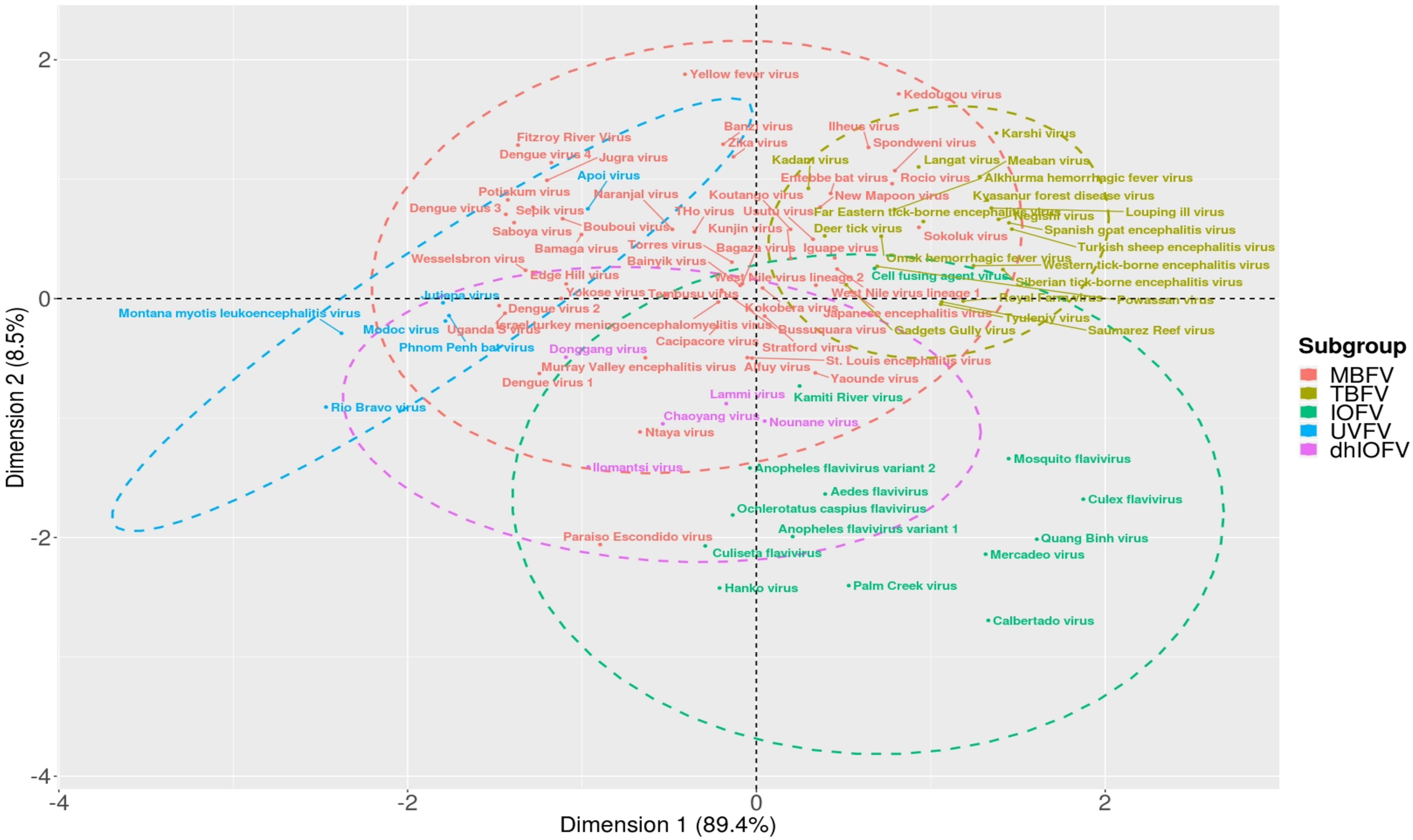
SUPPLEMENTARY FIGURES

An unsupervised algorithm for host identification in flaviviruses

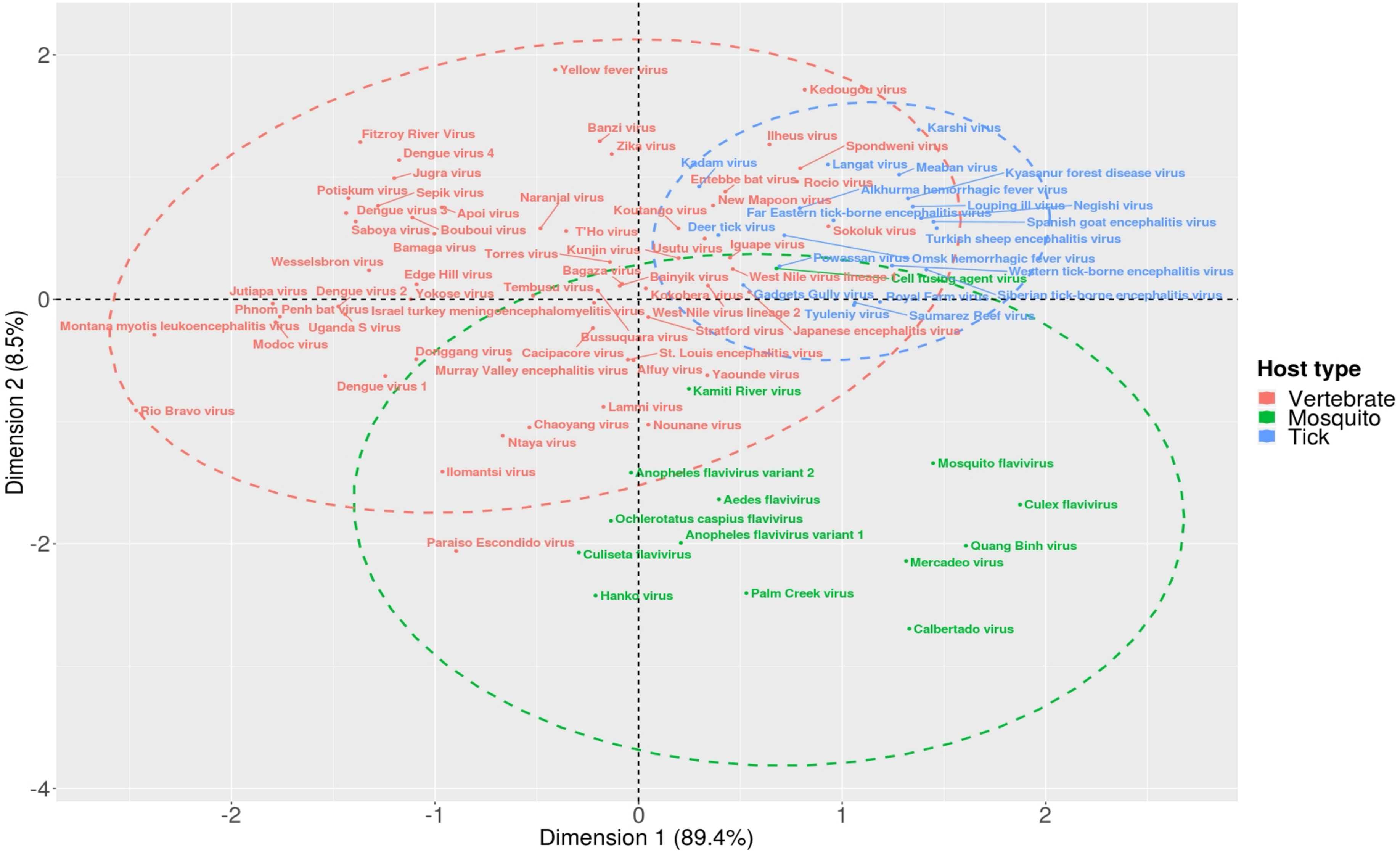
Phuoc Truong Nguyen [1, 2], Santiago Garcia-Vallve [3], Pere Puigbò [1, 4, 5, *]

[1] *Department of Biology, University of Turku, Turku, Finland.*
[2] *Currently at the Department of Virology, Medicum, Faculty of Medicine, University of Helsinki, Helsinki, Finland.*
[3] *Research Group in Cheminformatics & Nutrition, Department of Biochemistry and Biotechnology, Rovira i Virgili University, Tarragona, Catalonia, Spain*
[4] *Currently at the Department of Biochemistry and Biotechnology, Rovira i Virgili University, Tarragona, Catalonia, Spain*
[5] *Currently at the Nutrition and Health Unit, Eurecat Technology Centre of Catalonia, Reus, Catalonia, Spain*

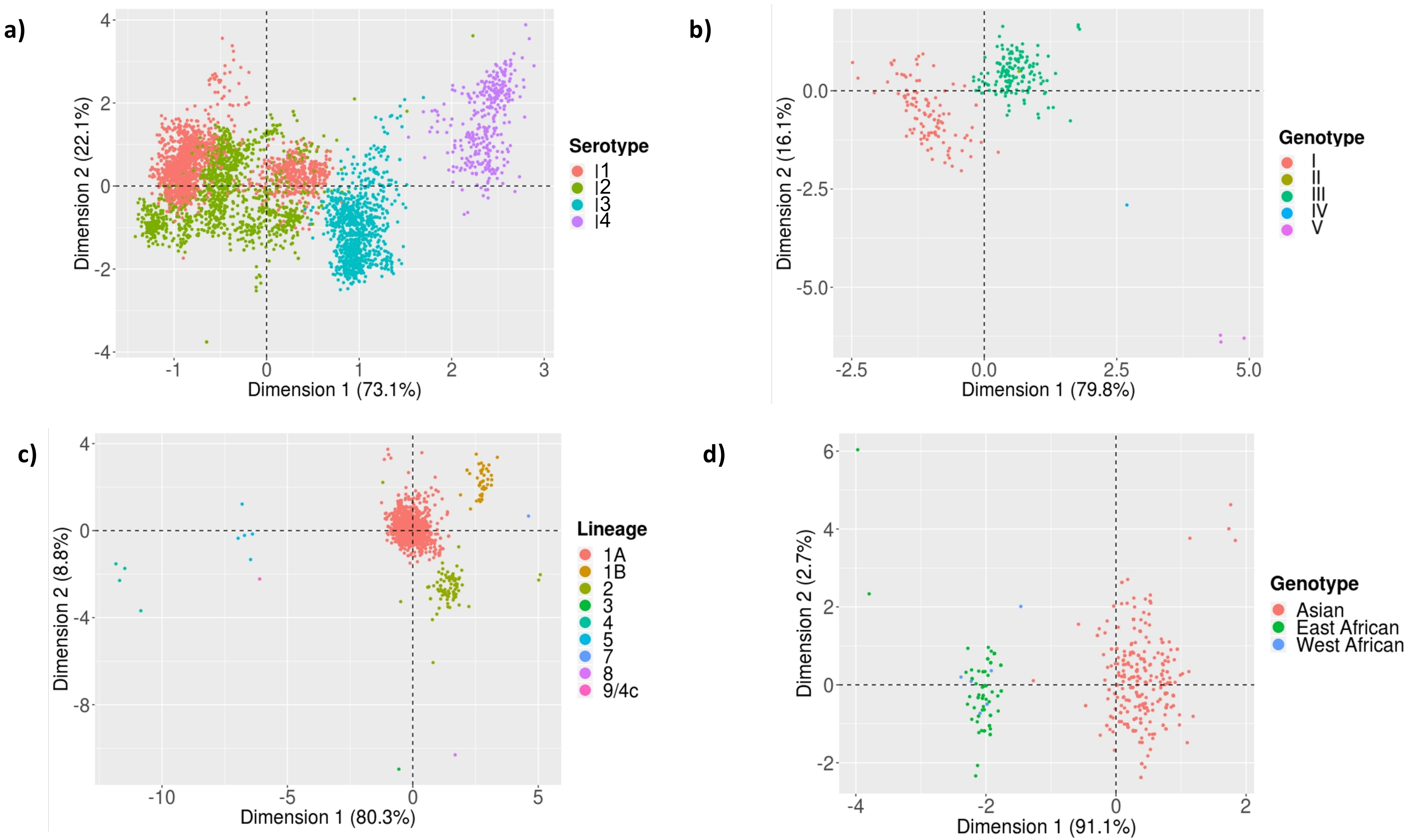
[*] *Corresponding author: pepuav@utu.fi*



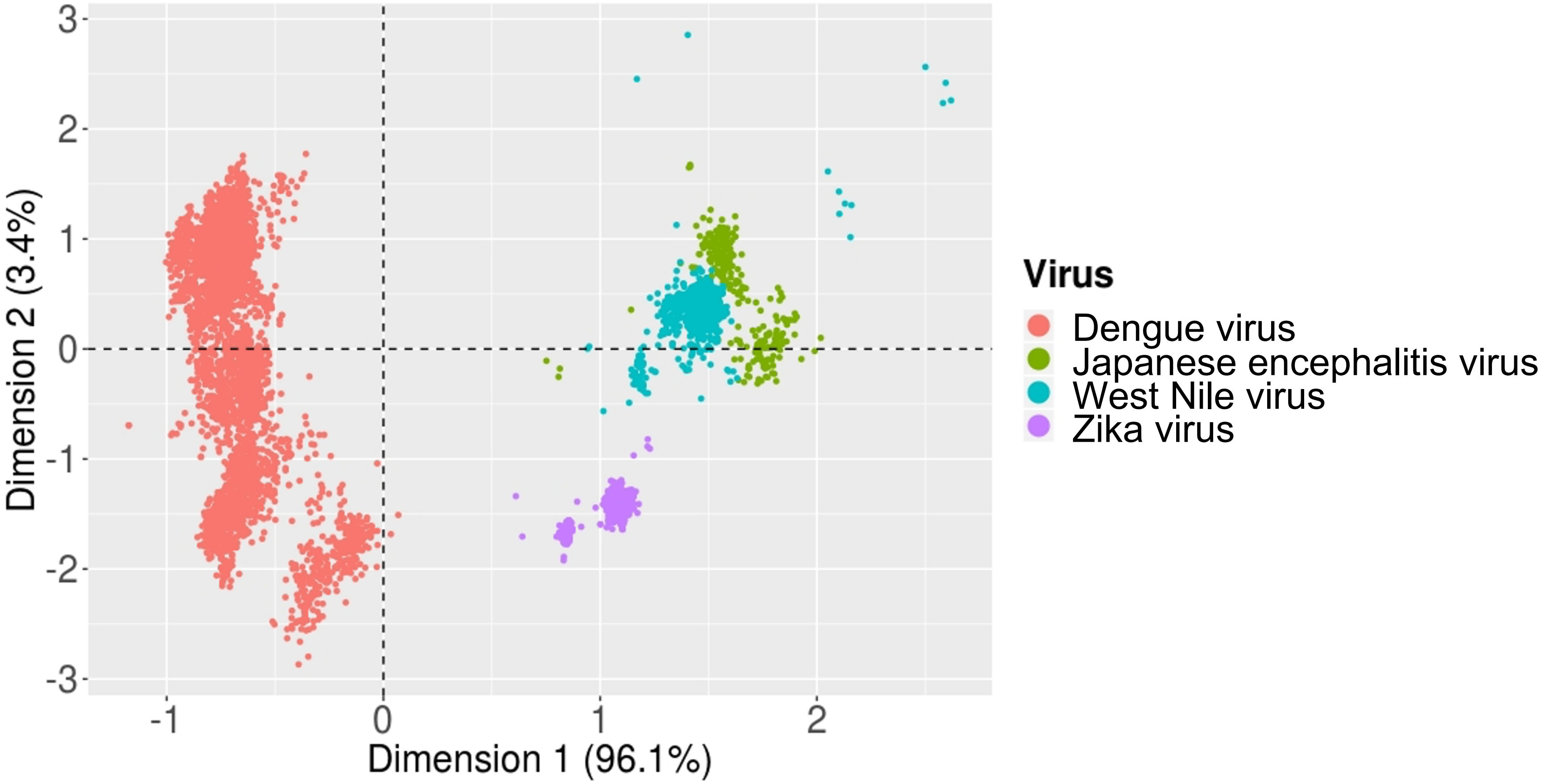
Supplementary figure S1. Interspecies correspondence analysis and subgroup centroids of normalized Codon Adaptation Index (nCAI) values of flaviviruses, genus *Flavivirus* (N = 94). While there is a distinct separation between the clusters of insect-only and the other flavivirus subgroups, the clouds (shown as dashed circles with colors corresponding to a flavivirus subgroup) computed from them show major overlap, indicating that most of the subgroups share similar host preferences, e.g. unknown vector flaviviruses most likely have similar hosts as mosquito-borne flaviviruses. An interesting case can be observed with Paraiso Escondido virus, which is located outside the mosquito-borne cloud, which suggests that it does not have a vertebrate host. This is supported by previous studies, in which the virus is described to infect only sandflies. MBFV = Mosquito-borne flavivirus, TBFV = Tick-borne flavivirus, IOFV = Insect-only flavivirus, UVFV = Unknown vector flavivirus, and dhIOFV = Dual-host insect-only flavivirus. The centroids were computed based on the multivariate normal distribution of each subgroup with a confidence level of 0.95. Dimension 1 explains 89.4 percent and dimension 2 contributes to 8.5 percent of the variation.



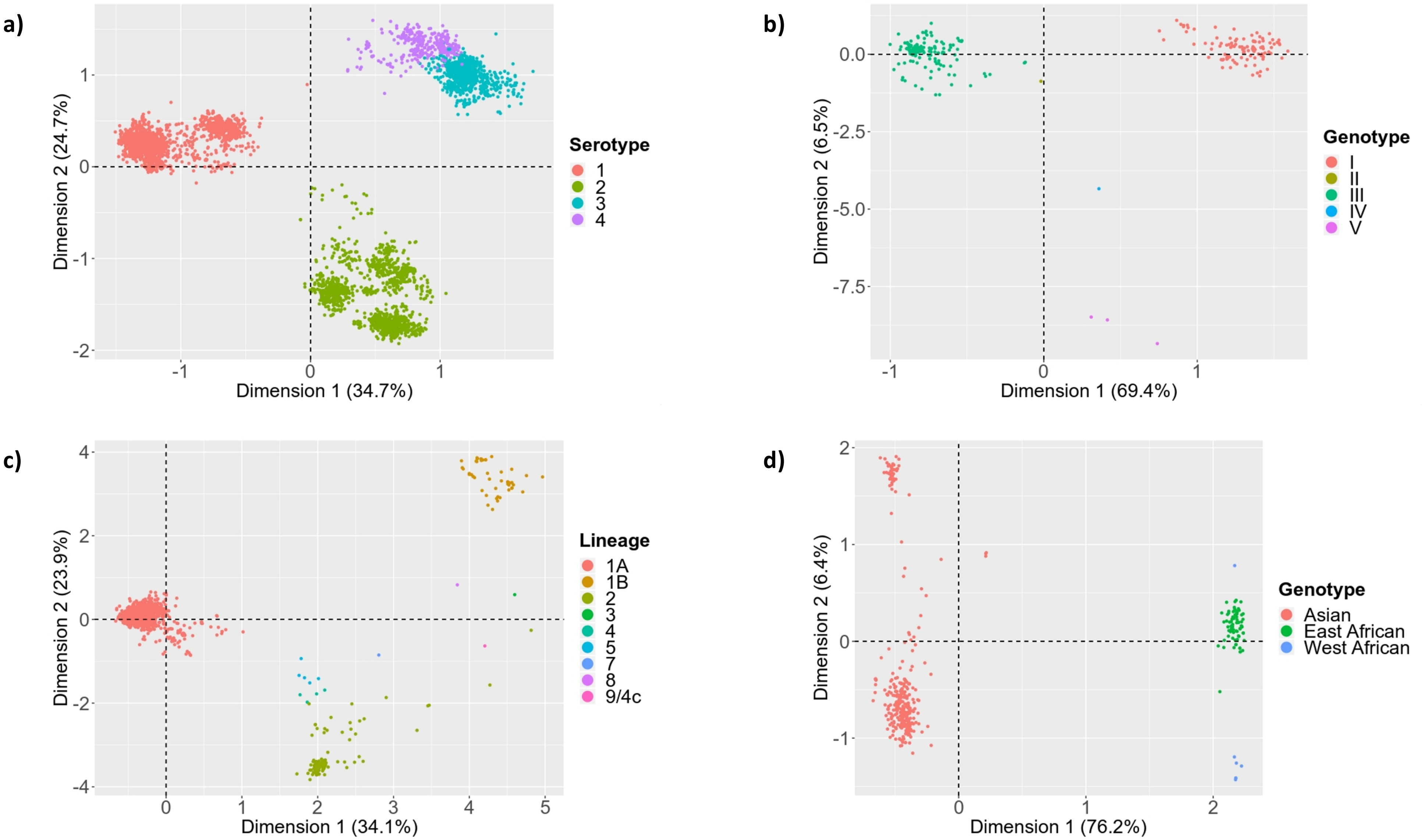
Supplementary figure S2. Interspecies correspondence analysis with host type centroids of normalized Codon Adaptation Index (nCAI) values of flaviviruses, genus *Flavivirus* (N = 94). When the flaviviruses are categorized based on their preferred host type, we can observe two major clouds (dashed circles) and clusters; viruses that have a vertebrate host and those that have a mosquito-host. Tick-borne viruses largely overlap with the vertebrate host cluster, suggesting that while tick-borne flaviviruses cluster separately, they have the same type of host as mosquito-borne viruses. An interesting exception in the vertebrate cluster is Paraiso Escondido virus, which is located outside the cloud. This may be due to the virus infecting mainly sandflies rather than vertebrates. The centroids were computed based on the multivariate normal distribution of each host type with a confidence level of 0.95. Dimension 1 explains 89.4 percent and dimension 2 contributes to 8.5 percent of the variation.



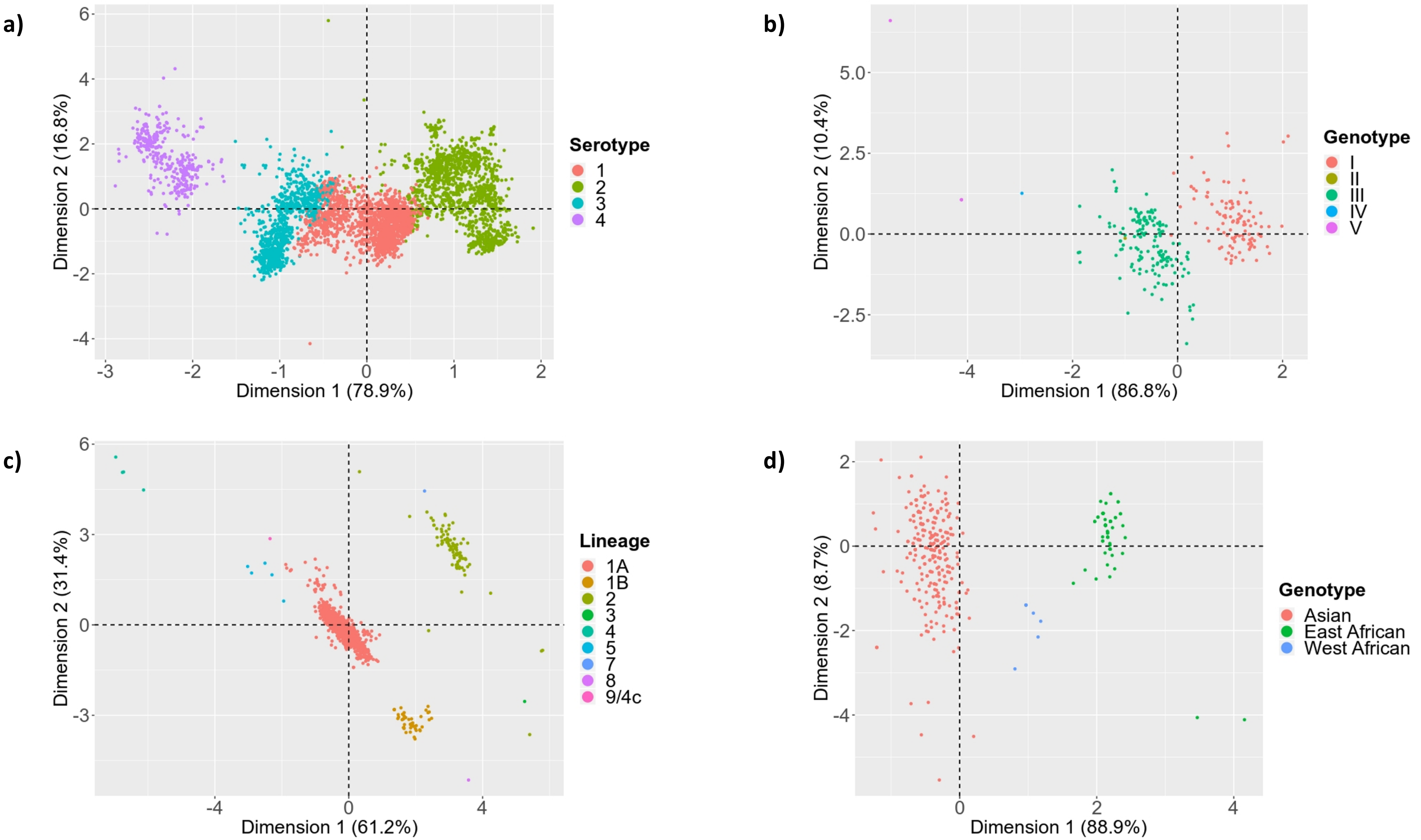
Supplementary figure S3. Intraspecies correspondence analyses of normalized Codon Adaptation Index (nCAI) values of four major mosquito-borne flaviviruses (genus *Flavivirus*). The results show that nCAI is able to discriminate between the different categories of (a) Dengue viruses (N = 4865), (b) Japanese encephalitis viruses (N = 297), (c) West Nile viruses (N = 1619) and Zika viruses (N = 494). Additionally, the distances between separate clusters mirror actual phylogeny. For example, Dengue virus serotypes 1–3, which are clustered close together, are more related to each other than to serotype 4, and the same can be observed with the results of the Zika virus genotypes, in which the Asian and African genotypes cluster separately, but nCAI does not distinguish East and West African genotypes from each other. In panel a, Dimension 1 explains 73.1 percent and Dimension 2 contributes to 22.1 percent of the variability. In panel b, Dimension 1 explains 79.8 percent and Dimension 2 contributes to 16.1 percent of the variability. In panel c, Dimension 1 explains 80.3 percent and Dimension 2 contributes to 8.8 percent of the variability. In panel d, Dimension 1 explains 91.1 percent and Dimension 2 contributes to 2.7 percent of the variability.



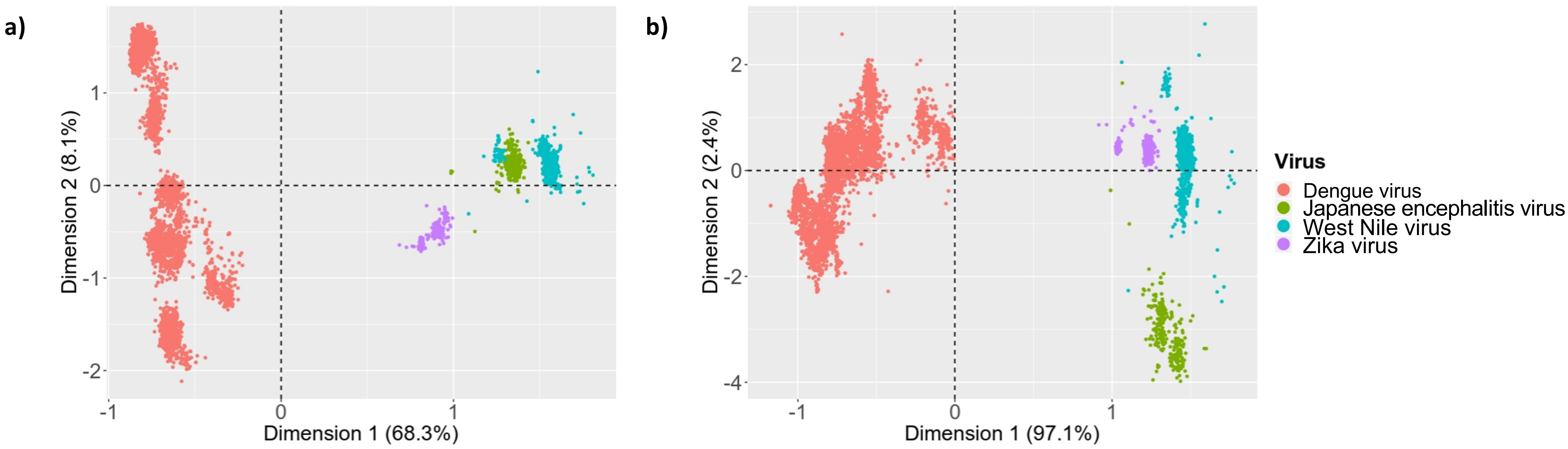
Supplementary figure S4. Interspecies correspondence analysis of normalized Codon Adaptation Index (nCAI) values of four major mosquito-borne flaviviruses, genus *Flavivirus* (N = 7275). The different virus genomes form distinct clusters based on computed nCAI values. The distance between these clusters mirror some phylogenetic relationships among the viruses accurately, e.g. Japanese encephalitis viruses and West Nile viruses share a more recent ancestor than to the other viruses. Of the variation, 96.1 percent is due to Dimension 1 and 3.4 percent is because of Dimension 2.



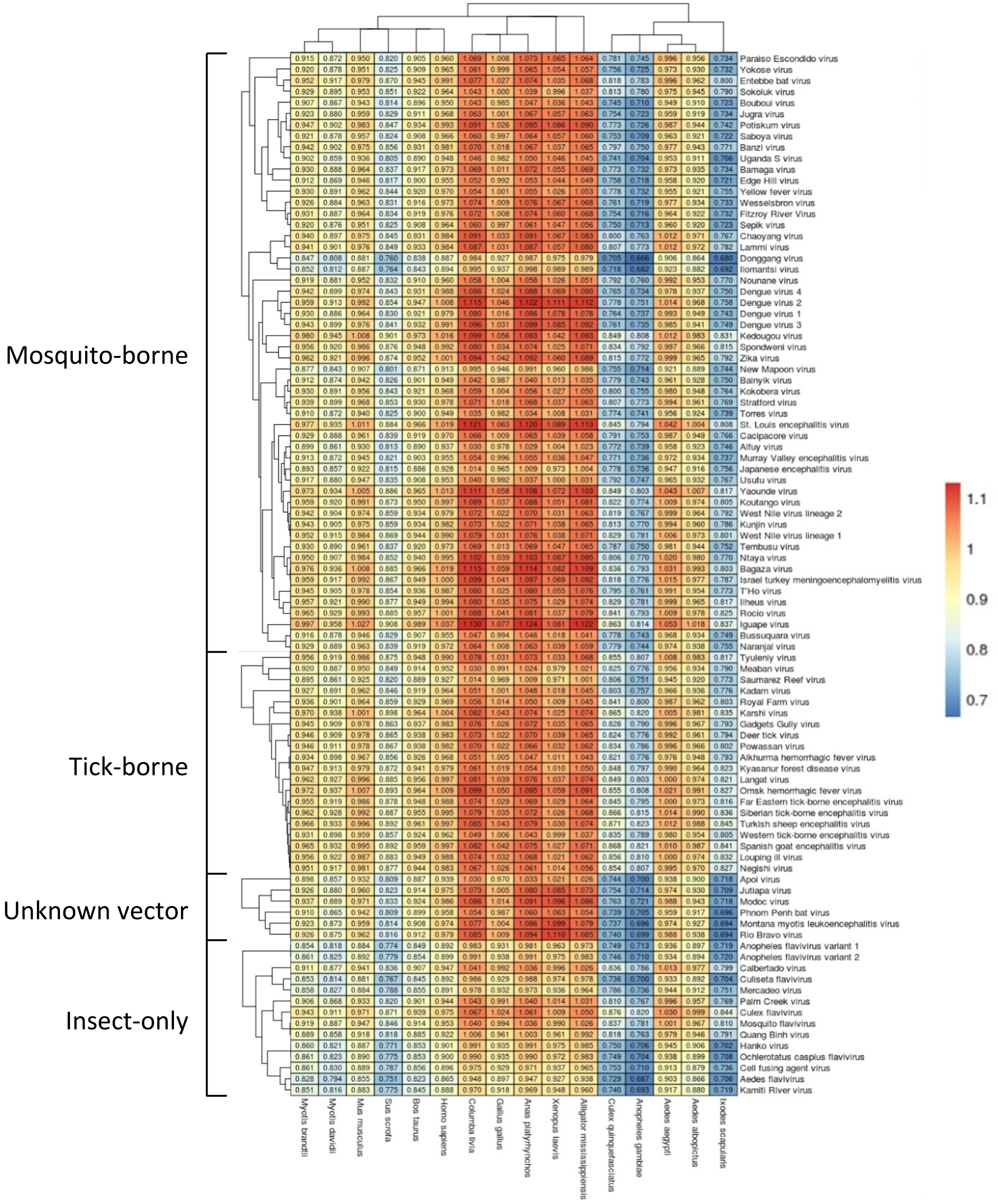
Supplementary figure S5. Intraspecies correspondence analysis of the Relative Synonymous Codon Usage (RSCU) values of major mosquito-borne flaviviruses. With (a) Dengue viruses (N = 4865), (b) Japanese encephalitis viruses (N = 297), (c) West Nile viruses (N = 1619) and Zika viruses (N = 494), different groups of viruses are distinguishable based on their codon usages, although the amount of separation does not necessarily mirror evolutionary relationships with the exception of Zika virus. Each virus outgroup is noted in the legends. Dimension 1 explains 34.7 percent and Dimension 2 attributes to 24.7 percent of the variation in panel a. Dimension 1 explains 69.4 percent and Dimension 2 attributes to 6.5 percent of the variation in panel b. Dimension 1 explains 34.1 percent and Dimension 2 attributes to 23.9 percent of the variation in panel c. Finally, Dimension 1 explains 76.2 percent and Dimension 2 attributes to 6.4 percent of the variation in panel d.



Supplementary figure S6. Intraspecies correspondence analysis of the guanine and cytosine content based on the third nucleotide of each codon (%G3+%C3) of major mosquito-borne flaviviruses. Based on nucleotide compositions, the genomic sequences of (a) Dengue viruses (N = 4865), (b) Japanese encephalitis viruses (N = 297), (c) West Nile viruses (N = 1619) and Zika viruses (N = 494) form distinct cluster that match established categories of each virus. Additionally, the distances between clusters is similar to their phylogenetic relatedness. Each outgroup virus of a respective flavivirus is mentioned in the legend. In panel a, 78.9 percent of the variation is explained by Dimension 1 and 16.8 percent by Dimension 2. In panel b, 86.8 percent of the variation is explained by Dimension 1 and 10.4 percent by Dimension 2. In panel c, 61.2 percent of the variation is explained by Dimension 1 and 31.4 percent by Dimension 2. In panel d, 88.9 percent of the variation is explained by Dimension 1 and 8.7 percent by Dimension 2.

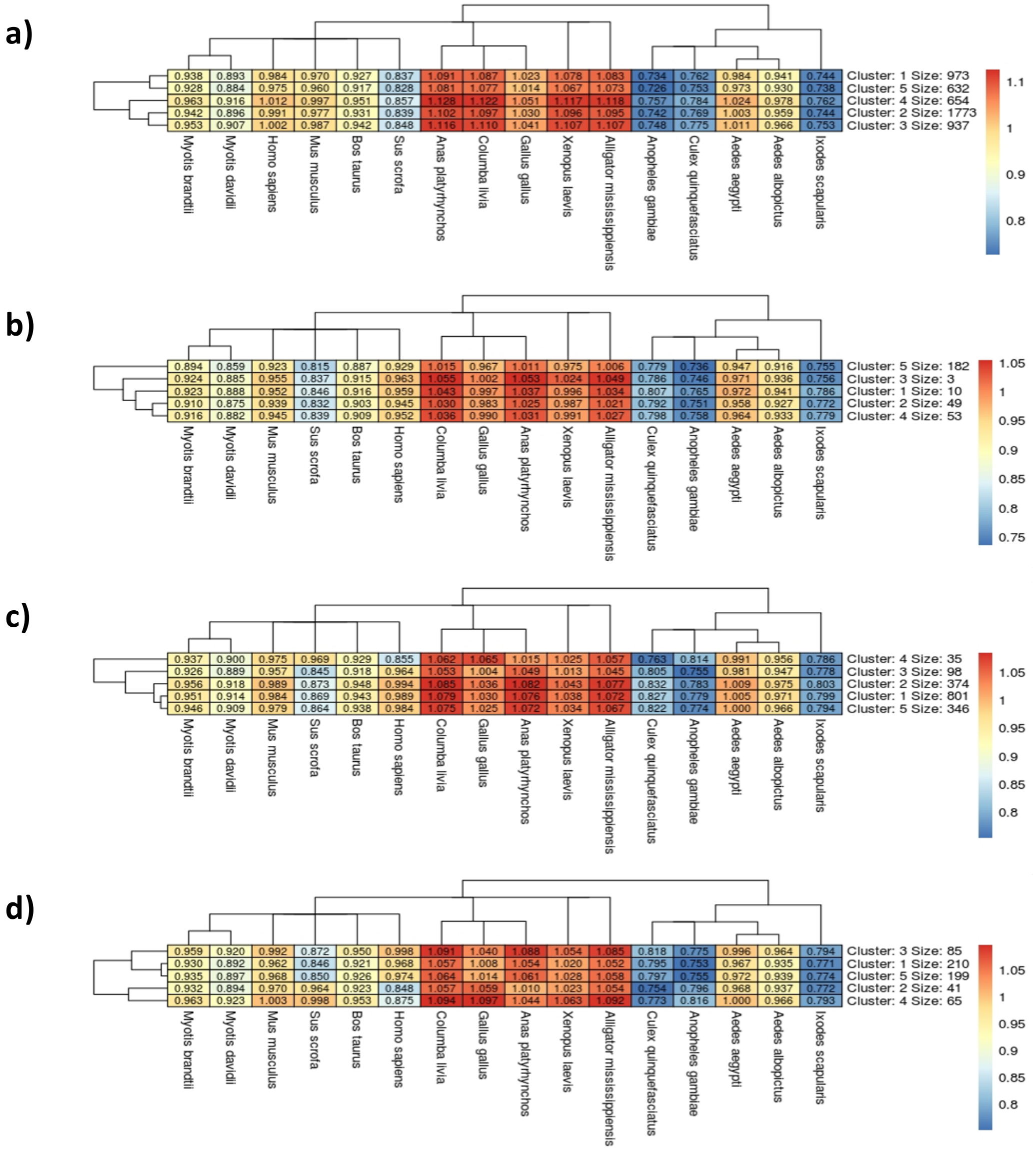


Supplementary figure S7. Interspecies correspondence analysis of the genomic compositions of major mosquito-borne flaviviruses (N = 7275). The viruses form species specific clusters based on their (a) Relative Synonymous Codon Usage (RSCU) values and (b) guanine and cytosine contents based on the third nucleotide of each codon (%G3+%C3). While both methods are capable of distinguishing subgroups within a species, RSCU has a greater ability to separate these groups compared to %G3+%C3. The distances between clusters do not perfectly equate to phylogenetic relationships, e.g. in panel b, Zika viruses are not more related to West Nile viruses than to Dengue viruses. With both analyses, Cell fusing agent virus was used as an outgroup. In panel a, Dimension 1 and 2 explain 68.3 and 8.1 percent of the variation respectively. In panel b, Dimension 1 contributes to 97.1 percent of the variance, while Dimension 2 contributes to 2.4 percent.



	Myotis brandtii	Myotis davidii	Mus musculus	Sus scrofa	Bos taurus	Homo sapiens	Columba livia	Gallus gallus	Anas platyrhynchos	Xenopus laevis	Alligator mississippiensis	Culex quinquefasciatus	Anopheles gambiae	Aedes aegypti	Aedes albopictus	Ixodes scapularis
Paraiso Escondido virus	0.915	0.872	0.950	0.820	0.905	0.960	1.069	1.008	1.073	1.065	1.064	0.781	0.745	0.996	0.956	0.734
Yokose virus	0.920	0.878	0.951	0.825	0.909	0.965	1.061	0.999	1.065	1.054	1.057	0.756	0.725	0.973	0.930	0.732
Entebbe bat virus	0.952	0.917	0.979	0.870	0.945	0.991	1.077	1.027	1.074	1.035	1.068	0.818	0.783	0.996	0.962	0.800
Sokoluk virus	0.929	0.895	0.953	0.851	0.922	0.964	1.043	1.000	1.039	0.996	1.037	0.813	0.780	0.975	0.945	0.790
Bouboui virus	0.907	0.867	0.943	0.814	0.896	0.950	1.043	0.985	1.047	1.036	1.043	0.745	0.710	0.949	0.910	0.723
Jugra virus	0.923	0.880	0.959	0.829	0.911	0.968	1.063	1.001	1.067	1.057	1.063	0.754	0.723	0.959	0.919	0.734
Potiskum virus	0.947	0.902	0.983	0.847	0.934	0.993	1.091	1.026	1.095	1.086	1.090	0.773	0.726	0.987	0.944	0.742
Saboya virus	0.921	0.878	0.957	0.824	0.908	0.966	1.060	0.997	1.064	1.057	1.060	0.753	0.709	0.963	0.921	0.722
Banzi virus	0.942	0.902	0.975	0.856	0.931	0.981	1.070	1.018	1.067	1.037	1.065	0.797	0.750	0.977	0.943	0.771
Uganda S virus	0.902	0.859	0.936	0.805	0.890	0.948	1.046	0.982	1.050	1.046	1.045	0.741	0.704	0.953	0.911	0.706
Bamaga virus	0.930	0.888	0.964	0.837	0.917	0.973	1.069	1.011	1.072	1.055	1.069	0.773	0.732	0.973	0.935	0.734
Edge Hill virus	0.912	0.869	0.946	0.817	0.900	0.955	1.052	0.992	1.053	1.044	1.049	0.758	0.718	0.958	0.920	0.721
Yellow fever virus	0.930	0.891	0.962	0.844	0.920	0.970	1.054	1.001	1.055	1.026	1.053	0.778	0.732	0.955	0.921	0.755
Wesselsbron virus	0.926	0.884	0.963	0.831	0.916	0.973	1.074	1.009	1.076	1.067	1.068	0.761	0.719	0.977	0.934	0.733
Fitzroy River Virus	0.931	0.887	0.964	0.834	0.919	0.976	1.072	1.008	1.074	1.060	1.068	0.754	0.716	0.964	0.922	0.732
Sepik virus	0.920	0.876	0.951	0.825	0.908	0.964	1.060	0.997	1.061	1.047	1.056	0.750	0.713	0.960	0.920	0.723
Chaoyang virus	0.940	0.897	0.975	0.845	0.931	0.984	1.091	1.033	1.091	1.067	1.083	0.800	0.763	1.012	0.971	0.767
Lammi virus	0.941	0.901	0.976	0.849	0.933	0.984	1.087	1.031	1.087	1.057	1.080	0.807	0.773	1.012	0.972	0.782
Donggang virus	0.847	0.808	0.881	0.760	0.838	0.887	0.984	0.927	0.987	0.975	0.979	0.705	0.666	0.906	0.864	0.680
Ilomantsi virus	0.852	0.812	0.887	0.764	0.843	0.894	0.995	0.937	0.998	0.989	0.989	0.718	0.682	0.923	0.882	0.692
Nounane virus	0.919	0.881	0.952	0.832	0.910	0.960	1.058	1.004	1.058	1.026	1.051	0.792	0.760	0.992	0.953	0.770
Dengue virus 4	0.942	0.899	0.974	0.843	0.931	0.988	1.086	1.024	1.088	1.069	1.080	0.765	0.734	0.978	0.937	0.750
Dengue virus 2	0.959	0.913	0.992	0.854	0.947	1.008	1.115	1.046	1.122	1.111	1.112	0.778	0.751	1.014	0.968	0.758
Dengue virus 1	0.930	0.886	0.964	0.830	0.921	0.979	1.080	1.016	1.086	1.078	1.078	0.764	0.737	0.993	0.949	0.743
Dengue virus 3	0.943	0.899	0.976	0.841	0.932	0.991	1.096	1.031	1.099	1.085	1.092	0.761	0.735	0.985	0.941	0.749
Kedougou virus	0.980	0.945	1.008	0.901	0.973	1.016	1.099	1.056	1.093	1.042	1.093	0.849	0.808	1.012	0.983	0.831
Spondweni virus	0.956	0.920	0.986	0.876	0.948	0.992	1.080	1.034	1.074	1.025	1.071	0.834	0.792	0.997	0.966	0.815
Zika virus	0.962	0.921	0.996	0.874	0.952	1.001	1.094	1.042	1.092	1.060	1.089	0.815	0.772	0.999	0.965	0.792
New Mapoon virus	0.877	0.843	0.907	0.801	0.871	0.913	0.995	0.946	0.991	0.960	0.986	0.755	0.714	0.921	0.889	0.744
Bainyik virus	0.912	0.874	0.942	0.826	0.901	0.949	1.042	0.987	1.040	1.013	1.035	0.779	0.743	0.961	0.928	0.750
Kokobera virus	0.930	0.891	0.956	0.843	0.921	0.968	1.059	1.004	1.056	1.027	1.050	0.800	0.755	0.980	0.948	0.764
Stratford virus	0.939	0.899	0.968	0.853	0.930	0.978	1.071	1.018	1.068	1.037	1.063	0.807	0.773	0.994	0.961	0.769
Torres virus	0.910	0.872	0.940	0.825	0.900	0.949	1.035	0.982	1.034	1.008	1.031	0.774	0.741	0.956	0.924	0.739
St. Louis encephalitis virus	0.977	0.935	1.011	0.884	0.966	1.019	1.121	1.063	1.120	1.089	1.113	0.845	0.794	1.042	1.004	0.808
Cacipacore virus	0.929	0.888	0.961	0.839	0.919	0.970	1.066	1.009	1.065	1.039	1.058	0.791	0.753	0.987	0.949	0.766
Alfuy virus	0.899	0.861	0.930	0.813	0.890	0.937	1.030	0.978	1.029	1.004	1.023	0.772	0.739	0.958	0.923	0.746
Murray Valley encephalitis virus	0.913	0.872	0.945	0.821	0.903	0.955	1.054	0.996	1.055	1.036	1.047	0.771	0.736	0.972	0.934	0.737
Japanese encephalitis virus	0.893	0.857	0.922	0.815	0.886	0.928	1.014	0.965	1.009	0.973	1.004	0.778	0.736	0.947	0.916	0.756
Usutu virus	0.917	0.880	0.947	0.835	0.908	0.953	1.040	0.992	1.037	1.000	1.031	0.792	0.747	0.965	0.932	0.767
Yaounde virus	0.973	0.934	1.005	0.886	0.965	1.013	1.111	1.058	1.108	1.072	1.103	0.849	0.803	1.043	1.007	0.817
Koutango virus	0.959	0.920	0.991	0.873	0.950	0.997	1.089	1.037	1.088	1.051	1.081	0.822	0.774	1.009	0.974	0.805
West Nile virus lineage 2	0.942	0.904	0.974	0.859	0.934	0.979	1.072	1.022	1.070	1.031	1.063	0.819	0.767	0.999	0.964	0.792
Kunjin virus	0.943	0.905	0.975	0.859	0.934	0.982	1.073	1.022	1.071	1.038	1.065	0.813	0.770	0.994	0.960	0.786
West Nile virus lineage 1	0.952	0.915	0.984	0.869	0.944	0.990	1.079	1.031	1.076	1.038	1.071	0.829	0.781	1.006	0.973	0.801
Tembusu virus	0.930	0.890	0.961	0.837	0.920	0.973	1.069	1.013	1.069	1.047	1.065	0.787	0.750	0.981	0.944	0.752
Ntaya virus	0.950	0.907	0.984	0.852	0.940	0.995	1.102	1.039	1.103	1.087	1.095	0.806	0.770	1.020	0.980	0.770
Bagaza virus	0.976	0.936	1.008	0.885	0.966	1.019	1.115	1.059	1.114	1.082	1.109	0.836	0.793	1.031	0.993	0.803
Israel turkey meningoencephalomyelitis virus	0.959	0.917	0.992	0.867	0.949	1.000	1.099	1.041	1.097	1.069	1.092	0.818	0.776	1.015	0.977	0.787
T'Ho virus	0.945	0.905	0.978	0.854	0.936	0.987	1.080	1.025	1.080	1.055	1.076	0.795	0.761	0.991	0.954	0.773
Ilheus virus	0.957	0.921	0.990	0.877	0.949	0.994	1.080	1.035	1.075	1.029	1.074	0.829	0.781	0.999	0.965	0.817
Rocio virus	0.965	0.929	0.993	0.885	0.957	1.001	1.088	1.041	1.081	1.037	1.079	0.841	0.793	1.009	0.978	0.825
Iguape virus	0.997	0.958	1.027	0.908	0.989	1.037	1.130	1.077	1.124	1.081	1.122	0.863	0.814	1.053	1.018	0.837
Bussuquara virus	0.916	0.878	0.946	0.829	0.907	0.955	1.047	0.994	1.046	1.018	1.041	0.778	0.743	0.968	0.934	0.749
Naranjal virus	0.929	0.889	0.963	0.839	0.919	0.972	1.064	1.008	1.063	1.039	1.059	0.779	0.744	0.974	0.938	0.755
Tyuleniy virus	0.956	0.919	0.986	0.875	0.948	0.990	1.078	1.031	1.073	1.033	1.068	0.855	0.807	1.008	0.983	0.817
Meaban virus	0.920	0.887	0.950	0.849	0.914	0.952	1.030	0.991	1.024	0.979	1.021	0.825	0.776	0.956	0.934	0.790
Saumarez Reef virus	0.895	0.861	0.925	0.820	0.889	0.927	1.014	0.969	1.009	0.971	1.001	0.806	0.751	0.945	0.920	0.773
Kadam virus	0.927	0.891	0.962	0.846	0.919	0.964	1.051	1.001	1.048	1.018	1.045	0.803	0.757	0.966	0.936	0.776
Royal Farm virus	0.936	0.901	0.964	0.859	0.929	0.969	1.056	1.014	1.050	1.009	1.045	0.841	0.800	0.987	0.962	0.803
Karshi virus	0.970	0.938	1.001	0.898	0.964	1.004	1.082	1.043	1.074	1.025	1.074	0.865	0.820	1.005	0.981	0.835
Gadgets Gully virus	0.945	0.909	0.978	0.863	0.937	0.983	1.076	1.026	1.072	1.035	1.065	0.828	0.790	0.996	0.967	0.793
Deer tick virus	0.946	0.909	0.978	0.865	0.938	0.983	1.073	1.022	1.070	1.039	1.065	0.824	0.776	0.992	0.961	0.794
Powassan virus	0.946	0.911	0.978	0.867	0.938	0.982	1.070	1.022	1.066	1.032	1.062	0.834	0.786	0.996	0.966	0.802
Alkhurma hemorrhagic fever virus	0.934	0.898	0.967	0.856	0.926	0.968	1.051	1.005	1.047	1.011	1.043	0.821	0.776	0.976	0.948	0.793
Kyasanur forest disease virus	0.947	0.913	0.979	0.872	0.941	0.979	1.061	1.019	1.054	1.010	1.050	0.848	0.797	0.990	0.964	0.823
Langat virus	0.962	0.927	0.996	0.885	0.956	0.997	1.081	1.039	1.076	1.037	1.074	0.849	0.803	1.000	0.974	0.821
Omsk hemorrhagic fever virus	0.972	0.937	1.007	0.893	0.964	1.009	1.099	1.050	1.095	1.059	1.091	0.855	0.808	1.021	0.991	0.827
Far Eastern tick-borne encephalitis virus	0.955	0.919	0.986	0.878	0.948	0.988	1.074	1.029	1.069	1.029	1.064	0.845	0.795	1.000	0.973	0.816
Siberian tick-borne encephalitis virus	0.962	0.928	0.992	0.887	0.955	0.995	1.079	1.035	1.072	1.026	1.068	0.866	0.815	1.014	0.990	0.836
Turkish sheep encephalitis virus	0.966	0.933	0.996	0.892	0.961	0.997	1.085	1.043	1.079	1.030	1.074	0.871	0.823	1.012	0.988	0.845
Western tick-borne encephalitis virus	0.931	0.898	0.959	0.857	0.924	0.962	1.049	1.006	1.043	0.999	1.037	0.835	0.789	0.980	0.954	0.805
Spanish goat encephalitis virus	0.965	0.932	0.995	0.892	0.959	0.997	1.082	1.042	1.075	1.027	1.071	0.868	0.821	1.010	0.987	0.841
Louping ill virus	0.956	0.922	0.987	0.883	0.949	0.988	1.074	1.032	1.068	1.021	1.062	0.856	0.810	1.000	0.974	0.832
Negishi virus	0.951	0.917	0.981	0.877	0.944	0.983	1.067	1.026	1.061	1.014	1.056	0.854	0.807	0.995	0.970	0.827
Apoi virus	0.898	0.857	0.932	0.809	0.887	0.939	1.030	0.970	1.033	1.021	1.026	0.744	0.700	0.938	0.900	0.718
Jutiapa virus	0.926	0.880	0.960	0.823	0.914	0.975	1.073	1.005	1.080	1.085	1.073	0.754	0.714	0.974	0.930	0.709
Modoc virus	0.937	0.889	0.971	0.833	0.924	0.986	1.086	1.014	1.091	1.096	1.086	0.763	0.721	0.988	0.943	0.718
Phnom Penh bat virus	0.910	0.865	0.942	0.809	0.899	0.958	1.054	0.987	1.060	1.063	1.054	0.739	0.705	0.959	0.917	0.696
Montana myotis leukoencephalitis virus	0.923	0.873	0.959	0.814	0.908	0.974	1.077	1.004	1.086	1.099	1.079	0.737	0.696	0.974	0.927	0.694
Rio Bravo virus	0.926	0.875	0.962	0.816	0.912	0.979	1.085	1.009	1.094	1.110	1.085	0.740	0.699	0.988	0.938	0.694
Anopheles flavivirus variant 1	0.854	0.818	0.884	0.774	0.849	0.892	0.983	0.931	0.981	0.963	0.973	0.749	0.713	0.936	0.897	0.719
Anopheles flavivirus variant 2	0.861	0.825	0.892	0.779	0.854	0.899	0.991	0.938	0.991	0.975	0.983	0.746	0.710	0.934	0.894	0.720
Calbertado virus	0.911	0.877	0.941	0.836	0.907	0.947	1.041	0.992	1.036	0.996	1.026	0.836	0.786	1.013	0.977	0.799
Culiseta flavivirus	0.853	0.814	0.881	0.767	0.845	0.892	0.986	0.929	0.988	0.974	0.978	0.736	0.700	0.933	0.892	0.704
Mercadeo virus	0.858	0.827	0.884	0.788	0.855	0.891	0.978	0.932	0.973	0.936	0.964	0.786	0.736	0.944	0.912	0.751
Palm Creek virus	0.906	0.868	0.933	0.820	0.901	0.944	1.043	0.991	1.040	1.014	1.031	0.810	0.767	0.996	0.957	0.769
Culex flavivirus	0.943	0.911	0.971	0.871	0.939	0.975	1.067	1.024	1.061	1.009	1.050	0.876	0.820	1.030	0.999	0.844
Mosquito flavivirus	0.919	0.887	0.947	0.846	0.914	0.953	1.040	0.994	1.036	0.990	1.026	0.837	0.781	1.001	0.967	0.810
Quang Binh virus	0.889	0.858	0.918	0.818	0.885	0.922	1.006	0.961	1.003	0.961	0.992	0.818	0.763	0.979	0.946	0.791
Hanko virus	0.860	0.821	0.887	0.771	0.853	0.901	0.991	0.935	0.991	0.975	0.985	0.750	0.706	0.945	0.906	0.702
Ochlerotatus caspius flavivirus	0.861	0.823	0.890	0.775	0.853	0.900	0.990	0.935	0.990	0.972	0.983	0.749	0.704	0.938	0.899	0.708
Cell fusing agent virus	0.861	0.830	0.889	0.787	0.856	0.896	0.975	0.929	0.971	0.937	0.965	0.753	0.710	0.913	0.879	0.736
Aedes flavivirus	0.828	0.794	0.855	0.751	0.823	0.865	0.948	0.897	0.947	0.927	0.938	0.729	0.687	0.903	0.866	0.706
Kamiti River virus	0.851	0.816	0.883	0.775	0.845	0.888	0.970	0.918	0.969	0.948	0.960	0.740	0.693	0.917	0.880	0.719

Supplementary figure S8. Normalized Codon Adaptation Index (nCAI) heatmap of flavivirus (genus *Flavivirus*) subgroups (N = 94). The nCAI values show overall over-optimization (nCAI > 1.05) for avians, reptiles and amphibians, and under-optimization for *Culex* and *Anopheles* mosquitoes (nCAI < 0.95). Optimal hosts tend to be mammals and *Aedes* mosquitoes (nCAI 0.95–1.05). The columns are sorted according to taxonomic classification of hosts, and the rows are in accordance with the phylogeny of flaviviruses.



Supplementary figure S9. Intraspecies k-means (5) heatmaps of normalized Codon Adaptation Index (nCAI) values of major mosquito-borne flaviviruses (genus *Flavivirus*). When the nCAI values of (a) Dengue viruses, (b) Japanese encephalitis viruses, (c) West Nile viruses and (d) Zika viruses are plotted in heatmaps, they display similar overall adaptation levels to different host organisms, although there are differences between these viruses. On average the viruses are optimized for mice *(Mus musculus)* and humans *(Homo sapiens),* and *Aedes* mosquitoes, especially *Aedes aegypti* (nCAI 0.95–1.05), thus being likely hosts. The unlikely hosts are the other mammalian hosts, avians, reptiles and amphibians due to over-optimization (nCAI > 1.05), and *Culex* and *Anopheles* mosquitoes, and ticks *(Ixodes scapularis)* due to under-optimization (nCAI < 0.95). There are however exceptions; Japanese encephalitis viruses in panel b have as more likely optimal hosts mostly birds, reptiles and amphibians. The number and size of k-means clusters do not match the current classification of these flaviviruses.