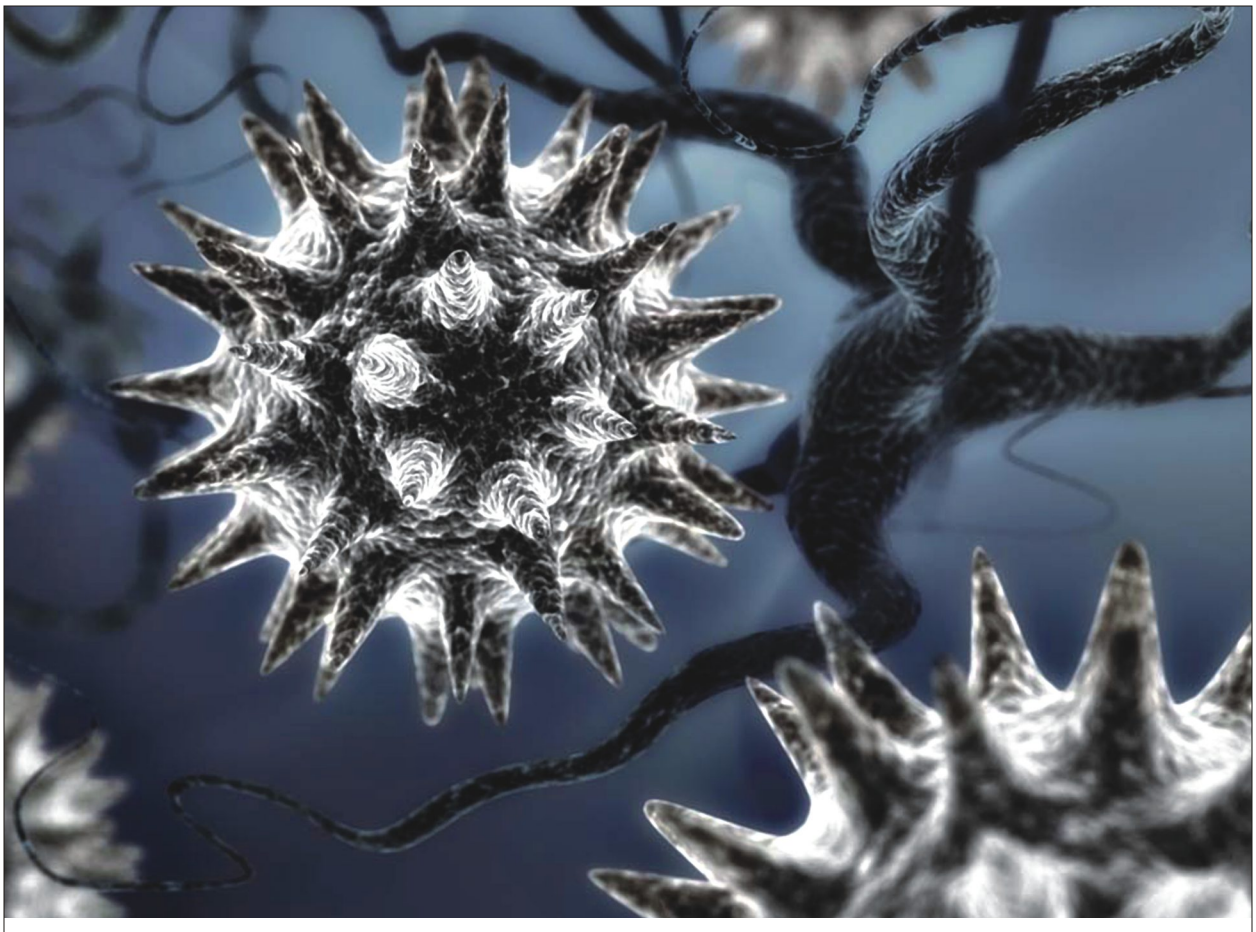

International Journal of Virology

ISSN 1816-4900

Academic Journals Inc.

www.academicjournals.com

OPEN ACCESS International Journal of Virology

ISSN 1816-4900
DOI: 10.3923/ijv.2025.1.23

Research Article
Repeating Amino Acids Point to Important Structural Motifs in NS3 and NS5 Proteins of Flaviviruses

[1]Ulyana Valerievna Potapova, [2]Artem Nikolaevich Bondaryuk, [3]Renat Viktorovich Adelshin and [4]Anzhelika Viktorovna Yudinceva

[1]Department of Biotechnology, Belarusian State Technological University, 13a Sverdlova Str., Minsk, Minskaya Vablasc 220006, Republic of Belarus
[2]Laboratory of Natural Focal Viral Infections, Irkutsk Antiplague Research Institute of Siberia and the Far East, 78 Trilisser Str., Irkutsk Oblast 664047, Russia
[3]Faculty of Biology and Soil Sciences, Irkutsk State University, 1 Karl Marx Str., Irkutsk Oblast 664033, Russia
[4]Limnological Institute Siberian Branch of the Russian Academy of Sciences, 3 Ulan-Batorskaya Str., Irkutsk Oblast 664033, Russia

Abstract

Background and Objective: Many studies point to conserved motifs in various proteins of the *Orthoflavivirus* genus. This study aimed to uniformly designate conserved motifs in two nonstructural proteins based on the "amino acid repeat" model and point out the connection, on the one hand, with their role in the functions of proteins and their role in the ecological strategy of the virus. **Materials and Methods:** It was supposed in the study, that some of the repeating amino acids in the protein sequence of the viruses remain conservative within a genus or a group and point to meaningful segments of the protein. The repeating amino acids were detected in all proteins of flavivirus, in the genomes of more than 70 virus types. The location of the repeats in the tertiary structure was analyzed. The position of conservative repeats in groups of viruses was associated, according to its vectors. **Results:** It was observed that NS3 and NS5 proteins have conservative repeats in different regions of the proteins. The genus-specific repeats QQ352-353 and KK461-462 of the interdomain region of the NS5 protein may be involved in blocking innate immunity through the STAT2 system. Also, group-specific repeats of a previously unexplored interface from the back side of EE438-439 helicase, RR450-451 and DDDD481-484, may be responsible for the difference in virus reproduction in ticks and mosquitoes. **Conclusion:** A method of amino acid repeats proves its effectiveness. Conservation of repeat within a genus may be important for virus survival. On the other hand, conservation of repeat within a group may have implications for viral propagation in a particular vector or host.

Key words: Flavivirus, RNA-dependent RNA polymerase, helicase, structural motifs, tick-borne infection, mosquito-borne infection

Citation: Potapova, U.V., A.N. Bondaryuk, R.V. Adelshin and A.V. Yudinceva, 2025. Repeating amino acids point to important structural motifs in NS3 and NS5 proteins of flaviviruses. Int. J. Virol., 20: XX-XX.

Corresponding Author: Ulyana Valerievna Potapova, Department of Biotechnology, Belarusian State Technological University, Minsk, Minsk Region 220006, Belarus

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

INTRODUCTION

Viruses of the genus *Orthoflavivirus* are a major global health threat. It includes many important human pathogens, such as Dengue virus, Zika virus and West Nile virus and are extensively studied[1-3]. The sequences of these viruses are diverse and there is a need to find a way that uniformly describes these differences. The treatment of flavivirus infection needs specific anti-viral substances. It requires a precise study of viral proteins to identify binding sites for inhibitors. There are several ways to detect sequence-structural relationships, here an original method of "amino acid repeats" is proposed. It assumes that a simultaneous occurrence of the same two residues in one site points to the importance of that site and area.

Viruses of the genus *Orthoflavivirus* have a single-stranded genome with positive polarity RNA of about 11000 nucleotides in size. The genome contains a single open reading frame (ORF) flanked by 5' and 3' untranslated regions[4]. ORF encodes a polyprotein, the processing of which produces 3 structural proteins-capsid (C), premembrane/membrane (prM/M) and envelope protein (E) and 7 non-structural proteins NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5[5]. The multifunctional proteins NS3 and NS5 have enzymatic activities (protease/helicase-NS3, methyltransferase/RNA-dependent RNA polymerase-NS5) required for viral RNA synthesis. The NS3 and NS5 proteins are identified as the main targets for drug discovery, due to their central role in the life cycle of the virus[6,7]. The identification of binding sites for allosteric inhibitors requires the characterization of functional interfaces in NS3 and NS5[8]. The family Flaviviridae is 1 of 3 viral families that encode RNA helicase[9,10]. The helicase domain consists of 8 conserved motifs I, Ia, Ib, II, III, IV, V and VI[11]. Conserved motifs form the functional core of the helicase and are involved in the binding of ATP and RNA molecules, ATP hydrolysis and unwinding of RNA chains.

The sequence of the MTase domain is poorly conserved among flaviviruses. The core globule of the MTase domain consists of a 7-component antiparallel β-sheet, which is surrounded by 10 α-helices and 2 short β-strands. The β-strands 5, 7 and 8 form the RNA binding site. Binding sites for the reaction cofactor SAH and GTP have also been described by Dong *et al.*[12].

The sequence and structure of the RNA-dependent RNA polymerase of flaviviruses is highly conserved and has now been well studied. Functional structural elements are known and these are typical for the protein family RdRp. They include nuclear localization sequence NLS (β/NLS and α/βNLS), priming loop and two zinc ion binding pockets (Zn1 and Zn2)[13]. Viral genomes are mutating and are under evolutionary pressure. Segments in the tertiary structure of viral proteins are moving under the influence of substitutions, but the activity of viruses must be preserved. Therefore, in some cases, the appearance of repeating amino acids in viral proteins can be meaningful and it indicates protein regions that are active in the catalytic function or the binding to other proteins.

In the present study, an attempt is made to identify repeating amino acids in the large corpus of sequences of flaviviruses, to associate the appearance of repeats with the properties of host-vector of the virus and to associate the position of repeat on 3D structure with the function of virus protein. The main objective of this study was to discover new functional interfaces in viral proteins of the genus *Orthoflavivirus*. The use of the method of conservative amino acid repeats in the study of the sequence and structure of viral proteins made it possible to identify previously unknown and poorly studied regions of the NS3 and NS5 proteins that are critical for the replication of viruses of the genus *Orthoflavivirus*.

MATERIALS AND METHODS

Study area: The scope of the study is flavivirus sequences isolated all over the world. A corpus of mosquito virus sequences was taken mostly from the African and Asian continents and tick-borne virus sequences were taken mostly from Russia, isolated from the beginning of 1990th. The study started in May, 2018 and lasted until May, 2023.

Sequences: A total of 1452 amino acid sequences for 78 species were taken from the GenBank database[14]. The reference sequences that were presented in the figures in this article were listed in Table 1, the remaining sequences were given in Table 2. The classification of viruses and taxon nomenclature is presented in accordance with the report of the International Committee on Taxonomy of Viruses (ICTV)[15].

The group of tick viruses TBFV was represented by 247 sequences for 14 species, the group of mosquito viruses MBFV is represented by 1090 sequences for 30 species, the group of viruses specific to vertebrates VSFV is represented by 22 sequences for 9 species, the group of viruses specific to insects ISFV is represented by 93 sequences for 25 species (Table 1 and 2).

Table 1: Reference GenBank accession codes of sequences analyzed in this study and presented in the alignment

Mammalian tick-borne flavivirus group					
Species	Abbreviation	Strain/Isolate	Host	GenBank No.	
Tick-borne encephalitis virus	TBEV		-		
Far Eastern subtype	TBEV-FE	Sofjn	-	AFV41132	1a
Siberian subtype	TBEV-Sib	Vasilchenko	-	AAF82240	1b
Baikalian subtype	TBEV-Baik	886-84	-	ABS00285	1c
European subtype	TBEV-Eur	Neudoerfl	-	NP_043135	1d
Himalayan subtype	TBEV-Him	Himalaya-1	-	AWK48861	1e
Louping ill virus	LIV		-		
British subtype	LIV-Brit	LI3/1	-	AKS47982	2
Greek goat encephalitis virus	GGEV	Vergina	-	ABB90677	3
Omsk hemorrhagic fever virus	OHFV	Bogoluvovska	-	NP_878909	4
Langat virus	LGTV	TP21	-	NP_620108	5
Kyasanur forest disease virus	KFDV	MCL_16_T_363	-	AXB87773	6
Powassan virus	POWV	Pow-24	-	AXY54925	7
Gadgets gully virus	GGYV	CSIRO122	*Ixodes uriae*	ABB90669	8
Royal farm virus	RFV	Eg Art 371	*Argas hermanni*	YP_009513190	9
Probably tick-borne					
Kadam tick-borne flavivirus group					
Kadam virus	KADV	AMP 6640	*Rhipicephalus pravus*	YP_009345035	10
Seabird tick-borne flavivirus group					
Kama virus	KAMV	LEIV-20776Tat	*Ixodes lividus*	AHH82587	11
Meaban virus	MEAV	Brest/Ar/T70	*Ornithodoros maritimus*	YP_009345031	12
Saumarez reef virus	SREV	CSIRO4	*Ornithodoros capensis*	YP_009345037	13
Tyuleniy virus	TYUV	6017	*Ixodes Uriae*	YP_009001464	14
Mosquito-borne flavivirus group					
Dengue virus group					
Dengue virus 1	DENV-1	SUB-120A	-	ATP66564	15
Dengue virus 2	DENV-2	BR/SJRP/779/2013	-	AKQ00025	16
Dengue virus 3	DENV-3	D83-144	-	AJA37731	17
Dengue virus 4	DENV-4	EHI310A129CY10	-	AGE13481	18
Japanese encephalitis virus group					
Japanese encephalitis virus	JEV	057434	-	ABU94628	19
Murray Valley encephalitis virus	MVEV	VIDRL	-	AXN90908	20
Usutu virus	USUV	AS201600083	*Strix nebulosa*	QCH41415	21
Koutango virus	KOUV	Dak Ar D 5443	-	ABW76844	22
West Nile virus	WNV	CHVir-205/2018	Isolation source-CSF, *Homo sapiens*	QCF41975	23
Yaounde virus	YAOV	Dak Ar Y276	-	YP_009350103	24
Cacipacore virus	CPCV	BeAn 3276000	*Formicarius analis*	YP_009126874	25
St. louis encephalitis virus	SLEV	Kern217	-	YP_001008348	26

Table 1: Continue

Mammalian tick-borne flavivirus group					
Species	Abbreviation	Strain/Isolate	Host	GenBank No.	
Ntaya virus group					
Rocio virus	ROCV	SPH 34675	-	YP_009553341	27
Bagaza virus	BAGV	DakAr B209	-	YP_002790883	28
Israel Turkey meningoencephalitis virus	ITV	ME 30502	*Meleagris gallopavo*	AIU94741	29
Ntaya virus	NTAV	IPDIA	-	YP_006846328	30
Tembusu virus	TMUV	FQ-C1	-	AQY45771	31
Zika virus	ZIKV	Zika virus homo sapiens Cuba 2017	-	ASK51714	32
Aroa virus group					
Aroa virus	AROAV	Strain bussuquara virus, isolate BeAn 4073	-	YP_001040004	33
Probably mosquito-borne					
Kedougou virus group					
Kedougou virus	KEDV	DakAar D1470	-	YP_002790882	34
Kokobera virus group					
Kokobera virus	KOKV	AusMRM 32	-	YP_001040007	35
Dual-host affiliated					
Insect-specific flaviviruses					
ISFV-like					
Lammi virus	LAMV	Strain mekrijarvi 2007, isolate M0719	*Aedes cinereus* (COI based identification)	YP_009056848	36
Chaoyang virus	CHAOV	HLD115	*Aedes vexans*	YP_005454257	37
Marisma mosquito virus	MMV	HU 4528/07	*Ochlerotatus caspius*	ATL63283	38
Donggang virus	DONV	DG0909	*Aedes* spp.	YP_005352889	39
Ilomantsi virus	ILOV	M0724	Mosquito pool, likely *Ochlerotatus riparius* and/or *Anopheles* spp. (COI based identification)	YP_009056847	40
Barkedji virus	BJV	SQU29/Oman/2016	*Culex quinquefasciatus*	AUS91146	41
Nounane virus	NOUV	Nounane_B3	*Uranotaenia mashonaensis*	YP_009345019	42
Yellow fever virus group					
Yellow fever virus	YFV	ArD149194	*Aedes taylori*	AFU76909	43
Sepik virus	SEPV	MK7148	Mosquito	YP_950478	44
Wesselsbron virus	WSLV	SAH177	-	YP_002922020	45
Probably mosquito-borne					
Edge Hill virus group					
Edge hill virus	EHV	YMP 48	-	YP_009256192	46
Banzi virus	BANV	SAH 336	-	ABI54472	47
Uganda S virus	UGSV	-	-	YP_009344968	48
Jugra virus	JUGV	P-9-314	-	YP_009344969	49
Saboya virus	SABV	Dak AR D4600	-	YP_009344967	50
Bouboui virus	BOUV	DAK AR B490	-	YP_009344961	51

Table 1: Continue

Mammalian tick-borne flavivirus group					
Species	Abbreviation	Strain/Isolate	Host	GenBank No.	
Species with no known arthropod vector					
Entebbe bat virus group					
Entebbe bat virus	ENTV	UgIL-30	Bat	YP_950477	52
Sokoluk virus	SOKV	LEIV-400K	*Pipistrellus sokolich*	YP_009126875	53
Yokose virus	YOKV	XYBX1332	*Pyotis daubentonii*	QAU20974	54
Species with no known arthropod vector					
Modoc virus group					
Modoc virus	MODV	M544	-	NP_619758	55
Jutiapa virus	JUTV	JG-128	*Sigmodon hispidus*	YP_009126871	56
Apoi virus	APOIV	ApMAR	-	NP_620045	57
Species with no known arthropod vector					
Rio bravo virus group					
Rio bravo virus	RBV	RiMAR	-	NP_620044	58
Phnom penh bat virus	PPBV	30834_A38	*Cynopterus brachyotis* angulatus	YP_009350101	59
Montana myotis leukoencephalitis virus	MMLV	-	-	NP_689391	60
Classical insect-specific flaviviruses					
ISFV[sl]					
Quang binh virus	QBV	VN180	*Culex tritaeniorhynchus* (mosquito)	YP_002884239	61
Culex theileri flavivirus RP-2011	CTFV	153	*Culex theileri*	CCC55432	62
Culex flavivirus	CxFV	Toyama1849	*Culex tritaeniorhynchus*	BAM74420	63
Palm creek virus	PCV	56	*Coquillettidia xanthogaster*	YP_009344962	64
Cuacua virus	CuCuV	MZ2014-Cv	*Mansonia* sp.	AOR81460	65
Nakiwogo virus	NAKV	Uganda08	*Mansonia africana nigerrima* voucher H4A1	YP_009259488	66
Nienokoue virus	NIEV	C64/CI/2004	*Culex* sp.	APR74286	67
Cell fusing agent virus	CFAV	Galveston	Aedes aegypti Galveston colony	YP_009259257	68
Kamiti river virus	KRV	SR-82	*Aedes mcintoshi*	NP_891560	69
Aedes flavivirus	AEFV	Narita-21	*Aedes albopictus*	YP_003029843	70
La tina virus	LTNV	49_LT96	*Host ochlerotatus scapularis*	ATL63282	71
Menghai flavivirus	MFV	MHAedFV1	*Aedes albopictus*	YP_009351861	72
Xishuangbanna aedes flavivirus	XFV	XSBNAeFV	*Aedes albopictus*	YP_009350102	73
Hanko virus	HANKV	-	*Mosquito*	YP_009259489	74
Sabethes flavivirus	SbFV	Ms42	*Sabethes belisarioi*	AZB73874	75
Haslams creek virus	HaCV	HaCV 115734	*Anopheles annulipes*	ASD50161	76
Karumba virus	KRBV	KRBV WBay2	*Anopheles meraukensis*	ARD06144	77
Mac Peak virus	McPV	McPV 150840	*Anopheles farauti*	YP_009389296	78

Table 2: All GenBank accession codes of sequences analyzed in this study

Group of viruses by vector	Virus group	Abbreviation		GenBank No.
Tick-borne flavivirus group	Mammalian group	TBEV	142	AEQ77277, AAF82240, AAD34205, AGI05090, AEQ77278, AIB53033, AEQ77280, AEQ77279, AKL59775, BAQ25430, AIL33471, AIB52942, AFU65175, AKP16366, BAQ25431, ALP82433, BAQ08281, AIB53035, AAO43537, ABF46836,AMD82540, ALP82432, AKL71380, ADQ00973, ADQ00971, ABS00285, ACN42746,ABS00284, AFP25086, AKP16371, ADX07734, ACJ38115, AKC88486, ACQ99330, AFP25085, AFP25082, ACF33496, AFP25084, AII96827, AFP25089, AKC88488, ALA09088, AFP25090, ACF33497, AGO86677, AEK94233, ALA09090, ABJ74160, AEP20480, AKP16370, AKC88490, AIL33469, AEP25267,BAMa62429, BAB72162, AFV48384, AIK20543, AIL33470, ADE93003, AGI19265, ACJ38114, AHL20249, AKC88489, AEK94232, AEQ77276, AFV41132, BAB71943, ACR49228, AFP25080, ACF33498, AKP16369, AFP25094, ADX87374, AFP25081, AGC31653, ACF33495, AAN73266, ADT80553, ACT32141, AIG24410, ACF33494, AFP25097, ADX87375, AFP25079, ANN44512, AFP25092, AFP25083, NP_043135, ABI31771, ACF33493, AHL20222, AIL83862, ALA09089, BAV60898, AIL83864, AFJ97024, AGP05331, ABD62793, AIG24412, AHF27216, AFI49403, AMQ36837, ADQ00970, ALP82437, ALA09091, AIL33472, AHL20221, ACL97686, ADM63091, AFJ97025, CAM82856, ADQ00969, AKE50910, AHM02467, AIL83860, ADE22271, AKE50894, ADQ00968, ADQ00972, AMQ49166, AHF27217, AFP25096, AAN87009,AHF27215, AKE50895, ADY80020, AKE50893, AIL83863, AAZ80455, AAA86739, AKP16367, AKP16368, ADY80019, AIG24411, AIL83861, Q01299.1, AIG51207, AHC30238, AHC30237, AHC30239, ADT80552, P07720.3
		LIV	7	AYG86395.1, NP_44677.1, ALP82434.1, AKS47982.1, AHY99586.1, AGN32859.1, CAA69190.1
		GGEV	1	ABB90677.1
		OHFV	4	NP_878909.1, AAR98531.1, BAH78736.1, AAQ91606.1
		LGTV	6	NP_620108.1, AWI66602.1, AAB22165.1, AAF75260.1, AAF75259.1, ACH42698.1
		KFDV	47	AXB87773.1, AXB87772.1, AXB87770.1, AXB87766.1, AXB87774.1, AXB87771.1, AXB87775.1, AXB87769.1, AXB87764.1, AXB87757.1, AXF50740.1, AXB87754.1, AXB87768.1, AXB87734.1, AXB87738.1, AXB87735.1, ACA50033.1, AXB87762.1, AXB87743.1, AXB87752.1, AFF18435.1, AFF18436.1, AXB87737.1, AFF18434.1, AXB87748.1, YP_009513189.1, AXB87731.1, AXB87765.1, AXB87751.1, AXB87760.1, AXB87750.1,AXB87763.1, AXB87746.1, AXB87740.1, AXB87745.1, AXB87747.1, AXB87744.1, AXB87733.1, AXB87767.1, AXB87756.1, AXB87761.1, AXB87730.1, AXB87753.1, AXB87755.1, AXB87739.1, AXB87742.1, AXB87732.1
		POWV	24	AXY54925.1, NP_620099.1, AWI66604.1, AUX13145.1, AUX13144.1, AUX13143.1, AUX13142.1, AUX13141.1, AIG95651.2, AMY50382.1, ALP82431.1, ALP82430.1, AEN83682.1, AEN83681.1, ADK37757.1, ADK37756.1, ADK37755.1, ADK37754.1, ADK37753.1, ADK37752.1, ACF05267.1, ACD88752.1, ACD67873.1, AAA02739.1
		GGYV	2	YP_009345034.1, ABB90669.1
		RFV	2	YP_009513190.1, ABB90673.1
	Kadam group	KADV	2	YP_009345035.1, ABB90670.1
	Seabird group	MEAV	2	YP_009345031.1, ABB90668.1
		SREV	2	YP_009345037.1, ABB90674.1
		KAMV	2	YP_009001771.1, AHH82587.1
		TYUV	4	YP_009001464.1, ALP82436.1, AHH82586.1, ABB90672.1
Mosquito-borne Flavivirus group	Dengue group	DENV-1	100	ATP66564.1, ATP66563.1, ABW82076.1, AHG06317.1, ABW82116.1, AHG06308.1, AQV12188.1, ABW82062.1, ABW82063.1, ABY65724.1, ABW82066.1, ABW82084.1, ABW82110.1, AQV12189.1, QBB90021.1, ALI16134.1, ABW82083.1, ABW82081.1, ABW82068.1, AHG06305.1, AWV55246.1, ADC92350.1, ABW82091.1, ACS12902.1, AHG06309.1, AWV55250.1, AHG06328.1, AHG06325.1, ABW82112.1, AQV12158.1, ABW82085.1, ABW82117.1, ABW82074.1, AHG06333.1, AQV12190.1, ABW82090.2, AHG06307.1, AHG06306.1, ABW82067.1, AHG06323.1, AQV12164.1, ABW82099.1, ACS12901.1, AQV12193.1, ACS12900.1, AHG06318.1, AHG06327.1, AHG06319.1, AJQ21325.2, ABW82070.1, AQV12163.1, AKC89102.1, ABA39233.1, ABW82096.1, AQV12191.1, ABW82089.1, AUZ41758.1, AQV12187.1, AHG06329.1, AHG06326.1, AHG06311.1, ABW82104.1, AQV12343.1, AHG06336.1, AQV12159.1, ACL99187.1, AQV12151.1, AQV12153.1, QBK46929.1, AWV55234.1, AQV12152.1, AHG06315.1, ABW82115.1, AHG06330.1, AUZ41774.1, ABW82064.1,

Table 2: Continue

Group of viruses by vector	Virus group	Abbreviation		GenBank No.
		DENV-3	100	ABW82093.1, AQV12363.1, AQV12310.1, ADU03653.1, ADC92349.1, AQV12341.1, AQV12154.1, AQV12360.1, ABW82075.1, AHG06331.1, ATP66556.1, ABW82111.1, AQV12166.1, BAD16759.1, ABW82078.1, AQV12162.1, AQV12157.1, AQV12349.1, AQV12195.1, ABW82071.1, AJQ21337.2, AQV12132.1, AQV12338.1, ACL99190.1 AJA37731.1, AAV88386.1, AAV88387.1, AAV88385.1, AAV88384.1, AQV12643.1, AQV12640.1, AJA37730.1, QBA29682.1, ACL99006.1, AQV12580.1, ACL99125.1, AQV12600.1, ACL99004.1, AQV12624.1, AQV12642.1, AQV12546.1, ABG73586.1, AHG23226.1, AUY62012.1, AQV12623.1, AQV12571.1, ACD13318.1, ACL99183.1, AQV12625.1, ACL99242.1, AQV12545.1, AAV88382.1, ACD13322.1, AQV12601.1, AQV12544.1, AQV12609.1, AUZ41821.1, ACW82883.1, ACA48726.1, AQV12572.1, ACJ04271.1, ACA48728.1, AHG23225.1, AQV12570.1, ACA48723.1, ADZ13190.1, AFD28531.1, ACL99007.1, AQV12595.1, ACL99244.1, AQV12533.1, ACN42695.1, ACL99130.1, ACW82917.1, ACL98979.1, AQV12627.1, AUZ41812.1, ACW82916.1, AQV12632.1, AAV88383.1, ACY70808.1, AFD28533.1, AQV12593.1, ACL99168.1, AQV12586.1, AHG23224.1, AQV12597.1, ACL99179.1, ACA48720.1, AQV12621.1, ACY70814.1, ACN42730.1, AQV12622.1, ACL99246.1, ACY70805.1, AQV12561.1, ACW82922.1, AQV12608.1, AQV12556.1, AQV12553.1, AQV12611.1, AQV12602.1, ACN42716.1, AQV12619.1, AQV12628.1, AQV12598.1, ACL99009.1, ACN42727.1, ACL99247.1, AQV12541.1, ACN42722.1, ACA48722.1, AQV12594.1, ACA48725.1, ACO06137.1, ACD13320.1, ABG73588.1, AQV12535.1, ACN42724.1, AQV12585.1, AQV12574.1, AQV12562.1, AQV12534.1, AQV12582.1
		DENV-2	92	AKQ00025.1, AKQ00027.1, AOE23002.1, AKQ00020.1, AKQ00024.1, AKQ00022.1, AKQ00026.1, AGK36297.1, ACO06171.1, ACW82873.1, AGK36292.1, AGX15368.1, ACK57811.1, ACW82875.1, AGX15374.1, AGK36293.1, AGK36298.1, AGK36291.1, AGX15373.1, AGN94882.1, AGX15383.1, AGX15376.1, AOE23003.1, ACK57809.1, ACK57806.1, ADI80655.1, ACK57812.1, ACK57817.1, AGX15378.1, AGX15386.1, AGX15385.1, ACK57814.1, AGX15379.1, AGX15367.1, AGX15370.1, AGK36289.1, AGN94883.1, ACS32033.1, AGX15382.1, AGN94881.1, AGX15371.1, AGX15372.1, AGX15375.1, AKQ00040.1, ACS32031.1, AKQ00021.1, ACD13400.1, ACQ44490.1, ACK57805.1, ACK57808.1, ACA48998.1, ACK57816.1, ACK57804.1, AGN94880.1, ACB87132.1, ACL99070.1, AAW31407.1, AAW31409.1, ACK57800.1, ACQ44499.1, ACD13352.1, ACK57801.1, AAW31412.1, ACK57796.1, ACD13396.1, ACH61726.1, ACL99075.1, AAG30730.1, ACE63535.1, AHI43688.1, ACL99109.1, ACQ44505.1, AFV95787.1, ACK57798.1, ACA48813.1, ACA49001.1, ACQ44506.1, ACW82981.1, ACA48812.1, ACW82985.1, AET43239.1, AEA50943.1, ACW82984.1, ACK57797.1, ACK57799.1, ACL99019.1, ACA48874.1, ACL99021.1, ACA48871.1, ACW82982.1, ACO06209.1, AET43250.1
		DENV-4	100	AGE13481.1, AHG06383.1, AEJ33672.1, AHG06378.1, AHG06380.1, AFK82574.1, AFY10034.1, AKQ62916.1, ARM59249.1, AXG50965.1, AHG06382.1, ANW72071.1, BAU45389.1, AFY10035.1, AAW30973.1, AFD53008.1, AHG06381.1, AFY10038.1, AFY10036.1, AHG06379.1, AAW51421.1, AFY10031.1, AFY10030.1, AFY10033.1, AFY10037.1, QBB90024.1, AQV12653.1, AAU89380.1, QBA29684.1, AFY10039.1, ADK37472.1, ANW72070.1, AGI95993.1, ANC57614.1, QCE20767.1, QCE20762.1, ARX96505.1, ACS32014.1, ACS32013.1, ACJ04171.1, ADA00410.1, AQV12652.1, ANK35835.1, AVY51410.1, AIG60035.1, AQV12650.1, ACO06145.1, ACS32037.1, ACO06140.1, AQV12651.1, AAK01233.1, AHN50410.1, Q5UCB8.1, AAK58017.1, ACQ44403.1, ACH61725.1, AEX09559.1, AAG45437.1, ACL99023.1, AQV12649.1, AEW50182.1, AHK05342.1, ACH61689.1, NP_73286.1, P09866.2, AAG45436.1, ACL99032.1, ACQ44402.1, ACS32018.1, ACQ44405.1, ACL99029.1, ACH61688.1, ACS32017.1, ACO06147.1, ACH61714.1, ACL99026.1, ACH61687.1, ACL99025.1, AET43240.1, ACS32016.1, AEX09560.1, ACO06199.1, ACQ44393.1, ACS32010.1, ACQ44389.1, ACH61693.1, ACL99035.1, ACH61691.1, AEA50929.1, AFX65871.1, AET43237.1, ACL99024.1, ACL99051.1, ACH99663.1, ACW83011.1, ACQ44390.1, ACQ44395.1, ACQ44408.1, ACL99060.1, ACS32015.1

Table 2: Continue

Group of viruses by vector	Virus group	Abbreviation		GenBank No.
	Japanese encephalitis group	JEV	99	ABU94628.1, ABU94629.1, ADN94471.1, ABU94627.1, AEO86778.1, AEO86789.1, AOT86205.1, AOT85914.1, AIN36638.1, AVP82543.1, ANV81277.1, AHK05344.1, AEO72437.1, AOT85917.1, AOT86206.1, AOT85937.1, AOT85929.1, ASM90791.1, BAJ51960.2, AIN36637.1, AAQ73508.1, AVP82544.1, AOT85932.1, AEO86792.1, AGW82423.1, BAI99561.1, AAQ73507.1, AAT00231.1, AAD20233.1, BAD81039.1, AEO72438.1, BAJ51953.2, AOT85920.1, AOT85936.1, AEO72440.1, AZS59363.1, AEO72439.1, BAJ51957.2, AOT85931.1, BAJ51954.2, BAJ51955.2, AKP06314.1, AZS59358.1, AEO86784.1, CCI69572.1, AEO72432.1, AEO72436.1, BAJ51952.2, AGT17711.1, AGT38389.1, BAI99560.1, BAF02840.1, AEY64199.1, AOT85921.1, AEO72435.1, AOT86203.1, BAI99562.1, AAC27708.1, AAA21436.1, AOT86222.1, AAS79438.1, AOT86221.1, BAJ51958.2, BAJ51950.2, AOT85927.1, AOT86201.1, AAM27885.1, NP_059434.1, AOT85941.1, AOT86211.1, AAQ73511.1, AOT85922.1, AOT86202.1, AEY64200.1, AAB66485.1, AOT86223.1, AEO86794.1, AOT85925.1, AOT85924.1, AQX45499.1, AOT86212.1, AOT85930.1, AFP33182.1, ACN39715.1, AOT85923.1, AAQ73512.1, AOT85926.1, AAM27886.1, AAQ73513.1, AZS59362.1, AAQ73514.1, AGL09208.1, AAQ73510.1, AEO72421.1, AEO72434.1, AAS79437.1, AOT86198.1, ANC28182.1, AEO72433.1
		MVEV	17	AXN90908.1, NP_051124.1, APW77688.1, AIN35081.1, AIA58171.1, AIA58170.1, AIA58169.1, AHF27232.1, AHF27231.1, AHF27230.1, AHF27229.1, AHF27228.1, AHF27227.1, AHF27226.1, AHF27225.1, AFY22653.1, AAF05296.1
		KOUV	1	ABW76844.2
		USUV	100	QCH41415.1, AID60245.1, AID60185.1, AID60220.1, AID60248.1, AID60182.1, AID60250.1, AID60196.1, AID60184.1, APT69980.1, AID60217.1, AID60231.1, AQT03216.1, AID60207.1, AID60199.1, AID60222.1, AID60195.1, APT69981.1, AID60234.1, AID60225.1, AID60211.1, AQT03209.1, AID60212.1, AID60186.1, AQT03210.1, AID60189.1, AID60257.1, APU52397.1, AID60230.1, AID60215.1, AID60198.1, AID60255.1, AID60206.1, AID60228.1, AQT03207.1, AQM55265.1, AID60194.1, AQT03220.1, AID60205.1, AQT03227.1, AID60240.1, AQT03212.1, AID60236.1, AID60244.1, AID60252.1, AID60227.1, AID60239.1, AID60216.1, AID60251.1, AQT03217.1, AQT03223.1, AQT03219.1, AQM55263.1, AQT03224.1, AQT03208.1, AID60233.1, AID60235.1, AID60232.1, AID60229.1, AQT03225.1, AID60226.1, AIN76231.1, AID60243.1, AID60242.1, AID60181.1, AID60237.1, AID60241.1, AID60201.1, AFE85504.1, AID60238.1, YP_164264.1, AID60253.1, AID60221.1, AID60256.1, ABP88817.1, AQP25640.1, APZ88659.1, APC23732.1, ATL75636.1, AQP25642.1, AEF13245.1, AWC68492.1, APZ88658.1, APC23727.1, AEK21245.1, APC23731.1, AGP50647.1, AQP25639.1, ATW75952.1, APZ88660.1, APZ88661.1, APQ43025.1, APZ88655.1, AGP50646.1, AXU25876.1, APC23730.1, AQP25637.1, APZ88657.1, AOP04274.1, AGP50648.1
		YAOV	2	YP_009350103.1, ABW76843.2
		WNV	100	QCF41975.1, AJR27895.1, AJR27897.1, AHH30722.1, AJR27896.1, AJR27892.1, AJR27894.1, AJR27893.1, AJR27898.1, AVL25111.1, AQP25647.1, AVL25113.1, AWT62803.1, AVL25117.1, AIJ03962.1, AVL25108.1, AIJ03959.1, AWT62802.1, AVL25115.1, AQP25649.1, AVL25109.1, AVL25112.1, AGX89731.1, AWT62805.1, AWT62804.1, AVL25121.1, AKH14861.1, AZK36004.1, AIJ03961.1, AVL25116.1, AIW82672.1, AVL25122.1, AHB37629.1, ALA10709.1, AVL25118.1, AVL25114.1, AGI15890.1, AGY34434.1, AVL25119.1, AJR27899.1, AMK52070.1, AVL25110.1, AHB37631.1, AAZ91684.1, ALA10710.1, AVL25123.1, AED99787.1, AVP26556.1, AHA58709.1, AZK36003.1, AMK51624.1, ARF20086.1, AEV45200.1, ARF20087.1, AMK51628.1, BBD13921.1, AIJ03960.1, AHA58708.1, AJR27178.1, AIA23853.1, AMK51627.1, ABU41789.1, AMK51626.1, AGX89730.1, AQY10112.1, AJR27177.1, AJR27176.1, AJR27179.1, AMK51629.1, ADZ96249.1, ABR19639.1, ABC49717.1, AIS74833.1, ANS10043.1, AFN70436.1, AIA23852.1, ACJ67802.1, ALA09311.1, AIQ82561.1, ALA10712.1, ABC49716.1, ABR19638.1, ABR19637.1, AAT02759.1, AFO64353.1, ABR19636.1, ADZ96248.1, NP_041724.2, AAT95390.1, ALA10711.1, AQP25650.1, ADZ96247.1, AQP25648.1, ADN65133.1, ABB01532.2, AYV87273.1, AVC05641.1, ADI33161.1, AAM81753.1, AAG02039.1

Table 2:Continue

Group of viruses by vector	Virus group	Abbreviation		GenBank No.
		SLEV	41	YP_001008348.1, ARR96288.1, ARR96287.1, APE73919.1, APE73918.1, APE73917.1, ABN11823.2, ABN11832.1, ABN11831.1, ABN11830.1, ABN11829.1, ABN11828.1, ABN11827.1, ABN11826.1, ABN11825.1, ABN11824.1, ABN11822.1, ABN11821.1, ABN11820.1, ABN11819.1, ABN11818.1, ABN11817.1, ABN11816.1, ABN11815.1, ABN11814.1, ABN11813.1, ABN11812.1, ABN11811.1, ABN11810.1, AMO02733.1, AIW82235.1, AEN02430.1, AHB17738.1, ACT31739.1, ACT31738.1, AGE00047.1, AGE00046.1, ACB58159.1, ABG37975.1, ABC87765.1, AAV34160.1
		CPCV	3	YP_009126874.1, CRI72595.1, AIU94740.1
	Ntaya group	BAGV	21	YP_002790883.1, AWA45345, AWA45344, AWA45343, AWA45342, AWA45341, AWA45340, AWA45339, AWA45338, AWA45337, AWA45336, AWA45335, BBD13922, ALD18905, ALD18904, ALD18903, AFX83954.1, AEI27245, AEI27244, AAV34161, ACG60714
		ITV	6	AIU94741.1, AGV15509.1, AGV15508.1, AGV15507.1, AGV15506.1, AGV15505.1
		NTAV	3	YP_006846328.2, AIU94743.1, AFS89612.2
		TMUV	97	AQY45771.1, AQY45770.1, AQY45769.1, AQY45768.1, AQY45767.1, AQY45766.1, AXY93835.1, YP_004734464.1, AWX59350.1, AWW43687.1, AWW43686.1, AWV66903.1, AWV66902.1, AQY77526.1, AQY77525.1, AQY77524.1, AQY77523.1, AQY77522.1, AQY77521.1, AQY77520.1, AQY77519.1, AQY77518.1, AQY77517.1, AQY77516.1, AQY77515.1, AQY77514.1, AQY77513.1, AQY77512.1, AQY77511.1, AQY77510.1, AQY77509.1, AQY77508.1, AQY77507, AQY77506, AQY77505, AQY77504, AQY77503, AQY77502, AQY77501, AQY77500, AQY77499, AOS50839, AOS50838, AOS50837, AOS50836, AOS50835, AOS50834, AOS50833, AOS50832, AOS50831, AOS50830, AOS50829, ANK79133, ANK79132, AMO44420, AIF73121, AIU44176, ALL27018, ALB36912, AKZ17853, AJR29358, AIX09855, BAQ19608, BAQ19607, BAQ19606, AIX09854, AIR72260, AIA08683.2, AIF73122, AIF73120, AIF73119, AIE89862, AIE89861, AIE89860, AIE89859, AGR42654, AHI18145, AGZ84629, AGZ84628, AGU68254, AFP95928.2, AGO86375, AFR68657, AEJ87340, AGE13899, AGC02839, AFV95076, AFR68851, AFJ15567, AFP19893, AFP19891, AEU04526, AEU04525, AET97637, AFB77241, AFB77240, AEX15510
		ROCV	3	YP_009553341.1, ATG32103.1, AAV34158.1
		ZIKV	97	ASK51714.1, AMN14620.1, ARB07953.1, AOY08523.3, AMN14619.1, AOY08525.1, ANH10698.1, AOY08535.1, ANZ46736.1, ASW34087.1, AOY08542.1, APB03020.1, AOO53984.1, APB03022.1, AOS90220.1, ATI25888.1, APB03024.2, AOY08538.2, AVZ25035.1, A0A024B7W1.1, ARU07074.1, AMA12084.1, ARB07976.1, AXF50016.1, APO39237.1, AYV62653.1, YP_009428568.1, AOO53982.1, AOS90223.1, AMU04506.1, AMS00611.1, ANN44857.1, ANO46306.1, AOY08526.2, ANO46305.1, ARB07934.1, AMR39831.1, ARB07967.1, AML81028.1, AOY08520.2, AOG18295.1, AXF50014.1, ANO46307.1, AMQ48982.1, AMC13911.1, AMK79468.2, APY24198.1, ALX35659.1, APO36913.1, ART29823.1, AOY08518.2, APG56457.1, AOX49264.1, AMZ03556.1, AMQ48981.1, ANO46303.1, ARU07076.1, AMD16557.1, ARB07932.1, AQU12485.1, ARU07075.1, ANO46308.1, AMK49165.1, APY24200.1, AMX81919.1, AYV74916.1, AOP04276.1, AOX49265.1, AMM39805.1, ARB07980.1 AND01116.2, AYI50273.1, AYV74932.1, ANF04752.1, AYI50274.1, AMB37295.1, AOY08539.3, AOG18296.1, ASU55416.1, ALL27019.1, APH11501.1, ALU33341.1, AQS26814.1, BBA85762.1, ANG09399.1, AOC50652.1, AOI20067.1, APH11549.1, AQM74761.1, APH11511.1, AUI42329.1, APH11492.1, ATG29287.1, AYV74919.1, ATG29291.1, ANB66184.1, APH11497.1
	Aroa group	AROAV	2	YP_001040004.1, AIU94739.1
	Kokobera group	KOKV	2	YP_001040007.1, AAV34157.1
	Kedougou group	KEDV	3	YP_002790882.1, AAV34156.1, ABI54477.1
Insect-specific Flavivirus related MBFV	Nounane group	LAMV	2	YP_009056848.1, ACR56717.1
		ILOV	1	YP_009056847.1
		MMV	2	ATL63283.1, ATD87259.1
		CHAOV	4	YP_005454257.1, AFD50747.1, AFE86156.1, ACP43327.1
		DONV	2	YP_005352889.1, AFD61560.1
		BJV	5	AUS91146.1, BBN23635.1, AGS41451.1, ABW74531.1, BBN23634.1
		NOUV	3	YP_009345019.1, ACN73462.1, ACM68470.1

Table 2: Continue

Group of viruses by vector	Virus group	Abbreviation		GenBank No.
Mosquito-borne Flavivirus group	Yellow fever group	YFV	79	AFU76909.1, AAS78199.1, AFU76908.1, AFU76910.1, AFU76912.1, QBA30232.1, AFU76903.1, AMZ00441.1, AFU76911.1, AFU76905.1, AXY92190.1, AMZ00440.1, AFU76904.1, AAA99713.1, AMZ00439.1, AXY92189.1, AFU76906.1, AXY92187.1, AXY92186.1, AVT50844.1, NP_041726.1, AVD96634.1, AFQ32463.1, AGO04419.1, AAZ31436.1, AAC54267.1, AAC35906.1, AEQ75453.1, AFQ32465.1, AAC35907.1, AFQ32464.1, AAC35908.1, AAC35899.1, AAC35900.1, AAZ07885.1, ACN41908.1, AXY92188.1, ASY08194.1, AFH35044.1, AIZ07887.1, ASY08198.1, AIZ07885.1, AIZ07888.1, ASY08196.1, AFH35033.1, AFH35037.1, ASY08193.1, ASY08199.1, ARM37842.1, AFH35042.1, ASY08192.1, ADK47994.1, ASY08202.1, ASY08195.1, AVQ67777.1, AWB15003.1, AWB15001.1, ASY08200.1, AWK57896.1, ARQ19026.1, AFH35041.1, AWB14999.1, AFH35034.1, AWB14998.1, AFH35036.1, ASY08189.1, AIZ07886.1, AWB14992.1, ATN45408.1, AFH35038.1, AWB14996.1, AFH35039.1, AWB14993.1, AFH35040.1, AWB15000.1, AIZ07889.1, AWB15002.1, AWB14991.1, AFH35043.1
		SEPV	4	YP_950478.1, ABI54479.1, AAV34159.1, ABI23562.1
		WSLV	7	YP_002922020.1, ASC48302.1, ASC48301.1, ASC48300.1, AFK88571.1, ABI54474.1, ACH70606.1
	Edge Hill group	EHV	2	YP_009256192.1, ABI54476.1
		JUGV	2	YP_009344969.1, ABI54482.1
		UGSV	2	YP_009344968.1, ABI54481.1
		BOUV	2	YP_009344961.1, ABI54473.1
		BANV	1	ABI54472.1
		SABV	2	YP_009344967.1, ABI54478.1
Vertebrate-specific Flavivirus group	Entebbe bat group	ENTV	4	YP_950477.1, AKP24039.1, AAV34153.1, ABI23561.1
		SOKV	2	YP_009126875.1, AIU94745.1
		YOKV	3	QAU20974.1, NP_872627.1, BAC79364.1
	Modoc group	APOIV	2	NP_620045.1, AAF70829.1
		MODV	2	NP_619758.1, CAC82912.1
		JUTV	2	YP_009126871.1, AJA91182.1
	Rio bravo group	RBV	3	NP_620044.1, AFG73004.1, AAF37322.1
		MMLV	2	NP_689391.1, CAC82713.1
		PPBV	2	YP_009350101.1, AJA91183.1
Classical insect-specific Flavivirus group		QBV	3	AYN44130.1, YP_002884239.1, ACQ55297.1
		CTFV	2	YP_009553010.1, AVM80372.1
		CxFV	37	AYW01749.1, AXQ04779.1, YP_899469.2, ASL70040.1, AWX63832, AWX63831, ARQ79718.1, AMK47901, API61892.1, API61891.1, AOX47506, AML84517, ACV04604.1, AGS14350.1, BAM74427, BAM74426, BAM74425, BAM74424, BAM74423, BAM74422, BAM74421, BAM74420, BAM74419, BAM74418, BAM74417, AFI49451.1, AFE86161, AFE86160, AFE86159, AFE86158, AFE86157, ADX94786, ACT09359, ACL54956, ACJ64914, BAG06229, BAF36740.2
		PCV	3	YP_009344962.1, ALC76577.1, AGG76014.1
		NAKV	1	YP_009259488.1
		NIEV	3	APR74286.1, YP_009041466.1, APR74285.1
		CuCuV	1	AOR81460.1
		KRV	3	NP_891560.1, AAO24117.1, AAO24118.1
		CFAV	6	YP_009259257.1, AWK27463.1, AWJ96713.1, AOE46763.1, ACV04606.1, AIM49245.1
		AEFV	4	YP_003029843.1, AIM49244.1, BAH83667.1, AGJ91136.1
		MFV	1	YP_009351861.1
		XFV	1	YP_009350102.1
		HANKV	2	YP_009259489.1, AEY84723.1
		LTNV	1	ATL63282.1
		SbFV	1	AZB73874.1
		HaCV	1	ASD50161.1
		KRBV	3	ARD06144.1, YP_009388577.1, ASD50159.1
		McPV	1	YP_009389296.1

Software: Amino acid sequence alignment was performed by the MAFFT ver. 7 (Multiple Alignment with Fast Fourier Transform)[16]. The structures that were used in the analysis: 2j7w, 4k6m, 4v0r, 7d6m, 7d6n, 2jlx and 7jno, which were obtained from the Protein Data Bank[17]. Visualization and analysis of the NS3 and NS5 structures were done with the UCSF ChimeraX 1.1[18].

RESULTS

Repeats in viral proteins: Sequence analysis has revealed repeats from two to six or more repeating amino acid positions. It has been established that the repeats differ in genus-, group-, species- and subtype-specificity. Repeats are found in structural and non-structural viral proteins, while repeats in NS3 and NS5 (as well as NS1, not considered in this study) are genus-specific. This study focuses on genus-specific repeats and some unique group-specific repeats are also highlighted.

According to the present study results, 6 genus-specific repeats were found, either to be concentrated in the vicinity of conservative motifs or to be part of a conservative motif. In the NS3 protein, a genus-specific repeat Pro324 was found in the helicase domain. In the NS5 protein, a genus-specific repeat was found in the methyltransferase domain-Gly85; in the RNA-dependent RNA polymerase domain, four genus-specific repeats were found Gln352, Lys461, Asp534 and Asp664.

NS3 Protein: Significant repeats were found in the helicase domain of the NS3 protein. The NS3 helicase domain of Flavivirus contains 6 sites that include significant amino acid repeats. The 5 of these 6 regions overlap with the conserved helicase motifs Ib, II, III, V, VI and have the same name (Fig. 1).

The basis of Ib's motif (267-277) is the invariant positions of CH*T267-270, which are flanked on the right by a double repeat of RRLL (Fig. 2, 3a). The RRLL274-277 repeat is conserved among the mammalian tick-borne flavivirus group, as well as the Meaban and Saumarez Reef viruses. Single repeat LL276-277 is characteristic of the Dengue and Entebbe bat group of viruses. This repeat is also found in the Japanese encephalitis virus and Ntaya groups. An atypical repeat RRMM274-277 was found in the Apoi virus. Structurally, motif Ib is localized along the ssRNA binding tunnel and forms an α-helix and a β-strand.

A 2016 study of DENV's conserved surface helicase residues suggests that R274 remains are needed for functional binding to viral RNA and RNA-stimulated ATPase activity. A characteristic feature of amino acid position R274 is that in the classical insect-specific flavivirus group, the arginine residue is completely absent, while in the Insect-Specific related MBFV group the arginine residue is retained for all viruses except Nounane (Fig. 2).

Motif II (289-309) has conserved MDE*H and A*RG patterns (Fig. 2). The AA301-302 repeat is located between two conserved patterns and is conserved, with some exceptions, in all groups, with one exception the Entebbe bat virus. In turn, the Entebbe bat virus group contains a unique group-specific repeat QQ307-308 (Fig. 2). Structurally, motif II is localized to the right of the ssRNA binding tunnel and forms an α-helix (Fig. 3b).

In this study, the variations of the classical DEAD catalytic residues of the NS3 helicase are observed. Positions DEAH 290-293 of motif II are conserved for the mammalian tick-borne flavivirus group, mosquito-borne flavivirus group, insect-specific related MBFV and vertebrate-specific flavivirus.

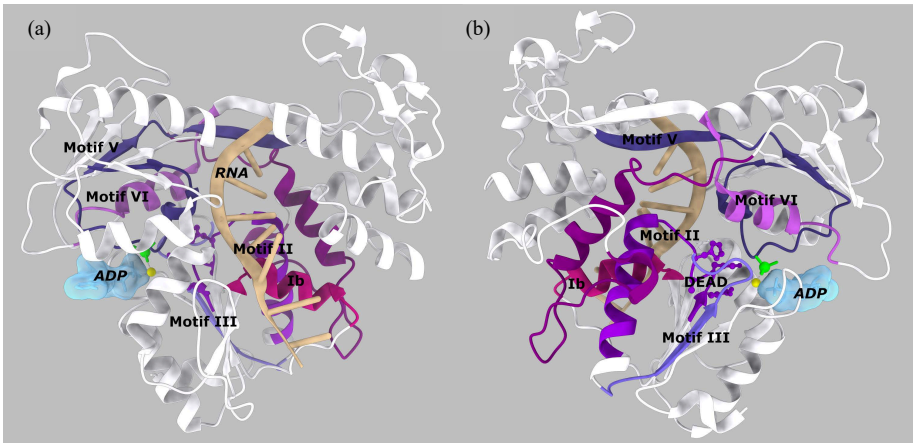

Fig. 1(a-b): Location of conserved motifs in the NS3 helicase domain of DENV4, (a) Front view and (b) Rear view

Conserved motifs are shown in color and RNA and ADP binding sites are indicated, DEAD catalytic residues are indicated as atoms, Mn and VO4 are the yellow and lime atoms, respectively and the pdb 2jlx has been used

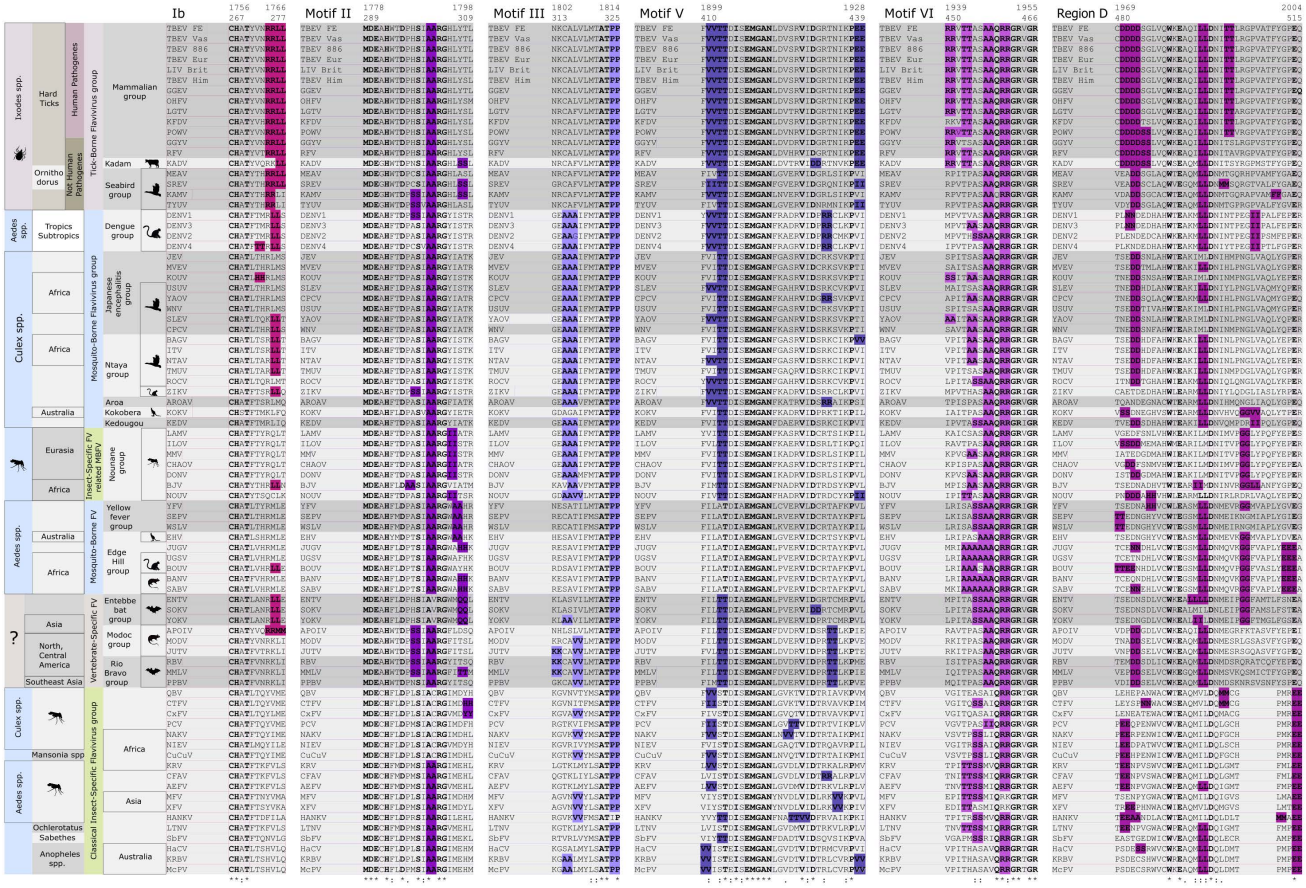

Fig. 2: Repeats of helicase domain NS3 in the genus *Orthoflavivirus*

Amino acid repeats are highlighted in color, conservative positions are indicated in bold and with an asterisk symbol and virus names are given in accordance with the ICTV taxonomy

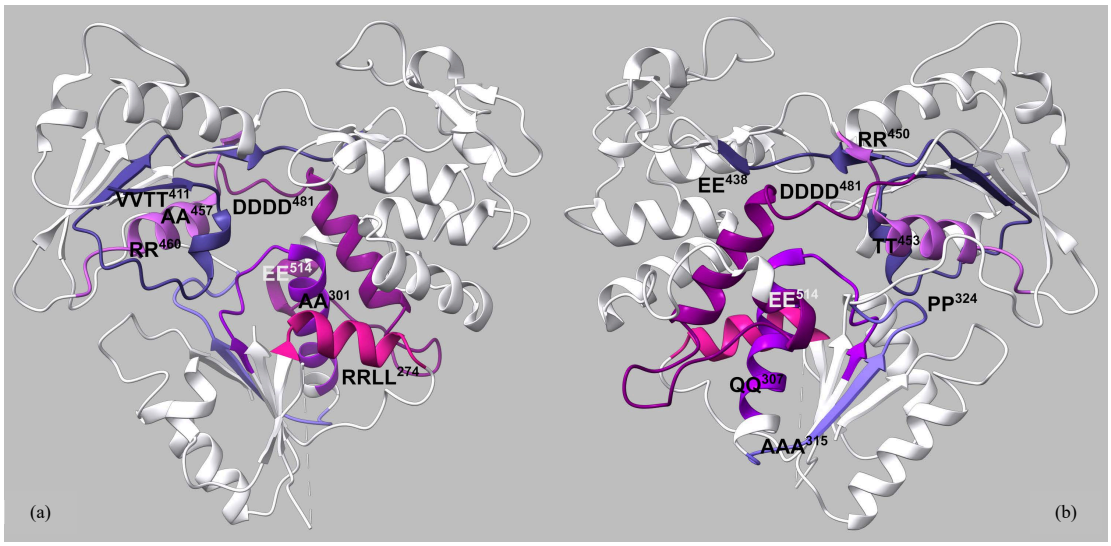

Fig. 3(a-b): Location of repeats in the NS3 helicase domain

Conserved motifs are presented in color and the position of genus-specific and group-specific repeats is indicated. The pdb 7jno has used

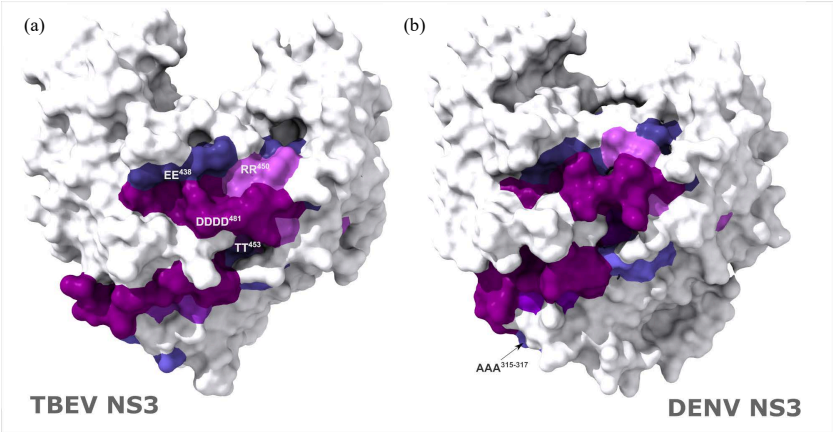

Fig. 4(a-b): Repeats of a hypothetical NS3 helicase domain binding site, (a) TBEV NS3 and (b) DENV NS3

In the Seabird and Kadam viruses subgroup of the tick-borne flavivirus group, the DEGH combination is retained, while in the classical insect-specific flavivirus group, the DECH combination is stable.

The group-specific amino acid substitution at position 292, Ala-Gly-Cys, is likely to be important for helicase function (Fig. 2). Analysis of modifications of the helicase catalytic center represents a challenge for new studies.

Motif III is based on invariant positions ATPP322-325, with the PP322-325 repeat being the most conserved in the flaviviruses (Fig. 2). Structurally, motif III is the 5 β-strand of a six-membered parallel β-sheet, with the P324-325 repeat lying in the same plane as the catalytic DEAD residues (Fig. 1b, 3b). Within motif III, a repeat A315-317 was found, characteristic of the Mosquito-Borne Flavivirus group, except for the Kokobera virus, with some variations that also occur in the Nounane group (Fig. 2). Structurally, repeat A315-317 is located at the bottom of the helicase domain (Fig. 3, 4a-b).

Motif V (410-439) is based on the invariant region EMGAN418-422 and five conserved positions are dispersed within the region (Fig. 2). On the left, the invariant region EMGAN418-422 is flanked by a doubled repeat VVTT 411-414, with T414 being a conserved position. In many groups, the repeat of VVTT 411-414 is reduced to TT413-414, in a single instance to VV411-412. The Yellow fever virus and Edge Hill virus groups are completely free from this repeat.

Motif V has a complex structural organization. On the front side of the helicase, it includes two β-strands of a five-membered parallel β-sheet and an adjacent α-helix and on the back side it forms a twisted β-hairpin. The motif V is critical for flavivirus helicase function; the contribution of the VVTT 411-414 repeats has not yet been clarified (Fig. 3a). However, for the E438-439 repeat from motif V, it was possible to find a putative connection with other repeats of the hard ticks group, tick-borne flavivirus group, forming an interface on the back side of the helicase, the possible role of which is described above (Fig. 4).

Motif VI (450-466) is a sequence region with an invariant pattern QR*GR*GR459-466, which overlaps with conserved repeats AA457-458, RR460-461 (Fig. 2). Moreover, the AA457-458 repeat is characteristic of all groups except the classical insect-specific flavivirus, while the RR460-461 repeat is conservative for the entire genus with the exception of two representatives from the Asia classical insect-specific flavivirus group. In the mammalian tick-borne flavivirus group and Kadam group, repeats form a cluster from positions 450 to 461 inclusive. To the already described repeats AA457-458 and RR460-461, RR450-451 and TT453-454 have been added. Structurally, all four repeats form an α-helix, which is localized between ssRNA and ADP (Fig. 3, 4a).

The last region that has amino acid repeats but is not associated with conserved helicase motifs, is named region D, according to a distinctive repeat of DDDD481-484 from the hard ticks sub-group of the tick-borne flavivirus group (Fig. 2).

Region D (480-515) was found in this study. This region contains four conserved positions-W490, E492, D498 and E514. It is known that replacement of the E492 residue leads to disruption of the activity of RTPase and ATPase and the E492 residue also serves to maintain the integrity of the NS3 helicase structure (Fig. 2). The group-specific repeats DD481-484, LL496-497, TT501-502 and EE514-515 are of greatest interest (Fig. 2). Currently, the significance of region D among flaviviruses has not been studied.

Group-specific repeats of amino acids from hard ticks of tick-borne flavivirus group E438-439, R450-451 and D481-484 localized on the back side of the helicase form a deep cavity (Fig. 4a). Probably, this structural region serves as a binding

site for some macromolecule characteristic of the life cycle of the TBE group of viruses. It is characteristic that for the DENV group, the same region of space does not have clusters of repeats (Fig. 4b).

One of the common viral strategies is based on the use of cellular RNA helicases in viral replication. Cellular RNA helicases are capable of acting as a scaffold for alternative protein-protein interactions. By analogy with cellular RNA helicases, viral RNA helicase can also serve as a scaffold for protein-protein interactions. Based on the assumption that invariant repeats preserved in the viral population are more often found at the site of contact of two macromolecules, such as protein-protein, protein-RNA and protein-cofactor, then, significant protein interfaces can be determined by clusters of repeats within the same area.

This study revealed that structurally significant regions of helicase Ib, VI and D have group-specific repeats that separate the group of tick-borne viruses from the mosquito-borne viruses. In contrast to previously described amino acid repeats characteristic of vertebrate viruses, the classical insect-specific flavivirus group has virtually no unique repeats of its own. As a rule, amino acid repeats of vertebrate viruses are reduced or absent in the classical insect-specific flavivirus group. The exception is the deletion of sequence from positions 504-510 and the unique repeat EE514-515 in region D (Fig. 2 and 3). This fact may indicate the existence of modified, truncated versions of the NS3 helicase. This may be valuable for comparative functional analysis with other flavivirus proteins.

Methyltransferase domain of NS5 protein: To indicate a specific region in the MTase domain, it is customary to rely on the basic elements of secondary structure-α-helices and β-sheets in the order they appear in the polypeptide chain. Group- and genus-specific repeats were found in the vicinity of the three indicated binding pockets (Fig. 5).

The GTP binding site is delimited by 1st and 2nd α-helices in the region 12-25. The group-specific repeat EEFF22-25 is a distinctive feature of the mammalian group viruses. The exceptions are Japanese encephalitis virus and Murray Valley encephalitis virus, which also contain a repeat of EEFF22-25. Structurally, the EEFF22-25 repeat forms the upper arch of the GTP binding pocket (Fig. 5, 6).

The binding site for the cofactor of the SAH reaction is conserved among flaviviruses. This site is formed from β-strand 2 and α-helix 5. In the region 79-93 there is a conserved motif DLG*GRGGW, which includes the genus-specific repeat GG85-86 (Fig. 5). Structurally, the GG85-86 repeat forms contact with the oxygen atoms of the SAH molecule (Fig. 6).

The repeat YYAA89-92 is adjacent to the conservative DLG*GRGGW motif and varies significantly in virus groups. Repeats GG85-86 and YYAA89-92 form α-helix 5. Structurally, the YYAA89-92 repeat forms the upper arch of the SAH molecule binding pocket (Fig. 5, 6).

The RNA substrate binding site is formed from two conserved regions. The first region (143-159) is formed by β-strand 5 and α-helix 8 and the second region (204-222) is formed by 9 α-helix, 7th and 8th β-strands.

At the sites 143-159 there is a conserved CD*GE motif, which is the catalytic site of the enzyme. The conserved CD*GE motif is adjacent to the SSSS150-153 repeat. The SSSS150-153 repeat is extremely variable and occupies from 2 to 6 positions in different groups of viruses (Fig. 5). Repeat SSSS150-153 is characteristic of all groups except the classical insect-specific flavivirus group. Structurally, the SSSS150-153 repeat in the MTase active site forms the lower roof of the binding pocket of the GTP molecule (Fig. 6).

The final active site repeat cluster of the MTase domain includes the GGG204-206 repeat located upstream of the conserved SRNS213-216 motif and the YY221-222 repeat following the catalytic residue Glu219. The GGG204-206 repeat varies from 2 to 3 positions in different groups. Repeat YY221-222 was found in the TBFV, yellow fever and classical insect-specific flavivirus groups (Fig. 5). Structurally, both repeats are located in the core of the RNA binding site and are localized in the adjacent β-strand 6 and 7 of the same β-sheet (Fig. 6).

Region between NS5 protein domains: The IDI includes polymerase and methyltransferase regions located immediately at the domain boundary. In this study, structural analysis reveals that the binding site of NS5 protein to STAT2 protein is located in IDI.

Activation of STAT2 is a necessary step for the induction of type I interferon (IFN) signaling. The innate immune response mediated by type I IFN constitutes the first line of host defense against viral infection. To successfully infect, viruses suppress the activation of IFN responses triggered by STAT2.

However, the solved structures of 6ux2 and 6wcz do not provide a complete understanding of the contact surface of the NS5 and STAT2 proteins, since atomic coordinates are missing for extended sections of the STAT2 chain. For example, from the N end in the 6wcz structure, there are no coordinates for 139 amino acid positions that correspond to the N domain. On the contrary, in the 1yvl structure, the atomic coordinates corresponding to the N-domain of the homologous STAT1 protein are known.

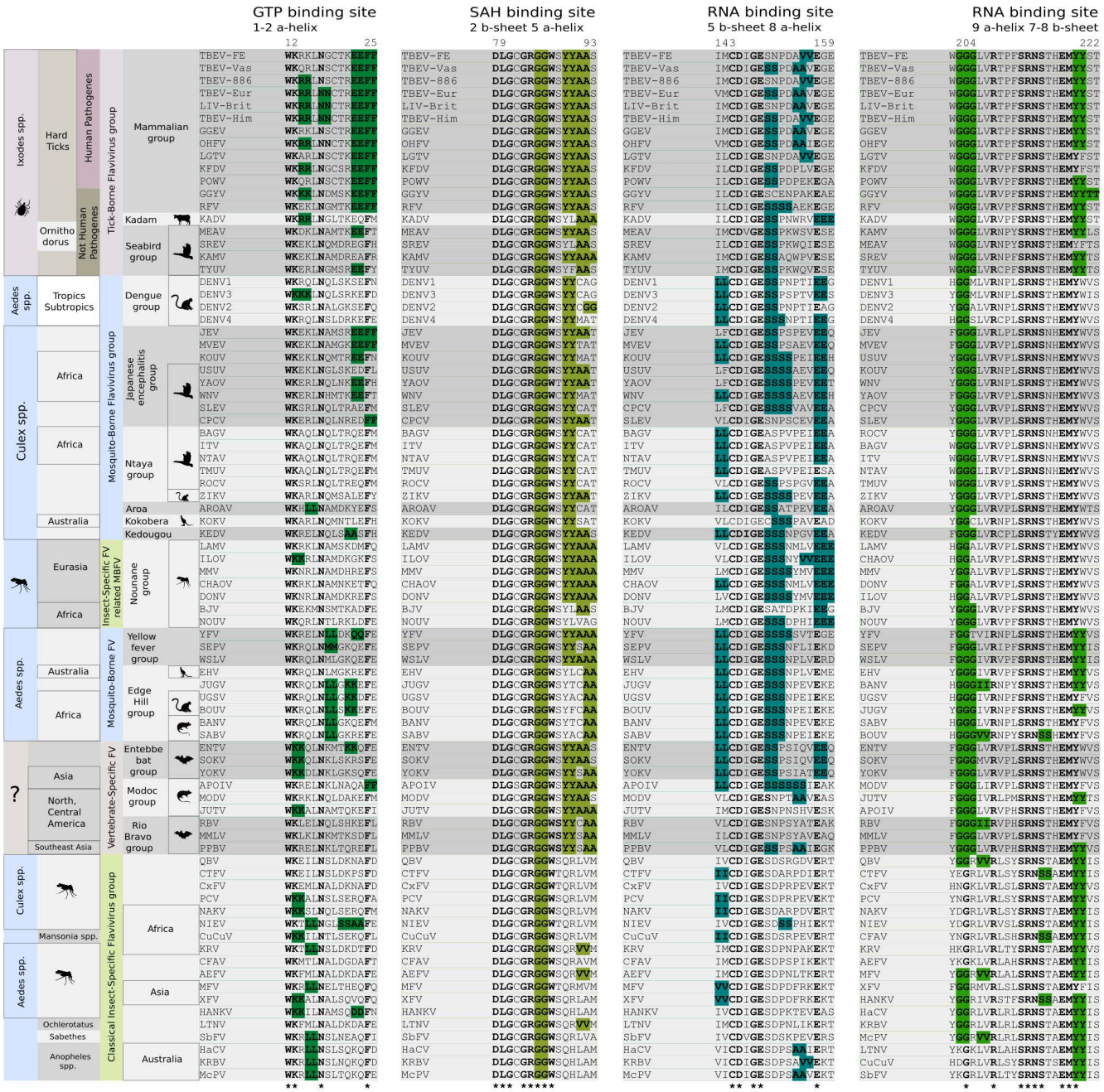

Fig. 5: Repeats of methyltransferase domain NS5 for the genus *Orthoflavivirus*

Conservative motives signed, repeats of amino acids are highlighted, conservative positions are in bold and an asterisk and abbreviations are given according to ICTV taxonomy

The IDI repeat study may help identify the unique flavivirus motifs involved in binding to STAT2. The IDI repeat group includes four structural elements. On the polymerase side, this is the nuclear localization sequence βNLS and the F motif. On the methyltransferase side, 2 regions with repeats were found that have not previously been described in the literature.

The motif (343–362) overlaps with the βNLS region of the RdRp domain and is named the NLS motif. This motif includes the following conserved regions: MTD344-346 and KVDT359-362. The NLS motif contains repeats: TT347-348 and QQ352-353. The TT347-348 repeat is conserved in all groups except the classical insect-specific flavivirus. QQ352-353 is a genus-specific repeat, invariant in all groups of

flaviviruses. The NLS motif is an α-helix that is located in the same plane of space as the three IDI motifs (Fig. 7-8).

The motif (451-473) coincides with the F motif of the RdRp domain and has the same name. Motif F contains 23 positions, 11 of which are conserved. Motif F contains repeats MM455-456 and KK461-462. The MM455-456 repeat is invariant in all groups except the vertebrate-specific Flavivirus and classical insect-specific flavivirus groups. The KK461-462 repeat is a genus-specific repeat that is enhanced by the KK458-459 repeat in the classical insect-specific Flavivirus group (Fig. 7).

Structurally, the motif F has two stable conformations. The first conformation is found in the structure of the full-length NS5 protein with polymerase activity, with the motif F forming a β-hairpin that is deeply embedded in the interdomain space and serves as a steric constraint when the domains approach each other. In the second conformation, the motif F is brought outside the protein globule and removed from the IDI, this position is associated with the presence of methyl transferase activity 43 (Fig. 8a-b). In the first conformation, the motif F is involved in the interaction with STAT2 (Fig. 9).

The cISFV-delete region (37-59) contains an extended deletion of 14 amino acid positions in the classical insect-specific flavivirus group (Fig. 7). The deletion is followed by the invariant region VSRG55-58 and the specific repeat GG58-59 found in the classical insect-specific flavivirus group. Structurally, the cISFV-delete region represents either a disordered loop in the interdomain space (Fig. 8a) or forms an α-helix (Fig. 8b).

The GG-torsion is a region of the sequence of 11 amino acids (104-113), three of which are strictly conservative. The GG-torsion contains two repeats-GG106-107 and E111-112. Repeat GG106-107 is maintained in many groups; with exceptions of Seabird, mosquito-borne flavivirus and vertebrate-specific flavivirus (Fig. 7).

Structurally, the GG-torsion forms a hairpin and takes on distinct conformations depending on the orientation of the NS5 protein domains (Fig. 8). The GG-torsion is located in the region of contact with the STAT-2 protein (Fig. 9). In the absence of any data on the sequence that corresponds to the GG-torsion, the role of this region remains unclear and needs to be further explored.

RNA-dependent RNA polymerase domain: The polymerase domain contains seven classical motifs: A, B, C, D, E, F, G, which are conserved within the genus and are typical for RNA-dependent RNA polymerases. The catalytic triad of RdRp is formed by aspartic acid residues D535 and DD664-665.

Two regions with genus-specific repeats were found that coincide with catalytic motif A and C (Fig. 10). Structurally, motif A and C are located in the center of the palm subdomain (Fig. 11).

Motif A (528-541) contains the genus-specific repeat DD534-535 and the repeat GG528-529 is conserved for all groups except Edge Hill group, vertebrate-specific flavivirus and classical insect-specific flavivirus. Motif A repeats are adjacent to a cluster of conserved residues AGWDT537-541 (Fig. 10). The conserved cluster AGWDT537-541 structurally forms a loop and is located in the same region with three residues of the catalytic triad (Fig. 11).

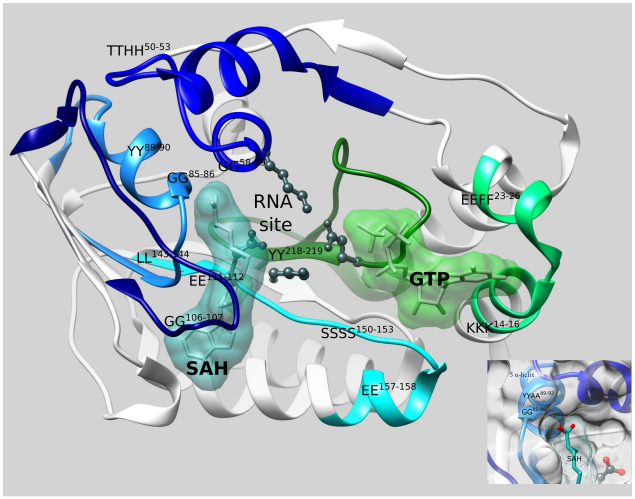

Fig. 6: Location of repeats in the NS5 MTase domain

Conserved motifs are presented in color and the position of group- and genus-specific repeats is indicated, catalytic residues are indicated as gray atoms, below right is a fragment of the SAH cofactor binding site in the MTase active site and the pdb 4v0r has used

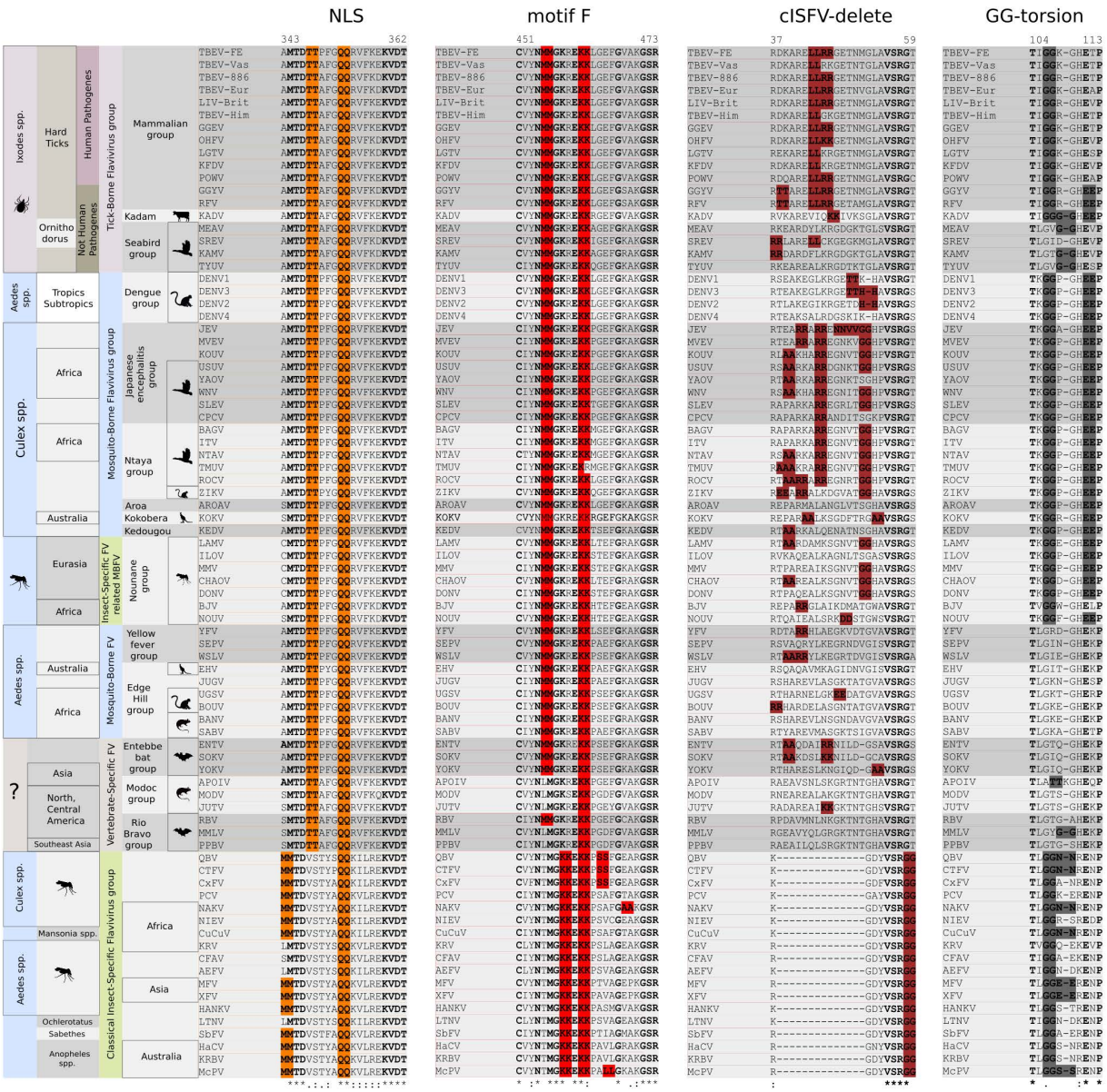

Fig. 7: IDI NS5 repeats of the genus *Orthoflavivirus*

Conserved motifs are indicated, repeats of amino acids are highlighted, conserved positions are indicated by bold script and asterisks, the NLS and F regions belong to the RdRp domain, the cISFV-delete and GG-torsion regions belong to the MTase domain and abbreviations are given according to ICTV taxonomy

Motif C (663-673) contains the catalytic repeat DD664-665 and repeats VV667-668 and DD672-673 (Fig. 10). The VV667-668 repeat is preserved in all groups except the Edge Hill group and in two groups specific only to insects it merges into one extended repeat DDVV664-668. The DD672-673 repeat is conserved in the TBFV, Nounane, Yellow fever, Entebbe bat group and is structurally an α-helix. Repeats GG528-529 and DD672-673 lie in the same region and form the lower border of the palm subdomain (Fig. 11).

DISCUSSION

The genomic organization of flaviviruses is similar, but their host range and modes of transmission differ fundamentally. Most viruses in the genus *Orthoflavivirus* are transmitted horizontally between hematophagous arthropods and vertebrate hosts, for this reason, they are called viruses with two hosts[19]. Dual-host flaviviruses are further divided into mosquito- and tick-borne viruses that infect vertebrates. In turn, the Tick-Borne Flavivirus group (TBFV) is separated into

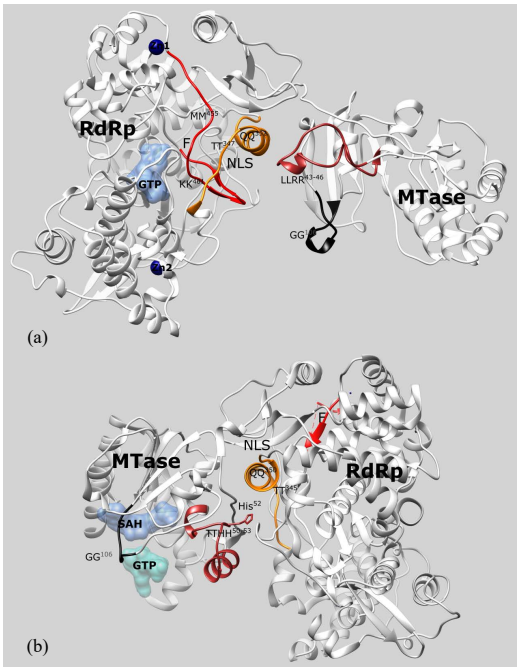

Fig. 8(a-b): IDI repeats location in the NS5 structure, (a) Foreground: The full-length model of the TBEV NS5 protein is taken from[27] and (b) Background: Structure of NS5 DENV3, pdb 4v0r

Structural elements with conservative genus- and group-specific repeats are highlighted in color. From the RdRp side: NLS-orange, F-red. From the MTase side: cISFV-delete-brown, GG-torsion-black. Cofactors SAH-blue, GTP-green, presented as a surface and Zinc ions are represented as a dark blue sphere

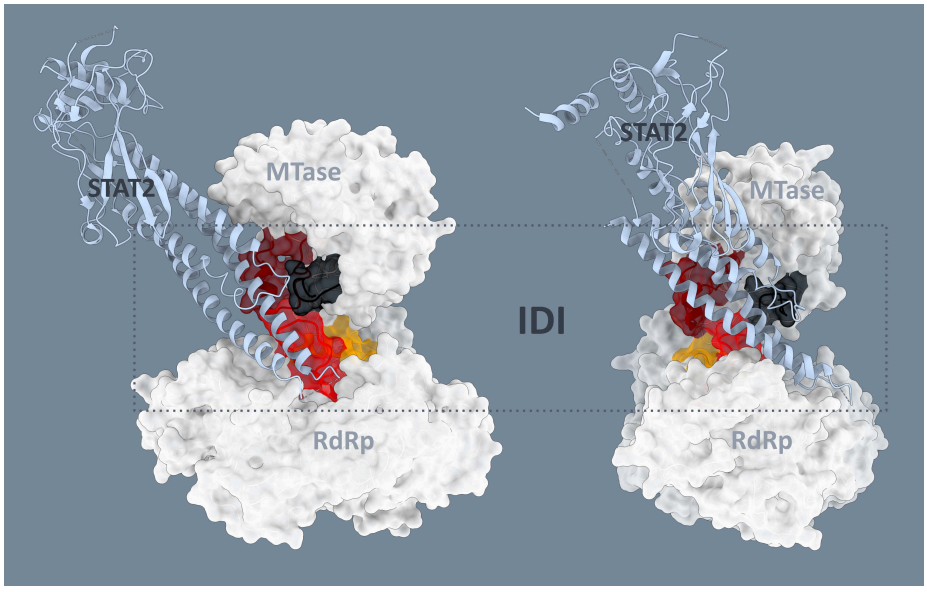

Fig. 9: Structure of the full-length NS5 protein in complex with STAT2

NS5-white, surface representation, STAT2-blue, representation of secondary structure, NS5 protein motifs: NLS-yellow, F-red, cISFV-delete-brown, GG-torsion-black

flaviviruses pathogenic for humans associated with hard ticks of the genus Ixodes and apathogenic flaviviruses associated with soft ticks of the genus *Ornithodorus*[20]. Mosquito-Borne Flavivirus group (MBFV) are separated into viruses associated with mosquitoes of the genus *Culex* (ornithophilic) and viruses associated with mosquitoes of

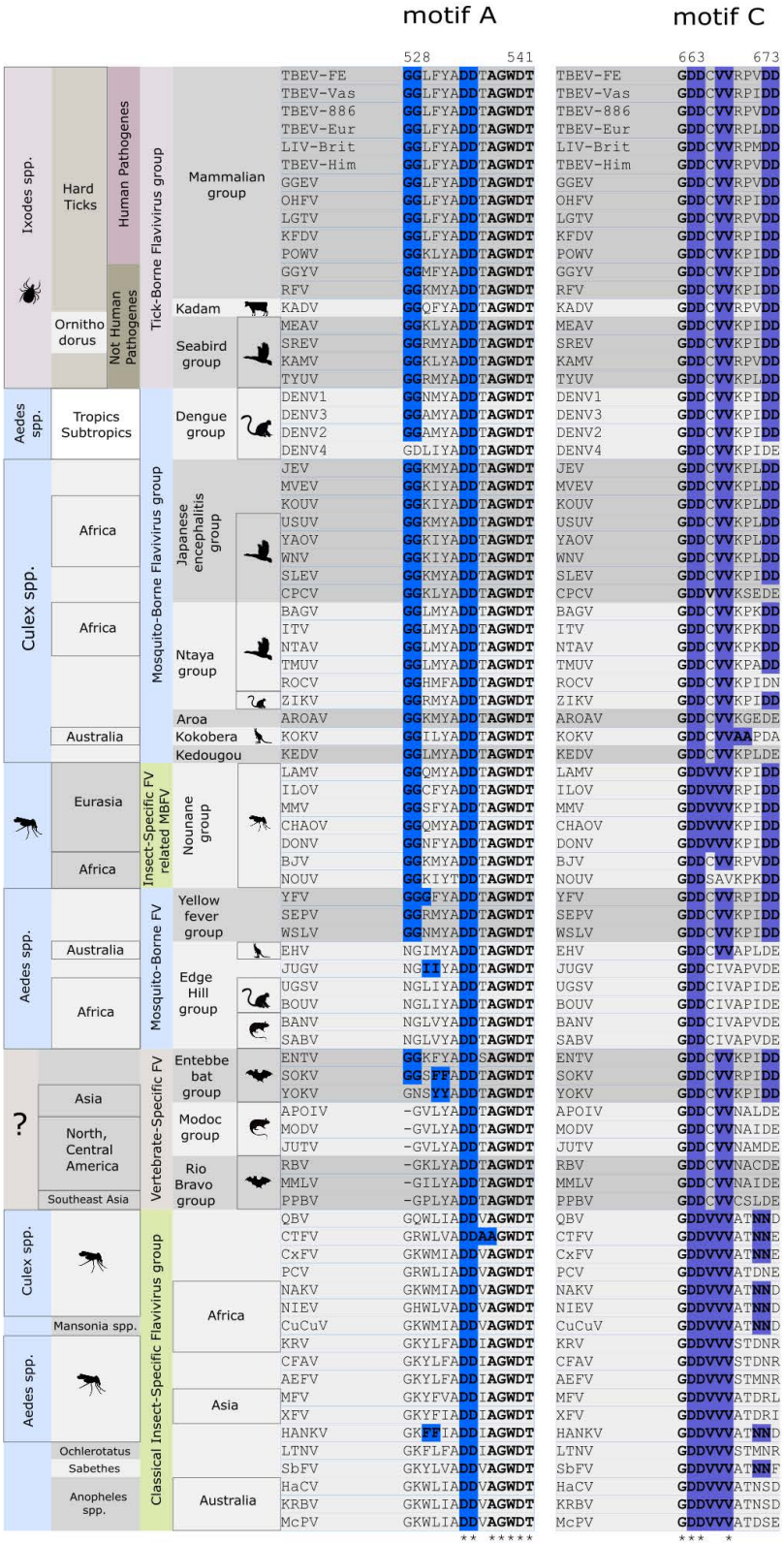

Fig. 10: Repeats of RdRp NS5 the genus *Orthoflavivirus*

Conserved motifs are indicated, repeats of amino acids are highlighted, conservative positions are in bold and an asterisk and abbreviations are given according to ICTV taxonomy

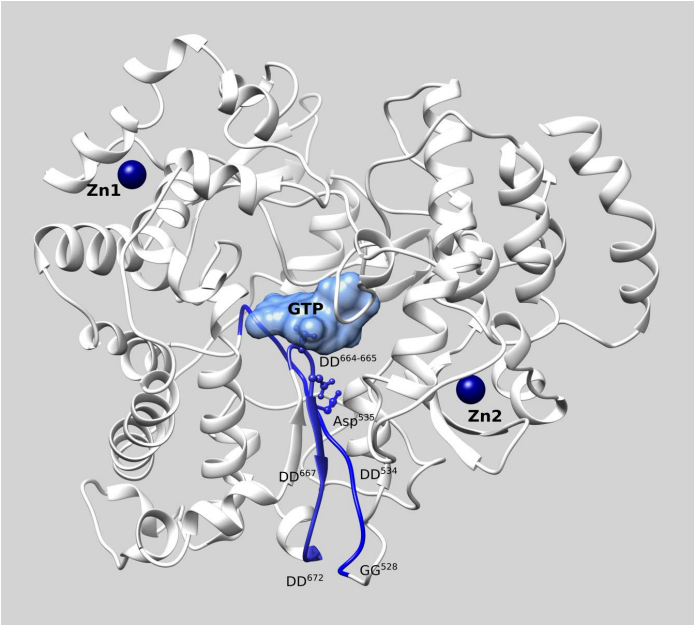

Fig. 11: Location of repeats in the NS5 RdRp domain

Conserved motifs are presented in color and the position of group- and genus-specific repeats is indicated, catalytic residues are indicated as blue atoms, zinc ions are represented as a dark blue sphere and the pdb 4v0r has used

the genus Aedes (mammophilic). However, not all flaviviruses circulate between arthropods and vertebrates; some of them are specific to insects[19].

Vertebrate-Specific Flaviviruses make up the VSFV group. These include rodent viruses (e.g., Modoc virus; MODV) and bat viruses (e.g., ENTV and RBV). In recent years, Insect-specific flaviviruses (ISFV) have attracted great interest. This growing group of species has a global distribution and is divided into two subgroups by phylogenetic analysis. The classical insect-specific flavivirus group was discovered first and differs from all known flaviviruses. At the moment, the insect-specific flaviviruses from the MBFV group are phylogenetically related to Mosquito Flaviviruses Infecting Vertebrates and are represented by viruses of the Nounane group[3]. The phenomenon of different host ranges of flavivirus can be clarified by comparing their genomes. The study of amino acid repeats makes it possible to trace the structural relationship between changes in protein folding while maintaining its main function and the adaptation of the virus to new conditions.

Viral proteins, viral RNA and cellular host factors combine to form a replicative complex found in virus-induced vesicles. Replication complex macromolecules are involved in RNA-RNA, RNA-protein and protein-protein interactions. Viral RNA replication efficiency depends on these interactions[21,22]. In the present work, the analysis of viral proteins was carried out based on positions in the amino acid sequence where repeating amino acids occur. A comparative analysis of flavivirus genomes found that genus-specific repeats occur in the NS3 and NS5 proteins.

Until now, structural units of a protein molecule have included domains, motifs and conserved amino acid positions. A protein domain usually corresponds to a specific enzyme function implemented in a living system. A protein motif has a local structure and function within the molecule. However, the amino acid composition of a motif varies significantly from species to species. A motif in a wide range of species may not have conserved amino acid positions at all. The question arises, by what mechanisms do organisms distant in origin retain common motifs in proteins? The difficulty of detecting a common motif core has led to the search for the structural basis of the motif. In this study, an attempt was made to search for a structural unit of the motif, for which conservative amino acid repeats were proposed.

It is known about the motifs of the NS3 protein that motif II functions as an ATP-binding domain, motif III is associated with the unwinding of RNA chains and motif V interacts with both RNA and ATP. The motif V and Ib are involved in RNA binding and possibly helicase activity. Motif VI (Fig.1) is responsible for protein interaction with RNA during ATP unwinding and hydrolysis[11,23,24]. The two domains of NS5 protein undergo a rearrangement in the life cycle of the

virus[25-27]. The crystal structures of the protein provide several possible variants of domain orentation[28,29]. A number of models have been proposed for interactions between NS5 domains to explain the movement of the RNA product into the MTase active site[30-33]. Crystallographic studies of recent decades have expanded the understanding of the spatial organization of the full-length flaviviral NS5 protein. It turned out that the interdomain interface (IDI) region plays a central role in the process of switching the enzymatic activities of this protein. The region between the domains of the NS5 protein forms the functional domain.

The IDI includes an extended interdomain linker that provides additional degrees of freedom for domain reorientation[29]. However, due to the high mobility of IDI, the structural elements of this region remain poorly studied[34-36]. The recently solved structure of 6wcz shows a complex of the full-length DENV NS5 protein with a fragment of the STAT2 protein. The authors suggest that there is a conserved mechanism of interaction with the STAT2 protein among flaviviruses[37].

Five genus-specific repeats were identified in the NS5 protein: GG85-86, QQ352-353, KK461-462, DD534-535 and DD664-665. Repeat-rich regions were found to overlap with or be adjacent to the classical methyltransferase (MTase) and RNA-dependent RNA polymerase (RdRp) motifs of flaviviruses. Repeat-rich regions in the structure of the NS5 protein are localized in three non-overlapping planes of space. Clusters of repeats were found in the vicinity of the active centers of MTase and RdRp and the region between the domains.

In previous work, Belikov and Potapova examined TBEV strains that are pathogenic and weakly pathogenic for humans[38,39]. As a result, differences were found in the conformation of the full-length NS5 protein. Four substitutions in the palm and thumb subdomain were found to have an allosteric effect on IDI conformation. This study revealed that in a group of viruses that are unable to reproduce in the cells of a vertebrate host, the IDI region has a number of significant specific differences, which include amino acid repeats and shortening of the main protein chain due to reduction. It is hypothesized that multiple sequence differences in the IDI region, which likely lead to significant changes in IDI conformation in the classical insect-specific flavivirus group, may prevent the formation of a complex with the STAT2 protein and as a result, the inability to suppress the innate immune response.

By analogy, it is hypothesized that changes in the conformation of the IDI of NS5 of TBEV strains with different pathogenicity may lead to a weakening of the pathogenic properties of TBEV for humans. Having an idea of the genus- and group-specific repeats in a certain sample under study, it is possible to identify important regions of the protein, the blockage of which could lead to inhibition of enzymatic activity.

Comparing the helicase repeats of NS3 and full-length NS5 protein, it can be noted that helicase repeats distinguish tick-borne viruses from those carried by mosquitoes. Structurally, these repeats form a cluster and create various functional interfaces on the surface of the helicase. On the contrary, in the protein NS5, there is no such dependence, there is a division into more extensive groups, such as viruses capable of affecting vertebrate cells and viruses that are not able to multiply in vertebrate cells.

CONCLUSION

This work shifts the emphasis for new researchers from well-studied pathogens (mammalian tick-borne flavivirus group, Mosquito-Borne Flavivirus group) towards a detailed study of little-known and poorly pathogenic viruses for humans (Kadam tick-borne flavivirus group, dual-host affected insect-specific flaviviruses ISFV-like, classical insect-specific flaviviruses). Understanding the differences between the mechanisms implemented in proteins of weakly or non-pathogenic viruses for humans' sheds light on critical areas of the structure, which affect reproduction in the cells of the vertebral host. So successful blocking of the innate immunity system or a number of other key stages of the life cycle may be possible. The relatedness of certain types of viruses to a particular host or vector is often correlated with repeats in proteins.

SIGNIFICANCE STATEMENT

Insect-specific and tick-borne flaviviruses have attracted great interest. Many studies point to conserved motifs in various proteins of the *Orthoflavivirus* genus. In this study, the "amino acid repeat" model was used to specify structural motifs in NS3 and NS5 proteins. Results proved that a structural motif in the interdomain region of the NS5 protein may be involved in blocking innate immunity through the STAT2 system. It can be used to screen for potential inhibitors of flavivirus infection.

REFERENCES

1. Calzolari, M., L. Zé-Zé, D. Růžek, A. Vázquez and C. Jeffries *et al*., 2012. Detection of mosquito-only flaviviruses in Europe. J. Gen. Virol., 93: 1215-1225.

2. Calzolari, M., L. Zé-Zé, A. Vázquez, M.P.S. Seco, F. Amaro and M. Dottori, 2016. Insect-specific flaviviruses, a worldwide widespread group of viruses only detected in insects. Infect. Genet. Evol., 40: 381-388.
3. Blitvich, B. and A.E. Firth, 2015. Insect-specific flaviviruses: A systematic review of their discovery, host range, mode of transmission, superinfection exclusion potential and genomic organization. Viruses, 7: 1927-1959.
4. Chambers, T.J., C.S. Hahn, R. Galler and C.M. Rice, 1990. *Flavivirus* genome organization, expression, and replication. Annu. Rev. Microbiol., 44: 649-688.
5. Wallner, G., C.W. Mandl, C. Kunz and F.X. Heinz, 1995. The flavivirus 3'-noncoding region: Extensive size heterogeneity independent of evolutionary relationships among strains of tick-borne encephalitis virus. Virology, 213: 169-178.
6. Lescar, J., D. Luo, T. Xu, A. Sampath, S.P. Lim, B. Canard and S.G. Vasudevan, 2008. Towards the design of antiviral inhibitors against flaviviruses: The case for the multifunctional NS3 protein from dengue virus as a target. Antiviral Res., 80: 94-101.
7. Sampath, A. and R. Padmanabhan, 2009. Molecular targets for flavivirus drug discovery. Antiviral Res., 81: 6-15.
8. Tay, M.Y.F. and S.G. Vasudevan, 2018. The Transactions of NS3 and NS5 in Flaviviral RNA Replication. In: Dengue and Zika: Control and Antiviral Treatment Strategies, Hilgenfeld, R. and S.G. Vasudevan (Eds.), Springer, Singapore, ISBN: 978-981-10-8727-1, pp: 147-163.
9. Gross, C.H. and S. Shuman, 1996. Vaccinia virus RNA helicase: Nucleic acid specificity in duplex unwinding. J. Virol., 70: 2615-2619.
10. Phoon, C.W., P.Y. Ng, A.E. Ting, S.L. Yeo and M.M. Sim, 2001. Biological evaluation of hepatitis c virus helicase inhibitors. Bioorg. Med. Chem. Lett., 11: 1647-1650.
11. Cordin, O., J. Banroques, N.K. Tanner and P. Linder, 2006. The DEAD-box protein family of RNA helicases. Gene, 367: 17-37.
12. Dong, H., B. Zhang and P.Y. Shi, 2008. Flavivirus methyltransferase: A novel antiviral target. Antiviral Res., 80: 1-10.
13. Yap, T.L., T. Xu, Y.L. Chen, H. Malet and M.P. Egloff *et al*., 2007. Crystal structure of the dengue virus RNA-dependent RNA polymerase catalytic domain at 1.85-angstrom resolution. J. Virol., 81: 4753-4765.
14. Benson, D.A., M. Cavanaugh, K. C lark, I. Karsch-Mizrachi, D.J. Lipman, J. Ostell and E.W. Sayers, 2013. GenBank. Nucleic Acids Res., 41: D36-D42.
15. Siddell, S.G., D.B. Smith, E. Adriaenssens, P. Alfenas-Zerbini and B.E. Dutilh *et al*., 2023. Virus taxonomy and the role of the International Committee on Taxonomy of Viruses (ICTV). J. Gen. Virol., Vol. 104. 10.1099/jgv.0.001840.
16. Katoh, K., K. Misawa, K.I. Kuma and T. Miyata, 2002. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res., 30: 3059-3066.
17. Berman, H.M., J. Westbrook, Z. Feng, G. Gilliland and T.N. Bhat *et al*., 2000. The protein data bank. Nucleic Acids Res., 28: 235-242.
18. Pettersen, E.F., T.D. Goddard, C.C. Huang, E.C. Meng and G.S. Couch *et al*., 2021. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci., 30: 70-82.
19. Blitvich, B.J. and A.E. Firth, 2017. A review of flaviviruses that have no known arthropod vector. Viruses, Vol. 9. 10.3390/v9060154.
20. Grard, G., G. Moureau, R.N. Charrel, J.J. Lemasson and J.P. Gonzalez *et al*., 2007. Genetic characterization of tick-borne flaviviruses: New insights into evolution, pathogenetic determinants and taxonomy. Virology, 361: 80-92.
21. Chen, C.J., M.D. Kuo, L.J. Chien, S.L. Hsu, Y.M. Wang and J.H. Lin, 1997. RNA-protein interactions: Involvement of NS3, NS5, and 3' noncoding regions of Japanese encephalitis virus genomic RNA. J. Virol., 71: 3466-3473.
22. Uchil, P.D. and V. Satchidanandam, 2003. Architecture of the flaviviral replication complex: Protease, nuclease, and detergents reveal encasement within double-layered membrane compartments. J. Biol. Chem., 278: 24388-24398.
23. Chen, Y., J.P. Potratz, P. Tijerina, M. del Campo, A.M. Lambowitz and R. Russell, 2008. DEAD-box proteins can completely separate an RNA duplex using a single ATP. Proc. Natl. Acad. Sci. U.S.A., 105: 20203-20208.
24. Kukhanova, M.K., I.L. Karpenko and A.V. Ivanov, 2020. DEAD-box RNA helicase DDX3: Functional properties and development of DDX3 inhibitors as antiviral and anticancer drugs. Molecules, Vol. 25. 10.3390/molecules25041015.
25. Malet, H., M.P. Egloff, B. Selisko, R.E. Butcher and P.J. Wright *et al*., 2007. Crystal structure of the RNA polymerase domain of the West Nile virus non-structural protein 5. J. Biol. Chem., 282: 10678-10689.
26. Bussetta, C. and K.H. Choi, 2012. Dengue virus nonstructural protein 5 adopts multiple conformations in solution. Biochemistry, 51: 5921-5931.
27. Potapova, U., S. Feranchuk, G. Leonova and S. Belikov, 2018. The rearrangement of motif F in the flavivirus RNA-directed RNA polymerase. Int. J. Biol. Macromol., 108: 990-998.
28. Lu, G. and P. Gong, 2013. Crystal structure of the full-length Japanese encephalitis virus NS5 reveals a conserved methyltransferase-polymerase interface. PLoS Pathog., Vol. 9. 10.1371/journal.ppat.1003549.
29. Zhao, Y., T.S. Soh, J. Zheng, K.W.K. Chan and W.W. Phoo *et al*., 2015. A crystal structure of the dengue virus Ns5 protein reveals a novel inter-domain interface essential for protein flexibility and virus replication. PLoS Pathog., Vol. 11. 10.1371/journal.ppat.1004682.
30. Klema, V.J., R. Padmanabhan and K.H. Choi, 2015. Flaviviral replication complex: Coordination between RNA synthesis and 5'-RNA capping. Viruses, 7: 4640-4656.

31. Klema, V.J., M. Ye, A. Hindupur, T. Teramoto, K. Gottipati, R. Padmanabhan and K.H. Choi, 2016. Dengue virus nonstructural protein 5 (NS5) assembles into a dimer with a unique methyltransferase and polymerase interface. PLoS Pathog., Vol. 12. 10.1371/journal.ppat.1005451.
32. Tay, M.Y.F., K. Smith, I.H.W. Ng, K.W.K. Chan and Y. Zhao *et al*., 2016. The C-terminal 18 amino acid region of dengue virus NS5 regulates its subcellular localization and contains a conserved arginine residue essential for infectious virus production. PLoS Pathog., Vol. 12. 10.1371/journal.ppat.1005886.
33. Teramoto, T., A. Balasubramanian, K.H. Choi and R. Padmanabhan, 2017. Serotype-specific interactions among functional domains of dengue virus 2 nonstructural proteins (NS) 5 and NS3 are crucial for viral RNA replication. J. Biol. Chem., 292: 9465-9479.
34. Wu, J., W. Liu and P. Gong, 2015. A structural overview of RNA-dependent RNA polymerases from the flaviviridae family. Int. J. Mol. Sci., 16: 12943-12957.
35. Upadhyay, A.K., M. Cyr, K. Longenecker, R. Tripathi, C. Sun and D.J. Kempf, 2017. Crystal structure of full-length *Zika virus* NS5 protein reveals a conformation similar to *Japanese encephalitis virus* NS5. Acta Crystallogr. F Struct. Biol. Commun., 73: 116-122.
36. Lu, G. and P. Gong, 2017. A structural view of the RNA-dependent RNA polymerases from the *Flavivirus* genus. Virus Res., 234: 34-43.
37. Wang, B., S. Thurmond, K. Zhou, M.T. Sánchez-Aparicio and J. Fang *et al*., 2020. Structural basis for STAT2 suppression by flavivirus NS5. Nat. Struct. Mol. Biol., 27: 875-885.
38. Belikov, S.I., I.G. Kondratov, U.V. Potapova and G.N. Leonova, 2014. The relationship between the structure of the tick-borne encephalitis virus strains and their pathogenic properties. PLoS ONE, Vol. 9. 10.1371/journal.pone.0094946.
39. Potapova, U.V., S.I. Feranchuk, V.V. Potapov, N.V. Kulakova, I.G. Kondratov, G.N. Leonova and S.I. Belikov, 2012. NS2B/NS3 protease: Allosteric effect of mutations associated with the pathogenicity of tick-borne encephalitis virus. J. Biomol. Struct. Dyn., 30: 638-651.