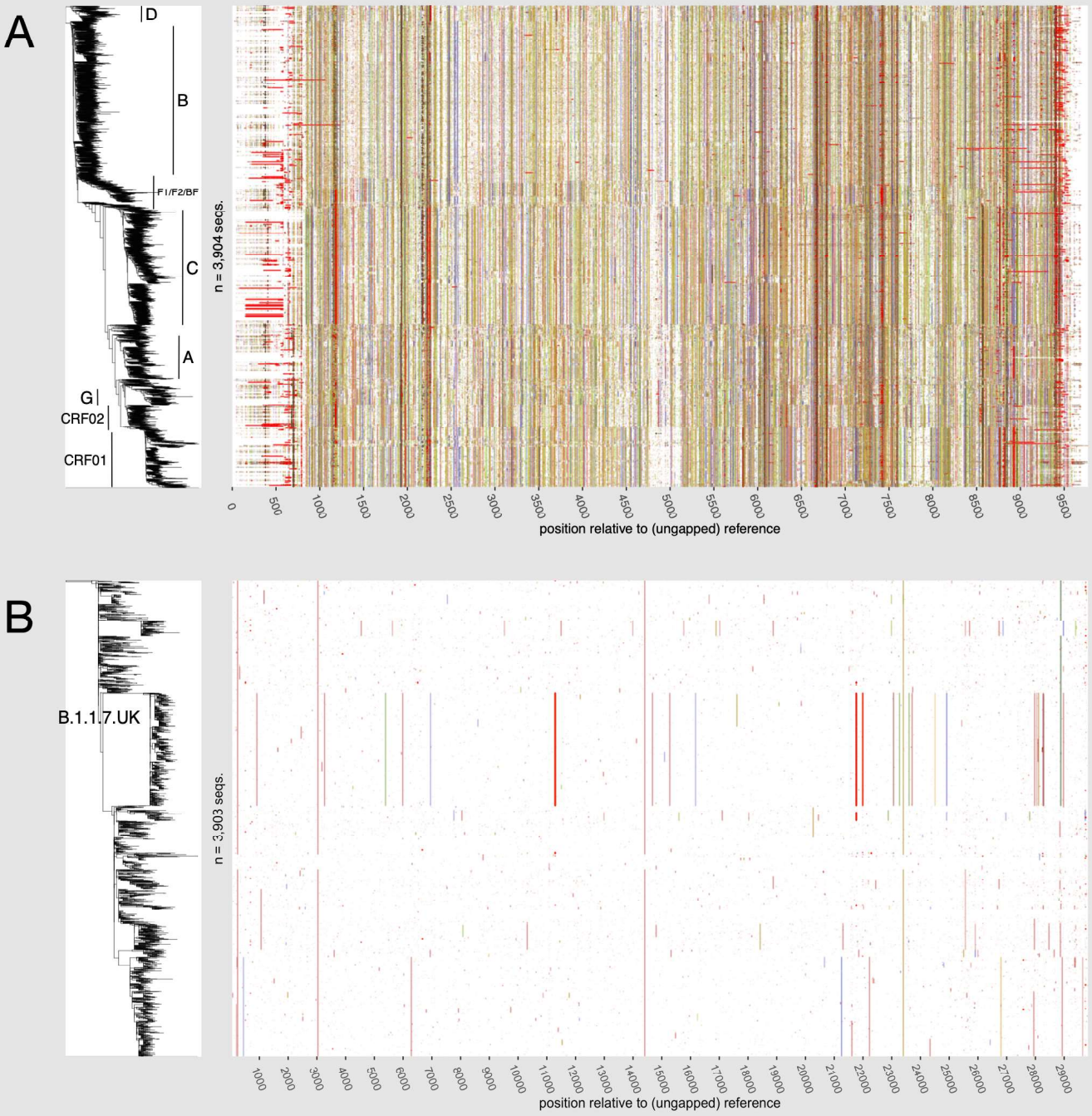
Cell Host & Microbe, Volume *29*

Supplemental information

HIV-1 and SARS-CoV-2: Patterns

in the evolution of two pandemic pathogens

Will Fischer, Elena E. Giorgi, Srirupa Chakraborty, Kien Nguyen, Tanmoy Bhattacharya, James Theiler, Pablo A. Goloboff, Hyejin Yoon, Werner Abfalterer, Brian T. Foley, Houriiyah Tegally, James Emmanuel San, Tulio de Oliveira, Network for Genomic Surveillance in South Africa (NGS-SA), Sandrasegaram Gnanakaran, and Bette Korber


A
D
B
F1/F2/BF
C
A
G
CRF02
CRF01
n = 3,904 secs.
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
5500
6000
6500
7000
7500
8000
8500
9000
9500
position relative to (ungapped) reference
B
B.1.1.7.UK
n = 3,903 seqs.
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
11000
12000
13000
14000
15000
16000
17000
18000
19000
20000
21000
22000
23000
24000
25000
26000
27000
28000
29000
position relative to (ungapped) reference
C
A.23.1.UG
B.1.429.US-CA
B.1.1.7.UK
B.1.1.248.BR
B.1.526.US-NY
B.1.352.ZA
n = 599 seqs.
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
11000
12000
13000
14000
15000
16000
17000
18000
19000
20000
21000
22000
23000
24000
25000
26000
27000
28000
29000
position relative to (ungapped) reference
replacements
A to -
C to -
G to -
T to -
- to A
C to A
G to A
T to A
- to C
A to C
G to C
T to C
- to G
A to G
C to G
T to G
- to T
A to T
C to T
G to T

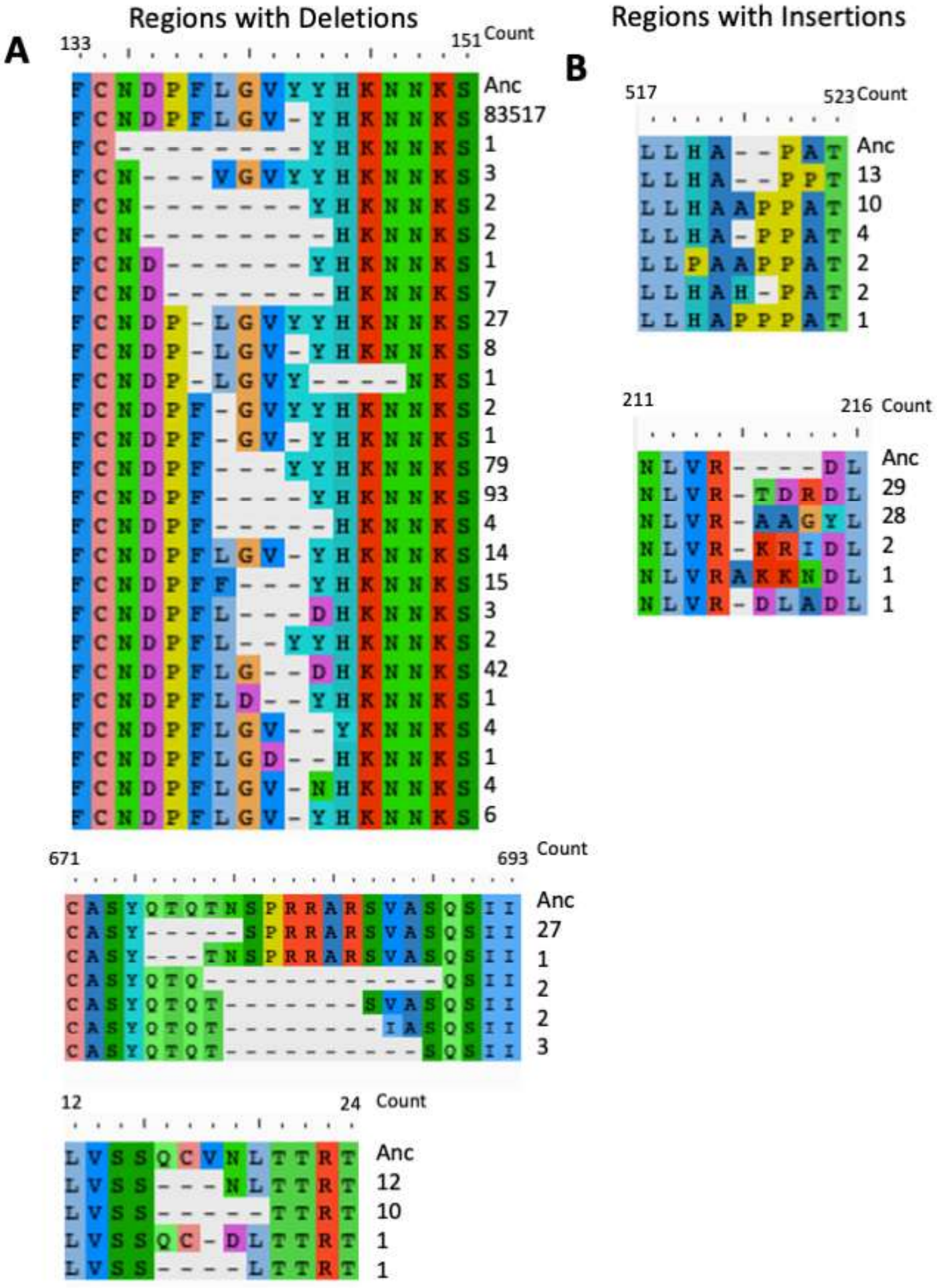

A
Regions with Deletions
133
151
Count
Anc
83517
1
3
2
2
1
7
27
8
1
2
1
79
93
4
14
15
3
2
42
1
4
1
4
6
671
693
Count
Anc
27
1
2
2
3
12
24
Count
Anc
12
10
1
1
B
Regions with Insertions
517
523
Count
Anc
13
10
4
2
2
1
211
216
Count
Anc
29
28
2
1
1

Figure legends

Figure S1. Variability of HIV-1 and SARS-CoV-2 genomes at the nucleotide level. This figure is associated with Figure 1 in the main text.

Right panel: Variant-visualized whole-genome sequence alignments of HIV-1 (a), and SARS-CoV-2 (b). The colored panels are a matrix where each row represents a single sequence, and the columns are positions in a sequence alignment, where colored marks ("+") denote positions that vary compared to a reference sequence. Reference-identical positions are shown as white. A "plurality" consensus sequence, with the most common alignment base at each position, serves as the reference for HIV-1; the outbreak strain (NC 045512) is the reference sequence for SARS-CoV-2. Nucleotide changes are colored based on the mutant base. Sequences are ordered top-to-bottom according to the phylogenetic tree in the left panel; consequently, continuous vertical stripes indicate lineage-specific mutations that are shared by related sequences (see text). The trees in the left panel are in each case derived from a whole-genome nucleotide alignment: an approximation to the maximum-likelihood tree, generated with RAxML-NG (Kozlov et al., 2019) for HIV-1 Env, and for SARS-CoV-2 Spike, a neighbor-joining tree inferred with PAUP (Swofford, 2003) using Log-Determinant (logdet) distances with uninformative sites removed (see Swofford et al., 1996, pp. 459-462), saved with parsimony branch lengths.

Figure S2. Rare but repeated indel patterns in SARS-CoV-2. This figure is associated with Figure 2B in the main text.

Five short regions of the Spike protein are shown, with their positions in Spike relative to the Ancestral Wuhan reference sequence (NC_045512) indicated above each the alignment. All indels are aligned to the reference sequence, indicated as "Anc" for Ancestor in the figure. Regions with deletions are shown on the left (A), and with insertions on the right (B). Most of these indels are rare (the counts are from the GISAID sample of 487,073 Spike sequences available in the cov.lanl.gov alignment on 2/25/2021) but they all come in several forms and are found repeated in several geographic regions, and recur, so are likely to be viable. (A, upper)

Unique repeated deletions patterns in the region between positions 133-151. There were many Spike sequence patterns of deletions in this region, the only common one being the loss of Y144, Δ144.This deletion was found 82,017 times in the context of the B.1.1.7 lineage, and over 1,500 times in the context of other variants and lineages. There was a total of 313 additional sequences with other deletion patterns in this region. To illustrate this, we included a representative example of each distinct deletion pattern between positions 133-151 in Spike, and include a count of how many times the form was repeated in the data set. (A, middle) Repeated deletion patterns in the region between positions 671-693. This small set of deletion patterns includes Spike sequences in which the furin cleavage site motif, RRxR, is deleted (Walls et al., 2020). These may represent sequences from cultures or viruses that are attenuated in vivo (Johnson et al., 2021). (A, lower) Repeated deletion patterns near the Spike signal peptide cleavage site between positions 12 and 24. . (B, Upper) The alignment between positions 517-523. Additional prolines are occasionally added to Spike near 520/521, which are located at the end of the RBD. In the top case (shown in the row beneath Anc), we found a repeated A522P change, no insertion, but a distinctive way to add two prolines in this local domain, so we included it here. In the other cases, either a single P, or a two-amino-acid duplication, AP, were introduced. The insertions in this region were all sampled from one source in Houston, Texas. (B. Lower) Distinctive repeated insertions found between Spike positions 211-216. These insertions are carried in lineages that are rare but increasing in frequency. In the May 12 sampling of GISAID, the insertion 214 TDR, which is carried in the lineage B.1.214.2, was found in 12 countries, and had been sampled 602 times. The insertion 215 AGG, which is carried in the lineage A.2.5.2, was found in 7 countries and sampled 122 times. Note that additional indel patterns have been recently observed in these regions and others that are discussed in the text, including several distinctive deletions patterns near the 242-244 deletion commonly found in the B.1.351 variant, the deletion at 156-157 found in B.1.617.2, and a four amino acid insertion near the furin cleavage site.

Table S1: This table is associated with Figure 1 in the main text.

Estimated mutational parameters based on maximum likelihood trees generated using RAxML-NG (Kozlov et al., 2019), with a general time-reversible model with rate categories estimated using an 8-category Gamma rate distribution (GTR+G8).

	HIV-1	SARS-CoV-2
Base frequencies:		
A	0.4456	0.2900
C	0.1741	0.1757
G	0.1974	0.1626
T	0.1828	0.3716
Exchangeabilities (R):		
AC	1.0546	0.2127
AG	3.0375	0.6778
AT	0.5612	0.1490
CG	0.7638	0.3050
CT	4.2699	2.1128
GT	1.0000	1.0000
Gamma shape	0.4804	0.2989

Network for Genomic Surveillance in South Africa (NGS-SA) author list and affiliations

Eduan Wilkinson, Nokukhanya Msomi[2], Arash Iranzadeh[4], Vagner Fonseca[1], Deelan Doolabh[5], Emmanuel James San[1], Koleka Mlisana[7,8], Anne von Gottberg[9,10], Sibongile Walaza[9,11], Mushal Allam[9], Arshad Ismail[9], Thabo Mohale[9], Allison J Glass[10,12], Susan Engelbrecht[13], Gert Van Zyl[13], Wolfgang Preiser[13], Francesco Petruccione[14,15], Alex Sigal[16,17,18], Diana Hardie[19], Gert Marais[19], Marvin Hsiao[19], Stephen Korsman[19], Mary-Ann Davies[20,21], Lynn Tyers[5], Innocent Mudau[5], Denis York[22], Caroline Maslo[23], Dominique Goedhals[24], Shareef Abrahams[25], Oluwakemi Laguda-Akingba[25,26], Arghavan Alisoltani-Dehkordi[27,28], Adam Godzik[28], Constantinos Kurt Wibmer[9], Bryan Trevor Sewell[29], José Lourenço[30], Sergei L Kosakovsky Pond[31], Steven Weaver[31], Marta Giovanetti[32], Luiz Carlos Junior Alcantara[32],Darren Martin[4,5], Jinal N Bhiman[9,10], Carolyn Williamson[5,8,19]

Affiliations:

[1] KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP), Department of Laboratory Medicine & Medical Sciences, University of KwaZulu-Natal, Durban, South Africa

[2] Discipline of Virology, University of KwaZulu-Natal, School of Laboratory Medicine and Medical Sciences and National Health Laboratory Service, Durban, South Africa

[4] Computational Biology Division, Department of Integrative Biomedical Sciences, University of Cape Town, Cape Town, 7925, South Africa

[5] Division of Medical Virology, Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa

[7] National Health Laboratory Service, Johannesburg, South Africa

[8] Centre for the AIDS Programme of Research in South Africa (CAPRISA), Durban, South Africa

[9] National Institute for Communicable Diseases of the National Health Laboratory Service, Johannesburg, South Africa

[10] School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

[11] School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

[12] Department of Molecular Pathology, Lancet Laboratories, Johannesburg, South Africa

[13] Division of Medical Virology at NHLS Tygerberg Hospital and Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

[14] Centre for Quantum Technology, University of KwaZulu-Natal, Durban, South Africa [15] National Institute for Theoretical Physics (NITheP), KwaZulu-Natal, South Africa

[16] Africa Health Research Institute, Durban, South Africa

[17] School of Laboratory Medicine and Medical Sciences, University of KwaZulu-Natal, Durban, South Africa

[18] Max Planck Institute for Infection Biology, Berlin, Germany

[19] Division of Medical Virology at NHLS Groote Schuur Hospital, University of Cape Town, Cape Town, South Africa

[20] Centre for Infectious Disease Epidemiology and Research, University of Cape Town, Cape Town, South Africa

[21] Western Cape Government: Health, Cape Town, South Africa

[22] Molecular Diagnostics Services, Durban, South Africa

[23] Department of Quality Leadership, Netcare Hospitals, Johannesburg, South Africa

[24] Division of Virology at NHLS Universitas Academic Laboratories, University of The Free State, Bloemfontein, South Africa

[25] National Health Laboratory Service, Port Elizabeth, South Africa

[26] Department of Laboratory Medicine and Pathology, Faculty of Health Sciences, Walter Sisulu University, Mthatha, South Africa

[27] Division of Medical Virology, Department of Pathology, University of Cape Town, Cape Town, South Africa

[28] Division of Biomedical Sciences, University of California Riverside School of Medicine, Riverside, California, USA

[29] Structural Biology Research Unit, Department of Integrative Biomedical Sciences, University of Cape Town, Rondebosch, South Africa

[30] Department of Zoology, University of Oxford, Oxford, United Kingdom

[31] Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, Pennsylvania, USA

[32] Laboratorio de Flavivirus, Fundacao Oswaldo Cruz, Rio de Janeiro, Brazil