

Cell Host & Microbe, Volume 22

Supplemental Information

Intrahost Selection Pressures Drive

Rapid Dengue Virus Microevolution

in Acute Human Infections

Poornima Parameswaran, Chunling Wang, Surbhi Bharat Trivedi, Meghana Eswarappa, Magelda Montoya, Angel Balmaseda, and Eva Harris

Title

Intrahost selection pressures drive rapid dengue virus microevolution in acute human infections

Authors

Poornima Parameswaran^{1,3}, Chunling Wang^{1,3}, Surbhi Bharat Trivedi¹, Meghana Eswarappa¹, Magelda Montoya¹, Angel Balmaseda² and Eva Harris^{1,4,*}

Author affiliations

¹Division of Infectious Diseases and Vaccinology, School of Public Health, University of California, Berkeley, Berkeley, CA, 94720-3370 USA

²Laboratorio Nacional de Virología, Centro Nacional de Diagnóstico y Referencia, Ministry of Health, Managua, 16064 Nicaragua

³These authors contributed equally

⁴Lead Contact

*Corresponding author: Dr. Eva Harris, Division of Infectious Diseases and Vaccinology, School of Public Health, University of California, Berkeley, 185 Li Ka Shing Center, 1951 Oxford Street, Berkeley, CA, 94720-3370, USA; Tel. 510-642-4845; FAX 510-642-6350; Email: eharris@berkeley.edu

Running Title

Intrahost virus microevolution in human dengue

Figure S1

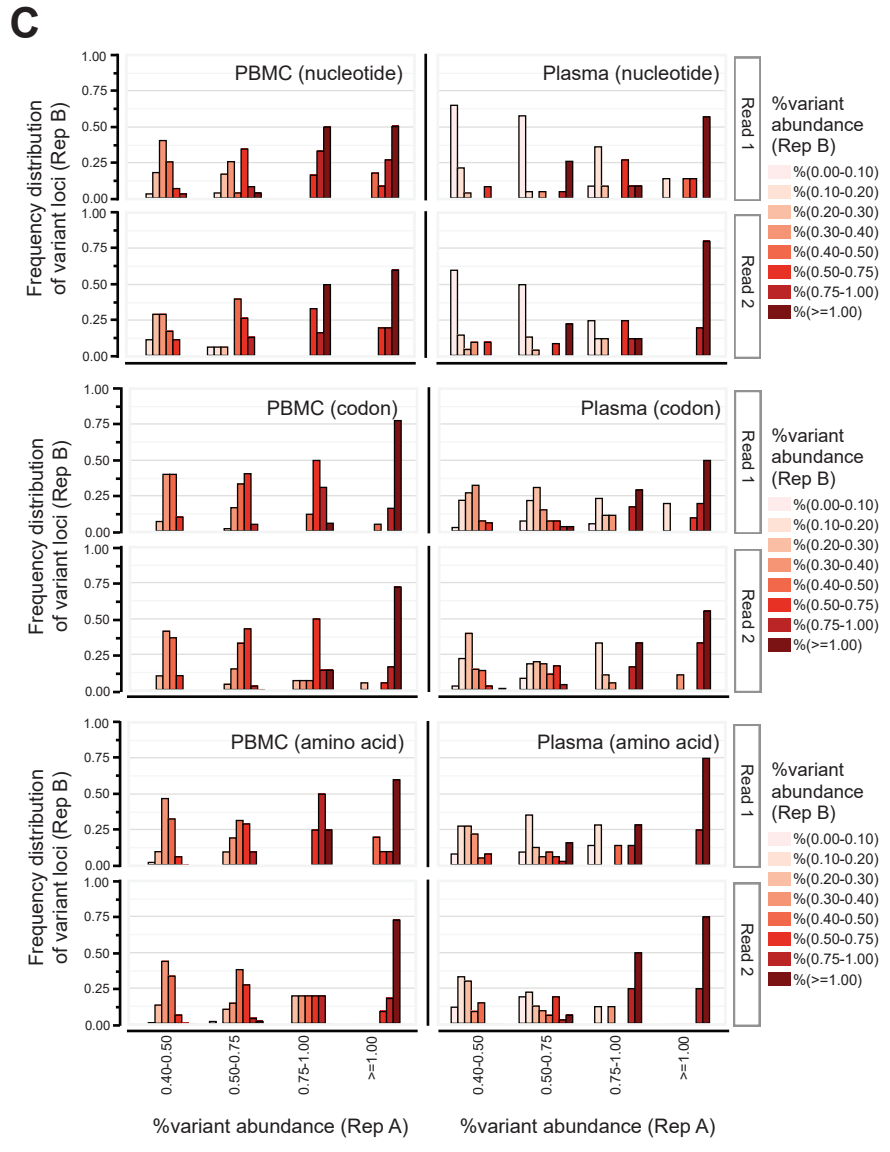
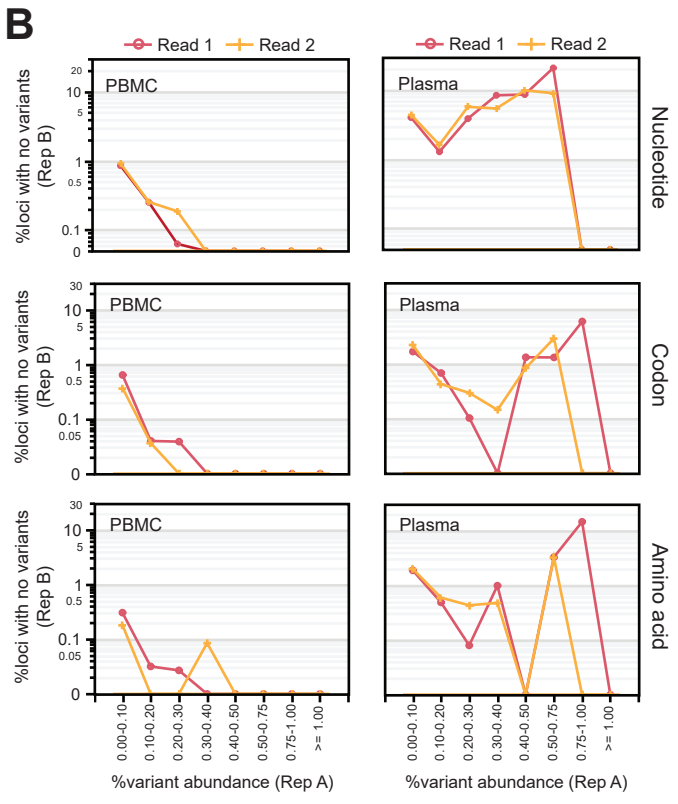
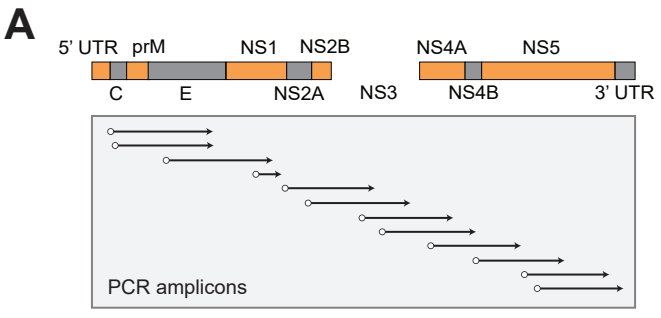


Figure S2

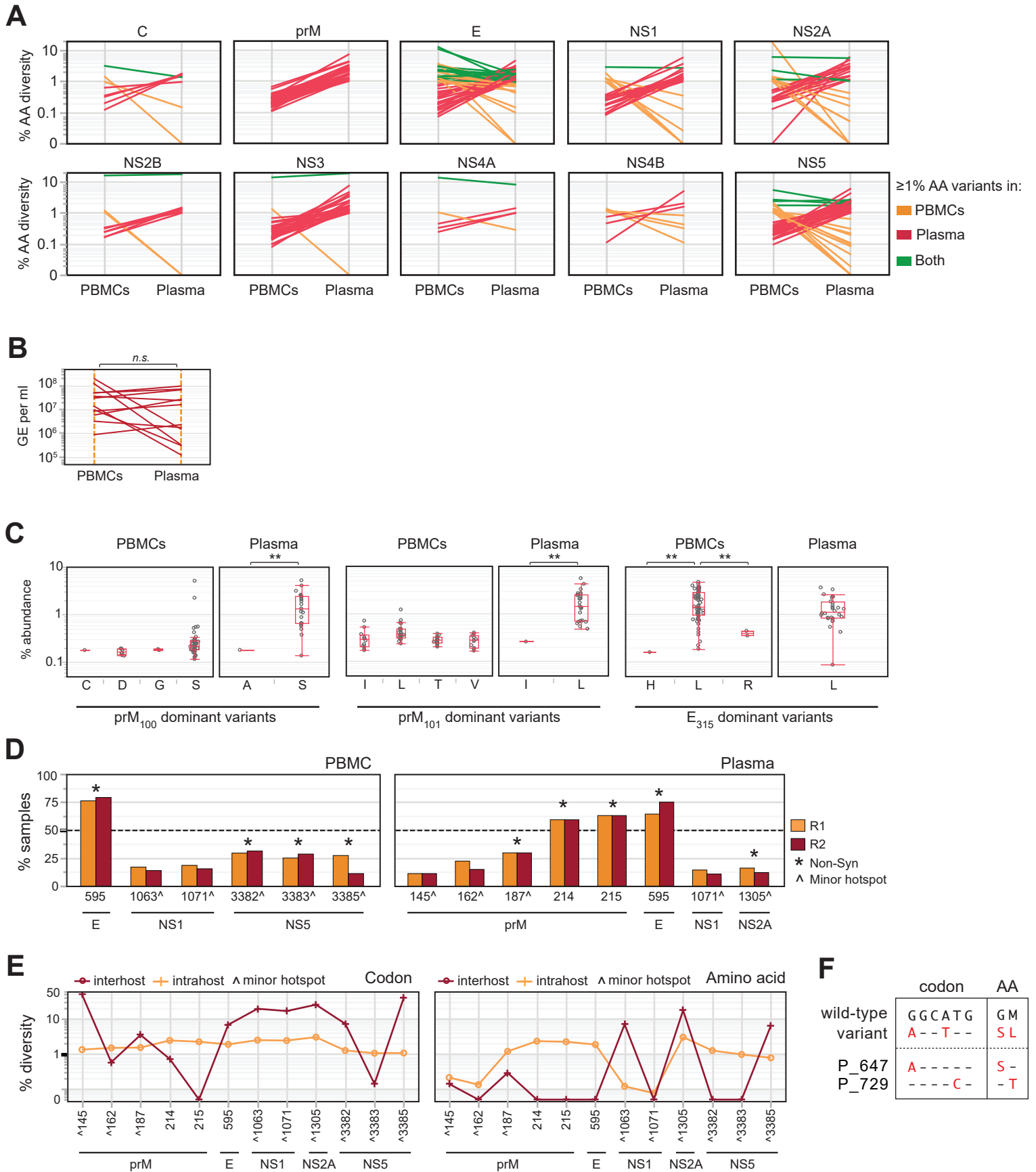


Figure S3

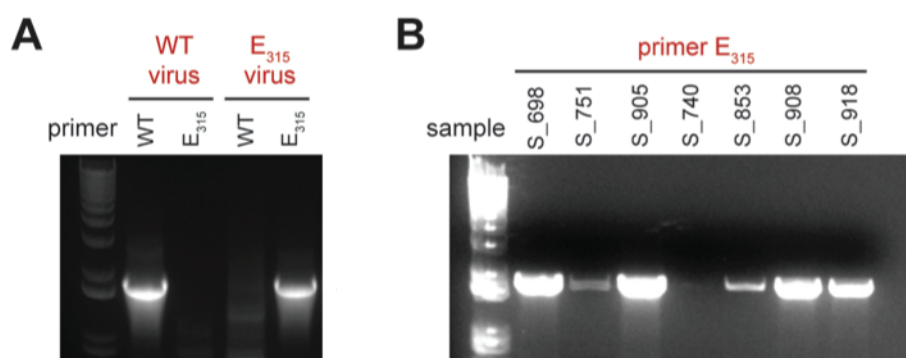


Figure S4

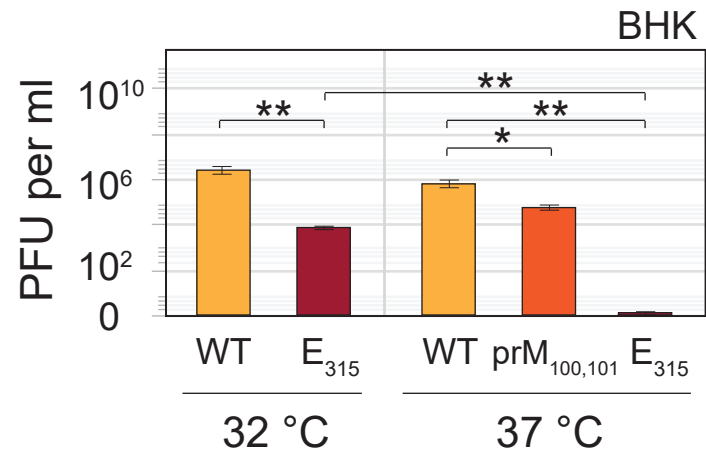
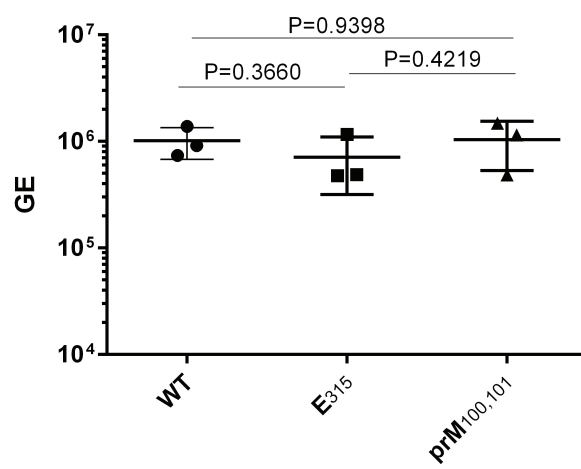


Figure S5



Supplemental Figure Legends

Figure S1. Defining thresholds for intrahost DENV-3 diversity, related to STAR Methods. (A) Sequence-directed capture of intrahost diversity in the DENV-3 coding region using twelve overlapping PCR amplicons. **(B)** For each subset of loci associated with discrete categories of percent variant diversity in replicate A (x-axis), the percent loci in replicate B exhibiting no variants is shown (y-axis). Data is pooled from replicate datasets for four PBMC (*left panels*) and three plasma (*right panels*) samples. Variant diversity is binned by read orientation (read 1 – red and read 2 - yellow), and by type of locus - nucleotide (*top panel*), codon (*middle panel*) or amino acid (*bottom panel*). The threshold for variant detection was chosen based on the minimum percent variant diversity in replicate A at which variants were also detected in replicate B (i.e., the % loci with no variants in replicate B equaled zero), for both reads 1 and 2. **(C)** Frequency distributions of loci in replicate B (y-axis) categorized by percent variant diversity (shades of red), for different cutoffs of percent variant diversity in replicate A (x-axis). Variant diversity is shown for replicate data from four PBMC (*left panels*) and three plasma (*right panels*) samples, and is binned by read orientation (read 1 or read 2) and type of locus - nucleotide (*top panel*), codon (*middle panel*) or amino acid (*bottom panel*).

Figure S2. Diversity profiles in PBMC and plasma samples at the gene and codon levels, related to Figure 1-3. (A) Gene-specific comparisons of percent amino acid variants in paired PBMC and plasma samples, colored by whether individual loci display $\geq 1\%$ AA diversity in PBMC samples only (yellow), in plasma samples only (red), or in both PBMC and plasma samples (green). **(B)** Comparisons of genome equivalents per milliliter (GE per ml) of extracted RNA in paired PBMC and plasma samples. 'n.s.' non-significant (Wilcoxon signed-rank test for paired samples). **(C)** Percent abundance of dominant non-synonymous variants observed in PBMC and plasma samples at the prM₁₀₀, prM₁₀₁ and E₃₁₅ hotspot loci, median \pm SD, boxes represent 25th and 75th percentiles, and whiskers are 10th and 90th percentiles. ** $P < 0.0001$ (Wilcoxon test for unpaired samples). **(D)** All hotspot loci (i.e., loci with $\geq 1\%$ variant abundance in $\geq 10\%$ of samples) in PBMC (*left panel*) and plasma (*right panel*) samples in

both read orientations (read 1 in yellow, and read 2 in red). Locus numbering reflects codon position in the DENV-3 polyprotein such that the dominant hotspots prM₁₀₀, prM₁₀₁ and E₃₁₅ are assigned codon numbers 214, 215 and 595, respectively. Hotspots associated with non-synonymous variants (i.e., mutations that change amino acid sequence) are highlighted with asterisks (*). The dotted line indicates the threshold for identifying the prominent hotspots (i.e., presence of hotspot in ≥50% of samples) in prM and E. Minor hotspot variants are indicated with '^'. **(E)** Average percent intrahost (red) and percent interhost (yellow) diversity in codon (*left panel*) or amino acid (*right panel*) coordinates for each hotspot locus. Interhost diversity was calculated across all global DENV-3 isolates for which full-genome sequences were available ($n = 690$). Minor hotspot variants are indicated with '^'. **(F)** Intermediate and/or alternate nucleotide and amino acid variants for the prM_{100,101} hotspot loci.

Figure S3. Detection of E₃₁₅ variant virus using Reverse Transcription and PCR (RT-PCR), related to Figure 2. (A) RT-PCR using reverse primers specific for wild-type virus (WT) or E₃₁₅ variant virus (E₃₁₅) yielded an amplified product that was 1865 nucleotide (nt) in length, from WT virus RNA template only, or E₃₁₅ variant virus RNA template only, respectively. Both WT and E₃₁₅ viruses were generated from infectious clones. **(B)** RT-PCR performed using the reverse primer specific for E₃₁₅ variant virus (primer E₃₁₅) on RNA templates extracted from a randomly-chosen subset of 7 serum samples, yielded amplified product of 1865 nt from all samples.

Figure S4. Phenotype of wild-type, prM_{100,101} and E₃₁₅ variant viruses, related to Figure 4. Plaque forming units per milliliter (PFU/ml) of C6/36-passaged WT and E₃₁₅ viruses at 32°C and 37°C, and prM_{100,101} virus at 37°C, in BHK-21 cells. No plaques were detected for the E₃₁₅ variant virus at 37°C. Data represent the mean ± SD. * $P < 0.01$ and ** $P < 0.0001$ (t -test).

Figure S5. Entry capabilities of wild-type, prM_{100,101} and E₃₁₅ variant viruses, related to Figure 4. Entry assays were performed by incubating the WT, E₃₁₅ variant, and prM_{100,101} variant viruses with U937 cells at 37°C for two hours, followed by extensive washing. Viruses were quantified by qRT-PCR

and the Genome Equivalent (GE) values were plotted on the Y-axis, mean \pm SD. *P*-values for the various comparisons were computed using a *t*-test.

Supplemental Tables

Table S1. Percent coverage across the DENV-3 genome for each patient sample, related to Table 1.

Sample type	Patient ID [^] #	Age in Years	Immune status	Disease severity	Sex	Sample collection date	% coverage (read 1)*	Average codon coverage (read 1)*	% coverage (read 2)*	Average codon coverage (read 2)*
PBMC	645	1	Primary	DF	M	09/06/2009	98.32	24357	98.10	20514
PBMC	698	7	Primary	DF	M	09/17/2009	99.71	23355	99.69	21379
PBMC	708	8	Primary	DF	F	09/25/2009	94.88	12353	94.30	10541
PBMC	716	3	Primary	DF	F	10/01/2009	98.73	7972.5	98.28	7655
PBMC	718	9	Primary	DF	F	10/04/2009	99.22	27086.5	99.22	24837.5
PBMC	729	14	Primary	DF	F	10/11/2009	99.10	19225	99.10	17433
PBMC	748	10	Primary	DF	F	09/25/2009	56.61	6716	55.50	6308
PBMC	751	10	Primary	DF	M	09/30/2009	98.33	26492	98.26	23950
PBMC	760	2	Primary	DF	M	10/05/2009	99.37	20094	99.36	18601.5
PBMC	765	5	Primary	DF	M	10/13/2009	99.72	29693.5	99.72	26486.5
PBMC	789	1	Primary	DF	M	10/13/2009	91.38	27174	90.88	25142.5
PBMC	828	5	Primary	DF	F	10/21/2009	90.18	8497	88.94	7891.5
PBMC	831	4	Primary	DF	F	10/20/2009	70.62	9696	66.54	7539
PBMC	832	3	Primary	DF	M	10/20/2009	98.13	29959	98.00	22980
PBMC	839	3	Primary	DF	F	10/23/2009	99.77	27094	99.66	22461.5
PBMC	844	2	Primary	DF	M	10/25/2009	99.49	23128	99.30	18748
PBMC	849	12	Primary	DF	M	10/26/2009	99.52	23127	99.47	19292
PBMC	905	6	Primary	DF	M	11/18/2009	99.70	20021	99.49	15207
PBMC	921	3	Primary	DF	F	11/28/2009	88.41	9714.5	87.74	8234
PBMC	922	7	Primary	DF	F	12/07/2009	88.73	8719	87.30	7408
PBMC	924	2	Primary	DF	M	12/07/2009	99.26	21616	98.99	17800.5
PBMC	927	6	Primary	DF	M	11/19/2009	99.71	18250	99.62	14752
PBMC	934	10	Primary	DF	M	11/28/2009	99.11	16879	98.93	14792.5
PBMC	938	9	Primary	DF	M	11/29/2009	99.72	32155.5	99.69	27281
PBMC	939	10	Primary	DF	F	12/07/2009	99.49	41328.5	99.43	35533.5
PBMC	944	8	Primary	DF	M	12/11/2009	99.88	22253	99.86	19311
PBMC	947	4	Primary	DF	F	12/12/2009	99.88	34318.5	99.88	30020
PBMC	950	4	Primary	DF	M	12/12/2009	98.68	32089	98.65	27836
PBMC	950R	4	Primary	DF	M	12/12/2009	98.57	28363	98.53	25061
PBMC	955	3	Primary	DF	M	12/19/2009	99.24	31680	99.24	27881.5
PBMC	955R	3	Primary	DF	M	12/19/2009	98.44	23452	98.40	21361
PBMC	956	1	Primary	DF	F	12/20/2009	99.89	34099	99.89	31041
PBMC	957	3	Primary	DF	F	12/18/2009	99.88	46762	99.88	42269.5
PBMC	957R	3	Primary	DF	F	12/18/2009	98.82	79693.5	98.81	69010.5
PBMC	1005	9	Primary	DF	M	01/17/2010	99.87	57352	99.86	50548
PBMC	1005R	9	Primary	DF	M	01/17/2010	98.67	30928	98.65	28224
PBMC	647	6	Primary	DHF	M	09/18/2009	77.77	5968	76.93	5009
PBMC	710	6	Primary	DHF	M	09/25/2009	99.52	20294	99.52	17935

PBMC	801	2	Primary	DHF	F	10/26/2009	98.58	32990.5	98.56	30312.5
PBMC	699	7	Primary	DSS	M	09/18/2009	97.54	15235	97.50	13659.5
PBMC	949	9	Primary	DSS	M	12/12/2009	99.64	22635	99.62	20165
PBMC	631	6	Secondary	DF	M	09/07/2009	66.10	6100	62.62	5751
PBMC	632	6	Secondary	DF	M	09/07/2009	98.05	18262	97.95	16261
PBMC	646	5	Secondary	DF	F	09/19/2009	95.93	5822	94.40	4982.5
PBMC	652	8	Secondary	DF	F	09/29/2009	96.80	8873	96.59	7706.5
PBMC	655	11	Secondary	DF	M	10/04/2009	65.73	6552	61.93	6103
PBMC	656	9	Secondary	DF	F	10/05/2009	99.03	18862	99.01	16160
PBMC	694	10	Secondary	DF	M	09/23/2009	98.53	23765	98.51	20980
PBMC	697	6	Secondary	DF	M	09/20/2009	99.69	15933	99.67	14163.5
PBMC	714	8	Secondary	DF	F	09/29/2009	98.24	26650	98.17	23905
PBMC	715	10	Secondary	DF	F	09/28/2009	96.94	20799	96.88	19146
PBMC	722	12	Secondary	DF	F	10/09/2009	75.16	3192	71.88	3037
PBMC	726	6	Secondary	DF	M	10/10/2009	98.94	18862	98.96	16855
PBMC	737	12	Secondary	DF	M	10/12/2009	98.98	25059	98.83	21842
PBMC	740	4	Secondary	DF	F	10/12/2009	99.44	14024	99.09	12111.5
PBMC	792	6	Secondary	DF	F	10/19/2009	57.97	6185	57.35	5689
PBMC	793	7	Secondary	DF	M	10/19/2009	99.22	20080	99.14	18858.5
PBMC	825	10	Secondary	DF	F	10/19/2009	99.63	41146	99.35	36957
PBMC	830	10	Secondary	DF	F	10/20/2009	99.24	16291.5	96.02	12751
PBMC	843	10	Secondary	DF	M	10/24/2009	95.98	44036	95.95	38892
PBMC	853	12	Secondary	DF	F	10/26/2009	99.72	22125	99.59	17492
PBMC	918	8	Secondary	DF	F	11/16/2009	99.12	14107	98.73	11335
PBMC	946	7	Secondary	DF	F	12/12/2009	99.71	22971.5	99.70	19424
PBMC	989	5	Secondary	DF	M	01/27/2010	99.59	11506.5	99.50	10864
PBMC	728	12	Secondary	DHF	M	10/09/2009	88.70	14226	88.67	12810
PBMC	731	5	Secondary	DHF	F	10/10/2009	86.95	36923	86.83	33841
PBMC	833	10	Secondary	DHF	M	10/22/2009	99.57	28447.5	99.43	23109.5
PBMC	848	8	Secondary	DHF	M	10/26/2009	99.58	22216.5	99.36	16733.5
PBMC	937	12	Secondary	DHF	M	11/29/2009	99.72	25700	99.71	21772
PBMC	959	4	Secondary	DHF	F	01/04/2010	98.59	26513.5	98.54	23619
PBMC	717	12	Secondary	DSS	F	10/04/2009	99.26	37144	99.26	30921
PBMC	847	10	Secondary	DSS	F	10/24/2009	97.57	28295	97.56	23012
Plasma	698	7	Primary	DF	M	09/17/2009	99.32	30860	99.04	24554
Plasma	718	9	Primary	DF	F	10/04/2009	99.65	22817.5	99.63	20161
Plasma	751	10	Primary	DF	M	09/30/2009	97.55	8484.5	97.45	7150
Plasma	763	7	Primary	DF	F	10/07/2009	88.76	15506	88.76	13640
Plasma	832	3	Primary	DF	M	10/20/2009	85.72	9447	77.49	19887
Plasma	844	2	Primary	DF	M	10/25/2009	99.34	28214	99.32	22459.5
Plasma	849	12	Primary	DF	M	10/26/2009	76.70	7388.5	76.56	5946.5
Plasma	863	6	Primary	DF	F	10/31/2009	50.75	41668	50.72	30857
Plasma	887	2	Primary	DF	M	11/09/2009	51.42	56349	51.35	48253
Plasma	897	4	Primary	DF	F	11/15/2009	97.34	11099	96.92	9293
Plasma	905	6	Primary	DF	M	11/18/2009	99.67	25622	99.67	23884
Plasma	909	5	Primary	DF	M	11/02/2009	97.46	40600.5	97.43	36203
Plasma	927	6	Primary	DF	M	11/19/2009	99.77	15265	99.71	13156

Plasma	938	9	Primary	DF	M	11/29/2009	99.92	26900.5	99.90	22439
Plasma	939	10	Primary	DF	F	12/07/2009	99.74	21931	99.72	20123
Plasma	944	8	Primary	DF	M	12/11/2009	99.82	36336	99.82	33207
Plasma	646	5	Secondary	DF	F	09/19/2009	99.80	9317	99.75	7654.5
Plasma	656	9	Secondary	DF	F	10/05/2009	99.79	23477	99.72	19973
Plasma	697	6	Secondary	DF	M	09/20/2009	89.64	29521	89.35	25767
Plasma	722	12	Secondary	DF	F	10/09/2009	99.47	21656	99.44	18554.5
Plasma	722R	12	Secondary	DF	F	10/09/2009	92.62	13972	91.91	12277.5
Plasma	737	12	Secondary	DF	M	10/12/2009	88.33	14093	88.27	11744
Plasma	740	4	Secondary	DF	F	10/12/2009	95.24	10473	92.92	9816
Plasma	742	9	Secondary	DF	F	10/13/2009	98.52	19365.5	98.43	16830
Plasma	830	10	Secondary	DF	F	10/20/2009	51.28	41408	51.33	36802
Plasma	853	12	Secondary	DF	F	10/26/2009	99.42	83233	99.35	73919
Plasma	868	13	Secondary	DF	M	11/02/2009	84.87	58334	83.43	51339
Plasma	908	9	Secondary	DF	F	11/04/2009	93.88	18502	91.27	16125
Plasma	918	8	Secondary	DF	F	11/16/2009	99.95	43136	99.96	40261
Plasma	936	14	Secondary	DF	M	11/30/2009	98.50	31324.5	98.50	27792
Plasma	936R	14	Secondary	DF	M	11/30/2009	99.71	28621.5	99.71	26402
Plasma	937	12	Secondary	DHF	M	11/29/2009	99.82	22978.5	99.81	20074
Plasma	847	10	Secondary	DSS	F	10/24/2009	99.74	27811	99.71	24865.5
Plasma	847R	10	Secondary	DSS	F	10/24/2009	99.90	57468	99.90	51485

^The suffix 'R' denotes replicate sample used in defining thresholds for percent intrahost diversity.

#All samples used for this study were from individuals who presented with DENV-3 infection during the 2009-2010 epidemic season in Nicaragua. Over 100 other clinical and laboratory parameters are available for these samples, but not shown due to space constraints.

*Only nucleotide loci with coverage of ≥ 1000 reads and locus-specific Sanger quality scores of ≥ 30 were considered for calculating genome coverage. Samples with $\geq 50\%$ genome coverage are reported.

Table S2. Interhost diversity at hotspots for intrahost diversity, related to Figure 2.

Codon position	Primary [#] codons (intrahost)	Variant* codons (intrahost)	Primary [#] codons (interhost [^])	Variant* codons (interhost [^])	Primary [#] AA (intrahost)	Variant* AA (intrahost)	Primary [#] AA (interhost [^])	Variant* AA (interhost [^])
145	ATT, ATC	ATC, ATT	ATC	ATT, (ACT)	I	V, (T)	I	T
162	ACG, ACA	ACA, ACG	ACG	ACA	T	A	T	-
187	ACA	GCT, GCA, (ACT)	ACA	ACG, (GCA, GCG)	T	A	T	A
214	GGC	AGC	GGC	GGT	G	S	G	-
215	ATG	TTG	ATG	-	M	L	M	-
595	CAT	CTT	CAT	CAC	H	L	H	-
1063	AAT, AAC	AAC, AAT	AAC	AAT, AGC, (GAC, GAT)	N	S, (K, D, H, I, T)	N	S, D
1071	TTA, TTG	TTG, TTA	TTG	CTG, TTA	L	F, S, (*, V, I)	L	-
1305	GTG, ATG	GCG, GTG, ATG, GTA	GTG	CTG, ATG, (TTG, GTA)	V, M	A, M, V	V	L, M
3382	GAG	GGG	GAG	GAA	E	G	E	-
3383	GAG	GGG	GAG	GAA	E	G	E	-
3385	TCG	TTG, TCA	TCG	TCA, TTG, (ACA, TTA, CCA)	S	L	S	L, T, (P)

[^]For calculating interhost diversity, all available full-length DENV-3 genomes from global isolates ($n = 690$) were used

[#]Codons and amino acids with the highest prevalence are listed as 'primary'

*Minor codons and amino acids are listed as 'variant' (in decreasing order of prevalence)

Table S3. Frequencies of various haplotypes linked to variants at the prM_{100,101} and E₃₁₅ hotspot loci, related to Figure 3.

Variant	Haplotypes linked to variant	Frequency (predicted) [^]	Frequency (observed) [#]	Number of samples with variant haplotype identical to wild-type consensus haplotype (percent of all samples with variant haplotype)
<i>prM</i> _{100,101}	C-G-C	0.702	0.688	3 (100%)
<i>prM</i> _{100,101}	T-G-C	0.170	0.195	2 (100%)
<i>prM</i> _{100,101}	C-A-C	0.093	0.104	2 (100%)
<i>prM</i> _{100,101}	C-A-T	0.001	0.013	1 (100%)
<i>E</i> ₃₁₅	G-G-A	0.533	0.636	24 (100%)
<i>E</i> ₃₁₅	A-G-A	0.304	0.195	7 (100%)
<i>E</i> ₃₁₅	A-A-A	0.040	0.117	4 (100%)
<i>E</i> ₃₁₅	A-G-G	0.017	0.052	4 (100%)

[^]Product of the frequencies of alleles at each of the three loci, calculated from Nicaraguan consensus DENV-3 genomes used in this study.

[#]Frequencies of haplotypes observed in Nicaraguan consensus DENV-3 genomes used in this study.

Table S4. Sequences of primers used for amplifying the DENV-3 genome, related to STAR**Methods.**

Sequence	Description	Tm	Amplicon#	Comments
GACTCGGAAGCTTGCTTAAC	DENV3_32_F	62.00	DENV3-Amplicon1	
TGATCCAAAGTCCCAAGCTG	DENV3_2196_R	61.18		
TCAATATGCTGAAACGCGTG	DENV3_128_F	60.81	DENV3-Amplicon2	
TGATCCAAAGTCCCAAGCTG	DENV3_2196_R	61.18		
AAGCTATGCATTGAGGGGAA	DENV3_1106_F	60.04	DENV3-Amplicon3	
GTATCGCAAGGGAGGRAGTG	DENV3_3378_R	59.69		
CATATGGCTGAAACTCCGAG	DENV3_2907_F	58.33	DENV3-Amplicon4	
CCAAGCACAAAGACACCCATT	DENV3_3511_R	60.95		
GGARGTGAAAGGTGGACAA	DENV3_3466_F	59.94	DENV3-Amplicon5	R=A or G
ATRCTGGCTGGGTCTGTGAA	DENV3_5393_R	58.16		
TTGACTGTTGCTGGAGAAC	DENV3_3946_F	58.85	DENV3-Amplicon6	
CGTCTATGGCGCYGACTT	DENV3_6073_R	59.46		
AAACGAACGCAGAACCAGAT	DENV3_5009_F	59.74	DENV3-Amplicon7	
GCGTACAATGTCCAGGCTGA	DENV3_6941_R	60.47		
CTCGTGTTGGAATGGGAGAG	DENV3_5417_F	60.17	DENV3-Amplicon8	
ACAYAAGGCCCATGATGTTC	DENV3_7389_R	60.79		
GTCAATCRCCCTTGATCTTG	DENV3_6366_F	60.60	DENV3-Amplicon9	R=G or A
ATGTTACCTGTRCCATTGGA	DENV3_8240_R	58.85		
CGGTRGATGGGATAATGACA	DENV3_7253_F	61.53	DENV3-Amplicon10	R=G or A
CTCCGGGTATCTTGAAAATG	DENV3_9139_R	59.38		
ATCCAATGGYACAGGTAACA	DENV3_8220_F	58.85	DENV3-Amplicon11	Y=T or C
GATTGTCCTCGATCCACACC	DENV3_10024_R	60.33		
ATGGATCTTATGAAGTCAAAGC	DENV3_8478_F	58.86	DENV3-Amplicon12	
ACTGTGGCTTAARTGGCCT	DENV3_10302_R	59.62		