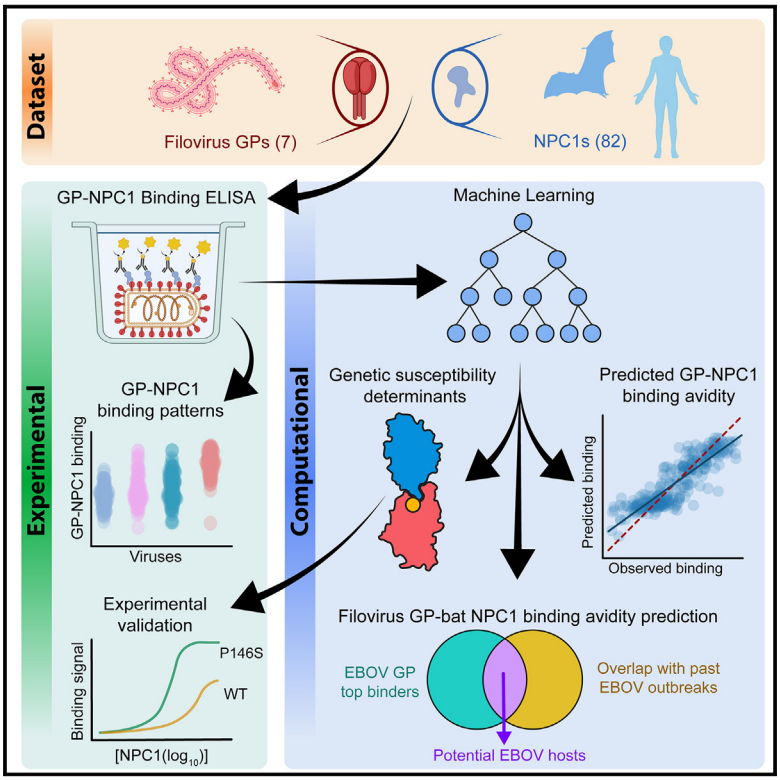
Article

Cell Host & Microbe

Decoding the blueprint of receptor binding by filoviruses through large-scale binding assays and machine learning

Graphical abstract



Highlights

- Differential binding propensities observed among filovirus GPs and bat NPC1s
- Amino acid properties predict GP-NPC1-binding avidity across filoviruses and bats
- Known bat hosts of filoviruses rank high in GP-NPC1-binding avidity analysis
- Combining receptor avidity and outbreak overlap refines predictions of EBOV hosts

Authors

Gorka Lasso, Michael Grodus, Estefania Valencia, ..., Rohit K. Jangra, Kartik Chandran, Simon J. Anthony

Correspondence

gorka.lasso@gmail.com (G.L.), rohit.jangra@lsuhs.edu (R.K.J.), kartik.chandran@einsteinmed.edu (K.C.), sjanthony@ucdavis.edu (S.J.A.)

In brief

Lasso et al. combine large-scale binding assays and machine learning to uncover molecular rules of filovirus-receptor binding, predict receptor-binding strength, and identify bats with the potential to act as filovirus hosts. These findings offer actionable data to design targeted surveillance strategies and anticipate potential animal-to-human transmission events.

Lasso et al., 2025, Cell Host & Microbe *33*, 294–313
February 12, 2025
https://doi.org/10.1016/j.chom.2024.12.016

CellPress

CellPress

Cell Host & Microbe

Article

Decoding the blueprint of receptor binding by filoviruses through large-scale binding assays and machine learning

Gorka Lasso,[1,25,*] Michael Grodus,[2,25,26] Estefania Valencia,[1,25,27] Veronica DeJesus,[1,28] Eliza Liang,[2] Isabel Delwel,[1,29] Rob H. Bortz III,[1,30] Dmitry Lupyan,[3] Hanna Y. Ehrlich,[4] Adrian A. Castellanos,[5] Andrea Gazzo,[6] Heather L. Wells,[7] Supaporn Wacharapluesadee,[8] Alexandre Tremeau-Bravard,[4] Janine F.R. Seetahal,[9] Tom Hughes,[10,11] Jimmy Lee,[10,11] Mei-Ho Lee,[10,11] Anna R. Sjodin,[12] Marike Geldenhuys,[13] Marinda Mortlock,[13] Isamara Navarrete-Macias,[7] Kirsten Gilardi,[4] Michael R. Willig,[12,14] Alessandra F.D. Nava,[15] Elisabeth H. Loh,[16] Makda Asrat,[7] Tierra Smiley-Evans,[4] Walter S. Magesa,[17] Sijali Zikankuba,[17] David Wolking,[4] Gerardo Suzán,[18] Rafael Ojeda-Flores,[18] Christine V.F. Carrington,[19] Ariful Islam,[20] Jonathan H. Epstein,[11] Wanda Markotter,[13] Christine K. Johnson,[4] Tracey Goldstein,[21] Barbara A. Han,[5] Jonna A.K. Mazet,[4] Rohit K. Jangra,[22,23,24,*] Kartik Chandran,[1,31,*] and Simon J. Anthony[7,*]

[1]Department of Microbiology and Immunology, Albert Einstein College of Medicine, Bronx, New York, NY 10461, USA
[2]Center for Infection and Immunity, Mailman School of Public Health, Columbia University, New York, NY 10032, USA
[3]Schrödinger Inc., Cambridge, MA 02142, USA
[4]One Health Institute, School of Veterinary Medicine, University of California, Davis, CA 95616, USA
[5]Cary Institute of Ecosystem Studies, Millbrook, NY 12545, USA
[6]Department of Pathology and Laboratory Medicine, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[7]Department of Pathology, Microbiology, and Immunology, School of Veterinary Medicine, University of California, Davis, CA 95616, USA
[8]Thai Red Cross Emerging Infectious Diseases Clinical Center, King Chulalongkorn Memorial Hospital, Faculty of Medicine, Chulalongkorn University, Pathumwan, Bangkok 10330, Thailand
[9]Department of Diagnostic Medicine/Pathobiology, College of Veterinary Medicine, Kansas State University, Manhattan, KS 66502, USA
[10]Conservation Medicine, 47000 Sungai Buloh, Selangor, Malaysia
[11]EcoHealth Alliance, New York, NY 10018, USA
[12]Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT 06269, USA
[13]Centre for Viral Zoonoses, Department of Medical Virology, University of Pretoria, Pretoria 0001, South Africa
[14]Center for Environmental Sciences and Engineering, Institute of the Environment, University of Connecticut, Storrs, CT 06269, USA
[15]Fundação Oswaldo Cruz-Fiocruz, Instituto Leônidas & Maria Deane, Laboratório de Ecologia de Doenças Transmissíveis na Amazônia – EDTA, Manaus 69.057-070, AM, Brazil
[16]Division of Natural Sciences and Mathematics, Transylvania University, Lexington, KY 40508, USA
[17]College of Veterinary Medicine and Biomedical Sciences, Sokoine University of Agriculture, P.O. Box 3021, Morogoro, Tanzania

(Affiliations continued on next page)

SUMMARY

Evidence suggests that bats are important hosts of filoviruses, yet the specific species involved remain largely unidentified. Niemann-Pick C1 (NPC1) is an essential entry receptor, with amino acid variations influencing viral susceptibility and species-specific tropism. Herein, we conducted combinatorial binding studies with seven filovirus glycoproteins (GPs) and NPC1 orthologs from 81 bat species. We found that GP-NPC1 binding correlated poorly with phylogeny. By integrating binding assays with machine learning, we identified genetic factors influencing virus-receptor-binding and predicted GP-NPC1-binding avidity for additional filoviruses and bats. Moreover, combining receptor-binding avidities with bat geographic distribution and the locations of previous Ebola outbreaks allowed us to rank bats by their potential as Ebola virus hosts. This study represents a comprehensive investigation of filovirus-receptor binding in bats (1,484 GP-NPC1 pairs, 11 filoviruses, and 135 bats) and describes a multidisciplinary approach to predict susceptible species and guide filovirus host surveillance.

INTRODUCTION

Filoviruses in the genera *Orthoebolavirus* and *Orthomarburgvirus* have caused sporadic but increasingly frequent outbreaks of fatal hemorrhagic fever in humans for decades.[1] However, for many of these viruses, most notably the ebolaviruses, the particular animal species that act as reservoir or bridge hosts remain poorly defined, limiting efforts to forecast zoonotic spillover

[18]Laboratorio de Ecología de Enfermedades y Una Salud, Departamento de Etología, Fauna Silvestre y Animales de Laboratorio, Facultad de Medicina Veterinaria y Zootecnia, Universidad Nacional Autónoma de México, Ciudad de México, 04510, Mexico
[19]Department of Preclinical Sciences, Faculty of Medical Sciences, The University of the West Indies, St. Augustine, Republic of Trinidad and Tobago
[20]Gulbali Research Institute, Charles Sturt University, Wagga Wagga, NSW 2678, Australia
[21]One Health Institute, Colorado State University, Fort Collins, CO 80523, USA
[22]Department of Microbiology and Immunology, Louisiana State University Health Sciences Center-Shreveport, Shreveport, LA 71103, USA
[23]Center for Applied Immunology and Pathological Processes, Louisiana State University Health Sciences Center-Shreveport, Shreveport, LA 71103, USA
[24]Center of Excellence for Emerging Viral Threats, Louisiana State University Health Sciences Center-Shreveport, Shreveport, LA 71103, USA
[25]These authors contributed equally
[26]Present address: Rockefeller University, New York, NY 10065, USA
[27]Present address: Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[28]Present address: Cancer Research UK Beatson Institute, Glasgow G61 1BD, Scotland
[29]Present address: Stanford University, Stanford, CA 94305, USA
[30]Present address: University of Pittsburgh Medical Center, Pittsburgh, PA 15219, USA
[31]Lead contact
*Correspondence: gorka.lasso@gmail.com (G.L.), rohit.jangra@lsuhs.edu (R.K.J.), kartik.chandran@einsteinmed.edu (K.C.), sjanthony@ucdavis.edu (S.J.A.)
https://doi.org/10.1016/j.chom.2024.12.016

events. The largest filovirus outbreak to date, caused by Ebola virus (EBOV; genus *Orthoebolavirus*), occurred in West Africa in 2014–2016, infecting over 28,000 people and resulting in over 11,000 deaths.[2] This epidemic was unprecedented in terms of its size and location and highlighted clear gaps in our understanding of the host and geographic range of EBOV. The most *recent* filovirus outbreak, caused by Marburg virus (MARV; genus *Orthomarburgvirus*), began in September 2024 in Rwanda—an area where MARV had not previously been reported. This outbreak, which as of December 6th has resulted in 66 cases and 15 deaths (https://rbc.gov.rw/marburg/), serves as a reminder that filoviruses continue to spill over to humans in regions not historically associated with these pathogens.

Multiple lines of evidence suggest that bats are the primary hosts of filoviruses.[3–12] The most robust data are for MARV and Lloviu virus (LLOV; genus *Cuevavirus*), which have been isolated from Egyptian fruit bats (*Rousettus aegyptiacus*) and Schreiber's bat (*Miniopterus schreibersii*), respectively.[13–15] Bombali virus (BOMV; genus *Orthoebolavirus*) has been repeatedly detected in Angolan free-tailed bats (*Mops condylurus*), although it has not yet been isolated from wild-caught individuals.[16–20] Similarly, Měnglà virus (MLAV; genus *Dianlovirus*) has been identified in Chinese rousette bats (*Rousettus amplexicaudatus*),[21,22] and Dehong virus (DEHV; genus *Delovirus*) has been detected in Chinese Leschenault's rousette bats (*Rousettus leschenaultii*).[23] By contrast, the evidence for the prototypic ebolaviruses—EBOV, Sudan virus (SUDV), Tai Forest virus (TAFV), Bundibugyo virus (BDBV), and Reston virus (RESTV)—is less robust. For instance, RESTV RNA has been detected in *Miniopterus schreibersii* bats in the Philippines,[24] but this evidence remains limited, and significant taxonomic revisions complicate these findings. Historically considered a single, widespread species, *M. schreibersii* populations exhibit substantial genetic diversity, making it unlikely that the bats in the Philippines represent the same species that hosts LLOV in Europe.

EBOV, in particular, highlights the persistent gaps in our understanding of filovirus ecology. Despite being isolated in 1976 and causing over 24 human outbreaks, this virus continues to emerge in unexpected places at unexpected times. Although the hosts of EBOV are not definitively known, a wealth of circumstantial evidence nonetheless points to bats. Short fragments of EBOV RNA have been detected in fruit-eating bats (*Myonycteris torquata*, *Epomops franqueti*, *Hypsignathus monstrosus*).[3] Serological surveys have also revealed the presence of EBOV-reactive antibodies in various bat species.[3,6,25–31] Epidemiologic evidence links past outbreaks of EBOV with the contact or consumption of bats,[9] and ecological investigations indicate geographic co-occurrence of presumptive host species and EBOV outbreaks.[10–12] However, low viral RNA detection rates, failure to recover infectious virus from wild-caught individuals, and reliance on highly cross-reactive serologic data mean that no species have been definitively linked to the virus. Other non-bat mammals (e.g., primates, duikers, swine, and rodents) can also serve as susceptible hosts of EBOV, likely facilitating virus amplification and zoonotic transmission, although their role in the ecology of EBOV remains unclear.[1,32–43]

Various computational strategies have been employed to predict potential hosts for filoviruses.[12,44,45] Biological and ecological traits from filovirus-positive bats were used to predict the likelihood that different bat species harbor filoviruses.[44] The model predicted potential bat hosts in Southeast and East Asia, which was later supported by studies reporting diverse filoviruses in bats from China.[21,46] Thus, predictive models have clear utility in guiding surveillance for unknown filoviruses. However, these models lack sufficient depth to make predictions for specific filoviruses, such as hosts of EBOV versus hosts of BOMV. For instance, *Eidolon helvum* bats were predicted to host filoviruses[44]; however, *E. helvum* cells are refractory to EBOV infection.[47] Thus, while *E. helvum* may indeed host a filovirus, it is unlikely that this filovirus is EBOV. This lack of granularity stems from knowledge gaps, reliance on cross-reactive serologic data, and significant sampling bias, leaving most bat species unexamined. This underscores the need for expanded surveillance to identify filovirus hosts, particularly for EBOV. However, without data to constrain and prioritize sampling goals, such surveys quickly become costly and impractical.

Molecular data can overcome some of the limitations of ecological models. As a case in point, the cellular Niemann-Pick C1 (NPC1) protein is a critical molecular trait governing filovirus susceptibility by virtue of its role as an indispensable entry

receptor for all known filoviruses.[48,49] Two protruding loops in human NPC1 domain C (NPC1-C) interact with a hydrophobic pocket and a hydrophilic patch on EBOV GP.[50] Amino acid variation at this interface significantly influences cellular susceptibility and host species-specific tropism for multiple filoviruses.[47,51–55] Therefore, the comprehensive characterization of GP-NPC1 binding across different filoviruses and bat species could inform higher-resolution models predicting filovirus-specific hosts.

Here, we conducted large-scale GP-NPC1-C (hereafter referred to as GP-NPC1) experimental binding studies across seven divergent filovirus GPs and NPC1 orthologs (from human and 81 bat species). To our knowledge, this is the largest initiative to experimentally characterize receptor binding in filoviruses. Experimental binding data was used to train a random forest regressor to identify and understand the genetic determinants of viral susceptibility and develop a model that predicts GP-NPC1-binding avidity. Our model captures general principles underpinning receptor binding, affording predictions across additional bat species and viruses. Combining it with bat geographical distribution and locations of past EBOV outbreaks afforded the identification of potential bat hosts for EBOV. Such an approach could be extended to other filoviruses and updated as new filovirus and animal sequences are reported, thereby informing future surveillance efforts.

RESULTS

GPs and NPC1-Cs show differential binding propensities

We expressed and purified soluble NPC1-Cs from humans and 81 bat species and engineered recombinant vesicular stomatitis viruses (rVSVs) displaying different filovirus GPs on their surfaces (EBOV, SUDV, RESTV, BDBV, TAFV, BOMV, and MLAV). The GPs on the rVSVs were cleaved with thermolysin to expose their NPC1-binding site.[56,57] We then performed ELISA binding experiments to evaluate interactions between the rVSVs and the NPC1-Cs, using human NPC1-C as a control (Tables S1 and S2). The bat subset contains representatives for 12 of the 19 known bat families from four different continents (North America, South America, Africa, and Asia) (Figures S1A and S2B). Bat NPC1-Cs showed high sequence identity to each other (Figure S1C), with good agreement between the NPC1-C sequence-based dendrogram and bat phylogeny. Similarly, the viral GPs clustered according to taxonomic assignments but showed greater sequence variability (Figure S1D).

We observed different binding propensities between bat NPC1-Cs and between virus GPs (Figures 1A–1C). From a virus-centric perspective, BOMV GP bound bat NPC1s poorly, whereas TAFV GP and SUDV GP showed the highest binding propensities (Figure 1B). From a host-centric perspective, NPC1s from rhinolophid and phyllostomid bats showed the poorest binding, whereas NPC1s from molossid and pteropodid bats generally showed the strongest binding (Figure 1C). We also observed that NPC1s from African bats bound filovirus GPs more strongly than bats from the Americas or Asia (Figure 1D). Exceptions included the GPs of BOMV, MLAV (poor binders), and SUDV (good binder), which showed similar binding to NPC1s in African and non-African bats (Figure S2).

Filovirus susceptibility correlates with receptor-binding avidity

Our dataset includes the NPC1 from *Mops condylurus*, the known host for BOMV. The binding avidity for this interaction ranks in the 93rd percentile (Figure 1E; Table S2), suggesting that the receptor of bat hosts binds with high-binding avidity to the corresponding filovirus GPs. To further support the relationship between NPC1 binding and filovirus susceptibility, we cross-referenced our dataset with bat species with known evidence of filovirus susceptibility.[44,58] Out of the 70 bat species reported to test positive for at least one filovirus (mostly by serology and/or PCR), 16 (plus human) are present in our dataset (Table S3). NPC1s from susceptible organisms generally bind filovirus GPs more strongly than those categorized as possessing "unknown susceptibility" ($p_{\text{Mann-Whitney}} = 2 \times 10^{-4}$; Figure 1F). We also cross-referenced our binding data with published *in vitro* infectivity assays using bat-derived cell lines.[47,54,55] Our ELISA binding curves aligned with 23 of 24 reported infectivity assays (Figure S3). The only disagreement was *E. helvum* NPC1-RESTV GP, to which our binding ELISA reported poor binding; however, the *Eh*-derived bat cell line is susceptible to infection.[47,55]

GP-NPC1-binding avidity agrees poorly with overall amino acid sequence identity

We observed a weak, yet significant, positive correlation between sequence identity and binding similarity. For GPs (Spearman's rho: 0.49, *p*val 4.9×10^{-2}), sequence identities typically lie between 55% and 60%, a range where binding similarity can vary considerably (Figure 2A). Similarly, we observed dissimilar binding avidities among NPC1-Cs (Spearman's rho: 0.27, *p*val 1.3×10^{-54}) at pairwise sequence identities as high as $\geq$95% (Figure 2B). Such poor agreement is also evident when comparing the corresponding dendrograms (Figures 2C and 2D). For instance, BDBV and TAFV GPs showed the highest pairwise sequence identity (74.3%), yet their distinct NPC1-binding profiles caused them to segregate in a binding-based dendrogram (Figure 2C). On the whole, clustering bat NPC1s based on sequence identity and binding similarity revealed two different dendrograms: although the sequence-based dendrogram agrees with bat phylogeny, the binding-based dendrogram shows NPC1s from different bat families with similar binding profiles and vice versa (Figure 2D). These findings provide evidence that GP-NPC1 binding cannot be predicted based on overall sequence identity or phylogenetic distance and that physicochemical differences between particular amino acid residues distinguishing two orthologs are important.

Physicochemical description of amino acids predicts GP-NPC1-binding avidity

We trained a random forest regressor to predict binding avidity between filovirus GPs and NPC1s using the experimentally generated dataset and physicochemical properties of amino acids (Figures 3A and S4; Table S4).[59–64] Rather than describing all GP and NPC1 residues at once, we used the known EBOV

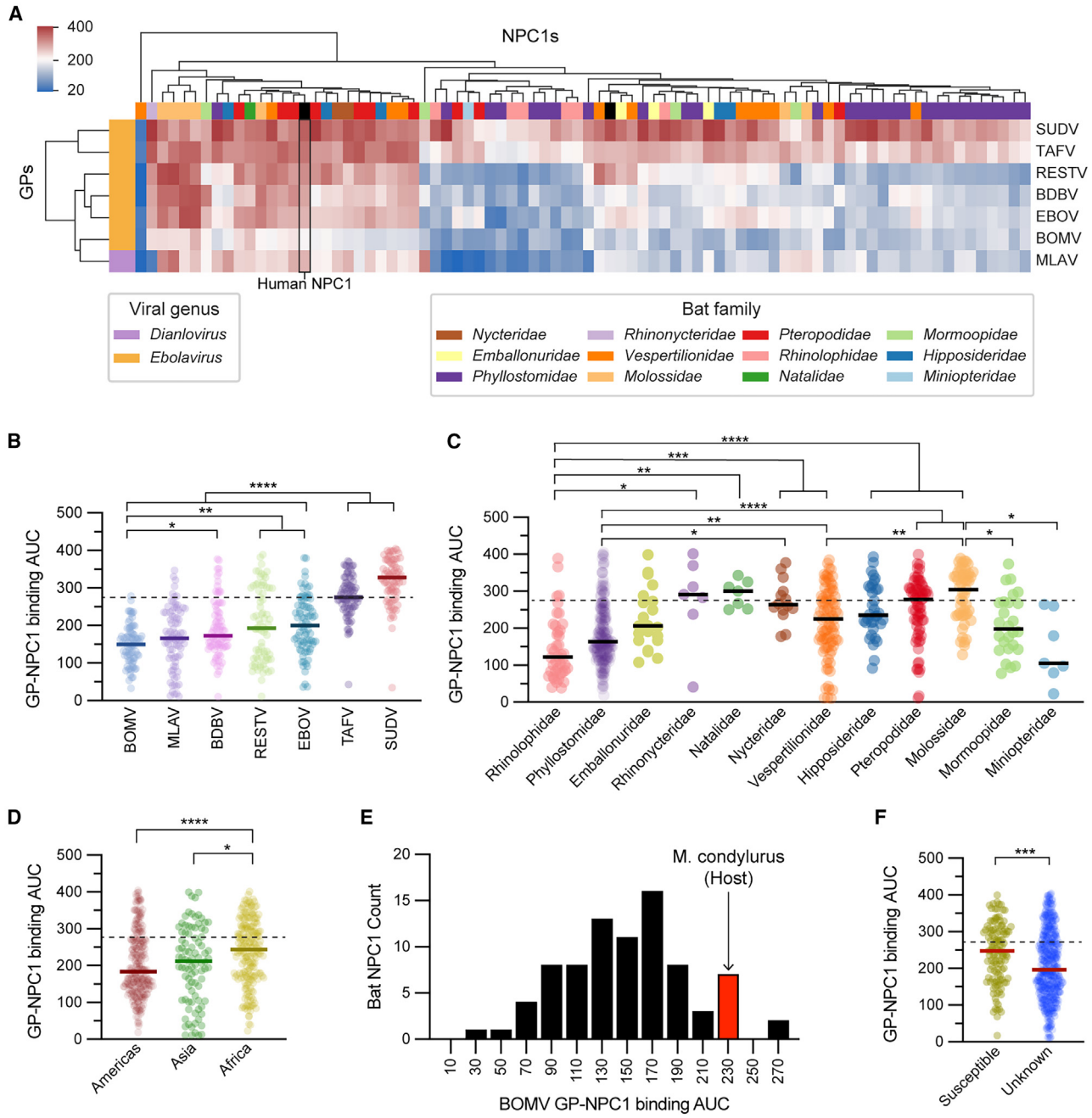


Figure 1. Filovirus GPs and bat NPC1-Cs show differential binding propensities

(A) Hierarchical clustering of NPC1s and filovirus GPs based on the experimental binding avidity (binding area under the curve [AUC]; Table S2).

(B) NPC1-binding avidity for each filovirus GP.

(C and D) GP-binding avidity for each bat NPC1, grouped by family (C) or geographic location (D).

(E) Distribution of BOMV GP-NPC1-binding avidities. Red bar: binding avidity range encompassing the known BOMV host.

(F) GP-binding avidities for each bat NPC1 (median binding AUC), grouped by evidence of filovirus susceptibility.

(B–E) Horizontal bars correspond to median values. (B–D and F) Horizontal dotted line corresponds to the reference binding AUC (EBOV GP-human NPC1: 277.2). Kruskal-Wallis corrected for multiple comparisons using Dunn's test (Levels of significance: *0.01 < p < 0.05; **0.001 < p < 0.01; ***0.0001 < p < 0.001; ****$p \leq$ 0.0001, see STAR Methods). See also Figures S1–S3 and Tables S1, S2, S3, and S4.

GP-human NPC1 complex[50] to identify residues within a variying inter-subunit distance cutoff (from 5 to 20 Å). At the lowest distance cutoff, only the residues in the immediate binding interface were considered. However, as we increased the distance cutoff, residues that contribute to binding through long-range effects were also considered. We used 90% of the experimental data

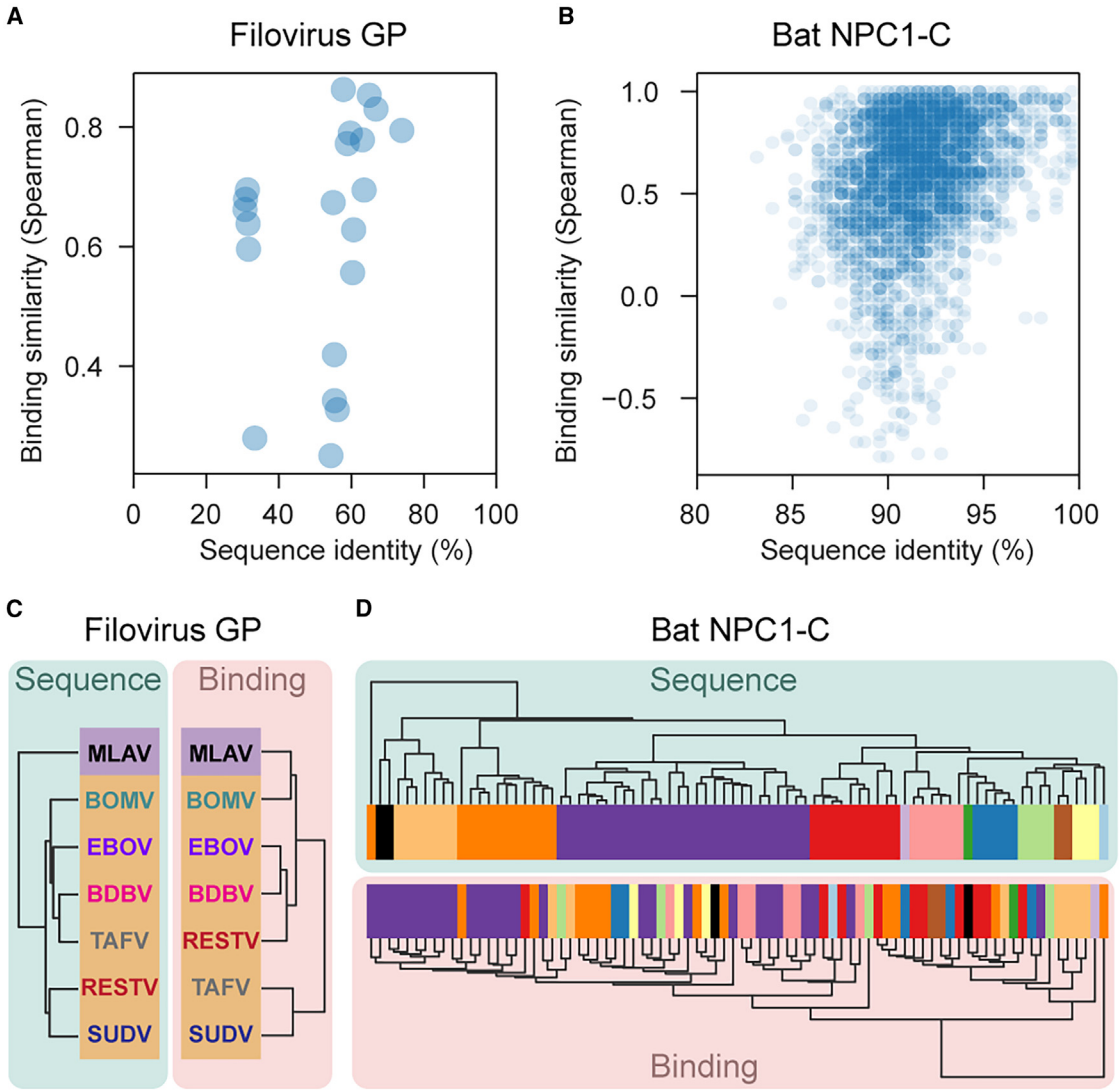


Figure 2. Pairwise sequence identity poorly correlates with binding

(A and B) Pairwise comparison between filovirus GPs (A) and bat NPC1s (B) based on sequence identity and binding similarity (as dictated by Spearman correlation).

(C and D) Comparison of sequence-identity-based dendrogram (green) and binding-based similarity dendrogram (red). Colored labels correspond to filovirus genera (C) or bat families (D), as in Figure 1A. See also Figure S1 and Table S1.

for training, hyperparameter optimization, and 10-fold cross-validation (CV) and set aside 10% of the data for model testing. Since the GPs and NPC1s in the test set are also considered in the training set (though not the exact GP-NPC1 combinations), there is a risk of overestimating the performance of our regression model. To address this, we assembled a separate set comprising LLOV GP-NPC1-C binding data. LLOV GP shares a sequence identity of 37% ± 3.5 with GPs in the training set and has unique interfacial residues mediating the interaction with NPC1 (Figure S5A).

First, we evaluated the predictive performance of the models trained using only a single protein scale at a time (Figure 3A). Hydrophobicity was the most important protein scale, as shown by 10-fold CV and evaluation on the test and the LLOV datasets. Next, we trained the random forest regressor on each possible combination involving 2–9 amino acid scales at five different inter-subunit distances (2,555 models; Figure S5B). We selected the best-performing models based on their evaluation across different evaluation schemes (CV, test, LLOV) and performance consistency at varying inter-subunit distance cutoffs. When choosing the second-best-performing model, we also prioritized models trained on unique features not shared with the best model. The best-performing model was trained using 25 features describing the aromatic, coil, and hydrophobicity propensities of 21 residues located up to 7 Å away from the binding partner (Figures 3B and S5C; Table S4). The second-best-performing model was trained on 50 features describing the aromatic, charge, flexibility, and hydrophobicity of 45 residues located up to 15 Å away from the binding partner (Figures 3B and S5D; Table S4). We found that although hydrophobicity is the most

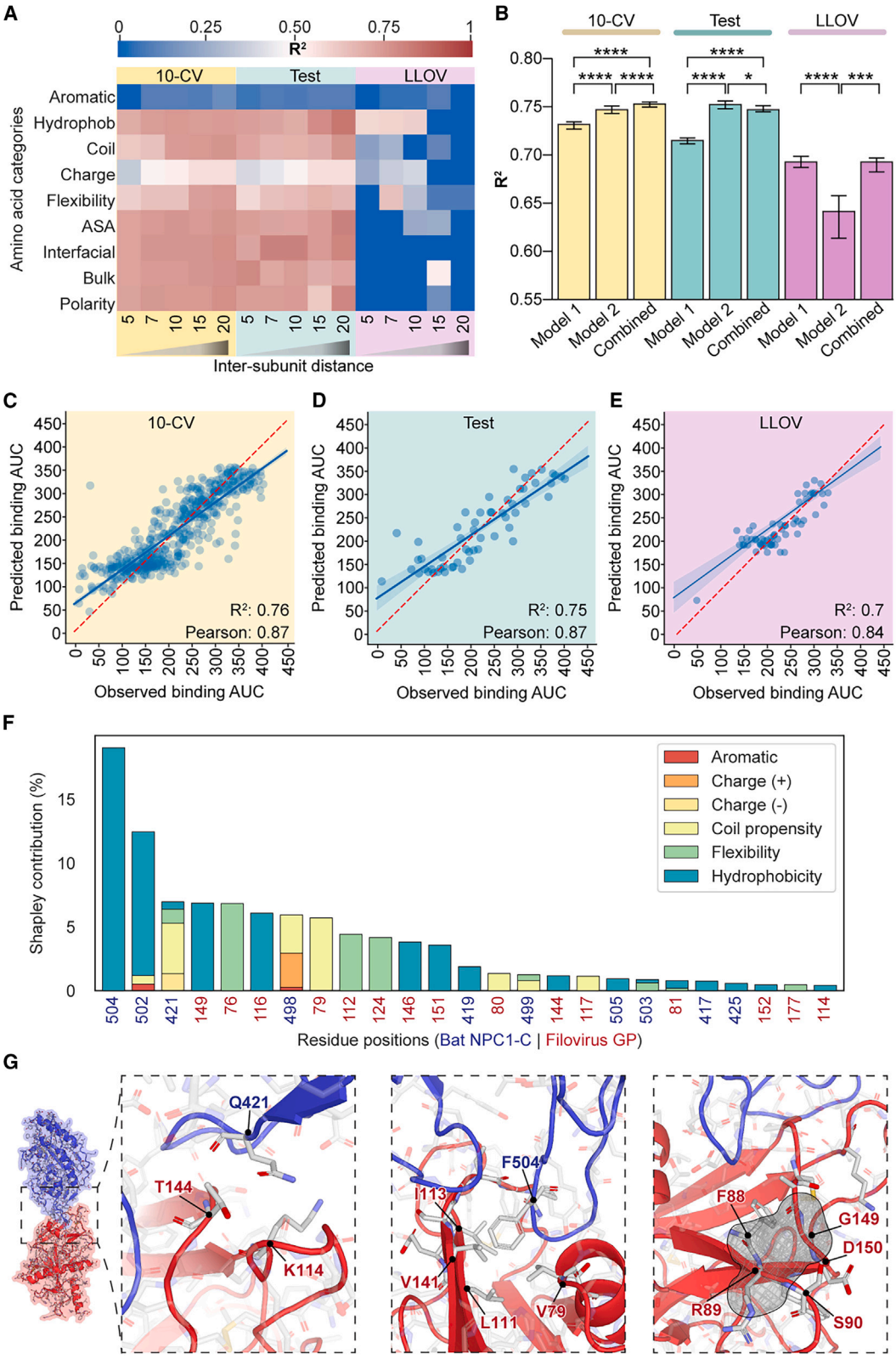


Figure 3. Physicochemical properties of amino acids predict binding avidity between filovirus GPs and bat NPC1-Cs

(A–E) Model evaluation on three different schemes/datasets (yellow: 10-fold cross validation, CV; teal: test; purple: LLOV rVSV dataset). (A) Evaluation of random forest regressors trained on one single amino acid scale at a time (rows). Only residues within a distance cutoff of the binding partner are considered (columns). (B)

(legend continued on next page)

relevant property for predicting binding, other amino acid properties with lower predictive power on their own (e.g., flexibility; Figure 3A) become important when combined with orthogonal information (Figure 3B). Since the majority of features leveraged by each of the two models are unique, we combined both models into a single meta-predictor maximizing the overall performance and consistency of the predictions (R^2 for 10-fold CV, test set, and LLOV dataset were 0.75, 0.75, and 0.69, respectively; Figures 3B–3E). We further evaluated the meta-predictor using per-NPC1 and per-GP cross-validation (Figure S6). Specifically, we iteratively withheld the binding data for one NPC1 or GP at a time during training and used it to evaluate the meta-predictor. The predictive performance is consistent, suggesting the model can be applied to filovirus GPs and bat NPC1s outside the experimental datasets. On a per-NPC1 basis, the model showed good performance ($R^2 > 0.6$ and/or Pearson correlation > 0.7) for 73 out of 82 NPC1s, with a median Pearson correlation of 0.92. On a per-GP basis, the model performed well for four viruses (EBOV, BDBV, RESTV, and TAFV; $R^2 > 0.6$ and/or Pearson correlation > 0.7) and moderately well for BOMV (Pearson correlation: 0.68), with a median Pearson correlation of 0.76. The generally higher performance observed in per-NPC1 cross-validation is likely due to the larger number of NPC1 orthologs, which facilitates the deconvolution of co-occurring amino acid variations in specific NPC1s. The random forest meta-predictor, also referred to as meta-predictor or model for simplicity hereafter, was used in the subsequent analyses described below.

Model interpretation reveals general principles dictating GP-NPC1 binding

The meta-predictor leverages physicochemical properties for 52 residues (31 and 21 residues from GP and NPC1, respectively), the majority of which sit at the GP-NPC1 interface (73% of selected residues are within 7 Å of the binding partner). We used Shapley additive explanations (SHAP) to identify the most relevant residues and properties in our model.[65,66] NPC1 residues 421, 502, and 504 contributed the most to binding prediction (human NPC1 residue numbering; Figure 3F; Table S4). Previous structural studies have highlighted the roles of these residues on filovirus GP binding.[47,50,67] Human NPC1 Q421 is in the protruding loop 1 that engages with the hydrophilic patch of EBOV GP. This residue is in close contact with two EBOV GP residues, K114 and T144, that are also utilized by our model, suggesting that the interplay between these three residues is an important determinant of binding (Figure 3G). Human NPC1 D502 is in the protruding loop 2, which engages with the hydrophobic pocket of GP, and its substitution with Phe in *E. helvum* NPC1 abrogates binding to EBOV GP.[47] Human NPC1 F504 is also in the protruding loop 2 and makes key inter-subunit contacts (Figure 3G). In line with these observations, the hydrophobicity of NPC1 residues 502 and 504 is the physicochemical property that plays the largest role in our model (Figure 3F).

On the viral side, the overall contribution of selected residues is more evenly distributed (Figure 3F; Table S4). Several residues (79, 114, 115, 140, 151, and 156; EBOV GP residue numbering) have been shown experimentally to influence NPC1 binding.[68–70] Our model predicts residue 149 to be one of the most relevant residues in GP. To our knowledge, however, its potential role as a genetic determinant of NPC1 binding and susceptibility has not been described. In EBOV, G149 sits at one of the edges where the two halves of the hydrophobic NPC1-binding cavity meet. *In silico* mutagenesis suggested that larger residues at this position (e.g., BOMV: Ser; BDBV: Glu; MLAV: Lys) would cause an opening of the binding pocket or affect the conformation of neighboring residues that mediate the interaction with NPC1 (Figure 3G).

The flexibility propensity of GP residue 76 is the second most important viral feature in the meta-predictor. Among the GPs used for training, only SUDV GP (S76) differs from the others at this position (A76, Figure S5A). We hypothesize that this variation, along with other amino acid differences, may contribute to the high-binding propensity observed for SUDV GP (Figure 1B), as we describe later.

S146 decreases binding avidity of BOMV GP-mediated interactions with bat NPC1-Cs

BOMV GP showed the lowest propensity to bind NPC1s (Figure 1B). Model interpretation suggests a key role for residue 146 (EBOV GP residue numbering; Figure 4A). Residue identity at position 146 differs among BOMV strains, with a Ser present in the BOMV GP in *M.condylurus* (the primary host species[16–20]) and a Pro present in the BOMV GP of *C. pumilus* (a likely secondary host[16]). In our experiments, the BOMV isolate contains Ser, whereas all other filovirus GPs have Pro (Figure S5A). We hypothesized that GP P146S may at least partially explain the overall lowest binding avidity of BOMV GP (Figure 1B). We generated single amino acid-mutant viruses, rVSV-BOMV GP S146P and rVSV-EBOV GP P146S, and assessed their binding to human NPC1. This single amino acid change did not affect protein expression (Figure 4B) but was sufficient to reproduce the NPC1-binding phenotypes of the rVSVs bearing BOMV and EBOV GP (Figure 4C). These findings were mirrored by VSV-based single-cycle infectivity assays in wild-type (WT) and NPC1-overexpressing cells (Figure 4D), where viral infection aligned with the identity of the residue at position 146, regardless of GP background (S146 > P146). This strongly suggests that the amino acid change at this position is determinative.

We performed molecular dynamics free energy perturbation on the EBOV GP-human NPC1 complex to understand how P146S affects binding. P146S is predicted to increase

Comparison of the best predictive models (each model integrates features from various amino acid scales) and the combined model (averaging the predictions made by models 1 and 2). Kruskal-Wallis corrected for multiple comparisons using Dunn's test (Levels of significance: *0.01 < p < 0.05; **0.001 < p < 0.01; ***0.0001 < p < 0.001; ****$p \leq 0.0001$, see STAR Methods). (C–E) Comparison of observed and predicted binding avidities.

(F) Contribution of the 25 most relevant residues to the meta-predictor. Stacked bars describe the importance of the physicochemical properties (inset legend) of selected residues (x axis: red and blue labels correspond to filovirus GP residues and bat NPC1 residues, respectively).

(G) Selected residues that contribute to the model as shown in the EBOV GP (red)-human NPC1 (blue) complex.[50] Right: density cloud represents the area of naturally occurring residues (Ser, Glu, Lys) at position 149, depicting the likely steric clashes with F88 and R89. See also Figures S4–S6 and Table S4.

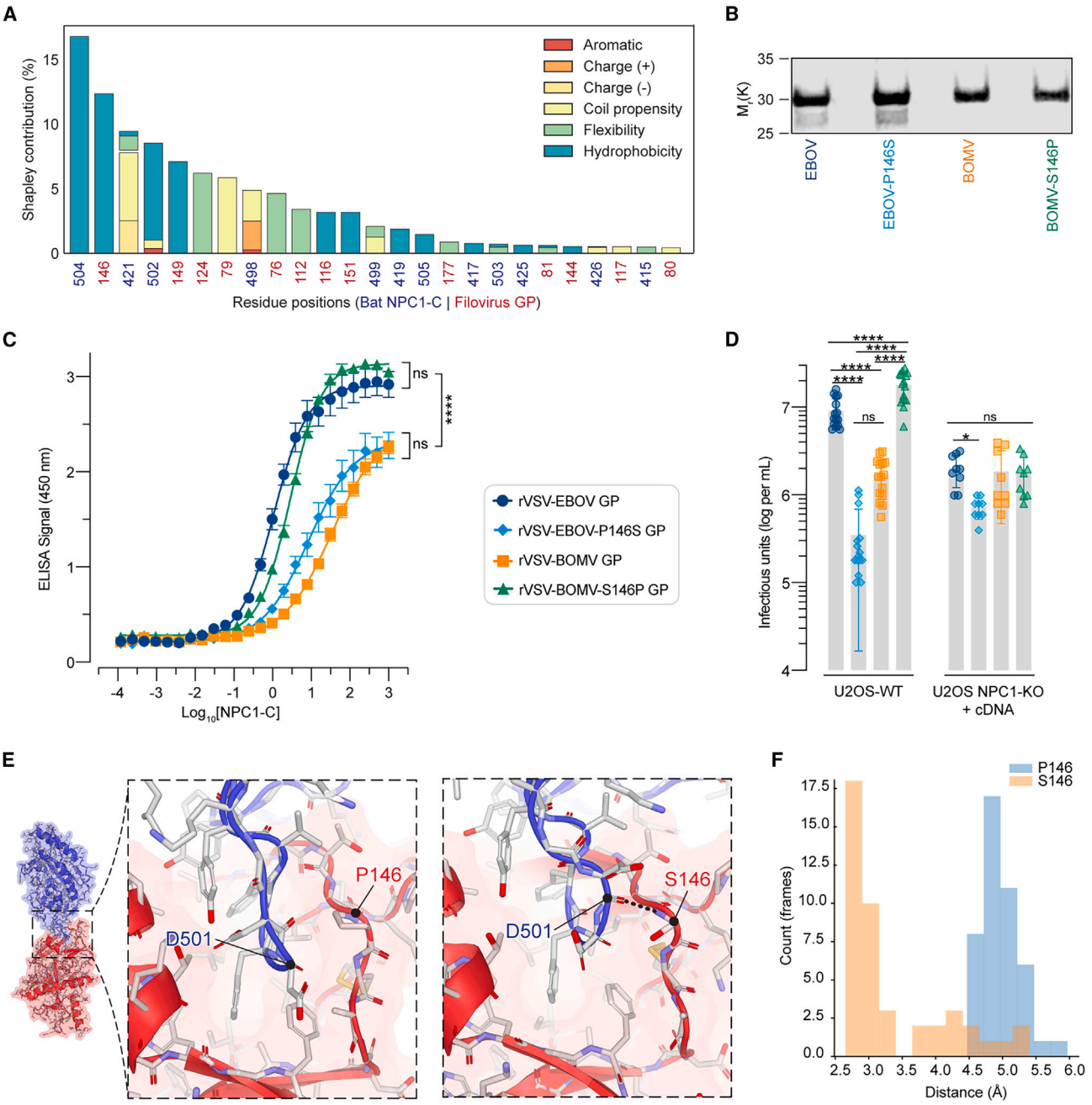


Figure 4. S146 in BOMV GP is a key phenotypic determinant of NPC1 binding and cell entry

(A) Contribution of the 25 most relevant residues to the meta-predictor in BOMV GP-mediated interactions. Stacked bars describe the importance of physicochemical properties (inset legend) of selected residues (x axis: red and blue labels correspond to filovirus GP and bat NPC1-C residues, respectively).

(B) Western blot of VSV M signal to normalize rVSV input for infectivity assay.

(C) Binding ELISA between human NPC1 and rVSV-EBOV GP (dark blue), rVSV-EBOV P146S GP (light blue), rVSV-BOMV GP (orange), and rVSV-BOMV S146P GP (green).

(D) Infection assay in U2OS cells (left) and U2OS NPC1 KO + cDNA cells (right) with rVSV-GPs.

(E) Left: EBOV GP (red)-human NPC1 (blue) complex. Middle and right: representative MD trajectory snapshots showing a close-up view of human NPC1 (blue) bound to EBOV GP (middle; red) and EBOV GP P146S (right; red), and a backbone-to-backbone hydrogen bond between NPC1 D501 and EBOV P146S GP is shown.

(F) Distance distribution between heavy atoms involved in the backbone-to-backbone hydrogen bond between NPC1 D501 and EBOV GP 146 throughout the simulation in the context of P146 and S146. (C) Two-way ANOVA, corrected for multiple analyses using Tukey. (D) One-way ANOVA, corrected for multiple analyses using Tukey.Levels of significance: *$0.01 < p < 0.05$; **$0.001 < p < 0.01$; ***$0.0001 < p < 0.001$; ****$p \leq 0.0001$ (see STAR Methods).

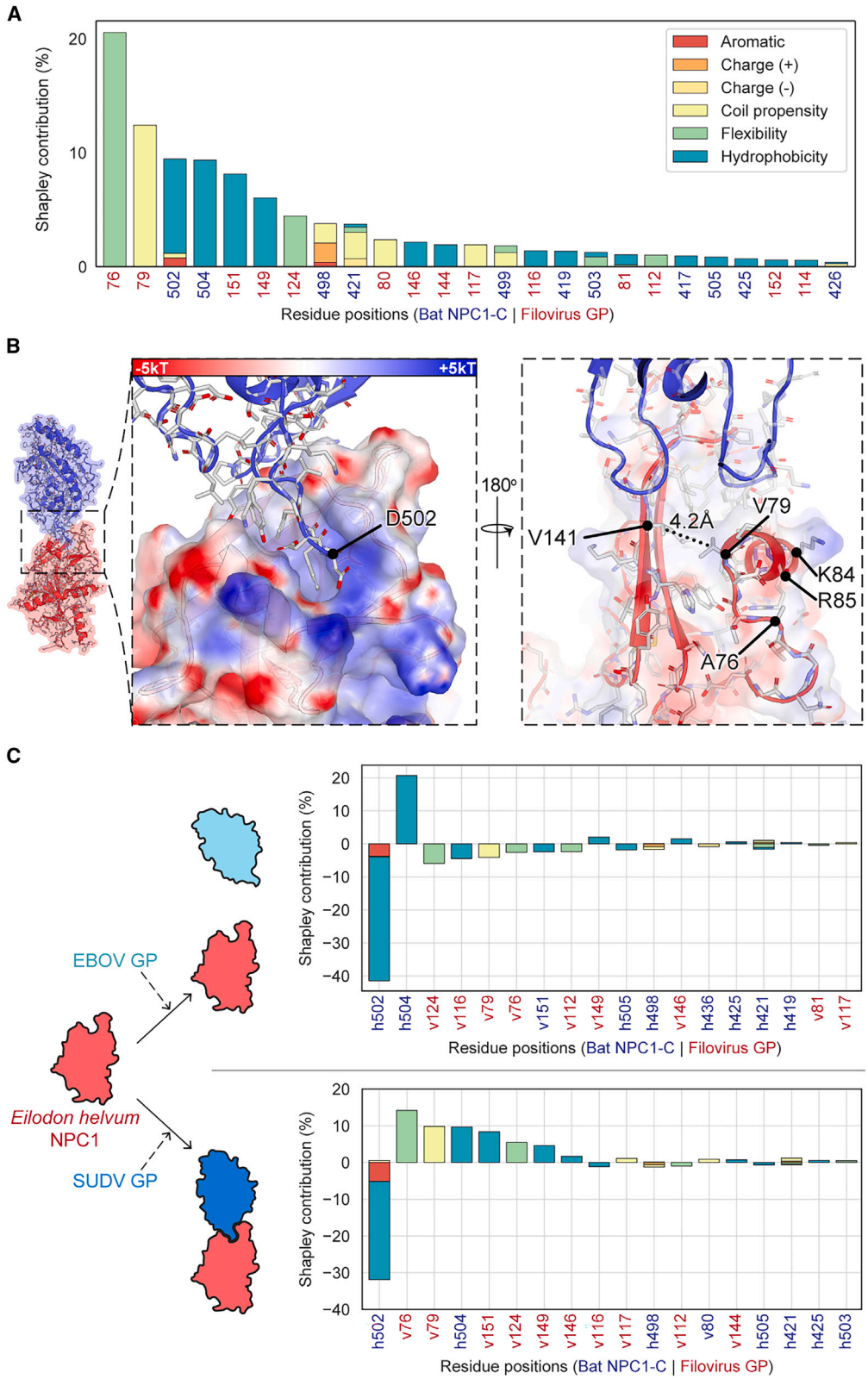


Figure 5. Amino acid variants govern specific susceptibility to particular filoviruses

(A) Contribution of the 25 most relevant residues to the meta-predictor in SUDV GP-mediated interactions. Stacked bars describe the importance of physico-chemical properties (inset legend) of selected residues. x axis: red and blue labels correspond to filovirus GP and bat NPC1-C residues, respectively.

(legend continued on next page)

the free energy of binding, destabilizing the complex (ΔΔG = +2.0 kcal/mol).[71] Further analysis of the MD trajectories revealed that the backbone nitrogen of S146 forms a hydrogen bond with the backbone oxygen of human NPC1 D501 (Figures 4E and 4F). This newly formed bond brings the respective protein backbones closer by ∼2 Å and causes a kink in the human NPC1 loop that mediates key interactions with EBOV GP, potentially causing differences in the binding mode that result in a lower binding affinity. These results, together with other recent work,[72,73] demonstrate that model interpretation can be leveraged to identify amino acid variations that can significantly impact GP-NPC1 binding, viral entry, and spread.

Spatially proximal amino acid variants likely underpin the increased binding avidity of SUDV GP for bat NPC1-Cs

SUDV GP showed the highest binding propensity (Figure 1B). Model interpretation, via SHAP, identified the flexibility of residue 76 and the coil propensity of residue 79 as the most relevant viral features driving NPC1 binding (EBOV GP residue numbering, Figure 5A). These residues do not directly engage NPC1 (Figure 5B), suggesting that variants at these positions trigger a conformational change in the cavity that affects binding in a generalized manner. In EBOV GP, V79 forms a hydrophobic contact with V141 in the opposite half of the pocket (Figure 5B). This interaction likely plays a structural role in stabilizing the cavity, so mutations at either position may promote cavity opening. Indeed, the EBOV GP V141A mutation was shown to enhance binding to *E. helvum* NPC1 and suggested to open up the binding pocket.[47,74] In SUDV, residues at positions 79 and 141 are unique (I79, A141; Figure S5A). Although both contribute to predicting binding avidity, the SHAP value for A141 is minimal compared with I79 (Table S4), which may be the result of feature redundancy rather than a biophysical phenomenon (either one alone is sufficient for accurate binding prediction). On the other hand, *in silico* mutagenesis suggests that S76 can establish hydrogen bonds with K84 or R85, which might also contribute to cavity opening (Figure 5B). In this case, residues identified at positions 84 and 85 do not differ between filovirus GPs and do not contribute toward binding prediction (Table S4). Altogether, our results suggest SUDV residues S76 and I79 work in a concerted manner with A141 to promote an opening of the hydrophobic cavity, which might in turn promote a generalized increased in NPC1 binding.

Random forest model identifies genetic determinants dictating species-specific susceptibility in *Eidolon helvum*

Previous cell-based infectivity assays showed that *E. helvum* cells are refractory to EBOV infection due to a single amino acid variation in their NPC1—D502F—that abrogates EBOV GP binding.[47] However, the same cells were susceptible to SUDV, and further mutagenesis experiments showed that V141A EBOV GP compensates for the effect of D502F in *E. helvum* NPC1, suggesting a potential avenue for host range expansion.[47] We sought to evaluate whether our model could successfully recapitulate these known intersubunit relationships. SHAP analysis revealed that the hydrophobicity of residue 502 is the strongest negative contributor to the lower binding predicted between *E. helvum* NPC1 and EBOV GP (Figure 5C). Moreover, SUDV S76 and I79 are predicted to counteract the effect of F502. Since residues 141 and 79 interact, our results align well with the published literature and further support the capacity of our model to identify the likely residues underlying receptor binding and host susceptibility.[47]

NPC1 residues under positive selection contribute to GP-NPC1-binding predictions

Positive selection represents adaptive genetic changes that accrue due to selection pressures, including those imposed by viruses.[75] Previous work has shown that NPC1 residues crucial for GP binding exhibit distinctive patterns of accelerated evolution, suggesting that mutations at these locations may confer some resistance to infection in bats.[47] However, amino acid residues in host proteins that interact with viral proteins tend to be evolutionarily constrained since they typically interact with other host proteins. In NPC1, a substantial proportion of the GP-interacting residues are also involved in the interaction with host NPC2, as part of the lysosomal cholesterol trafficking pathway (Figure 6A).[76] We identified 10 residues under positive selection (Table S5), six of which are at the GP-binding interface (human NPC1 positions 421, 502, 505, and 528) or its vicinity (<11 Å from binding partner, human NPC1 positions 417 and 436).[77,78] Our model utilizes physicochemical features derived from six residues under positive selection (enrichment fold: 6.1, hypergeometric *p*val: 6.9×10^{-5}), suggesting that our model captured amino acid variants that influence GP binding. Of the four residues under positive selection that localize to the GP-binding interface, three (421, 502, and 505) are also at the NPC2-binding interface, albeit with lower buried surface area (BSA), suggesting a greater role of these residues in GP binding (Figure 6B). The remaining residue under positive selection (528) is uniquely found in the GP-binding interface. These results provide a potential roadmap of NPC1 residues that evolved to decrease binding avidity to filovirus GPs without significantly impairing binding to NPC2.

Model predictions of GP-NPC1 binding for additional filoviruses and bat species

We predicted GP-NPC1 binding for additional 54 bat species (bats outside the experimental dataset with known NPC1 sequences), LLOV, and three additional filoviruses (MARV, RAVV, and DEHV[23]), with a total of 135 bat species from 14 families and 11 filoviruses. This yielded 627 experimentally generated and 858 predicted paired binding avidities (hereafter termed the extended dataset; Table S4). We employed a

(B) EBOV GP (red)- human NPC1 (blue) complex and closed-up views of the binding interface, the surface of EBOV GP is colored by the electrostatic potential. (C) Left: binding summary of *E. helvum* NPC1 (red) for EBOV GP (cyan) and SUDV GP (blue). Right: contributions of the most relevant residues used by the meta-predictor to predict the binding avidity *of E. helvum* NPC1 for EBOV GP (top) and SUDV GP (bottom). Stacked bars describe the importance of physicochemical properties of selected residues (x axis: red and blue labels correspond to filovirus GP and bat NPC1-C residues, respectively).

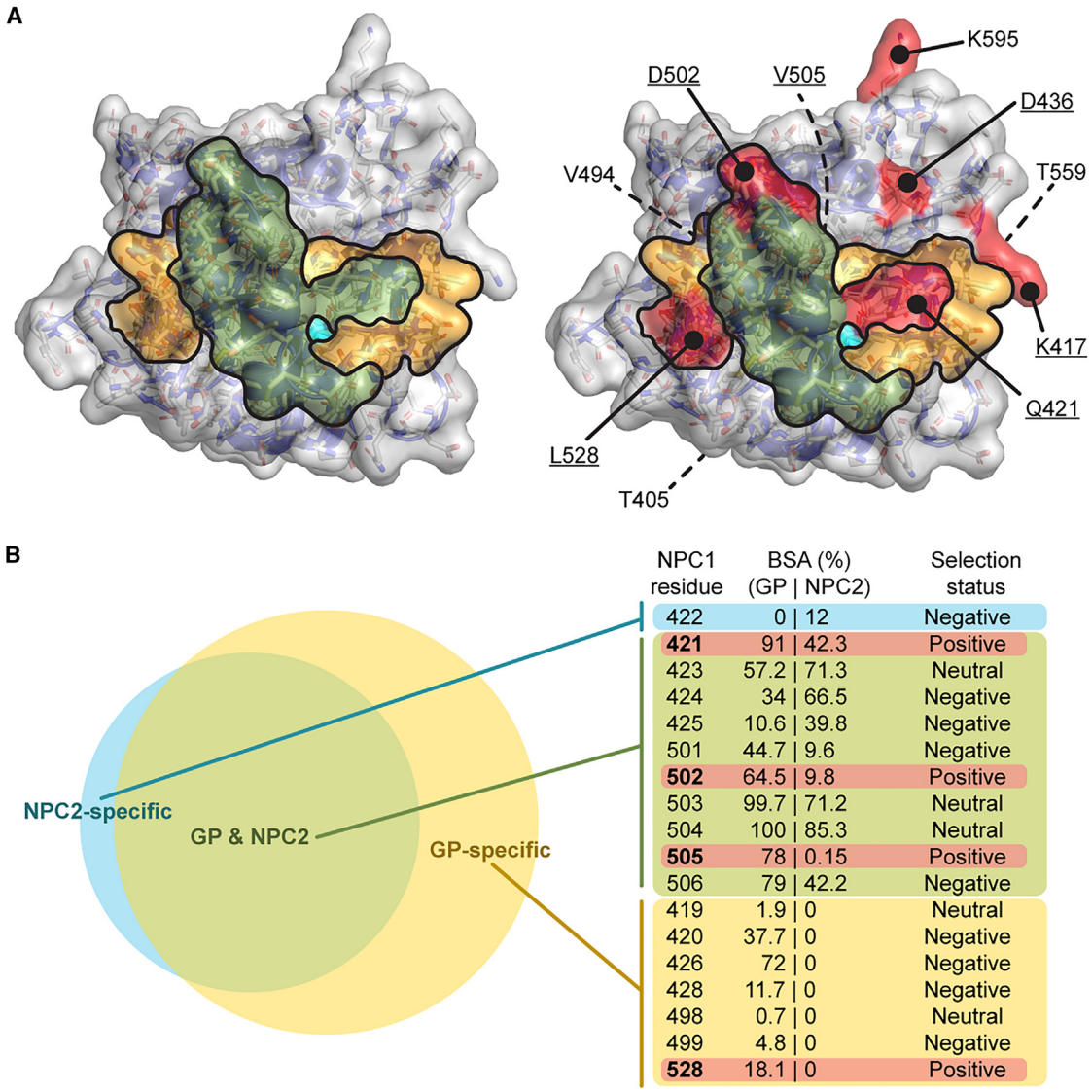


Figure 6. The random forest regressor leverages NPC1 residues under positive selection

(A) Left: surface representation of human NPC1 showing the EBOV GP and human NPC2 binding interfaces (light blue: NPC2-specific, golden: GP-specific, green: GP and NPC2). Right: residues under positive selection (Mixed Effects Model of Evolution [MEME] or Fubar) are highlighted in red, and residues used by the meta-predictor are underlined.

(B) Left: Venn-diagram representation of the interfacial NPC1-C residues mediating the interaction with EBOV GP and NPC2. Right: description of each interfacial residue in NPC1, including the buried surface area (BSA%) in the interaction with GP and NPC2 and the evolutionary selection status. See also Table S5.

hierarchical generalized linear mixed effects model to study overall trends, controlling for potential dependencies and confounders among variables (Figure 7A). The observations from the extended dataset aligned with the experimental-only dataset (Spearman r: 0.91). In addition, divergent GPs were predicted to bind with high avidity to NPC1s from vespertilionid (RESTV, LLOV, and EBOV) and emballonurid bats (SUDV) (Figure S7A). Unexpectedly, American molossid bats displayed a uniquely high-binding profile (Figures 7B, 7C, and S7B–S7F). Removing this family from the mixed effects model recapitulated the significant differences observed between African bats as compared with American and Asian bats (Figures 1D and S7G).

Integrating GP-NPC1-binding avidity, bat geographical distributions, and historical outbreak information informs on potential EBOV hosts in Africa

Our experimental results indicate that bat filovirus receptors and those in other susceptible bats generally bind their respective filovirus GPs with higher avidity (Figures 1E and 1F). Predicted binding avidities showed that host NPC1 interactions for MARV and DEHV ranked above the 80th percentile (Figures 7D and 7E). These findings support the meta-predictor as a tool to identify potential bat hosts for filoviruses, extending beyond those used for model training and optimization.

However, the receptor of a host may not be the strongest binder among orthologs due to factors like differences at other

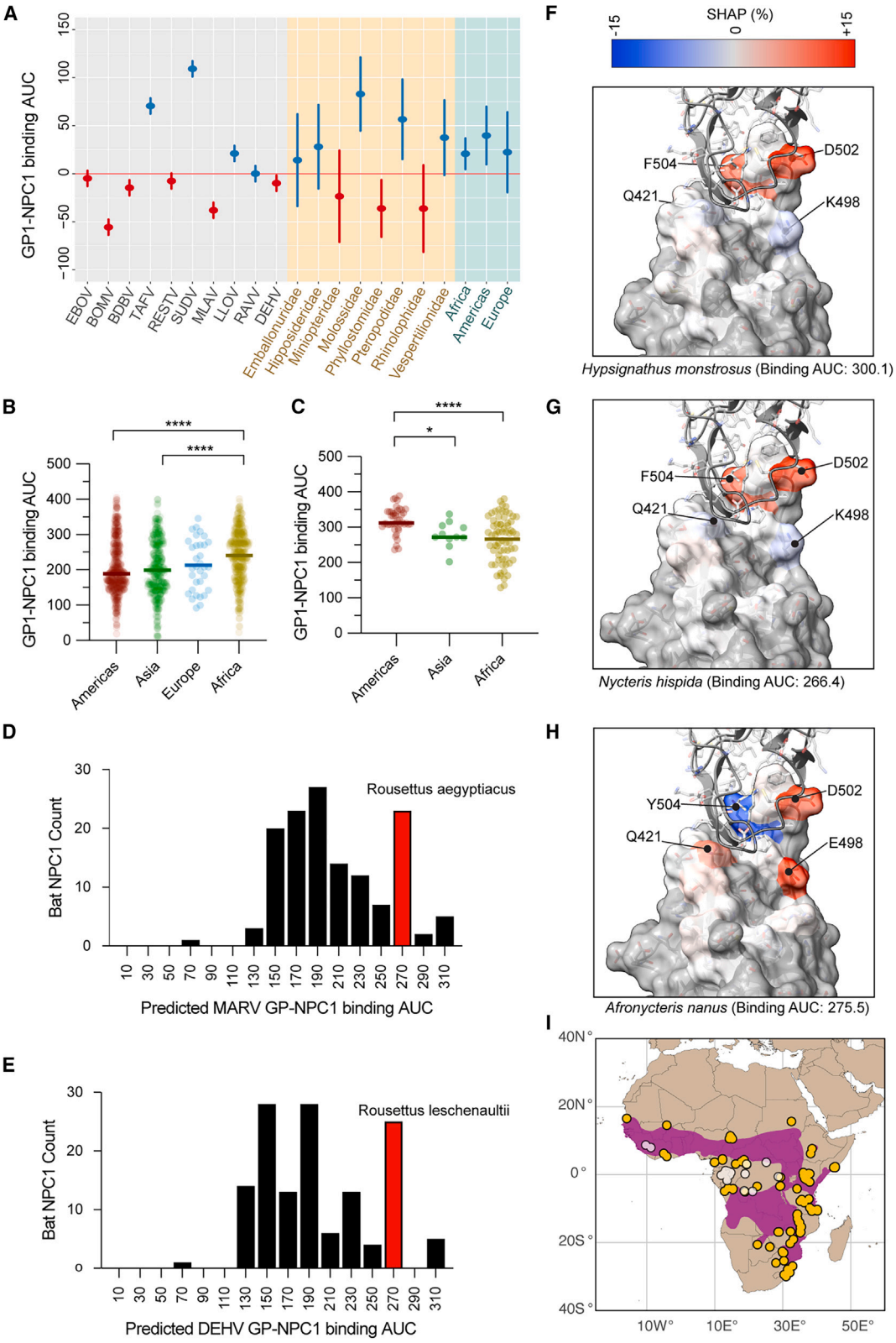


Figure 7. Experimental and predicted GP-NPC1-binding avidities inform on bats with the potential to act as host

(A) Hierarchical mixed effects model, accounting for random effects of bat family, for experimental and predicted binding avidities. Ellipses and whiskers (lines) represent the estimated coefficients and 95% confidence intervals, respectively, with colors corresponding to positive (blue) or negative (red) effects. Results are

(legend continued on next page)

viral entry steps, immune evasion capabilities, and evolutionary adaptations favoring lower binding affinity. Consequently, relying solely on receptor-binding avidities may produce false positives. Accordingly, we reasoned that bats in outbreak-prone regions whose NPC1 strongly binds EBOV GP would be high-priority surveillance targets. To this end, we cross-referenced EBOV GP-NPC1-binding data with geographic information, including bat species distributions and past human outbreak locations (Table S6).

We found 17 bat species whose geographic distribution overlapped >75% with the location of previous human outbreaks of EBOV. Six of these are represented in our extended dataset, observing a range of binding avidities. Strong binders include *Hypsignathus monstrosus* (*Pteropodidae*), *Nycteris hispida* (*Nycteridae*), and *Afronycteris nanus* (*Vespertilionidae*).

We modeled the EBOV GP-NPC1 complexes for all three host candidates and used SHAP to identify key residues contributing to these interactions (Figures 7F–7H). In *H. monstrosus* and *N. hispida*, NPC1 residues D502 and F504 enhance binding avidity. By contrast, *A. nanus* NPC1 Y504 is predicted to decrease binding avidity, as the loss of hydrophobicity at this position is known to disrupt GP binding.[50] Although F504Y would be expected to destabilize the complex, our binding assays show binding comparable to human NPC1 (Table S2). SHAP analysis suggests that both Q421 and E498 counteract the negative effect of Y504 in *A. nanus*, highlighting how the local environment can influence the contribution of specific amino acid variations to GP-NPC1 binding.

Further, several other bat species in our extended dataset have the genetic *potential* to host EBOV based on GP binding, but their geographic distribution poorly overlaps with known EBOV outbreaks. Among those, *Mops condylurus* is the known host of BOMV and a suspected host for EBOV.[16–20,84] The range of *M. condylurus* extends when including additional surveillance data, covering additional regions where EBOV outbreaks have occurred (Figure 7I; Table S6).[27,30,31,81,82] Therefore, it is likely that the geographic range of this species is incomplete.

DISCUSSION

Zoonotic viruses continue to have profound impacts on public health, the global economy, and society at large. Yet predicting and mitigating their transmission to humans remains a formidable challenge. A promising avenue to prevent animal-to-human spillovers is to survey and monitor animals serving as host for pathogens, especially if those are found near human populations. For many filoviruses, including those within the genus *Orthoebolavirus*, their natural hosts remain unknown. Moreover, our ability to predict the functional consequences of sequence variation or the identity of species that could serve as intermediate hosts for human transmission is limited. The 2014–2016 EBOV disease (EVD) epidemic in West Africa, the deadliest ever recorded, highlights the urgent need to identify and survey the natural host(s) of ebolaviruses causing fatal disease in humans.

Receptor binding plays a critical role in filovirus infection as it is essential for viral entry. Accordingly, we sought to uncover the molecular mechanisms governing the interaction between the filovirus glycoprotein (GP) and its receptor (NPC1) to support more informed predictions of filovirus host range. We integrated large-scale experimental binding assays and machine learning to estimate GP-NPC1-binding avidity across a panel of filoviruses and bats and identify the key genetic determinants of this indispensable virus-receptor interaction.

Patterns of GP-NPC1 binding

Bat NPC1s from the molossids and pteropodids exhibit higher binding propensities to filovirus GPs, whereas those from the rhinolophids and phyllostomids display lower binding propensities. These trends are corroborated by accumulating knowledge about filovirus host ecology. Molossids and pteropodis have been linked to filoviruses based on molecular, serological, or virological data.[3–7,13,16–20,24–26,28–31,46,85–103] By contrast, no filovirus has been identified in rhinolophids or phyllostomids to date, although serological evidence of filovirus exposure has been reported in both bat families.[7,85,98,100,103,104] We observed similar patterns of GP-NPC1 binding when taking an EBOV-centric view. EBOV GP showed significantly higher binding avidity to NPC1s from molossid and pteropodid bats. This finding is well supported by previous studies.[3–7,25–28,30,31,84,85,91,97,101] However, no full EBOV genomes have yet been recovered from any of these species, and no infectious virus has been isolated. This study reaffirms that surveillance efforts to identify the host(s) of EBOV should prioritize molossid and pteropodid bats. Importantly, binding propensities did not correlate with the number or significance of human outbreaks, reinforcing that genetic and ecological factors beyond GP-NPC1 binding contribute to spillover, amplification, and spread of zoonotic viruses.

Importantly, binding propensities did not correlate with the number or significance of human outbreaks. Although TAFV GP binds bat NPC1s with high avidity, most human outbreaks were caused by EBOV or SUDV, with TAFV linked to a single human case.[105] This apparent contradiction suggests that other factors also influence the zoonotic outbreak potential. This is

grouped by fixed effect category (gray: virus-centric; yellow: bat-centric, reference; cyan: bat-centric, grouped by geographical location). Points and bars reflect estimated differences in binding avidity compared with a reference (MARV, *Mormoopidae*, and Asia, for each category, respectively). The horizontal red line at zero represents no difference in binding compared with the reference, and whiskers that do not intersect with the dashed line demonstrate statistical significance. (B and C) GP-binding avidities (experimental or predicted) for all bat NPC1s in the extended dataset (B) and for each molossid NPC1 considered in the extended dataset (C), grouped by geographical location. Horizontal bars correspond to median values. Kruskal-Wallis corrected for multiple comparisons using Dunn's test (Levels of significance: *$0.01 < p < 0.05$; **$0.001 < p < 0.01$; ***$0.0001 < p < 0.001$; ****$p \leq 0.0001$, see STAR Methods).

(D and E) Distribution of predicted GP-NPC1-binding avidities for MARV (D) and DEHV (E). Red bar: binding avidity bin encompassing the known hosts for the corresponding filoviruses.

(F–H) Interaction models of EBOV GP (cartoon representation) with NPC1s (surface representation) from *H. monstrosus* (F), *N. hispida* (G), and *Afronycteris nanus* (H). The NPC1 surface is colored by their corresponding SHAP contributions.

(I) Accepted distribution of *M. condylurus* (purple shade) in Africa,[79,80] additional locations were obtained from the USAID PREDICT[81] and African Chiroptera Report[82] (yellow circles). The map also shows the location of past human outbreaks of EBOV (white circles).[83] See also Figure S7 and Table S6.

unsurprising, given how many other genetic and ecological factors are involved in spillover, amplification, and spread of zoonotic viruses.

GP-NPC1 binding can refine existing models leveraging broader features

Because NPC1 binding is crucial for filoviruses to infect host cells, poorer receptor binding is expected to result in lower cellular susceptibility, with implications for viral host range.[47,52,55] Indeed, we found good agreement between receptor binding and host susceptibility.[44,47,54,55,58] This suggests that current strategies to predict filovirus hosts based on broader features (e.g., environmental data and genomic properties) would benefit from including protein interactions dictating viral entry.[44,58,106–108] For instance, a model that integrates biological, environmental, and ecological traits ranked *Pteropus medius* (previously *Pteropus giganteus*), an Asian pteropodid bat, in the 90th percentile probability as a potential host.[44] However, our ELISA binding assays showed poor binding between its NPC1 and the GPs of filoviruses found in Asia (RESTV and MLAV), suggesting that *P. medius* is unlikely to act as their natural host. While this does not preclude *P. medius* from being the host of a different, unknown, filovirus, our results provide important mechanistic context for serological positives, which are common due to high cross-reactivity.[104] Moreover, our results demonstrate that receptor-binding propensities do not correlate with phylogenetic distances and that single non-synonymous variations can have a profound effect on binding.

Molecular determinants of GP-NPC1 binding

We trained a random forest regressor to predict receptor-binding avidity by integrating binding experiments, amino acid sequence information, and structural information. Selected features leveraged by the model align with structural and biochemical data, suggesting that the model at least partially captured the physicochemical nature of the contacts known to drive filovirus-receptor binding.[50] Using SHAP for model interpretation, we identified residues known to play a major role in binding and host tropism and uncovered others whose functional relevance has not yet been proven. We also identified specific amino acid changes that are predicted to induce structural changes that affect NPC1 binding in a generalized manner, likely explaining the overall lower and higher binding avidities observed for BOMV and SUDV GPs, respectively. Further, our model incorporates physicochemical properties of residues under positive selection, which may result from long-term co-evolution between filoviruses and bats, promoting variations that hinder filovirus GP binding.[47] These outputs align well with previous independent research exploring filovirus biology, serving as additional validation for our machine-learning model.

An approach to refine surveillance

When the model was applied to other filoviruses outside the experimental datasets (MARV and DEHV), the predicted binding avidities for the NPC1s from known hosts ranked above the 80th percentile. This suggests that the model can be used to assess receptor binding for novel filoviruses or newly sequenced bats. To explore the utility of the model in identifying surveillance priorities, we integrated our GP-NPC1-binding data with the geographic distributions of African bat species, focusing on potential EBOV hosts. This exercise demonstrated both the value and current limitations of our approach: most bat species overlapping with past EBOV outbreaks lack an available NPC1 sequence. This gap highlights the need to prioritize sequencing efforts for specific bats, including three pteropodis (*Myonycteris angolensis*, *Epomops franqueti*, and *Myonycteris torquata*) and a hipposiderid (*Macronycteris gigas*). We also identified bats whose range overlaps with EBOV and whose NPC1s bind to EBOV GP with high avidity, one of which (*Hypsignathus monstrosus*) is proposed to harbor EBOV based on previous serological and molecular evidence.[3,4,27,29–31] Similarly, this approach can be used to de-prioritize bat species whose geographic distribution overlaps with past EBOV outbreaks but whose NPC1s bind poorly to EBOV GP (e.g., *E. helvum*). Although geographical data adds a much-needed dimension to the interpretation of molecular binding data, however, predictions on EBOV hosts are necessarily provisional, since bat distribution maps are often incomplete or inaccurate and will undergo refinement over time.

Another potential application is identifying bat species in non-endemic regions that could act as hosts during virus expansion, as we assessed NPC1-GP binding across bats globally. Notably, molossid bats in the Americas stand out for their uniquely high-binding avidity. While this observation might reflect sampling bias, it is also possible that these bats have evolved stochastically in ways that enhance their affinity for filovirus GPs. These findings warrant further investigation and suggest that American molossid bats may be valuable targets for surveillance.[100] Finally, our tool could also be used to assess the zoonotic potential of novel filoviruses for spillover into humans, as human NPC1 is included in our dataset.

Limitations of the study

Our approach has several limitations. First, because both GPs and NPC1s differ at multiple residues, feature redundancy minimization may have selected residues based on statistical associations rather than causation. High-throughput experimental methods, such as deep mutational scanning, could help disentangle the contributions of co-occurring amino acid variants. Second, the naturally occurring amino acid variants in GP and NPC1 exceed those represented in our dataset, potentially limiting the scope of the model. However, extensive evaluations suggest that the predictor can effectively assess receptor binding for filoviruses and hosts beyond the experimental datasets. Third, while receptor binding is a critical step, other biological factors—such as host-virus interactions during post-entry stages of the viral lifecycle and host immune responses—likely also influence the susceptibility of bat hosts to filovirus infection.[47,109,110]

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by Simon Anthony (sjanthony@ucdavis.edu) or Kartik Chandran (lead contact, kartik.chandran@einsteinmed.edu).

Materials availability

Plasmids in this study, together with documenting information, will be made available upon request and following execution of a UBMTA between Albert

Einstein College of Medicine or UC Davis (as appropriate, depending on the resource being requested) and the recipient institution.

Data and code availability

- The genomic data have been deposited in GenBank (accession numbers GenBank: PQ137083–PQ137194) and are publicly available as of the date of publication. See key resources table for further details.
- All original code has been deposited at GitHub and is publicly available as of the date of publication. DOI (https://doi.org/10.6084/m9.figshare.28144118) is also listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This study was made possible in part by the generous support of the American people through the United States Agency for International Development (USAID) Emerging Pandemic Threats PREDICT program (Cooperative Agreement Numbers GHN-A-00-09-00010-00 and AID-OAA-A-14-00102). PREDICT supported the collection of samples used to generate NPC1-C sequences and the construction of NPC1-C expression constructs. This work was also supported by both the National Institutes of Health (R01AI101436, R21AI156482, P20GM134974, and P30CA008748) and the National Science Foundation (NSF) Predictive Intelligence for Pandemic Prevention program (Phase I #2200221 and Phase II #2412522). NSF supported data analysis and manuscript preparation but not sample acquisition or laboratory experimentation. The contents are the responsibility of the authors and do not necessarily reflect the views of USAID, NIH, NSF, or the United States Government. We thank the governments of participating countries for their collaboration and permissions to conduct this study.

AUTHOR CONTRIBUTIONS

Conceptualization: G.L., J.A.K.M., K.C., and S.J.A. Methodology: G.L., M.G., E.V., D.L., H.Y.E., A.A.C., A.G., H.L.W., R.K.J., K.C., and S.J.A. Software: G.L., H.Y.E., and A.A.C. Validation: G.L., M.G., and E.V. Formal analysis: G.L., M.G., E.V., D.L., H.Y.E., A.A.C., H.L.W., B.A.H., and S.J.A. Investigation: G.L., M.G., E.V., V.D., E.L., I.D., R.H.B., D.L., H.Y.E., A.A.C., S.W., A.T.-B., T.H., J.L., M.-H.L., M.G., M.M., I.N.-M., T.G., R.K.J., and S.J.A. Resources: D.L., B.A.H., K.C., J.A.K.M., and S.J.A. Resources (access to samples): T.H., A.S., K.G., M.R.W., A.F.D.N., E.H.L., T.S.-E., W.S.M., S.Z., D.W., G.S., R.O.-F., C.V.F.C., A.I., J.H.E., W.M., C.K.J., S.J.A., and J.F.R.S. Data curation: G.L., M.G., E.V., A.A.C., M.A., H.L.W., and S.J.A. Writing – Original draft: G.L., D.L., H.Y.E., A.A.C., K.C., and S.J.A. Writing – review & Editing: all authors. Visualization: G.L., M.G., H.Y.E., and A.A.C. Supervision: G.L., B.A.H., J.A.K.M., R.K.J., K.C., and S.J.A. Funding acquisition: K.C., J.A.K.M., and S.J.A. Project administration, G.L., R.K.J., K.C., and S.J.A.

DECLARATION OF INTERESTS

D.L. is employed by Schrodinger Inc. and holds company stocks. K.C. holds shares in Integrum Scientific, LLC, and Eitr Biologics Inc.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.chom.2024.12.016.

Received: August 30, 2024
Revised: November 6, 2024
Accepted: December 11, 2024
Published: January 15, 2025

REFERENCES

1. Feldmann, H., and Geisbert, T.W. (2011). Ebola haemorrhagic fever. Lancet *377*, 849–862. https://doi.org/10.1016/S0140-6736(10)60667-8.
2. World Health Organization. (2016). Ebola outbreak, West Africa, March 2014-2016. https://www.who.int/emergencies/situations/ebola-outbreak-2014-2016-West-Africa.
3. Leroy, E.M., Kumulungui, B., Pourrut, X., Rouquet, P., Hassanin, A., Yaba, P., Délicat, A., Paweska, J.T., Gonzalez, J.P., and Swanepoel, R. (2005). Fruit bats as reservoirs of Ebola virus. Nature *438*, 575–576. https://doi.org/10.1038/438575a.
4. Hayman, D.T.S., Yu, M., Crameri, G., Wang, L.F., Suu-Ire, R., Wood, J.L.N., and Cunningham, A.A. (2012). Ebola virus antibodies in fruit bats, Ghana, West Africa. Emerg. Infect. Dis. *18*, 1207–1209. https://doi.org/10.3201/eid1807.111654.
5. Laing, E.D., Mendenhall, I.H., Linster, M., Low, D.H.W., Chen, Y., Yan, L., Sterling, S.L., Borthwick, S., Neves, E.S., Lim, J.S.L., et al. (2018). Serologic Evidence of Fruit Bat Exposure to Filoviruses, Singapore, 2011-2016. Emerg. Infect. Dis. *24*, 114–117. https://doi.org/10.3201/eid2401.170401.
6. Ogawa, H., Miyamoto, H., Nakayama, E., Yoshida, R., Nakamura, I., Sawa, H., Ishii, A., Thomas, Y., Nakagawa, E., Matsuno, K., et al. (2015). Seroepidemiological Prevalence of Multiple Species of Filoviruses in Fruit Bats (Eidolon helvum) Migrating in Africa. J. Infect. Dis. *212* (*Suppl 2*), S101–S108. https://doi.org/10.1093/infdis/jiv063.
7. Yuan, J., Zhang, Y., Li, J., Zhang, Y., Wang, L.F., and Shi, Z. (2012). Serological evidence of ebolavirus infection in bats, China. Virol. J. *9*, 236. https://doi.org/10.1186/1743-422X-9-236.
8. Swanepoel, R., Leman, P.A., Burt, F.J., Zachariades, N.A., Braack, L.E., Ksiazek, T.G., Rollin, P.E., Zaki, S.R., and Peters, C.J. (1996). Experimental inoculation of plants and animals with Ebola virus. Emerg. Infect. Dis. *2*, 321–325. https://doi.org/10.3201/eid0204.960407.
9. Leroy, E.M., Epelboin, A., Mondonge, V., Pourrut, X., Gonzalez, J.P., Muyembe-Tamfum, J.J., and Formenty, P. (2009). Human Ebola outbreak resulting from direct exposure to fruit bats in Luebo,

Democratic Republic of Congo, 2007. Vector Borne Zoonotic Dis. *9*, 723–728. https://doi.org/10.1089/vbz.2008.0167.

10. Alexander, K.A., Sanderson, C.E., Marathe, M., Lewis, B.L., Rivers, C.M., Shaman, J., Drake, J.M., Lofgren, E., Dato, V.M., Eisenberg, M.C., et al. (2015). What factors might have led to the emergence of Ebola in West Africa? PLoS Negl. Trop. Dis. *9*, e0003652. https://doi.org/10.1371/journal.pntd.0003652.
11. Peterson, A.T., Bauer, J.T., and Mills, J.N. (2004). Ecologic and geographic distribution of filovirus disease. Emerg. Infect. Dis. *10*, 40–47. https://doi.org/10.3201/eid1001.030125.
12. Pigott, D.M., Golding, N., Mylne, A., Huang, Z., Henry, A.J., Weiss, D.J., Brady, O.J., Kraemer, M.U.G., Smith, D.L., Moyes, C.L., et al. (2014). Mapping the zoonotic niche of Ebola virus disease in Africa. eLife *3*, e04395. https://doi.org/10.7554/eLife.04395.
13. Towner, J.S., Amman, B.R., Sealy, T.K., Carroll, S.A.R., Comer, J.A., Kemp, A., Swanepoel, R., Paddock, C.D., Balinandi, S., Khristova, M.L., et al. (2009). Isolation of genetically diverse Marburg viruses from Egyptian fruit bats. PLoS Pathog. *5*, e1000536. https://doi.org/10.1371/journal.ppat.1000536.
14. Kemenesi, G., Tóth, G.E., Mayora-Neto, M., Scott, S., Temperton, N., Wright, E., Mühlberger, E., Hume, A.J., Suder, E.L., Zana, B., et al. (2022). Isolation of infectious Lloviu virus from Schreiber's bats in Hungary. Nat. Commun. *13*, 1706. https://doi.org/10.1038/s41467-022-29298-1.
15. Tóth, G.E., Hume, A.J., Suder, E.L., Zeghbib, S., Ábrahám, Á., Lanszki, Z., Varga, Z., Tauber, Z., Földes, F., Zana, B., et al. (2023). Isolation and genome characterization of Lloviu virus from Italian Schreibers's bats. Sci. Rep. *13*, 11310. https://doi.org/10.1038/s41598-023-38364-7.
16. Goldstein, T., Anthony, S.J., Gbakima, A., Bird, B.H., Bangura, J., Tremeau-Bravard, A., Belaganahalli, M.N., Wells, H.L., Dhanota, J.K., Liang, E., et al. (2018). The discovery of Bombali virus adds further support for bats as hosts of ebolaviruses. Nat. Microbiol. *3*, 1084–1089. https://doi.org/10.1038/s41564-018-0227-2.
17. Karan, L.S., Makenov, M.T., Korneev, M.G., Sacko, N., Boumbaly, S., Yakovlev, S.A., Kourouma, K., Bayandin, R.B., Gladysheva, A.V., Shipovalov, A.V., et al. (2019). Bombali Virus in Mops condylurus Bats, Guinea. Emerg. Infect. Dis. *25*, 1774–1775. https://doi.org/10.3201/eid2509.190581.
18. Forbes, K.M., Webala, P.W., Jääskeläinen, A.J., Abdurahman, S., Ogola, J., Masika, M.M., Kivistö, I., Alburkat, H., Plyusnin, I., Levanov, L., et al. (2019). Bombali Virus in Mops condylurus Bat, Kenya. Emerg. Infect. Dis. *25*, 955–957. https://doi.org/10.3201/eid2505.181666.
19. Kareinen, L., Ogola, J., Kivistö, I., Smura, T., Aaltonen, K., Jääskeläinen, A.J., Kibiwot, S., Masika, M.M., Nyaga, P., Mwaengo, D., et al. (2020). Range Expansion of Bombali Virus in Mops condylurus Bats, Kenya, 2019. Emerg. Infect. Dis. *26*, 3007–3010. https://doi.org/10.3201/eid2612.202925.
20. Lebarbenchon, C., Goodman, S.M., Hoarau, A.O.G., Le Minter, G., Dos Santos, A., Schoeman, M.C., Léculier, C., Raoul, H., Gudo, E.S., and Mavingui, P. (2022). Bombali Ebolavirus in Mops condylurus bats (Molossidae), Mozambique. Emerg. Infect. Dis. *28*, 2583–2585. https://doi.org/10.3201/eid2812.220853.
21. Yang, X.-L., Tan, C.W., Anderson, D.E., Jiang, R.-D., Li, B., Zhang, W., Zhu, Y., Lim, X.F., Zhou, P., Liu, X.-L., et al. (2019). Characterization of a filovirus (Měnglà virus) from Rousettus bats in China. Nat. Microbiol. *4*, 390–395. https://doi.org/10.1038/s41564-018-0328-y.
22. Xie, S.Z., Yao, K., Li, B., Peng, C., Yang, X.L., and Shi, Z.L. (2024). Development of a Měnglà virus minigenome and comparison of its polymerase complexes with those of other filoviruses. Virol. Sin. *39*, 459–468. https://doi.org/10.1016/j.virs.2024.03.011.
23. He, B., Hu, T., Yan, X., Pa, Y., Liu, Y., Liu, Y., Li, N., Yu, J., Zhang, H., Liu, Y., et al. (2024). Isolation, characterization, and circulation sphere of a filovirus in fruit bats. Proc. Natl. Acad. Sci. USA *121*, e2313789121. https://doi.org/10.1073/pnas.2313789121.
24. Jayme, S.I., Field, H.E., de Jong, C., Olival, K.J., Marsh, G., Tagtag, A.M., Hughes, T., Bucad, A.C., Barr, J., Azul, R.R., et al. (2015). Molecular evidence of Ebola Reston virus infection in Philippine bats. Virol. J. *12*, 107. https://doi.org/10.1186/s12985-015-0331-3.
25. Brook, C.E., Ranaivoson, H.C., Broder, C.C., Cunningham, A.A., Héraud, J.M., Peel, A.J., Gibson, L., Wood, J.L.N., Metcalf, C.J., and Dobson, A.P. (2019). Disentangling serology to elucidate henipa- and filovirus transmission in Madagascar fruit bats. J. Anim. Ecol. *88*, 1001–1016. https://doi.org/10.1111/1365-2656.12985.
26. MBA Djomsi, D.M., Mba Djonzo, F.A., Ndong Bass, I., Champagne, M., Lacroix, A., Thaurignac, G., Esteban, A., De Nys, H., Bourgarel, M., Akoachere, J.F., et al. (2022). Dynamics of Antibodies to Ebolaviruses in an Eidolon helvum Bat Colony in Cameroon. Viruses *14*, 560. https://doi.org/10.3390/v14030560.
27. Pourrut, X., Souris, M., Towner, J.S., Rollin, P.E., Nichol, S.T., Gonzalez, J.P., and Leroy, E. (2009). Large serological survey showing cocirculation of Ebola and Marburg viruses in Gabonese bat populations, and a high seroprevalence of both viruses in Rousettus aegyptiacus. BMC Infect. Dis. *9*, 159. https://doi.org/10.1186/1471-2334-9-159.
28. Hayman, D.T.S., Emmerich, P., Yu, M., Wang, L.F., Suu-Ire, R., Fooks, A.R., Cunningham, A.A., and Wood, J.L.N. (2010). Long-term survival of an urban fruit bat seropositive for Ebola and Lagos bat viruses. PLoS One *5*, e11978. https://doi.org/10.1371/journal.pone.0011978.
29. Pourrut, X., Délicat, A., Rollin, P.E., Ksiazek, T.G., Gonzalez, J.P., and Leroy, E.M. (2007). Spatial and temporal patterns of Zaire ebolavirus antibody prevalence in the possible reservoir bat species. J. Infect. Dis. *196* (*Suppl 2*), S176–S183. https://doi.org/10.1086/520541.
30. De Nys, H.M., Kingebeni, P.M., Keita, A.K., Butel, C., Thaurignac, G., Villabona-Arenas, C.J., Lemarcis, T., Geraerts, M., Vidal, N., Esteban, A., et al. (2018). Survey of Ebola Viruses in Frugivorous and Insectivorous Bats in Guinea, Cameroon, and the Democratic Republic of the Congo, 2015-2017. Emerg. Infect. Dis. *24*, 2228–2240. https://doi.org/10.3201/eid2412.180740.
31. Seifert, S.N., Fischer, R.J., Kuisma, E., Badzi Nkoua, C., Bounga, G., Akongo, M.J., Schulz, J.E., Escudero-Pérez, B., Akoundzie, B.J., Ampiri, V.R.B., et al. (2022). Zaire ebolavirus surveillance near the Bikoro region of the Democratic Republic of the Congo during the 2018 outbreak reveals presence of seropositive bats. PLoS Negl. Trop. Dis. *16*, e0010504. https://doi.org/10.1371/journal.pntd.0010504.
32. Leroy, E.M., Rouquet, P., Formenty, P., Souquière, S., Kilbourne, A., Froment, J.M., Bermejo, M., Smit, S., Karesh, W., Swanepoel, R., et al. (2004). Multiple Ebola virus transmission events and rapid decline of central African wildlife. Science *303*, 387–390. https://doi.org/10.1126/science.1092528.
33. Rouquet, P., Froment, J.M., Bermejo, M., Kilbourn, A., Karesh, W., Reed, P., Kumulungui, B., Yaba, P., Délicat, A., Rollin, P.E., et al. (2005). Wild animal mortality monitoring and human Ebola outbreaks, Gabon and Republic of Congo, 2001-2003. Emerg. Infect. Dis. *11*, 283–290. https://doi.org/10.3201/eid1102.040533.
34. Leroy, E.M., Telfer, P., Kumulungui, B., Yaba, P., Rouquet, P., Roques, P., Gonzalez, J.P., Ksiazek, T.G., Rollin, P.E., and Nerrienet, E. (2004). A serological survey of Ebola virus infection in central African nonhuman primates. J. Infect. Dis. *190*, 1895–1899. https://doi.org/10.1086/425421.
35. Nidom, C.A., Nakayama, E., Nidom, R.V., Alamudi, M.Y., Daulay, S., Dharmayanti, I.N.L.P., Dachlan, Y.P., Amin, M., Igarashi, M., Miyamoto, H., et al. (2012). Serological evidence of Ebola virus infection in Indonesian orangutans. PLoS One *7*, e40740. https://doi.org/10.1371/journal.pone.0040740.
36. Rollin, P.E., Williams, R.J., Bressler, D.S., Pearson, S., Cottingham, M., Pucak, G., Sanchez, A., Trappier, S.G., Peters, R.L., Greer, P.W., et al. (1999). Ebola (subtype Reston) virus among quarantined nonhuman primates recently imported from the Philippines to the United States. J. Infect. Dis. *179* (*Suppl 1*), S108–S114. https://doi.org/10.1086/514303.

37. Miranda, M.E.G., and Miranda, N.L.J. (2011). Reston ebolavirus in humans and animals in the Philippines: a review. J. Infect. Dis. *204* (*Suppl 3*), S757–S760. https://doi.org/10.1093/infdis/jir296.

38. Bermejo, M., Rodríguez-Teijeiro, J.D., Illera, G., Barroso, A., Vilà, C., and Walsh, P.D. (2006). Ebola outbreak killed 5000 gorillas. Science *314*, 1564. https://doi.org/10.1126/science.1133105.

39. Sayers, R.D., Watt, P.A., Muller, S., Bell, P.R., and Thurston, H. (1991). Structural and functional smooth muscle injury after surgical preparation of reversed and non-reversed (in situ) saphenous vein bypass grafts. Br. J. Surg. *78*, 1256–1258. https://doi.org/10.1002/bjs.1800781035.

40. Sayama, Y., Demetria, C., Saito, M., Azul, R.R., Taniguchi, S., Fukushi, S., Yoshikawa, T., Iizuka, I., Mizutani, T., Kurane, I., et al. (2012). A seroepidemiologic study of Reston ebolavirus in swine in the Philippines. BMC Vet. Res. *8*, 82. https://doi.org/10.1186/1746-6148-8-82.

41. Weingartl, H.M., Embury-Hyatt, C., Nfon, C., Leung, A., Smith, G., and Kobinger, G. (2012). Transmission of Ebola virus from pigs to non-human primates. Sci. Rep. *2*, 811. https://doi.org/10.1038/srep00811.

42. Fischer, K., Jabaty, J., Suluku, R., Strecker, T., Groseth, A., Fehling, S.K., Balkema-Buschmann, A., Koroma, B., Schmidt, K.M., Atherstone, C., et al. (2018). Serological Evidence for the Circulation of Ebolaviruses in Pigs From Sierra Leone. J. Infect. Dis. *218* (*Suppl 5*), S305–S311. https://doi.org/10.1093/infdis/jiy330.

43. Morvan, J.M., Deubel, V., Gounon, P., Nakouné, E., Barrière, P., Murri, S., Perpète, O., Selekon, B., Coudrier, D., Gautier-Hion, A., et al. (1999). Identification of Ebola virus sequences present as RNA or DNA in organs of terrestrial small mammals of the Central African Republic. Microbes Infect. *1*, 1193–1201. https://doi.org/10.1016/s1286-4579(99)00242-7.

44. Han, B.A., Schmidt, J.P., Alexander, L.W., Bowden, S.E., Hayman, D.T.S., and Drake, J.M. (2016). Undiscovered Bat Hosts of Filoviruses. PLoS Negl. Trop. Dis. *10*, e0004815. https://doi.org/10.1371/journal.pntd.0004815.

45. Nyakarahuka, L., Ayebare, S., Mosomtai, G., Kankya, C., Lutwama, J., Mwiine, F.N., and Skjerve, E. (2017). Ecological Niche Modeling for Filoviruses: A Risk Map for Ebola and Marburg Virus Disease Outbreaks in Uganda. PLoS Curr. *9*, ecurrents.outbreaks.07992a87522e1f229c7cb023270a2af1. https://doi.org/10.1371/currents.outbreaks.07992a87522e1f229c7cb023270a2af1.

46. Yang, X.L., Zhang, Y.Z., Jiang, R.D., Guo, H., Zhang, W., Li, B., Wang, N., Wang, L., Waruhiu, C., Zhou, J.H., et al. (2017). Genetically Diverse Filoviruses in Rousettus and Eonycteris spp. Bats, China, 2009 and 2015. Emerg. Infect. Dis. *23*, 482–486. https://doi.org/10.3201/eid2303.161119.

47. Ng, M., Ndungo, E., Kaczmarek, M.E., Herbert, A.S., Binger, T., Kuehne, A.I., Jangra, R.K., Hawkins, J.A., Gifford, R.J., Biswas, R., et al. (2015). Filovirus receptor NPC1 contributes to species-specific patterns of ebolavirus susceptibility in bats. eLife *4*, e11785. https://doi.org/10.7554/eLife.11785.

48. Carette, J.E., Raaben, M., Wong, A.C., Herbert, A.S., Obernosterer, G., Mulherkar, N., Kuehne, A.I., Kranzusch, P.J., Griffin, A.M., Ruthel, G., et al. (2011). Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. Nature *477*, 340–343. https://doi.org/10.1038/nature10348.

49. Côté, M., Misasi, J., Ren, T., Bruchez, A., Lee, K., Filone, C.M., Hensley, L., Li, Q., Ory, D., Chandran, K., et al. (2011). Small molecule inhibitors reveal Niemann-Pick C1 is essential for Ebola virus infection. Nature *477*, 344–348. https://doi.org/10.1038/nature10380.

50. Wang, H., Shi, Y., Song, J., Qi, J., Lu, G., Yan, J., and Gao, G.F. (2016). Ebola Viral Glycoprotein Bound to Its Endosomal Receptor Niemann-Pick C1. Cell *164*, 258–268. https://doi.org/10.1016/j.cell.2015.12.044.

51. Martinez, O., Ndungo, E., Tantral, L., Miller, E.H., Leung, L.W., Chandran, K., and Basler, C.F. (2013). A Mutation in the Ebola Virus Envelope Glycoprotein Restricts Viral Entry in a Host Species- and Cell-Type-Specific Manner. J. Virol. *87*, 3324–3334. https://doi.org/10.1128/JVI.01598-12.

52. Ndungo, E., Herbert, A.S., Raaben, M., Obernosterer, G., Biswas, R., Miller, E.H., Wirchnianski, A.S., Carette, J.E., Brummelkamp, T.R., Whelan, S.P., et al. (2016). A Single Residue in Ebola Virus Receptor NPC1 Influences Cellular Host Range in Reptiles. mSphere *1*, e00007-16. https://doi.org/10.1128/mSphere.00007-16.

53. Urbanowicz, R.A., McClure, C.P., Sakuntabhai, A., Sall, A.A., Kobinger, G., Müller, M.A., Holmes, E.C., Rey, F.A., Simon-Loriere, E., and Ball, J.K. (2016). Human Adaptation of Ebola Virus during the West African Outbreak. Cell *167*, 1079–1087.e5. https://doi.org/10.1016/j.cell.2016.10.013.

54. Takadate, Y., Manzoor, R., Saito, T., Kida, Y., Maruyama, J., Kondoh, T., Miyamoto, H., Ogawa, H., Kajihara, M., Igarashi, M., et al. (2020). Receptor-Mediated Host Cell Preference of a Bat-Derived Filovirus, Lloviu Virus. Microorganisms *8*, 1530. https://doi.org/10.3390/microorganisms8101530.

55. Takadate, Y., Kondoh, T., Igarashi, M., Maruyama, J., Manzoor, R., Ogawa, H., Kajihara, M., Furuyama, W., Sato, M., Miyamoto, H., et al. (2020). Niemann-Pick C1 Heterogeneity of Bat Cells Controls Filovirus Tropism. Cell Rep. *30*, 308–319.e5. https://doi.org/10.1016/j.celrep.2019.12.042.

56. Schornberg, K., Matsuyama, S., Kabsch, K., Delos, S., Bouton, A., and White, J. (2006). Role of endosomal cathepsins in entry mediated by the Ebola virus glycoprotein. J. Virol. *80*, 4174–4178. https://doi.org/10.1128/JVI.80.8.4174-4178.2006.

57. Wong, A.C., Sandesara, R.G., Mulherkar, N., Whelan, S.P., and Chandran, K. (2010). A forward genetic strategy reveals destabilizing mutations in the Ebolavirus glycoprotein that alter its protease dependence during cell entry. J. Virol. *84*, 163–175. https://doi.org/10.1128/JVI.01832-09.

58. Sundaram, M., Schmidt, J.P., Han, B.A., Drake, J.M., and Stephens, P.R. (2022). Traits, phylogeny and host cell receptors predict Ebolavirus host status among African mammals. PLoS Negl. Trop. Dis. *16*, e0010993. https://doi.org/10.1371/journal.pntd.0010993.

59. Kyte, J., and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. *157*, 105–132. https://doi.org/10.1016/0022-2836(82)90515-0.

60. Deléage, G., and Roux, B. (1987). An algorithm for protein secondary structure prediction based on class prediction. Protein Eng. *1*, 289–294. https://doi.org/10.1093/protein/1.4.289.

61. Bhaskaran, R., and Ponnuswamy, P.K. (1984). Dynamics of amino acid residues in globular proteins. Int. J. Pept. Protein Res. *24*, 180–191. https://doi.org/10.1111/j.1399-3011.1984.tb00944.x.

62. Tien, M.Z., Meyer, A.G., Sydykova, D.K., Spielman, S.J., and Wilke, C.O. (2013). Maximum allowed solvent accessibilites of residues in proteins. PLoS One *8*, e80635. https://doi.org/10.1371/journal.pone.0080635.

63. Hwang, H., Petrey, D., and Honig, B. (2016). A hybrid method for protein-protein interface prediction. Protein Sci. *25*, 159–165. https://doi.org/10.1002/pro.2744.

64. Zimmerman, J.M., Eliezer, N., and Simha, R. (1968). The characterization of amino acid sequences in proteins by statistical methods. J. Theor. Biol. *21*, 170–201. https://doi.org/10.1016/0022-5193(68)90069-6.

65. Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S.I. (2020). From Local Explanations to Global Understanding with Explainable AI for Trees. Nat. Mach. Intell. *2*, 56–67. https://doi.org/10.1038/s42256-019-0138-9.

66. Lundberg, S.M., and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems (Curran Associates Inc).

67. Zhao, Y., Ren, J., Harlos, K., and Stuart, D.I. (2016). Structure of glycosylated NPC1 luminal domain C reveals insights into NPC2 and Ebola virus interactions. FEBS Lett. *590*, 605–612. https://doi.org/10.1002/1873-3468.12089.

68. Wang, J., Manicassamy, B., Caffrey, M., and Rong, L. (2011). Characterization of the receptor-binding domain of Ebola glycoprotein

in viral entry. Virol. Sin. *26*, 156–170. https://doi.org/10.1007/s12250-011-3194-9.

69. Bornholdt, Z.A., Ndungo, E., Fusco, M.L., Bale, S., Flyak, A.I., Crowe, J.E., Jr., Chandran, K., and Saphire, E.O. (2016). Host-Primed Ebola Virus GP Exposes a Hydrophobic NPC1 Receptor-Binding Pocket, Revealing a Target for Broadly Neutralizing Antibodies. mBio *7*, e02154-15. https://doi.org/10.1128/mBio.02154-15.
70. Dube, D., Brecher, M.B., Delos, S.E., Rose, S.C., Park, E.W., Schornberg, K.L., Kuhn, J.H., and White, J.M. (2009). The Primed Ebolavirus Glycoprotein (19-Kilodalton $GP_{1,2}$): Sequence and Residues Critical for Host Cell Binding. J. Virol. *83*, 2883–2891. https://doi.org/10.1128/JVI.01956-08.
71. Chipot, C., and Pohorille, A. (2007). Calculating Free Energy Differences Using Perturbation Theory. In Free Energy Calculations: Theory and Applications in Chemistry and Biology, C. Chipot and A. Pohorille, eds. (Springer), pp. 33–75. https://doi.org/10.1007/978-3-540-38448-9_2.
72. Bentley, E.M., Richardson, S., Derveni, M., Rijal, P., Townsend, A.R., Heeney, J.L., Mattiuzzo, G., and Wright, E. (2021). Cross-Neutralisation of Novel Bombali Virus by Ebola Virus Antibodies and Convalescent Plasma Using an Optimised Pseudotype-Based Neutralisation Assay. Trop. Med. Infect. Dis. *6*, 155. https://doi.org/10.3390/tropicalmed6030155.
73. Li, L., Gan, T., Ma, Z., Huang, Y., and Zhong, J. (2024). Assessing risk of Bombali virus spillover to humans by mutagenesis analysis of viral glycoprotein. hLife *2*, 32–42. https://doi.org/10.1016/j.hlife.2023.11.003.
74. Igarashi, M., Hirokawa, T., Takadate, Y., and Takada, A. (2021). Structural Insights into the Interaction of Filovirus Glycoproteins with the Endosomal Receptor Niemann-Pick C1: A Computational Study. Viruses *13*, 913. https://doi.org/10.3390/v13050913.
75. Sironi, M., Cagliani, R., Forni, D., and Clerici, M. (2015). Evolutionary insights into host-pathogen interactions from mammalian sequence data. Nat. Rev. Genet. *16*, 224–236. https://doi.org/10.1038/nrg3905.
76. Li, X., Saha, P., Li, J., Blobel, G., and Pfeffer, S.R. (2016). Clues to the mechanism of cholesterol transfer from the structure of NPC1 middle lumenal domain bound to NPC2. Proc. Natl. Acad. Sci. USA *113*, 10079–10084. https://doi.org/10.1073/pnas.1611956113.
77. Murrell, B., Wertheim, J.O., Moola, S., Weighill, T., Scheffler, K., and Kosakovsky Pond, S.L. (2012). Detecting Individual Sites Subject to Episodic Diversifying Selection. PLoS Genet. *8*, e1002764. https://doi.org/10.1371/journal.pgen.1002764.
78. Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S.L., and Scheffler, K. (2013). FUBAR: a fast, unconstrained bayesian approximation for inferring selection. Mol. Biol. Evol. *30*, 1196–1205. https://doi.org/10.1093/molbev/mst030.
79. Marsh, C.J., Sica, Y.V., Burgin, C.J., Dorman, W.A., Anderson, R.C., del Toro Mijares, I., Vigneron, J.G., Barve, V., Dombrowik, V.L., Duong, M., et al. (2022). Expert range maps of global mammal distributions harmonised to three taxonomic authorities. J. Biogeogr. *49*, 979–992. https://doi.org/10.1111/jbi.14330.
80. Mammal Diversity Database (2024). Mammal Diversity Database. Version v1.12.1. Zenodo. https://doi.org/10.5281/zenodo.10595931.
81. PREDICT. (2020). PREDICT. https://ohi.vetmed.ucdavis.edu/programs-projects/predict-project.
82. Cakenberghe, V.V., and Seamark, E.C.J. (2022). African Chiroptera Report 2022. www.africanbats.org.
83. CDC. (2024). Ebola Disease Outbreaks by Species and Size, Since 1976. https://www.cdc.gov/ebola/outbreaks/index.html.
84. Marí Saéz, A., Weiss, S., Nowak, K., Lapeyre, V., Zimmermann, F., Düx, A., Kühl, H.S., Kaba, M., Regnaut, S., Merkel, K., et al. (2015). Investigating the zoonotic origin of the West African Ebola epidemic. EMBO Mol. Med. *7*, 17–23. https://doi.org/10.15252/emmm.201404792.
85. Barr, J., Boyd, V., Todd, S., Smith, I., Prada, D., O'Dea, M., Jackson, B., Pearce, L., Adams, T.E., Vanderduys, E., et al. (2022). Detection of filovirus-reactive antibodies in Australian bat species. J. Gen. Virol. *103*. https://doi.org/10.1099/jgv.0.001785.
86. Suu-Ire, R., Obodai, E., Bel-Nono, S.O., Ampofo, W.K., Mazet, J.A.K., Goldstein, T., Johnson, C.K., Smith, B., Boaatema, L., Asigbee, T.W., et al. (2022). Surveillance for potentially zoonotic viruses in rodent and bat populations and behavioral risk in an agricultural settlement in Ghana. One Health Outlook *4*, 6. https://doi.org/10.1186/s42522-022-00061-2.
87. Towner, J.S., Pourrut, X., Albariño, C.G., Nkogue, C.N., Bird, B.H., Grard, G., Ksiazek, T.G., Gonzalez, J.P., Nichol, S.T., and Leroy, E.M. (2007). Marburg virus infection detected in a common African bat. PLoS One *2*, e764. https://doi.org/10.1371/journal.pone.0000764.
88. Kuzmin, I.V., Niezgoda, M., Franka, R., Agwanda, B., Markotter, W., Breiman, R.F., Shieh, W.J., Zaki, S.R., and Rupprecht, C.E. (2010). Marburg virus in fruit bat, Kenya. Emerg. Infect. Dis. *16*, 352–354. https://doi.org/10.3201/eid1602.091269.
89. Taniguchi, S., Watanabe, S., Masangkay, J.S., Omatsu, T., Ikegami, T., Alviola, P., Ueda, N., Iha, K., Fujii, H., Ishii, Y., et al. (2011). Reston Ebolavirus antibodies in bats, the Philippines. Emerg. Infect. Dis. *17*, 1559–1560. https://doi.org/10.3201/eid1708.101693.
90. Amman, B.R., Carroll, S.A., Reed, Z.D., Sealy, T.K., Balinandi, S., Swanepoel, R., Kemp, A., Erickson, B.R., Comer, J.A., Campbell, S., et al. (2012). Seasonal pulses of Marburg virus circulation in juvenile Rousettus aegyptiacus bats coincide with periods of increased risk of human infection. PLoS Pathog. *8*, e1002877. https://doi.org/10.1371/journal.ppat.1002877.
91. Olival, K.J., Islam, A., Yu, M., Anthony, S.J., Epstein, J.H., Khan, S.A., Khan, S.U., Crameri, G., Wang, L.F., Lipkin, W.I., et al. (2013). Ebola virus antibodies in fruit bats, bangladesh. Emerg. Infect. Dis. *19*, 270–273. https://doi.org/10.3201/eid1902.120524.
92. Amman, B.R., Nyakarahuka, L., McElroy, A.K., Dodd, K.A., Sealy, T.K., Schuh, A.J., Shoemaker, T.R., Balinandi, S., Atimnedi, P., Kaboyo, W., et al. (2014). Marburgvirus resurgence in Kitaka Mine bat population after extermination attempts, Uganda. Emerg. Infect. Dis. *20*, 1761–1764. https://doi.org/10.3201/eid2010.140696.
93. He, B., Feng, Y., Zhang, H., Xu, L., Yang, W., Zhang, Y., Li, X., and Tu, C. (2015). Filovirus RNA in Fruit Bats, China. Emerg. Infect. Dis. *21*, 1675–1677. https://doi.org/10.3201/eid2109.150260.
94. Pawęska, J.T., Jansen van Vuren, P., Kemp, A., Storm, N., Grobbelaar, A.A., Wiley, M.R., Palacios, G., and Markotter, W. (2018). Marburg Virus Infection in Egyptian Rousette Bats, South Africa, 2013-2014[1]. Emerg. Infect. Dis. *24*, 1134–1137. https://doi.org/10.3201/eid2406.172165.
95. Changula, K., Kajihara, M., Mori-Kajihara, A., Eto, Y., Miyamoto, H., Yoshida, R., Shigeno, A., Hang'ombe, B., Qiu, Y., Mwizabi, D., et al. (2018). Seroprevalence of Filovirus Infection of Rousettus aegyptiacus Bats in Zambia. J. Infect. Dis. *218* (*Suppl 5*), S312–S317. https://doi.org/10.1093/infdis/jiy266.
96. Kajihara, M., Hang'ombe, B.M., Changula, K., Harima, H., Isono, M., Okuya, K., Yoshida, R., Mori-Kajihara, A., Eto, Y., Orba, Y., et al. (2019). Marburgvirus in Egyptian Fruit Bats, Zambia. Emerg. Infect. Dis. *25*, 1577–1580. https://doi.org/10.3201/eid2508.190268.
97. Dovih, P., Laing, E.D., Chen, Y., Low, D.H.W., Ansil, B.R., Yang, X., Shi, Z., Broder, C.C., Smith, G.J.D., Linster, M., et al. (2019). Filovirus-reactive antibodies in humans and bats in Northeast India imply zoonotic spillover. PLoS Negl. Trop. Dis. *13*, e0007733. https://doi.org/10.1371/journal.pntd.0007733.
98. Zhang, C., Wang, Z., Cai, J., Yan, X., Zhang, F., Wu, J., Xu, L., Zhao, Z., Hu, T., Tu, C., et al. (2020). Seroreactive Profiling of Filoviruses in Chinese Bats Reveals Extensive Infection of Diverse Viruses. J. Virol. *94*, e02042-19. https://doi.org/10.1128/JVI.02042-19.
99. Amman, B.R., Bird, B.H., Bakarr, I.A., Bangura, J., Schuh, A.J., Johnny, J., Sealy, T.K., Conteh, I., Koroma, A.H., Foday, I., et al. (2020). Isolation of Angola-like Marburg virus from Egyptian rousette bats from West

Africa. Nat. Commun. *11*, 510. https://doi.org/10.1038/s41467-020-14327-8.

100. Schulz, J.E., Seifert, S.N., Thompson, J.T., Avanzato, V., Sterling, S.L., Yan, L., Letko, M.C., Matson, M.J., Fischer, R.J., Tremeau-Bravard, A., et al. (2020). Serological Evidence for Henipa-like and Filo-like Viruses in Trinidad Bats. J. Infect. Dis. *221* (*Suppl 4*), S375–S382. https://doi.org/10.1093/infdis/jiz648.

101. Lacroix, A., Mbala Kingebeni, P., Ndimbo Kumugo, S.P., Lempu, G., Butel, C., Serrano, L., Vidal, N., Thaurignac, G., Esteban, A., Mukadi Bamuleka, D., et al. (2021). Investigating the Circulation of Ebola Viruses in Bats during the Ebola Virus Disease Outbreaks in the Equateur and North Kivu Provinces of the Democratic Republic of Congo from 2018. Pathogens *10*, 557. https://doi.org/10.3390/pathogens10050557.

102. Zhao, K., Zhang, W., Li, B., Xie, S.Z., Yi, F., Jiang, R.D., Luo, Y., He, X.Y., Zhang, Y.Z., Shi, Z.L., et al. (2022). Ecological study of cave nectar bats reveals low risk of direct transmission of bat viruses to humans. Zool. Res. *43*, 514–522. https://doi.org/10.24272/j.issn.2095-8137.2021.480.

103. Tian, J., Sun, J., Li, D., Wang, N., Wang, L., Zhang, C., Meng, X., Ji, X., Suchard, M.A., Zhang, X., et al. (2022). Emerging viruses: Cross-species transmission of coronaviruses, filoviruses, henipaviruses, and rotaviruses from bats. Cell Rep. *39*, 110969. https://doi.org/10.1016/j.celrep.2022.110969.

104. Kainulainen, M.H., Harmon, J.R., Karaaslan, E., Kyondo, J., Whitesell, A., Twongyeirwe, S., Malenfant, J.H., Baluku, J., Kofman, A., Bergeron, É., et al. (2024). A public, cross-reactive glycoprotein epitope confounds Ebola virus serology. J. Med. Virol. *96*, e29946. https://doi.org/10.1002/jmv.29946.

105. Formenty, P., Hatz, C., Le Guenno, B., Stoll, A., Rogenmoser, P., and Widmer, A. (1999). Human infection due to Ebola virus, subtype Côte d'Ivoire: clinical and biologic presentation. J. Infect. Dis. *179* (*Suppl 1*), S48–S53. https://doi.org/10.1086/514285.

106. Babayan, S.A., Orton, R.J., and Streicker, D.G. (2018). Predicting reservoir hosts and arthropod vectors from evolutionary signatures in RNA virus genomes. Science *362*, 577–580. https://doi.org/10.1126/science.aap9072.

107. Luo, W., Roy, A., Guo, F., Irwin, D.M., Shen, X., Pan, J., and Shen, Y. (2020). Host Adaptation and Evolutionary Analysis of Zaire ebolavirus: Insights From Codon Usage Based Investigations. Front. Microbiol. *11*, 570131. https://doi.org/10.3389/fmicb.2020.570131.

108. Olivero, J., Fa, J.E., Real, R., Farfán, M.Á., Márquez, A.L., Vargas, J.M., Gonzalez, J.P., Cunningham, A.A., and Nasi, R. (2017). Mammalian biogeography and the Ebola virus in Africa. Mamm. Rev. *47*, 24–37. https://doi.org/10.1111/mam.12074.

109. Fels, J.M., Bortz, R.H., 3rd, Alkutkar, T., Mittler, E., Jangra, R.K., Spence, J.S., and Chandran, K. (2021). A Glycoprotein Mutation That Emerged during the 2013-2016 Ebola Virus Epidemic Alters Proteolysis and Accelerates Membrane Fusion. mBio *12*, e03616-20. https://doi.org/10.1128/mBio.03616-20.

110. Riesle-Sbarbaro, S.A., Wibbelt, G., Düx, A., Kouakou, V., Bokelmann, M., Hansen-Kant, K., Kirchoff, N., Laue, M., Kromarek, N., Lander, A., et al. (2024). Selective replication and vertical transmission of Ebola virus in experimentally infected Angolan free-tailed bats. Nat. Commun. *15*, 925. https://doi.org/10.1038/s41467-024-45231-0.

111. Keck, Z.-Y., Enterlein, S.G., Howell, K.A., Vu, H., Shulenin, S., Warfield, K.L., Froude, J.W., Araghi, N., Douglas, R., Biggins, J., et al. (2016). Macaque Monoclonal Antibodies Targeting Novel Conserved Epitopes within Filovirus Glycoprotein. J. Virol. *90*, 279–291. https://doi.org/10.1128/JVI.02172-15.

112. Wec, A.Z., Herbert, A.S., Murin, C.D., Nyakatura, E.K., Abelson, D.M., Fels, J.M., He, S., James, R.M., de La Vega, M.A., Zhu, W., et al. (2017). Antibodies from a Human Survivor Define Sites of Vulnerability for Broad Protection against Ebolaviruses. Cell *169*, 878–890.e15. https://doi.org/10.1016/j.cell.2017.04.037.

113. Flyak, A.I., Ilinykh, P.A., Murin, C.D., Garron, T., Shen, X., Fusco, M.L., Hashiguchi, T., Bornholdt, Z.A., Slaughter, J.C., Sapparapu, G., et al. (2015). Mechanism of human antibody-mediated neutralization of Marburg virus. Cell *160*, 893–903. https://doi.org/10.1016/j.cell.2015.01.031.

114. Wec, A.Z., Nyakatura, E.K., Herbert, A.S., Howell, K.A., Holtsberg, F.W., Bakken, R.R., Mittler, E., Christin, J.R., Shulenin, S., Jangra, R.K., et al. (2016). A "Trojan horse" bispecific-antibody strategy for broad protection against ebolaviruses. Science *354*, 350–354. https://doi.org/10.1126/science.aag3267.

115. Ng, M., Ndungo, E., Jangra, R.K., Cai, Y., Postnikova, E., Radoshitzky, S.R., Dye, J.M., Ramírez de Arellano, E., Negredo, A., Palacios, G., et al. (2014). Cell entry by a novel European filovirus requires host endosomal cysteine proteases and Niemann-Pick C1. Virology *468–470*, 637–646. https://doi.org/10.1016/j.virol.2014.08.019.

116. Wirchnianski, A.S., Wec, A.Z., Nyakatura, E.K., Herbert, A.S., Slough, M.M., Kuehne, A.I., Mittler, E., Jangra, R.K., Teruya, J., Dye, J.M., et al. (2021). Two Distinct Lysosomal Targeting Strategies Afford Trojan Horse Antibodies With Pan-Filovirus Activity. Front. Immunol. *12*, 729851. https://doi.org/10.3389/fimmu.2021.729851.

117. Misasi, J., Chandran, K., Yang, J.Y., Considine, B., Filone, C.M., Côté, M., Sullivan, N., Fabozzi, G., Hensley, L., and Cunningham, J. (2012). Filoviruses require endosomal cysteine proteases for entry but exhibit distinct protease preferences. J. Virol. *86*, 3284–3292. https://doi.org/10.1128/JVI.06346-11.

118. Spence, J.S., Krause, T.B., Mittler, E., Jangra, R.K., and Chandran, K. (2016). Direct Visualization of Ebola Virus Fusion Triggering in the Endocytic Pathway. mBio *7*, e01857-15. https://doi.org/10.1128/mBio.01857-15.

119. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2.

120. Waskom, M.L. (2021). seaborn: statistical data visualization. J. Open Source Software *6*, 3021. https://doi.org/10.21105/joss.03021.

121. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.

122. Raschka, S. (2018). MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack. J. Open Source Software *3*, 3. https://doi.org/10.21105/joss.00638.

123. R Core Team (2016). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).

124. Brooks, M.E., Kristensen, K., Van Benthem, K.J., Magnusson, A., Berg, C.W., Nielsen, A., Skaug, H.J., Mächler, M., and Bolker, B.M. (2017). glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. R J. *9*, 9. https://doi.org/10.32614/RJ-2017-066.

125. Massicotte, P., and South, A. (2023). rnaturalearth: World Map Data from Natural Earth. https://cran.r-project.org/web/packages/rnaturalearth/rnaturalearth.pdf.

126. Pebesma, E. (2018). Simple features for R: standardized support for spatial vector data. R J. *10*, 439–446. https://doi.org/10.32614/RJ-2018-009.

127. Madeira, F., Pearce, M., Tivey, A.R.N., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N., Kolesnikov, A., and Lopez, R. (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic Acids Res. *50*, W276–W279. https://doi.org/10.1093/nar/gkac240.

128. Li, L., Li, C., Sarkar, S., Zhang, J., Witham, S., Zhang, Z., Wang, L., Smith, N., Petukh, M., and Alexov, E. (2012). DelPhi: a comprehensive suite for DelPhi software and associated resources. BMC Biophys. *5*, 9. https://doi.org/10.1186/2046-1682-5-9.

129. Xiang, Z., Soto, C.S., and Honig, B. (2002). Evaluating conformational free energies: the colony energy and its application to the problem of loop prediction. Proc. Natl. Acad. Sci. USA *99*, 7432–7437. https://doi.org/10.1073/pnas.102179699.

130. Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kalé, L., and Schulten, K. (2005). Scalable molecular dynamics with NAMD. J. Comput. Chem. *26*, 1781–1802. https://doi.org/10.1002/jcc.20289.

131. Mitternacht, S. (2016). FreeSASA: An open source C library for solvent accessible surface area calculations. F1000Res *5*, 189. https://doi.org/10.12688/f1000research.7931.1.

132. Weaver, S., Shank, S.D., Spielman, S.J., Li, M., Muse, S.V., and Kosakovsky Pond, S.L. (2018). Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes. Mol. Biol. Evol. *35*, 773–777. https://doi.org/10.1093/molbev/msx335.

133. Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., and Ferrin, T.E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. *30*, 70–82. https://doi.org/10.1002/pro.3943.

134. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., et al. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. *46*, W296–W303. https://doi.org/10.1093/nar/gky427.

135. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. Nucleic acids research *28*, 235–242. https://doi.org/10.1093/nar/28.1.235.

136. Apweiler, R., Bairoch, A., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., et al. (2004). UniProt: the Universal Protein knowledgebase. Nucleic Acids Res. *32*, D115–D119. https://doi.org/10.1093/nar/gkh131.

137. Sayers, E.W., Bolton, E.E., Brister, J.R., Canese, K., Chan, J., Comeau, D.C., Connor, R., Funk, K., Kelly, C., Kim, S., et al. (2022). Database resources of the national center for biotechnology information. Nucleic Acids Res. *50*, D20–D26. https://doi.org/10.1093/nar/gkab1112.

138. Nielsen, G.K., Dagnaes-Hansen, F., Holm, I.E., Meaney, S., Symula, D., Andersen, N.T., and Heegaard, C.W. (2011). Protein replacement therapy partially corrects the cholesterol-storage phenotype in a mouse model of Niemann-Pick type C2 disease. PLoS One *6*, e27287. https://doi.org/10.1371/journal.pone.0027287.

139. Devi, G.R., Byrd, J.C., Slentz, D.H., and MacDonald, R.G. (1998). An insulin-like growth factor II (IGF-II) affinity-enhancing domain localized within extracytoplasmic repeat 13 of the IGF-II/mannose 6-phosphate receptor. Mol. Endocrinol. *12*, 1661–1672. https://doi.org/10.1210/mend.12.11.0192.

140. Beukers, M.W., Oh, Y., Zhang, H., Ling, N., and Rosenfeld, R.G. (1991). [Leu27] insulin-like growth factor II is highly selective for the type-II IGF receptor in binding, cross-linking and thymidine incorporation experiments. Endocrinology *128*, 1201–1203. https://doi.org/10.1210/endo-128-2-1201.

141. Hashimoto, R., Fujiwara, H., Higashihashi, N., Enjoh-Kimura, T., Terasawa, H., Fujita-Yamaguchi, Y., Inagaki, F., Perdue, J.F., and Sakano, K. (1995). N-terminal deletion mutants of insulin-like growth factor-II (IGF-II) show Thr7 and Leu8 important for binding to insulin and IGF-I receptors and Leu8 critical for all IGF-II functions. J. Biol. Chem. *270*, 18013–18018. https://doi.org/10.1074/jbc.270.30.18013.

142. Kleinfelter, L.M., Jangra, R.K., Jae, L.T., Herbert, A.S., Mittler, E., Stiles, K.M., Wirchnianski, A.S., Kielian, M., Brummelkamp, T.R., Dye, J.M., et al. (2015). Haploid Genetic Screen Reveals a Profound and Direct Dependence on Cholesterol for Hantavirus Membrane Fusion. mBio *6*, e00801. https://doi.org/10.1128/mBio.00801-15.

143. Bortz, R.H., 3rd, Wong, A.C., Grodus, M.G., Recht, H.S., Pulanco, M.C., Lasso, G., Anthony, S.J., Mittler, E., Jangra, R.K., and Chandran, K. (2020). A Virion-Based Assay for Glycoprotein Thermostability Reveals Key Determinants of Filovirus Entry and Its Inhibition. J. Virol. *94*, e00336-20. https://doi.org/10.1128/JVI.00336-20.

144. Clare, E.L., Lim, B.K., Fenton, M.B., and Hebert, P.D.N. (2011). Neotropical Bats: Estimating Species Diversity with DNA Barcodes. PLoS One *6*, e22648. https://doi.org/10.1371/journal.pone.0022648.

145. Townzen, J.S., Brower, A.V.Z., and Judd, D.D. (2008). Identification of mosquito bloodmeals using mitochondrial cytochrome oxidase subunit I and cytochrome b gene sequences. Med. Vet. Entomol. *22*, 386–393. https://doi.org/10.1111/j.1365-2915.2008.00760.x.

146. Slough, M.M., Chandran, K., and Jangra, R.K. (2019). Two Point Mutations in Old World Hantavirus Glycoproteins Afford the Generation of Highly Infectious Recombinant Vesicular Stomatitis Virus Vectors. mBio *10*, e02372-18. https://doi.org/10.1128/mBio.02372-18.

147. Slough, M.M., Li, R., Herbert, A.S., Lasso, G., Kuehne, A.I., Monticelli, S.R., Bakken, R.R., Liu, Y., Ghosh, A., Moreau, A.M., et al. (2023). Two point mutations in protocadherin-1 disrupt hantavirus recognition and afford protection against lethal infection. Nat. Commun. *14*, 4454. https://doi.org/10.1038/s41467-023-40126-y.

148. Ortigoza, M.B., Yoon, H., Goldfeld, K.S., Troxel, A.B., Daily, J.P., Wu, Y., Li, Y., Wu, D., Cobb, G.F., Baptiste, G., et al. (2022). Efficacy and Safety of COVID-19 Convalescent Plasma in Hospitalized Patients: A Randomized Clinical Trial. JAMA Intern. Med. *182*, 115–126. https://doi.org/10.1001/jamainternmed.2021.6850.

149. McKinney, W. (2010). Data Structures for Statistical Computing in Python. In Proceedings of the 9th Python in Science Conference, pp. 56–61. https://doi.org/10.25080/Majora-92bf1922-00a.

150. Wang, L., Deng, Y., Wu, Y., Kim, B., LeBard, D.N., Wandschneider, D., Beachy, M., Friesner, R.A., and Abel, R. (2017). Accurate Modeling of Scaffold Hopping Transformations in Drug Discovery. J. Chem. Theor. Comput. *13*, 42–54. https://doi.org/10.1021/acs.jctc.6b00991.

151. Clark, A.J., Negron, C., Hauser, K., Sun, M., Wang, L., Abel, R., and Friesner, R.A. (2019). Relative Binding Affinity Prediction of Charge-Changing Sequence Mutations with FEP in Protein–Protein Interfaces. J. Mol. Biol. *431*, 1481–1493. https://doi.org/10.1016/j.jmb.2019.02.003.

152. Harder, E., Damm, W., Maple, J., Wu, C., Reboul, M., Xiang, J.Y., Wang, L., Lupyan, D., Dahlgren, M.K., Knight, J.L., et al. (2016). OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. J. Chem. Theor. Comput. *12*, 281–296. https://doi.org/10.1021/acs.jctc.5b00864.

153. Mackerell, A.D., Feig, M., and Brooks, C.L., 3rd (2004). Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J. Comp. Chem. *25*, 1400–1415. https://doi.org/10.1002/jcc.20065.

154. Warwicker, J., and Watson, H.C. (1982). Calculation of the electric potential in the active site cleft due to alpha-helix dipoles. J. Mol. Biol. *157*, 671–679. https://doi.org/10.1016/0022-2836(82)90505-8.

155. Pebesma, E., and Bivand, R. (2023). Spatial Data Science with Applications in R, First Edition (Chapman and Hall/CRC). https://doi.org/10.1201/9780429459016.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
FVM09 monoclonal antibody	Keck et al.[111]	N/A
ADI-15878 monoclonal antibody	Wec et al.[112]	RRID:AB_2750594
MR72 monoclonal antibody	Bornholdt et al.[69]; Flyak et al.[113]	N/A
Anti-Flag M2-Peroxidase (HRP) antibody	Sigma	Cat# A8592; RRID:AB_439702
Alexa-680 goat anti-mouse IgG secondary antibody	Thermo Fisher Scientific	Cat# A21057; RRID:AB_2535723
Alexa-680 goat anti-rabbit IgG secondary antibody	Thermo Fisher Scientific	Cat# A-21109; RRID:AB_2535758
anti-human IgG conjugated to horseradish peroxidase secondary antibody	Thermo Fisher Scientific	Cat# A18811; RRID:AB_2535588
Mouse anti-VSV M antibody (mAb 23H12)	Kerafast	RRID: AB_2734773
Bacterial and virus strains		
rVSV-EBOV/Mayinga GP (EBOV/H.sap-tc/COD/76/Yambuku-Mayinga)	Wong et a[57]	N/A
rVSV-BDBV GP (BDBV/H.sap/UGA/07/But-811250)	Wec et al.; Wec et al.[112,114]	N/A
rVSV-BOMV GP (BOMV/Mops condylurus/SLE/2016/PREDICT_SLAB000156)	Goldstein et al.[16]	N/A
rVSV-SUDV GP (SUDV/C.por-lab/SSD/76/Boneface)	Wec et al.[112]; Wec et al.[114]	N/A
rVSV-RESTV GP (RESTV/M.fas-tc/USA/89/Phi89-AZ-1435)	Wec et al.[112]; Wec et al.[114]	N/A
rVSV-LLOV GP (LLOV/M.sch-wt/ESP/03/Asturias-Bat86)	Ng et al.[115]	N/A
rVSV-MLAV GP (MLAV/Rousettus-wt/CHN/2015/Sharen-Bat9447-1)	Wirchnianski et al.[116]	N/A
rVSV-TAFV GP (TAFV/H.sap-tc/CIV/94/CDC807212)	Wec et al.[112]; Misasi et al.[117]	N/A
rVSV-EBOV GP P146S (parental strain: EBOV/H.sap-tc/COD/76/Yambuku-Mayinga)	This paper	N/A
rVSV-BOMV GP S146S (parental strain: BOMV/Mops condylurus/SLE/2016/PREDICT_SLAB000156)	This paper	N/A
Biological samples		
Rabbit polyclonal sera	Ng et al.[115]	N/A
Chemicals, peptides, and recombinant proteins		
1X protease inhibitor	Millipore Sigma	Cat# 11836170001
Benzonase nuclease	Millipore Sigma	Cat# E1014-5KU
Filipin complex	Sigma-Aldrich	Cat# F9765-25MG
Thermolysin	Sigma-Aldrich	Cat# P1512-25MG
Phosphoramidon	Sigma-Aldrich	Cat# R7385-5MG
Trypsin inhibitor	Millipore sigma	Cat# T6522
PNGase F	New England Biolabs	Cat # P0704S
AEBSF protease inhibitor tablets	Thermo Fisher Scientific	Cat# 78431
DMEM	Thermo Fisher Scientific	Cat# 11965-118
GlutaMAX	Thermo Fisher Scientific	Cat# 35050-061

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Pen-Strep	Thermo Fisher Scientific	Cat# 15140-122
FreeStyle 293 expression media	Thermo Fisher Scientific	Cat# 12338-026
McCoy's 5A medium	Thermo Fisher Scientific	Cat# 16600108
RIPA buffer	Thermo Fisher Scientific	Cat# 89900
10X Bolt Reducing Agent	Thermo Fisher Scientific	Cat# B009
4X Bolt LDS Sample Buffer	Thermo Fisher Scientific	Cat# B0007
1-Step Ultra TMB-ELISA Substrate Solution	Thermo Fisher Scientific	Cat# 34029
Gentle Binding Buffer	Thermo Fisher Scientific	Cat# 21012
Elution buffer	Thermo Fisher Scientific	Cat# 21013
PD10 Sephadex column	Cytiva	Cat# 17085101
1-Step Ultra TMB-ELISA Substrate Solution	Thermo Fisher Scientific	Cat# 34029
Critical commercial assays		
Ni-NTA Agarose beads	Qiagen	Cat# 30230
Chromatography column	Bio-rad	Cat# 9704652
Dialysis cassette	Thermo Fisher Scientific	Cat# 66380
Protein A beads	Genscript	Cat# L00210
Anti-penta-HIS biosensors	ForteBio	Cat# 18-5120
DNA Gel Extraction kit	Millipore	Cat# LSKGEL050
One-step RT-PCR	Qiagen	Cat# 210212
Deposited data		
All code generated	Github	https://doi.org/10.6084/m9.figshare.28144118; https://github.com/chandranlab/filo_GP-bat_NPC1
NPC1-C (*Ametrida centurio* | Brazil)	GenBank	PQ137124
NPC1-C (*Anoura geoffroyi* | Trinidad)	GenBank	PQ137125
NPC1-C (*Artibeus concolor* | Brazil)	GenBank	PQ137126
NPC1-C (*Artibeus glaucus* | Brazil)	GenBank	PQ137127
NPC1-C (*Artibeus jamaicensis* | Puerto Rico)	GenBank	PQ137128
NPC1-C (*Artibeus lituratus* | Brazil)	GenBank	PQ137129
NPC1-C (*Artibeus obscurus* | Brazil)	GenBank	PQ137130
NPC1-C (*Balantiopteryx plicata* | Mexico)	GenBank	PQ137083
NPC1-C (*Brachyphylla cavernarum* | Puerto Rico)	GenBank	PQ137131
NPC1-C (*Carollia brevicauda* | Brazil)	GenBank	PQ137132
NPC1-C (*Carollia perspicillata* | Brazil)	GenBank	PQ137133
NPC1-C (*Carollia sowelli* | Mexico)	GenBank	PQ137134
NPC1-C (*Centurio senex* | Mexico)	GenBank	PQ137135
NPC1-C (*Chaerephon nigeriae* | Senegal)	GenBank	PQ137104
NPC1-C (*Chaerephon plicatus* | Myanmar)	GenBank	PQ137105
NPC1-C (*Chaerephon pumilus*| Tanzania)	GenBank	PQ137107
NPC1-C (*Chaerephon pumilus*| Sierra Leone)	GenBank	PQ137106
NPC1-C (*Coleura afra* | Tanzania)	GenBank	PQ137084
NPC1-C (*Cynopterus sphinx* | Bangladesh)	GenBank	PQ137155
NPC1-C (*Desmodus rotundus* | Belize)	GenBank	PQ137136
NPC1-C (*Diaemus youngi* | Brazil)	GenBank	PQ137137
NPC1-C (*Eidolon helvum* | Saudi Arabia)	GenBank	PQ137156
NPC1-C (*Eonycteris spelaea* | Thailand)	GenBank	PQ137157
NPC1-C (*Epomophorus labiatus* | Sierra Leone)	GenBank	PQ137158
NPC1-C (*Epomophorus minor* | Sierra Leone)	GenBank	PQ137159
NPC1-C (*Epomophorus wahlbergi* | Tanzania)	GenBank	PQ137160

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
NPC1-C (*Epomops buettikoferi* | Sierra Leone)	GenBank	PQ137161
NPC1-C (*Eptesicus fuscus wetmorei* | Puerto Rico)	GenBank	PQ137179
NPC1-C (*Erophylla bombifrons* | Puerto Rico)	GenBank	PQ137138
NPC1-C (*Glauconycteris poensis* | Sierra Leone)	GenBank	PQ137180
NPC1-C (*Glossophaga leachii* | Mexico)	GenBank	PQ137139
NPC1-C (*Glossophaga soricina* | Mexico)	GenBank	PQ137140
NPC1-C (*Hipposideros armiger* | Thailand)	GenBank	PQ137088
NPC1-C (*Hipposideros beatus* | Republic of Congo)	GenBank	PQ137089
NPC1-C (*Hipposideros bicolor* | Malaysia)	GenBank	PQ137090
NPC1-C (*Hipposideros cineraceus* | Malaysia)	GenBank	PQ137091
NPC1-C (*Hipposideros diadema* | Malaysia)	GenBank	PQ137092
NPC1-C (*Hipposideros jonesi* | Sierra Leone)	GenBank	PQ137093
NPC1-C (*Hipposideros larvatus* | Bangladesh)	GenBank	PQ137094
NPC1-C (*Hipposideros lekaguli* | Thailand)	GenBank	PQ137095
NPC1-C (*Hipposideros pomona* | Thailand)	GenBank	PQ137096
NPC1-C (*Hipposideros ruber* complex | Uganda)	GenBank	PQ137097
NPC1-C (Hipposideros ruber complex | Tanzania)	GenBank	PQ137098
NPC1-C (*Kerivoula cuprosa* | Sierra Leone)	GenBank	PQ137181
NPC1-C (*Kerivoula phalaena* | Republic of Congo)	GenBank	PQ137182
NPC1-C (*Kerivoula smithii* | Republic of Congo)	GenBank	PQ137183
NPC1-C (*Lonchorhina aurita* | Trinidad)	GenBank	PQ137141
NPC1-C (*Macroglossus sobrinus* | Malaysia)	GenBank	PQ137162
NPC1-C (*Macrotus waterhousii* | Mexico)	GenBank	PQ137142
NPC1-C (*Megaderma sp* | Bangladesh)	GenBank	PQ137099
NPC1-C (*Megaloglossus woermanni* | Liberia)	GenBank	PQ137163
NPC1-C (Micropteropus pusillus | Sierra Leone)	GenBank	PQ137164
NPC1-C (*Mimon crenulatum* | Brazil)	GenBank	PQ137143
NPC1-C (*Miniopterus magnater* | Thailand)	GenBank	PQ137100
NPC1-C (*Miniopterus minor* | Tanzania)	GenBank	PQ137101
NPC1-C (*Miniopterus nimbae* | Liberia)	GenBank	PQ137102
NPC1-C (*Miniopterus pusillus* | Thailand)	GenBank	PQ137103
NPC1-C (*Molossops temminckii* | Brazil)	GenBank	PQ137108
NPC1-C (*Molossus molossus* | Trinidad)	GenBank	PQ137109
NPC1-C (*Molossus rufus* | Trinidad)	GenBank	PQ137110
NPC1-C (*Monophyllus redmani* | Puerto Rico)	GenBank	PQ137144
NPC1-C (*Mops cf nanulus* | Sierra Leone)	GenBank	PQ137111
NPC1-C (Mops condylurus | Sierra Leone)	GenBank	PQ137112
NPC1-C (*Mormoops blainvillii* | Puerto Rico)	GenBank	PQ137115
NPC1-C (*Mormoops megalophylla* | Mexico)	GenBank	PQ137116
NPC1-C (*Myonycteris angolensis* | Uganda)	GenBank	PQ137165
NPC1-C (*Myonycteris angolensis smithii* | Sierra Leone)	GenBank	PQ137166
NPC1-C (*Myopterus daubentoni* | Senegal)	GenBank	PQ137113
NPC1-C (Myotis bocagii | Sierra Leone)	GenBank	PQ137184
NPC1-C (*Natalus stramineus* | Mexico)	GenBank	PQ137120
NPC1-C (*Neoromicia nanus* | Uganda)	GenBank	PQ137185
NPC1-C (*Neoromicia rendalli* | Sierra Leone)	GenBank	PQ137186
NPC1-C (*Noctilio leporinus* | Puerto Rico)	GenBank	PQ137121
NPC1-C (*Nycteris hispida* | Sierra Leone)	GenBank	PQ137122

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
NPC1-C (*Nycteris sp.* | Sierra Leone)	GenBank	PQ137123
NPC1-C (*Nycticeinops schlieffeni* | CIV)	GenBank	PQ137187
NPC1-C (*Nyctinomops macrotis* | Mexico)	GenBank	PQ137114
NPC1-C (*Phylloderma stenops* | Brazil)	GenBank	PQ137145
NPC1-C (*Phyllostomus elongatus* | Brazil)	GenBank	PQ137146
NPC1-C (*Phyllostomus hastatus* | Trinidad)	GenBank	PQ137147
NPC1-C (*Pipistrellus coromandra* | Bangladesh)	GenBank	PQ137188
NPC1-C (*Pipistrellus coromandra* | Malaysia)	GenBank	PQ137189
NPC1-C (*Pipistrellus nanulus* | Sierra Leone)	GenBank	PQ137190
NPC1-C (*Pteronotus parnellii* | Trinidad)	GenBank	PQ137117
NPC1-C (*Pteronotus quadridens* | Puerto Rico)	GenBank	PQ137118
NPC1-C (*Pteronotus rubiginosus* | Brazil)	GenBank	PQ137119
NPC1-C (*Pteropus lylei* | Thailand)	GenBank	PQ137168
NPC1-C (*Pteropus medius* | Bangladesh)	GenBank	PQ137167
NPC1-C (*Pteropus sp* | Thailand)	GenBank	PQ137169
NPC1-C (*Rhinolophus affinis* | Malaysia)	GenBank	PQ137171
NPC1-C (*Rhinolophus hipposideros* | Pakistan)	GenBank	PQ137172
NPC1-C (*Rhinolophus shameli* | Malaysia)	GenBank	PQ137173
NPC1-C (*Rhinolophus shortridgei* | Bangladesh)	GenBank	PQ137174
NPC1-C (*Rhinolophus sp* | Uganda)	GenBank	PQ137177
NPC1-C (*Rhinolophus sp* | Sierra Leone)	GenBank	PQ137175
NPC1-C (*Rhinolophus sp* | Sierra Leone)	GenBank	PQ137176
NPC1-C (*Rhinophylla pumilio* | Brazil)	GenBank	PQ137148
NPC1-C (*Rousettus leschenaultii* | Bangladesh)	GenBank	PQ137170
NPC1-C (*Scotoecus albofuscus* | Sierra Leone)	GenBank	PQ137191
NPC1-C (*Scotophilus kuhlii* | Bangladesh)	GenBank	PQ137192
NPC1-C (*Scotophilus viridis nigritellus* | Sierra Leone)	GenBank	PQ137193
NPC1-C (*Stenoderma rufum* | Puerto Rico)	GenBank	PQ137149
NPC1-C (*Sturnira lilium* | Brazil)	GenBank	PQ137150
NPC1-C (*Sturnira luisi* | Brazil)	GenBank	PQ137151
NPC1-C (*Taphozous longimanes* | Bangladesh)	GenBank	PQ137086
NPC1-C (*Taphozous mauritianus* | Tanzania)	GenBank	PQ137087
NPC1-C (*Triaenops afer* | Republic of Congo)	GenBank	PQ137178
NPC1-C (*Unidentified bat* | Brazil)	GenBank	PQ137085
NPC1-C (*Unidentified Phyllostomid Bat* | Trinidad)	GenBank	PQ137152
NPC1-C (*Unidentified Vespertillionid Bat* | Republic of Congo)	GenBank	PQ137194
NPC1-C (*Uroderma bilobatum* | Trinidad)	GenBank	PQ137153
NPC1-C (*Vampyressa bidens* | Brazil)	GenBank	PQ137154
Experimental models: Cell lines		
NPC1 KO U2OS cells	Spence et al.[118]	N/A
African green monkey kidney (VERO) cells	ATCC	Cat# CCL-81; RRID:CVCL_0059
Freestyle 293-F cells	Thermo Fisher Scientific	Cat# R79007
Human osteosarcoma (U2OS) cells	ATCC	Cat# HTB-96
Oligonucleotides		
Primer: NPC1-FWDa Forward: TTCTCCTTGGTCTTCATTGCT	This paper	N/A
Primer: NPC1-FWDb Forward: TTCTCCCTGGTCTTTATCGCC	This paper	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Primer: NPC1-FWDc Forward: TCCTTCTGTGTCCGAAACCC	This paper	N/A
Primer: NPC1-FWDd Forward: TTCTCCCTGGTCTTTATCGCC	This paper	N/A
Primer: NPC1-RVSa Reverse: AGGCTATGGAAATATACAGGAACA	This paper	N/A
Primer: NPC1-RVSb Reverse: AGCTTTTGATGTGCCCCAAG	This paper	N/A
Primer: NPC1-RVSc Reverse: AGGCGATGGAAATGTACAGAAACA	This paper	N/A
Primer: NPC1-RVSd Reverse: AGCTTTTGATGTGCCCCAAG	This paper	N/A
Recombinant DNA		
Plasmids: bat NPC1-C (81 orthologs)	This paper	N/A
Plasmid: *Hs*NPC1	This paper	N/A
Plasmid: ADI-15878	Wec et al.[112]	N/A
Plasmid: MR72	Bornholdt et al.[69]; Flyak et al.[113]	N/A
Software and algorithms		
Python (version 2.8)	Python Software Foundation	https://www.python.org
SciPy package	Virtanen et al.[119]	https://scipy.org
Seaborn	Waskom[120]	https://seaborn.pydata.org
Scikit-learn	Pedregosa et al.[121]	https://scikit-learn.org/stable/
Mlxtend	Raschka[122]	https://rasbt.github.io/mlxtend/
SHapley Additive exPlanations	Lundberg et al.[65]; Lundberg and Lee[66]	https://shap.readthedocs.io/en/latest/
R	R Core Team[123]	https://www.r-project.org
R package glmmTMB	Brooks et al.[124]	https://cran.r-project.org/web/packages/glmmTMB/index.html
R package rnaturalearth	Massicotte and South[125]	https://cran.r-project.org/web/packages/rnaturalearth/index.html
R package sf	Pebesma[126]	https://cran.r-project.org/web/packages/sf/index.html
Clustal Omega	Madeira et al.[127]	https://www.ebi.ac.uk/jdispatcher/msa/clustalo
Schrödinger's Suite	Schröedinger	https://www.schrodinger.com/
Delphi	Li et al.[128]	http://compbio.clemson.edu/lab/delphisw/
Loopy	Xiang et al.[129]	https://honig.c2b2.columbia.edu/loopy
NAMD	Phillips et al.[130]	https://www.ks.uiuc.edu/Research/namd/
FreeSASA	Mitternacht[131]	https://freesasa.github.io
FUBAR	Murrell et al.[78]	http://www.hyphy.org
MEME	Murrell et al.[77]	http://www.hyphy.org
Pymol (version 2.5.5)	Schröedinger	https://www.pymol.org
PRISM (version 10.2.3)	Graphpad	https://www.graphpad.com
Adobe Illustrator (version 28.6)	Adobe	https://www.adobe.com
DataMonkey	Weaver et al.[132]	http://www.datamonkey.org
LICOR Image Studio software (version.1.0.19)	LI-COR Biosciences	https://www.licor.com/bio/image-studio
ChimeraX	Pettersen et al.[133]	https://www.cgl.ucsf.edu/chimerax
SWISS-MODEL	Waterhouse et al.[134]	https://swissmodel.expasy.org

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Other		
Mini Blot Module	Thermo Fisher Scientific	Cat# B1000
iBright Imaging System	Thermo Fisher Scientific	Cat# A44241
Cytation 5 cell imaging multi-mode reader	Agilent BioTek	N/A
iBlot transfer stacks	Thermo Fisher Scientific	Cat# IB301002
iBlot Gel Transfer Device	Thermo Fisher Scientific	Cat# IB1001
30K MWCO Amicon concentrator	Merck-millipore	Cat# UFC903086
LI-COR Fc fluorescent imager	LI-COR Biosciences	N/A
6-well plates	Corning	Cat# 3506
24-well plate	Corning	Cat# 3524
96-well plates	Corning	Cat# 3690
Database: PDB	Berman et al.[135]	https://www.rcsb.org
Database Uniprot	Apweiler et al.[136]	https://www.uniprot.org
Database: NCBI database	Sayers et al.[137]	https://www.ncbi.nlm.nih.gov/gene/
Database: The Mammal Diversity Database	Marsh et al.[79]; Mammal Diversity Database[80]	https://www.mammaldiversity.org
Database: USAID PREDICT	N/A	https://ohi.vetmed.ucdavis.edu/programs-projects/predict-project
Database: 2022 African Chiroptera Report[82]	Cakenberghe and Seamark[82]	https://www.africanbats-npc.org/acr/african-chiroptera-report-2022

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell lines

Clonal human U2OS osteosarcoma cell lines knocked out for human NPC1 (NPC1-KO) and NPC1-KO cell populations ectopically expressing human NPC1 have been described previously.[118]

African green monkey kidney cells (Vero, ATCC) were cultured in high-glucose Dulbecco's modified Eagle medium (DMEM, Thermo Fisher Scientific) and supplemented with 2% fetal bovine serum (FBS, Atlanta Biologicals), 1% GlutaMAX (Thermo Fisher Scientific), and 1% penicillin-streptomycin (Pen-Strep, Thermo Fisher Scientific) and kept in a humidified incubator at 37°C with 5% CO_2. Freestyle 293-F cells (Thermo Fisher Scientific) were grown in FreeStyle 293 expression media (Thermo Fisher Scientific) supplemented with 1% Pen-Strep and incubated in a shaking incubator at 37°C with 8% CO_2. Human osteosarcoma cells (U2OS, ATCC) were cultured in modified McCoy's 5A medium (Thermo Fisher Scientific), supplemented with 10% FBS, 1% GlutaMAX, and 1% Pen-Strep. U2OS cells were cultured in a humidified incubator at 37°C with 5% CO2.

Viruses

Recombinant vesicular stomatitis viruses (rVSVs) encoding enhanced green fluorescent protein (eGFP) in the first position and replacing VSV G with glycoproteins from EBOV/Mayinga (EBOV/H.sap-tc/COD/76/Yambuku-Mayinga), TAFV (TAFV/H.sap-tc/CIV/94/CDC807212), BDBV (BDBV/H.sap/UGA/07/But-811250), BOMV (BOMV/Mops condylurus/SLE/2016/PREDICT_SLAB000156), SUDV/Boneface (SUDV/C.por-lab/SSD/76/Boneface), RESTV (RESTV/M.fas-tc/USA/89/Phi89-AZ-1435), LLOV (LLOV/M.sch-wt/ESP/03/Asturias-Bat86) and MLAV GP (MLAV/Rousettus-wt/CHN/2015/Sharen-Bat9447-1) were propagated as described previously.[16,57,112,114,116,117,138–143] All viruses were propagated on Vero cells. rVSV-EBOV P146S and rVSV-BOMV S146P (EBOV residue numbering) viruses were generated using the same method as the WT rVSV-GPs described above. Briefly, the VSV G protein, in the context of the VSV genomic plasmid, was replaced with either EBOV GP P146S or BOMV GP S146P and rescued by co-transfection with plasmids encoding VSV N, P and L in 293FT cells and subsequently propagated in Vero cells. Virus stocks of these mutants were generated and confirmed by plaque purification and Sanger sequencing.

METHOD DETAILS

Generation of NPC1 sequences

NPC1 mRNA sequences were generated for 112 bat species. An additional bat NPC1 sequences were obtained from GenBank (Table S1). To generate bat NPC1 sequences we accessed bio-banked nucleic acids collected and stored as part of the USAID PREDICT project.[81] Bat species were identified based on morphological traits (i.e., 'field' identification) and confirmed in the laboratory by DNA barcoding of cytochrome c oxidase I (COI) gene or cytochrome B (cytB).[144,145] All COI bands of expected size were gel extracted from 1% agarose gel and purified using Millipore DNA Gel Extraction Kit (Millipore). NPC1 sequencing was performed using

a nested RT-PCR approach targeting mRNA using one step RT-PCR (Qiagen). An initial set of primers flanking domain C of NPC1 was used to amplify a target of 1,128 bases. The primers used were: NPC1con/FWD: TTYGAGGGCTGYTTGAGG and NPC1con/RVS: ACCAGRATGAAGATGTTGTCCACC. If no band was observed on gel electrophoresis, template from the round 1 PCR was used in a subsequent nested PCR reaction using 4 different pairs of primers. The forward primers were: NPC1-FWDa: TTCTCCTTGGTCTTCATTGCT; NPC1-FWDb: TTCTCCCTGGTCTTTATCGCC; NPC1-FWDc: TCCTTCTGTGTCCGAAACCC; NPC1-FWDd: TTCTCCCTGGTCTTTATCGCC. The reverse primers were NPC1-RVSa: AGGCTATGGAAATATACAGGAACA; NPC1-RVSb: AGCTTTTGATGTGCCCCAAG; NPC1-RVSc: AGGCGATGGAAATGTACAGAAACA, and NPC1-RVSd: AGCTTTTGATG TGCCCCAAG. Expected amplicon size was 850-871bp. Bands of the expected size were cloned into Stataclone PCR cloning vector and 8 white colonies form each construct were sequenced.

GP Proteolytic cleavage in rVSVs

Cleavage reactions for the rVSV-filovirus GPs were performed in NT buffer (10 mM Tris Cl, pH 7.5, 135 mM NaCl) with thermolysin (THL, Sigma-Aldrich) for 1 hour in a 37°C water bath and terminated by adding 0.5 mM phosphoramidon (Sigma-Aldrich) followed by a 20 min incubation on ice. The optimal concentration of thermolysin used in the cleavage reactions varied for each rVSV-filovirus GP and was determined via the detection of cleaved GP (GP_{CL}) via immunoblotting and subsequent functional ELISA assays. The concentrations of thermolysin used for each rVSV-filovirus GP cleavage reaction were as follows: rVSV-EBOV-GP, 0.5 mg/mL; rVSV-EBOV-GP P146S, 0.5 mg/mL; rVSV-SUDV-GP, 0.5 mg/mL; rVSV-BOMV-GP, 0.25 mg/mL; rVSV-BOMV-GP S146P, 0.25 mg/mL; rVSV-TAFV-GP, 0.4 mg/mL; rVSV-RESTV-GP, 0.2 mg/mL; rVSV-BDBV-GP, 0.25 mg/mL; rVSV-LLOV-GP, 0.1 mg/mL; rVSV-MLAV-GP, 0.21 mg/mL. For rVSV-MARV-GP, 150 ug/mL of trypsin was incubated with the virus for 10 min at 37°C and stopped using (11.25 μL trypsin inhibitor, Millipore sigma).

Detection of cleaved GP in rVSVs

To detect GP_{CL} for the rVSV-filovirus GPs, cleavage products (using conditions described in the section above, GP proteolytic cleavage) were denatured by adding Glycoprotein Denaturing Buffer at 1X concentration and incubating at 95C for 5 min. Denatured products were incubated on ice for 5 min and subsequently deglycosylated by treating with 1000 U PNGase F, 1X NP-40 (1% NP-40 in MiliQ-H2O), 1X GlycoBuffer 2 (50 mM Sodium Phosphate, pH 7.5) (New England Biolabs) for 1 h in a 37°C water bath. Samples were prepared for western blot using the 10X Bolt Reducing Agent (Thermo Fisher Scientific) and 4X Bolt LDS Sample Buffer (Thermo Fisher Scientific) and incubated at 70°C for 10 minutes. Samples were run on a 4-12% Bolt Bis-Tris precast gel (Thermo Fisher Scientific) for 1 hour at 110V. Proteins were transferred onto a 0.2 μM Nitrocellulose membrane from the iBlot transfer stacks (Thermo Fisher Scientific) for 7 minutes at 23V using the iBlot Gel Transfer Device (Thermo Fisher Scientific). GP1 was detected using rabbit polyclonal sera (1:2000 dilution) recognizing the N-terminal of (residues 83-97; TKRWGFRSDVIPKIV) EBOV GP[115] in 1% blocking milk (1X PBS) for 1 h at room temperature or 4C overnight followed by Alexa-680 goat anti-rabbit IgG (Thermo Fisher Scientific) (1:10,000 dilution) in 1% blocking milk. LI-COR Fc fluorescent imager was used to image western blot.

Normalization of GP amount per well plate

To normalize for GP amount per well in ELISA binding experiments, viral particles were subjected to 3-fold serial dilutions and directly immobilized on high-binding 96-well plates (Corning), followed by overnight incubation at 4°C. The plate was washed with PBS and then blocked with a blocking buffer consisting of PBS supplemented with 3% bovine serum albumin (BSA, Fisher Scientific) for 1 hour at 37°C. After blocking, the plates were washed again with PBS, and each well was incubated with 2 μg/mL of FVM09 antibody (a macaque monoclonal antibody targeting a highly conserved filovirus GP epitope) diluted in blocking buffer for 1 hour at 37°C.[111] Following incubation, plates were washed with PBS, and FVM09 binding was detected using an anti-human IgG conjugated to horseradish peroxidase secondary antibody (HRP, Thermo Fisher Scientific) diluted at 1:9000 in blocking buffer for 1 hour at 37°C. After washing with PBS, 1-Step Ultra TMB-ELISA Substrate Solution (Thermo Fisher Scientific) was added and allowed to develop for 10 minutes, followed by quenching with 0.5 M H2SO4 (Sigma-Aldrich). Absorbance was measured at 450 nm using a Cytation 5 cell imaging multi-mode reader (Agilent Biotek). The binding data were analyzed in GraphPad Prism using Non-linear regression (Sigmoidal, 4PL model).

Normalization of virus particles for infection input using VSV-M immunoblotting

To normalize for viral particles as input for infection assays with rVSV-EBOV-GP WT, rVSV-EBOV-GP P146S, rVSV-BOMV-GP WT, and rVSV-BOMV-GP S146P, 10 μL of each virus stock was used for immunoblotting of the VSV M protein. The primary antibody, mouse anti-VSV M (mAb 23H12),[146] was used at a 1:500 dilution followed by incubation with the secondary antibody Alexa Goat anti-mouse 680 (Thermo Fisher Scientific). LI-COR Fc fluorescent imager was used to image western blot and for signal analysis/quantification (LI-COR Biosciences).

Expression and purification of soluble NPC1-C

293 FreeStyle cells were seeded at a density of 500,000 cells/mL and incubated in a 37°C shaker under 8% CO_2 overnight. The following day, transfection of plasmids expressing soluble NPC1-C was carried out using 0.67 μg of DNA per mL of the culture and 3 μg of PEI transfection reagent per μg of plasmid DNA. Briefly, plasmid DNA was mixed with the PEI reagent, vortexed, and incubated at room temperature for 20 minutes to facilitate the formation of PEI-DNA complexes. The transfection mixtures were

then added to the cells and incubated at 37°C under 8% CO_2 in a shaker for 6 days. The supernatant containing secreted NPC1-C was harvested 6 days post-transfection.

For purification, supernatants containing soluble NPC1-C were supplemented with 100 mM HEPES (pH 7.6) and two AEBSF protease inhibitor tablets (final concentration 0.2 mM, Thermo Fisher Scientific) before centrifugation at 4000 x g for 15 minutes at 4°C. The supernatant was decanted into a new container, and centrifugation was repeated. Ni-NTA Agarose beads (5 mL, Qiagen) were resuspended and transferred to a chromatography column (Bio-rad), which was then washed with 20 mL 1X PBS. The supernatant containing NPC1-C and Ni-NTA agarose beads were mixed in a 1 L beaker, and a magnetic stir rod was inserted. The mixture was continuously stirred at 4°C overnight to allow for binding of the beads and HIS-tagged NPC1-C.

The following day, the beaker was placed on ice for 10 minutes to allow the beads to settle, and the mixture was decanted into a chromatography column for gravity filtration. The beads were washed twice with 20 mL low imidazole nickel wash (20 mM imidazole) and once with 20 mL nickel wash buffer (500 mM NaCl, 20 mM HEPES pH 7.6, 10% glycerol; no imidazole) in the chromatography column. Elution of NPC1-C was performed by adding 2 mL of nickel elution buffer (500 mM NaCl, 500 mM imidazole, 20 mM HEPES pH 7.6, 10% glycerol) to the chromatography column and collecting the eluent in a 10 mL beaker. A second elution was carried out with an additional 1 mL of elution buffer. The 3 mL of protein elution was transferred to a dialysis cassette (Thermo Fisher Scientific) and placed in a 4 L beaker containing dialysis buffer, followed by overnight incubation at 4°C on a magnetic stir plate. The dialyzed, purified soluble NPC1-C was collected the next day and transferred to Eppendorf tubes. Purified NPC1-C was quantified using biolayer interferometry with anti-penta-HIS biosensors (ForteBio) and subsequently normalized for GP-binding ELISAs.

Generation of IgGs

Plasmids expressing the light and heavy chains of ADI-15878[112] and MR72[69,113] were added to 45 mL of PBS (200 ug each). 1.2 mL of PEI transfection reagent was added to the PBS DNA mixture, vortexed, and incubated at room temperature for 20 min. 293 FreeStyle cells were transfected with the PEI:DNA mixture at a density of $1e^6$ cells per mL. Cultures were harvested for purification 6 days post transfection. For purification, cells were centrifuged at 4000 rpm for 20 min at 4°C and, supernatant was decanted into a new beaker. Supernatant pH was brought to pH 8 and one AEBSF protease inhibitor tablet was added. Protein A beads (Genscript) were added (0.5 mL) to a disposable 15 mL chromatography column with membrane and washed with 20 mL Gentle Binding Buffer (Thermo Fisher Scientific). The protein A beads were transferred to the supernatant containing the antibody and incubated at 4°C for 2 hours to allow for binding. After 2 hours, the supernatant bead mixture was decanted in a membrane column, washed twice with 10 mL Gentle Binding Buffer, and eluted using 12.5 mL elution buffer (Thermo Fisher Scientific). Two PD10 Sephadex columns (Cytiva) were washed with 20 mL of HEPES buffer. The elute was transferred to the PD10 columns, and the flowthrough was collected and supplemented with 3.5 mL HEPES buffer. The flowthrough was then concentrated to a final volume of 1-1.5 mL using a 30K MWCO Amicon concentrator (Merck-millipore) and spinning at 4000 rpm for 2–10 min.

Single-cycle infectivity

rVSVs expressing GFP and filovirus GPs in place of the VSV G were added to cells and incubated at 37C for 1 hr. NH_4Cl (20 mM) was added to block viral spread from subsequent rounds of infection. Individual eGFP-positive cells were counted at 12-14 hours post-infection using the Cytation-5 (Agilent BioTek).

Binding ELISA

We used two antibodies that recognize a conserved epitope in the base of the filovirus GP (ADI-15878 and MR72) to immobilize a cleaved form of filovirus GP carrying rVSVs onto ELISA plates.[112] ADI-15878[112] and MR72[69,113] antibodies were diluted in PBS (Fisher Scientific, pH 7.4) to 2 μg/mL and 5 μg/mL, respectively, coated onto high-binding half-area 96-well plates (Corning) and incubated at 37°C for 1 hour. Plates were then blocked in a blocking buffer containing PBS and 3% bovine serum albumin (BSA) for 1 hour at 37°C. Blocked plates were exposed to GP-normalized concentrations of thermolysin-cleaved rVSV-GP particles for 1 hour at 37°C. After washing with PBS and the plates were exposed to a 3-fold serial dilution of soluble NPC1-C proteins starting at 1 μM for another hour at 37°C. Subsequently, plates were washed with PBS, and incubated with a monoclonal anti-Flag M2-Peroxidase (HRP) antibody (Sigma) for 1 hour at 37°C. Plates were washed with PBS, then 1-Step Ultra TMB-ELISA Substrate Solution (Thermo Fisher Scientific) was added for 10-15 minutes and quenched with 0.5 M H2SO4 (Sigma-Aldrich). Absorbance was read at 450 nm on a Cytation-5 cell imaging multi-mode reader (Agilent Biotek). We assessed the binding avidity between 7 rVSV filovirus GPs (BDBV, BOMV, EBOV, TAFV, RESTV, SUDV, MLAV) and the NPC1-C orthologs from 81 different bats and human (total number of GP-NPC1 pairs: 574). We also assessed the binding avidity between rVSV LLOV GP and 61 NPC1-C orthologs (60 orthologs from bat and human NPC1-C).

Binding curves were generated by fitting a sigmoidal curve, and the corresponding Area-Under-the-Curve was used to describe binding strength (as binding AUC increases, so does the binding strength). We used an *in-house* algorithm to estimate sigmoidal curves and calculate the corresponding binding AUC.[147,148] We first processed each biological replicate independently by subtracting the corresponding background noise and normalizing the Optical Densities (OD) relative to a positive control (EBOV GP rVSV binding to human human NPC1). Background noise was inferred by considering the lowest average OD readout using the corresponding technical replicates for a given sample. Denoised ODs were then normalized using the highest average ODs for the positive

control included in the same ELISA plate. Following this, we jointly processed denoised and normalized ODs to fit a single sigmoidal curve [1] and estimate its area under the curve (binding AUC). Curve fit was implemented using the non-linear least square function as implemented in the SciPy package.[119]

$$y = \frac{y_{min} + (y_{max} - y_{min})}{\left[1 + 10^{((log_{10}EC_{50} - x) \times Hill)}\right]} \quad \text{(Equation 1)}$$

Where y corresponds to the experimental readout (ER); y_{min} and y_{max} are the minimum and maximum ERs, respectively; EC_{50} is the titer that gives half-maximum ER, y_{max}; Hill describes the slope of the curve, and x is the $\log_{10}$ ([rVSV]).

Sequence comparison

All-against-all pairwise sequence comparisons for full-length GP proteins and NPC1-Cs were carried out with Clustal Omega.[127]

Hierarchical clustering

Sequence and binding-based hierarchical clustering and heatmap assembly were computed using the clustermap function (method=average; metric= Euclidean distance) as implemented in the seaborn python package.[120]

Machine Learning

Dataset

The experimental dataset (574 GP-NPC1-C pairs) was initially evaluated for consistency by comparing the individual fitted sigmoidals for each of the technical and biological replicate in the binding assays. We removed ten GP-NPC1-C pairs (SUDV: 1; BOMV: 3; RESTV: 6) with larger disagreement between replicates (Binding AUC_{Max} – binding AUC_{Min} > 100 and coefficient of variation −stdev / mean > 0.3). The remaining dataset (564 GP-NPC1-C pairs) was randomly split into training (90%) and testing (10%), ensuring that binding data from each of the seven viruses was considered in each subset.

Feature extraction

We identified GP and NPC1-C interfacial residues at various inter-subunit distance cutoffs (5, 7, 10, 15, and 20 Å) based on the crystal structure of Zaire ebolavirus bound to human NPC1-C.[50,135] For this, we first identified all residues in the crystal structure with ≥ 1 heavy atom within a given distance cutoff of the binding partner. Then, we inferred interfacial residues in all other GP and NPC1-C sequences by identifying the corresponding aligned residues in a MSA obtained with Clustal Omega.[127] Each GP and NPC1-C FASTA sequence was then reduced to a sequence of interfacial residues based on a given distance cutoff. Then, we paired the interfacial FASTA sequences to describe the experimental binding pairs (564 GP-NPC1 pairs). Residues in paired FASTA sequences were translated into nine physicochemical features and all possible combinations involving 2-9 amino acid scales (511 unique combinations in total). The amino acid scales used to describe each interfacial residue were obtained from the UniProt database[136]: (1) aromatic residues, (2) accessible surface area (ASA),[62] (3) charge, (4) bulkiness,[64] (5) coil,[60] (6) flexibility,[61] (7) hydrophobicity,[59] (8) polarity, and (9) interfacial propensity.[63] Categorical features (Aromatic -True, False-, Charge -Positive, Negative-) were one-hot-encoded.[149]

Machine Learning

We employed a Random Forest (RF) regressor given its robustness to overfitting, and its interpretability. The RF regressor ('RandomForestRegressor') from scikit-learn[121] was trained on physicochemical properties describing interfacial residues (at a given distance cutoff) to predict binding strength (as dictated by the binding AUC of the sigmoidal curves describing the GP-NPC1-C binding profiles). RF hyper-parameters were optimized by randomized search and 10-fold cross-validation (CV) on the training set (n_estimators; max_depth; min_samples_split) using the ''RandomizedSearchCV'' to sample 100 parameter settings.[121] To facilitate model interpretation and reduce the risk of overfitting, we applied backward and forward sequential feature selection using the ''SequentialFeatureSelector'' function (SFS) implemented in Mlxtend.[122] SFS removes (backward) or add (forward) features in a greedy fashion based on the CV score of an estimator (r^2) until a feature subset of size k is reached. First, backward selection was applied to recursively remove features. The first subset of selected features was then subjected to a second round of selection, this time using forward selection. The value k ranged between 20 and 50 features, based on the initial size of the feature space (which differed depending on the inter-subunit distance cut-off and the number physicochemical features per residue). We evaluated the RF regression model by 10-fold CV on the training set and on the test set. In addition, we evaluated the RF regression model on a separate dataset involving Lloviu filovirus (LLOV) GP binding to 60 different bat NPC1-C and human NPC1-C. LLOV GP contains unique interfacial residues not shared with the filovirus GPs in the training set. Model evaluation was assessed by measuring the R^2 between observed and predicted binding AUCs. The best performing models were selected based on i) their predictive performance by cross-validation and on the LLOV dataset; ii) consistency of the model across varying distance cutoffs (penalizing models with significance performance drops between consecutive distance cutoffs) and; iii) when choosing the second-best performing model, we prioritized models that complement the best performing model (that is, models trained using different features whenever possible). The two best performing models (trained mostly on separate features) were integrated into a single meta-predictor by averaging the predicted binding AUCs. In addition, the two best performing models and the meta-predictor were further evaluated 100 times, the reported R^2 for each evaluation scheme (CV, test and LLOV test) corresponds to the median R^2 (Figure 3B). The meta-predictor regressor showed a more consistent performance than the individual models and was selected as the final model to be carried forward for model interpretation

and predicting GP-NPC1 binding on the extended dataset. Model interpretation was carried out with SHapley Additive exPlanations (SHAP).[65,66] SHAP assess the relationship between the value of a given feature (e.g., hydrophobicity of residue 504 in NPC1-C from *Mops condylurus*) and binding strength, demonstrating how a single feature impacts the models' output. For this, SHAP estimates the marginal contribution of each feature over all possible feature orderings and subsets, accounting for the interaction between features. The Shapley value corresponds to the weighted average of all marginal contributions. When assessing feature importance in the meta-predictor that integrates the two best-performing models, we averaged the corresponding Shapley values.

Molecular visualization and mutagenesis

Visualization of crystal structures and *in silico* mutagenesis was carried out with Pymol (Schröedinger).

Modeling the interaction between EBOV GP and NPC1-C of candidate hosts

We model the NPC1-C domains of interisualization of *Hypsignathus monstrosus*, Nycteris hispida, and Afronycteris nanus using as template the GP-bound conformation of human with Swiss-model.[134] The EBOV GP-NPC1-C complex was modeled by superimposing the modeled NPC1-C to the crystal structure of the EBOV GP-human NPC1-C complex[50] using ChimeraX.[133]

Free energy perturbation

Free Energy Perturbation (FEP)[71] was carried out to evaluate the energetic contribution of P146S EBOV GP on binding human NPC1. Schrödinger's Suite, release version 2019-4 was used to set up and run FEP. The structure of EBOV GP and human NPC1-C[50] was prepared with Protein Preparation Wizard, and Protein FEP panel was used to set up FEP calculations. Default core- and charge-hopping protocols were used to account for ring breaking involving a proline amino acid or charge mutation, respectively.[150,151] The simulation time for each mutation was 10 ns, with SPC water model and OPLS3 force field.[152]

Electrostatic calculations

Electrostatic calculations on the crystal structure of EBOV GP were performed with Delphi.[128] First, missing residues in disordered regions were modeled using Loopy.[129] Then EBOV GP was subjected to a two-step minimization using the conjugate gradient algorithm implemented in NAMD[130] and the CHARMM forcefield.[153] First, hydrogen atoms were minimized for 3,000 steps, while heavy atoms were spatially fixed. Second, backbone atoms were spatially fixed and sidechain heavy atoms were minimized with spatial constraints for 10,000 steps.

Electrostatic potentials were computed with the finite difference Poisson-Boltzmann (FDPB) method,[154] implemented in Delphi.[128] Atomic charges and radii were extracted from the CHARMM forcefield. The dielectric constant of the protein interior and the solvent were set to four and 80, respectively. The ion exclusion parameter was set to two and the ionic strength to 145mM. Electrostatic calculations were carried out using a lattice with 3.7 grids per Å and a series of focusing runs of increasing percentage fill (perfil) was performed from 20% to 85%. Calculations were iterated until they reached convergence, defined $1x10^{-4}$ as the point at which the final maximum energy change is less than 10 kTe^{-1}. Visualization of electrostatic surfaces was carried out with Pymol (Schröedinger).

Protein interface detection

EBOV GP and human NPC2 binding interfaces in human NPC1-C were estimated by calculating the accessible surface area (ASA) of the corresponding unbound and bound crystal structures.[50,76] Surface residues with a buried surface area ($ASA_{unbound}$ – ASA_{bound}) > 0 constitute the interface. Surface area was calculated with FreeSASA.[131]

Retrieval of additional NPC1 bat orthologs

We searched public repositories to retrieve the amino acid sequences for bat NPC1 orthologs not described in the experimental dataset. First, we blasted human NPC1-C against the swissprot and trembl databases from Uniprot,[136] selecting as NPC1 orthologs those with an e-value < $1x10^{-6}$. Whenever multiple hits were obtained for a given organism, the hit with the lowest e-value was selected. Second, we searched the NCBI database for NPC1 orthologs.[137] Third, we included additional NPC1 sequences reported herein. Finally, we combined the separate sets of NPC1 orthologs, and we filtered out NPC1 orthologs from non-bats and bat NPC1s already included in the experimental dataset. Finally, we chose a single representative NPC1 ortholog per bat specie. We found 56 additional bat NPC1 orthologs not described in the experimental dataset.

Selection analysis

A total of 135 NPC1 Domain C sequences were used to evaluate signatures of positive selection. Each bat species was represented by a single NPC1 sequence. A codon alignment was generated using Clustal Omega with default parameters,[127] and inspected manually to correct misalignments and ensuring the reading frame was maintained throughout. Selection analysis was then performed using the DataMonkey web server (http://www.datamonkey.org), which offers a suite of tools for detecting selection at the molecular level.[132] We used two tests of selection, FUBAR (Fast, Unconstrained Bayesian AppRoximation) to identify sites under pervasive selection, and MEME (Mixed Effects Model of Evolution) to identify episodic diversifying selection.[77,78] These tests were selected as they can both detect selection at individual sites in the alignment. For FUBAR, sites with a positive dN-dS value and a posterior probability greater than 0.90 for Prob[dN > dS] were considered to be under positive selection. Sites with a negative dN-dS

value and a posterior probability greater than 0.90 for Prob[dN < dS] were considered to be under negative selection. For MEME, sites with $p \leq 0.05$ and a dN>dS were classified as under positive selection.

Hierarchical Generalized Linear Mixed Model

To account for the hierarchical structure in our dataset, in which bat species are nested within taxonomic families as well as geographic regions of natural occurrence, we used mixed effects models with random effects to determine significant variability in host and virus binding at the levels of family and region. In a forward selection process, we considered multiple models with virus species, bat species, bat taxonomic family, and geographic region of origin as predictors, with potential inclusion of both interaction terms and nested random effects. In the regression models, we excluded bat families with < 3 species (*Megadermatidae*, *Rhinonycteridae*, *Nycteridae*, *Noctilionidae*, *Natalidae*). The final models for the experimental and predicted datasets were selected based on standard metrics for goodness of fit, including AIC minimization and analysis of residuals. The best-fitting model for the overall analysis included a random effect for bat species nested within its respective family and all other variables as independent predictors. Best-fitting models for virus-stratified analyses differed by virus species, but included bat family and region as fixed effects (with interaction terms for MARV, MLAV, and RAVV), and species as a random effect (for LLOV, SUDV, and TAFV). Models were fit using the glmmTMB package in R.[123,124]

Geographic distribution of African bats

We gathered range distribution polygons for all bats with taxonomy harmonized to the Mammal Diversity Database.[79,80] Bats were filtered to only those species that intersected with the African continent using the rnaturalearth and sf packages in R.[123,125,126,155] For each species, we calculated the percent intersection of the distribution with reported outbreaks of EBOV using the st_intersects function in the sf package. Additional geographic location for *Mops condylurus* was collected from the USAID PREDICT project[81] and the 2022 African Chiroptera Report,[82] The location for past human outbreaks of EBOV was obtained from the CDD.[83]

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical parameters, including the definition of the center, dispersion, and associated significance, are reported in the main text, figure legends, and supplementary material. We have used Kruskal Wallis, one-way ANOVA and two-way ANOVA tests as well as Hierarchical Generalized Linear Mixed Models to determine significance. Statistical analysis were carried out in PRISM (GraphPad) and R.[123] Whenever appropriate, p values were adjusted for multiple comparisons. Data are statistically significant when $p \leq 0.05$ (Levels of significance: *$0.01 < p \leq 0.05$; **$0.001 < p \leq 0.01$; ***$0.0001 < p \leq 0.001$; ****$p \leq 0.0001$).