

Appendix Supplementary Information for

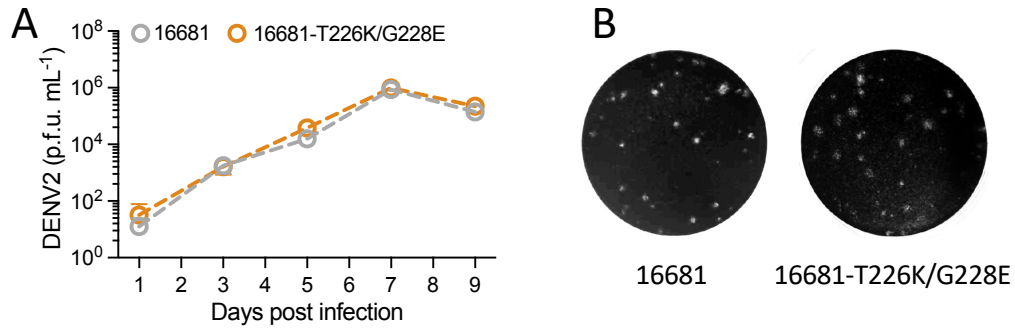
Neighboring mutation-mediated enhancement of dengue virus infectivity and spread

Lu Chen^{1,#}, Xianwen Zhang^{2,#}, Xuan Guo^{1,#}, Wenyu Peng^{1,#}, Yibin Zhu¹, Zhaoyang Wang¹, Xi Yu¹, Huicheng Shi¹, Yuhan Li¹, Liming Zhang¹, Lei Wang², Penghua Wang³, and Gong Cheng^{1,2,*}

*Correspondence: gongcheng@mail.tsinghua.edu.cn

Table of Contents:

Appendix Figure S1.....	2
Appendix Figure S2.....	3
Appendix Table S1.....	4
Appendix Table S2.....	5
Appendix Table S3.....	6
Appendix Table S4.....	8
Appendix Table S5.....	9
Appendix Table S6.....	10
Appendix Supplementary Text 1.....	11

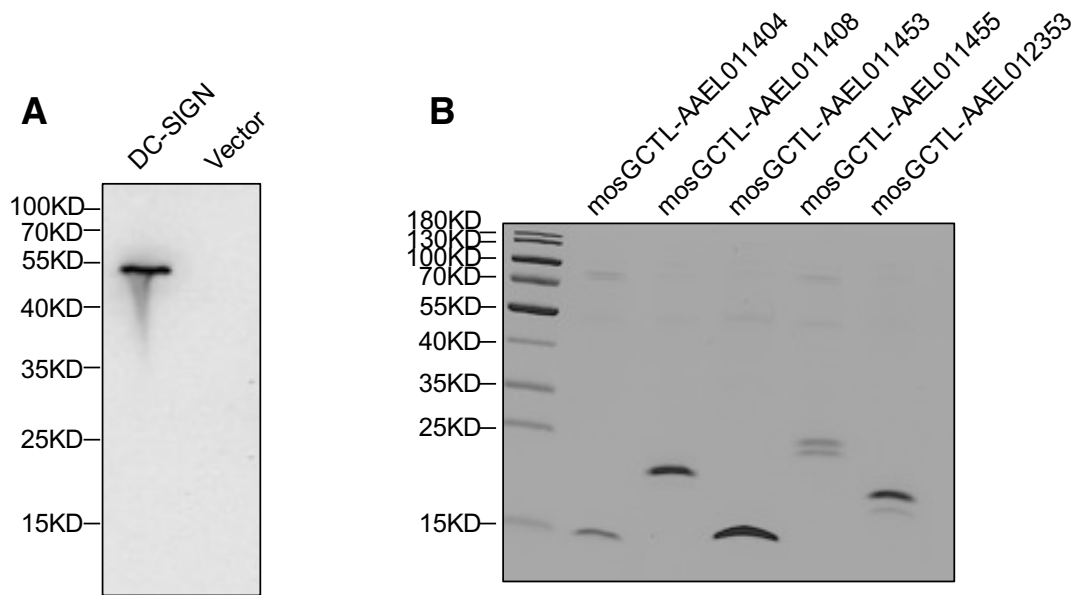


Appendix Figure S1. Growth kinetics and plaque morphology of the 16681-T226K/G228E mutant.

A Growth kinetics of the 16681-T226K/G228E mutant in Vero cells. Vero cells were infected with 0.01 MOI of viruses, and the cell supernatant was collected at Days 1, 3, 5, 7, and 9 for the plaque assay.

B Plaque morphology of the 16681 and 16681-T226K/G228E strains.

Data information: In (A), data are presented as mean \pm SEM of n = 4 biological replicates.



Appendix Figure S2. Expression of DC-SIGN and mosquito C-type lectins.

A Expression of human DC-SIGN in 293T cells. The 293T cells were transfected with a pcDNA3.1 recombinant plasmid with a human *DC-SIGN* gene. Thirty-six hours after transfection, the cells were lysed, and the expression of human DC-SIGN was detected by western blotting with an anti-flag antibody.

B Purification of mosquito C-type lectins in S2 cells. The protein concentration was measured using protein assay dye, and the protein purity was verified by SDS-polyacrylamide gel electrophoresis (PAGE).

Appendix Table S1. Sequence number of the DENV1-4 E proteins.

Serotype	Genotype	<1995	1996-2000	2001-2005	2006-2010	2011-2015	2016-2020	N/A
DENV1	I	72	42	240	1534	1339	1366	378
	II	3	0	0	0	1	0	0
	III	1	0	0	0	0	0	2
	IV	10	4	132	223	255	86	79
	V	77	116	110	470	1366	409	20
DENV2	Asian I	94	38	146	313	467	54	134
	Asian II	9	0	1	17	2	3	8
	Asian American	125	133	277	371	226	90	112
	Cosmopolitan American	27	42	160	612	1147	932	9
	American	49	2	1	1	0	0	24
	Sylvatic	15	1	0	2	0	0	4
DENV3	I	79	25	29	126	245	140	34
	II	55	57	86	193	196	0	58
	III	24	62	394	440	269	133	138
	IV	0	0	0	0	0	0	3
	V	4	0	3	4	3	0	0
DENV4	I	12	0	70	78	200	210	58
	II	34	60	46	288	649	56	158
	III	0	2	0	0	0	0	5
	Sylvatic	3	0	0	0	0	0	3

Appendix Table S2. Time-series data of the AOF of T226K and G228E substitutions in the DENV2 Asian I genotype from 1995 to 2019.

Year	t	The AOF of T226K (%)	The AOF of G228E (%)
1995	0	28.57	0
1996	1	27.27	0
1997	2	20	0
1998	3	11.11	0
1999	4	14.29	14.29
2000	5	42.86	14.29
2001	6	60.71	46.43
2002	7	66.67	66.67
2003	8	100	69.23
2004	9	92.86	85.71
2005	10	100	88.89
2006	11	98.33	100
2007	12	100	96.39
2008	13	100	92.86
2009	14	100	100
2010	15	100	98.7
2011	16	99.4	99.4
2012	17	98.66	98.66
2013	18	95.59	95.59
2014	19	85	85
2015	20	80.65	79.03
2016	21	100	100
2017	22	100	100
2018	23	100	100
2019	24	94.44	100

Appendix Table S3. Primers for qPCR and clones.

Primers for SYBR RT-QPCR	Upper primer	Lower primer
DENV2 <i>capsid</i>	CAGATCTCTGATGAATAACCAACG	CATTCCAAGTGAGAATCTCTTTGTCA
DENV3 <i>envelope</i>	ATGGAATGTGTGGGAGGTGG	GGCTTTCTATCCARTAGCCCATG
DENV4 <i>capsid</i>	GCAGATCTCTGGAAAAATGAACCA	GAGAATCTCTTCACCAACCCYTG
<i>Aedes aegypti Actin</i>	GAACACCCAGTCCTGCTGACA	TGCGTCATCTTCTCACGGTTAG
<i>Homo sapiens GAPDH</i>	AGCCTCAAGATCATCAGCAATG	ATGGACTGTGGTCATGAGTCCTT
<i>Mus musculus GAPDH</i>	TGCTGAGTATGTCGTGGAGTC	GGTTCACACCCATCACAAAC
Primers for cloning into pcDNA3.1 (+)	Upper primer	Lower primer
DC-SIGN	GGCTAGTTAAGCTTGGTACCATGAGTGACTCCAAGGAACC	TCGTCGTCATCCTTGTAATCCGCAGGAGGG GGGTTTGGGG
Primers for cloning into pET28a	Upper primer	Lower primer
DC-SIGN ectodomain	GTCGACAAGCTTGCGGCCGCGTGTCCAAGGTCCCCAGCTC	TGGTGGTGGTGGTGCTCGAGCGCAGGAGGG GGGTTTGGGG
Primers for infectious clones	Upper primer	Lower primer
BID-V1164 fragment1	ATATATCCTGTATCACATATTCTGCGGCCGCTAATACGACTCACT ATAGAGTTGTTAGTC	CTAATGTCTCCTGTCATAATGGTCAACTTTA CCTC
BID-V1164 fragment2	GACCATTATGACAGGAGACATTAGAGGAATC	TCTACTCGAGTTGAAATGTATCCTCTAGC
BID-V1164 fragment3	GAGGATACATTTCAACTCGAGTAGAGATGGGTGAAGCAG	CGACGAACCATCTCAGTTTTGC
BID-V1164 fragment4	GCAAAACTGAGATGGTTCGTCTG	CCATTCCATTTTCTGGCGTTCTG

Appendix Table S3. Primers for qPCR and clones (Continued).

Primers for infectious clones	Upper primer	Lower primer
16681 fragment1	TTAATACGACTCACTATAGAGTTGTTAGTCTACGTG GAC	CTGGTCCAGCGAGATTCTTTGGAATTATCATCTCAC
16681 fragment2	GAGATGATAATTCCAAAGAATCTCGCTGGACCAGTG	CTCTTGCCTTCTGAGTCATGAAGGTTGGGAGCC
16681 fragment3	GGCTCCCAACCTTCATGACTCAGAAGGCAAGAG	TGTCCAAATGGAGTCGTGTCTGTCATTG
16681 fragment4	CAATGACAGACACGACTCCATTTGGACA	TGCCTGCAGGTCGACTCTAGAGAACCTGTTGATTCAA CAG
PF89/27643 fragment1	CATTATACGAAGTTATATTCGATGCGGCCGCTAATAC GACTCACTATAGAGTTGTTAGTC	CATTATACGAAGTTATATTCGATGCGGCCGCTAATAC GACTCACTATAGAGTTGTTAGTC
PF89/27643 fragment2	GTGTGCGGAATCAGGTCGAC	CTCATCCATTACTATCAAGTTGTAGTTTGG
PF89/27643 fragment3	CTACAACCTTGATAGTAATGGATGAGGC	GGATCCATGGTAAGCCCACG
PF89/27643 fragment4	CGTGCGGCTTACCATGGATCC	CAGAACCTGTTGATTCAACAGCAC
H241 fragment1	CATTATACGAAGTTATATTCGATGCGGCCGCTAATAC GACTCACTATAGGGTCGTGTGG	GCTTGTCTGCATAAAGGAATGAGTG
H241 fragment2	CATTCCTTTATGCAGGACAAGC	GACCAAACAATGTTGGAATGATCCC
H241 fragment3	GGGATCATTCCAACATTGTTTGGTCCG	GATGTTGTGTTTACAGAGCTCACGA
H241 fragment4	TCGTGAGCTCTGTAAACACAACATC	CTCAACAACACCAATCCATCTCG

Appendix Table S4. The hyperparameters of the support vector machine regression (SVR) and Gaussian process regression (GPR) models.

Model	Hyperparameters	Description and function equation
SVR	Kernel function	Polynomials kernel function: $(1+xx_k)^p$; Radial basis function (Gaussian): $\exp(-\ x-x_k\ ^2/2\sigma^2)$; Multilayer perceptron or sigmoid: $\tanh(p_1xx_k+p_2)$, where x and x_k are vectors in the input space, $p > 0$, $p_1 > 0$, $p_2 < 0$, $\sigma > 0$ are parameters, and p is an integer.
	Epsilon	Half the width of the epsilon-insensitive band, stored as a nonnegative scalar value.
	Kernel scale	Numeric scale factor used to divide predictor values.
	Box constraint	Box constraints for dual problem alpha coefficients, stored as a numeric vector containing n elements, where n is the number of observations in x .
GPR	Kernel function	Exponential kernel function: $\sigma^2\exp(-(x-x')/l)$; Rational quadratic kernel function: $\sigma^2(1+(x-x')^2/2al)^{-a}$; Squared Exponential kernel function: $\sigma^2\exp(-(x-x')^2/l^2)$; Matern kernel function: $\sigma^22^{1-\nu}/\Gamma(\nu)((2\nu)^{0.5}(x-x')/l)^\nu K_\nu((2\nu)^{0.5}(x-x')/l)$; where x and x' are vectors in the input space, l is the characteristic length scale, a is the positive-valued scale-mixture parameter, σ^2 is the signal standard deviation, K_ν is the modified Bessel function and ν is a positive parameter, and Γ is the gamma function.
	Sigma	Predictor standard deviations, stored as a vector of numeric values.
	Kernel scale	Numeric scale factor used to divide predictor values.

Appendix Table S5. The model parameters and statistical parameters.

Model	Model parameters	Pearson r	p
Numerical kinetics model for the AOF of T226K	$K_{a226}=0.00364;$ $K_{d226}=0.168; k_{t226}=2.86$	0.975	< 0.0001
Numerical kinetics model for the AOF of G228E	$K_{a228}=0.0657;$ $K_{d228}=0.024; k_{t228}=2.36$	0.979	< 0.0001
GPR-1 model for the AOF of T226K	-	0.975	< 0.0001
GPR-1 model for the AOF of G228E	-	0.985	< 0.0001
SVR model for the AOF of T226K	-	0.973	< 0.0001
SVR model for the AOF of G228E	-	0.986	<0.0001

Appendix Table S6. The input data and output data for training the Gaussian process regression (GPR-2) model.

t	Input data		Output data
	The AOF of T226K (%)	The AOF of G228E (%)	The prevalence of the DENV2-Asian I (%)
3	11.11	0	13.79
4	14.29	14.29	19.44
5	42.86	14.29	13.73
8	100	69.23	14.44
9	92.86	85.71	32.18
10	100	88.89	11.92
11	98.33	100	35.29
12	100	96.39	26.77
13	100	92.86	20.51
14	100	100	18.97
15	100	98.7	20.92
16	99.4	99.4	45.9
17	98.66	98.66	40.71
18	95.59	95.59	21.18
19	85	85	8.3
20	80.65	79.03	11.31
21	100	100	5.39
22	100	100	2.18
23	100	100	6.42
24	94.44	100	12.59

Appendix Supplementary Text 1

Numerical kinetics modeling of the AOF of T226K/G228E substitutions

The kinetics of the AOF of the T226K/G228E substitution in the DENV2 Asian I genotype are described by Equations 1-2:

$$\frac{(\text{The AOF of T226K})}{dt} = R_{a226} - R_{b226} \quad 1$$

$$\frac{(\text{The AOF of G228E})}{dt} = R_{a228} - R_{b228} \quad 2$$

where t is the number of years from 1995, R_{a226} (year^{-1}) is the substitution rate of T with K at the 226th residue, R_{a228} (year^{-1}) is the substitution rate of G with E at the 228th residue, R_{b226} (year^{-1}) is the substitution rate of K with another amino acid at the 226th residue, and R_{b228} (year^{-1}) is the substitution rate of E with another amino acid at the 228th residue.

R_{a226} is proportional to the AOF of 226T:

$$R_{a226} \propto (\text{The AOF of 226WT}) \quad 3$$

$$\text{The AOF of 226WT} = 100\% - (\text{The AOF of T226K}) \quad 4$$

We assumed that R_{a226} is proportional to the power of t :

$$R_{a226} \propto t^{k_{t226}} \quad 5$$

where k_{t226} is the exponential constant of t .

Combining Equations 3-5 yields:

$$R_{a226} = k_{a226}(100\% - (\text{The AOF of T226K}))t^{k_{t226}} \quad 6$$

where k_{a226} ($\text{year}^{-1-k_{t226}}$) is the rate constant of the T226K substitution.

R_{b226} is proportional to the AOF of T226K substitution:

$$R_{b226} = k_{b226}(\text{The AOF of T226K}) \quad 7$$

where k_{b226} (year^{-1}) is the rate constant of the substitution of K with another amino acid at the 226th residue.

Combining Equations 1, 6, and 7 yields:

$$\frac{d(\text{The AOF of T226K})}{dt} = k_{a226}(100\% - (\text{The AOF of T226K}))t^{k_{t226}} -$$

$$k_{b226}(\text{The AOF of T226K}) \quad 8$$

Equation 8 is the theoretical kinetics model for the AOF of the T226K substitution.

As with the R_{a226} , R_{a228} is proportional to the AOF of 228G and the power of t :

$$R_{a228} \propto (\text{The AOF of 228WT}) \quad 9$$

$$\text{The AOF of 228WT} = 100\% - (\text{The AOF of G228E}) \quad 10$$

$$R_{a228} \propto t^{k_{t228}} \quad 11$$

where k_{t228} is the exponential constant of t .

Given that the G228E substitution compensatorily occurs based on the T226K mutation in the DENV2 Asian I genotype, we assumed that the AOF of the G228E substitution is also proportional to the AOF of the T226K substitution and the difference between the AOF of T226K and the AOF of G228E substitutions:

$$R_{a228} \propto (\text{The AOF of T226K}) \quad 12$$

$$R_{a228} \propto ((\text{The AOF of T226K}) - (\text{The AOF of G228E})) \quad 13$$

Combining Equations 9-13 yields:

$$R_{a228} = k_{a228}((\text{The AOF of T226K}) - (\text{The AOF of G228E}))(\text{The AOF of T226K})(100\% - (\text{The AOF of G228E}))t^{k_{t228}} \quad 14$$

where k_{a228} ($\text{year}^{-1-k_{t228}}$) is the rate constant of G228E substitution.

R_{b228} is proportional to the AOF of G228E substitution:

$$R_{b228} = k_{b228}(\text{The AOF of G228E}) \quad 15$$

where k_{b228} (year^{-1}) is the rate constant of the substitution of E with another amino acid at the 228th residue.

Combining Equations 1, 14, and 15 yields

$$\frac{d(\text{The AOF of G228E})}{dt} = k_{a228}((\text{The AOF of T226K}) - (\text{The AOF of G228E}))(\text{The AOF of T226K})(100\% - (\text{The AOF of G228E}))t^{k_{t228}} - k_{b228}(\text{The AOF of G228E}) \quad 16$$

Equation 16 is the theoretical kinetics model for the AOF of the G228E substitution.