CATNAP: a tool to compile, analyze and tally neutralizing antibody panels

Hyejin Yoon[1], Jennifer Macke[1], Anthony P. West, Jr[2], Brian Foley[1], Pamela J. Bjorkman[2], Bette Korber[1] and Karina Yusim[1,*]

[1]Los Alamos National Laboratory, Los Alamos, NM, USA and [2]California Institute of Technology, Pasadena, CA, USA

ABSTRACT

CATNAP (Compile, Analyze and Tally NAb Panels) is a new web server at Los Alamos HIV Database, created to respond to the newest advances in HIV neutralizing antibody research. It is a comprehensive platform focusing on neutralizing antibody potencies in conjunction with viral sequences. CATNAP integrates neutralization and sequence data from published studies, and allows users to analyze that data for each HIV Envelope protein sequence position and each antibody. The tool has multiple data retrieval and analysis options. As input, the user can pick specific antibodies and viruses, choose a panel from a published study, or supply their own data. The output superimposes neutralization panel data, virus epidemiological data, and viral protein sequence alignments on one page, and provides further information and analyses. The user can highlight alignment positions, or select antibody contact residues and view position-specific information from the HIV databases. The tool calculates tallies of amino acids and N-linked glycosylation motifs, counts of antibody-sensitive and -resistant viruses in conjunction with each amino acid or N-glycosylation motif, and performs Fisher's exact test to detect potential positive or negative amino acid associations for the selected antibody. Website name: CATNAP (Compile, Analyze and Tally NAb Panels). Website address: http://hiv.lanl.gov/catnap.

INTRODUCTION

Despite more than 30 years of focused scientific efforts worldwide, creating an effective HIV vaccine has proven difficult. HIV is extraordinary variable (1) and designing a vaccine able to stimulate immunological responses that will broadly cross-react with circulating variants is a challenge. Inducing neutralizing antibodies that can block viral infection of a target cell is considered essential for an effective vaccine. Several breakthrough experimental techniques developed in recent years allowed highly efficient interrogation of human memory B cells and plasma cells and consequently the isolation of many broadly neutralizing antibodies (bNAbs) (2). These bNAbs are able to neutralize multiple circulating HIV-1 strains, and large panels of pseudotyped viruses are used to assess the neutralization breadth and potency of these antibodies (3–7).

Neutralization panel data is usually available in the supplemental materials of the published studies and includes IC_{50}, and for some studies IC_{80}, neutralization values (the concentration at which infectivity is reduced by 50% or 80%, respectively (7)) for multiple antibodies and hundreds of viruses. Depending on the study, antibody structure and antibody sequences may also be available. The accumulation of these new monoclonal antibodies together with the vast related neutralization panel information requires storage, comparison and analysis tools. Several databases and servers to organize and access this data have become available in recent years (8–10). The standalone program AntibodyDatabase (8) provides an integrated platform for examining sequence, structure, and neutralization data in a holistic way; however this tool is not a web server. The bNAber database (9) collects neutralization scores and available antibody structures and sequences of the most important bNAbs, and has very useful visualization and analysis tools. The Neutralization-based Epitope Prediction (NEP) server predicts antibody-specific epitopes at the residue level based on neutralization panels of viral strains, using the user's data (10,11). These important web-based resources, however, do not contain the viral sequence data, and so do not readily enable the exploration of how HIV Envelope (Env) sequence variation is correlated with the neutralization sensitivity. The combination of published neutralization scores, antibody sequences, and viral sequences commonly used for the evaluation of neutralizing antibodies, together with initial analysis of bNAb associations with viral sequence mutations has first become available on the web through our new tool CATNAP (Compile, Analyze and Tally NAb Panels), available at the Los Alamos HIV Database. In addition, many large Env panels are published

*To whom correspondence should be addressed. Tel: +1 505 665 2313; Email: kyusim@lanl.gov

without accession numbers, and with wide discrepancies in sequence names used by different laboratories, making subsequent comparisons between studies and meta-analysis difficult. Given that Los Alamos HIV Database project has a mission of bringing together global HIV sequence and immunology data, we worked with the primary investigators to systematically determine the exact viruses used for neutralization studies in different laboratories. This has enabled an integrated view of HIV Envelope sequences, neutralizing antibody IC_{50} and IC_{80} data, and descriptions of HIV Env/antibody contact residues, collected from multiple studies and put into the framework of our database tools and web services. Links to Antibody/Env structures and complete information of antibody sequence data are also provided, and better visualization tools for this data are under development.

In this report, we focus on CATNAP, a web-based portal of neutralizing antibody IC_{50} and IC_{80} values in conjunction with viral data, inspired by studies by West *et al.* (8) and Gnanakaran *et al.* (12). CATNAP allows users to superimpose neutralization results and virus sequences from published sources and their own data, and perform initial analysis to find potential neutralization antibody signatures. This tool is a part of our larger neutralization antibody resources page at the Los Alamos HIV Database, which includes several other useful sources of information, such as a table of best neutralizing antibodies, a list of external tools for germline antibody reconstruction (13–18), an HIV genome browser, and the neutralizing antibody contexts and features database. The latter is compiled from multiple references and contains coordinates of important neutralizing antibody contact sites (3,4,19–28), mutations affecting neutralization sensitivity (22,29), CD4 contacts (4,26) cytoplasmic tail interactions (30), and antibody sensitivity signature predictions (8,10,11). CATNAP taps into the neutralizing antibody contexts and features database and links to our HIV Immunology and Sequence databases, allowing the rich data collected in these databases to be directly incorporated into the analysis.

MATERIALS AND METHODS

The neutralization panel data (IC_{50} and IC_{80} values for specific monoclonal antibodies and pseudotyped viruses) were collected from 49 published neutralization studies, mostly from tables in PDF format provided in the supplemental materials. The viral data was collected from Los Alamos HIV Database and in some cases personal communication with the authors, and required careful consideration and systematization to resolve sequence name ambiguities between different laboratories. Antibody sequences were downloaded from GenBank, and links to the structures in Protein Data Bank are provided. Standard statistics are applied to tally and analyze neutralization results. Specifically, antibody associations with viral mutations are evaluated by Fisher's exact test, counting the number of antibody-resistant or antibody-sensitive viruses (above or below threshold of detection) and the presence or absence of specific amino acids or N-glycosylation motif in the viral sequence alignment position (12).

RESULTS

Data retrieval

CATNAP includes neutralization data from published studies, as well as curated HIV-1 Env alignments corresponding to neutralization panels with carefully standardized virus sequence names. The current collection comprises 172 antibodies and 722 HIV-1 viruses (529 with sequences, and the remaining 193 viruses were published by companies, and so the sequences are proprietary) from 49 published studies, with more being added frequently.

The input page allows the user to retrieve details about selected HIV-1 Env sequences and antibodies used in neutralization panels, retrieve available details about the donor from whom the antibody was isolated, and select viruses and antibodies both individually and by study, for analysis (Figure 1). The *antibody details* option provides, for the specified antibody(s), links to the corresponding immunology database records, notes, references, links to crystal structures in the Protein Data Bank, antibody donor ID and clonal lineage, data from our Neutralizing Antibody Contexts and Features Database, and heavy and light chain antibody variable region mRNA sequences. The *virus details* option provides HIV subtype, sampling country, disease stage information, accession number, neutralization tier, and Los Alamos Database comments. (In some cases, the sequence of the virus used in experiments does not exactly correspond to the GenBank sequence. We resolved these issues to the extent possible via personal communication with the authors, and the comments associated with each sequence document whether the GenBank sequence or the unpublished one from the authors is used.) Importantly, a sequence name stored in CATNAP corresponds to the most frequently used name we identified in published neutralization studies. In many cases, however, different names are used in studies by different groups, and they frequently do not exactly correspond to the database sequence name. We provide a list of sequence name aliases and HIV database name for an easy one-to-one correspondence between the various short names, database names and accession numbers, thereby facilitating meta-studies. Finally, a set of accession numbers of the selected viruses can be automatically uploaded to the Sequence Search Interface of our HIV Sequence Database, to obtain information from many additional sequence database fields. The *assay details* option provides a downloadable table of available IC_{50} and IC_{80} data for selected viruses and antibodies from each study. When multiple studies assayed the same antibody–virus combination, both the original data and the geometric mean of the repeated data points is shown.

Analysis page

The analysis page superimposes virus data, neutralization data, and aligned viral sequences on one page. For each virus-antibody pair, the user can see virus information (tier, subtype, country, GenBank accession number, sequence aliases), neutralization values, both per study and the geometric means, and the virus sequences with highlighted N-linked glycosylation motifs (Figure 2). Analysis and ad-

Env pos.	Feature	AA	Entropy Group M	Entropy Subtype B	Entropy Subtype C	Annotation
160	gp120, V2	N	0.401	0.452	0.358	N160A confers resistance to some but not all members of the CAP256-VRC26 clonal lineage, so the antibody evolved away from dependency on this glycan.
162	gp120, V2	S	0.289	0.445	0.179	The CAP256-VRC26 lineage appears to have initially responded to a superinfecting virus that had T162; the earlier founder virus had I162, so it lacked the glycan at 160, and was resistant to this lineage of antibodies.
166	gp120, V2	R	1.127	1.398	0.772	R166A mutation confers complete resistance to members of the CAP256-VRC26 lineage
167	gp120, V2	G	0.601	1.122	0.185	D167N confers at least partial resistance to all members of CAP256-VRC26 clonal lineage tested.
168	gp120, V2	K	0.547	0.666	0.396	K168A confers resistance to some but not all members of the CAP256-VRC26 clonal lineage.

A) Antibody details

Antibody	Structure	Donor	Clonal lineage	Isolation paper	Neutralizing antibody feature	Heavy chain	Light chain
VRC26.01	4OCR	CAP256	VRC26	Doria-Rose2014	CAP256-VRC26 lineage resistance sites	Homo sapiens anti-HIV-1 immunoglobulin CAP256-VRC26.01 heavy chain variable region mRNA, partial cds	Homo sapiens anti-HIV-1 immunoglobulin CAP256-VRC26.01 light chain variable region mRNA, partial cds

EVQVVESGGGVVQPGRSLRLSCTASGFTFSNFAMGWVRQAPGKGLEWVAF
ISSDGSNKNYGDSVKGRFTISRDNSKNTVFLQMNSLRVEDTALYYCAKDV
DYKSDEWGTEYYDISISYPIQDPRAMVGAFDLWGQGTMVTVSX

U134866
AAGTGCAGGTGGTGGAGTCTGGGGGAGGCGTGGTCCAGCCTGGG
CTGAGACTCTCCTGTACAGCCTCTGGATTCACTTTCAGCAACTT
GGGCTGGGTCCGCCAGGCTCCGGGCAAGGGTCTGGAGTGGGTGG
ATATCATCTGATGGAAGTAATAAAAACTATGGAGACTCCGTGAAG
ATTCACCATCTCCAGAGACAATTCCAAGAACACGGTGTTCCTGCA
ACAGTCTGAGAGTTGAGGACACGGCTTTGTATTACTGTGCGAAAG
GGAGACTACAAAAGTGATGAGTGGGGGACGGAATATTATGATATT
TTCTTACCCTATTCAAGACCCACGGGCCATGGTTGGAGCTTTTGA
GGGGGCCAAGGGACAATGGTCACCGTCTCTCC

QSVLTQPPSVSAAPGQKVTISCSGSSSTIGNNYVSWYRLLPGTAPKLLIY

Patient Code	CAP256
Patient Sex	Female
Risk Factor	
Infection Country	
Infection City	
Infection Year	2005
HLA Type	A*2902, A*6601, B*1503, B*5802, Cw*0401, Cw*0602
Patient Ethnicity	
Progression	Rapid progressor (RP)
Species	human
Patient Note	Donor infected with subtype C; superinfected 13 to 15 weeks after infection with a second unrelated subtype C virus; diagnosed with extrapulmonary tuberculosis in November 2009 and started on ARV January 2010, 4 years 7 months postinfection; showed high level of cross-neutralizing antibody activity at 3 years of infection.
CTL CD8+ Records	
T-Helper CD4+ Records	
Antibody Records	2843, 3038, 3039, 3040, 3041, 3042, 3043, 3044, 3045, 3046, 3047, 3048, 3049
Sequence Database Patient ID Record	28376

B) Virus details

Virus name	Subtype	Country	Patient health	Days post infection	Days from seroconversion	Fiebig	Risk factor	Accession	Tier	Alias	HIV DB name	Seq data	LANL comments
25710_2_43	C	INDIA	acute infection	45		5	Heterosexual	EF117271	2	25710, 25710-2.43, HIV-25710-2.43	HIV_25710_2	Yes	This virus belongs to a 12-virus panel that was shown optimal for standardized assessments of neutralizing antibodies (deCamp2014 PMID 24352443).
3016_V5_C45	D	TANZANIA	acute infection		early	1 or 2	Heterosexual	HM215283	2	3016, 3016.v5.c45	3016_v5_c45	Yes	

Figure 1. Data retrieval. Examples of antibody and virus retrievable and downloadable details are shown. Green arrows and green text shown on Figures 1–3 represent comments added on the figure, but not present on the web site. (**A**) Antibody details include antibody structures in the Protein Data Bank; clickable donor ID from whom the antibody was isolated, which leads to donor information stored in the HIV database; Env positions related to antibody neutralization; and antibody variable chain sequences. PG9 is the antibody used in the example. The inset on the top of the figure shows an example of Env positions related to PG9 neutralization; Env positions shown in the inset have mutations that affect PG9 and related bNAb sensitivity. Inset on the right shows patient details for the donor from whom the antibody was isolated, and provides the link to HIV-1 sequences from that donor. (**B**) Virus details include virus name, subtype, country of isolation, patient health and risk status, GenBank accession number, neutralization tier, different names and aliases used in the literature, and our comments.
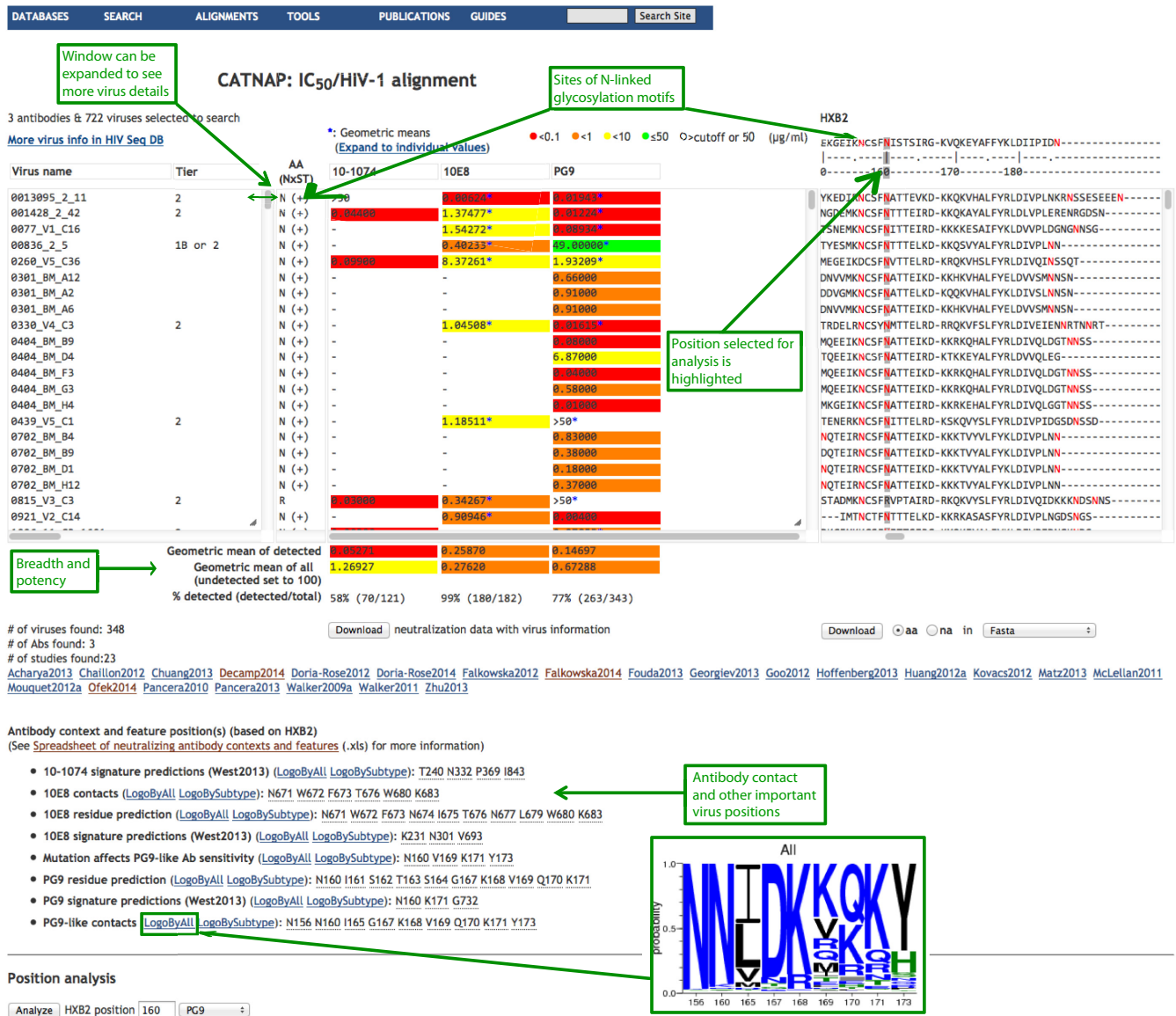
Figure 2. Analysis page. Three aligned windows are shown side-by-side. Each window has vertical and horizontal scroll bars, and each window can be expanded vertically and horizontally by clicking and dragging its right corner. All fields in each window are sortable, and all three windows are sorted simultaneously. For example, clicking on MAb 10E8 in the center window will sort all 3 windows by 10E8 neutralization scores. The left window lists virus names included in the analysis as well as neutralization tier, subtype, country, aliases (expandable data), and the presence of potential N-linked glycosylation motifs (denoted as NxST). The center window lists antibodies with their color-coded IC_{50} and/or IC_{80} neutralization scores. When multiple studies assayed the same antibody–virus combination, the geometric mean of the data points is shown and marked with an asterisk (*). The 'Expand' link shows the data points from multiple studies used to calculate the geometric means. A mouseover on each individual geometric mean will also show data-points together with the study references. The right window shows aligned Env sequences, with the HXB2 reference sequence on top. Potential N-glycosylation sites are highlighted in red. Additional information is provided under the three aligned windows above: breadth of neutralization and potency of each antibody over the selected group of viruses (potency is calculated two ways: for sensitive viruses that are below neutralization threshold of detection and for all tested viruses where resistant viruses are set to 100 μg/ml.); antibody contact HIV position(s) and a link to the sequence logo of these positions showing their variability in HIV-1 circulating sequences (see the inset). The option to analyze an individual HIV protein sequence position is on the bottom of the figure (this position is highlighted on the Env alignment in the right window with a gray vertical line). When a position of interest is entered, a new 'AA NxST' column appears between the virus and antibody windows. This column shows the amino acid found in each sequence at the position of interest, and the presence of an N-linked glycosylation motif in that position is indicated by '+'. N-linked glycosylation motifs are also highlighted in the alignment in red.
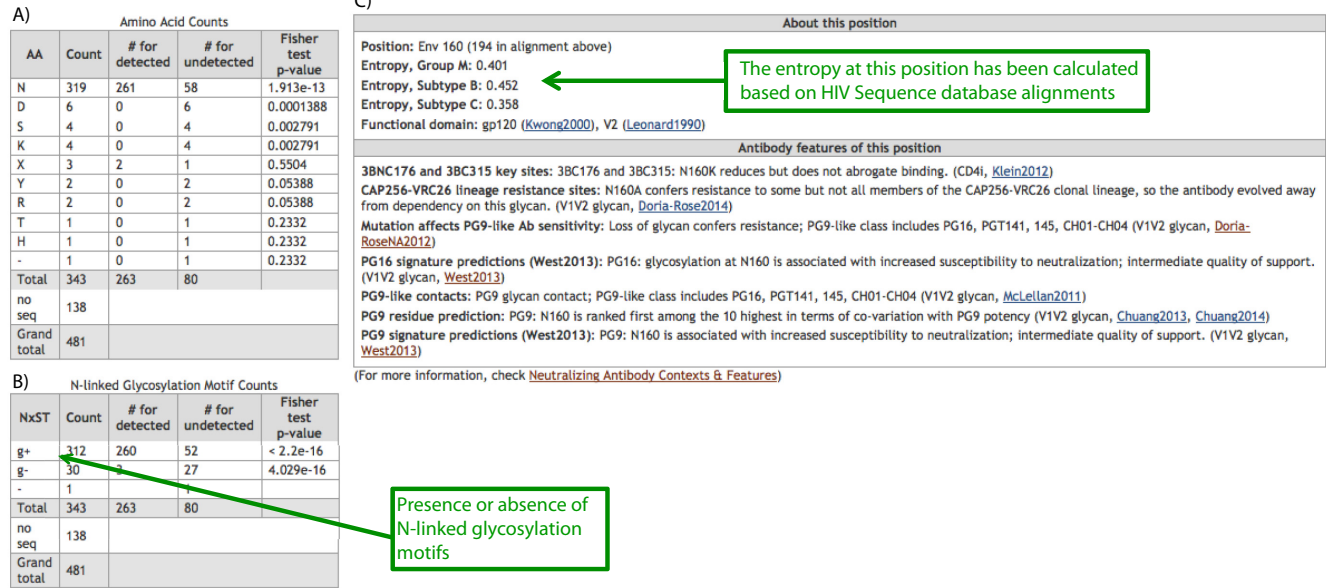
Analysis at HXB2 position 160 for Ab PG9

A)

Amino Acid Counts

AA	Count	# for detected	# for undetected	Fisher test p-value
N	319	261	58	1.913e-13
D	6	0	6	0.0001388
S	4	0	4	0.002791
K	4	0	4	0.002791
X	3	2	1	0.5504
Y	2	0	2	0.05388
R	2	0	2	0.05388
T	1	0	1	0.2332
H	1	0	1	0.2332
-	1	0	1	0.2332
Total	343	263	80	
no seq	138			
Grand total	481			

B)

N-linked Glycosylation Motif Counts

NxST	Count	# for detected	# for undetected	Fisher test p-value
g+	312	260	52	< 2.2e-16
g-	30	3	27	4.029e-16
-	1			
Total	343	263	80	
no seq	138			
Grand total	481			

Presence or absence of N-linked glycosylation motifs

C)

About this position

Position: Env 160 (194 in alignment above)
Entropy, Group M: 0.401
Entropy, Subtype B: 0.452
Entropy, Subtype C: 0.358
Functional domain: gp120 (Kwong2000), V2 (Leonard1990)

The entropy at this position has been calculated based on HIV Sequence database alignments

Antibody features of this position

3BNC176 and 3BC315 key sites: 3BC176 and 3BC315: N160K reduces but does not abrogate binding. (CD4i, Klein2012)
CAP256-VRC26 lineage resistance sites: N160A confers resistance to some but not all members of the CAP256-VRC26 clonal lineage, so the antibody evolved away from dependency on this glycan. (V1V2 glycan, Doria-Rose2014)
Mutation affects PG9-like Ab sensitivity: Loss of glycan confers resistance; PG9-like class includes PG16, PGT141, 145, CH01-CH04 (V1V2 glycan, Doria-RoseNA2012)
PG16 signature predictions (West2013): PG16: glycosylation at N160 is associated with increased susceptibility to neutralization; intermediate quality of support. (V1V2 glycan, West2013)
PG9-like contacts: PG9 glycan contact; PG9-like class includes PG16, PGT141, 145, CH01-CH04 (V1V2 glycan, McLellan2011)
PG9 residue prediction: PG9: N160 is ranked first among the 10 highest in terms of co-variation with PG9 potency (V1V2 glycan, Chuang2013, Chuang2014)
PG9 signature predictions (West2013): PG9: N160 is associated with increased susceptibility to neutralization; intermediate quality of support. (V1V2 glycan, West2013)

(For more information, check Neutralizing Antibody Contexts & Features)

Figure 3. Statistical analysis. (**A**) Counts of different amino acids at the chosen HXB2 position of ENV are shown, together with the counts of bNAb PG9 sensitive (detected, that is below threshold of detection) and resistant (undetected, that is above threshold of detection) viruses for each amino acid. (**B**) Counts of viruses with and without an N-linked glycosylation motif at the chosen HXB2 position of Env are shown. In (A) and (B), Fisher's exact test is based on a contingency table of 'AA/not AA' (or 'NxST/not NxST') and 'virus sensitive/not sensitive'. (**C**) Information from the HIV database about the position of interest is shown, and includes entropy calculations using HIV Database tool Entropy and HIV-1 alignments for M group, and B and C subtypes, information about functional domains, and antibody features of this position, such as how mutations in this position affect various bNAbs neutralization sensitivity, signature predictions at this position, mutations affecting antibody binding, etc.

ditional information are provided under the superimposed data (Figures 2 and 3), and include several options.

- Breadth and potency of each antibody/serum over the selected group of viruses are assessed, with potency calculated as a geometric mean of neutralization scores of both antibody-sensitive viruses and all viruses (Figure 2).
- List of Env-antibody contact positions is displayed for the selected antibodies, with the link to another HIV database tool, QuickAlign, which aligns these positions to a large set of sequences from the HIV Sequence Database and shows sequence logos (amino acid compositions and frequencies) and summaries of how variable these positions are in HIV circulating sequences (Figure 2).
- Antibody signature analysis options highlight a position of interest in the alignment (Figure 2), and for each sequence report the presence or absence of N-glycosylated motif for in this position (Figure 2), extract further information from the HIV databases and perform basic antibody signature analysis on this position (Figure 3). The extracted information includes entropy scores as a measure of amino acid variability in this position in M group, B clade and C clade database alignments, protein features such as functional domains associated with this position, antibody contacts and other NAb-associated features of this position, such as mutations affecting antibody sensitivity, signature predictions, and other features (Figure 3c). Statistical analysis displays the amino acid makeup of that position in the alignment (Figure 3a), the count of viruses with and without N-linked glycosylation

motif (Figure 3b), the count of antibody sensitive and resistant viruses for each amino acid or N-glycosylation motif found in this position of the alignment, together with the corresponding Fisher's exact test results to detect a potential positive or negative amino acid signature for the selected antibody (12). In Figure 3b, for example, an N-glycosylation motif at Env position 160 is enriched in bNAb PG9-detected viruses ($P < 2.2e-16$), indicating, in agreement with published studies (8,31), an association of a glycosylation site at position 160 with increased susceptibility to PG9 neutralization. This analysis should be interpreted with caution, as the counts are not phylogenetically corrected, and significant correlations that are not directly related to antibody sensitivity may arise as a consequence of structure in the phylogenetic tree. Still such associations can be informative, particularly for hypothesis forming or validation of a pre-existing hypothesis. A signature analysis tool that incorporates a phylogenetic correction (12) is in progress at the database.

Custom CATNAP

Custom CATNAP allows users to superimpose and analyze any table of numerical data in conjunction with genetic sequences (Figure 2), and to look for associated genetic signatures (Figure 3). Examples include IC_{50} neutralization values for monoclonal antibodies, like those used in the basic CATNAP tool, or ID_{50} titers for plasma neutralization data, but the tool can be used for *any* numerical data (binding kinetic data, functional data) in conjunction with sequence data. As neutralization data frequently use cut-

offs, with results below or above a threshold being considered undetected, for the statistical analysis the user should choose the type that describes the most potent scores in the data. For example, IC_{50} scores for the monoclonal antibodies reflect the concentration needed to achieve 50% neutralization, and so values with a '>' symbol are considered undetected. ID_{50} scores for the plasma antibodies, on the other hand, reflect the number of dilutions required to reduce the infectivity by 50%, and the high scores are the most potent, so the values with a '<' symbol are considered undetected. Like in basic CATNAP, potential genetic signatures are defined by Fisher's exact test ([12]). Users can use their own alignments, or can use a premade alignment that contains subsets of Env sequences in the CATNAP tool, including HIV-1 viral strains commonly used in neutralization panels.

DISCUSSION

Overall, we believe CATNAP is a useful tool for the HIV research community, particularly good for integrating and superimposing complete HIV sequence data and broadly neutralizing antibody data across many studies. It provides initial signature analysis on the fly, and it can be used to analyze custom data including genetic sequences in conjunction with numerical data. We consider CATNAP to be a foundation for a new neutralizing antibody relational database, and we plan to expand this resource to contain many more analysis tools and data tables, such as incorporating phylogenetic correction in evaluating antibody signatures in viral sequences ([12]), listing of germline antibody sequences, antibody variable sequence alignments, HIV-antibody 3D structure visualization tools, and within-patient studies with autologous viral and antibody sequence data.

FUNDING

REFERENCES

1. Gaschen,B., Taylor,J., Yusim,K., Foley,B., Gao,F., Lang,D., Novitsky,V., Haynes,B., Hahn,B.H., Bhattacharya,T. *et al.* (2002) Diversity considerations in HIV-1 vaccine selection. *Science*, **296**, 2354–2360.
2. Burton,D.R., Poignard,P., Stanfield,R.L. and Wilson,I.A. (2012) Broadly neutralizing antibodies present new prospects to counter highly antigenically diverse viruses. *Science*, **337**, 183–186.
3. Scheid,J.F., Mouquet,H., Ueberheide,B., Diskin,R., Klein,F., Oliveira,T.Y., Pietzsch,J., Fenyo,D., Abadir,A., Velinzon,K. *et al.* (2011) Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science*, **333**, 1633–1637.

4. Wu,X., Zhou,T., Zhu,J., Zhang,B., Georgiev,I., Wang,C., Chen,X., Longo,N.S., Louder,M., McKee,K. *et al.* (2011) Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. *Science*, **333**, 1593–1602.
5. deCamp,A., Hraber,P., Bailer,R.T., Seaman,M.S., Ochsenbauer,C., Kappes,J., Gottardo,R., Edlefsen,P., Self,S., Tang,H. *et al.* (2014) Global panel of HIV-1 Env reference strains for standardized assessments of vaccine-elicited neutralizing antibodies. *J. Virol.*, **88**, 2489–2507.
6. Seaman,M.S., Janes,H., Hawkins,N., Grandpre,L.E., Devoy,C., Giri,A., Coffey,R.T., Harris,L., Wood,B., Daniels,M.G. *et al.* (2010) Tiered categorization of a diverse panel of HIV-1 Env pseudoviruses for assessment of neutralizing antibodies. *J. Virol.*, **84**, 1439–1452.
7. Montefiori,D.C. (2005) Evaluating neutralizing antibodies against HIV, SIV, and SHIV in luciferase reporter gene assays. *Curr. Protoc. Immunol.*, Chapter 12, Unit 12.11.
8. West,A.P. Jr, Scharf,L., Horwitz,J., Klein,F., Nussenzweig,M.C. and Bjorkman,P.J. (2013) Computational analysis of anti-HIV-1 antibody neutralization panel data to identify potential functional epitope residues. *Proc. Natl .Acad. Sci. U.S.A.*, **110**, 10598–10603.
9. Eroshkin,A.M., LeBlanc,A., Weekes,D., Post,K., Li,Z., Rajput,A., Butera,S.T., Burton,D.R. and Godzik,A. (2014) bNAber: database of broadly neutralizing HIV antibodies. *Nucleic Acids Res.*, **42**, D1133–D1139.
10. Chuang,G.Y., Acharya,P., Schmidt,S.D., Yang,Y., Louder,M.K., Zhou,T., Kwon,Y.D., Pancera,M., Bailer,R.T., Doria-Rose,N.A. *et al.* (2013) Residue-level prediction of HIV-1 antibody epitopes based on neutralization of diverse viral strains. *J. Virol.*, **87**, 10047–10058.
11. Chuang,G.Y., Liou,D., Kwong,P.D. and Georgiev,I.S. (2014) NEP: web server for epitope prediction based on antibody neutralization of viral strains with diverse sequences. *Nucleic Acids Res.*, **42**, W64–W71.
12. Gnanakaran,S., Daniels,M.G., Bhattacharya,T., Lapedes,A.S., Sethi,A., Li,M., Tang,H., Greene,K., Gao,H., Haynes,B.F. *et al.* (2010) Genetic signatures in the envelope glycoproteins of HIV-1 that associate with broadly neutralizing antibodies. *PLoS Comput. Biol.*, **6**, e1000955.
13. Kepler,T.B. (2013) Reconstructing a B-cell clonal lineage. I. Statistical inference of unobserved ancestors. *F1000Res*, **2**, 103.
14. Ye,J., Ma,N., Madden,T.L. and Ostell,J.M. (2013) IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res.*, **41**, W34–W40.
15. Giudicelli,V., Brochet,X. and Lefranc,M.P. (2011) IMGT/V-QUEST: IMGT standardized analysis of the immunoglobulin (IG) and T cell receptor (TR) nucleotide sequences. *Cold Spring Harb. Protoc.*, 695–715.
16. Brochet,X., Lefranc,M.P. and Giudicelli,V. (2008) IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Res.*, **36**, W503–W508.
17. Gaeta,B.A., Malming,H.R., Jackson,K.J., Bain,M.E., Wilson,P. and Collins,A.M. (2007) iHMMune-align: hidden Markov model-based alignment and identification of germline genes in rearranged immunoglobulin gene sequences. *Bioinformatics*, **23**, 1580–1587.
18. Souto-Carneiro,M.M., Longo,N.S., Russ,D.E., Sun,H.W. and Lipsky,P.E. (2004) Characterization of the human Ig heavy chain antigen binding complementarity determining region 3 using a newly developed software algorithm, JOINSOLVER. *J. Immunol.*, **172**, 6790–6802.
19. Liao,H.X., Lynch,R., Zhou,T., Gao,F., Alam,S.M., Boyd,S.D., Fire,A.Z., Roskin,K.M., Schramm,C.A., Zhang,Z. *et al.* (2013) Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature*, **496**, 469–476.
20. Diskin,R., Scheid,J.F., Marcovecchio,P.M., West,A.P. Jr, Klein,F., Gao,H., Gnanapragasam,P.N., Abadir,A., Seaman,M.S., Nussenzweig,M.C. *et al.* (2011) Increasing the potency and breadth of an HIV antibody by using structure-based rational design. *Science*, **334**, 1289–1293.
21. Pejchal,R., Doores,K.J., Walker,L.M., Khayat,R., Huang,P.S., Wang,S.K., Stanfield,R.L., Julien,J.P., Ramos,A., Crispin,M. *et al.* (2011) A potent and broad neutralizing antibody recognizes and penetrates the HIV glycan shield. *Science*, **334**, 1097–1103.
22. Falkowska,E., Ramos,A., Feng,Y., Zhou,T., Moquin,S., Walker,L.M., Wu,X., Seaman,M.S., Wrin,T., Kwong,P.D. *et al.*

(2012) PGV04, an HIV-1 gp120 CD4 binding site antibody, is broad and potent in neutralization but does not induce conformational changes characteristic of CD4. *J. Virol.*, **86**, 4394–4403.

23. Blattner,C., Lee,J.H., Sliepen,K., Derking,R., Falkowska,E., de la Pena,A.T., Cupo,A., Julien,J.P., van Gils,M., Lee,P.S. *et al.* (2014) Structural delineation of a quaternary, cleavage-dependent epitope at the gp41-gp120 interface on intact HIV-1 Env trimers. *Immunity*, **40**, 669–680.

24. Scharf,L., Scheid,J.F., Lee,J.H., West,A.P. Jr, Chen,C., Gao,H., Gnanapragasam,P.N., Mares,R., Seaman,M.S., Ward,A.B. *et al.* (2014) Antibody 8ANC195 reveals a site of broad vulnerability on the HIV-1 envelope spike. *Cell Rep.*, **7**, 785–795.

25. McLellan,J.S., Pancera,M., Carrico,C., Gorman,J., Julien,J.P., Khayat,R., Louder,R., Pejchal,R., Sastry,M., Dai,K. *et al.* (2011) Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. *Nature*, **480**, 336–343.

26. Zhou,T., Georgiev,I., Wu,X., Yang,Z.Y., Dai,K., Finzi,A., Kwon,Y.D., Scheid,J.F., Shi,W., Xu,L. *et al.* (2010) Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. *Science*, **329**, 811–817.

27. Zhou,T., Xu,L., Dey,B., Hessell,A.J., Van Ryk,D., Xiang,S.H., Yang,X., Zhang,M.Y., Zwick,M.B., Arthos,J. *et al.* (2007) Structural definition of a conserved neutralization epitope on HIV-1 gp120. *Nature*, **445**, 732–737.

28. Kwong,P.D., Wyatt,R., Robinson,J., Sweet,R.W., Sodroski,J. and Hendrickson,W.A. (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature*, **393**, 648–659.

29. O'Rourke,S.M., Schweighardt,B., Phung,P., Mesa,K.A., Vollrath,A.L., Tatsuno,G.P., To,B., Sinangil,F., Limoli,K., Wrin,T. *et al.* (2012) Sequences in glycoprotein gp41, the CD4 binding site, and the V2 domain regulate sensitivity and resistance of HIV-1 to broadly neutralizing antibodies. *J. Virol.*, **86**, 12105–12114.

30. Postler,T.S. and Desrosiers,R.C. (2013) The tale of the long tail: the cytoplasmic domain of HIV-1 gp41. *J. Virol.*, **87**, 2–15.

31. Doria-Rose,N.A., Georgiev,I., O'Dell,S., Chuang,G.Y., Staupe,R.P., McLellan,J.S., Gorman,J., Pancera,M., Bonsignori,M., Haynes,B.F. *et al.* (2012) A short segment of the HIV-1 gp120 V1/V2 region is a major determinant of resistance to V1/V2 neutralizing antibodies. *J. Virol.*, **86**, 8319–8323.