Supplementary Information

Title: Investigating antibody cross-reactivity and transmission dynamics of alphaviruses and flaviviruses using a multiplex serological assay

Author list: Victor Yman[1,2,3†], Jason Rosado[1,4†], Noé Ochida[5], Laura Garcia[1], Marie-Fabrice Gasasira[1], Gaëlle Baudemont[1], Estee Cramer[1], Myrielle Dupont-Rouzeyrol[6], Karl Huet[6], Maylis Douine[7], Alice Sanna[7], Yann Lambert[7], Gabriel Carrasco-Escobar[8], Oscar Nolasco[8,9], Dionicia Gamboa[8,9], Gamou Fall[10], Oumar Ndiaye[10], Oumar Faye[10], Cheikh Loucoubar[11], Aissatou Toure-Balde[12], Makhtar Niang[12], Ines Vigan-Womas[12,13], Simon Cauchemez[5], Michael T White[1*]

Affiliations:

[1] Infectious Disease Epidemiology & Analytics G5 Unit, Department of Global Health, Institut Pasteur, Université Paris Cité, INSERM U1347; Paris, France.

[2] Division of Infectious Diseases, Department of Medicine Solna, Karolinska Institutet; Stockholm, Sweden.

[3] Department of Infectious Diseases, Södersjukhuset; Stockholm, Sweden

[4] Université Paris Cité, IRD, Inserm, MERIT, F-75006; Paris, France

[5] Mathematical Modelling of Infectious Diseases Unit, Institut Pasteur, Université Paris Cité, INSERM U1332, CNRS UMR2000; Paris, France

[6] URE Dengue and Arboviruses, Institut Pasteur de Nouvelle-Calédonie; Noumea, New Caledonia

[7] Centre d'Investigation Clinique Guyane, Inserm 1424, Centre Hospitalier Universitaire de Guyane, UA17 Santé des Populations en Amazonie; Cayenne, Guyane

[8] Instituto de Medicina Tropical Alexander von Humboldt, Universidad Peruana Cayetano Heredia; Lima, Peru

[9] Laboratorio de Malaria: Parásitos y Vectores, Laboratorios de Investigación y Desarrollo, Facultad de Ciencias e Ingeniería, Universidad Peruana Cayetano Heredia; Lima, Peru

[10] Virology Department, Institut Pasteur de Dakar; Dakar, Sénégal

[11] Epidemiology, Clinical Research and Data sciences Department, Institut Pasteur de Dakar; Dakar, Sénégal

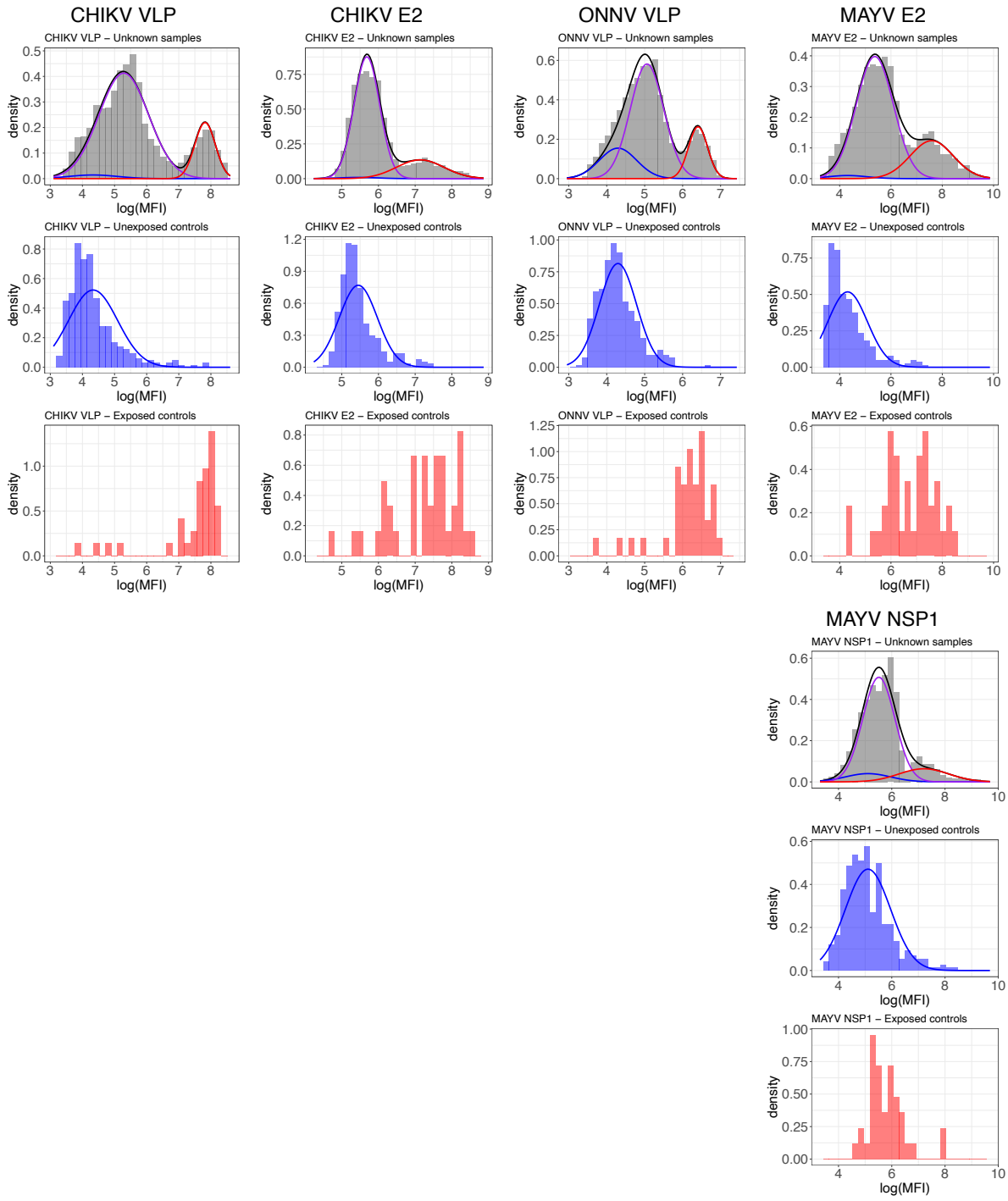
[12] Immunophysiopathology and Infectious Diseases Department, Institut Pasteur de Dakar; Dakar, Sénégal

[13] Institut Pasteur Paris, Pasteur Network; Paris, France

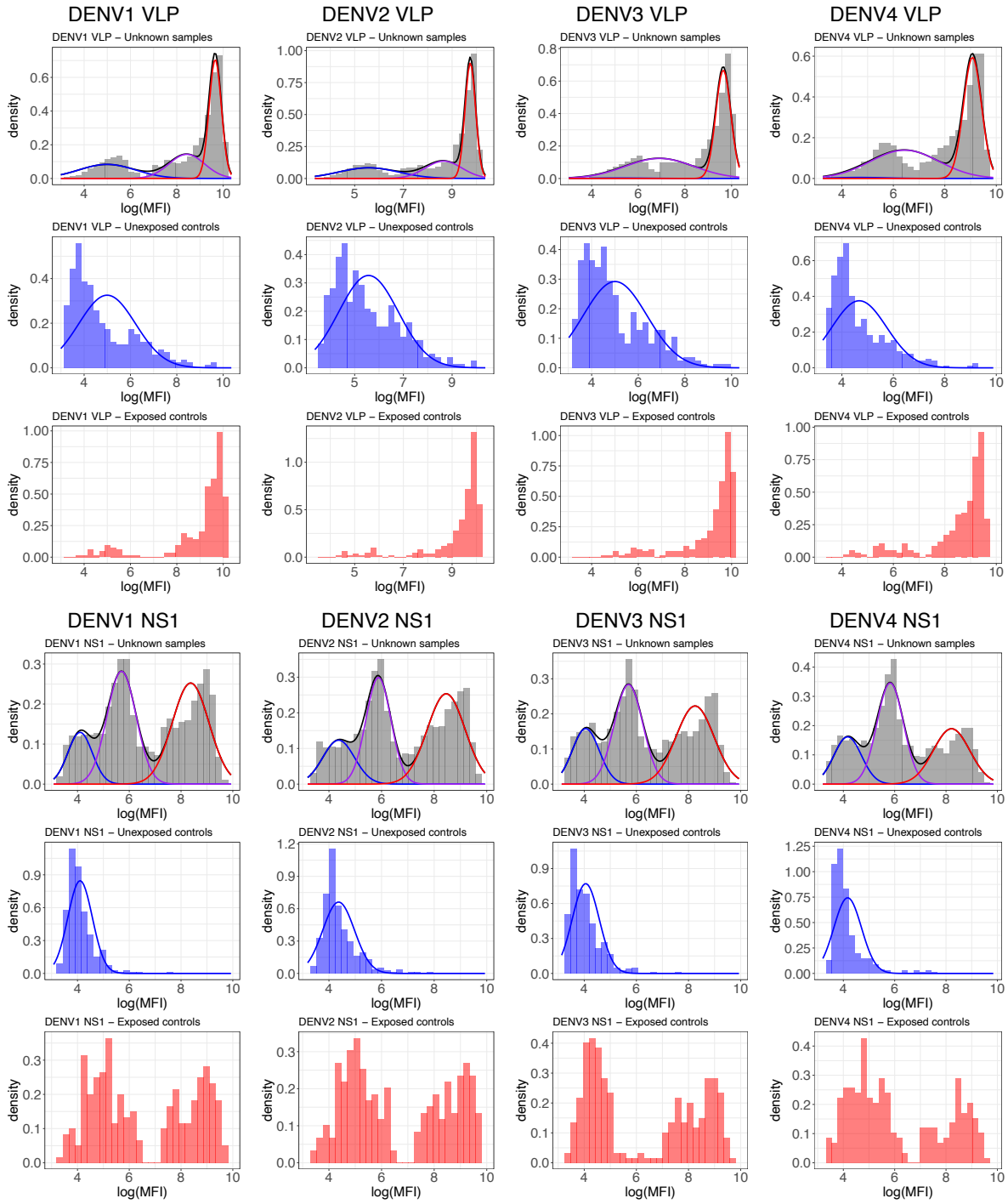
[*] Corresponding author email: michael.white@pasteur.fr

[†] These authors contributed equally to this work: Victor Yman, Jason Rosado

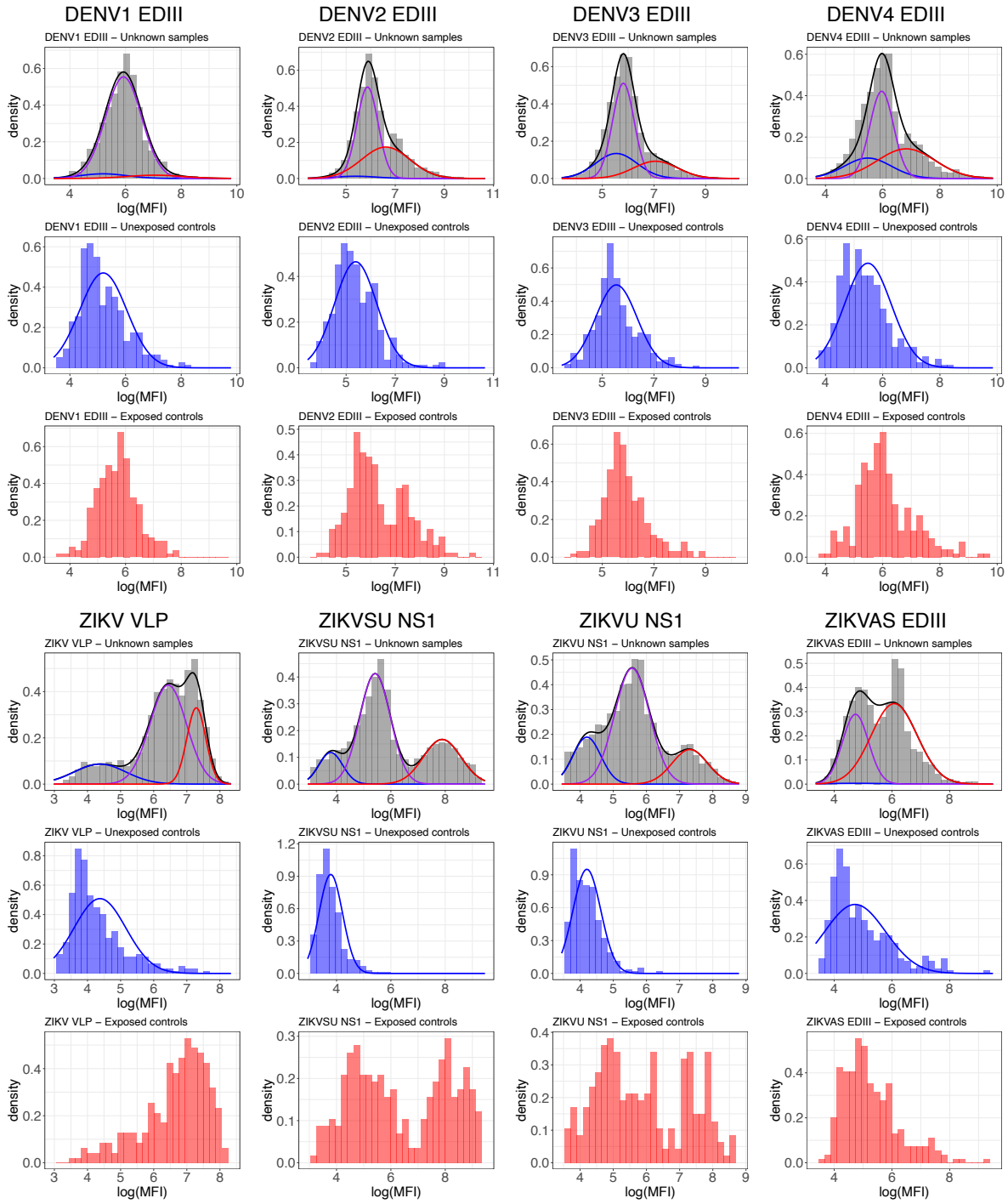
Supplementary Figures



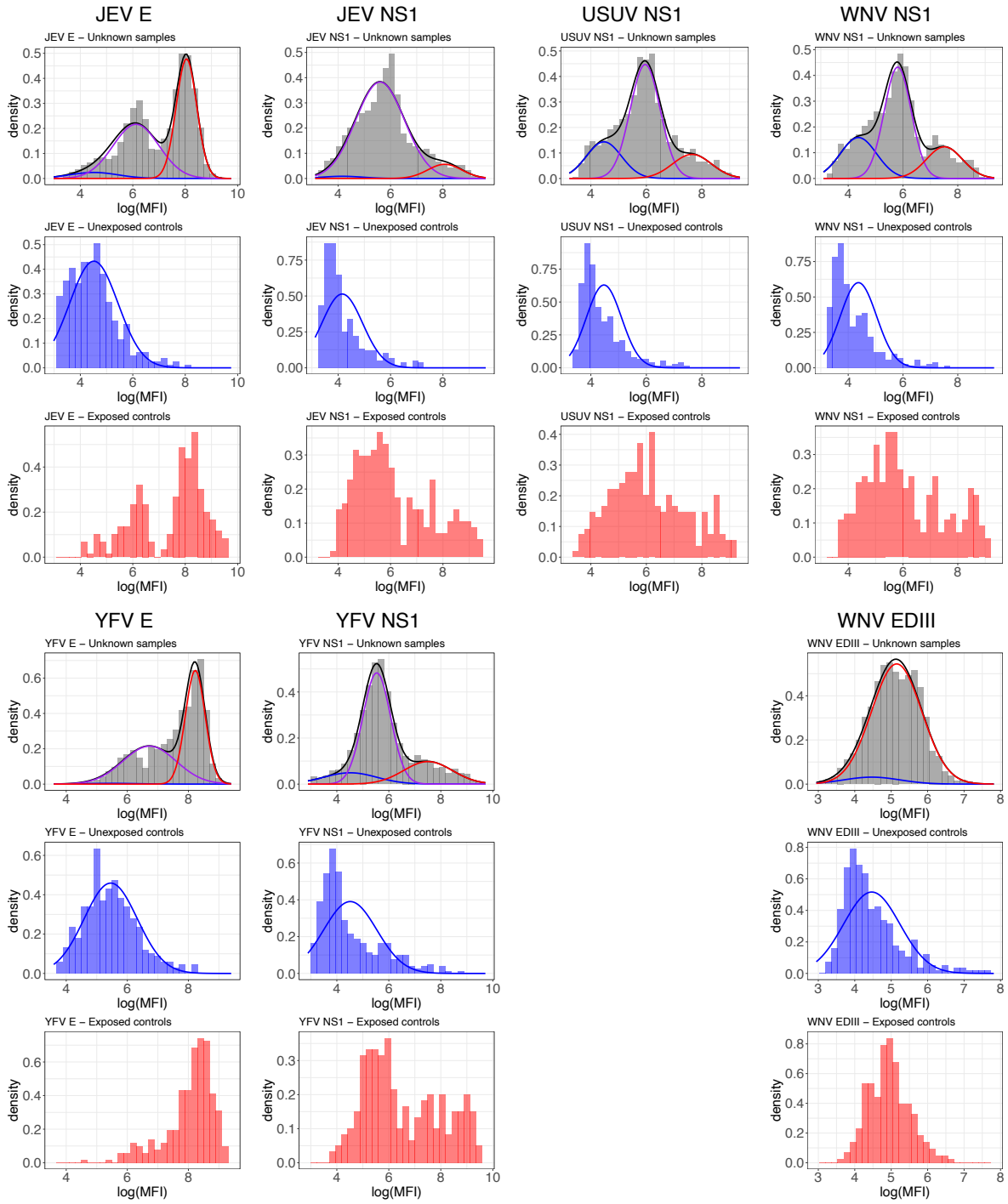
Supplementary Figure 1. Gaussian mixture model (GMM) fits. For each antibody response, a Gaussian finite mixture model was fitted jointly to data from both the cross-sectional survey samples where serostatus unknown (Light grey – Top panel) and the unexposed controls known to be seronegative (Light blue – Middle panel). The finite mixture model was not fitted to data from the exposed controls (Light red – Bottom panel). The solid lines represent the model fits of the three-component Gaussian mixtures (Blue – Lowest component, Purple – Middle component, Red – Upper component). The lowest component (blue) represents the seronegative population while the upper component (red) represents the seropositive population.



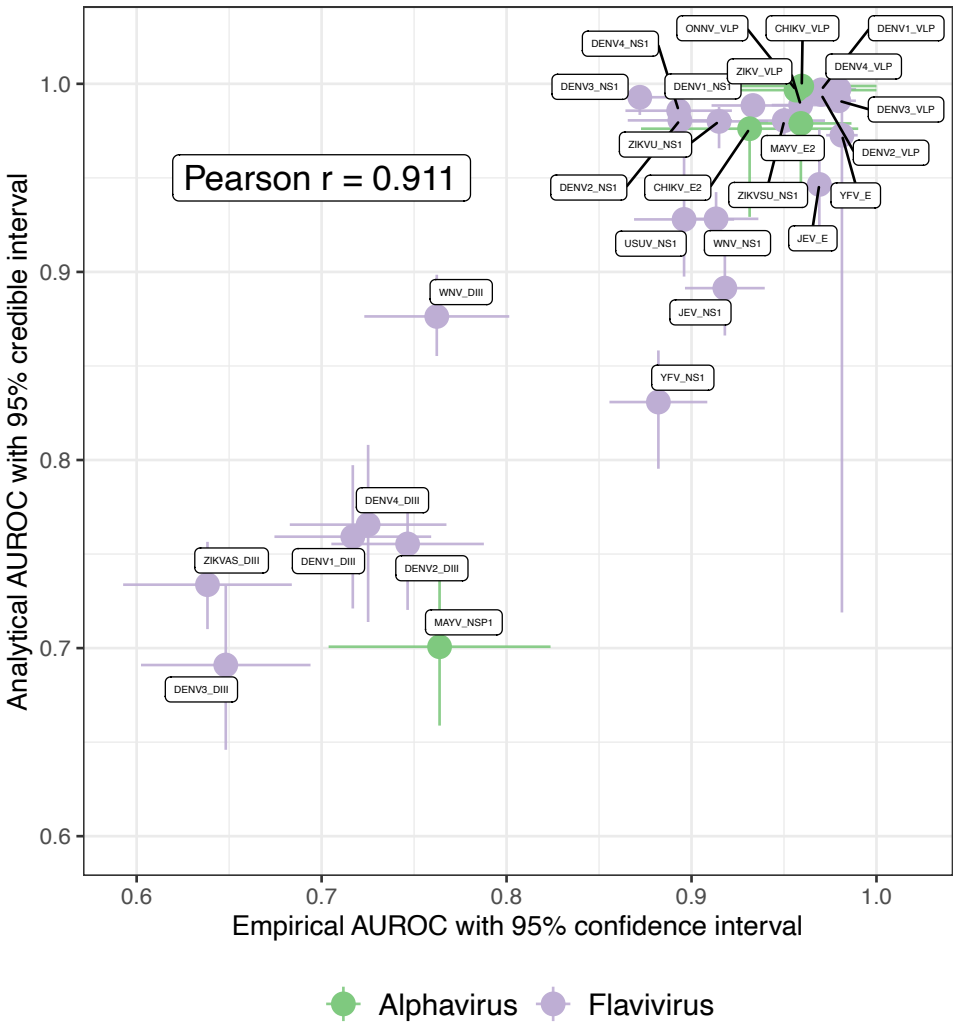
Supplementary Figure 2. Gaussian mixture model (GMM) fits. For each antibody response, a Gaussian finite mixture model was fitted jointly to data from both the cross-sectional survey samples where serostatus unknown (Light grey – Top panel) and the unexposed controls known to be seronegative (Light blue – Middle panel). The finite mixture model was not fitted to data from the exposed controls (Light red – Bottom panel). The solid lines represent the model fits of the three-component Gaussian mixtures (Blue – Lowest component, Purple – Middle component, Red – Upper component). The lowest component (blue) represents the seronegative population while the upper component (red) represents the seropositive population.



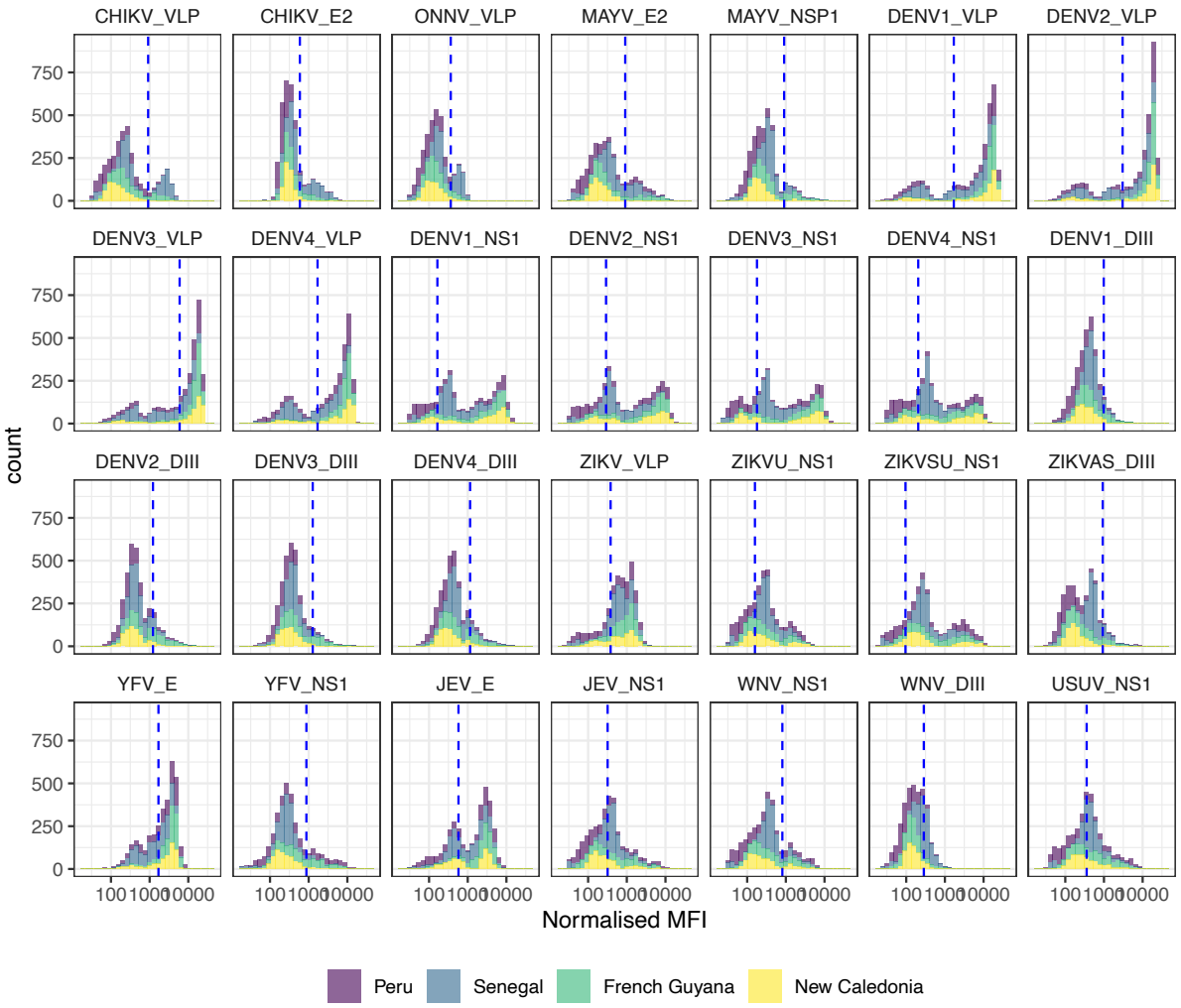
Supplementary Figure 3. Gaussian mixture model (GMM) fits. For each antibody response, a Gaussian finite mixture model was fitted jointly to data from both the cross-sectional survey samples where serostatus unknown (Light grey – Top panel) and the unexposed controls known to be seronegative (Light blue – Middle panel). The finite mixture model was not fitted to data from the exposed controls (Light red – Bottom panel). The solid lines represent the model fits of the three-component Gaussian mixtures (Blue – Lowest component, Purple – Middle component, Red – Upper component). The lowest component (blue) represents the seronegative population while the upper component (red) represents the seropositive population.



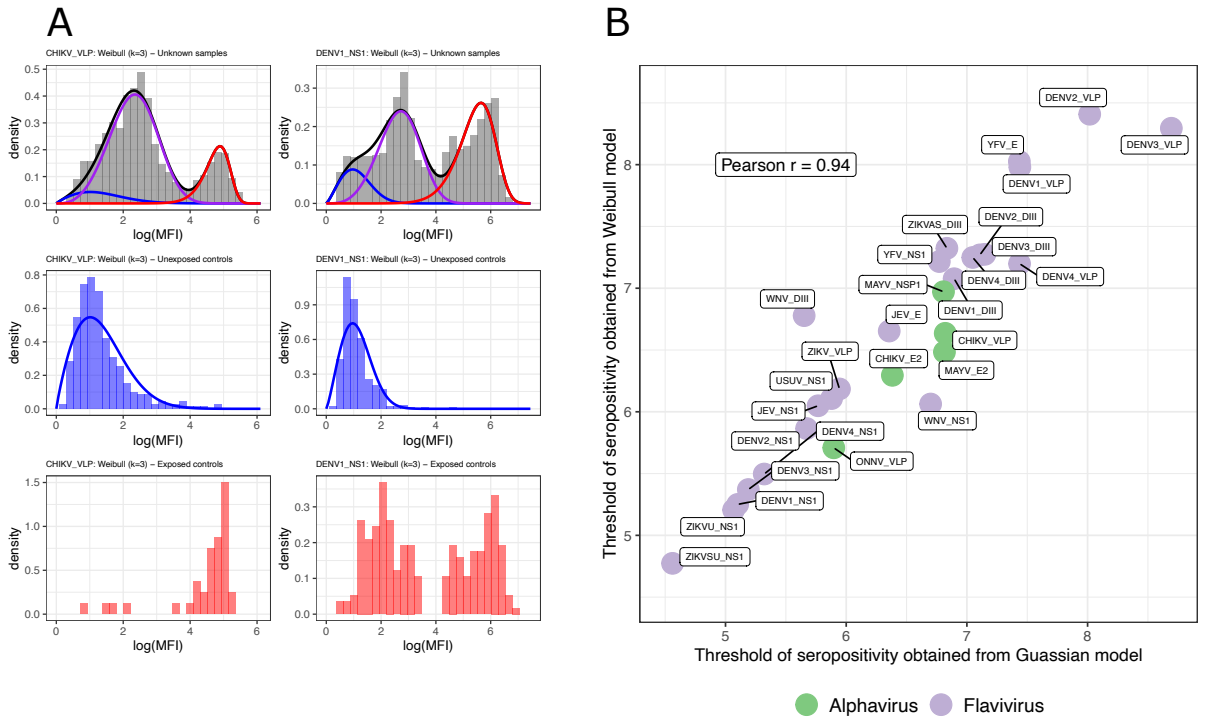
Supplementary Figure 4. Gaussian mixture model (GMM) fits. For each antibody response, a Gaussian finite mixture model was fitted jointly to data from both the cross-sectional survey samples where serostatus unknown (Light grey – Top panel) and the unexposed controls known to be seronegative (Light blue – Middle panel). The finite mixture model was not fitted to data from the exposed controls (Light red – Bottom panel). The solid lines represent the model fits of the three-component Gaussian mixtures (Blue – Lowest component, Purple – Middle component, Red – Upper component). The lowest component (blue) represents the seronegative population while the upper component (red) represents the seropositive population.



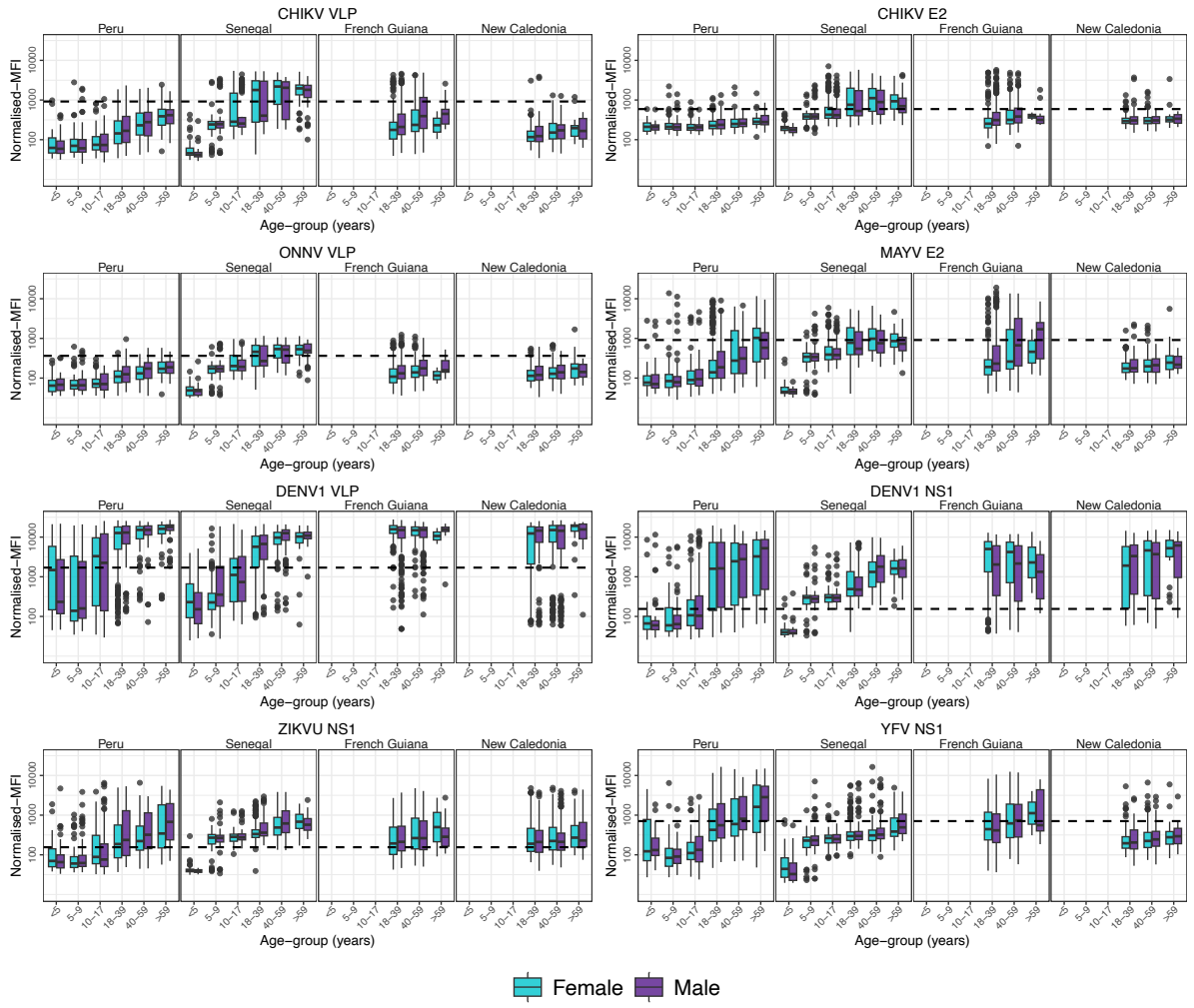
Supplementary Figure 5. Scatterplot of the empirical versus the analytical AUROC for each antibody response. X-axis error bars represent non-parametric 95% Confidence Interval of the empirical AUROC calculated using the DeLong method. Y-axis error bars represent the 95% Credible Interval of the analytical AUROC from the Bayesian Finite Mixture Model. Sample size refers to individual samples: Cross-sectional survey samples (n = 3486), Unexposed controls (n = 330), Suspected alphavirus exposed controls (n = 38), Suspected flavivirus exposed controls (n = 222). Abbreviations: *virus-like particle (VLP), envelope protein (E), envelope protein domain III (EDIII), non-structural protein 1 (NS1/NSP1).*



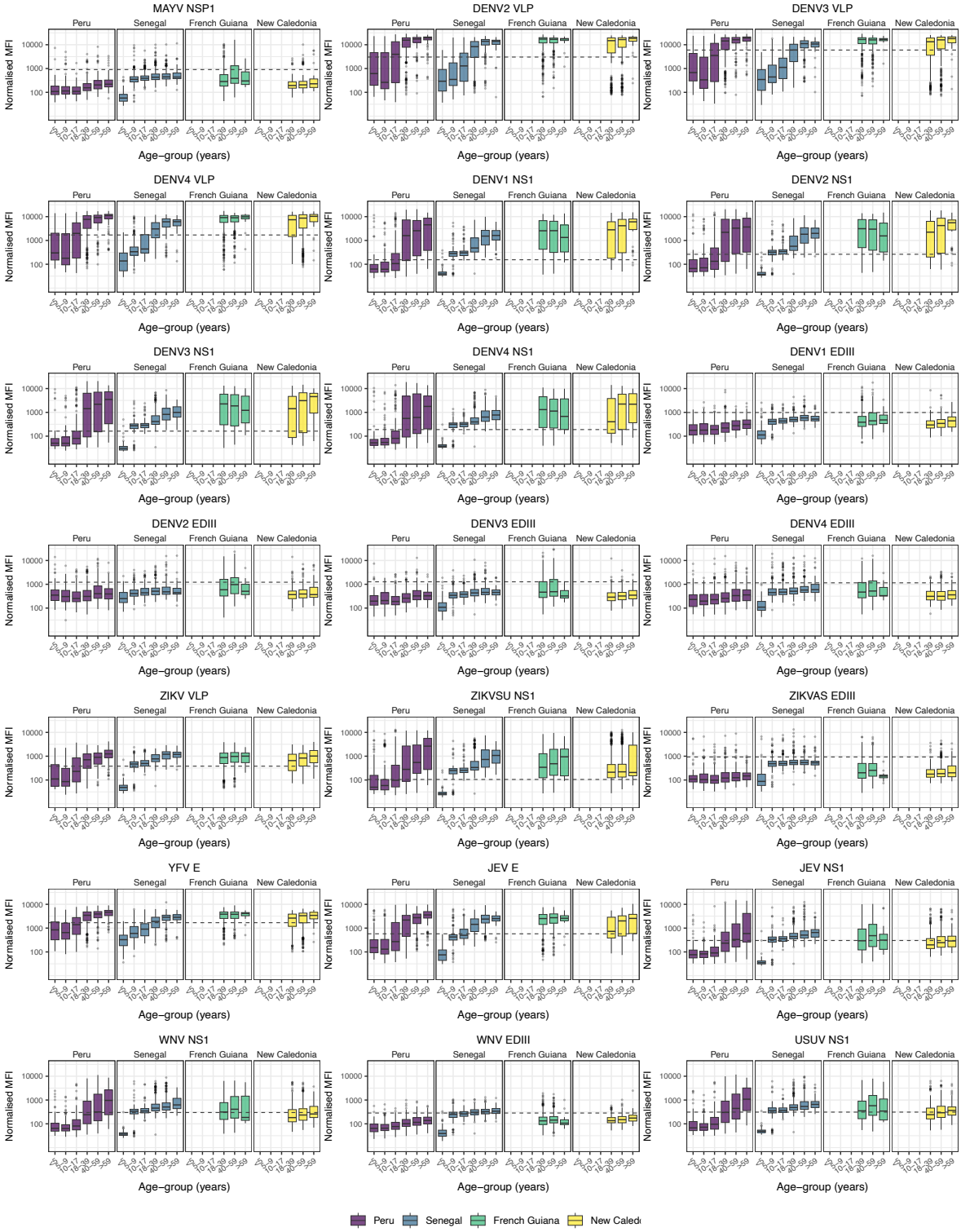
Supplementary Figure 6. Histogram of the distribution of antibody data for all arboviral antigens. Histogram bars are coloured by the study site origin of the samples. The dashed blue line represents the threshold of seropositivity which controls assay specificity at 97·7% derived from the GMM model estimates. Sample size refers to individual samples: Peru (n = 880), Senegal (n = 1077), French Guiana (n = 787), New Caledonia (n = 600). Abbreviations: *virus-like particle (VLP), envelope protein (E), envelope protein domain III (EDIII), non-structural protein 1 (NS1/NSP1).*



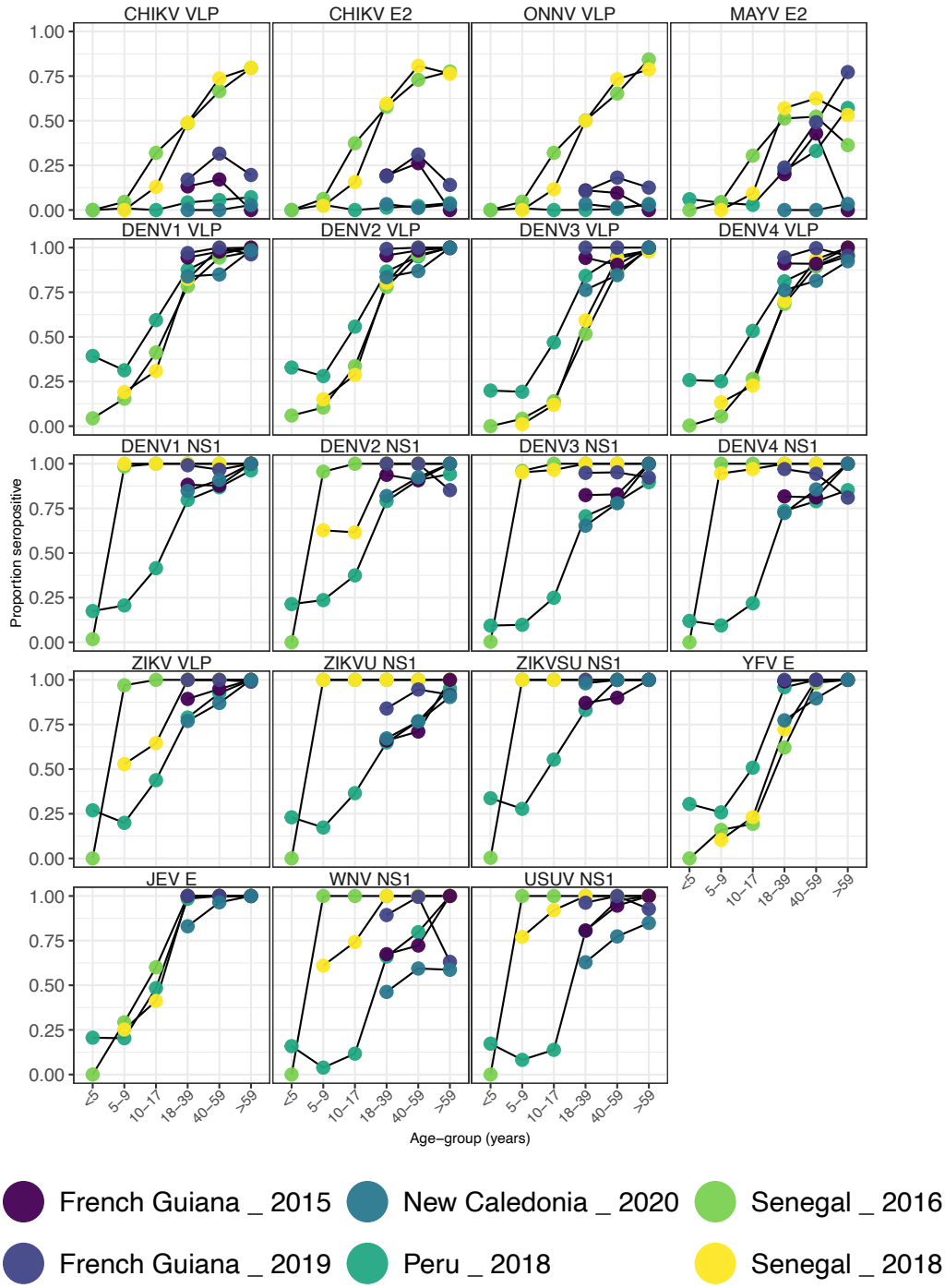
Supplementary Figure 7. Weibull mixture models (WMMs). Due to skewness, immunoassay data can sometimes not be accurately described by a mixture of gaussian distributions. Therefore, WMMs based on two- and three-component mixtures of Weibull distributions were also evaluated. The greater flexibility of the Weibull distribution to accommodate skewness and kurtosis in the data led to improved model fits (with lower LOOIC) of the three-component WMM for most antibody responses (Supplementary Table S2). As for the Gaussian WMMs thresholds of seropositivity were selected from the point of the analytical ROC curve where specificity is 97.7%. **A)** For each antibody response, a Weibull finite mixture model was fitted jointly to data from both the cross-sectional survey samples, where serostatus unknown (Light grey – Top panel), and the unexposed controls, known to be seronegative (Light blue – Middle panel). The finite mixture model was not fitted to data from the suspected exposed controls (Light red – Bottom panel). The solid lines represent the model fit of the three-component Weibull mixtures **B)** Scatterplot comparing the estimated seropositivity thresholds of 97·7% specificity obtained from the Gaussian mixture model with those obtained from the Weibull mixture model. Points colours denote the individual target antigens included in the multiplex assay. Sample size refers to individual samples: Cross-sectional survey samples (n = 3486), Unexposed controls (n = 330), Suspected alphavirus exposed controls (n = 38), Suspected flavivirus exposed controls (n = 222). *Abbreviations: virus-like particl1e (VLP), envelope protein (E), envelope protein domain III (EDIII), non-structural protein 1 (NS1/NSP1).*



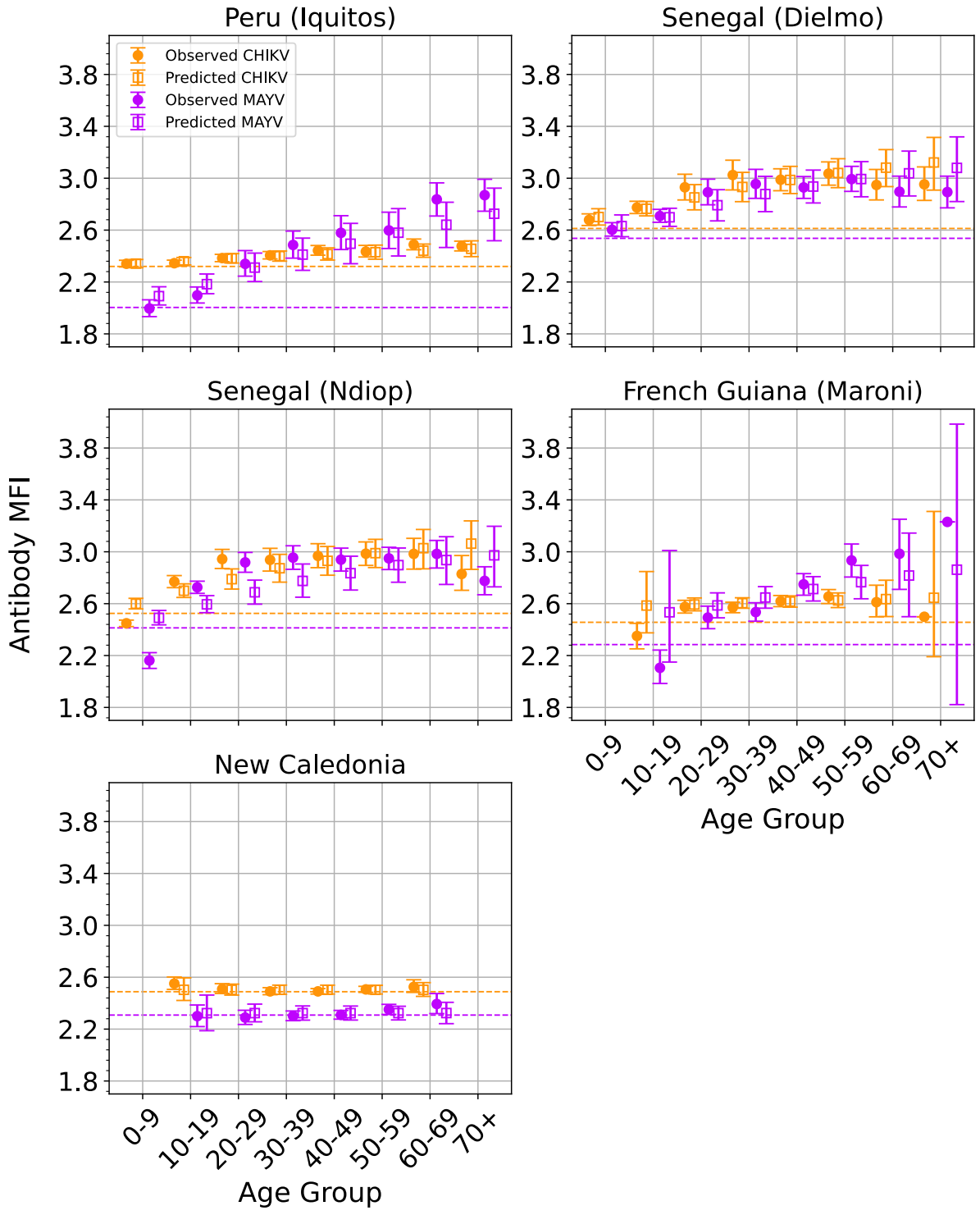
Supplementary Figure 8. Box-plots of the age-specific antibody reactivity stratified by sex at each study site for selected antibody responses. Box colours denote the sex. Boxes represent the median and interquartile range of the data. Whiskers extend to 1.5 times the height of the boxes. Outliers beyond this range are indicated by individual points. The dashed black line indicates a Gaussian finite mixture model-derived seropositivity threshold of 97.7% specificity. Sample size refers to individual samples: Peru (n = 880), Senegal (n = 1077), French Guiana (n = 787), New Caledonia (n = 600). *Abbreviations: virus-like particle (VLP), envelope protein (E), non-structural protein 1 (NS1).*



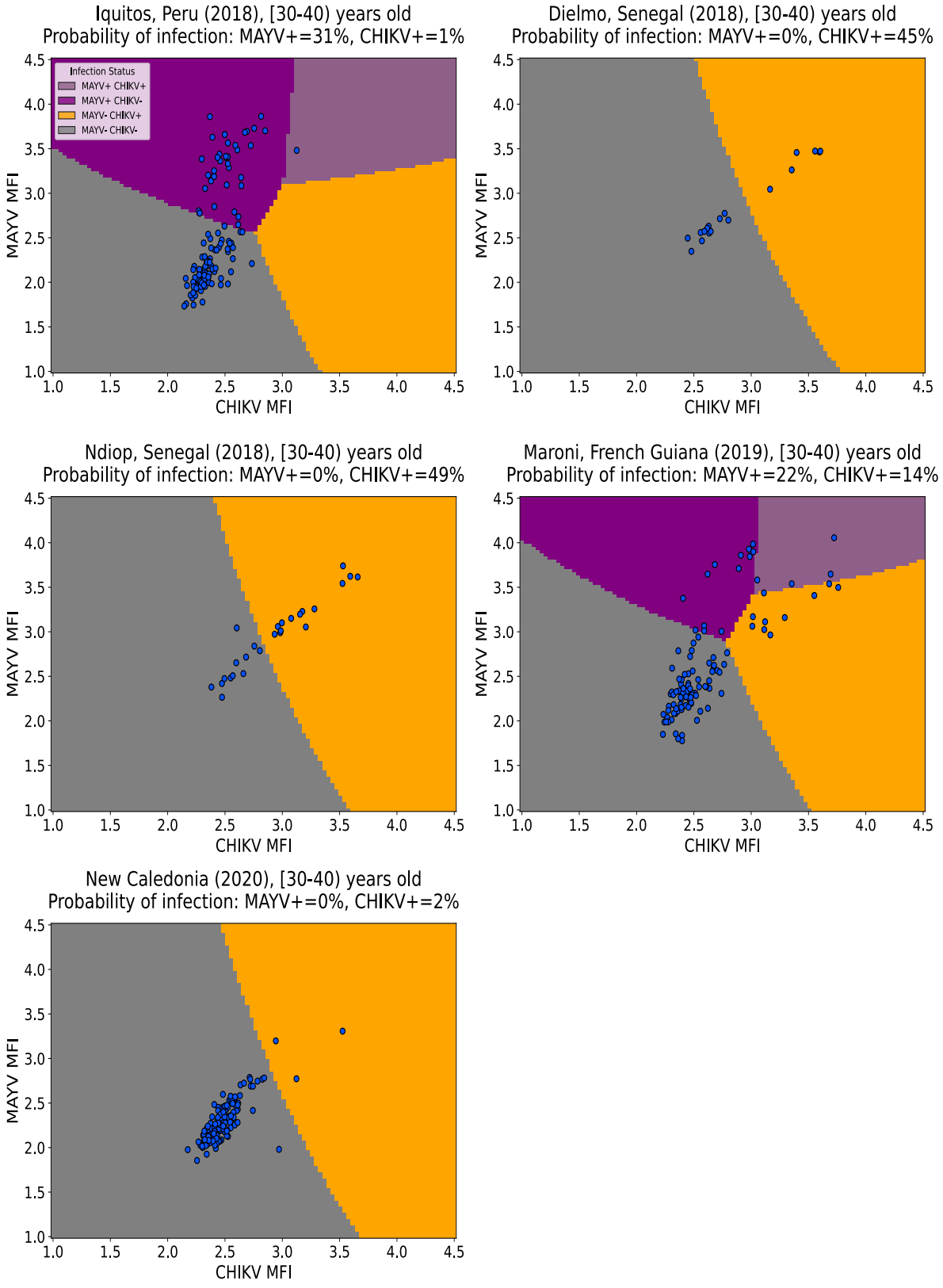
Supplementary Figure 9. Box-plots of the age-specific antibody reactivity. Data are stratified by age group and study site. Box colours denote the study site. Boxes represent the median and inter quartile range of the data. Whiskers extend to 1.5 times the height of the boxes. Outliers beyond this range are indicated by individual points. The dashed blue line indicates the GMM derived threshold of seropositivity of 97.7% specificity. Sample size refers to individual samples: Peru (n = 880), Senegal (n = 1077), French Guiana (n = 787), New Caledonia (n = 600). *Abbreviations: virus-like particle (VLP), envelope protein (E), envelope protein domain III (EDIII), non-structural protein 1 (NS1/NSP1).*



Supplementary Figure 10. The age-specific adjusted seroprevalence for all antigens at each of the study sites and survey time-points. The plot includes data for all serological markers where the area under the analytical ROC curve exceeded 0.9. All estimates of seroprevalence have been adjusted for imperfect sensitivity and specificity of the assay (see methods). Point colours denote the study site and survey year. Note that these estimates do not account for serological cross-reactivity between related viruses and that, due to cross-reactivity, the seroprevalence of a given virus may not necessarily reflect the underlying epidemiology of that virus at a given survey site. Sample size refers to individual samples: Peru (n = 880), Senegal (n = 1077), French Guiana (n = 787), New Caledonia (n = 600). *Abbreviations: virus-like particle (VLP), envelope protein (E), non-structural protein 1 (NS1)*



Supplementary Figure 11. Boxplot of the observed (solid) and model predicted (open) age-specific antibody levels for CHIKV (orange) and MAYV (purple) at each stud site. The dashed lines represent the model-estimated mean reactivity in unexposed individuals.



Supplementary Fig. 12. A graphical representation of how the model classifies each samples most-likely exposure history given the antibody response data for MAYV E2 and CHIKV E2 and data on age and location. The plot includes a subset of the survey sites and cross-sectional surveys.

Supplementary Methods

Bayesian Finite mixture models – Weibull distribution

The Weibull distribution is a flexible two-parameter probability distribution defined by a shape-parameter α and a scale-parameter σ:

$$Weibull(Ab|\alpha,\sigma) = \frac{\alpha}{\sigma}\left(\frac{Ab}{\sigma}\right)^{-\alpha-1} exp\left(-\left(\frac{Ab}{\sigma}\right)^{\alpha}\right) \quad (1)$$

For each antibody response, a two-part Weibull finite mixture model with shared parameters was fitted jointly to log-transformed antibody data from the negative (unexposed) validation samples and the cross-sectional survey samples in a Bayesian framework. The negative validation samples were assumed to be described by the lowest component only while the cross-sectional survey samples were assumed to be described by all of the components such that:

$$p(Ab_{neg}) = Weibull(Ab_{neg}|\alpha_1,\sigma_1) \quad (2)$$
$$p(Ab_{cs}) = \theta_1 * Weibull(Ab_{cs}|\alpha_1,\sigma_1) + \cdots + \theta_k * Weibull(Ab_{cs}|\alpha_k,\sigma_k)$$

Where k represents the number of mixture components, Ab_{neg} and Ab_{cs} represent the antibody response in the negative control samples and the cross-sectional survey samples, respectively, and $\theta_1,\ldots,\theta_k$ are the mixing coefficients (or weights) for each of the k components. For all antibody responses we evaluated both two- ($k = 2$) and three-component ($k = 3$) mixtures.

Adjusted estimates of population-level seroprevalence

Seropositivity was defined individually for each antigen using the Gaussian finite mixture model derived threshold of 97.7% specificity. When the sensitivity or specificity of an assay is imperfect (i.e. <1) but both are well-characterized, the measured seroprevalence can be expressed as a function of the true underlying population prevalence:

$$M(T) = T * sensitivity + (1-T)(1-specificity \quad (4)$$

where M is the measured seroprevalence and T is the true population-level seroprevalence.
By solving the equation for T, the true underlying population prevalence can be estimated by adjusting the measured seroprevalence for the sensitivity and specificity of the test:

$$T = \frac{(M + specificity - 1)}{(sensitivity - 1 + specificity)} \quad (5)$$

The adjusted population-level seroprevalence was calculated for all antibody responses with an analytical AUROC>0·9.

Supplementary Table 1. Naming convention for the parameters of the cross-reactivity and virus transmission model

Code parameter name	Methods parameter name
mu0**V**_global	$\mu_{0,global}^{V}$
sigma_mu0**V**	ω_s^V
alpha**V_s**	λ_s^V, α_s^V
mu**V**	μ^V
muMXC	$\gamma^{C \to M}$
muCXM	$\gamma^{M \to C}$
eps**V**	σ^V

Suplementary Table 2. List of recombinant antigens, bead regions and coupling concentrations

Bead region	Antigen name	Pathogen	Protein	Company	Catalogue no.	Expression System	C-terminal tag
63	CHIKV E2	Chikungunya virus (Senegal 37997 strain)	Envelope protein 2 (E2)	Native Antigen	REC31617-100	HEK293	mouce-FC
27	CHIKV VLP	Chikungunya virus (Senegal 37997 strain)	Virus like particle consisting of Envelope proteins (E2 and E2) and capsid protein	Native Antigen	CHIKV-VLP-100	HEK293	untagged
12	DENV1 EDIII	Dengue virus 1	Envelope protein domain III (EDIII)	Native Antigen	REC31761-100	*E. coli*	His
53	DENV1 NS1	Dengue virus 1	Non-structural protein 1 (NS1)	Native Antigen	DENV1-NS1-100	HEK293	His
61	DENV1 VLP	Dengue virus 1	Virus-like particle consisting of Envelope, pre-Membrane and Membrane proteins	Native Antigen	DENV1-VLP-100	HEK293	Secretion aided by C-terminal JE (all internal, transmembrane)
14	DENV2 EDIII	Dengue virus 2	Envelope protein domain III (EDIII)	Native Antigen	REC31762-100	*E. coli*	His
54	DENV2 NS1	Dengue virus 2	Non-structural protein 1 (NS1)	Native Antigen	DENV2-NS1-100	HEK293	His
64	DENV2 VLP	Dengue virus 2	Virus-like particle consisting of Envelope, pre-Membrane and Membrane proteins	Native Antigen	DENV2-VLP-100	HEK293	Secretion aided by C-terminal JE (all internal, transmembrane)
15	DENV3 EDIII	Dengue virus 3	Envelope protein domain III (EDIII)	Native Antigen	REC31763-100	*E. coli*	His
55	DENV3 NS1	Dengue virus 3	Non-structural protein 1 (NS1)	Native Antigen	DENV3-NS1-100	HEK293	His
65	DENV3 VLP	Dengue virus 3	Virus-like particle consisting of Envelope, pre-Membrane and Membrane proteins	Native Antigen	DENV3-VLP-100	HEK293	Secretion aided by C-terminal JE (all internal, transmembrane)
18	DENV4 EDIII	Dengue virus 4	Envelope protein domain III (EDIII)	Native Antigen	REC31764-100	*E. coli*	His
57	DENV4 NS1	Dengue virus 4	Non-structural protein 1 (NS1)	Native Antigen	DENV4-NS1-100	HEK293	His
66	DENV4 VLP	Dengue virus 4	Virus-like particle consisting of Envelope, pre-Membrane and Membrane proteins	Native Antigen	DENV4-VLP-100	HEK293	Secretion aided by C-terminal JE (all internal, transmembrane)
76	JEV NS1	Japanese Encephalitis virus	Non-structural protein 1 (NS1)	Native Antigen	JEV-NS1-100	HEK293	His
77	JEV E	Japanese Encephalitis virus	Envelope protein (E)	Native Antigen	REC31688-100	HEK293	His
72	MAYV NSP1	Mayaro virus (Brazil strain)	Non-structural protein 1 (NSP1)	Alpha Diagnostics	MAYV41-R-25	*E. coli*	His
34	MAYV E2	Mayaro virus (Acre27 strain)	Envelope protein 2 (E2)	Native Antigen	REC31644-100	HEK293	His (glycine-serine linker)
47	ONNV VLP	O'nyong nyong virus	Virus like particle consisting of Envelope proteins (E2 and E2) and capsid protein	Native Antigen	REC31626-100	HEK293	untagged
26	USUV NS1	Usutu Virus	Non-structural protein 1 (NS1)	Native Antigen	USUV-NS1-100	HEK293	HIs
29	WNV NS1	West Nile Virus	Non-structural protein 1 (NS1)	Native Antigen	WNV-NS1-100	HEK293	His
73	WNV EDIII	West Nile Virus (lineage 1, strain NY99)	Envelope protein domain III (EDIII)	Sino Biologicals	40345-V08Y	Yeast	His
44	YFV E	Yellow Fever virus	Envelope protein	Native Antigen	REC31798-100	insect	sheep-FC

45	YFV NS1	Yellow Fever virus	Non-structural protein 1 (NS1)	Native Antigen	YFV-NS1-100	HEK293	His
52	ZIKV VLP	Zika Virus	Virus-like particle consisting of Envelope, pre-Membrane and Membrane proteins	Native Antigen	ZIKV-VLP-100	HEK293	untagged
74	ZIKVAS DIII	Zika Virus (Asian strain)	Envelope protein domain III (EDIII)	Native Antigen	REC31775-100	*E. coli*	His
62	ZIKVSU NS1	Zika Virus (Suriname strain Z1106033)	Non-structural protein 1 (NS1)	Native Antigen	ZIKVSU-NS1-100	HEK293	His
46	ZIKVU NS1	Zika Virus (Uganda Strain)	Non-structural protein 1 (NS1)	Native Antigen	ZIKV-NS1-100	HEK293	His