

Conserved Sequences of Dengue Virus Proteins as Candidate Targets for Diagnostics, Prophylactic and/or Therapeutic Purposes

PLoS Neglected Tropical Diseases, Aug. 2008, 2 (8), e272

Asif M. Khan

Joint International Tropical Medicine Meeting
October 13-14, 2008

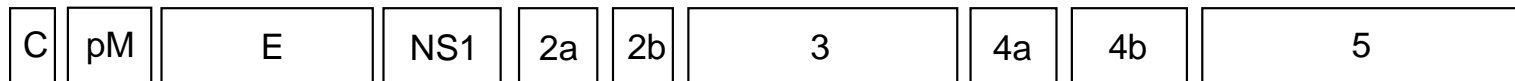
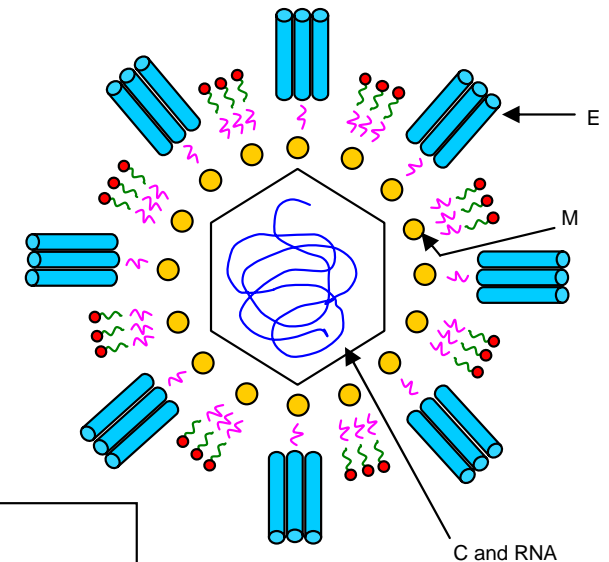
Bangkok, Thailand



Dengue virus (DENV)

- Family *Flaviviridae* and genus *Flavivirus*
- Genome (11kb) translated into a single polyprotein (~3390aa)
- Polyprotein cleaved to yield 10 different dengue proteins
- Four genetically distinct serotypes

Dengue genome/proteome:  Whitehead *et al.*, 2007



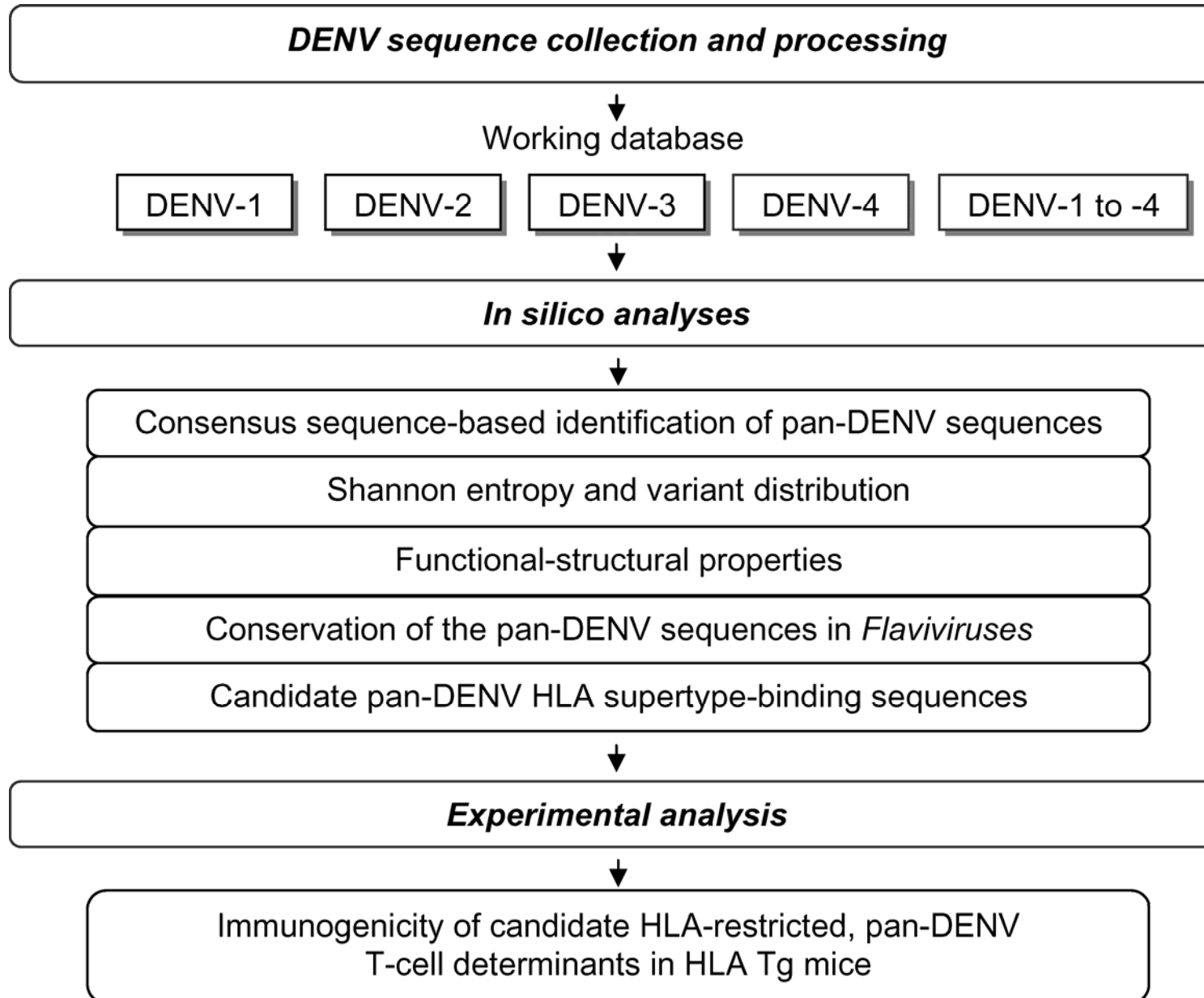
Motivations

- Genetic variation and rapid evolution are hallmarks of RNA viruses
- Variability is uneven across the genome
- RNA viruses are thus marked by protein sites permissive to multiple mutations and sites evolutionarily robust and highly conserved
- Identification and characterization of the historical dynamics of such conserved sites → potential targets for diagnostic, prophylactic and/or therapeutic purposes

Objectives

- Identify protein sequence regions common across the four serotypes and present in the majority of the sequences of each serotype (pan-DENV sequences)
- Characterize pan-DENV sequences for their application as candidate targets for diagnostics, prophylactic and/or therapeutic purposes

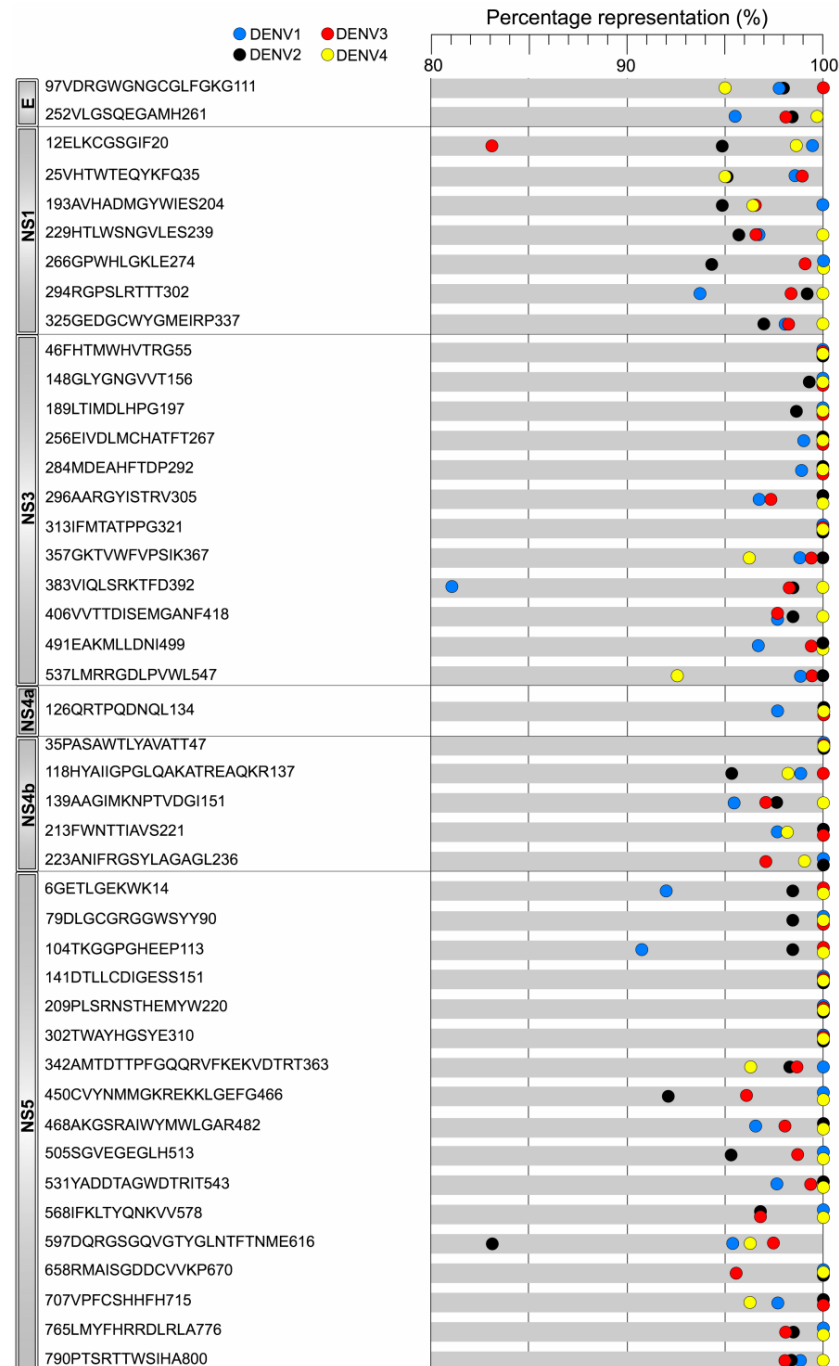
Methodology overview



More than 12,000 DENV protein sequences analyzed

Protein	DENV1 (#)	DENV2 (#)	DENV3 (#)	DENV4 (#)	DENV1-4 (#)
C	298	311	547	122	1278
prM	311	404	590	225	1530
E	1051	1518	910	366	3845
NS1	565	752	308	159	1784
NS2a	238	173	169	125	705
NS2b	224	163	183	44	614
NS3	186	178	297	34	695
NS4a	178	162	151	32	523
NS4b	176	163	150	113	602
NS5	179	187	267	195	828
<i>TOTAL</i>	<i>3406</i>	<i>4011</i>	<i>3572</i>	<i>1415</i>	12404

44 pan-DENV sequences identified



44 pan-DENV sequences identified

DENV protein	Size (aa)	Pan-DENV sequences ^a		
		No.	Size ^b	% of protein ^c
C	113-115	0	0	0
prM	166	0	0	0
E	493-495	2	25	5
NS1	352	7	74	21
NS2a	218	0	0	0
NS2b	130	0	0	0
NS3	618-619	12	122	20
NS4a	150	1	9	6
NS4b	245-249	5	69	28
NS5	900-904	17	215	24
<i>Total</i>	3387-3398	44	514	15

Entropy used to measure of the diversity of DENV and predicted evolutionary stability of the pan-DENV sequences

● Entropy H

$$H(x) =$$

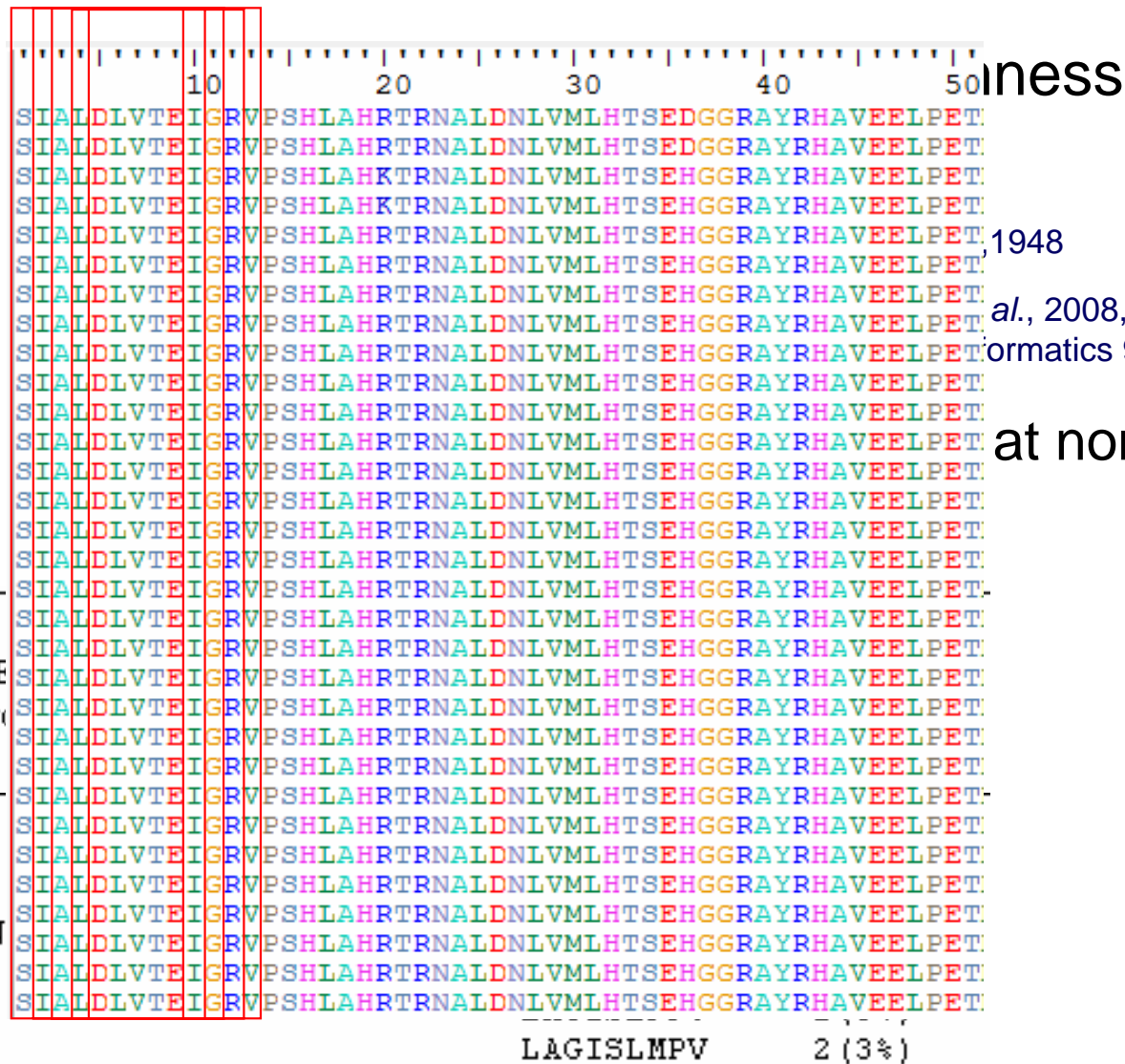
■ $p(i,x)$ is position x

● Entropy i

■ The n pri

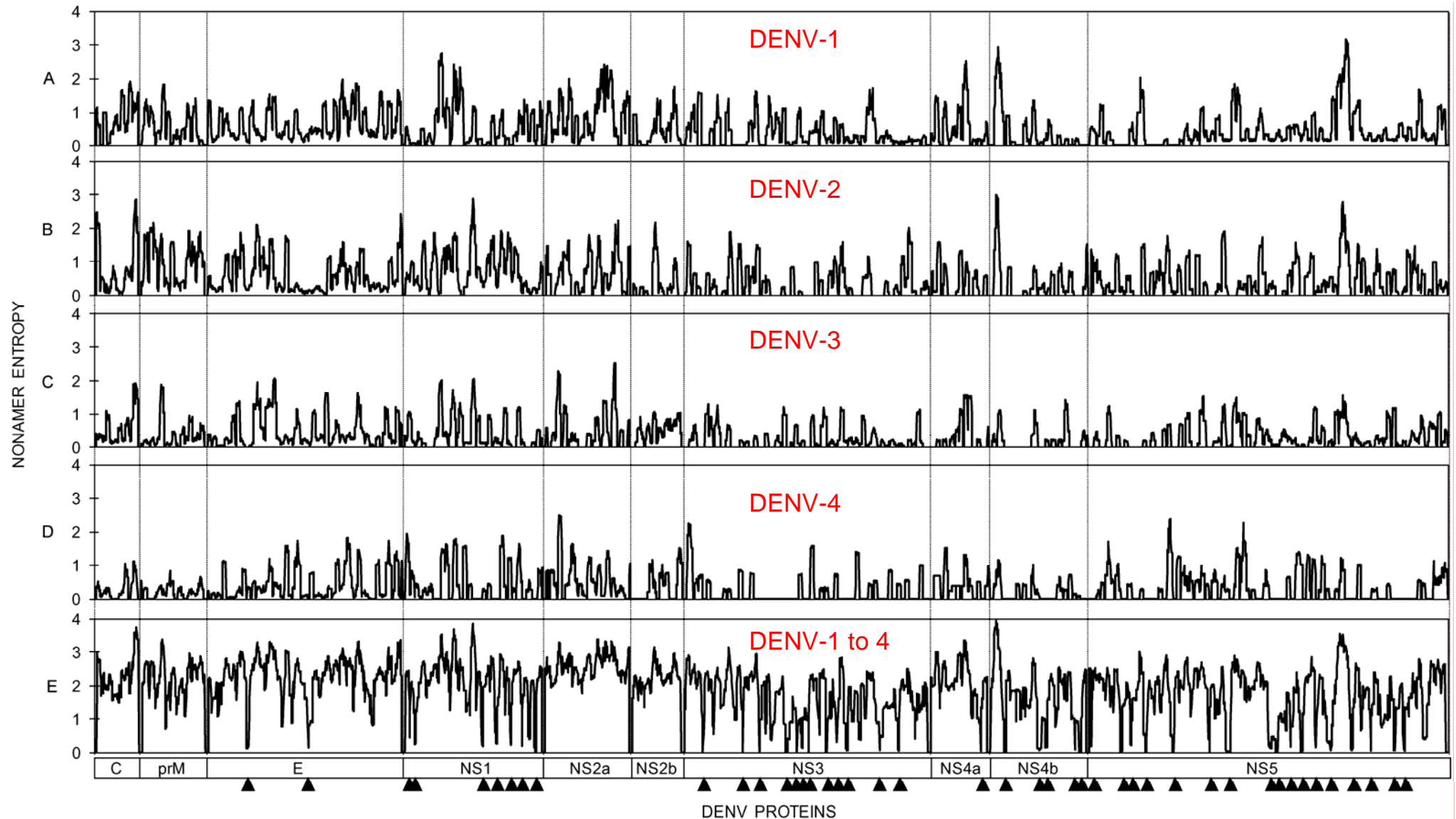
■ Their

● Lower en N

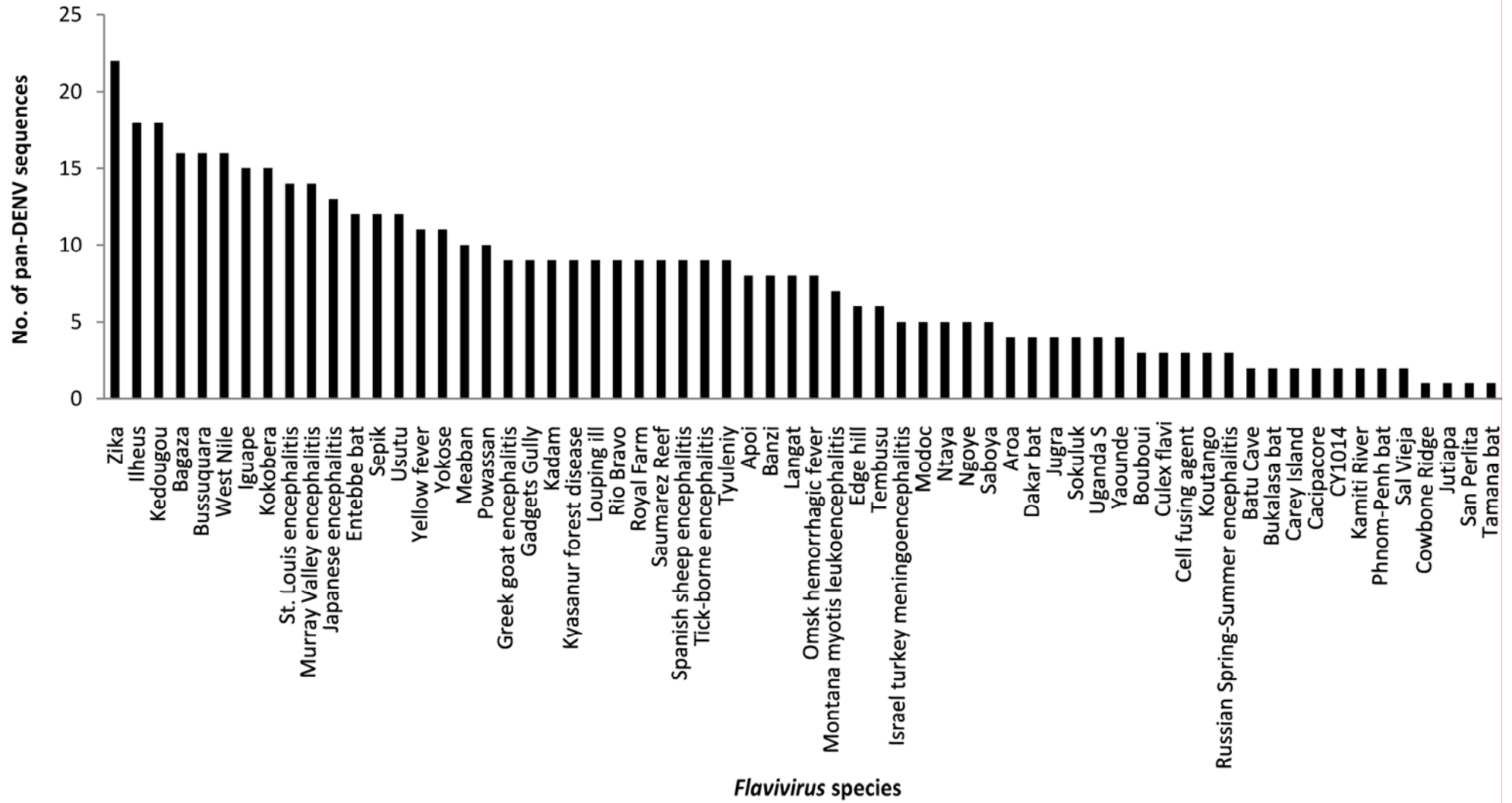


,1948
al., 2008,
informatics 9(1): S18
at nonamer

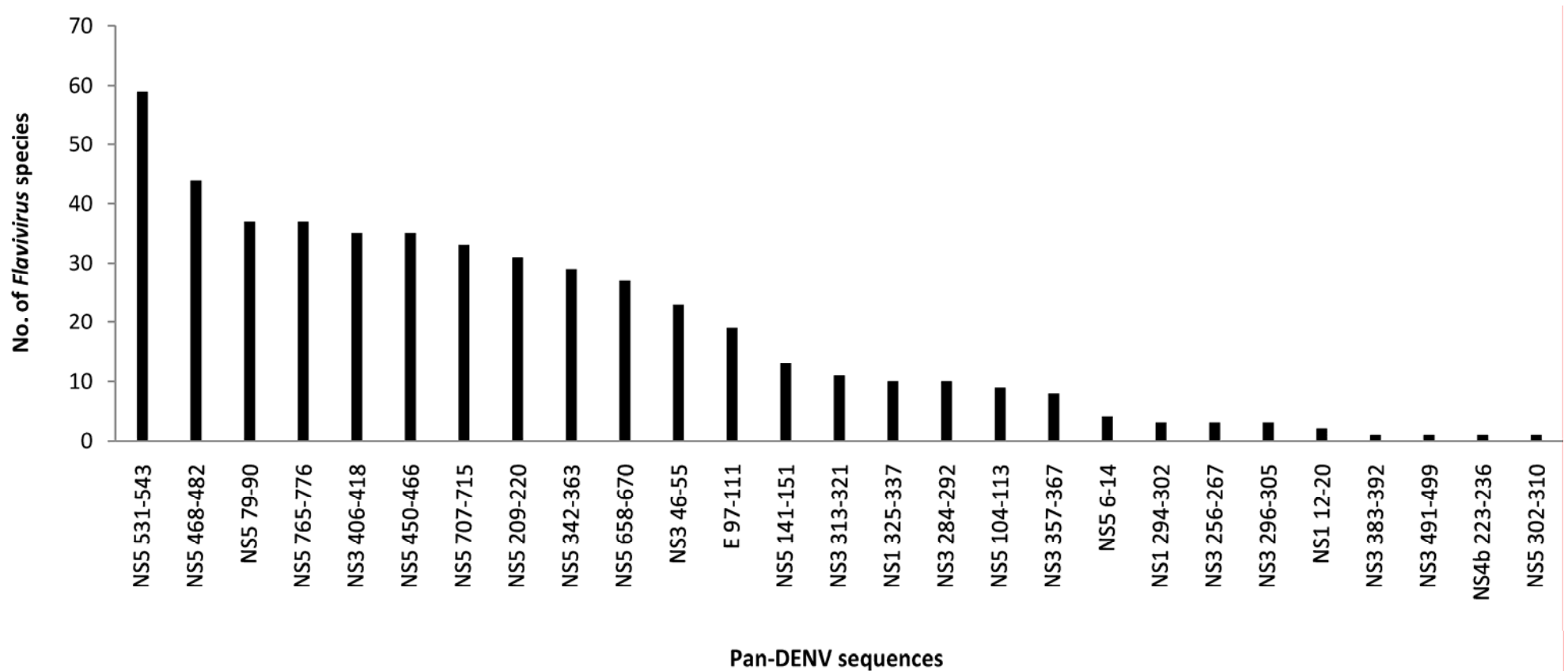
Pan-DENV sequences are localized in evolutionarily stable low peptide entropy regions



27 pan-DENV sequences shared by as many as 64 other *Flaviviruses*



Each of the 27 pan-DENV sequences was shared by at least one and as many as 57 *Flaviviruses*



27 pan-DENV sequences shared by other *Flaviviruses* exhibited low to high % representation

DENV protein	Pan-DENV sequence ^a	Species (#) ^b	Percentage representation (%) number of sequence analyzed ^c													
			LEV	WNV	JEV	MVE	UV	KBV	IH	TBEV	LV	OMSK	LIV	PV	KFDV	YFV
E	97VDRGWGNGCGLFGKG ₁₁₁	19	97 77	97 276	97 250											82 179
NS1	12ELKCGSGIF ₂₀	2			3 58											
	294RGPSLR ₃₀₂ TTT	3			3 58											
	325GEDGCWYGMEIRP ₃₃₇	10	90 30	96 138	31 186											
NS3	46FHTMWHVTRG ₅₅	23									89 19					87 23
	256EIVDLMCHATFT ₂₆₇	3														
	284MDEAHFTDP ₂₉₂	10		99 134	98 54											
	296AARGYISTRV ₃₀₅	3	9 98	4 137												
	313IFMTATPPG ₃₂₁	11	96 26	100 134	98 53											
	357GKTVWFVPSIK ₃₆₇	8		94 141					17 12							
	383VIQLSRKTFD ₃₉₂	1														
	406VTTDISEMGANF ₄₁₈	35	34 77	91 146	21 248					68 25						
491EAKMLLDNI ₄₉₉	1															
NS4b	223ANIFRGSYLAGAGL ₂₃₆	1														
NS5	6GETLGEKWK ₁₄	4														
	79DLGCGRGGWSYY ₉₀	37	77 35	96 140	20 244					82 38				83 18	40 10	78 27
	104TKGGPGHEEP ₁₁₃	9	93 28	89 148												
	141DTLLCDIGESS ₁₅₁	13		100 134	< 1 268											84 25
	209PLSRNSTHEMYW ₂₂₀	31	90 29	100 134						61 28				100 17		95 22
	302TWAYHGSYE ₃₁₀	1														
	342AMTDTTPFGQQRVFKEKVDTRT ₃₆₃	29	33 79	50 272	18 289	18 17	23 13		25 12	32 34				60 45		51 41
	450CVYNMMGKREKLGFEFG ₄₆₆	35	24 103	33 344	69 74	20 15		7 14	14 14	63 27			11 19	41 41		23 90
	468AKGSRAIWYMWLGAR ₄₈₂	44	32 19						62 13	82 22	42 12			41 41		11 208
	531YADDITAGWDTRIT ₅₄₃	59	90 30	73 193	88 58				80 10	69 26		93 14		84 19		85 26
658RMAISGDCCVVKP ₆₇₀	27	73 37	50 272	95 55	50 10								47 34		3 37	
707VPFCSHHFH ₇₁₅	33								100 19		100 13		76 17		95 22	
765LMYFHRRDLRLA ₇₇₆	37	64 42	91 151	83 59				44 18								

Pan-DENV sequences are relevant for structure and function

DENV protein	Pan-DENV sequence ^a	Functional domains and motifs ^b	Putative post-transcriptional modifications ^b
E	97VDRGWGNGCGLFGKG ₁₁₁ 252VLGSQEGAMH ₂₆₁	Dimerisation Domain, Fusion Peptide Dimerisation Domain	N-Myristoylation -
NS1	12ELKCGSGIF ₂₀ 25VHTWTEQYKFQ ₃₅ 294RGPSLR ₃₀₂ 325GEDGCWYGMEIRP ₃₃₇	- - - -	N-Myristoylation CKII PKC N-Myristoylation
NS3	46FHTMWHVTRG ₅₅ 148GLYGNGVVT ₁₅₆ 189LTIMDLHPG ₁₉₇ 256EIVDLMCHATFT ₂₆₇ 284MDEAHFTDP ₂₉₂ 296AARGYISTRV ₃₀₅ 313IFMTATPPG ₃₂₁ 357GKTVWFVPSIK ₃₆₇ 383VIQLSRKTFD ₃₉₂ 537LMRRGDLPVWL ₅₄₇	Peptidase S7 - - DEAD/H Domain DEAD/H Domain Microbodies C-Terminal Targeting Signal DEAD/H Domain - - Cell Attachment	- N-Myristoylation CKII - - PKC - PKC PKC -
NS4a	126QRTPQDNQL ₁₃₄	-	CKII
NS4b	213FWNTTIAVS ₂₂₁ 223ANIFRGSYLAGAGL ₂₃₆	- -	N-Glycosylation N-Myristoylation
NS5	6GETLGEKWK ₁₄ 79DLGCGRGGWSYY ₉₀ 209PLSRNSTHEMYW ₂₂₀ 450CVYNMMGKREKKLGEFG ₄₆₆ 505SGVEGEG ₅₁₃ 597DQRGSGQVGT ₆₁₆ 658RMAISGDDCVVKP ₆₇₀ 790PTSR ₈₀₀	- FtsJ-like Methyltransferase Domain - - - RdRp Catalytic Domain RdRp Catalytic Domain -	CKII N-Myristoylation N-Glycosylation, CKII Amidation CKII N-Myristoylation, CKII CKII PKC

Pan-DENV sequences overlap T-cell epitopes immunogenic in HLA transgenic mice

DENV protein	Pan-DENV sequence ^b	Ag-specific CD4 T-cell responses ^a			
		Peptide sequences (DENV-3) ^c	IFN- γ -SFC/10 ⁶ splenocytes \pm SD ^d		
			DR2	DR3	DR4
E	252 VLGSQEGAMH ₂₆₁	PEVV <u>VLGSQEGAMH</u> T	-	-	88 \pm 34
NS1	193 AVHADMGYWIES ₂₀₄	<u>AVHADMGYWIESQKN</u>	-	17 \pm 1	-
	229 HTLWSNGVLES ₂₃₉	<u>WPKSHTLWSNGVLES</u>	-	129 \pm 3*	-
		<u>HTLWSNGVLESDMII</u>	-	131 \pm 103	37 \pm 3
	266 GPWHLGKLE ₂₇₄	<u>HTQTAGPWHLGKLE</u>	-	333 \pm 6	-
	294 RGPSLRITTT ₃₀₂	<u>TRGPSLRITTTVSGKL</u>	-	-	11 \pm 4
NS3	189 LTIMDLHPG ₁₉₇	<u>KKRNLTIMDLHPGSG</u>	-	-	50 \pm 16
	206 AARGYISTRV ₃₀₅	<u>ASIAARGYISTRVGM</u>	40 \pm 14	-	-
		<u>ARGYISTRVGMGEAA</u>	9 \pm 4	-	-
	313 IFMTATPPG ₃₂₁	<u>EAAAIFMTATPPGTA</u>	-	-	474 \pm 116
		<u>IFMTATPPGTADAFP</u>	-	-	323 \pm 287
	357 GKTVWFVPSIK ₃₆₇	<u>TDFAGKTVWFVPSIK</u>	48 \pm 15	-	-
		<u>GKTVWFVPSIKAGND</u>	396 \pm 14	-	-
	383 VIQLSRKTFD ₃₉₂	<u>KKVIQLSRKTFDTEY</u>	-	21 \pm 3	-
	406 VVTTDISEMGANF ₄₁₈	<u>FVVTTDISEMGANFK</u>	-	-	408 \pm 104
		<u>TDISEMGANFKADR V</u>	-	152 \pm 33	-
NS5	302 TWAYHGSYE ₃₁₀	<u>DENPYKTWAYHGSYEVK</u>	126 \pm 10*	-	14 \pm 5
		<u>TWAYHGSYEVKATGSA</u>	161 \pm 20*	-	63 \pm 17
	342 AMTDTTTPFGQQRVFKEKVDTR ₃₆₃	<u>MVTQMAMTDTTTPFGQQR</u>	-	-	28 \pm 0*
	450 CVYNNMGMKREKLGEGF ₄₆₆	<u>GSCVYNNMGMKREKLGEG</u>	-	-	13 \pm 2
	505 SGVEGEGHL ₅₁₃	<u>NSYSGVEGEGHLKLYI</u>	-	-	184 \pm 15
	531 YADDTAGWDTRIT ₅₄₃	<u>KIPGGAMYADDTAGWDT</u>	-	-	46 \pm 3
	568 IFKLTQNKVV ₅₇₈	<u>ANAIFKLTQNKVVVKVQ</u>	577 \pm 384	-	24 \pm 9*
	597 DQRGSGQVGTYGLNFTNME ₆₁₆	<u>VMDIISRKQDQRGSGQVG</u>	-	88 \pm 1	-
	658 RMAISGDDCVVKP ₆₇₀	<u>VERLKRMAISGDDCVVK</u>	-	159 \pm 24	16 \pm 6
		<u>MAISGDDCVVKPIDDRF</u>	-	249 \pm 39	-
	707 VPFCSHHFH ₇₁₅	<u>DWQQVFPFCSHHFHELIM</u>	32 \pm 8*	34 \pm 11	-
	765 LMYFHRRDLRLA ₇₇₆	<u>MYFHRRDLRLASNAI</u>	75 \pm 16*	-	33 \pm 9
	790 PTSRTTWSIHA ₈₀₀	<u>VHWVPTSRTTWSIHAHH</u>	-	-	83 \pm 1
	<u>SRTTWSIHAHHQWMTE</u>	-	-	122 \pm 46	

● 30 ELISpot positive peptides → 22 pan-DENV sequences

Pan-DENV sequences overlap reported DENV T-cell epitopes immunogenic in human

DENV protein	Pan-DENV sequence ^a	Immunogenic T-cell determinants ^b			
		Sequence ^c	T subset	HLA Ag	Reference(s)
E	252VLGSQEGAMH ₂₆₁	<u>KKQDVVVVLGSQEGAM</u>	-	-	[76]
NS3	46FHTMWHVTRG ₅₅	<u>TFHTMWHVTRGAVLM</u>	CD4	-	[76]
	148GLYGNGVVT ₁₅₆	<u>KVVGLYGNGVVTRSG</u>	CD4	DR*15	[76]
	189LTIMDLHPG ₁₉₇	<u>KRLTIMDLHPGAGKT</u>	CD4	-	[72]
		<u>RKLTIMDLHPGSGKT</u>	CD4	-	[72]
		<u>RKLTIMDLHPGAGKT</u>	CD4	-	[72]
		<u>RNLTIMDLHPGSGKT</u>	CD4	-	[72]
	256EIVDLMCHATFT ₂₆₇	<u>EHTGREIVDLMCHAT</u>	CD4	-	[76]
		<u>EIVDLMCHATFTMRL</u>	CD4	-	[76]
		<u>EIVDLMCHAT</u>	CD4	DPw2	[77,78]
	284MDEAHFTDP ₂₉₂	<u>LIIMDEAHFTDPASI</u>	-	-	[76]
	313IFMTATPPG ₃₂₁	<u>AGIFMTATPPGSRDP</u>	-	-	[76]
	357GKTVWFVPSIK ₃₆₇	<u>TVWFVPSIK</u>	CD8	A*11	[16]
	383VIQLSRKTFD ₃₉₂	<u>KKVIQLSRKTFDSEY</u>	-	-	[76]
406VVTTDISEMGANF ₄₁₈	<u>NDWDFVVTTDISEMG</u>	-	-	[76]	

● 15 DENV human T-cell epitopes → 10 pan-DENV sequences

Summary

- The pan-DENV sequence have been relatively free of mutations and are predicted to remain conserved
- Provides critical data for rational vaccine development, structure-based design of candidate inhibitory compounds, and improvement of the current diagnostic methods
- A framework for similar analysis of any other virus

Acknowledgements & Thanks

National University of Singapore

- A/Prof. Tan Tin Wee
- Olivo Miotto
- Heiny Tan

Johns Hopkins University

- Prof. J. Thomas August
- Dr. Eduardo J.M. Nascimento
- Dr. K.N. Srinivasan
- Dr. J. Salmon
- Dr. E.T. Marques

Dana-Farber Cancer Institute

- Prof. Vladimir Brusic
- Dr. Zhang Guanglan

Funding:

National Institute of Allergy and Infectious Diseases, NIH, USA

Grant No. 5 U19 AI56541 & Contract No. HHSN2662-00400085C

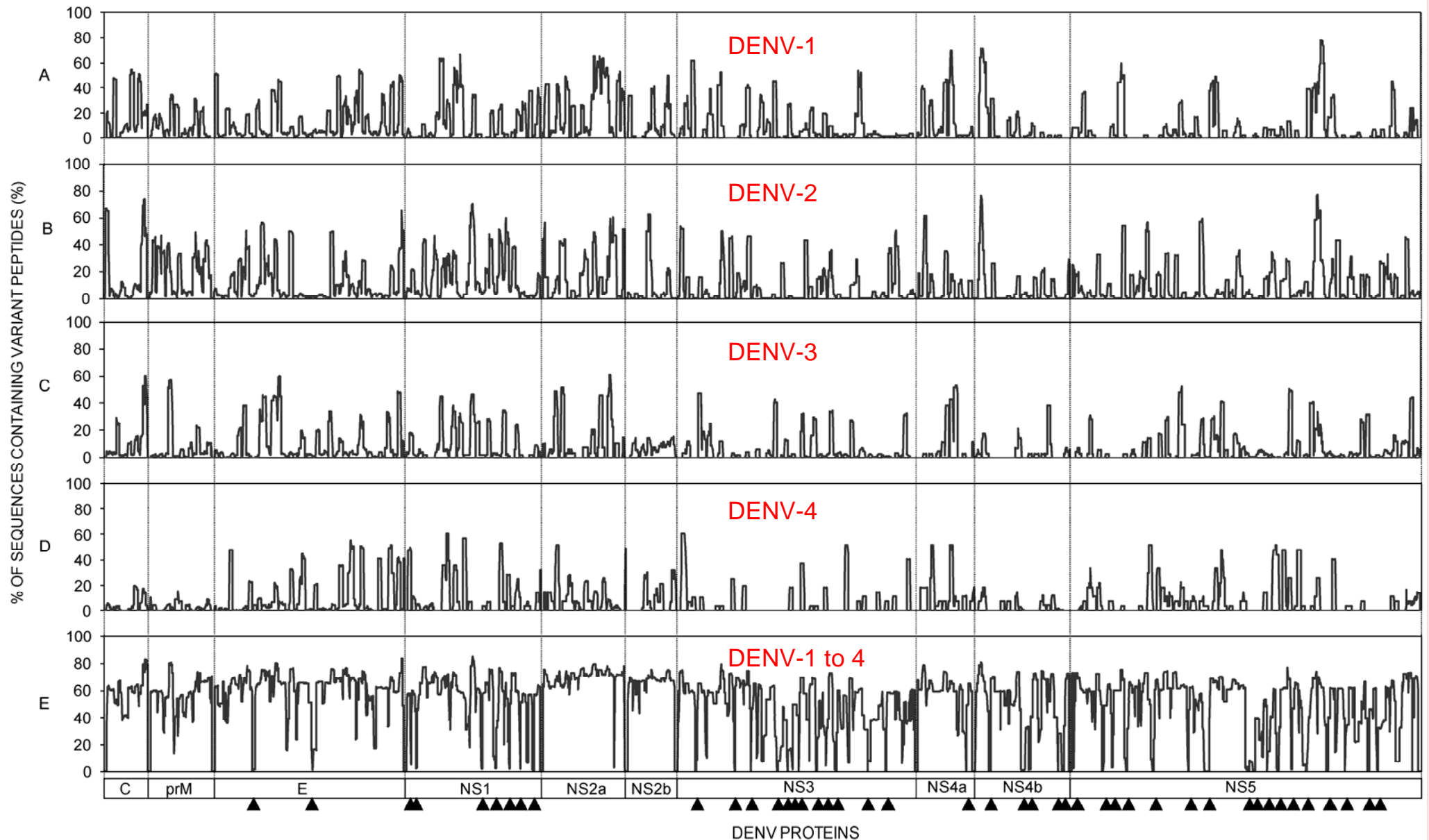


Supplementary Slides

Analysis of variant representation

DENV-3 protein	Nonamer position	No. of sequences	Nonamer peptides ^a	Representation of peptides	Combined % representation of variants ^b	Nonamer entropy ^c
E	14	479	<u>DFVEGLSGA</u>	479 (100%)	0	0
NS2a	176	64	<u>LAGISLLPV</u>	25 (39%)	61	2.4
			LAGVSLLPV	11 (17%)		
			LAGVSLLPL	9 (14%)		
			LAVISLLPV	9 (14%)		
			LAGISLLPL	6 (9%)		
			LAGISLFPV	2 (3%)		
			LAGISLMPV	2 (3%)		
			NS4a	86		
SIGLICVIA	19 (28%)					
SIGLICVIV	8 (13%)					
SIGLICVAA	2 (3%)					

Pan-DENV sequences are localized in regions of low variant representation



Pan-DENV sequences contained putative promiscuous binders that overcome human population HLA diversity

Protein	Position	Class I and II HLA Supertypes													
		A1	A2	A3	A24	A26	B7	B8	B27	B39	B44	B58	B62	DR	
E	97-111				●		●								
NS1	12-20					●		●					●		
	25-35	●			●	●		●	●			●	●		
	193-204	●			●	●				●		●		●	
	229-239	●						●		●			●	●	
	325-337										●				
NS3	46-55									●				●	
	189-197													●	
	256-267		●	●			●		●		●	●		●	
	296-305			●	●	●			●						
	313-321													●	
	357-367	●	●	●	●	●						●		●	
	383-392						●		●				●	●	
	406-418	●				●							●	●	
	537-547	●	●						●	●	●		●	●	
NS4a	126-134							●	●						
NS4b	35-47	●	●	●										●	
	118-137		●	●	●			●	●	●			●	●	
	139-151								●					●	
	223-236	●	●	●	●			●	●	●			●	●	
NS5	6-14									●					
	79-90	●				●							●		
	141-151		●											●	
	209-220	●					●	●	●			●	●	●	
	342-363	●		●	●	●	●	●	●					●	
	450-466	●	●	●	●	●		●	●		●			●	
	468-482	●	●						●	●		●		●	
	531-543	●	●			●								●	
	568-578	●	●	●	●	●								●	
	597-616	●	●			●	●			●		●	●	●	
	658-670	●	●	●	●	●			●					●	
	707-715			●										●	
	765-776	●	●	●	●	●			●	●			●	●	
790-800	●	●			●			●					●		

● 34/44 pan-DENV sequences predicted to contain 100 supertype-restricted T-cell epitopes

- NetCTL – Larsen et al., 2005
- Multipred – Zhang et al., 2005
- ARB – Bui et al., 2005
- Tepitope – Bian & Hammer, 2004