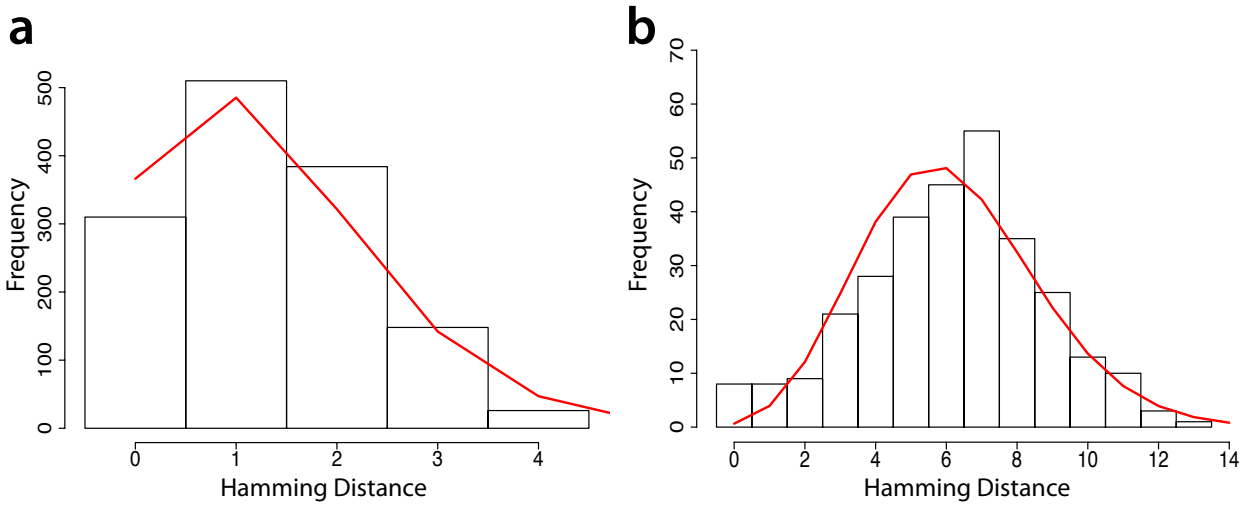
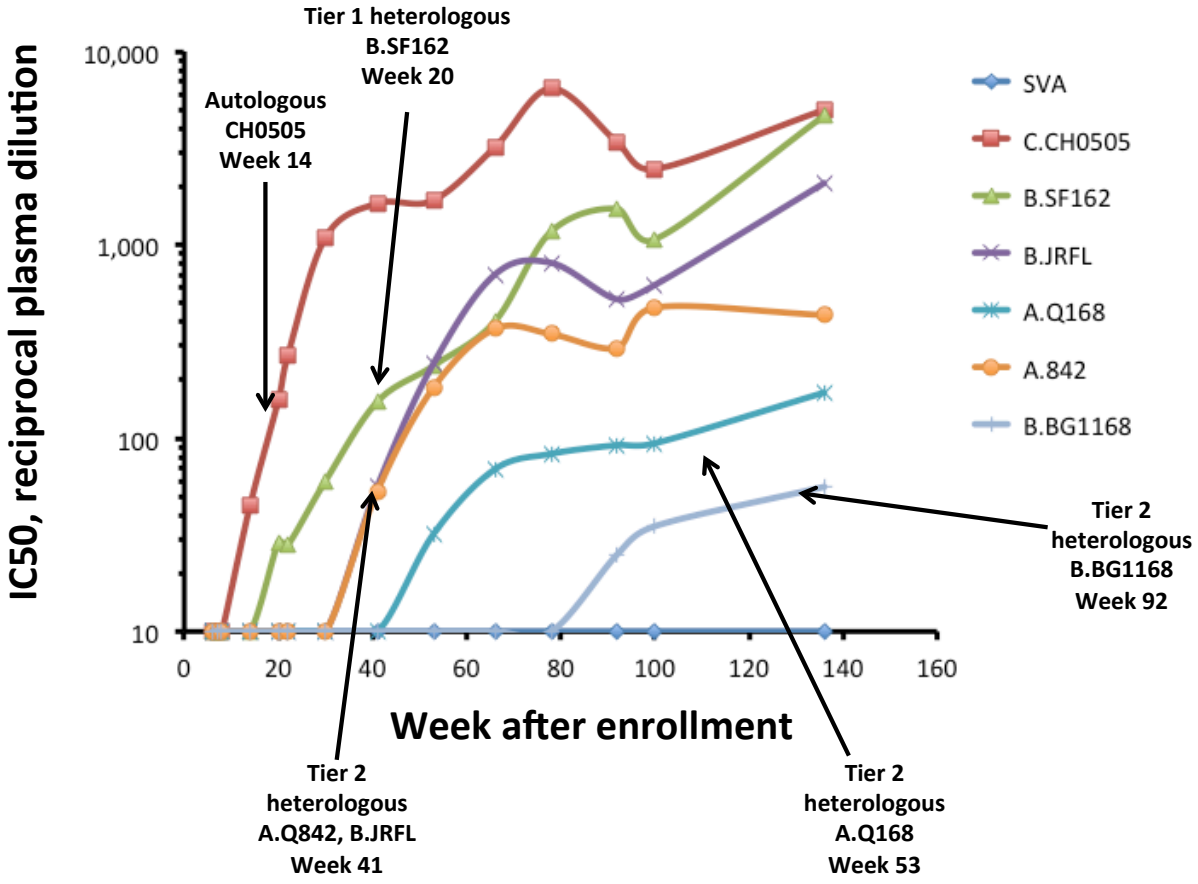
SUPPLEMENTARY INFORMATION

doi:10.1038/nature12053

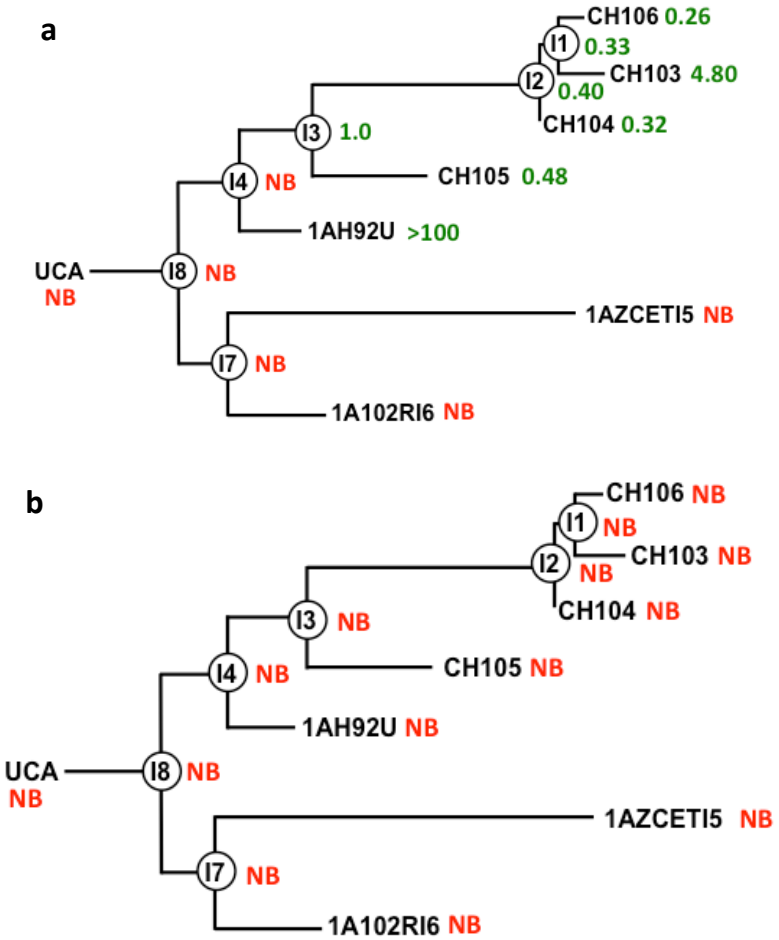


Supplementary Figure 1. Hamming distance frequency distributions of sequences at (a) week 4 and (b) week 14. A model of the best fit Poisson distribution is shown as a red line. Analysis of the sequence diversity in the first available sample (a) from subject CH505 using the Poisson Fitter tool (ref below) indicates that the sequences were a consistent with a star phylogeny and that the mutations were accumulating according to a Poisson distribution (goodness of fit p = 0.11). This is consistent with a single founder virus establishing the infection, with random accumulation of mutations prior to selection. The lambda parameter was 1.325, and assuming the mutation rate of 2.16 10^{-05}, the estimated time from the most recent common ancestor was 22 days (95% CI, 18-27). Given that the outer bound of this confidence interval is 27 days, it is highly like this sample was taken within 4 weeks of infection, thus we are calling this sampling time "week 4" as a conservative estimate. This timing estimate is further supported by Feibig staging at time of enrollment. By week 14 (b), the tree was no longer consistent with a star phylogeny or a Poisson distribution ($p << 10^{-10}$), indicating selection was well underway. Of note, although the mutation data at week 4 (a) is statistically consistent with a Poisson distribution, the observed number of pairwise sequence identities was somewhat reduced relative to expectation, and the observed number of Hamming distances of 1 and 2 are slightly more than expected. This is of interest as this shift is the a result of a single mutation in loop D, in a CH103 contact residue (N279K) -- so although the deviation from the Poisson was not significant, given its location it is possible that the site is a very early indicator of selection.

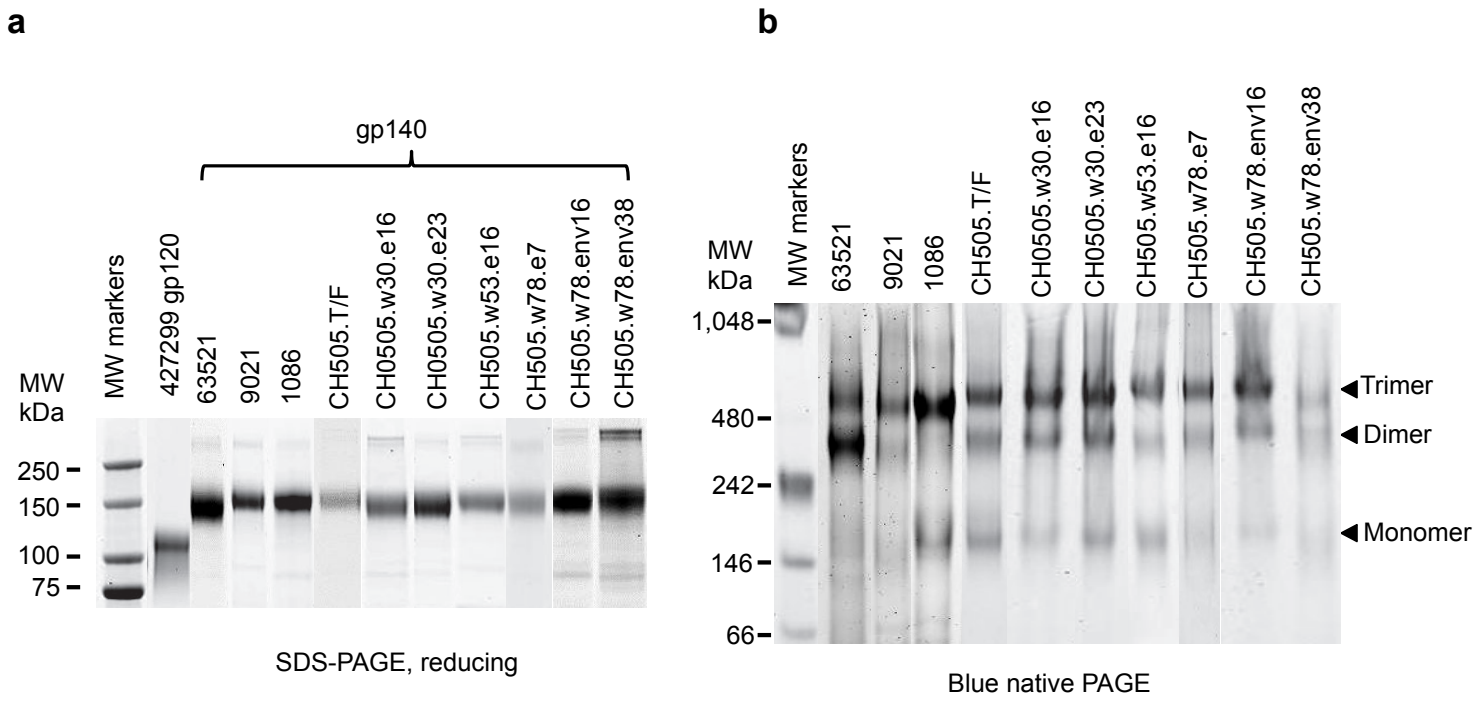
Reference: Giorgi EE, Funkhouser B, Athreya G, Perelson AS, Korber BT, Bhattacharya T. Estimating time since infection in early homogeneous HIV-1 samples using a Poisson model. BMC Bioinformatics 2010 Oct 25;11:532. PMID: 20973976
http://www.hiv.lanl.gov/content/sequence/POISSON_FITTER/poisson_fitter.html



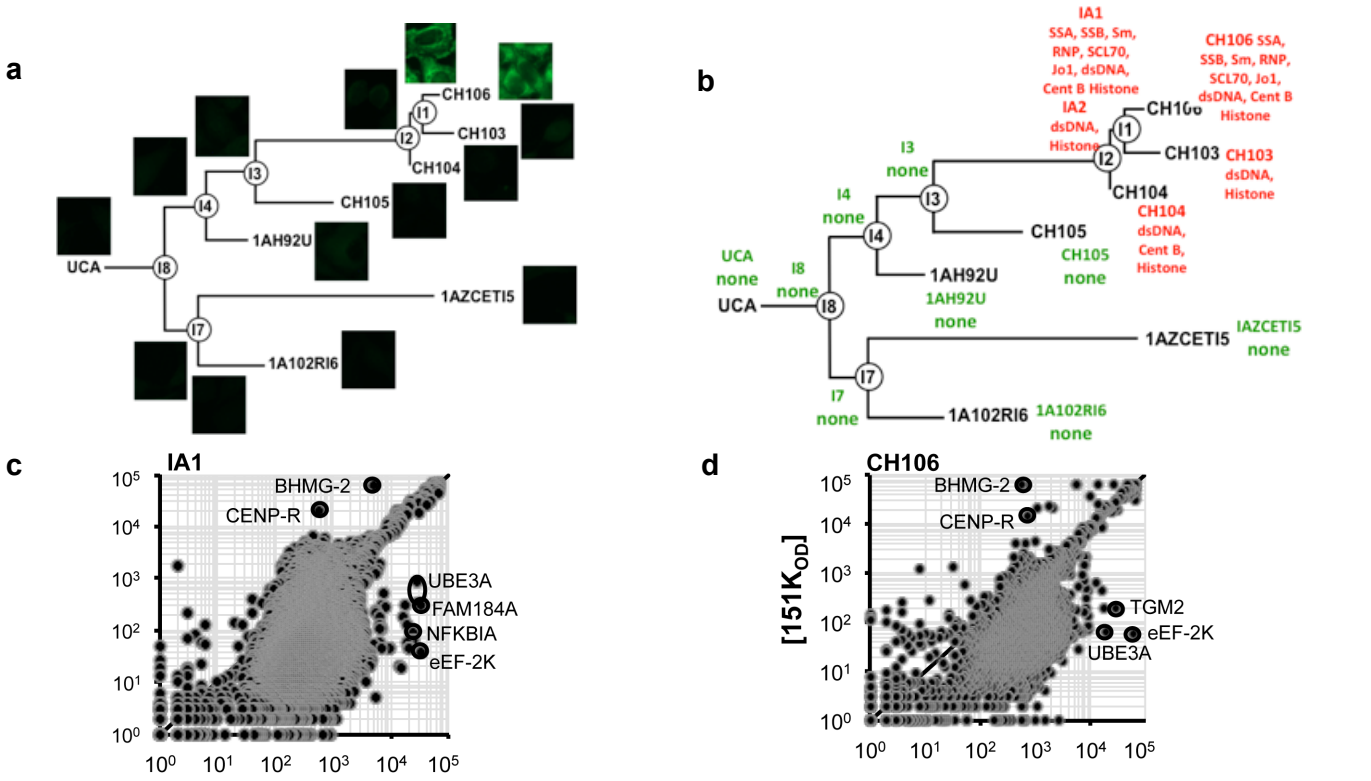
Supplementary Figure 2. Plasma neutralization of CH505 patient over time to autologous transmitted/founder (T/F) and heterologous HIV-1 viruses. Plasma samples were longitudinally collected from HIV-1 patient CH505 starting from time of infection(in x axis) and tested for neutralization activity against the autologous transmitted/founder (T/F) virus and heterologous HIV-1 Env pseudoviruses including subtype B (B) SF162, JRFL and BG1168) and subtype A (Q168 and 842) in TZM-bl cell-based neutralization assays. Results were expressed as IC50 (reciprocal plasma dilution) (in y axis).



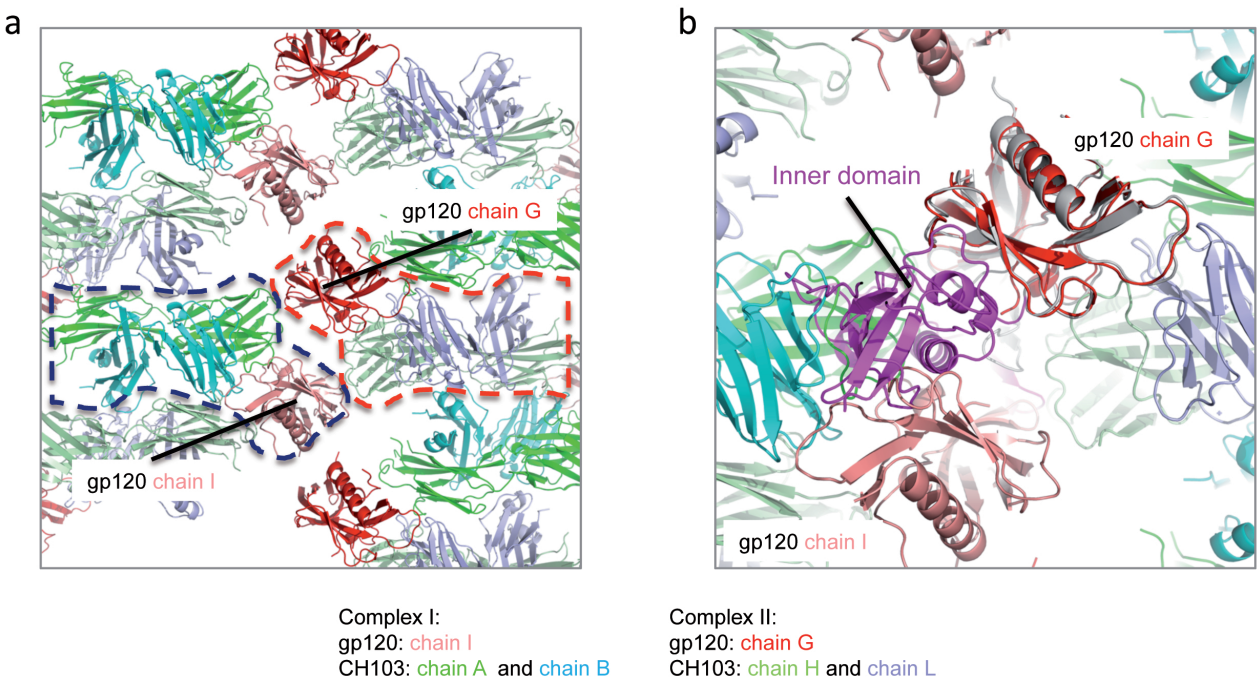
Supplementary Figure 3. Reactivity of antibodies in CH103 clonal lineage to HIV-1 Env resurfaced core3 (RSC3) and RSC3 mutant. Antibodies in CH103 clonal lineage were tested in dose range from 100ug to 0.0005ug/ml for binding to (**a**) HIV-1 Env RSC3 and (**b**) RSC3 mutant with P363N and D371I mutations in ELISA. Results are expressed as EC50 (ug/ml) and are indicated next to the individual antibodies. NB = no detectable binding.



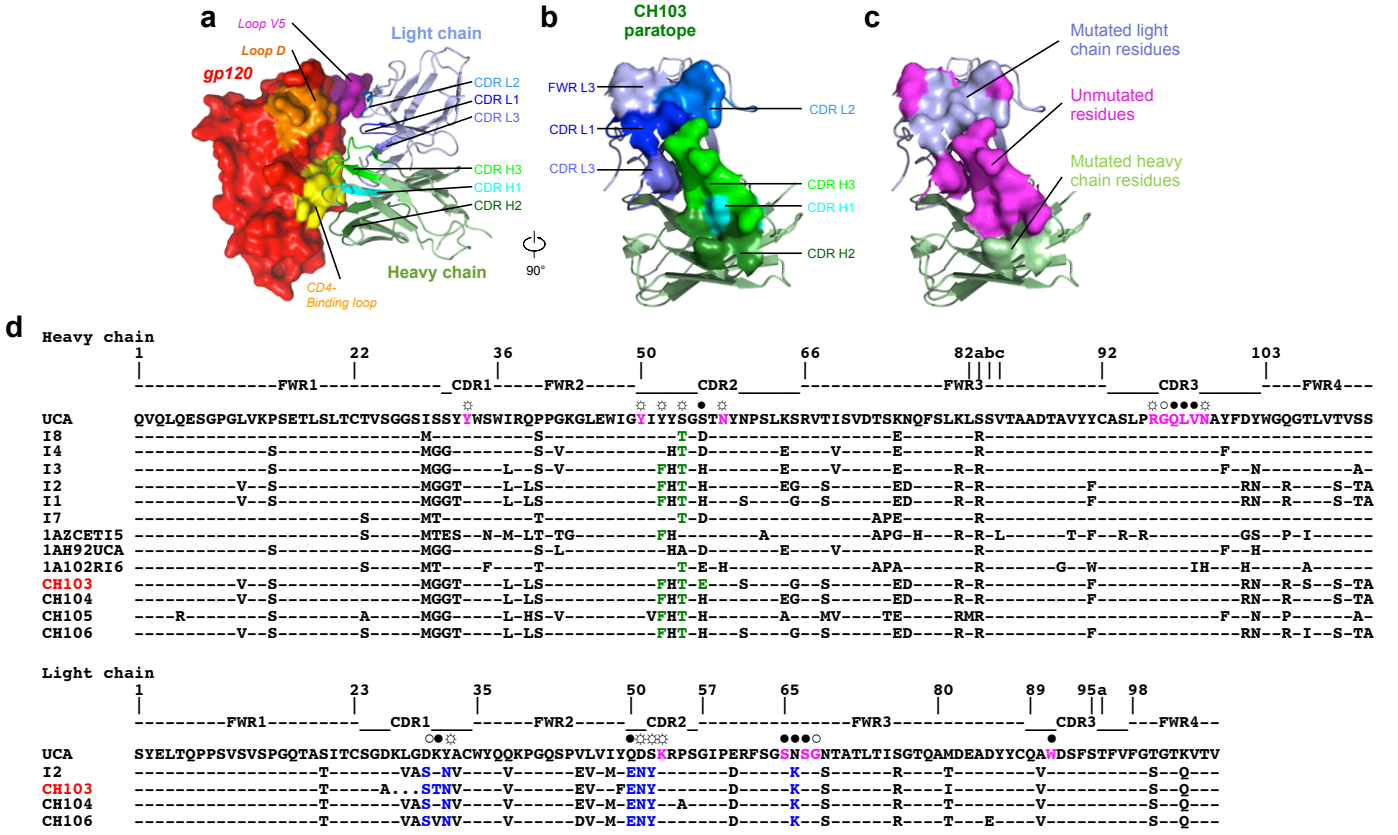
Supplementary Figure 4. SDS-PAGE analysis of recombinant HIV-1 Env gp140 and gp120 proteins. HIV-1 Env gp140 and gp120 proteins were analyzed on SDS-PAG under reducing condition (**a**) and gp140 proteins were analyzed on blue negative PAGE (**b**). Individual HIV-1 Env proteins are identified on the top of gels. **a,** The HIV-1 gp120 and gp140 used in the study had no degradation under reducing condition in SDS-PAGE. **b,** Most heterologous HIV-1 Env gp140 Envs and all autologous CH505 gp140 Envs migrated predominantly as trimers and also contain dimer and monomer forms.



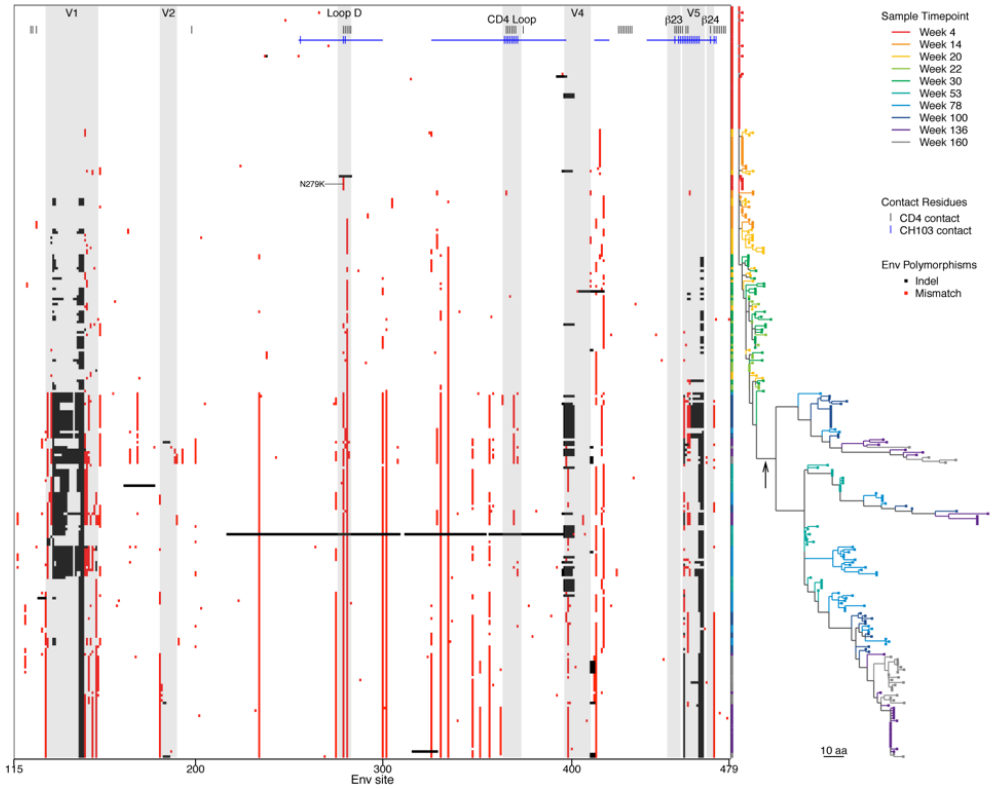
Supplementary Figure 5. Polyreactivity analysis of antibodies in CH103 clonal lineage by HEp-2 staining, ANA assays and protein array microchip analysis. Reactivity of antibodies in CH103 clonal lineage was assayed by indirect immunofluorescence HEp-2 staining (**a**) and by ANA assays (**b**). Pictures at magnification x 200 of immunofluorescence staining for individual antibodies are presented next to the antibody ID. Results of the reactivity of individual antibodies with panel of autoantigens assayed by ANA are indicated (**b**). The intermediate antibody (I1) and CH106 were identified as reactive with HEp-2 cells and then selected for further testing for reactivity with human host cellular antigens (**c** and **d**) using Invitrogen ProtoArrays™ . We found that I1 (**c**) and CH106 ((**d**) exhibit specific autoreactivity and robust polyreactivity. Bound antibody was determined by immunofluorescence and relative fluorescence intensities for 9,400 recombinant human proteins in the 151K array (y-axis) that were plotted against (x-axis) the homologous intensities in IA1 (**c**) and CH106 (**d**) arrays. All proteins were printed in duplicate on each array and each data point represents one fluorescence measurement. The diagonals in each graph represent equal fluorescence intensities (equivalent binding) by the I1, CH106 and 151K control Ab. Self-antigens bound by the I1 and CH106 are identified by high fluorescence intensity versus 151K and are indicated by circles. Polyreactivity was indicated by significant and general skewing from the diagonal. Autoantigens identified: BHMT2 (betaine-homocysteine methyltransferase 2); CENP-R (centromere protein R) [151K]; eEF-2K (eukaryotic elongation factor-2 kinase); UBE3A (ubiquitin-protein ligase E3A) [IA1 and CH106]; TGM2 (transglutaminase 2) [CH106]; NFKBIA (nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha); FAM184A (family with sequence similarity 184, member A) [I1].



Supplementary Figure 6. Crystal packing of the CH103-gp120 complex in P21 space group. a, A view of the crystal lattice. The two complexes in each asymmetric unit are marked with red and blue dashed lines and are shown in cartoon diagrams with gp120 in red and salmon, CH103 heavy chain in green and palegreen, and CH103 light chain in light blue and cyan. **b**, A close-up view of the lattice between two neighboring complexes. When extended core gp120 of clade C ZM176.66 from the VRC01 complex is superposed with its ordered corresponding portions in the CH103 complex, the inner domain shown in magenta clashes with the neighboring complex, indicating inner domain of gp120 is not present in the CH103-gp120 crystal due to proteolytic degradation during crystal growth.

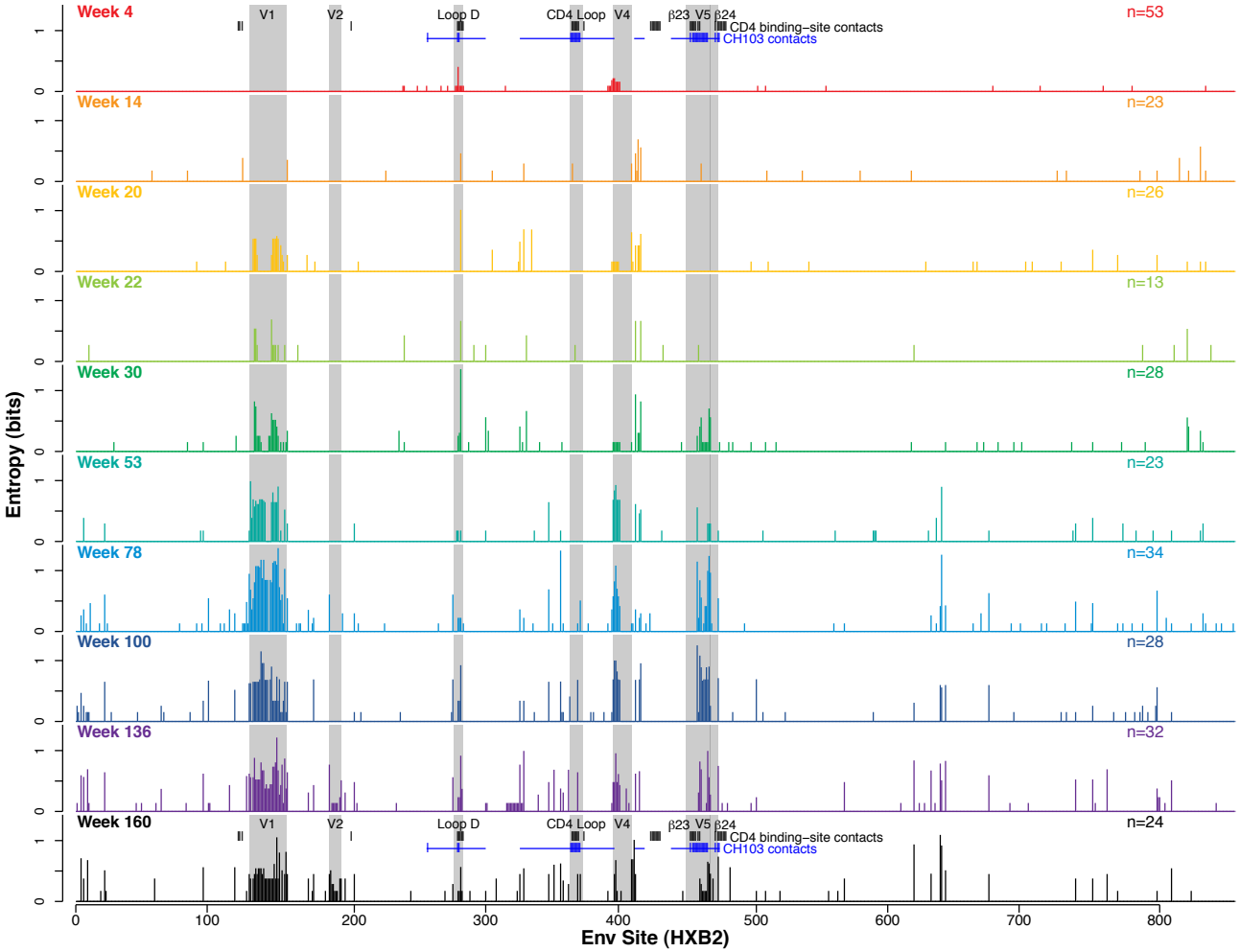


Supplementary Figure 7. **CH103 paratope, critical residues, and required immune precursors. a**, Overall structure of complex with variable domains of CH103 depicted in ribbon representation and gp120 shown as a molecular surface. The color scheme is the same as in Fig. 3a. **b**, CH103 paratope surface displayed on top of an underlying polypeptide ribbon. The surface is colored and labeled by contributing antibody components. **c**, CH103 paratope surface colored by maturation states of the underlying residues. Unmutated residues are colored magenta while affinity matured residues are colored green and light blue for heavy and light chains respectively. **d**, Sequence alignment of heavy and light chains of CH103 clonal lineage members. Framework and CDR residues are labeled, as are residues that interact with the gp120 (open circle, main chain interaction; open circle with rays, side chain interactions; filled circle, both main chain and side chain interactions). The unmutated paratope residues are highlighted in magenta and the maturation-gained paratope residues are highlighted in green for heavy chain and blue for light chain.

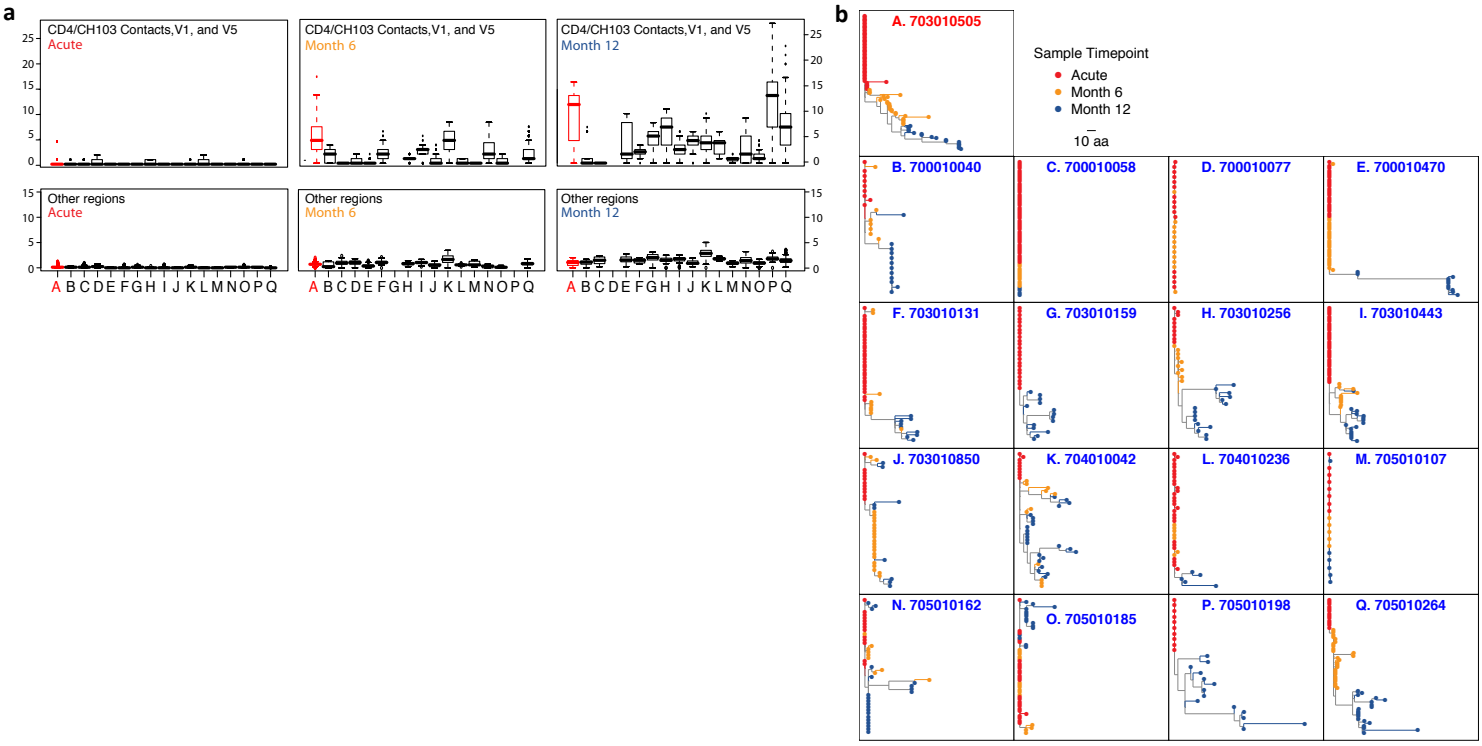


Supplementary Figure 8. Pixel map and phylogenetic tree of HIV-1 *env* gene evolution over time in CH505. The pixel tool (http://www.hiv.lanl.gov/content/sequence/pixel/pixel.html) was used to illustrate the amino acid changes in the V1 to V5 region of the envelope. we focused on this region as it is most critical for CD4bs antibody susceptibility, and includes all known CD4 binding contacts, which are indicated as black tic marks along the top of the figure. Blue tic marks indicated CH103 contact residues, and the horizontal blue line indicates that part of gp120 that was used for the CH103 crystal structure (although the contact surface is mostly there, still quite a bit is missing that is important for CD4 and VRC01, which is why we use CD4 contacts to help define bits that may be important for CH103 binding in those missing regions). Each row represents one sequence, and they are ordered according to the phylogeny. Red bits indicate amino acid changes relative to the TF virus, which was inferred as previously described [1,2]. Black bits indicate either an insertion or a deletion. The phylogenetic tree on the right was made with PhyML v2 [3] and the JTT substitution model [4] from the translated Env sequences. The tree was configured as a ladder and the T/F virus was reconstructed from the first time point sequences obtained at week 4 after transmission. Colors indicated the estimated number of weeks from infection. The tree was rendered with APE v3.0-6 [5] and both used R v2.15.1 [6]. The arrow indicates the week 30-53 selective bottleneck.

1. Keele BF et al. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc. Natl. Acad. Sci. USA, 105: 7552-7557 (2008).
2. Salazar-Gonzalez JF et al. Genetic identity, biological phenotype and evolutionary pathways of transmitted founder viruses in acute and early HIV-1 infection. J. Exp. Med. 206:1273-1289 (2009).
3. Guindon S, Gascuel O. A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52:696–704 (2003).
4. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. Comput Applic Biosci 8: 275–282 (1992).
5. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20: 289–290 (2004).
6. R Core Team. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

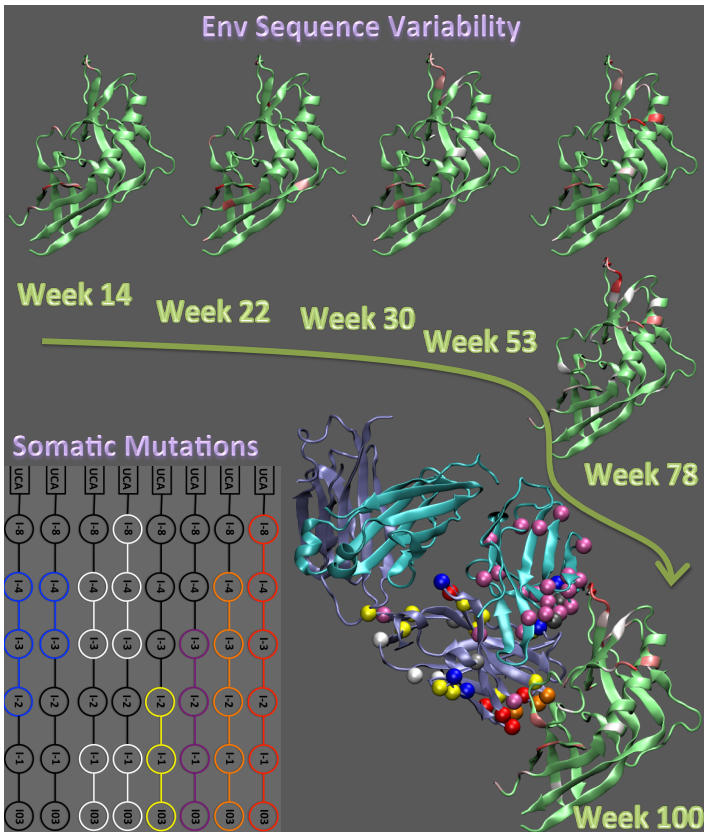


Supplementary Figure 9. Entropy map illustrating the per site diversity within each time point sampled in CH505. Full gp160 is shown, and CD4 and CH103 contact residues are highlighted. Shown is the Shannon entropy of each position in the alignment, where the observed frequency of all in a position characters is considered, and a gap is treated as a character (Korber J Virol. 1994;68(11):7467-81). This provides a map of regional within-time point diversity spanning Env, and illustrates where mutations are concentrated and the relative diversity of key regions over time.



Supplementary Figure 10. A comparison of the pace of viral sequence evolution in CH505 (indicated here by the 9-digit anonymous study-participant identifier 703010505) in regions relevant to the CH103 epitope with other subjects. The regions of interest include the CH103 contacts defined by the structure in this paper, as well as VRC01 contacts and CD4bs contacts[1], and the V1 and V5 loops immediately adjacent to these contacts. a, The distribution of sequence distances expressed as the percentage of amino acids that are different between two sequences, resulting from a pair-wise comparison of all sequences sampled in a given time point. Because these are all homogeneous (single-founder) infection cases, very few mutations appear in the CH103 relevant regions or other sites in the virus during acute infection (left hand panels). By 24 weeks after enrollment (week 30 from infection in (A) 703010505, labeled month 6 here as it is approximate), extensive mutations have begun to accrue, focused in CH103 relevant regions (top middle panel), but not in other regions of Env (bottom middle panel). Subject 703010505 has the highest ranked diversity among 15 subjects[2] (B-Q) sampled in this time frame (p=0.067), indicating a focused selective pressure began unusually early in this subject. By 1 year (month 12 indicates samples taken between 10-14 months from enrollment, due to variation in timing of patient visits), this region has begun to evolve in many individuals, possibly due to autologous NAb responses active later in infection. b, Phylogenetic trees based on concatenated CH103 relevant regions (HXB2 sites 124-127, 131, 132, 279-283, 364-371, 425-432, 455-465, 471-477) were created with PhyML3.0[3], using HIVw[4], a within-subject HIV protein substitution model, which was selected to be the optimum model for these sequences using ProtTest[5]. Indels were treated as an additional character state, rather than as missing information. In this view, the extensive evolution away from the T/F virus by month 6, shown in gold, is particularly striking. Distances between sequences sampled in 703010505 (A) at month 6 and the T/F ancestral state were significantly greater than the sequences in the next most variable individual (L) designated by the 9-digit identifier 704010042 (Wilcoxon rank sum, p = 0.0003: CH505, median = 0.064, range = 0.019 – 0.13, N = 25, and 704010042, median = 0.0271, range = 0.009 – 0.056, N = 26).

1. Zhou, T., *et al*. Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. *Science* **329**, 811-817 (2010).
2. Liu MK, *et al*. Vertical T cell immunodominance and epitope entropy determine HIV-1 escape. *J Clin Invest*. **123**, 380-93 (2013).
3. Guindon, S., *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* **59**, 307-321 (2010).
4. Nickle, D.C., *et al.* HIV-specific probabilistic models of protein evolution. *PLoS ONE* **2**, e503 (2007).
5. Abascal, F., Zardoya, R. & Posada, D. ProtTest: selection of best-fit models of protein evolution. Bioinformatics **21**, 2104-2105 (2005).



Supplementary Figure 11. Co-evolution of virus and antibody - interplay between maturation of antibody CH103 and sequence variability epitope in gp120. The sequence variability (within sample) at each time point is mapped on a gp120 structure that tracks the viral evolution over time from 14 weeks through 100 weeks post-transmission. Entropy at each residue is color-coded as green to white to red to indicate no sequence variation to slight variation to high sequence variation. This extensive virus within-time point diversity coincides with maturing antibody lineages that ultimately develop breadth. Here we capture the somatic mutations along the CH103 clonal lineage beginning with unmutated common ancestor (UCA) to I-8 to I-4 to I-3 to I-2 to I-1 and to mature CH103. The color balls in heavy (violet) and light (cyan) chains of antibody indicate the appearance/disappearance of somatic mutations during the evolutionary path according to the following scheme. Red balls: mutations appeared in I-8 and remained all the way thru maturation to CH103, Orange balls: mutations appeared in I-4 and remained thru maturation to CH103, Blue balls: mutations that appeared early but are lost before maturing to CH103, and Gray balls: mutations that appeared very late in maturation. Structure of Fab CH103-gp120 from ZM176.66 complex determined in this work is used to map these mutations. The sequence entropy at week 100 is used to pair with CH103 to simply illustrate the relative spatial locations of somatic mutations and sequence variability in gp120. As discussed in the text, viral evolution with time tracks with neutralization breadth and this simple mapping supports that (i) T/F virus began to diversify very early in regions in or proximal to the epitope, and (ii) Somatic mutations that occur early in evolution and remain fixed in heavy chain tend to cluster near the gp120 contact region unlike those mutations that appear later.

Supplementary Table 1. Plasma neutralization activity developed over time of in patient CH505 against the autologous transmitted/funder (T/F) and heterologous viruses.

CH505 Plasma Week after Infection	EC50, reciprocal dilution						
	MuLV	CH505	B.SF162	B.JRFL	A.Q168	A.842	B.BG1168
6	<20	<20	<20	<20	<20	<20	<20
7	<20	<20	<20	<20	<20	<20	<20
8	<20	<20	<20	<20	<20	<20	<20
14	<20	**45**	<20	<20	<20	<20	<20
20	<20	**157**	29	<20	<20	<20	<20
22	<20	**267**	28	<20	<20	<20	<20
30	<20	**1,291**	**60**	<20	<20	<20	<20
41	<20	**1,636**	**154**	**56**	<21	**53**	<20
53	<20	**1,701**	**237**	**244**	**32**	**183**	<20
66	<20	**3,193**	**401**	**701**	**69**	**367**	<20
78	<20	**6,428**	**1,172**	**806**	**83**	**345**	<20
92	<20	**3,396**	**1,534**	**522**	**92**	**293**	**25**
100	<20	**2,464**	**1,066**	**619**	**94**	**473**	**35**
136	<20	**4,985**	**4,651**	**2,085**	**172**	**433**	**56**
138	<20	**3,586**	**5,081**	**1,368**	**138**	**326**	**51**
140	<20	**3,374**	**13,407**	**1,287**	**148**	**237**	**56**
144	<20	**4,665**	**8,354**	**905**	**118**	**237**	**40**
152	<20	**1,789**	**3,122**	**1,612**	**108**	**234**	**35**
160	<20	**2,684**	**9,761**	**2,482**	**144**	**230**	**55**
176	<20	**2,003**	**5,148**	**2,243**	**91**	**150**	**58**
208	<20	**1,353**	**5,850**	**1,303**	**60**	**95**	**31**
233	<20	**3,279**	**3,612**	**895**	**107**	**151**	**37**
234	<20	**3,033**	**4,887**	**1,712**	**103**	**232**	**60**
236	<20	**1,969**	**4,417**	**1,354**	**107**	**299**	**57**
mAb 2F5*	>50	>50	**0.69**	**2.26**	**1.62**	**11.71**	**1.43**

*EC50 values for positive control antibody 2F5 are presented as ug/ml. MuLV = murine leukemia virus as negative control.

Supplementary Table 2. V(D)J rearrangement of the matured, and reverted unmutated ancestor and intermediate antibodies in CH103 clonal lineage.

Antibody ID	IgH_ID	VH	DH	JH	Mutation frequency	CDRH3 length[1]	Isotype	VL ID	VL	JL	Mutation frequency	CDRL3 length[1]
UCA	UCAVH	4-59*01	3-16*01	4*02	0.0%	15	IgG1	UCAVL	3-1*01	1*01	0%	10
I8	I8VH	4-59*01	3-16*01	4*02	3.6%	15	IgG1	UCAVL	3-1*01	1*01	0%	10
I7	I7VH	4-59*01	3-16*01	4*02	5.0%	15	IgG1	UCAVL	3-1*01	1*01	0%	10
I4	I4VH	4-59*01	3-16*01	4*02	6.9%	15	IgG1	UCAVL	3-1*01	1*01	0%	10
I3	I3VH	4-59*01	3-16*01	4*02	9.1%	15	IgG1	I2VL	3-1*01	1*01	10.0%	10
I2	I2VH	4-59*01	3-16*01	4*02	14.9%	15	IgG1	I2VL	3-1*01	1*01	10.0%	10
I1	I1VH	4-59*01	3-16*01	4*02	15.2%	15	IgG1	I2VL	3-1*01	1*01	10.0%	10
1AZCETI5	1AZCETI5VH	4-59*01	3-16*01	4*02	15.2%	15	IgG1	I2VL	3-1*01	1*01	10.0%	10
1AH92U	1AH92UVH	4-59*01	3-16*01	4*02	8.3%	15	IgG1	UCAVL	3-1*01	1*01	0%	10
1A102RI6	1A102RI6VH	4-59*01	3-16*01	4*02	7.7%	15	IgG1	I2VL	3-1*01	1*01	10.0%	10
CH106	CH106VH	4-59*01	3-16*01	4*02	16.0%	15	IgG1	CH106VL	3-1*01	1*01	11.2%	10
CH103	CH103VH	4-59*01	3-16*01	4*02	16.8%	15	IgG1	CH103VL	3-1*01	1*01	10.6%	10
CH104	CH104VH	4-59*01	3-16*01	4*02	14.9%	15	IgG1	CH104VL	3-1*01	1*01	10.6%	10
CH105	CH105VH	4-59*01	3-16*01	4*02	12.7%	15	IgG1	I2VL	3-1*01	1*01	10.0%	10

[1]The HCDR3 and LCDR3 lengths of the CH103 lineage are similar to the median of HCDR3 and LCDR3 lengths of unrelated antibodies in pyrosequencing database or Genbank. Using the same 454 pyrosequencing dataset derived from three HIV infected subjects unrelated to the CH505 patient as the source of comparison, we find that the CH103 CDRH3 length of 45 nucleotides (15 aa) is the median value. The interquartile range is 39-54 nucleotides (13-18 aa). 9% of all heavy chains in this database have HCDR3 length = 45 nucleotides, this is the second most-frequent length, after 42 nucleotides. We used human L\lambda rearrangements from Genbank to compare the light chain. The CH103 light chain CDR3 is 30 nucleotides (10 aa) long. The median among Genbank hu man lambda chains is 33 nucleotides (11 aa). 24% of all human lambda chains have HCDR3 length = 30 nt, second-most frequent after 33nt.

Supplementary Table 3a. Comparison of neutralization activity of CH103, and other CD4bs mAbs against 25 clade A Env-pseudoviruses.

		IC50[a]				IC80[a]			
Virus ID	**Clade**	**VRC01**	**CH31**	**b12**	**CH103**	**VRC01**	**CH31**	**b12**	**CH103**
0260.v5.c36	A	0.53		>50	>50	1.48		>50	>50
0330.v4.c3	A	0.06		>50	7.84	0.23		>50	31.30
0439.v5.c1	A	0.05	0.06	>50	8.37	0.24	0.26	>50	>50
3365.v2.c20	A		0.01	18.90	4.00		0.03	>50	40.00
3415.v1.c1	A	0.09	0.06	6.93	2.46	0.26	0.18	13.70	4.73
3718.v3.c11	A	0.22	0.13	12.10	29.60	4.99	0.41	43.80	>50
398-F1_F6_20	A	0.06		0.06	>50	0.32		1.30	>50
BB201.B42	A	0.34	0.04	0.52	9.19	1.11	0.13	3.77	40.50
BB539.2B13	A	0.09	>50	1.42	22.20	0.33	>50	7.40	>50
BI369.9A	A	0.15	0.01	14.90	>50	0.66	0.04	>50	>50
BS208.B1	A	0.03	0.02	0.04	>50	0.10	0.05	0.33	>50
KER2008.12	A	0.56	0.11	>50	>50	1.74	0.41	>50	>50
KER2018.11	A	0.07	0.39	49.20	40.00	0.40	2.32	>50	>50
KNH1209.18	A	0.09	0.02	0.79	>50	0.30	0.10	3.39	>50
MB201.A1	A	0.24	0.02	>50	17.70	0.48	0.07	>50	>50
MB539.2B7	A	0.54	0.16	14.00	3.70	1.48	0.45	>50	14.20
MI369.A5	A	0.16	0.03	1.58	>50	0.77	0.08	36.40	>50
MS208.A1	A	0.15	0.07	0.13	>50	0.67	0.21	0.98	>50
Q23.17	A	0.09		>50	10.40	0.25		>50	12.90
Q259.17	A	0.05	5.09	>50	10.60	0.25	>50	>50	33.90
Q769.d22	A	0.02	0.03	>50	1.00	0.07	0.16	>50	2.67
Q842.d12	A	0.01	0.003	>50	1.46	0.02	0.01	>50	5.10
QH209.14M.A2	A	0.02	0.01	>50	1.16	0.08	0.07	>50	5.75
RW020.2	A	0.30	0.004	10.70	23.60	0.87	0.03	21.60	47.40
UG037.8	A	0.04	0.02	>50	1.76	0.13	0.09	>50	11.90
Breadth	**N=25 Titer <50**	**96**	**80**	**56**	**68**	**96**	**76**	**40**	**48**
Geometric Mean[b]		**0.095**	**0.038**	**2.180**	**6.569**	**0.373**	**0.12**	**5.34**	**0.39**

[a]Values <1μg/ml are indicated in red and values 1-50 μg/ml are in green.
[b]Geometric means were calculated for neutralization sensitive viruses with an IC_{50} or IC_{80} value <50μg/ml.
*Results of 118 isolates summarized in Tables 3a, b and c are representatives of total of 196 isolates tested.

Supplementary Table 3b. Comparison of neutralization activity of CH103, and other CD4bs mAbs against 39 clade B Env-pseudoviruses.

		IC50[a]				IC80[a]			
Virus ID	**Clade**	**VRC01**	**CH31**	**b12**	**CH103**	**VRC01**	**CH31**	**b12**	**CH103**
3988.25	B	2.10		2.21	9.84	>50		12.40	11.68
5768.04	B	0.10	0.10	5.62	7.08	0.94	2.37	>50	7.03
6101.10	B	0.10		>50	1.80	0.33		>50	2.25
6535.3	B	2.16		0.87	4.67	6.27		18.50	3.90
7165.18	B	>50		>50	>50	>50		>50	>50
45_01dG5	B				0.80				1.10
89.6.DG	B	0.46	0.06	0.08	1.40	1.58	0.33	0.49	1.56
AC10.29	B	1.43		1.86	>50	3.83		16.30	>50
ADA.DG	B	0.42	0.21	0.13	1.49	1.40	1.07	0.73	1.82
Bal.01	B	0.10		0.09	0.68	0.32		0.45	2.01
BG1168.01	B	0.45	0.65	>50	21.70	1.43	2.50	>50	>50
BL01.DG	B	>50	>50	>50	>50	>50	>50	>50	>50
BR07.DG	B	1.67	2.01	0.19	9.67	4.67	8.06	1.30	>50
BX08.16	B	0.28		0.92	1.15	0.68		8.16	1.07
CAAN.A2	B	1.06	>50	>50	>50	2.63	>50	>50	>50
CNE10	B	0.78		>50	26.90	1.87		>50	>50
CNE12	B	0.79		>50	26.40	2.19		>50	>50
CNE14	B	0.39		8.28	0.52	0.98		24.50	0.60
CNE4	B	0.87		>50	43.50	2.36		>50	>50
CNE57	B	0.54		16.60	>50	1.34		>50	>50
HO86.8	B	>50	>50	>50	>50	>50	>50	>50	>50
HT593.1	B	0.44	0.17	0.25	40.00	1.72	0.76	4.51	>50
HXB2.DG	B	0.04		0.00	0.12	0.08		0.01	0.23
JRCSF.JB	B	0.23		0.22	0.91	0.80		1.56	1.96
JRFL.JB	B	0.03		0.02	0.02	0.12		0.10	0.05
MN.3	B	0.03		<0.0006	0.19	0.07		0.002	0.54
PVO.04	B	0.39		>50	>50	0.99		>50	>50
QH0515.01	B	0.52	0.03	0.80	10.20	3.66	0.38	12.50	4.93
QH0692.42	B	1.16		0.79	25.20	3.00		3.05	>50
REJO.67	B	0.05		1.70	2.15	1.47		>50	5.38
RHPA.7	B	0.05		0.13	8.53	0.21		0.58	9.20
SC422.8	B	0.13		0.20	2.44	0.16		2.19	3.08
SF162.LS	B	0.24		0.03	0.84	0.33		0.15	1.19
SS1196.01	B	0.28		0.78	1.34	0.63		5.51	1.08
THRO.18	B	4.42		0.96	>50	0.65		8.76	>50
TRJO.58	B	0.08		>50	>50	15.10		>50	>50
TRO.11	B	0.34		>50	5.34	0.24		>50	5.15
WITO.33	B	0.11		9.67	7.35	1.09		>50	8.53
YU2.DG	B	0.06	0.072	1.46	1.80	0.27	0.23	6.34	2.00
Breadth	**N=39 Titer <50**	**90**	**21**	**64**	**77**	**87**	**21**	**56**	**59**
Geometric Mean[b]		**0.30**	**0.17**	**0.46**	**2.84**	**0.92**	**1.01**	**1.47**	**1.91**

[a]Values <1μg/ml are indicated in red and values 1-50 μg/ml are in green.
[b]Geometric means were calculated for neutralization sensitive viruses with an IC_{50} or IC_{80} value <50μg/ml.

Supplementary Table 3c. Comparison of neutralization activity of CH103, and other CD4bs mAbs against 54 clade C Env-pseudoviruses.

		IC50[a]				IC80[a]			
Virus ID	**Clade**	**VRC01**	**CH31**	**b12**	**CH103**	**VRC01**	**CH31**	**b12**	**CH103**
286.36	C	0.10		0.90	8.36	1.36		3.18	18.80
288.38	C	1.52		>50	8.45	0.38		>50	19.10
0013095-2.11	C	0.14		>50	5.71	5.86		>50	25.00
001428-2.42	C	0.01		>50	1.51	0.32		>50	4.64
0077_V1.C16	C	1.04		0.16	>50	0.04		2.42	>50
00836-2.5	C	0.13		>50	28.60	3.65		>50	>50
0921.V2.C14	C			3.30	>50	0.52		12.70	>50
16055-2.3	C	0.11		>50	>50			>50	>50
16845-2.22	C	2.41	4.55	>50	>50	0.37	27.80	>50	>50
16936-2.21	C	0.11		>50	5.57	9.07		>50	34.10
25710-2.43	C	0.55		>50	>50	0.47		>50	>50
25711-2.4	C	0.71		23.40	17.20	1.56		>50	46.00
25925-2.22	C	0.56		>50	26.70	1.70		>50	>50
26191-2.48	C	0.20		1.44	>50	1.39		7.25	>50
3168.V4.C10	C	0.13		>50	10.30	0.65		>50	>50
3637.V5.C3	C	4.09		>50	>50	0.28		>50	>50
3873.V1.C24	C	0.95		>50	>50	11.00		>50	>50
6322.V4.C1	C	>50		>50	>50	0.33		>50	>50
6471.V1.C16	C	>50		>50	>50	>50		>50	>50
6631.V3.C10	C	>50		>50	>50	>50		>50	>50
6644.V2.C33	C	0.16		0.03	0.38	>50		0.36	1.51
6785.V5.C14	C	0.33		16.50	8.95	0.53		>50	27.60
6838.V1.C35	C			>50	29.00	0.87		>50	>50
96ZM651.02	C	0.53		>50	8.99			>50	>50
BR025.9	C	0.27	1.13	>50	>50	1.94	49.00	>50	>50
CAP210.E8	C	>50		>50	>50	1.08		>50	>50
CAP244.D3	C	0.86		>50	>50	>50		>50	>50
CAP45.G3	C	9.47		0.11	>50	2.36		3.65	>50
CNE30	C	0.93		3.81	>50	>50		18.10	>50
CNE31	C	0.96		18.40	>50	2.59		>50	>50
CNE53	C	0.11		>50	6.92	2.24		>50	>50
CNE58	C	0.12		>50	0.54	0.28		>50	1.46
DU123.06	C	13.60		0.40	>50	0.31		5.02	>50
DU151.02	C	7.70		16.00	>50	>50		>50	>50
DU156.12	C	0.08		1.41	>50	>50		4.90	>50
DU172.17	C	>50		0.55	>50	0.24		2.95	>50
DU422.01	C	>50		0.40	>50	>50		2.17	>50
MW965.26	C	0.04		0.002	0.01	>50		2.34	0.09
SO18.18	C	0.07	0.07	14.30	1.88	0.12	0.32	47.20	7.48
TV1.29	C	>50	>50	>50	>50	0.17	>50	>50	>50
TZA125.17	C	>50	>50	>50	>50	>50	>50	>50	>50
TZBD.02	C	0.07	0.09	>50	5.55	>50	0.34	>50	23.70
ZA012.29	C	0.25	0.02	>50	28.00	0.23	0.10	>50	>50
ZM106.9	C	0.25		>50	4.82	0.83		>50	8.11
ZM109.4	C	0.13		>50	21.20	0.64		>50	>50
ZM135.10a	C	1.28		>50	>50	0.39		>50	>50
ZM176.66	C	0.04	0.01	>50	0.23	6.16	2.27	>50	1.25
ZM197.7	C	0.62		>50	35.40	1.62	0.06	>50	>50
ZM214.15	C	0.88		4.78	41.40	0.25		>50	>50
ZM215.8	C	0.28		>50	2.57	1.55		>50	>50
ZM233.6	C	4.25		>50	7.35	3.04		>50	>50
ZM249.1	C	0.08		2.73	7.23	0.83		14.80	37.60
ZM53.12	C	0.84	4.76	>50	>50	23.10		>50	>50
ZM55.28a	C	0.14		>50	4.59	0.24	17.00	>50	14.30
Breadth	**N=54 Titer <50**	**81**	**13**	**35**	**52**	**76**	**15**	**26**	**30**
Geometric Mean[b]		**0.366**	**0.219**	**1.224**	**5.23**	**0.870**	**1.574**	**4.881**	**8.11**

[a]Values <1μg/ml are indicated in red and values 1-50 μg/ml are in green.

[b]Geometric means were calculated for neutralization sensitive viruses with an IC_{50} or IC_{80} value <50μg/ml.

Supplementary Table 4. Binding of antibodies in CH103 clonal lineage to heterologous HIV-1 Env proteins.

	Binding to heterologous HIV-1 Env, EC_{50} (ug/ml)		
Antibody[a]	AE.427299 gp120	B.9021 gp140	C.1086 gp140
UCA	NB[b]	NB	NB
I8	NB	NB	NB
I7	NB	NB	NB
1A102RI6	NB	NB	NB
I4	NB	NB	36.2
1AZCETI5	NB	>10	4.5
I3	NB	0.086	0.11
I2	NB	0.03	0.06
I1	NB	0.066	0.12
1AH92U	NB	3.2	0.16
CH104	NB	0.063	0.06
CH103	NB	0.5	0.07
CH106	NB	0.06	0.22
CH105	NB	0.09	0.11

[a]The amino acid sequences and alignments of V_H and V_L gene segments of CH103 clonal lineage antibodies are shown in Supplementary Figure 7; DNA sequences and alignments of V_H and V_L gene segments of CH103 clonal lineage antibodies are shown below. [b]NB = No dateable binding.

V_H sequences and alignment:

```
UCA_VH          CAGGTGCAGC TGCAGGAGTC GGGCCCAGGA CTGGTGAAGC CTTCGGAGAC CCTGTCCCTC
I8_VH           ---------- ---------- ---------- ---------- ---------- ----------
I7_VH           ---------- ---------- ---------- ---------- ---------- ----------
I4_VH           ---------- ---------- ------G--- ---------T ---------- ----------
I3_VH           ---------- ---------- ------G--- ---------T ---------- ----------
I2_VH           ---------- -A-------- ------G--- G---------T ---------- ----------
I1_VH           ---------- -A-------- ------G--- G---------T ---------- ----------
1AZCETI5_VH     ---------- ---------- ---------- ---------- ---------- ----------
1A102RI6_VH     ---------- ---------- ---------- ---------- ---------- ----------
1AH92U_VH       ---------- ---------- ------G--- ---------T ---------- ----------
CH106_VH        ---------- -A-------- ------G--- G---------T ---------- ----------
CH103_VH        ---------- -A-------- ------G--- G---------T ---------- ----------
CH104_VH        ---------- -A-------- ------G--- G---------T ---------- ----------
CH105_VH        ---------- ---G------ ------G--- -----C---T ---------- ---------- 60

UCA_VH          ACCTGCACTG TCTCTGGTGG CTCCATCAGT AGTTACTACT GGAGCTGGAT CCGGCAGCCC
I8_VH           ---------- ---------- ------G--- -----T---- ---------- -------T--
I7_VH           -------G-- ---------- ------G-C- -----T---- ----T----- -------A--
I4_VH           ---------- ---------- ------GG-- G----T---- ---------- -------T--
I3_VH           ---------- ---------- ------GG-- G-G--T--T- --------C- -------T--
I2_VH           ---------- ---------- ------GG-- G-GACT--T- ----T---C- G--C-T-T--
I1_VH           ---------- ---------- ------GG-- G-GACT--T- ----T---C- G--C-T-T--
1AZCETI5_VH     -------G-- -------G-- ------G-C- GAG-CT---- ---AT----- G----T-A--
1A102RI6_VH     -------G-- ---------- ------G-C- -----T---- --TTT----- -------A--
1AH92U_VH       ---------- ---------- ------GG-A G----T---- ---------- -------T--
CH106_VH        ------T--- ---------- ------GG-- G-GACT--T- ----T---C- G--C-T-T--
CH103_VH        --T------- ---------- ------GG-- G-GACT--T- ----T---C- G--C-T-T--
CH104_VH        ---------- ---------- ------GG-- G-GACT--T- ----T---C- G--C-T-T--
```

```
CH105_VH      ------G--- ---------- ------GG-- G-G--T--T- --------C- ------TT-- 120

UCA_VH        CCAGGGAAGG GACTGGAGTG GATTGGGTAT ATCTATTACA GTGGGAGCAC CAACTACAAC
I8_VH         --C------- ---------- ------C--- ---------- C---AGA--- ----------
I7_VH         --C------- ---------- ------C--- ---------- C---AGA--- ----------
I4_VH         --C-T----- ---------- ------C--- ------C--- C---AGA--- G---------
I3_VH         --C-T----- ---------- ------C--- ----T-C--- C---ACA--- G---------
I2_VH         --C------- ---------- ------C--- ----T-C--- C---ACA--- ---------T
I1_VH         --C------- ---------- ------C--- ----T-C-T- C---ACA--- --------GT
1AZCETI5_VH   --CAC-GG-- ---------- ------C--- ----T-C-TT CC--A----- ---------T
1A102RI6_VH   --C------- ---------- ------C--- ---------- C---AGAG-- -C--------
1AH92U_VH     --CCT----- ---------- ------C--- ------C--G C---AGA--- G--T------
CH106_VH      --C------- ---------- ------C--- ----T-C-T- C---ACA--- --------GT
CH103_VH      --C------- -------A-- ------C--- ----T-C-T- C---AGAG-- ---T----GT
CH104_VH      --C------- ---------- ------C--- ----T-C--- C---ACA--- ---------T
CH105_VH      --C-T----- ---------- ------T--- G---T-C--- C---ACA--- G--------- 180

UCA_VH        CCCTCCCTCA AGAGTCGAGT CACCATATCA GTAGACACGT CCAAGAACCA ATTCTCCCTG
I8_VH         ---------- ---------- ---------- --G------- --G------- G---------
I7_VH         ---------- ---------- ---------- --G---G--C --G------- G---------
I4_VH         ---------G -------C-- ----G-C--- --G------- --G------- G---------
I3_VH         ---------G -------C-- ----G-C--- --G------- --G------- G---------
I2_VH         --------GG --G------- -T-------- --G------- --G--G---- G---------
I1_VH         --------G- --G------- -T-------- --G------- --G--G---- G---------
1AZCETI5_VH   ---G------ ----C----- ------C--- --G---G--C --GG------ T---------
1A102RI6_VH   ---------- ---------- ---------- --G---G--C --GC------ G---------
1AH92U_VH     ---------G -------C-- ----G-C--- --G------- ---------- ----------
CH106_VH      --------G- --G------- -T-------- --G------- --G--G---- G---------
CH103_VH      --------G- --G------- -T-------- --G------- --G--G---- G---------
CH104_VH      --------GG --G------- -T-------- --G------- --G--G---- G---------
CH105_VH      ---------G C------C-- --TGG-C--- --G------A --G------- G--------- 240

UCA_VH        AAGCTGAGCT CTGTGACCGC TGCGGACACG GCCGTGTATT ACTGTGCGAG CCTGCCCAGG
I8_VH         --------G- ---------- ---------- ---------- ---------- ----------
I7_VH         --------G- ---------- ---------- ---------- ---------- ----------
I4_VH         --------G- ---------- ---------- ---------- ---------- ----------
I3_VH         -G------G- ---------- ---------- ---------- ---------- ----------
I2_VH         -GA-----G- ---------- ---------- -----C---- TT-----C-- T---------
I1_VH         -GA-----G- ---------- ---------- -----C---- TT-----C-- T---------
1AZCETI5_VH   CGAT-A--G- --C------- ---------- ---ACA---- T-------C- -----G----
1A102RI6_VH   --------G- ---------- ---------- -G-------- GG-------- ----------
1AH92U_VH     --------G- ---------- ---------- ---------- ---------- ----------
CH106_VH      -GA---C-G- ---------- ---------- -----T---- TT-----C-- T---------
CH103_VH      -GA-----G- ---------- ---------- -----C---- TT-----C-- T---------
CH104_VH      -GA-----G- ---------- ---------- -----C---- TT-----C-- T---------
CH105_VH      -G-A----G- ---------- ---------- ---------- ---------- ---------- 300

UCA_VH        GGGCAGTTAG TCAATGCCTA CTTTGACTAC TGGGGCCAGG GAACCCTGGT CACCGTCTCC
I8_VH         ---------- ---------- ---------- ---------- ---------- ----------
I7_VH         ---------- ---------- ---------- ---------- ---------- ----------
I4_VH         ---------- ---------T ---------- ---------- ---------- ----------
I3_VH         -----A---- ---------T -------A-- ---------- ---------- -------G--
I2_VH         --C--A---- ---------- ----CG-A-T -------GC- ---------- -T-----A--
I1_VH         --C--A---- ---------- ----CG-A-T -------GC- ---------- -T-----A--
1AZCETI5_VH   ---------- ----C----- -----GT-C- -------C-- ---T------ ----------
1A102RI6_VH   ---------A --C------- ----C----- ---------- --G------- ----------
1AH92U_VH     ---------- ---------T -------C-- ---------- ---------- ----------
CH106_VH      --C--A---- ---------- ----CG-A-T -------GC- ---T------ -T-----A--
CH103_VH      --C--A---- ---------- ----CG-A-T -------GC- --T-T----- -T-----A--
CH104_VH      --C--A---- ---------- ----CG-A-T -------GC- ---------- -T-----A--
CH105_VH      --A--A---- ---------T ---C---A-- -------C-- ---------- -------G-- 360

UCA_VH        TCA
I8_VH         ---
I7_VH         ---
I4_VH         ---
```

```
I3_VH           ---
I2_VH           G--
I1_VH           G--
1AZCETI5_VH     ---
1A102RI6_VH     ---
1AH92U_VH       ---
CH106_VH        G--
CH103_VH        G--
CH104_VH        G--
CH105_VH        ---  363
```

V_L sequences and alignment:

```
UCA_VL     TCCTATGAGC TGACTCAGCC ACCCTCAGTG TCCGTGTCCC CAGGACAGAC AGCCAGCATC
PI1_VL     ---------- ---------- ---------- ---------- ---------- -----C---A
I2_VL      ---------- ---------- ---------- ---------- ---------- -----C---A
CH103_VL   ---------- ---------- ---------- ---------- ---------- -----C---A
CH106_VL   ---------- ---------- ---------- ---------- ---------- -----C---A
CH104_VL   ---------- ---------- ---------- ---------- ---------- -----C---A 60

UCA_VL     ACCTGCTCTG GAGATAAATT GGGGGATAAA TATGCTTGCT GGTATCAGCA GAAGCCAGGC
PI1_VL     --G------- -G------G- --CAAG---- A---T----- --------GT ------G---
I2_VL      --G------- -G------G- --CAAG---- A---T----- --------GT ------G---
CH103_VL   --G------- -G-CA..... ....AG--C- A---T----- --------GT ------G---
CH106_VL   --G------- -G------G- --CAAG-GT- A---T----- --------GT ------G---
CH104_VL   --G------- -G------G- --CAAG---- A---T---T- --------GT ------G--- 120

UCA_VL     CAGTCCCCTG TGCTGGTCAT CTATCAAGAT AGCAAGCGGC CCTCAGGGAT CCCTGAGCGA
PI1_VL     ---------- A-G------- ----G--A-- TAT------- ---------- ------C--G
I2_VL      ---------- A-G------- G---G--A-- TAT------- ---------- ------C--G
CH103_VL   ---------- A-G------- --T-G-GA-- TAT------- ---------- ------C--G
CH106_VL   ---------- ACG------- G---G--A-- TAT------- ---------- ------C--G
CH104_VL   ---------- A-G------- G---G--A-- TAT------G ---------- ------C—G 180

UCA_VL     TTCTCTGGCT CCAACTCTGG GAACACAGCC ACTCTGACCA TCAGCGGGAC CCAGGCTATG
PI1_VL     ---------- ----G----- --G------- ---------- --C------- ---------A
I2_VL      ---------- ----G----- --G------- ---------- --C------- --------CA
CH103_VL   ---------- ----G----- --G------- ---------- --C------- ---------A
CH106_VL   ---------- ----G----- --G------- ---------- --C------- --------CA
CH104_VL   ---------- ----G----- --G------- ---------- --C------- --------CA 240

UCA_VL     GATGAGGCTG ACTATTACTG TCAGGCGTGG GACAGCTTCT CCACCTTCGT CTTCGGAACT
PI1_VL     ---------- ---------- -----T---- ---------- ---------- -------T--
I2_VL      ---------- ---------- -----T---- ---------- ---------- -------T--
CH103_VL   ---------- -T-------- -----T---- ---------- ---------- -------T--
CH106_VL   ---------- -G-------- -----T---- ---------- ---------- -------T--
CH104_VL   ---------- ---------- -----T---- ---------- ---------- -------T-- 300

UCA_VL     GGGACCAAGG TCACCGTCCT C
PI1_VL     ------C--- ---------- -
I2_VL      ------C--- ---------- -
CH103_VL   ------C--- ---------- -
CH106_VL   ------C--- ---------- -
CH104_VL   ------C--- ---------- - 321
```

Supplementary Table 5. Affinity and kinetics of CH103 UCAs binding to autologous T/F CH505 gp140 [a].

	Binding affinity to autologous Envs		
CH103UCAs	k_a (x 10^3 $M^{-1}s^{-1}$)	k_d, (x 10^{-3} s^{-1})	K_d, nM
CH103UCA	26.7	0.926	37.5
CH103UCA-2,3,5[b]	20.5	2.9	141.5
CH103UCA-4	27.2	1.0	36.8
CH103UCA-6	25.0	6.6	264.0

[a] SPR binding rate constants and dissociation constant (K_d) was measured with each antibody captured on an anti-IgG (Fc specific) antibody surface and CH505 gp140 was injected in solution at concentrations ranging from 2 to 100 ug/mL and as described in the online Methods section. Data is representative of at least two independent measurements.
[b] Amino acid sequences encoded by V_HDJ_H of CH103UCAs-2, 3 - 5 are the same amino acid as shown in the alignment below.

DNA sequence alignment of V_HDJ_H CH103 UCAs:

```
CH103UCA      CAGGTGCAGC TGCAGGAGTC GGGCCCAGGA CTGGTGAAGC CTTCGGAGAC CCTGTCCCTC ACCTGCACTG TCTCTGGTGG CTCCATCAGT AGTTACTACT GGAGCTGGAT CCGGCAGCCC CCAGGGAAGG
CH103UCA-2    ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
CH103UCA-3    ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
CH103UCA-4    ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
CH103UCA-5    ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
CH103UCA-6    ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- 130

CH103UCA      GACTGGAGTG GATTGGGTAT ATCTATTACA GTGGGAGCAC CAACTACAAC CCCTCCCTCA AGAGTCGAGT CACCATATCA GTAGACACGT CCAAGAACCA ATTCTCCCTG AAGCTGAGCT CTGTGACCGC
CH103UCA-2    ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- G--------- ----------
CH103UCA-3    ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
CH103UCA-4    ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- G--------- ----------
CH103UCA-5    ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- G--------- ----------
CH103UCA-6    ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- G--------- ---------- 260

CH103UCA      TGCGGACACG GCCGTGTATT ACTGTGCGAG CCTGCCCAGG GGGCAGTTAG TCAATGCCTA CTTTGACTAC TGGGGCCAGG GAACCCTGGT CACCGTCTCC TCA
CH103UCA-2    ---------- ---------- ---------- ---------- ---G------ ---------- ---------- ---------- ---------- ---
CH103UCA-3    ---------- ---------- ---------- ---------- ---G------ ---------- ---------- ---------- ---------- ---
CH103UCA-4    ---------- ---------- ---------- ---------- -----A---A ---------- ---------- ---------- ---------- ---
CH103UCA-5    ---------- ---------- ---------- ---------- ---G------ ----C----- ---------- ---------- ---------- ---
CH103UCA-6    ---------- ---------- ---------- ---------- ---G------ --CG------ ---------- ---------- ---------- --- 260
```

Amino acid sequence alignment of V_HDJ_H CH103 UCAs:

```
CH103UCA      QVQLQESGPG LVKPSETLSL TCTVSGGSIS SYYWSWIRQP PGKGLEWIGY IYYSGSTNYN PSLKSRVTIS VDTSKNQFSL KLSSVTAADT AVYYCASLPR GQLVNAYFDY WGQGTLVTVS S
CH103UCA-2    ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- -E-------- ---------- -
CH103UCA-3    ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- -E-------- ---------- -
CH103UCA-4    ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---I------ ---------- -
CH103UCA-5    ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- -E-------- ---------- -
CH103UCA-6    ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- -E--R----- ---------- - 121
```

Supplementary Table 6. Reactivity of autologous Envs with antibodies in CH103 clonal lineage in ELISA.

Autologous Env	*Apparent binding affinity of autologous Envs to CH103 clonal lineage antibodies, EC50, ug/ml									
	UCA	I8	I4	I3	I2	I1	CH105	CH103	CH104	CH106
CH0505 T/F	2	1.1	0.3	0.12	0.09	0.11	0.1	0.08	0.12	0.08
CH505.w30.e16	NB	>10	2.1	0.07	0.047	0.06	0.064	0.055	0.05	0.05
CH0505.w30.e23	NB	NB	>20	0.14	0.07	0.09	0.08	0.044	0.07	0.053
CH505.w53.e16	NB	NB	NB	0.066	0.03	0.05	0.05	0.03	0.036	0.032
CH505.w78.e7	NB	NB	NB	0.13	0.054	0.083	0.09	0.043	0.1	0.13
CH505.w78.e16	NB	NB	NB	NB	0.2	>10	0.3	1.2	0.19	0.14
CH505.w78.e38	NB	NB	NB	>100	>100	>10	>10	>10	>10	>10

*Env proteins highlighted in green had 2-fold or greater loss of binding affinity to antibodies in CH103 clonal lineage compared with the binding of transmitted/founder (T/F) Env to the same antibodies. NB = No detectable binding.

Supplementary Table 7. V_HDJ_H sequences 2 genes (IZ95W and 02IV4) very similar to the CH103 VDJ genes, possible clonal members, identified by 454 sequencing and alignment with their UCA. V_HDJ_H genes of IZ95W and 02IV4 were produced as recombinant antibodies complemented with V_LJ_L genes of UCA and tested for binding to the autologous CH505 T/F Env and heterologous HIV-1 Envs in ELISA assays. MAb IZ95W bound CH505 T/F gp140 with end point titer of 11.1 ug/ml, but did not BIND with heterologous Envs, 6321, 9021, 1086C and 427299.

```
>UCA_VHDJH
QVQLQESGPGLVKPSETLSLTCTVSGGSISSYYWSWIRQPPGKGLEWIGYIYYSGSTNYNPSLKSRVTISVDTSKNQFSLKLSSVTAADTAVYYCASLPRGQ
LVNAYFDYWGQGTLVTVSS
>IZ95W_VHDJH
QVQLQESGPGLVKPSETLSLTCTVSGGSIVSYYWSWIRQPPGKGLEWIGYMYYSGSTNYNPSLKSRVTISIDTSKNQFSLKLRSVTAADTAVYYCASLPRGQ
LILGYFDYWGQGTLVTVSS
>02IV4_VHDJH
QVQLQESGSGLVKPSETLSLTCTVSGGSISSYYWSWIRQPPGKGLEWIGFIYYSGSTNYNPSLKSRVTISIDTSKNQFSLKLRSVTAADTAVYYCASLPRGQ
LILGYFDYWGQGTLVTVSS

UCA_VHDJH        QVQLQESGPG LVKPSETLSL TCTVSGGSIS SYYWSWIRQP PGKGLEWIGY IYYSGSTNYN
IZ95W_VHDJH      ---------- ---------- ---------V ---------- ---------- M---------
02IV4_VHDJH      --------S- ---------- ---------- ---------- ---------F ---------- 60

CUCA_VHDJH       PSLKSRVTIS VDTSKNQFSL KLSSVTAADT AVYYCASLPR GQLVNAYFDY WGQGTLVTVS
IZ95W_VH         ---------- I--------- --R------- ---------- ---ILG---- ----------
02IV4_VH         ---------- I--------- --R------- ---------- ---ILG---- ---------- 120

CUCA_VHDJH       S
IZ95W_VHDJH      -
02IV4_VHDJH      - 121
```

Supplementary Table 8. Crystallographic data collection and refinement statistics.

	CH103:gp120	Fab CH103
PDB accession code	4JAN	4JAM
Data collection		
Space group	$P2_1$	$P2_1$
Cell constants		
a, *b*, *c* (Å)	48.9, 208.7, 69.4	43.0, 146.4, 66.3
α, β, γ (°)	90.0, 107.2, 90.0	90.0, 97.7, 90.0
Wavelength (Å)	1.00	1.00
Resolution (Å)	46.8-3.25 (3.31-3.25)*	38.2-1.65(1.71-1.65)
R_{merge}	13.3 (36.0)	6.7(53.1)
$I / \sigma I$	9.4 (2.3)	17.1 (2.6)
Completeness (%)	93.7 (59.0)	98.0 (90.9)
Redundancy	3.5 (2.5)	3.4 (2.3)
Refinement		
Resolution (Å)	3.25	1.65
No. of unique reflections	19,564 (619)	95,531 (8,787)
R_{work} / R_{free} (%)	19.6/25.6	16.9/19.7
No. atoms	17823	13687
Protein	8843	6437
Ligand/ion	154	134
Water	22	726
B-factors (Å^2)	84.0	26.3
Protein	84.0	24.7
Solvent	44.2	36.9
R.m.s. deviations		
Bond lengths (Å)	0.002	0.007
Bond angles (°)	0.610	1.160
Ramachandran		
Most favored regions (%)	92.0	97.4
Additional allowed regions (%)	7.3	2.6
Disallowed regions (%)	0.7	0.0

* Values in parentheses are for highest-resolution shell.

Supplementary Table 9. Comparison of interactions between HIV-1 gp120 and CD4, CH103 and other CD4-binding site antibodies.

HIV-1 gp120			Residue-by-residue binding surface on HIV-1 gp120 (Å^2) * HIV-1 gp120 interacting molecule (PDB code)								
Region	Residue Number	Residue Type	CD4 (2NXY)	CH103 (4JAN)	b12 (2NY7)	b13 (3IDX)	F105 (3HI1)	VRC01 (3NGB)	VRC03 (3SE8)	VRC-PG04 (3SE9)	NIH45-46 (3U7Y)
	49	D									4.8
	96	W									7.1
	97	K						26.4	41.0	10.5	44.0
	99	D									22.1
	102	E									33.8
α1	105	H					2.9				
	108	I					2.8				
	109	I					28.7				
	112	W					69.9				
	122	L								3.1	3.3
	123	T							8.5		5.6
	124	P	39.9					36.8	81.7	31.2	33.8
	125	L	6.7								
V1/V2	126	C	61.5								
	127	V	30.8								
	196	C	5.5								
	198	T						12.7	51.3		
	199	S							13.4		
	210	F					5.3				
β4/5	226	L					11.8				
	244	T					5.7				
	255	V					18.1				
Loop B	256	S		56.6							
	257	T			15.0		10.9		6.9		
	275	V								26.2	11.1
	276	N						22.6	13.3	17.7	9.9
	278	T						84.4	73.0	35.2	36.7
Loop D	279	D	18.5	3.9			6.6	56.5	36.5	51.7	57.5
	280	N	51.1	94.9	13.0		14.3	70.2	76.3	72.0	69.7
	281	A	74.7		29.8	16.7	52.2	69.6	70.8	74.0	75.7
	282	K	31.6					30.5	17.2	31.9	49.7
	283	T	18.2			12.6		11.9	8.2	10.6	5.2
	364	S		18.9							
	365	S	65.6	85.9	38.3	11.5	25.5	61.5	58.0	46.0	59.1
	366	G	24.6	16.7	44.5	29.5	53.8	22.0	23.3	22.0	21.1
	367	G	38.6	71.1	74.4	61.7	37.4	24.1	26.4	22.9	26.2
β15/ α3	368	D	69.5	64.0	64.6	87.0	87.3	48.6	54.2	51.2	47.4
CD4-binging	369	P		19.4	59.7	62.9	16.6				
loop	370	E	14.7	30.1	14.7	26.4	53.7		16.0		
	371	I	39.6	39.8	80.8	61.8	47.5	44.1	35.2	44.8	60.4
	372	V			30.9	25.8					
	373	T			17.3	11.7					
	375	S				9.2	7.0		3.3		
	382	F					36.2				
β17	384	Y			4.6	16.2	7.2				
	386	N			35.9	2.8					
	417	P			16.3						
β18	418	C			2.2						
	419	R			84.7	64.8					

Supplementary Table 9 continued:

Residue-by-residue binding surface on HIV-1 gp120 (Å^2) *

HIV-1 gp120		HIV-1 gp120 interacting molecule (PDB code)								
Region	**Residue**		**CD4 (2NXY)**	**CH103 (4JAN)**	**b12 (2NY7)**	**b13 (3IDX)**	**F105 (3HI1)**	**VRC01 (3NGB)**	**VRC03 (3SE8)**	**VRC-PG04 (3SE9)**	**NIH45-46 (3U7Y)**
	Number	**Type**									
β20/21 Bridging sheet	421	K				40.2	54.1				
	424	I					4.3				
	425	N	24.5			60.7	18.2		12.1		
	426	M	14.4			12.5	79.1		7.1		
	427	W	28.2					7.6	35.1		12.8
	428	Q						4.7			
	429	K	14.9				63.0	2.1	49.7	47.6	23.7
	430	V	111.5		31.1			58.3			57.3
	431	G			13.2				13.8	7.0	11.3
	432	K			47.8						8.0
β 23	455	T	15.5	18.6	28.3	3.5	9.6	31.2	24.0	32.9	31.9
	456	R	3.6		5.4			5.8	6.4	2.4	6.8
	457	D	37.4	49.1				46.4	43.7	27.1	45.8
Loop V5	458	G		32.5				50.5	35.2	39.4	44.9
	459	G	32.2	48.9				69.1	62.9	56.8	68.8
	460	N	64.8	53.3				37.1	63.9		24.1
	461	S		66.1				67.8	51.2	66.3	54.9
	462	N		36.2					28.9	26.7	6.6
	463	N		16.9				13.7	15.4		10.2
β24	465	S						9.7	9.4		8.7
	466	G						6.0			2.1
	467	I						11.3	15.9		17.8
	469	R	13.5	59.4				23.3	21.8	17.2	21.0
Outer domain exiting loop	471	G		4.2			4.2				
	472	G	20.5	6.6	20.5	6.3		8.4	23.7	3.6	4.8
	473	G	23.4		53.5	23.8	50.2	27.6	29.2	18.3	22.8
	474	D	37.2		26.0	25.3	43.9	17.3	3.0	7.6	30.5
	475	M	2.8		65.0		33.6				4.2
α5	476	R	9.3				40.9			21.9	24.8
	477	D	3.7				3.1				
	480	R									16.4

*: Residues with interacting surface area less than 2.0 Å^2 are not listed. Accession codes for CH103 have been requested.

Supplementary Table 10. Interface between antibody CH103 and ZM176.66 gp120.

Supplementary Table 10a, Total buried surface areas across the interface of CH103 and HIV-1 gp120.

Antibody	Interface on gp120 (Å²)	Interface on CH103 (Å²)	
		Total antibody area	Area contributed by UCA residues (% total)
Heavy chain	429	493	414 (84%)
Light chain	378	377	164 (44%)
Total	807	870	578 (66%)

Supplementary Table 10b, Residue-by-residue buried surface area of gp120 residues that interact with CH103.

gp120 residue			Heavy chain interactions		Light chain interactions	
Region	Number	Type	Bond type*	Surface area (Å^2)	Bond type	Surface area (Å^2)
Loop B	256	SER	H	12		
Loop D	279	ASP				4
	280	ASN		12	H	40
CD4-binding loop	364	HIS		13		4
	365	SER	H	54		32
	366	GLY		18		4
	367	GLY	H	68		
	368	ASP	HS	65		
	369	LEU	H	19		
	370	GLU	HS	29		
	371	ILE		39		
Loop V5	455	THR		19		
	457	ASP		35		13
	458	GLY		10	H	24
	459	GLY				50
	460	ASN		51		51
	461	ASP			HS	70
	462	ASP				37
	463	ASN				15
β24	469	ARG		24		34
	471	GLY		5		
	472	GLY		6		

* Bond type: H: Hydrogen, S: Salt bridge.
Detailed gp120:CH103 interface data was calculated on the EBI PISA server (http://www.ebi.ac.uk/msdsrv/prot_int/cgi-bin/piserver).

Supplementary Table 10c. Residue-by-residue buried surface areas of the CH103 paratope residues.

Chain	Region	Residue		Bond type*	Buried surface area (Å^2)	Contribution by Region (%)
		Number (Kabat)	Type			
H	CDR H1	33	TYR	H	32	3.7
H	CDR H2	50	TYR	H	22	11.8
H		52	PHE		12	
H		54	THR		8	
H		56	GLU	H	59	
H		58	ASN		2	
H	CDR H3	97	ARG	HS	82	41.1
H		98	GLY		23	
H		99	GLN	H	120	
H		100	LEU	H	27	
H		100A	VAL	H	55	
H		100B	ASN		50	
L	CDR L1	27	SER		4	8.0
L		31	THR		14	
L		32	ASN	H	52	
L	CDR L2	50	GLU	H	38	14.0
L		51	ASN	H	28	
L		52	TYR		33	
L		53	LYS	H	22	
L	FWR L3	65	SER		9	13.1
L		66	LYS	HS	44	
L		67	SER		25	
L		68	GLY		36	
L	CDR L3	91	TRP		71	8.1

* Bond type: Hydrogen, D: Disulphide bond, S: Salt bridge, C: Covalent link.
Detailed gp120:CH103 interface data was calculated on the EBI PISA server (http://www.ebi.ac.uk/msdsrv/prot_int/cgi-bin/piserver).

Supplementary Table 11. Hydrogen bonds and salt bridges between CH103 and ZM176.66 gp120.

Hydrogen bonds							
Antibody CH103				Distance (Å)	**gp120**		
Chain	Number	Type	Atom		Atom	Type	Number
H	33	TYR	OH	2.2	OD1	ASP	368
H	50	TYR	OH	3.7	O	GLY	367
H	56	GLU	OE1	3.2	N	LEU	369
H	97	ARG	NH1	3.8	OG	SER	256
H	97	ARG	NH2	2.6	OE1	GLU	370
H	99	GLN	NE2	3.1	OG	SER	365
H	100	LEU	N	2.4	O	SER	365
H	100	LEU	N	3.2	OG	SER	365
H	100A	VAL	N	3.2	OG	SER	365
L	32	ASN	ND2	3.0	O	GLY	458
L	50	GLU	OE2	3.1	ND2	ASN	280
L	51	ASN	ND2	2.9	OD1	ASP	461
L	53	LYS	NZ	3.2	OD1	ASN	280
L	66	LYS	NZ	3.2	OD2	ASP	461

Salt bridges							
Antibody CH103				Distance (Å)	**gp120**		
Chain	Number	Type	Atom		Atom	Type	Number
H	97	ARG	NE	3.9	OD1	ASP	368
H	97	ARG	NE	2.6	OD2	ASP	368
H	97	ARG	NH2	2.6	OE1	GLU	370
H	97	ARG	NH2	2.7	OD2	ASP	368
L	66	LYS	NZ	3.7	OD1	ASP	461
L	66	LYS	NZ	3.2	OD2	ASP	461

Supplementary Table 12. Residue-by-residue specification of unmutated versus mutated residues on antibody CH103.

Heavy chain mutations

Region	ID	UCA	Mature	Paratope	Description and note
FWR 1	11	L	V		Shortens side chain on strand A
	14	P	S		Alters loop between strands A and B
	29	I	M		Enhances interactions with heavy chain Trp34
	30	S	G		Increases loop flexibility
CDR 1	31	S	G		Increases CDR 1 flexibility
	32	Y	T		Avoids clash with other heavy chain residues
FWR 1	37	I	L		Neutral mutation
	39	Q	L		Alters heavy/light chain interface
	40	P	S		Allows flexibility in strand C
CDR 2	52	Y	F	Yes	Polar to hydrophobic
	53	Y	H	Yes	Polar to basic
	54	S	T	Yes	Adds carbon to paratope interface
	56	**S**	**E**	**Yes**	**Forms hydrogen bond with backbone amide of Leu369 in the CD4-binding loop**
	60	N	S		Alters heavy/light chain interface
	65	S	G		Increases flexibility in loop between strand C" and D
FWR 3	68	T	S		Avoids clashes with neighboring residues
	75	K	E		Basic to acidic change in loop between strand D and E
	76	N	D		Polar to acidic change at the beginning of strand E
	81	K	R		Neutral mutation
	82A	S	R		Polar to basic change at C terminus of strand E
	91	Y	F		Polar to hydrophobic change at the heavy/light chain interface
CDR 3	101	D	R		Acidic to basic change at the first layer adjacent to paratope
	102	Y	N		Smaller side chain at the end of CDR H3
FWR 4	105	Q	R		Alters heavy/light chain interface
	107	T	S		Avoids clashes with neighboring residues
	110	T	S		Alters heavy/light chain interface
	112	S	T		Minor change at the end of strand G
	113	S	A		Avoids clashes with neighboring residues

(Supplementary Table12 continues to next page)

Supplementary Table12, continued:

Light chain mutations

Region	ID	UCA	Mature	Paratope	Description and note
FWR 1	20	S	T		Surface residue in strand B
CDR 1	26	D	A		Avoids clashes with neighboring residues
	27	K	.		Deletion reduces potential clashes with HIV-1 gp120
	27A	L	.		Deletion reduces potential clashes with HIV-1 gp120
	27B	G	.		Deletion reduces potential clashes with HIV-1 gp120
	27C	D	S	Yes	Acidic to polar change
	31	K	T	Yes	Shorter side chain reduces potential clashes with HIV-1 gp120
	32	Y	N	Yes	Smaller side chain reduces potential clashes with HIV-1 gp120
	33	A	V		Bulker side chain increases packing of light chain core
FWR 2	38	Q	V		Alters heavy/light chain interface
	45	V	E		Surface residue at strand C', hydrophobic to acidic change
	46	L	V		Alters heavy/light chain interface
	49	Y	F		Alters heavy/light chain interface
CDR 2	**50**	**Q**	**E**	**Yes**	**Forms hydrogen bond with Asn280 in gp120 loop D**
	51	**D**	**N**	**Yes**	**Forms hydrogen bond with Asn461 in gp120 loop V5**
	52	S	Y	Yes	Enhances interactions with gp120 loop D
	60	E	D		Surface residue in loop between strands C" and D
	66	**N**	**K**	**Yes**	**Forms hydrogen bond and salt bridges with Asn461 in loop V5**
FWR 3	69	N	S		Shortens side chain in loop between strands D and E
	76	S	R		Polar to basic change and longer side chain at C terminus of strand E
	81	M	I		Shortens side chain in loop between strands E and F
CDR 3	90	A	V		Bulker side chain increases packing of light chain core
FWR 4	100	T	S		Neutral mutation in strand G and near heavy/light chain interface

To determine the frequency of germline antibodies that could potentially serve as unmutated common ancestors of a lineage line CH103, we have interrogated a combined dataset of 454 pyrosequences of three HIV infected subjects unrelated to the CH505 patient. Gene segment frequencies in this dataset demonstrate that the frequency of the VH4-59 gene is 4.2%, the JH4 is 49.7% and the frequency of HCDR3 length of the CH103 VH length (a 15mer) is 8.9%. The proportion of sequences with all three characteristics, if independent is VH4-59/JH4/CDR3 Length=15 is 1/540 with the actual count in the analyzed data set of the combinantion = 637/386853 = 1/607.
This frequency is clearly very common. The question that remains regards the prevalence of the relevant characteristics of CDR3.
For example, the HC CDR3 contact residues (from figure 4 of the paper) are RGQLVN starting at position 4 in HCDR3 with the following conservative substitutions:

R: K; G: A; Q: E; L: I, V; V: I, V; N: D

We therefore use the HCDR3 motif: XXX(R/K)(G/A)(Q/E)(L/V/I)(L/V/I)(N/D)nX, and scanned our pyrosequencing heavy-chain dataset for its occurrence. This motif occurred 10 times among the 337567 in-frame HCDR3 in our pyrosequencing database.
If we allow positions other than the fourth (which contains the R/K necessary for the salt bridge) to vary we obtain the table below. The number of positions at which the observed HCDR3 differs from the CH103 HCDR3 motif is on

the left, and the number out of 337567 HCDR3 seqeunces is on the right. All of the CDR3 in this table have R or K at position 4.

distance	number of sequences out of 337567
0	10
1	71
2	1028

An appropriate light-chain UCA is also likely to be readily available. We downloaded 2312 rearranged human lambda V-region sequences from Genbank and analyzed them for comparison. The CH103 light chain uses IGLV3-1 and IGLJ1. These genes are found in 9.6% and 15.5% respectively of all sequences in the Genbank lambda database. The CH103 light chain is 30 nt long, as are 23.7% of the Genbank lambdas. The single contact residue in the light-chain CDR3 is tryptophan at the 3rd CDR3 position, which is encoded by the IGLV gene.Indeed 43% of all Genbank lambda chains have W at position 3 of CDR3. Thus, there is considerable evidence that the germlines of the CH103 lineage are relatively common by a variety of criteria.

Supplementary Table 13. Localization of sites under positive selection using the fixed effects likelihood (FEL)[1] (p-value < 0.10) and the mixed effects model of evolution (MEME)[2] (q-value < 0.1).

	Using FEL							Using MEME			
Week after infection	n Pos[a]	n Neg[b]	CH103 n Pos[c]	Non-CH103 n Pos[d]	*Fisher's p* (n Pos)[e]	*Fisher's p* (n Neg)[f]	Substitution per site[g]	selected sites	Positively selected sites inside CH103 footprint	Positively selected sites outside CH103 footprint	*Fisher's exact P* value
4	0	6	0	0	na	1	0.0035	0	0	0	na
14	0	14	0	0	na	1	0.0095	0	0	0	na
20	5	20	3	2	0.009	0.71	0.022	3	2/92	1/830	0.05
30	8	32	4	4	0.005	0.36	0.04	5	3/92	2/830	0.009
160	36	88	11	25	0.0004	0.6	0.057	34	11/92	23/830	0.0002

[a] Number of positively selected sites; [b] Number of negatively selected sites; [c]Number of positively selected sites among 92 sites inside CH103 binding regions (footprint); [d]Number of positively selected sites among 830 sites not in CH103 footprint; [e]P value from Fisher's exact test for positively selected sites inside vs. outside CH103 footprint; [f]P value from Fisher's exact test for negatively selected sites inside vs. outside CH103 footprint; and [g]Per-site substitution rate among 922 aligned sites.

The 922 codons in the CH505 alignment were considered as 2 sets: 92 codons (10%) were included in the candidate regions for CH103 selection (CH103, CD4, and VRC01 contact residues, as well as V1 and V5 hypervariable loops which border these contacts), and the 830 other codons remaining in the alignment. We used FEL[1] and MEME[2] methods to quantify selection in the CH505 codon-aligned sequences, implemented through the HyPhy package at the DATAMONKEY website (http://www.datamonkey.org). The full alignment was used for the initial analysis, and the codon sets defined above were used to see if positive selection was concentrated in the CH103 contact/CD4bs region. We used the strategy implemented at the DATAMONKEY web site to select optimal substitution models, with a $p < 0.10$ cutoff as evidence suggesting positive selection for the FEL model, and a $q < 0.10$ cutoff for the MEME model. Analysis by using both FEL and MEME methods showed that positive selection was enriched in CH103 binding regions by week 20, and this focus continued throughout the course of the study, through week 160. Fisher's exact test was used to test the null hypothesis that the positively selected sites are evenly distributed throughout Env; they are not, and are enriched in the CH103 region. In contrast, the number of sites under negative selection was evenly distributed between the two regions. The amino acids that are changing in the regions of interest for CH103 escape are shown in Fig 5. At week 4, using FEL [1] and MEME [2], there was no statistical evidence for positive selection anywhere in the CH505 codon-aligned sequences, though there was evidence for negative selection at 6 positions with p values below the cutoff. However, FEL and MEME will underestimate positive selection within a subject, as the frequencies of identical sequences are not considered, and thus changes in population frequency are not considered positive selection. Given this, it is of note that in the week-4 sample, a single mutation in the full alignment of 55 sequences occured more than once, and it was a N279K change in Loop D, found in 5 of the 55 sequences. There was also one instance of a short (7 residue) in-frame deletion spanning this position. This would produce just one ancestral change in the phylogenetic tree, so it could not provide statistical evidence of selection, but still coincidence of facts makes it of interest: 279 is located in a key contact position for CH103 in Loop D, in a region under clear strong subsequent selective pressure. Neighboring positions are mutating by week 14, a further indication that local positive selection might be underway, leaving open the possibility that these sites may targeted by the CH103 lineage very early in infection. Codon models also do not take into account insertions and deletions, an essential aspect of HIV env evolution, which is evident in CH505 in V1 by week 20.

1. Kosakovsky Pond, S.L. & Frost, S.D. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molecular biology and evolution* **22**, 1208-1222 (2005).
2. Murrell, B., *et al*. Detecting individual sites subject to episodic diversifying selection. *PLoS genetics* **8**, e1002764 (2012).

Supplementary Table 14. Autologous and heterologous neutralization activity of CH103 clonal lineage antibodies.

Antibody ID	Neutralization activity (IC50, ug/ml)							
		Heterologous viruses						
	C.CH505 T/F	**B.SF162**	**B.JRFL**	**A.Q168**	**A.Q842**	**B.BG1168**	**C.ZM106**	**MuLV**
UCA	> 50	> 50	> 50	> 50	> 50	> 50	> 50	> 50
I8	49.4	> 50	> 50	> 50	> 50	> 50	> 50	> 50
I7	14.7	> 50	> 50	> 50	> 50	> 50	> 50	> 50
I4	10.9	> 50	> 50	> 50	> 50	> 50	> 50	> 50
1AZCETI5	12.8	> 50	> 50	> 50	> 50	> 50	> 50	> 50
1A102RI6	4.6	> 50	> 50	> 50	> 50	> 50	> 50	> 50
1AH92U	> 50	22.0	> 50	> 50	> 50	> 50	> 50	> 50
I3	6.2	1.2	0.071	24.1	> 50	42.6	> 50	> 50
I2	1.5	1.1	< 0.023	2.6	3.7	14.7	17.0	> 50
I1	2.5	0.211	< 0.023	2.2	1.8	15.4	19.4	> 50
CH103	3.2	0.316	0.033	2.3	1.4	17.7	12.8	> 50
CH104	2.0	0.096	0.025	2.7	2.4	13.4	12.0	> 50
CH105	4.9	0.562	0.038	5.5	40.0	15.8	28.5	> 50
CH106	2.3	0.121	< 0.023	1.6	0.6	8.2	8.1	> 50

>50	>20 - <50	>10 - <20	>5 - <10	<5

Supplementary Table 15. Alignment of gp160 Env sequences of CH505 transmitted/funder (T/F) and tested heterologous HIV-1 viruses.

```
C.CH0505T/F   MRVMGIQRNY PQW.WIWSML GFWMLMICNG ....MWVTVY YGVPVWKEAK TTLFCASDAK
A.Q168        -K-R--K--L ....-K-GTM LLG---TYSV .AEQL----- -------D-E ----------
A.Q842        --A----M-C QNL.-R-GTM ILG-IIF-SA .VDNL----- ---------E ----------
B.BG1168      ---K--MK-C QHL.-R-GIM LLGI----SA .TEKL----- ---------T ----------
B.JRFL        ---K--RK-- QHL.-RGGT- LLGIIV--SA .VEKL----- ---------T ----------
B.SF162       ---K--RK-- QHL.-RGGT- LLG-----SA .VEKL----- ---------T ----------
C.ZM106       -K-RE-L--W R--.---GI- ---------V VVGNL----- ---------- ----------
B.HXB2        ---KEKYQHL WR-G-R-GTM LLG-----SA .TEKL----- ---------T ----------   60

C.CH0505T/F   AYEKEVHNVW ATHACVPTDP NPQEMVLKNV TENFNMWKND MVDQMHEDVI SLWDQSLKPC
A.Q168        --ST-K--I- ---------- ----IH-E-- --E------N --E---T-I- -------R--
A.Q842        ---T-K---- ---------- ----IH-E-- --E------N --E---T-I- ----------
B.BG1168      ---T------ ---------- ----VK-E-- --D--V---- --T-----I- ----------
B.JRFL        --DT------ ---------- ----V--G-- --K------N --E--Q--I- ----------
B.SF162       --DT------ ---------- ----I--E-- ---------N --E-----I- ----------
C.ZM106       ---R------ ---------- S------E-- ---------- --------I- ----------
B.HXB2        --DT------ ---------- ----V--V-- ---------- --E-----I- ----------   120

C.CH0505T/F   VKLTPLCVTL NCTNA..... .....T...A SNSSI...I. .......EGM KNCSFNITTE
A.Q168        ---------- ----VN.... ..NNT-...N V-NNT...G. .....WD-ER ----------
A.Q842        ---------- D-N-VT.... ..NNG-...S DM......R. ........-EI ------M---
B.BG1168      ---------- H--DVNTTCI TTNNS-MTNS ----TEGNCS SYNYNGR-EL R--------S
B.JRFL        ---------- --KDV..... ...NA-NTTN GSEGTM.... .....ERGEI ---------S
B.SF162       ---------- H---L..... ..KNA-NTKS --WKEM.... .....DRGEI -----KV--S
C.ZM106       ---------- K-V-V..... ...NA-SKSN ASATNDG... ......SGE- ---T------
B.HXB2        --------S- K--DL..... ..KND-NTNS -SGRMIM..E ......KGEI -------S-S   180

C.CH0505T/F   LRDKREKKNA LFYKLDIVQL ...DGNSSQ. .YRLINCNTS VITQACPKVS FDPIPIHYCA
A.Q168        -----Q-VYS ------V--I .....DN-S. .--------- A---------T -E--------
A.Q842        -----Q-VYS ---------I NEDQ----NN K----T---- A---------T -E--------
B.BG1168      IQ--VQ.DY- I--------PI KSDNSDNTS. .--------- --------I- -E--------
B.JRFL        I--EVQ-EY- ------V-PI ...-N-NTS. .----S-D-- --------I- -E--------
B.SF162       I-N-MQ-EY- ------V-PI ...-NDNTS. .-K------- ---------- -E--------
C.ZM106       I---KRNES- --------P- ...TNDNNSG E--------- AM-------- ----------

B.HXB2        I-G-VQ-EY- F------IPI D...NDTTS. .-K-TS---- ---------- -E--------   240

C.CH0505T/F   PAGYAILKCN NKTFTGTGPC NNVSTVQCTH GIKPVVSTQL LLNGSLAEGE IIIRSENITN
A.Q168        ---F-----K DEK-N----- K--------- ---------- --------K- VM-----F--
A.Q842        ---F-----K DEE-N-I--- K--------- ---------- --------K- VK--C-----
B.BG1168      ---F------ D-K-S-K-T- ---------- --R----L.T VV------EG VVL----F--
B.JRFL        ---F------ D---N-K--- K--------- --R------- --------E- VV---D-F--
B.SF162       ---F------ D-K-N-S--- T--------- --R------- --------EG VV-----F-D
C.ZM106       ---------- ----N----- Y--------- ---------- ---------- -------L-D
B.HXB2        ---F------ ----N----- T--------- --R------- --------E- VV---V-F-D   300

C.CH0505T/F   NVKTIIVHLN ESVKIECTRP NNKTRTSIRI ..GPGQAFYA TGQVIGDIRE AYCNINESKW
A.Q168        -A-N-L-QFK -P---N---- D-N------- ..-------- --.I-----Q ---TV-G-E-
A.Q842        -A-----Q-V NP---N---- --N--K--H- ..-------- --DI-----Q -H--V-RTE-
B.BG1168      -A-----Q-K DP-----E-- --N-IK--HL ..---R-WH- ---I-----K -F-TL-STN-
B.JRFL        -A-----Q-K ---E-N---- --N--K--H- ..---R---T --EI-----Q -H---SRA--
B.SF162       -A-----Q-K ---E-N---- --N--K--T- ..---R---- --DI-----Q -H---SGE--
C.ZM106       ---------- --IH-T---- --N--K---- ..----T--- --EI-----K -----S-E--
B.HXB2        -A-----Q-- T--E-N---- --N--KR--- QR---R--VT I-K.--NM-Q -H---SRA--   360

C.CH0505T/F   NETLQRVSKK LKEYFP.HKN ITFQPSSGGD LEITTHSFNC GGEFFYCNTS SLFNRTYMAN
A.Q168        -KA--K-VEQ -RSS-E.N-T -I-AN----- ---------- ---------- G--DS-WNDT
A.Q842        -N--HQ-VEQ -RKH-..N-T -N-AN-T--- ---------- ---------T N---S-WNHT
B.BG1168      TN--KQMVE- -R-Q-E.N-T -A-NQ-T--- P--VM-T--- ---------T Q---SIWYNT
B.JRFL        -D--KQIVI- -R-Q-E.N-T -V-NH----- P--VM----- --------ST Q---S-WNN-
B.SF162       -N--KQIVT- -QAQ-G.N-T -V-KQ----- P--VM----- --------ST Q---S-WNNT
C.ZM106       -KA--E-G-- ---H--.N-T -K-A------ ---------- R--------- K---S---H-
B.HXB2        -N--KQIAS- -R-Q-GNN-T -I-KQ----- P--V------ --------ST Q---S-WFNS   420
```

```
C.CH0505T/F   STDMANSTET NSTRTITIHC RIKQIINMWQ EVGRAMYAPP IAGNITCISN ITGLLLTRDG
A.Q168        .......DSR QENG---LP- ---------- RT-Q-I---- -Q-A-R-V-- ----I-----
A.Q842        A....SMNS- E-ND--ILP- ---------- R--Q------ -R-V-R-E-- ----I-----
B.BG1168      T.NSSWNNK- W-NN---LP- ---------- Q--K-I---- -K-K-K---- ----------
B.JRFL        T...EG-NN- EGN.---LP- ---------- ---K------ -R-Q-R-S-- ----------
B.SF162       ......IGPN -TNG---LP- -------R-- ---K------ -R-Q-R-S-- ----------
C.ZM106       A.....TSRN ATNA---LP- --R------- ---------- -------V-- ------V---
B.HXB2        TWSTEG-NN- EGSD---LP- ---------- K--K------ -S-Q-R-S-- ----------   480

C.CH0505T/F   GK...NNTET FRPGGGNMKD NWRSELYKYK VVEVKPLGVA PTNARRRVVE REKRAVGMGA
A.Q168        -NN.NSTN-- ------D-R- ---------- --KIE----- --K------G -------I--
A.Q842        -NT.NSTR-- ------D-R- ---------- --KIE----- --K-K----- -------I--
B.BG1168      -DT.N-G--I ------D-R- ---------- --QIE----- --K-K----Q -------L--
B.JRFL        -IN.E-G--I ------D--- ---------- --KIE----- --K-K----Q -------I--
B.SF162       --EIS-T--I ------D-R- ---------- --KIE----- --K-K----Q ------TL--
C.ZM106       -NGDT-D--- ------D--N ---------- ---I------ --E-K----- -------I--
B.HXB2        -NS.N-ES-I ------D-R- ---------- --KIE----- --K-K----Q -------I--   540

C.CH0505T/F   VFLGFLGAAG STMGAASITL TVQARQLLSG IVQQQSNLLK AIEAQQHMLK LTVWGIKQLQ
A.Q168        ---------- ---------- ---------- ---------- -------L-R ----------
A.Q842        --I------- ---------- ---------- ---------R -------L-- ----------
B.BG1168      M--------- ---------- ---------- -----N---R -------L-Q ----------
B.JRFL        ---------- -------M-- -----L---- -----N---R ------R--Q ----------
B.SF162       M--------- -------L-- ---------- -----N---R -------L-Q ----------
C.ZM106       -L-------- ---------- -A----V--- ---------R -------L-Q ----------
B.HXB2        L--------- -------M-- ---------- -----N---R -------L-Q ----------   600

C.CH0505T/F   ARVLALERYL KDQQLLGMWG CSGKLICTTN VYWNSSWSNK TYGDIWDNMT WMQWEREISN
A.Q168        -----V---- -------I-- ---------- -P-------- SQSE--E--- -L---K----
A.Q842        -----V---- -------I-- ---------S -P-------- SQNE------ -L--DK----
B.BG1168      -----V---- -------I-- ---------A -P--A----- SQEE----L- -----K--N-
B.JRFL        -----V---- G------I-- ---------A -P--A----- SLDR--N--- --E-----D-
B.SF162       -----V---- -------I-- ---------A -P--A----- SLDQ--N--- --E-----D-
C.ZM106       T--------- -------L-- ---R-----A -P-------- SLT------- ----DK-V--
B.HXB2        --I--V---- -------I-- ---------A -P--A----- SLEQ--NHT- --E-D---N-   660

C.CH0505T/F   YTEIIYELLE ESQNQQEKNE QDLLALDRWN SLWNWFNITN WLWYIKIFIM IVGGLIGLRI
A.Q168        --Q---T-I- ---------- -------K-A ------D-SK -----R---- ----------
A.Q842        --Q---D--- ---------- -------K-A N-----D-S- ---------- ----------
B.BG1168      --SV--T--- Q--------- -E--E--K-A ------D--K -----R---- ----------
B.JRFL        --SE--T-I- ---------- -E--E--K-A ------D--K ---------- ----------
B.SF162       --NL--T-I- ---------- -E--E--K-A ------D-SK ---------- -----V----
C.ZM106       --NT--R--- D--S------ K------S-K N--T--D-S- ---------- V---------
B.HXB2        --SL-HS-I- ---------- -E--E--K-A ---------- ------L--- -----V----   720


C.CH0505T/F   IFAVLSLVNR VRQGYSPLSL QTLIPSPRGP DRPGGIEEEG GEQDRNRSTR LVSGFLALVW
A.Q168        V-----V--- ---------F ---L-A---- ---D------ ---G-G--RQ --N--ST-I-
A.Q842        V-----VI-- ---------F --HT-N---L ---ER----- ----K---I- --------A-
B.BG1168      V---I-I--- ---------- --RF-A---- ---E------ -GR---I--- ------P-I-
B.JRFL        V-T---I--- ---------F ---L-A---- ---E------ --R--D--G- --N-----I-
B.SF162       V-T---I--- ---------F --RF-A---- ---E------ --R--D--SP --H-L---I-
C.ZM106       ----I-I--- ---------- ---TQ--G-- --L-R----- -----D--I- --N---T-A-
B.HXB2        V-----I--- ---------F --HL-T---- ---E------ --R--D--I- --N-S---I-   780

C.CH0505T/F   DDLRSLCLFI YHRLRDFILI AARAGELLGR SSLKGLRRGW EALKYLGSLV QYWGLELKRS
A.Q168        ----N----S ------L--- ---IV----. ......---- --I---WN-L ---IQ---N-
A.Q842        ---------S ---------- V--TV----H -------L-- -G-----N-L S---R--RI-
B.BG1168      ---------S -------L-- V--IV----. ......---- -----WWN-L ---SQ---N-
B.JRFL        V--------S ------LL-T VT-IV----. ......---- -V---WWN-L ---SQ---N-
B.SF162       ---------S ------L--- ---IV----. ......---- -----W-N-L ---IQ---N-
C.ZM106       ---------S ---------- ---VV----H ---R--QK-- ---------- ---C------
B.HXB2        ---------S ------LL-- VT-IV----. ......---- -----WWN-L ---SQ---N-   840

C.CH0505T/F   AISLLDTLAI AVGEGTDRIL EFVLGICRAI RNIPTRIRQG FETALL
A.Q168        -----N-T-- --A-----AI -IIQRAIT-V L--------- --R---
A.Q842        -TN----I-- VIAGW---VI -IGQRL---F L---R----- --R---
B.BG1168      -V---N-T-- V-A------I -ALQR----- LH-------- --R---
B.JRFL        -V---NAT-- --A------I -ALQRTY--- LH-------- L-R---
B.SF162       -V--F-AI-- --A------I -VAQR-G--F LH--R----- --R---
C.ZM106       ------SI-M --A------I -L-QR---G- YH--R----- --A---
B.HXB2        -V---NAT-- --A-----VI -V-Q-A---- -H--R----- L-RI--  886
```