mitovirus 1</i>	KJ485703
<i>Colletotrichum falcatum mitovirus 1</i>	MK279482
<i>Coquillettidia venezuelensis narnavirus 2</i>	MK285334
<i>Cryphonectria parasitica mitovirus 1</i>	NC_004046
<i>Culex narnavirus 1</i>	MK628543
<i>Entomophthora muscae mitovirus 6</i>	BK010734
<i>Entomophthora muscae mitovirus 7</i>	MK682519
<i>Erysiphe necator mitovirus 1</i>	NC_037054
<i>Erysiphe necator mitovirus 3</i>	NC_037056
<i>Fusarium poae mitovirus 4</i>	NC_030864
<i>Gigaspora margarita mitovirus 1</i>	NC_040702
<i>Grapevine associated narnavirus 1</i>	LN827948
<i>Haverford narna-like virus</i>	MT129703
<i>Leptosphaeria biglobosa mitovirus 1</i>	NC_040819
<i>Lutzomyia longipalpis mitovirus 1 (Lul-MV-1)</i>	BK013136
<i>Macrophomina phaseolina mitovirus 3</i>	KT823703
<i>Narnaviridae environmental sample clone</i>	KP642119
<i>Ochlerotatus-associated narna-like virus 2</i>	KF298276
<i>Plasmopara viticola lesion-associated mitovirus56</i>	MN539817
<i>Psorophora varipes narnavirus</i>	MK285336
<i>Rhizopus microsporus 20S narnavirus</i>	MK204624
<i>Sclerotinia sclerotiorum mitovirus 1</i>	NC_026510
<i>Sclerotinia sclerotiorum mitovirus 1-A2</i>	MF444234
<i>Setosphaeria turcica mitovirus 1</i>	MK279486
<i>Shahe narna-like virus 6</i>	KX883554
<i>Soybean leaf-associated mitovirus 2</i>	KT598239
<i>Tuber aestivum mitovirus</i>	NC_015629
<i>Tuber excavatum mitovirus</i>	JN222389
<i>Tynnyfer narna-like virus</i>	MT129705
<i>Wenling narna-like virus 9</i>	NC_033064
<i>Zhejiang mosquito virus 3</i>	MF176365



**Table S4.** Oligonucleotides designed and used in this study.

Forward primer (5'-3')	Reverse primer (5'-3')	Amplicon size (bp)
GGAAGATCCATTAGGTCCAG	CAGGTACACGAAGAGTCCTA	509

**Table S5.** Functional annotation of the sequences used to build the original vFam\* models utilized as seeds for progressive assembly with GenSeed-HMM program.

Profile HMM	Functional annotation
vFam_2927	DNA-directed RNA polymerase II subunit 1
vFam_2543	DNA-directed RNA polymerase II subunit RPB5
vFam_5985	RNA-dependent DNA polymerase
vFam_69	DNA polymerase
vFam_37	Reverse transcriptase
vFam_387	Capsid protein
vFam_3987	Reverse transcriptase
vFam_3011	Reverse transcriptase
vFam_561	Nucleoprotein N
vFam_1529	Nucleoprotein N
vFam_5143	DNA-directed RNA polymerase subunit alpha
vFam_3906	DNA-directed RNA polymerase subunit alpha
vFam_571	RNA-dependent RNA polymerase
vFam_1961	RNA-dependent RNA polymerase
vFam_1272	RNA-dependent RNA polymerase
vFam_87	Capsid protein
vFam_1118	Capsid protein
vFam_254	DNA polymerase

\*vFAM database of profile HMMs (<http://derisilab.ucsf.edu/software/vFam/>).

**Table S6.** Functional annotation of the contigs obtained by progressive assembly using vFam models as seeds.

<b>vFam seed</b>	<b>Number of assembled contigs</b>	<b>Function annotation of the coded protein</b>
vFam_2927	4	Contig 1 – RNA polymerase 2
vFam_2543	1	Contig 1 – DNA-dependent RNA polymerase
vFam_5985	71	Contig 1 – Reverse transcriptase Contig 3 – RNA-dependent DNA polymerase
vFam_69	8	Contig 1 – DNA polymerase subunit epsilon Contig 2 – DNA polymerase delta
vFam_37	66	Contig 1 – Reverse ribonuclease integrase
vFam_387	18	Contig 1 – Reverse transcriptase Contig 2 – DNA helicase
vFam_3987	34	Contig 1 – Reverse ribonuclease integrase
vFam_3011	40	Contig 1 – Reverse ribonuclease integrase
vFam_5143	7	Contig 1 – DNA-dependent RNA polymerase
vFam_3906	5	Contig 1 – DNA-dependent RNA polymerase
vFam_571	1	Contig 1 – RNA-dependent RNA polymerase
vFam_1961	3	Contig 1 – RNA-dependent RNA polymerase
vFam_1272	3	Contig 1 – RNA-dependent RNA polymerase
vFam_87	2	Contig 1 – Capsid protein
vFam_1118	2	Contig 1 – Capsid protein
vFam_254	5	Contig 1 – DNA polymerase
vFam_561	12	Contig 1 – Nucleoprotein N
vFam_1529	6	Contig 1 – Nucleoprotein N

**Table S7.** Use of UGA and UGG codons in coding sequences of some viruses and their putative hosts.

Accession number	Organism/source	# of UGA / UGG	% of UGA / UGG	AT content (%)
<b><i>Leviviridae</i></b>				
NC_001426	<i>Escherichia virus BZ13</i>	***1/26	3.7/96.3	52.14
NC_001417	<i>Escherichia virus MS2</i>	0/23	0.8/100	47.88
NC_000913	<i>Escherichia coli</i> str. K-12 substr. MG1655	***167/20886	0/100	50.80
<b><i>Botourmiaviridae/Ourmia-like viruses</i></b>				
MH192988	<i>Aspergillus fumigatus</i> mitovirus 1	***1/19	5.0/95.0	45.02
NC_017016	<i>Aspergillus fumigatus</i> mitochondrion	68/5	93.2/6.8	74.52
NC_013111	<i>Cassava virus C</i>	0/17	0/100	50.07
NC_011065	<i>Epirus cherry virus</i>	0/14	0/100	46.42
NC_011068	<i>Ourmia melon virus</i>	***1/14	6.7/93.3	48.47
<b><i>Narnaviridae</i></b>				
MK984721	<i>Aedes japonicus narnavirus 1</i>	0/11 (RdRp) 0/9 (HR)	0/100 (RdRp) 0/100 (HR)	40.83
MF194022	<i>Aedes aegypti</i> mitochondrion	99/3	97.1/2.9	78.96
MK628543	<i>Culex narnavirus 1</i>	0/18 (RdRp) 0/7 (HR)	0/100 (RdRp) 0/100 (HR)	38.54
GU188856	<i>Culex quinquefasciatus</i> mitochondrion	94/5	94.9/5.1	78.14
KF298276	<i>Ochlerotatus-associated narna-like virus 2</i>	0/15 (RdRp) 0/13 (HR)	0/100 (RdRp) 0/100 (HR)	40.24
MT993477	<i>Ochlerotatus caspius</i> mitochondrion**	15/0	100/0	70.66
NC_004051	<i>Saccharomyces 20S RNA narnavirus</i>	***1/18	5.3/94.7	41.73
NC_004050	<i>Saccharomyces 23S RNA narnavirus</i>	***1/20	4.8/95.2	41.02
MK628543	<i>Saccharomyces cerevisiae</i> mitochondrion	39/0	100/0	83.87
<b><i>Mitoviridae</i></b>				
NC_004046	<i>Cryphonectria mitovirus 1</i>	9/8	52.9/47.1	63.42
AF456838	<i>Cryphonectria parasitica</i> mitochondrion**	23/5	82.1/17.9	67.95
NC_026510	<i>Sclerotinia sclerotiorum</i> mitovirus 1	10/3	76.9/23.1	60.24
MF444234	<i>Sclerotinia sclerotiorum</i> mitovirus 1-A2	8/4	66.7/33.3	63.37
KT283062	<i>Sclerotinia sclerotiorum</i> mitochondrion	125/29	81.2/18.8	69.10
NC_004052	<i>Ophiostoma</i> mitovirus 4	11/2	84.6/15.4	73.26
KY084297	<i>Ophiostoma novo-ulmi</i> mitochondrion**	16/1	94.1/5.9	76.18
BK013136	<i>Lutzomyia longipalpis</i> mitovirus 1	11/4	73.3/26.7	69.74
BK013137	<i>Lutzomyia longipalpis</i> mitochondrion	99/1	99.0/1.0	78.07
MN539817	<i>Plasmopara viticola</i> associated mitovirus 56	11/11	50.0/50.0	57.53
NC_045922	<i>Plasmopara viticola</i> mitochondrion	***1/101	1.0/99.0	76.29
KX883554.1	<i>Shahe narna-like virus</i>	0/8	0/100	65.34
NC_033064	<i>Wenling narna-like virus 9</i>	13/1	92.9/7.1	63.84

\*Abbreviations: HR - Hypothetical gene coded in the reverse frame; RdRp – RNA-dependent RNA polymerase.

\*\*Incomplete sequence. \*\*\*UGA codon used for signaling translation termination.