Supplementary Information

Article: **Genetic characterization of 2006-2008 isolates Chikungunya virus from Kerala, South India by whole genome sequence analysis**

Journal name: Virus Genes

Authors: E.Sreekumar[a]*, Aneesh Issac[a#], Sajith Nair[a#], Ramkumar Hariharan[b], M.B. Janki[b], D.S. Arathy[a], R. Regu [c], Thomas Mathew[d], M. Anoop[a], K.P.Niyas[a] and M.R. Pillai[b]

[a]Molecular Virology Laboratory; [b]Translational Cancer Research Laboratory,
Rajiv Gandhi Centre for Biotechnology (RGCB), Thycaud P.O.,
Thiruvananthapuram, 695014, Kerala, India
[c]National Institute of Communicable Diseases (NICD), Kozhikkode, Kerala, India
[d]Department of Community Medicine, T.D Medical College, Alappuzha, Kerala, India

*corresponding author
Tel.: +91 471 2529517; Fax: +91 471 2348096

E-mail address: esreekumar@rgcb.res.in (E.Sreekumar)

Table 1: Details of clinical samples included in the study for sequence analysis

Patient Sample No.	Age	Sex	Date of collection	Place of Collection	Unique Mutations		
					E1 A226V	E2 K252Q	nsP2 L539S
RGCB03-06	48	F	07-10-06	Kayamkulam	N	N	N
RGCB 05-06	38	F	07-10-06	Kayamkulam	N	N	N
RGCB 40-07	50	F	12-07-07	Pathanamthitta	Y	N	N
RGCB 44-07	45	M	12-07-07	Pathanamthitta	N	N	N
RGCB 63-07	65	F	12-07-07	Punalur	Y	Y	Y
RGCB 65-07	13	F	12-07-07	Punalur	Y	N	N
RGCB 69-07	40	F	12-07-07	Punalur	Y	N	N
RGCB 70-07	35	F	12-07-07	Punalur	Y	Y	Y
RGCB71-07	42	M	12-07-07	Punalur	N	N	N
RGCB74-07	45	F	12-07-07	Punalur	Y	Y	Y
RGCB75-07	20	M	12-07-07	Punalur	Y	N	Y
RGCB79-07	N.A	F	12-07-07	Punalur	N	N	N
RGCB80-07	22	M	12-07-07	Punalur	Y	Y	Y
RGCB100-07	13	M	12-07-07	Kottayam	Y	N	N
RGCB102-07	64	M	12-07-07	Kottayam	Y	N	N
RGCB106-07	4	F	12-07-07	Kanjirapally	Y	N	N
RGCB112-07	8	F	13-07-07	Punalur	Y	Y	Y
RGCB116-07	46	F	13-07-07	Punalur	Y	Y	Y
RGCB120-07	17	M	13-07-07	Punalur	Y	Y	Y
RGCB153-07	54	M	13-07-07	Anchal	Y	Y	Y
RGCB156-07	28	M	13-07-07	Anchal	Y	Y	Y
RGCB157-07	22	F	13-07-07	Anchal	Y	N	N
RGCB176-07	25	M	13-07-07	Kanjirapally	Y	N	N
RGCB283-07	N.A	F	19-07-07	Trivandrum	Y	Y	Y
RGCB300-07	49	F	06-08-07	Trivandrum	Y	Y	Y
RGCB301-07	23	M	06-08-07	Trivandrum	Y	Y	Y
RGCB305-07	19 days	N.A	13-08-07	Trivandrum	Y	N	N
RGCB355-08	25	F	29-05-08	Panathady	Y	Y	Y
RGCB356-08	46	F	29-05-08	Panathady	Y	N	N
RGCB358-08	16	M	29-05-08	Panathady	Y	N	N
RGCB359-08	69	F	29-05-08	Panathady	Y	N	N
RGCB360-08	38	F	29-05-08	Panathady	Y	N	N
RGCB361-08	63	F	29-05-08	Panathady	Y	Y	Y
RGCB362-08	80	M	29-05-08	Panathady	Y	Y	Y
RGCB363-08	8	F	29-05-08	Panathady	Y	N	N
RGCB365-08	48	M	29-05-08	Panathady	Y	Y	Y
RGCB367-08	37	F	29-05-08	Panathady	Y	N	N
RGCB370-08	41	M	29-05-08	Panathady	Y	N	N

Samples used for whole genome analysis are highlighted.

Y-present; N-Absent; N.A-not available

Table 2. Location of nucleotide changes that were identified only in the 2006, 2007 & 2008 RGCB isolates among the various isolates studied. The GenBank Accession number and the strain name of the isolates, other than RGCB isolates, included in the comparative analysis are- AF369024 (S-27); EF027139 (IND-00-MH4); EU564335 (CHIK31); EF210157 (DRDE-06); EU2444823 (ITA07-RA1); EU037962 (Wuzerberg 1); EF012359 (D570/06); DQ443544 (LR2006_OPY1); EU564334 (TM25); EF027134 (IND-06-AP3); EF027135 (IND-06-KA15); EF027136 (IND-06-MH2); EF027137 (IND-06-RJ1); EF027138 (IND-06-TN1).

Isolate	**Nucleotide changes exclusively observed in RGCB isolates (Numbering with respect to S-27)**										
	5'NTR	**Coding region**									**3' NTR**
		Non-structural protein				**Structural protein**					
		nsP1	**nsP2**	**nsP3**	**nsP4**	**C**	**E3**	**E2**	**6K**	**E1**	
RGCB03/KL06	-	-	A4051G	**T4825A** (D250E)	A6310G T6388C T6517T G6589A **T7131A** (M489K)	G8214T	-	G9487A (E316K)	-	**A10343G** (E127G) **T10546C** (Y195H) A10812G G11142A **A11179G**(T396A)	-
RGCB05/KL06	-	-	-	**C4148T**(P26S)	T5980C A6106G	**G8162A** (G199D)	-	T8592A C8760T	-	**A10552G** (M197V)	-
RGCB80/KL07	-	T1381G[#] **A1668G** (D536G)	**T1824C** (V48A) A2209G **T2223C** (V181A) **A2666G** (K329E) **T3297C** (L539S) T3397C C4009T	**A4355C** (K94Q) C5014T	A6076G **T6546C** (V289A) C7450T[#]	C7983T[#]	-	**A9295C** (K252Q) G9654A G9681A[#]	-	A10356G **C10670T***(A226V) **T11290C** (C433R)	C11614T[#]
RGCB120/KL07	-	T1381G[#]	**T2391G** (L237R) **T2394C** (L238P) **T3297C** (L539S) T3397C	C5014T	A6076G T6805C C7450T[#]	C7983T[#]	-	T8895C **A9295C**(K252Q) **A9623G** (E361G) G9654A G9681A[#] **T9787G** (C416G)	**A9988G** (S61G)	A10356G **A10514G** (D184G) **C10670T***(A226V)	C11614T[#] T11641C
RGCB355/KL08	-	T1381G[#]	**T3297C** (L539S) T3397C **T3804C** (L708P)	**T4677G** (L201R) C5014T A5116G A5392G **A5394G** (E440G)	A5725G T5770C G5824A **T6011C** (S116P) A6076G T6136C **A6419G** (N252D) **A7100G** (M479V) C7450T[#]	C7983T[#]	-	**A8757G** (N72S) **A9295C** (K252Q) G9654A G9681A[#]	-	A10356G **C10670T*** (A226V) **G10736A** (G248E) **A11225G** (K411R)	T11487C C11614T[#]
RGCB356/KL08	-	**G389A** (G105R) C725T **T848C** (W258R) C1603T	A2912G G2944A A3283G	A4646G	C6397T A7273G	**T7649C** (I28T)	**A8479G** (R44G)	A9114G **G9226A** (V229I) **G9589A** (G350S)	-	**T10004C** (V14A) **C10670T*** (A226V) **C10904T** (P304L)	A11354G A11673G

[#] This substitution was also seen in ITA07-RA1

* This change was also present in the strains ITA07-RA1, Wuzerberg 1, D570/06, TM25 and LR2006_OPY1; but not in any of the Indian strains analysed in the study.

Non-synonymous changes are shown in bold letters, and resulting amino acid changes (numbered with respect to the corresponding protein position) is shown in parenthesis.

Table. 3 CHIKV Structural protein epitopes associated with **HLA-A** as predicted by Epi-Jen server

Mutation position in the polypeptide	**HLA-A alleles**										
	A*0101	**A*0201**	**A*0202**	**A*0203**	**A*0206**	**A*0301**	**A*1101**	**A*24**	**A*3101**	**A*6801**	**A*6802**
I28T	-	-	-	-	-	-	-	-	-	-	-
G199D	-	-	-	-	-	GAVQYSG**G**R (27.1)	GAVQYSG**G**R (26.98)	QYSG**G**RFTI (1.492)	-	GAVQYSG**G**R (18.88)	-
R305G	DNVM**R**PGYY (36.31)	VM**R**PGYYQL ((0.18))	VM**R**PGYYQL (44.36)	VM**R**PGYYQL (4.79)	-	-	-	-	-	-	-
V554I	**V**TNHKKWQY (17.7)	-	-	-	KVDQCHAA**V** (93.97)	HAA**V**TNHKK (25.7)	HAA**V**TNHKK (25.82)	-	-	-	KVDQCHAA**V** (8.91)
K577Q	-	-	-	-	-	RNAELGDR**K** (98.86)	RNAELGDR**K** (22.44)	-	-	-	-
E641K	-	VTHKK**E**VVL (9.42)	VTHKK**E**VVL (32.96)	-	-	-	-	-	-	-	WVTHKK**E**VV (8.65)
G675S	-	-	-	-	-	-	-	-	-	-	-
E686G	IILYYY**E**LY (4.1)	**E**LYPTMTVV (0.7) EIILYYY**E**L (0.78)	-	**E**LYPTMTVV (4.06)	-	IILYYY**E**LY (31.48)	-	LYYY**E**LYPT (1.017) YY**E**LYPTMT (1.364) YYY**E**LYPTM (1.05)	-	-	-
C741G	-	FLLSLI**C**CI (0.13) SLI**C**CIRTA (6.19)	FLLSLI**C**CI (36.56)	FLLSLI**C**CI (0.23) SLI**C**CIRTA (5.5)	FLLSLI**C**CI (27.67)	LI**C**CIRTAK (1.88)	-	PFLLSLI**C**C (1.007)	LLSLI**C**CIR (20.94)	-	SLI**C**CIRTA (82.6)
S808G	IGAHTV**S**AY (37.67)	HTV**S**AYEHV (6.84)	HTV**S**AYEHV (50.58)	-	HTV**S**AYEHV (29.92)	-	-	-	-	-	-
V813A	-	HTVSAYEH**V** (6.84)	HTVSAYEH**V** (50.58) H**V**TVIPNTV (61.38)	-	HTVSAYEH**V** (29.92)	-	-	AYEH**V**TVIP (1.362)	-	-	H**V**TVIPNTV (9.51)
E926G	-	-	-	-	-	KT**E**FASAYR (6.18)	KT**E**FASAYR (20.56)	-	-	KT**E**FASAYR (15.1)	-
D983G	-	-	WTPF**D**NKIV (17.58)	WTPF**D**NKIV (25)	WTPF**D**NKIV (80.17)	SAWTPF**D**NK (1.73)	SAWTPF**D**NK (3.36)	AWTPF**D**NKI (0.804)	-	SAWTPF**D**NK (28.05)	WTPF**D**NKIV (38.82)
Y994H	KGDV**Y**NMY (44.36)	-	-	-	-	-	-	VYKGDV**Y**M (1.167) V**Y**NMDYPPF (1.704)	-	-	-
A1035V	A**A**GTVHVPY (3.62)	QLVLQRPA**A** (5.45)	LQRPA**A**GTV (27.61)	LQRPA**A**GTV (30.97)	LQRPA**A**GTV (52.24)	A**A**GTVHVPY (74.99)	-	-	-	-	-
G1057E	-	WLKER**G**ASL (6.61)	WLKER**G**ASL (5.62)	WLKER**G**ASL (0.16)	-	-	-	KYWLKER**G**A (1.097)	-	-	-
T1205A	-	**T**LGVQDISA (3.59)	-	PASHT**T**LGV (20.56)	PASHT**T**LGV (29.17)	-	-	NYPASHT**T**L (1.509)	-	-	-
K1220R	-	-	WVQ**K**ITGGV (78.89)	-	WVQ**K**ITGGV (9.73)	ATAMSWV**K** (8.71)	ATAMSWV**K** (2.87)	-	-	-	WVQ**K**ITGGV (2.76)

The values given within the bracket indicates the 50% inhibitory concentration (IC_{50}) of the peptide, a measure of the binding affinity. An IC_{50} value <50 is considered as a good affinity. Mutation points in the predicted epitopes are marked as bold.

Table. 4 CHIKV Structural protein epitopes associated with **HLA-B** as predicted by Epi-Jen server

Mutation position in the polypeptide	**HLA-B alleles**						
	B07	**B27**	**B40**	**B44**	**B51**	**B53**	**B3501**
I28T	-	**I**RPRPRPQR (1.221) PRPTIQV**I**R (1.293)	-	-	TPRPTIQV**I** (1.52)	TPRPTIQV**I** (83.56)	-
G199D	AVQYSG**G**RF (16.94)	**G**RFTIPTGA (1.598)	-	-	-	-	-
R305G	-	M**R**PGYYQLL (1.555)	-	-	-	-	-
V554I	-	-	-	-	-	-	-
K577Q	-	-	AELGDR**K**GK (0.968)	AELGDR**K**GK (0.968)	-	-	-
E641K	-	-	VTHKK**E**VVL (0.619)	VTHKK**E**VVL (0.619)	-	-	-
G675S	-	-	H**G**HPHEIIL (0.507)	H**G**HPHEIIL (0.507)	TAH**G**HPHEI (4.49) H**G**HPHEIIL (47.1)	-	-
E686G	-	-	Y**E**LYPTMTV (0.812)	Y**E**LYPTMTV (0.812)	-	-	-
C741G	-	-	-	-	-	-	-
S808G	-	-	-	-	**S**AYEHVTVI (15.7)	-	-
V813A	-	-	-	-	SAYEH**V**TVI (15.7)	-	-
E926G	-	-	SESCKT**E**FA (1.139) T**E**FASAYRA (1.137)	SESCKT**E**FA (1.139) T**E**FASAYRA (1.137)	-	-	-
D983G	-	-	-	-	TPF**D**NKIVV (14.55)	-	-
Y994H	IVVYKGDV**Y** (92.47)	-	-	-	-	-	-
A1035V	RPA**A**GTVHV (13.37)	QRPA**A**GTVH (1.365)	-	-	RPA**A**GTVH (30.2)	-	-
G1057E	-	-	-	-	-	-	-
T1205A	-	-	-	-	PASHT**T**LGV (79.25)	-	-
K1220R	-	-	-	-	TAMSWVQ**K**I (19.32)	-	-

The values given within the bracket indicates the 50% inhibitory concentration (IC_{50}) of the peptide, a measure of the binding affinity. An IC_{50} value <50 is considered as a good affinity. Mutation points in the predicted epitopes are marked as bold.