

RESEARCH ARTICLE

Geospatial HIV-1 subtype C gp120 sequence diversity and its predicted impact on broadly neutralizing antibody sensitivity

Jyoti Sutar^{1,2}, Suprit Deshpande¹, Ranajoy Mullick^{1,2}, Nitin Hingankar¹, Vainav Patel³, Jayanta Bhattacharya^{1,2*}

1 HIV Vaccine Translational Research Laboratory, Translational Health Science & Technology Institute, NCR Biotech Science Cluster, Faridabad, Haryana, India, **2** International AIDS Vaccine Initiative, New Delhi, India, **3** ICMR-National Institute for Research in Reproductive Health, Mumbai, India

* JBhattacharya@iavi.org, JBhattacharya@thsti.res.in



OPEN ACCESS

Citation: Sutar J, Deshpande S, Mullick R, Hingankar N, Patel V, Bhattacharya J (2021) Geospatial HIV-1 subtype C gp120 sequence diversity and its predicted impact on broadly neutralizing antibody sensitivity. PLoS ONE 16(5): e0251969. <https://doi.org/10.1371/journal.pone.0251969>

Editor: William Anderson Paxton, University of Liverpool Institute of Infection and Global Health, UNITED KINGDOM

Received: December 11, 2020

Accepted: May 6, 2021

Published: May 24, 2021

Copyright: © 2021 Sutar et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its [Supporting Information](#) files.

Funding: The authors wish to acknowledge the funding support from the Wellcome Trust/DBT India Alliance Team Science Grant (IA/TSG/19/1/600019), Science & Engineering Research Board, Department of Science & Technology, Government of India (CRG/2019/002939) and Department of

Abstract

Evolving diversity in globally circulating HIV-1 subtypes presents a formidable challenge in defining and developing neutralizing antibodies for prevention and treatment. HIV-1 subtype C is responsible for majority of global HIV-1 infections. In the present study, we examined the diversity in genetic signatures and attributes that differentiate region-specific HIV-1 subtype C *gp120* sequences associated with virus neutralization outcomes to key bnAbs having distinct epitope specificities. A total of 1814 full length HIV-1 subtype C *gp120* sequence from 37 countries were retrieved from Los Alamos National Laboratory HIV database (www.hiv.lanl.gov). The amino acid sequences were assessed for their phylogenetic association, variable loop lengths and prevalence of potential N-linked glycosylation sites (pNLGS). Responses of these sequences to bnAbs were predicted with a machine learning algorithm 'bNAb-ReP' and compared with those reported in the CATNAP database. Subtype C sequences from Asian countries including India differed phylogenetically when compared with that from African countries. Variable loop lengths and charges within Indian and African clusters were also found to be distinct from each other, specifically for V1, V2 and V4 loops. Pairwise analyses at each of the 25 pNLG sites indicated distinct country specific profiles. Highly significant differences ($p < 0.001$ ***) were observed in prevalence of four pNLGS (N130, N295, N392 and N448) between South Africa and India, having most disease burden associated with subtype C. Our findings highlight that distinctly evolving clusters within global intra-subtype C *gp120* sequences are likely to influence the disparate region-specific sensitivity of circulating HIV-1 subtype C to bnAbs.

Importance

Broadly neutralizing antibodies that act on viral envelope protein and prevent infection have emerged as an attractive choice for HIV-1 treatment and prophylaxis over and above existing antiretroviral drug therapy. While HIV-1 subtype C accounts for nearly half of the global

Biotechnology, Government of India (BT/PR24520/MED/29/1222/2017). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

infection, the degree of intra-clade C *gp120* variability across different geographical boundaries is poorly understood. Through systematic computational approach, we compared the *gp120* sequences that represent globally circulating HIV-1 subtype C towards accurately predicting their degree of susceptibility to most broad and potent neutralizing monoclonal antibodies having different target specificities across *gp120*. Our data not only revealed significant intra-clade C diversity but also highlighted neutralization diversities. These observations provide insights that warrant greater and continued surveillance of evolving HIV-1 globally circulating subtype C variants that will inform selection of bnAbs that will best dissect the region-specific genetic and neutralization diversities towards selecting appropriate bnAb combination for clinical use.

Introduction

The extraordinary diversity of *env* targeting neutralizing antibodies is a barrier to achieving the desired vaccine-induced and antibody-mediated protection. Evolving antigenic diversity in global and region-specific circulating HIV-1 subtypes is complex which not only poses significant roadblocks to developing preventive vaccine but also poses a challenge in neutralizing antibody mediated prophylaxis and treatment. Broadly neutralizing antibodies (bnAbs) act solely on the HIV-1 envelope glycoprotein (Env) for neutralizing genetically distinct HIV-1 subtypes. Till date, a number of bnAbs have been discovered from elite neutralizers and a few of them have been found to be able to prevent acquisition as well as significantly reduce plasma viral loads, when tested both in animal models and humans [1–3], thus justifying their importance as products for prevention and treatment. Some of the bnAbs that are currently being evaluated through human clinical trials will provide additional possibilities for prevention [4, 5], such as their extent, in addition to virus neutralization, in persistent viral clearance [6, 7] and eliminating HIV-1 infected cells [2, 8, 9]. While some single bnAbs have been found to show significant breadth across subtypes, combination of bnAbs with distinct epitope-specificity is believed to provide the most effective response against globally diverse HIV-1 subtypes and also would likely prevent the development of antibody-escape variants [2]. The trimeric Env glycoproteins on the virus surface are the most diverse of all proteins encoded by HIV-1; which differs by greater than 20% of amino acids between matched subtypes [10–14] and which continues to diversify at a population level [14–17]. This has been substantiated by observation that several epitopes that are targeted by different bnAbs have been found to vary over time [15]. While bnAbs target both surface *gp120* and membrane proximal external region (MPER) of *gp41*, *gp120* exhibits extraordinary sequence divergence compared to *gp41* MPER and variation in this region is believed to represent distinct genetic subtypes, or clades, which are prevalent in distinct geographic regions [18, 19]. Although bnAbs isolated from individuals infected with one particular subtype are generally effective at neutralizing viruses belonging to other subtypes, antibody potency is often found to be correlated with matched subtypes as described elsewhere [14, 20–24]. Moreover, diversity has been found to have an impact even within matched subtypes, as demonstrated by the fact that the subtype-matched neutralization advantage was more apparent in regions with distinct viral diversities [14]. HIV-1 subtype C accounts for approximately half of the global infections [25], which predominates in India and South Africa. Yet, robust, comparative *env* sequence diversity and evolution analysis, critical for bnAb based intervention strategies in these regions is severely lacking [26]. While recent development of bnAbs has considerably improved our knowledge on conserved epitopes that they target, and Env structure associated with broad and potent virus neutralizing antibodies, a greater understanding of the antigenic diversity of global HIV-1 subtype C Env would facilitate understanding potential bnAb combination that could overcome the intra-

clade C diversity. In the present study, we examined the variation in signature sequences, loop length, N-linked glycosylation and key epitopes within the existing globally circulating HIV-1 subtype C *gp120* targeted by potent bnAbs and predicted their impact on virus neutralization.

Results

Evidence of phylogenetic divergences of globally circulating HIV-1 subtype C *gp120* sequences

Previous studies have demonstrated that HIV-1 subtype C is the most abundant globally circulating subtype responsible for approximately 46.6% of all HIV infections [27]. In the present study, we retrieved from HIV database (www.hiv.lanl.gov), a total of 23750 sequences covering the complete *gp120* gene (HXB2 coordinates: 6225–7758). To mitigate the intra-individual quasispecies bias, we applied ‘one sequence/individual filter’ retaining 1927 full length sequences. As shown in Table 1, there is uneven distribution of *gp120* sequences from different countries. Of the 37 countries analyzed, South Africa (ZA) was found to contribute in the database (www.hiv.lanl.gov) more than 1000 sequences, while Malawi (MW), Botswana (BW) and Zambia (ZM) contributed between 100 to 1000 sequences. Interestingly, complete *gp120* sequences were found from only 84 unique individuals from India (IN), while all other countries contributed approximately 50 or less sequences. Upon removal of identical sequences as well as those harbouring internal stop codons, a total of 1814 full length HIV-1C *gp120* protein sequences were retained for further analysis. We next analyzed the phylogenetic properties of the subtype C *gp120* sequences that uniquely represent globally circulating subtype C across different geographies. Average number of amino acid differences per site between sequences grouped by source country were estimated by MEGA [28]. Sequences from Asian Countries (IN, China (CN) and Nepal (NP)) were observed to be closer to each other (0.183–0.208) compared to African countries (ZA and MW: 0.207–0.234). This trend continued in the maximum

Table 1. Details of country-wise HIV-1 subtype C sequences retrieved from LANL-HIV database.

Country Name	Country code	No of sequences
South Africa	ZA	1015
Malawi	MW	210
Zambia	ZM	191
Botswana	BW	121
India	IN	84
United Republic of Tanzania	TZ	53
Sweden	SE	45
United Kingdom	GB	31
China	CN	24
Ethiopia	ET	24
Nepal	NP	23
Brazil	BR	22
Kenya	KE	13
Cyprus	CY	11
Others*	-	60

* - Republic of Angola (AO): 1, Belgium (BE): 2, Bulgaria (BG): 1, Burundi (BI): 8, Germany (DE): 1, Djibouti (DJ): 2, Denmark (DK): 1, Spain (ES): 6, Finland (FI): 5, France (FR): 3, Georgia (GE): 1, Gambia (GM): 2, Israel (IL): 4, Italy (IT): 1, Myanmar (MM): 1, Senegal (SN): 3, Somalia (SO): 1, Thailand (TH): 2, Uganda (UG): 1, United States (US): 7, Uruguay (UY): 1, Yemen (YE): 1, Zimbabwe (ZW): 1, No Country: 4

<https://doi.org/10.1371/journal.pone.0251969.t001>

likelihood tree constructed as indicated in Fig 1A, wherein sequences from Asian countries (IN, CN and NP) were observed to form a unique sub-cluster. To further validate these observations statistically, a subset of sequences (N = 251) forming the aforementioned node was re-analysed with 1000 ultrafast bootstrap replicates and SH-aLRT. As indicated in Fig 1B, sequences from Asian countries clustered together with 97% bootstrap and 92.1% SH-aLRT support. They also clustered distinctly from African sequences supported by 84.5% SH-aLRT and 79% ultrafast bootstrap replicates.

Geography based subtype C *gp120* loop length and charge variation

Changes in the lengths of variable loops within *gp120* has been previously documented to be an evasion mechanism of HIV-1 to escape neutralizing antibody driven humoral responses [29–31]. Therefore, we next assessed lengths of variable regions between sequences from different countries. To prevent low sampling bias, we included sequences from six countries in this analysis that could contribute more than 50 sequences each (Fig 2A). Comparison of variable region loop (V1–V5) lengths with Kruskal Wallis non-parametric test indicated several statistically significant differences between sequences from diverse countries. For the V1 loop, sequences from UK (GB) had significantly lower length (median:16) compared to all other countries (median range: 23–27). This trend continued in V2 loop with lower length in UK sequences (median: 34) compared to rest of the countries (median range: 42–48). Additionally, sequences from BW and NP were observed to be significantly distinct. However, this was an effect of outlier values and minimal standard deviation in the loop lengths respectively. These

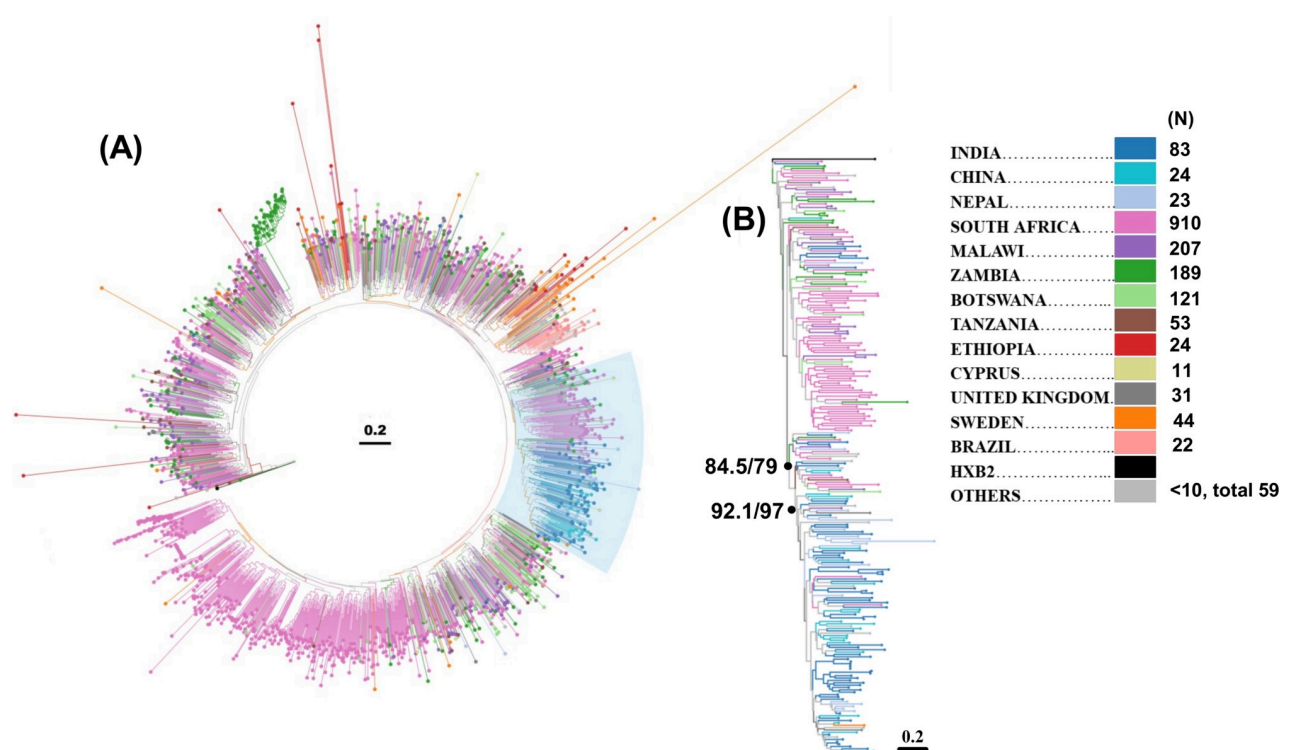


Fig 1. Phylogenetic Analysis of HIV-1 subtype C *gp120* amino acid sequence. A. A maximum likelihood tree depicting phylogenetic association of 1837 HIV-1 subtype C amino acid sequences depicted radially. The legend describes the country codes as well as the sequence distribution among the countries. B. A maximum likelihood subtree detailing phylogenetic association between sequences from South Africa and Those from India, China and Nepal. The black dots indicate nodes with corresponding SH-aLRT and bootstrap values.

<https://doi.org/10.1371/journal.pone.0251969.g001>

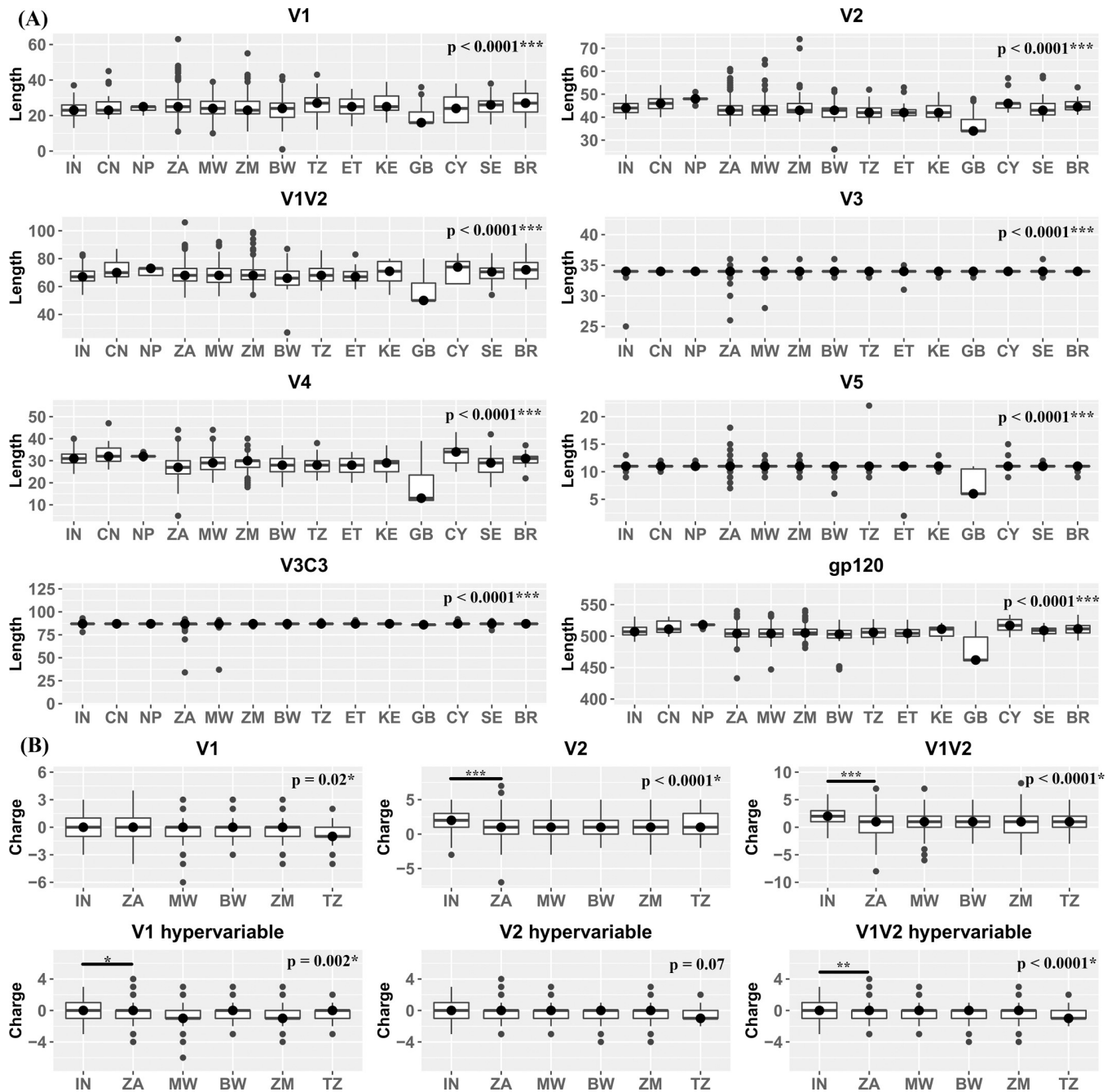


Fig 2. Assessment of Variable region characteristics. A. variable region length: gp120 variable region (V1, V2, V1+V2, V4, V5) as well as entire gp120 lengths have been plotted on the Y axis against the Countries of origin indicated on the X axis. B. Variable region charge: gp120 variable region Charges for V1, V2 and V1+V2 as well as hypervariable regions within them have been plotted on the Y axis against the Countries of origin indicated on the X axis. P values have been indicated following a statistical analysis by Kruskal-Wallis test followed by Dunn’s multiple comparison.

<https://doi.org/10.1371/journal.pone.0251969.g002>

differences remained consistent when V1V2 loop were taken together indicating no loop length compensation. The V3 loop length was observed to be highly conserved across all the countries with some outlier values in South Africa. Interestingly, V4 loop length was observed to be most variable among the variable domains. Sequences from UK were observed to have significantly shorter V4 loop (Median: 13) compared to rest of the countries (Median range:

27–34). Of note, sequences from India (IN) had longer V4 loop (Median: 31) than those from South Africa (Median: 27). Sequences from African countries (BW, ET, KE, MW, TZ, ZA and ZM) had shorter V4 loop length (Median range: 27–30) compared to the rest of the countries (Median range: 29–34) except UK. Similar to the V3 loop, V5 loop length was conserved across all the countries under consideration (Median: 11) except UK (median: 6). Overall, length of the gp120 protein was observed to be shortest in sequences from GB (Median: 462) and longest in those reported from NP (Median: 518). Consistent with the V4 loop observations, gp120 length was observed to be lower (Median range: 503–511) in countries from Africa (BW, ET, KE, MW, TZ, ZA and ZM) compared to rest of the countries (Median range: 507–518) except UK. A subset of six countries with more than 50 sequences (IN, ZA, MW, TZ and BW) was further selected for assessment of sequence charge distribution (Fig 2B). V1, V1-hypervariable region, V2, V2 hypervariable region as well as V1V2 cumulatively were found to have significantly different distributions. Of note, V1V2 hypervariable region charge was distinct in India as compared to South Africa.

Comparison of abundance of potential N linked glycosylation sites (pNLGs)

HIV-1 gp120 is a heavily glycosylated protein with host derived N-linked glycans making up ~50% of its total mass. These glycans play an important role in ensuring viral infectivity as well as evading neutralizing antibodies [32] emphasizing the importance of their assessment. In the present study, proportion of pNLGs were compared between sequences from 14 countries (Table 1). A median of 25 pNLGs were observed in the sequences (range: 11 to 33, 10–90 percentile: 22–28). There were no overall significant differences between the average number of pNLGs between countries (Kruskal Wallis test, $p > 0.05$). Out of a total of 3003 pairwise analyses of each pNLG position for all possible two country combinations performed with Fisher's exact test, 342 combinations were found to be statistically significant (Fisher's exact test, $p < 0.01$) involving 31 of the 33 pNLG sites. N88, N156, N160, N197, N276, N 301 and N386 were observed to be highly conserved across all countries with >70% abundance. N301 (abundance: 89–100%) was observed to be the most conserved site which plays a critical role in various envelope functions including membrane fusion and thus is highly conserved [33]. Fig 3 represents abundance at the 25 well characterized pNLG sites denoted along with their functional domains. As apparent in Fig 3 as well as through statistical comparisons, there was no significant difference between glycosylation profiles at any of the sites between India, China and Nepal. Similar observations were also made for a cluster of African countries South Africa, Tanzania, Kenya and Malawi. Furthermore, sequences from Ethiopia and Sweden were also observed to have similar NLG profiles. Upon comparison of sequences from India and south Africa, highly significant differences (Fisher's exact test, $p < 0.001^{***}$) were observed in 4 pNLG sites (N130, N295, N392 and N448), which are present in C1, C2, V4 and C4 domains respectively. These sites are important for integrity of the 'mannose patch' and interaction with several bnAbs such as 2G12, VRC-PG05, PGT135 and PGT151 [32].

Variation in Shannon entropy indicate significant intra-subtype C diversity

Shannon entropy is a measure of variability wherein higher entropy indicates higher variability. Entropy values were predicted with Entropy-One tool available through LANL HIV database. To assess the entropy differences associated with bnAb contact sites, we plotted entropy data for Subtype C overall (all sequences), South Africa, India, Malawi, Botswana, Zambia and Tanzania (Fig 4A). To prevent bias because of hypervariable nature of certain gp120 domains, the residue positions present in hypervariable regions were removed from the subsequent

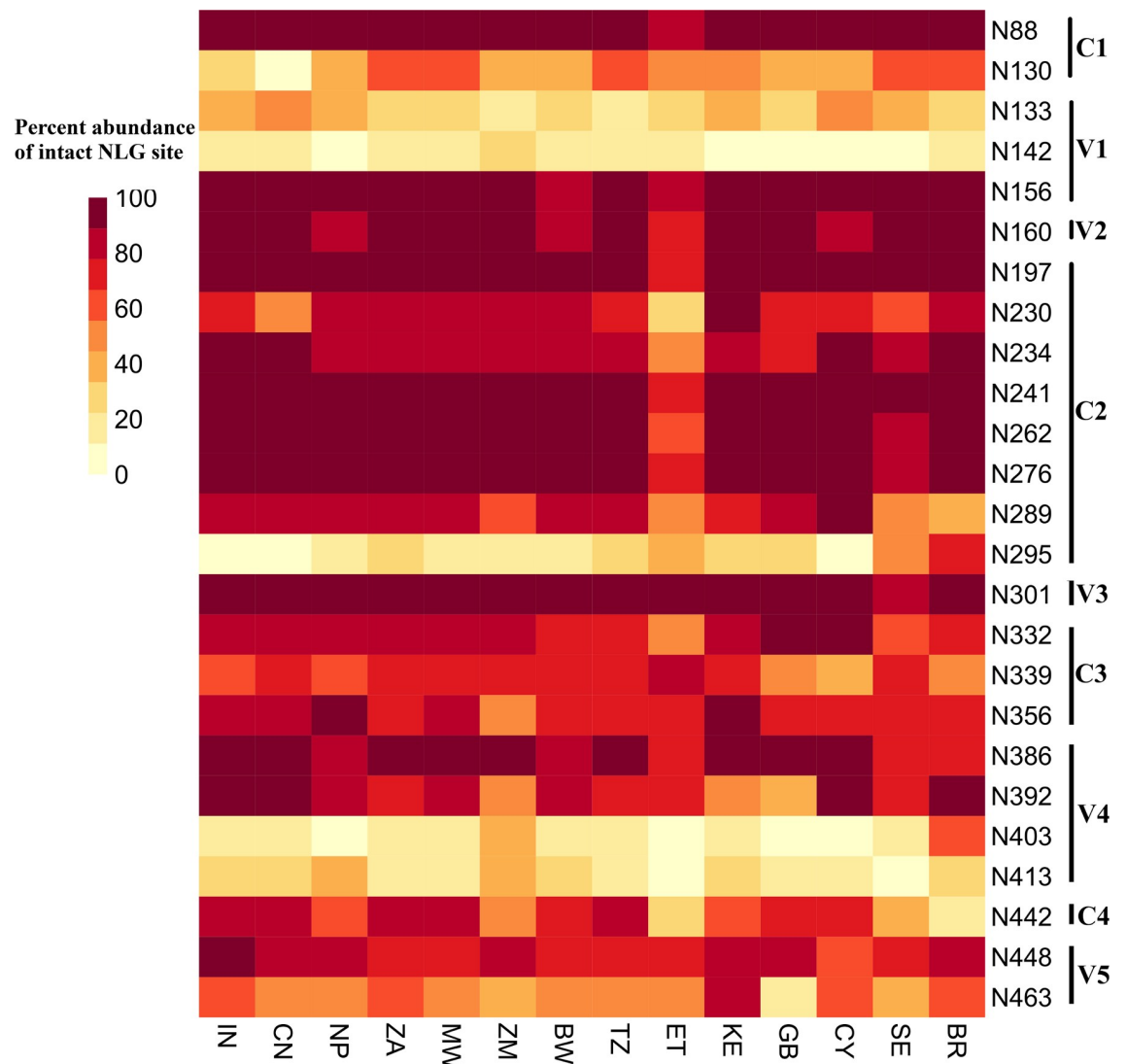


Fig 3. Assessment of potential N-linked glycosylation sites. A heatmap comparison of abundance of pNLG sites plotted on Y axis against countries plotted on X axis. Each pixel represents 1 pNLG site data from 1 country. Specific domains of pNLGs have been indicated along the Y axis. Color key represents correlation of color intensity with abundance of pNLGs ranging from 0 to 100.

<https://doi.org/10.1371/journal.pone.0251969.g003>

comparisons. While no difference was observed in entropy profiles at key bnAb sites between overall subtype C entropy and that observed in South Africa (Mann-Whitney test, $p = 0.0759$), corresponding entropy profile in India was significantly different both from overall subtype C ($p < 0.0001$) and South Africa ($p < 0.0001$). Difference between entropy profiles of South Africa and India were not significantly different in pairwise comparison with Malawi, Botswana, Zambia and Tanzania (Mann-Whitney test, $p > 0.05$). To further assess the entropy differences, Indian sequence ‘query’ data set ($N = 83$) was compared against South African sequence ‘background’ data set ($N = 910$) through Entropy-TWO tool on LANL HIV database which generated Shannon entropy values with statistical confidence measured through Monte-Carlo randomization with 100 replacements. Overall, 133 amino acid positions across gp120 were detected to have differential entropy between India and South Africa, of which 83 sites had higher entropy in South Africa while 50 sites had higher entropy in India (Fig 4B). As

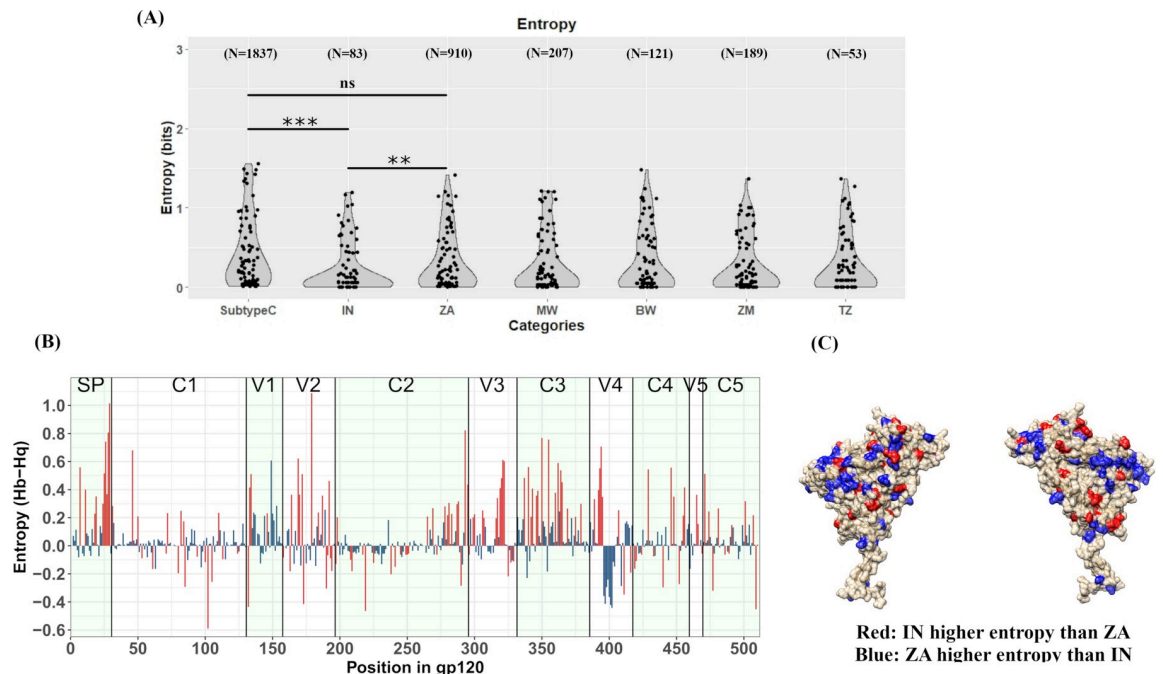


Fig 4. Entropy analysis. A. Shannon entropy (bits) at key sites for bnAbs VRC01, VRC03, VRC07, VRC13, CAP256:VRC26.25, PGDM1400, PG9, PG16, PGT121 and PGT128 excluding positions in hypervariable regions have been plotted for overall Subtype C, South Africa (ZA), India (IN), Malawi (MW), Botswana (BW), Zambia (ZM) and Tanzania (TZ). Statistical comparison of entropy distribution through p values has been performed following application of Mann-Whitney test. B. Shannon Entropy difference (H (background)- H (query) (unit: bits) has been plotted on Y axis against each amino acid position on X axis, where ZA dataset was the background while IN dataset was the query. Different domains of gp120 have been indicated. Bars with red color indicate positions with statistically significant entropy differences. Bars above 0 indicate higher entropy in South Africa while those below 0 indicate higher entropy in India. C. Variable entropy positions are plotted on prefusion gp120 envelope model derived from PDB:5U7O. Residue position highlighted in red indicate statistically significantly higher entropy in India compared to South Africa while those highlighted in blue indicate statistically significantly higher entropy in South Africa compared to India (C).

<https://doi.org/10.1371/journal.pone.0251969.g004>

indicated in Fig 4C, many of the differential entropy sites are located in the surface accessible V1-V2 and V3 regions, also targeted by several bnAbs. These observations suggest a differential pattern of conservation across the two populations and thus may result in variable breadth of neutralization by these bnAbs [34].

Variation in abundances of epitopes associated with resistance and susceptibility to bnAbs

Next, we examined the abundance of HIV-1 subtype C resistance phenotype as defined in the CATNAP database of key bnAbs having varied epitope specificities in gp120 across different countries. We assessed sequence datasets of different counties (www.hiv.lanl.gov) and calculated the frequency of key amino acid residues associated with the sensitivity and resistance to individual bnAbs. We examined the select key bnAbs targeting CD4bs (VRC01, VRC07, 3BNC117 and N6), V1V2 region (PG9, PG16, PGT145, PGDM1400 and CAP256:VRC26.25) and V3 supersite (PGT121, PGT128 and 10–1074). As shown in Fig 5, we found considerable variation in the abundance of amino acid residues that form epitopes associated with neutralization resistance to bnAbs with unique specificity. Variation in abundances of the following residues that form epitopes to bnAbs with unique specificities were observed in geographically divergent globally circulating HIV-1 subtype C (Fig 5). Furthermore, we performed Fisher's test for pairwise comparisons for resistance associated residue abundance at each of the contact sites for

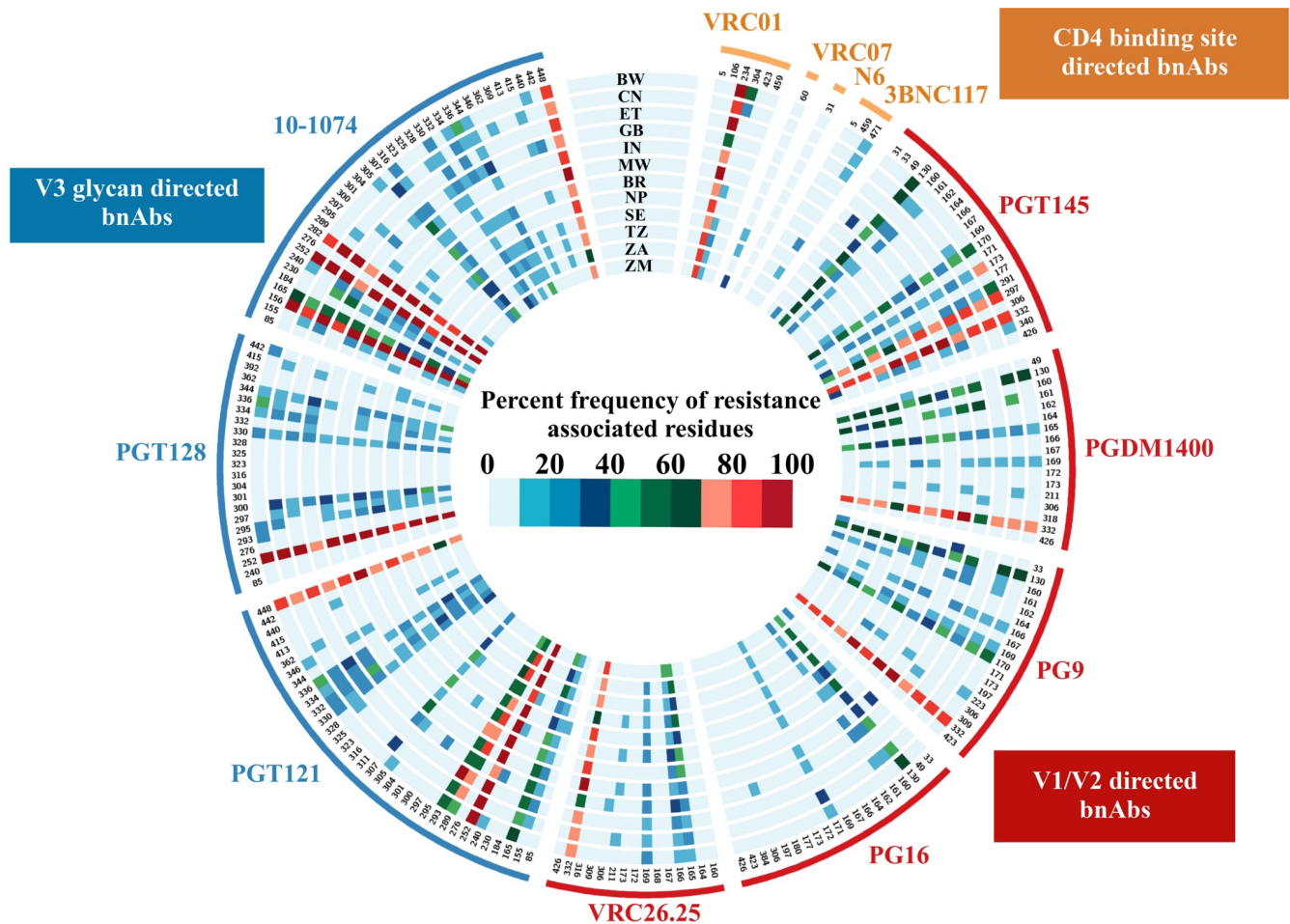


Fig 5. Abundance of bnAb resistance associated residues. A circular heatmap depicting abundance of bnAb resistance associated residues was plotted for 11 bnAbs (VRC01, VRC07, PGT121, PGT128, PGT145, PG9, PG16, VRC26.25, 3BNC117, 10-1074 and N6) wherein each track indicates the country of origin. Each pixel on the circular track indicates a specific residue position colored as per abundance of resistance causing residues at that position as per the color key.

<https://doi.org/10.1371/journal.pone.0251969.g005>

every bnAb across all countries (BW, CN, ET, GB, IN, MW, BR, NP, SE, TZ, ZA and ZM). For CD4bs directed bnAbs: variation in the sensitivity to global HIV-1 subtype C to VRC01 was found to be majorly associated with N234 and S364. Similarly, for V1V2 directed bnAbs: K130, I161, Q170, Y173, S291, T297, N332 and N340 associated with significant variation in PGT145 sensitivity; K130, I161, I165, V169 and N332 (most significant) associated with significant variation in PGDM1400 sensitivity; K130, I161, R166, V169, Q170 and N332 (most significant) associated with significant variation in PG9 sensitivity; K130, I161 and K171 associated with significant variation in PG16 sensitivity; I165, R166, V169 and N332 associated with significant variation in CAP256.VRC26.25 sensitivity was observed. Finally, for V3 directed bnAbs: K155, I165, R252, N289, E293, I307, H330, N332, S334, A336 and N448 associated with significant variation in PGT121 sensitivity; R252, N295, N300, H330, S334, A336, Q344, and T415 associated with significant variation in PGT128 sensitivity; K155, N156, I165, N230, T240, R252, K282, I307, A316, H330, S334, A336, Q344 and N448 associated with significant variation in 10-1074 sensitivity were observed. While the present study focuses on *gp120* sequences which are majorly relevant to the sensitivity to *gp120* targeting bnAbs, changes in sequences in *gp41* associated with resistance to bnAbs targeting *gp120* have also been reported [26, 35–38]. Upon

analysis of 1768 sequences representing different countries, we found high abundance of residues in *gp41* associated with resistance to bnAbs particularly those target V1V3 and V3 region (S1 Fig). Taken together, our study indicates the existence of variation in the abundance of key bnAb contact sites across global circulating HIV-1 subtype C. This observation highlights that the choice of bnAb combination for effective neutralization coverage against diverse region-specific circulating HIV-1 subtype C would likely vary significantly.

Evidence of accumulation of bnAb resistance phenotype in globally circulating HIV-1 subtype C over time

With abundance of bnAb neutralization data against specific envelope sequences obtained *in vitro* as well as through clinical trials, several machine learning-based algorithms are increasingly becoming available that can predict probable sensitivity to bnAbs on the basis of gp120 sequences. In the present study we employed one such recently published algorithm bNAb-ReP [39] to predict sensitivity of 1466 gp120 sequences selected in the present study from countries India, China, South Africa, Malawi, Zambia and Tanzania to following bnAbs: 3BNC117, VRC01, VRC07, PGT145, CAP256:VRC26.25, PGDM1400, PG9, PG16, PGT121, PGT128 and 10–1074. The prediction data were plotted along with country-wise *in vitro* data available through CATNAP database for a total of 283 sequences as indicated in Table 2. As indicated in Fig 6, probability values greater than 0.5 point towards sensitivity to bnAbs while those lower than 0.5 indicate probable resistance. For VRC01, while CATNAP database indicated unequivocal sensitivity of majority of sequences from all 6 countries, bNAb-ReP predicted a significant fraction of sequences from African countries to be resistant. Predictions for VRC26.25, 10–1074, PGT121 and PGDM1400 matched those reported in the CATNAP database. Similar to VRC01, predictions for 3BNC117, PGT128, PG9, PG16 and VRC07 did not match the data from CATNAP and indicated probable resistance in many of the sequences from all 6 countries. To assess if bnAb sensitivity differed over time, prediction data for each of the 11 bnAbs was plotted against three periods based on the reporting date of the sequences (Fig 7). The three periods considered were 1986–2000 (N = 244), 2001–2010 (N = 1187) and 2011–2019 (N = 333). Except for PGT121, PGT128 and 10–1074, all bnAbs showed significant decrease in sensitivity over time. Despite this decrease, most sequences were predicted to be susceptible to PGDM1400 and VRC26.25. However, susceptibility to 3BNC117, VRC01, VRC07, PGT145, PG9 as well as PG16 was predicted to have reduced significantly over the period of time assessed.

Discussion

Given that subtype C accounts for approximately half of the global HIV burden, limited information on the intra-clade C *env* diversity and its association with variation in their

Table 2. Details of sequences analyzed, and their sources used for predication analysis of HIV-1 subtype C to different bnAbs.

Country	Number of sequences	
	CATNAP	LANL-HIV (bNAb-ReP)
India	13	83
China	7	24
South Africa	146	910
Malawi	62	207
Zambia	23	189
Tanzania	32	53
Total	283	1466

<https://doi.org/10.1371/journal.pone.0251969.t002>

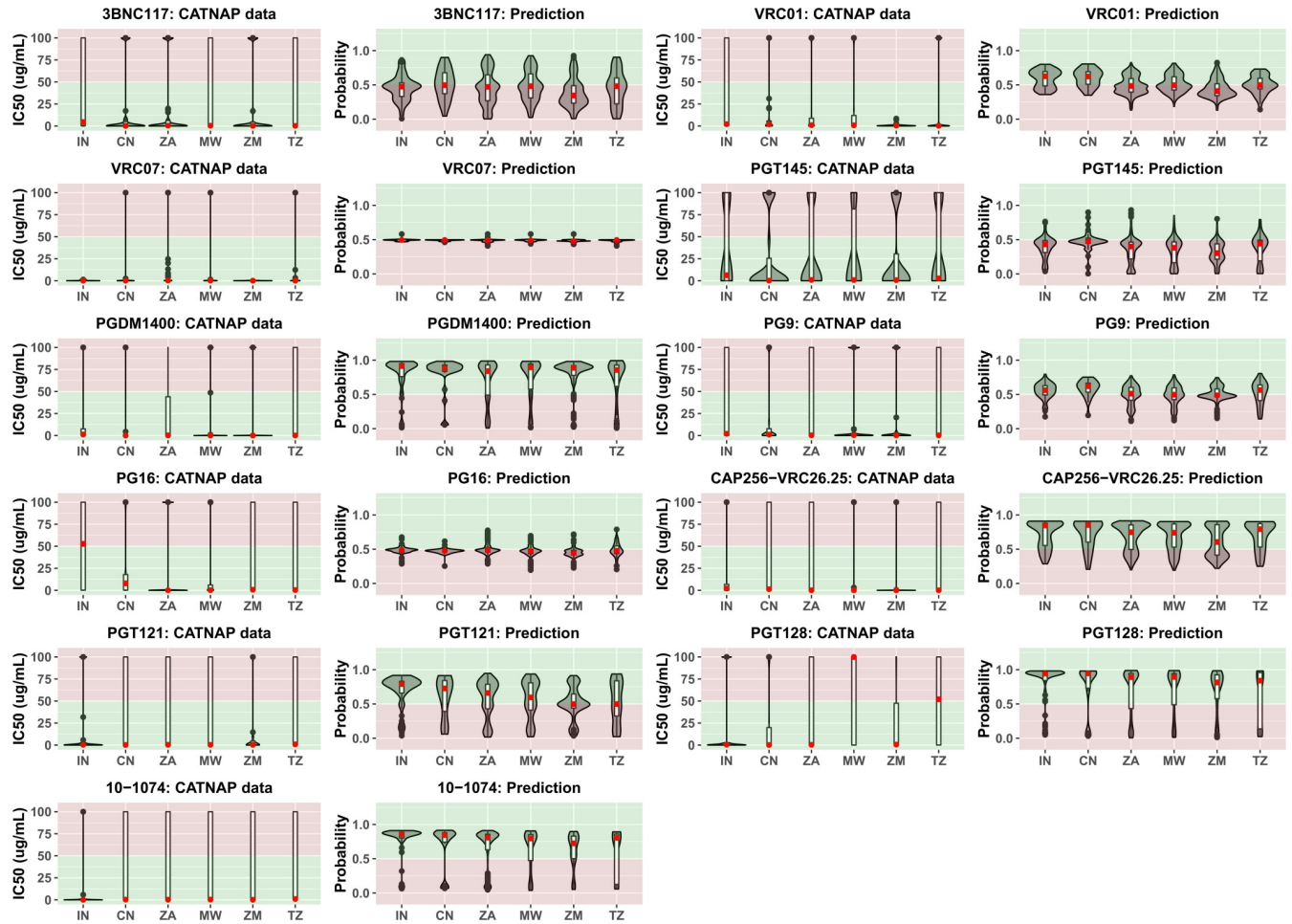


Fig 6. Prediction of bnAb sensitivity across different countries. Each panel indicates available country-wise CATNAP data plotted next to country-wise prediction data for available sequences for 3BNC117, VRC01, VRC03, VRC07, VRC13, CAP256-VRC26.25, PGDM1400, PG9, PG16, PGT145, PGT121, and PGT128 and 10-1074. In the CATNAP data panels, country-wise violin plots have been inlayed with boxplots against reported IC50 ($\mu\text{g/mL}$) values. Black dots indicate outliers while red dots indicate median values. Red background zone indicates bnAb resistance ($\text{IC}_{50} > 50 \mu\text{g/mL}$) while green background zone indicates bnAb sensitivity ($\text{IC}_{50} < 50 \mu\text{g/mL}$). In the bNAb-ReP data panels, country-wise violin plots have been inlayed with boxplots against probability of neutralization as predicted by bNAb-ReP. Black dots indicate outliers while red dots indicate median values. Red background zone indicates probable bnAb resistance (Neutralization probability < 0.5) while green background zone indicates probable bnAb sensitivity (Neutralization probability > 0.5).

<https://doi.org/10.1371/journal.pone.0251969.g006>

neutralization phenotypes exists. In the present study, we compared the genetic attributes of the globally circulating HIV-1 subtype C *gp120* sequences that differentiate the region-specific intra-clade HIV-1 subtype C neutralization diversity. For this, we used the available information in the HIV database (www.hiv.lanl.gov) and established algorithms [39] to suitably predict the association between genetic features that potentially affect the HIV-1 intra-clade C neutralization diversity. In spite of the disparity between the number of existing region-specific unique HIV-1 subtype C sequences in the database, region-specific distinct genetic clustering was observed by phylogenetic analysis of *gp120* amino acid sequences. It is to be noted that our analysis was based on one sequence per individual to avoid any sampling bias on an individual level. The region-specific subtype C *gp120* divergence could possibly be due to diversity in the population level across geography [14]. Indeed, studies have shown that HIV-1 can selectively incorporate broad range of biologically active host proteins in the process of viral egress, which can potentially exhibit altered pathogenicity and neutralization phenotypes [40]. This

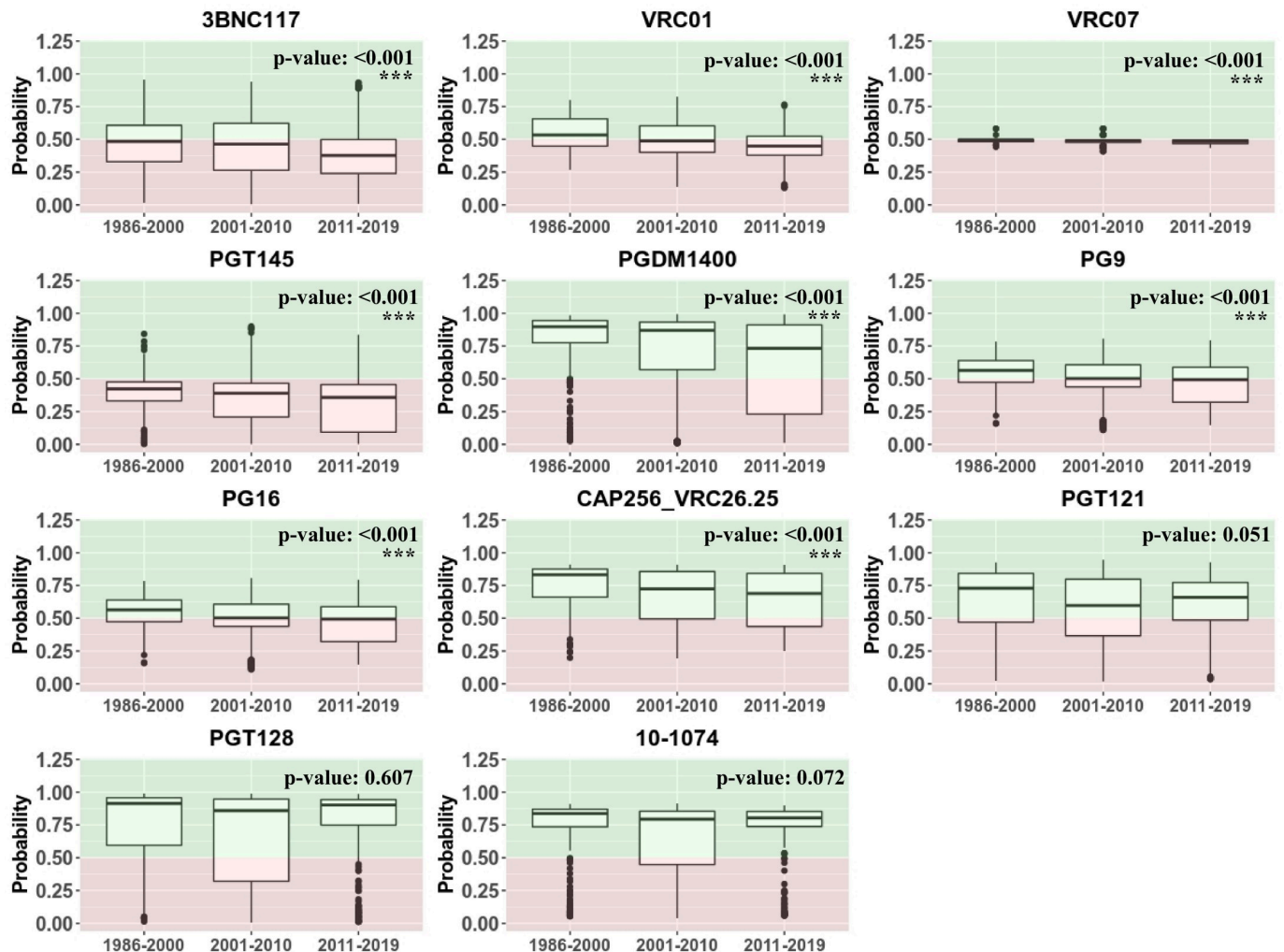


Fig 7. Assessment of predicted bnAb sensitivity over time. Each panel indicates cumulative prediction data for available sequences for 3BNC117, VRC01, VRC03, VRC07, VRC13, CAP256:VRC26.25, PGDM1400, PG9, PG16, PGT145, PGT121, and PGT128 and 10–1074 plotted against 3-time periods as follows: 1986–2000 (N = 244), 2001–2010 (N = 1187) and 2011–2019 (N = 333). P values indicate trend analysis performed by Jonckheere-Terpstra test. P values < 0.05 were considered statistically significant and have been indicated with ‘***’.

<https://doi.org/10.1371/journal.pone.0251969.g007>

also indicates possible association between intra clade C genetic diversity with ethnically distinct population.

The evolutionary genetic drift within subtype C that we observed from phylogenetic analysis is likely due to differential host characteristics which include immune response and differential genetic bottlenecks. For example, *gp120* sequences from Asian countries were found to demonstrate monophyletic clustering compared to that observed with those obtained from African countries. A number of studies [29–31] have demonstrated the role of loop length in the hypervariable regions, (with particular reference to V1V2 region), charge and N-linked glycosylation on altered neutralization phenotype. In the present study, we observed that while there is an existence of region-specific variation in V1V2 loop length, V4 loop length variation between subtype C sequences was found to be most profound across geographic boundaries. In addition, we also observed that the average *gp120* length of African countries was found to be smaller compared to other regions. Our data demonstrates subtle but significant variations in these attributes thereby indicating that a common set of bnAbs are not likely to be equally

effective against the HIV-1 subtype C circulating globally. The above conclusion was further substantiated by our finding of significant variation in the entropy profiles (variation in sites/positions associated with different bnAbs) between the geographically distinct *gp120* sequences. With limited data available in CATNAP, indeed we found evidence of variation in susceptibility of region-specific clade C to different bnAbs, which substantiate our observation. Interestingly, as reported elsewhere [16, 41, 42], our data also predicted potential likelihood of accumulation of resistance phenotype overtime to existing bnAbs. This observation indicate that it is necessary for continuous surveillance of evolving viruses in the context of subtype C along towards prioritizing bnAb combinations that will optimally dissect and overcome the evolving genetic diversity. Interestingly, a similar accumulation in ART resistance has been documented and studies that concurrently evaluate the interplay of these two evolutionary patterns, heretofore considered to be mutually exclusive, may highlight novel and synergistic therapeutic strategies. Furthermore, the recently concluded phase 2b HVTN 703/HPTN 081 Antibody Mediated Prevention (AMP) trial study also highlighted the discrepancy between the expected bnAb (VRC01) sensitivity of the viruses currently circulating in the population versus the sensitivity predicted *in vitro* through historically sampled viruses and the indispensability of continual virus surveillance [43–45].

Conclusion

The differences in HIV-1 clade C *gp120* sequences observed herein indicate disparate and distinctly evolving clusters within clade C with differential predicted responses to bnAbs. Elucidation of neutralization diversity of subtype C particularly in context of evolution of *gp120* over time will be essential for selecting appropriate bnAb combination for effective prophylaxis and treatment and also in informing rational vaccine design. Our study highlights that towards developing HIV-1 bnAbs as products for prevention and treatment, continued surveillance of the evolution of genetic features with particular reference to *env* gene that are targets of neutralizing antibodies of globally circulating HIV-1 remains crucial for the identification and prioritization of combination of bnAbs that would best provide maximal geography and population-specific neutralization coverage.

Methods

Retrieval of *gp120* sequences

Sequences for the *gp120* gene were retrieved from manually curated LANL HIV database (www.hiv.lanl.gov). Briefly, HIV-1 subtype C nucleotide sequences fully covering the genomic region 6225–7758 (as per HXB2 numbering) were retrieved and subsequently filtered with a one sequence per individual filter criterion. Sequence entries without any information regarding the sample source country were excluded. Multiple sequence alignment for the amino acid sequences along with HXB2 sequence (GenBank: K03455.1) was produced with Gene cutter (“Gene Cutter,” LANL). Gene Cutter clips the coding regions from unaligned nucleotide sequences and produces amino acid alignments based on Hmmer v 2.32 algorithm with a training set of the full-length genome alignment. Alignments were manually curated using Bioedit v7.2.5 [46]. Sequences with internal stop codons were discarded.

Phylogenetic analysis

The number of amino acid differences per site were estimated by averaging over all sequence pairs between different countries using Molecular Evolutionary Genetics Analysis software (MEGA v.10) [28]. The rate variation among sites was modelled with a gamma distribution

(shape parameter = 1). This analysis involved 1814 amino acid sequences. Phylogenetic trees were generated for the amino acid alignments with iqtree under 'HIVb' model with estimated γ parameters and number of invariable sites [47]. Robustness of the tree topology was further assessed by SH-aLRT as well as 1000 ultrafast bootstrap replicates implemented in iqtree. A subtree consisting of 251 sequences were again constructed as mentioned previously.

Estimation of the variable loop properties and potential N linked glycosylation sites

Variable loop regions for V1 (131–157: HXB2 numbering), V2 (158–196), V3 (296–331), V4 (386–417) and V5 (460–469) were retrieved from amino acid alignments with Bioedit v7.2.5. Each of the loop datasets were then processed with custom bash/awk scripts to generate length statistics. The length distributions were further assessed and compared by Kruskal Wallis test followed by Dunn's multiple comparison test. Cumulative variable loop charge values were predicted for each of the sequences with custom bash scripts wherein, Lysine (K), Arginine (R) and Histidine (H) residues were assigned +1 values each while Aspartic acid (D) and Glutamic acid (E) were assigned -1 values each. Potential N-linked glycosylation sites were predicted in amino acid sequence datasets with N-GlycoSite tool hosted at the HIV-LANL database [48]. Prevalence of each of the pNLG sites under study were calculated using custom bash/awk scripts and were further assessed statistically using Fisher's exact test. Country-wise pNLGs abundance heatmap was plotted using 'pheatmap' package in R.

Entropy analysis

Shannon entropy for selected sequence data sets was generated using Entropy-two tool following 100 randomizations with replacement [49, 50] (www.hiv.lanl.gov). Key sites for interaction with bnAbs VRC01, VRC03, VRC07, VRC13, CAP256:VRC26.25, PGDM1400, PG9, PG16, PGT121 and PGT128 were derived from CATNAP database. HIV-1 clade C overall entropy bed graph was derived from Genome browser on HIV-LANL database.

Analysis of bnAb epitope contact sites

Specific epitope contact sites/positions along with documented variants imparting sensitivity or resistance phenotype were retrieved from CATNAP database for bnAbs: VRC01, VRC07, PGT121, PGT128, PGT145, PG9, PG16, VRC26.25, 3BNC117, 10–1074 and N6. Frequency of such resistant variants were calculated across sequences from countries with more than 20 sequences available, using in-house bash scripts. Contact site resistance heatmaps were prepared with Circos v0.61 [51].

Prediction of bnAb sensitivity with bNAb-ReP

Sensitivity of contact sites to bnAbs (3BNC117, VRC01, VRC07, CAP256:VRC26.25, PGDM1400, PG9, PG16, PGT145, PGT121, PGT128 and 10–1074) were predicted with bNAb-ReP tool [39]. Neutralization data corresponding to the selected bnAbs was retrieved from CATNAP database (www.hiv.lanl.gov).

The temporal prediction data was stratified into approximately three decades as follows: 1986–2000 (N = 244), 2001–2010 (N = 1187) and 2011–2019 (N = 333).

Statistical analyses and data presentation

Statistical analyses for variable loop length distributions were performed using GraphPad Prism version 5.01 for Windows, GraphPad Software, San Diego California USA. Statistical

Comparison of Fisher's test for pNLG sites as well as abundance of bnAb resistance associated residues was performed through R statistical computing software (v3.4.0) and R studio v1.0.143 [52, 53]. Phylogenetic trees were visualised and edited with the R package 'Graphlan' [54]. Variable entropy positions were plotted on prefusion gp120 envelope model derived from PDB:5U7O in Chimera v1.14 [55]. Plots depicting variable region characteristics, entropy differences and bnAb sensitivity predictions were prepared using 'ggplot2' package in R [56]. Trend analysis for predicted bnAb sensitivity was performed by Jonckheere-Terpstra test implemented in R statistical software.

Supporting information

S1 Fig. Abundance of amino acid residues in gp41 associated with bnAb. A circos heatmap was plotted based on analysis of 1768 sequences from the HIV database (www.hiv.lanl.gov) to depict the abundance of known amino acid residues associated with resistance to the following bnAbs: PGT145, PGDM1400, PG9, PG16, CAP256-VRC26.25, PGT121, PGT128 and 10–1074. Each track indicates the country of origin and each pixel on the circular track indicates position of specific residue associated with resistance to indicated bnAbs. (TIF)

Acknowledgments

We thank our laboratory members for providing valuable inputs and suggestions. ICMR-NIRRH is acknowledged for server support for computation analysis. IAVI's work was made possible by generous support from many donors, including the Bill & Melinda Gates Foundation, the Ministry of Foreign Affairs of Denmark, Irish Aid, the Ministry of Finance of Japan, the Ministry of Foreign Affairs of the Netherlands, the Norwegian Agency for Development Cooperation (NORAD), the United Kingdom Department for International Development (DFID), and the United States Agency for International Development (USAID). The full list of IAVI donors is available at www.iavi.org. The contents are the responsibility of the International AIDS Vaccine Initiative and do not necessarily reflect the views of USAID or the United States Government. We thank Prof Gagandeep Kang, Translational Health Science & Technology Institute for support.

Author Contributions

Conceptualization: Suprit Deshpande, Nitin Hingankar, Jayanta Bhattacharya.

Data curation: Jyoti Sutar, Vainav Patel.

Formal analysis: Jyoti Sutar, Suprit Deshpande, Ranajoy Mullick, Vainav Patel, Jayanta Bhattacharya.

Funding acquisition: Jayanta Bhattacharya.

Investigation: Jyoti Sutar, Ranajoy Mullick, Nitin Hingankar, Jayanta Bhattacharya.

Methodology: Jyoti Sutar, Nitin Hingankar, Jayanta Bhattacharya.

Project administration: Jayanta Bhattacharya.

Resources: Jayanta Bhattacharya.

Supervision: Jayanta Bhattacharya.

Validation: Vainav Patel.

Writing – original draft: Jyoti Sutar.

Writing – review & editing: Jyoti Sutar, Vainav Patel, Jayanta Bhattacharya.

References

1. Mendoza P, Gruell H, Nogueira L, Pai JA, Butler AL, Millard K, et al. 2018. Combination therapy with anti-HIV-1 antibodies maintains viral suppression. *Nature* 561:479–484. <https://doi.org/10.1038/s41586-018-0531-2> PMID: 30258136
2. Kumar R, Qureshi H, Deshpande S, Bhattacharya J. 2018. Broadly neutralizing antibodies in HIV-1 treatment and prevention. *Ther Adv Vaccines Immunother* 6:61–68. <https://doi.org/10.1177/2515135518800689> PMID: 30345419
3. Escolano A, Dosenovic P, Nussenzweig MC. 2017. Progress toward active or passive HIV-1 vaccination. *J Exp Med* 214:3–16. <https://doi.org/10.1084/jem.20161765> PMID: 28003309
4. Sok D, Burton DR. 2016. HIV Broadly Neutralizing Antibodies: Taking Good Care Of The 98. *Immunity* 45:958–960. <https://doi.org/10.1016/j.immuni.2016.10.033> PMID: 27851923
5. Sok D, Burton DR. 2018. Recent progress in broadly neutralizing antibodies to HIV. *Nat Immunol* 19:1179–1188. <https://doi.org/10.1038/s41590-018-0235-7> PMID: 30333615
6. Caskey M, Klein F, Nussenzweig MC. 2019. Broadly neutralizing anti-HIV-1 monoclonal antibodies in the clinic. *Nat Med* 25:547–553. <https://doi.org/10.1038/s41591-019-0412-8> PMID: 30936546
7. Nishimura Y, Martin MA. 2017. Of Mice, Macaques, and Men: Broadly Neutralizing Antibody Immunotherapy for HIV-1. *Cell Host Microbe* 22:207–216. <https://doi.org/10.1016/j.chom.2017.07.010> PMID: 28799906
8. Caskey M, Klein F, Lorenzi JC, Seaman MS, West AP Jr., Buckley N, et al. 2015. Viraemia suppressed in HIV-1-infected humans by broadly neutralizing antibody 3BNC117. *Nature* <https://doi.org/10.1038/nature14411> PMID: 25855300
9. Caskey M, Schoofs T, Gruell H, Settler A, Karagounis T, Kreider EF, et al. 2017. Antibody 10–1074 suppresses viremia in HIV-1-infected individuals. *Nat Med* 23:185–191. <https://doi.org/10.1038/nm.4268> PMID: 28092665
10. Han C, Johnson J, Dong R, Kandula R, Kort A, Wong M, et al. 2020. Key Positions of HIV-1 Env and Signatures of Vaccine Efficacy Show Gradual Reduction of Population Founder Effects at the Clade and Regional Levels. *mBio* 11. <https://doi.org/10.1128/mBio.00126-20> PMID: 32518179
11. Korber B, Gaschen B, Yusim K, Thakallapally R, Kesmir C, Detours V. 2001. Evolutionary and immunological implications of contemporary HIV-1 variation. *Br Med Bull* 58:19–42. <https://doi.org/10.1093/bmb/58.1.19> PMID: 11714622
12. Lynch RM, Shen T, Gnanakaran S, Derdeyn CA. 2009. Appreciating HIV type 1 diversity: subtype differences in Env. *AIDS Res Hum Retroviruses* 25:237–48. <https://doi.org/10.1089/aid.2008.0219> PMID: 19327047
13. Gnanakaran S, Lang D, Daniels M, Bhattacharya T, Derdeyn CA, Korber B. 2007. Clade-specific differences between human immunodeficiency virus type 1 clades B and C: diversity and correlations in C3-V4 regions of gp120. *J Virol* 81:4886–91. <https://doi.org/10.1128/JVI.01954-06> PMID: 17166900
14. Hraber P, Korber BT, Lapedes AS, Bailer RT, Seaman MS, Gao H, et al. 2014. Impact of clade, geography, and age of the epidemic on HIV-1 neutralization by antibodies. *J Virol* 88:12623–43. <https://doi.org/10.1128/JVI.01705-14> PMID: 25142591
15. DeLeon O, Hodis H, O'Malley Y, Johnson J, Salimi H, Zhai Y, et al. 2017. Accurate predictions of population-level changes in sequence and structural properties of HIV-1 Env using a volatility-controlled diffusion model. *PLoS Biol* 15:e2001549. <https://doi.org/10.1371/journal.pbio.2001549> PMID: 28384158
16. Bouvin-Pley M, Morgand M, Moreau A, Jestin P, Simonnet C, Tran L, et al. 2013. Evidence for a continuous drift of the HIV-1 species towards higher resistance to neutralizing antibodies over the course of the epidemic. *PLoS Pathog* 9:e1003477. <https://doi.org/10.1371/journal.ppat.1003477> PMID: 23853594
17. Bunnik EM, Euler Z, Welkers MR, Boeser-Nunnink BD, Grijns ML, Prins JM, et al. 2010. Adaptation of HIV-1 envelope gp120 to humoral immunity at a population level. *Nat Med* 16:995–7. <https://doi.org/10.1038/nm.2203> PMID: 20802498
18. Buonaguro L, Tornesello ML, Buonaguro FM. 2007. Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications. *J Virol* 81:10209–19. <https://doi.org/10.1128/JVI.00872-07> PMID: 17634242
19. Taylor BS, Sobieszczyk ME, McCutchan FE, Hammer SM. 2008. The challenge of HIV-1 subtype diversity. *N Engl J Med* 358:1590–602. <https://doi.org/10.1056/NEJMra0706737> PMID: 18403767

20. Seaman MS, Janes H, Hawkins N, Grandpre LE, Devoy C, Giri A, et al. 2010. Tiered categorization of a diverse panel of HIV-1 Env pseudoviruses for assessment of neutralizing antibodies. *J Virol* 84:1439–52. <https://doi.org/10.1128/JVI.02108-09> PMID: 19939925
21. Binley JM, Wrin T., Korber B., Zwick M. B., Wang M., Chappey C., et al. 2004. Comprehensive cross-clade neutralization analysis of a panel of anti-human immunodeficiency virus type 1 monoclonal antibodies. *J Virol* 78:13232–13252. <https://doi.org/10.1128/JVI.78.23.13232-13252.2004> PMID: 15542675
22. Bures R, Morris L, Williamson C, Ramjee G, Deers M, Fiscus SA, et al. 2002. Regional clustering of shared neutralization determinants on primary isolates of clade C human immunodeficiency virus type 1 from South Africa. *J Virol* 76:2233–44. <https://doi.org/10.1128/jvi.76.5.2233-2244.2002> PMID: 11836401
23. Li M, Salazar-Gonzalez JF, Derdeyn CA, Morris L, Williamson C, Robinson JE, et al. 2006. Genetic and neutralization properties of subtype C human immunodeficiency virus type 1 molecular env clones from acute and early heterosexually acquired infections in Southern Africa. *J Virol* 80:11776–90. <https://doi.org/10.1128/JVI.01730-06> PMID: 16971434
24. Kulkarni SS, Lapedes A, Tang H, Gnanakaran S, Daniels MG, Zhang M, et al. 2009. Highly complex neutralization determinants on a monophyletic lineage of newly transmitted subtype C HIV-1 Env clones from India. *Virology* 385:505–20. <https://doi.org/10.1016/j.virol.2008.12.032> PMID: 19167740
25. Novitsky V, Smith UR, Gilbert P, McLane MF, Chigwedere P, Williamson C, et al. 2002. Human immunodeficiency virus type 1 subtype C molecular phylogeny: consensus sequence for an AIDS vaccine design? *J Virol* 76:5435–51. <https://doi.org/10.1128/jvi.76.11.5435-5451.2002> PMID: 11991972
26. Sutar J, Padwal V, Sonawani A, Nagar V, Patil P, Kulkarni B, et al. 2019. Effect of diversity in gp41 membrane proximal external region of primary HIV-1 Indian subtype C sequences on interaction with broadly neutralizing antibodies 4E10 and 10E8. *Virus Res* 273:197763. <https://doi.org/10.1016/j.virusres.2019.197763> PMID: 31553924
27. Hemelaar J, Elangovan R, Yun J, Dickson-Tetteh L, Fleminger I, Kirtley S, et al. 2019. Global and regional molecular epidemiology of HIV-1, 1990–2015: a systematic review, global survey, and trend analysis. *Lancet Infect Dis* 19:143–155. [https://doi.org/10.1016/S1473-3099\(18\)30647-9](https://doi.org/10.1016/S1473-3099(18)30647-9) PMID: 30509777
28. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol* 35:1547–1549. <https://doi.org/10.1093/molbev/msy096> PMID: 29722887
29. Deshpande S, Patil S, Kumar R, Hermanus T, Murugavel KG, Srikrishnan AK, et al. 2016. HIV-1 clade C escapes broadly neutralizing autologous antibodies with N332 glycan specificity by distinct mechanisms. *Retrovirology* 13:60. <https://doi.org/10.1186/s12977-016-0297-2> PMID: 27576440
30. Ringe R, Phogat S, Bhattacharya J. 2012. Subtle alteration of residues including N-linked glycans in V2 loop modulate HIV-1 neutralization by PG9 and PG16 monoclonal antibodies. *Virology* 426:34–41. <https://doi.org/10.1016/j.virol.2012.01.011> PMID: 22314018
31. van Gils MJ, Bunnik EM, Boeser-Nunnink BD, Burger JA, Terlouw-Klein M, Verwer N, et al. 2011. Longer V1V2 region with increased number of potential N-linked glycosylation sites in the HIV-1 envelope glycoprotein protects against HIV-specific neutralizing antibodies. *J Virol* 85:6986–95. <https://doi.org/10.1128/JVI.00268-11> PMID: 21593147
32. Doores KJ, Kong L, Krumm SA, Le KM, Sok D, Laserson U, et al. 2015. Two classes of broadly neutralizing antibodies within a single lineage directed to the high-mannose patch of HIV envelope. *J Virol* 89:1105–18. <https://doi.org/10.1128/JVI.02905-14> PMID: 25378488
33. Ogert RA, Lee MK, Ross W, Buckler-White A, Martin MA, Cho MW. 2001. N-linked glycosylation sites adjacent to and within the V1/V2 and the V3 loops of dualtropic human immunodeficiency virus type 1 isolate DH12 gp120 affect coreceptor usage and cellular tropism. *J Virol* 75:5998–6006. <https://doi.org/10.1128/JVI.75.13.5998-6006.2001> PMID: 11390601
34. Bai H, Li Y, Michael NL, Robb ML, Rolland M. 2019. The breadth of HIV-1 neutralizing antibodies depends on the conservation of key sites in their epitopes. *PLoS Comput Biol* 15:e1007056. <https://doi.org/10.1371/journal.pcbi.1007056> PMID: 31170145
35. Blish CA, Nguyen MA, Overbaugh J. 2008. Enhancing exposure of HIV-1 neutralization epitopes through mutations in gp41. *PLoS Med* 5:e9. <https://doi.org/10.1371/journal.pmed.0050009> PMID: 18177204
36. Ringe R, Bhattacharya J. 2012. Association of enhanced HIV-1 neutralization by a single Y681H substitution in gp41 with increased gp120-CD4 interaction and macrophage infectivity. *PLoS One* 7:e37157. <https://doi.org/10.1371/journal.pone.0037157> PMID: 22606344
37. Shen X, Dennison SM, Liu P, Gao F, Jaeger F, Montefiori DC, et al. 2010. Prolonged exposure of the HIV-1 gp41 membrane proximal region with L669S substitution. *Proc Natl Acad Sci U S A* 107:5972–5977. <https://doi.org/10.1073/pnas.0912381107> PMID: 20231447

38. Sutar J, Padwal V, Nagar V, Patil P, Patel V, Bandivdekar A. 2020. Analysis of sequence diversity and selection pressure in HIV-1 clade C gp41 from India. *Virusdisease* 31:277–291. <https://doi.org/10.1007/s13337-020-00595-x> PMID: 32904888
39. Rawi R, Mall R, Shen CH, Farney SK, Shiakolas A, Zhou J, et al. 2019. Accurate Prediction for Antibody Resistance of Clinical HIV-1 Isolates. *Sci Rep* 9:14696. <https://doi.org/10.1038/s41598-019-50635-w> PMID: 31604961
40. Burnie J, Guzzo C. 2019. The Incorporation of Host Proteins into the External HIV-1 Envelope. *Viruses* 11. <https://doi.org/10.3390/v11010085> PMID: 30669528
41. Bouvin-Pley M, Morgand M, Meyer L, Goujard C, Moreau A, Mouquet H, et al. 2014. Drift of the HIV-1 envelope glycoprotein gp120 toward increased neutralization resistance over the course of the epidemic: a comprehensive study using the most potent and broadly neutralizing monoclonal antibodies. *J Virol* 88:13910–7. <https://doi.org/10.1128/JVI.02083-14> PMID: 25231299
42. Hake A, Pfeifer N. 2017. Prediction of HIV-1 sensitivity to broadly neutralizing antibodies shows a trend towards resistance over time. *PLoS Comput Biol* 13:e1005789. <https://doi.org/10.1371/journal.pcbi.1005789> PMID: 29065122
43. Corey L, Gilbert PB, Juraska M, Montefiori DC, Morris L, Karuna ST, et al. 2021. Two Randomized Trials of Neutralizing Antibodies to Prevent HIV-1 Acquisition. *N Engl J Med* 384:1003–1014. <https://doi.org/10.1056/NEJMoa2031738> PMID: 33730454
44. Walker BD. 2021. The AMP Trials—A Glass Half Full. *N Engl J Med* 384:1068–1069. <https://doi.org/10.1056/NEJMe2101131> PMID: 33730459
45. Rademeyer C, Korber B, Seaman MS, Giorgi EE, Thebus R, Robles A, et al. 2016. Features of Recently Transmitted HIV-1 Clade C Viruses that Impact Antibody Recognition: Implications for Active and Passive Immunization. *PLoS Pathog* 12:e1005742. <https://doi.org/10.1371/journal.ppat.1005742> PMID: 27434311
46. Hall T. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp*
47. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268–74. <https://doi.org/10.1093/molbev/msu300> PMID: 25371430
48. Zhang M, Gaschen B, Blay W, Foley B, Haigwood N, Kuiken C, et al. 2004. Tracking global patterns of N-linked glycosylation site variation in highly variable viral glycoproteins: HIV, SIV, and HCV envelopes and influenza hemagglutinin. *Glycobiology* 14:1229–46. <https://doi.org/10.1093/glycob/cwh106> PMID: 15175256
49. Efron B, Tibshirani R. 1991. Statistical data analysis in the computer age. *Science* 253:390–5. <https://doi.org/10.1126/science.253.5018.390> PMID: 17746394
50. Gaschen B, Taylor J, Yusim K, Foley B, Gao F, Lang D, et al. 2002. Diversity considerations in HIV-1 vaccine selection. *Science* 296:2354–60. <https://doi.org/10.1126/science.1070441> PMID: 12089434
51. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res* 19:1639–45. <https://doi.org/10.1101/gr.092759.109> PMID: 19541911
52. Team RC. 2018. R: A language and environment for statistical computing.
53. Team R. 2015. RStudio: Integrated Development for R.
54. Asnicar F, Weingart G, Tickle TL, Huttenhower C, Segata N. 2015. Compact graphical representation of phylogenetic data and metadata with GraPhlAn. *PeerJ* 3:e1029. <https://doi.org/10.7717/peerj.1029> PMID: 26157614
55. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. 2004. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–12. <https://doi.org/10.1002/jcc.20084> PMID: 15264254
56. Analysis. WHgEGfD. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York