HIV vs. HPC

Los Alamos Computer Science Symposium, 2009

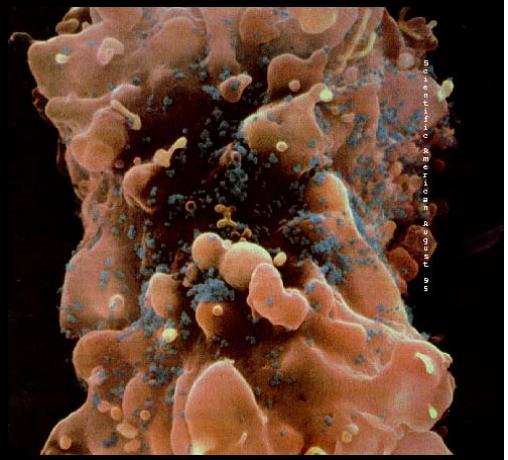
Bette Korber

Marcus Daniels, Tanmoy Bhattacharya, Gnana Gnanakaran

Los Alamos
NATIONAL LABORATORY
EST. 1943

The HIV-1 Pandemic:

UNAIDS 2008 report



- Global: 33,000,000 living with HIV-1
  - 25,000,000 have died of AIDS since 1981
  - 2.7 million new infections:
  - 7,400 people are infected every day
  - 2 million died of AIDS in 2007
  - 15 million orphans
- N America: 1,200,000 living with HIV
  - 23,000 deaths
  - 54,000 new infections
- Some African nations have a prevalence approaching 25% of the adult population
- India, at 1% prevalence, has largest number

The problem:
HIV is an extremely variable virus, and we need a vaccine that protects against the spectrum of diversity

I. Up to 35% of amino acids can differ between two HIV Env proteins

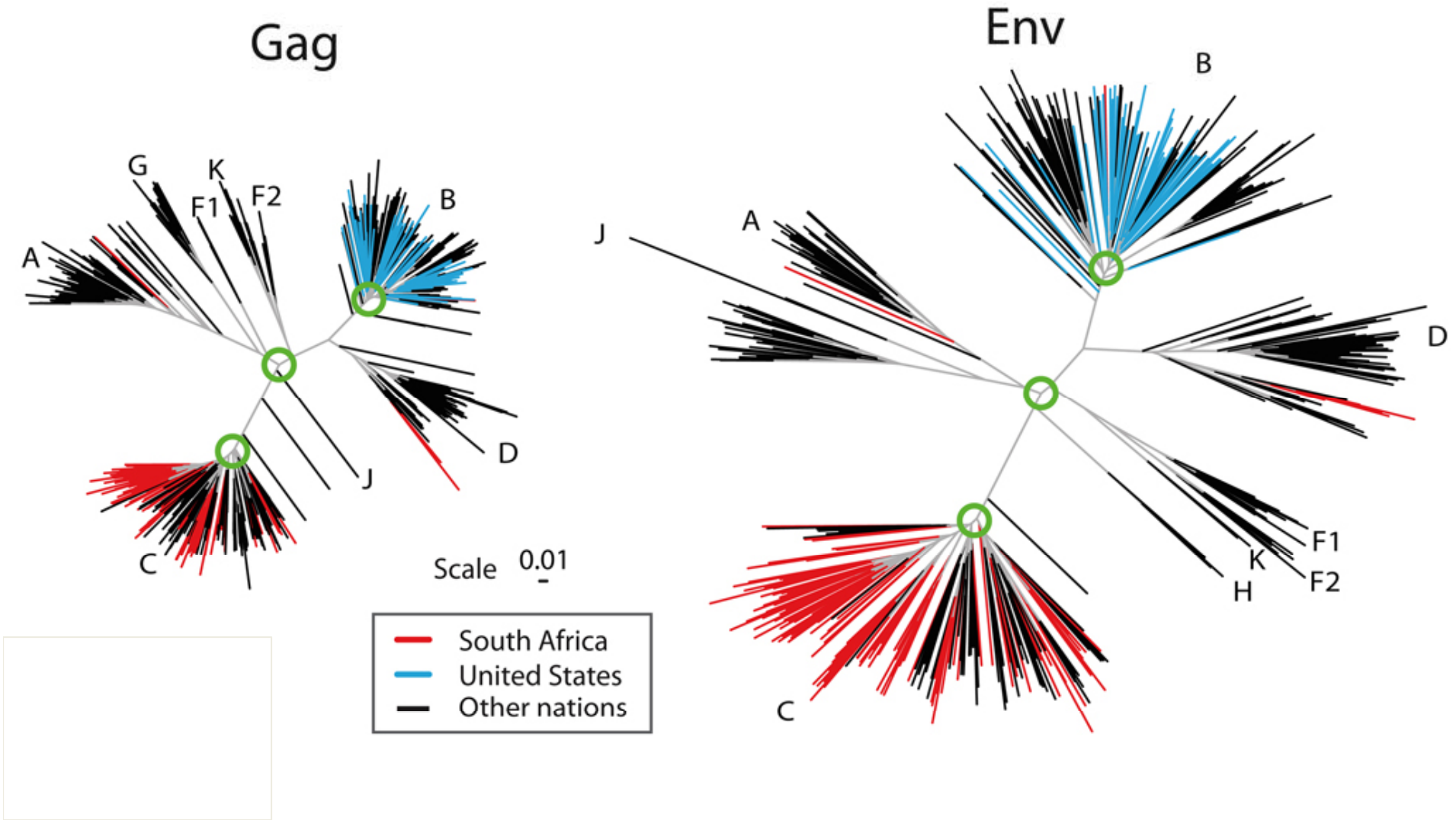
II. 10% amino acid variation can occur in a single individual over time

III. HIV changes it immunological face by:

- base mutations
- recombination
- insertion and deletion
- modification of glycosylation sites

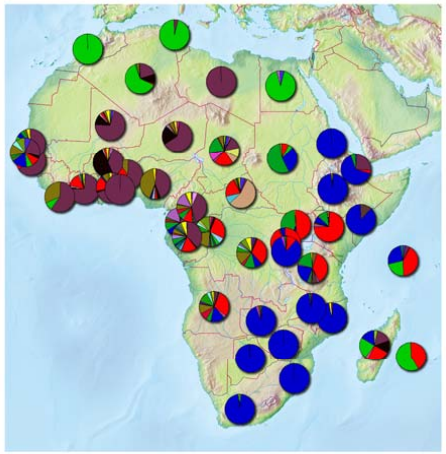
Global versus regional diversity

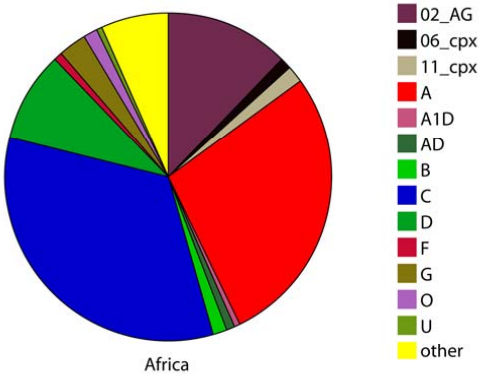
Korber et al., JV 2009



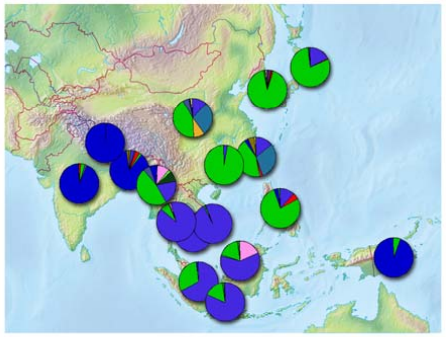
Global versus regional diversity

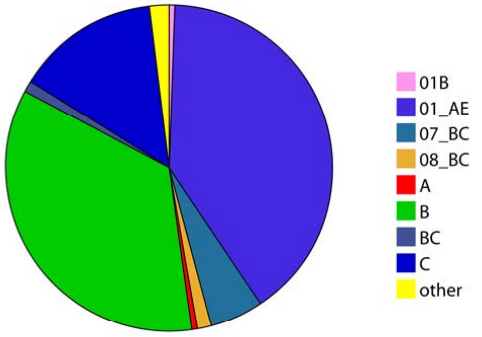
Korber et al., JV 2009



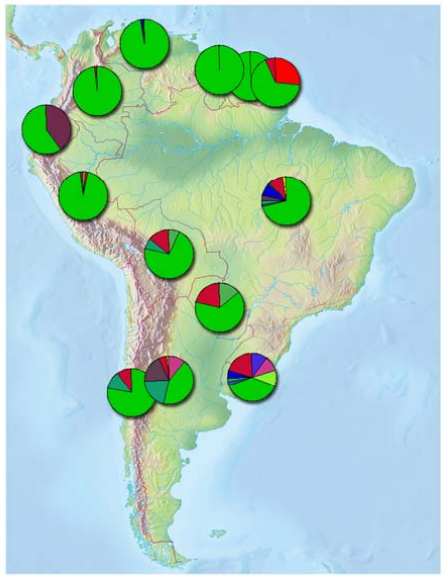


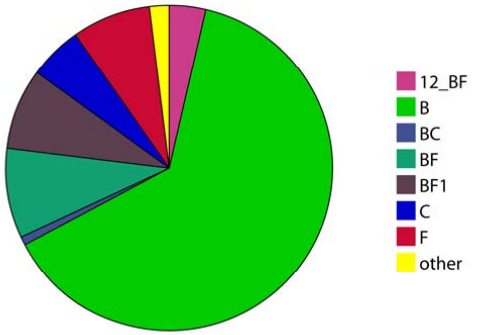
Africa
36,135 sequences





Asia
18,233 sequences

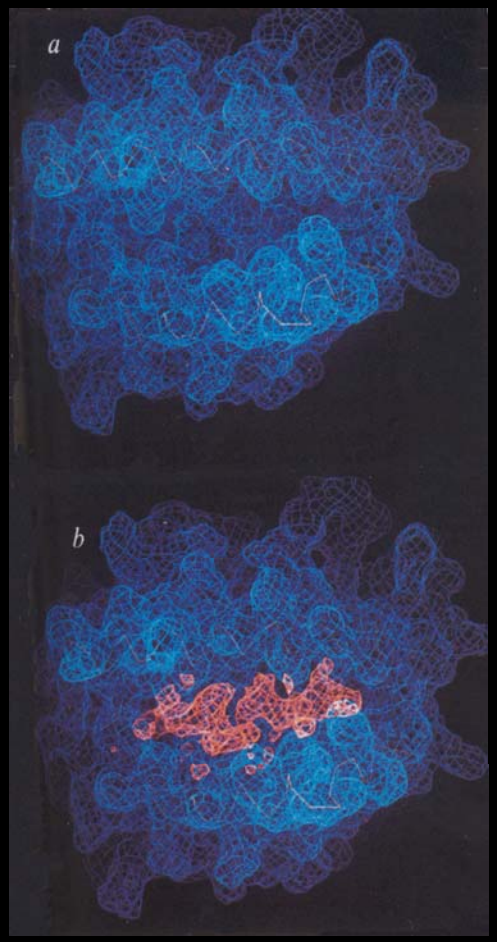




South America
8,160 sequences

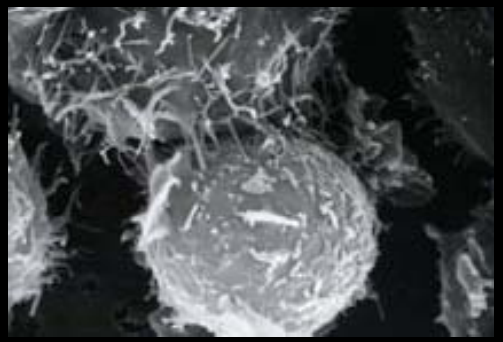
How to make an HIV vaccine that can contend with the diversity?

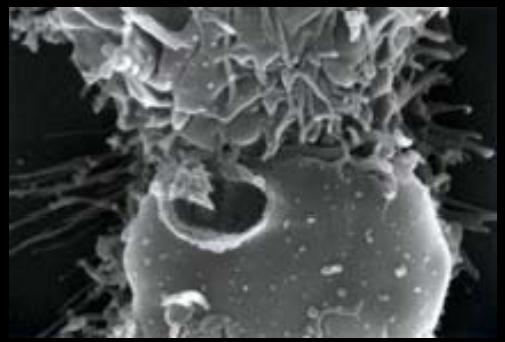
T cell responses are important for an AIDS vaccine



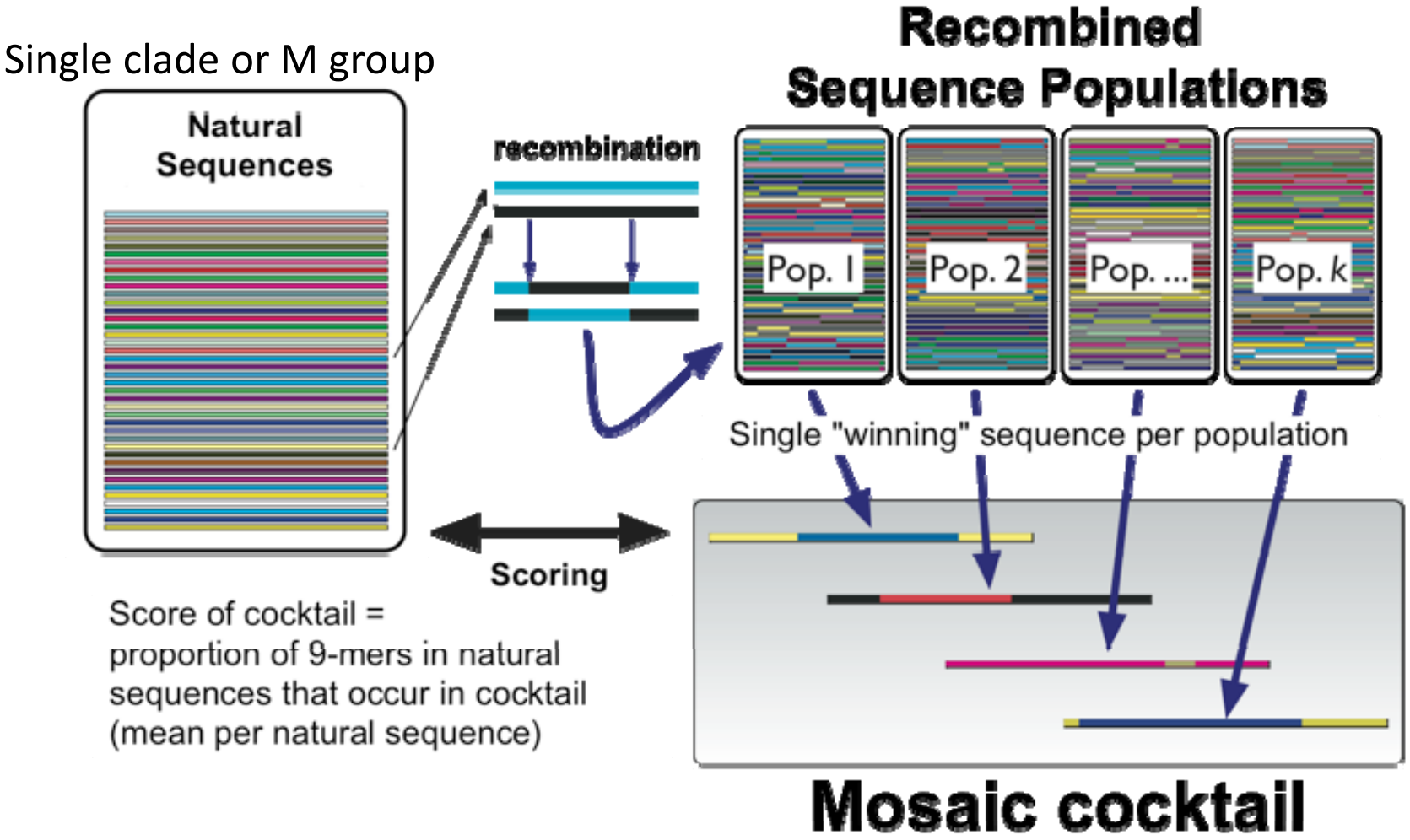
Killer T cells identify HIV infected cells
by recognizing small fragments of viral protein carried
to the cell surface by human class I proteins.

They trigger infected cells to self destruct.



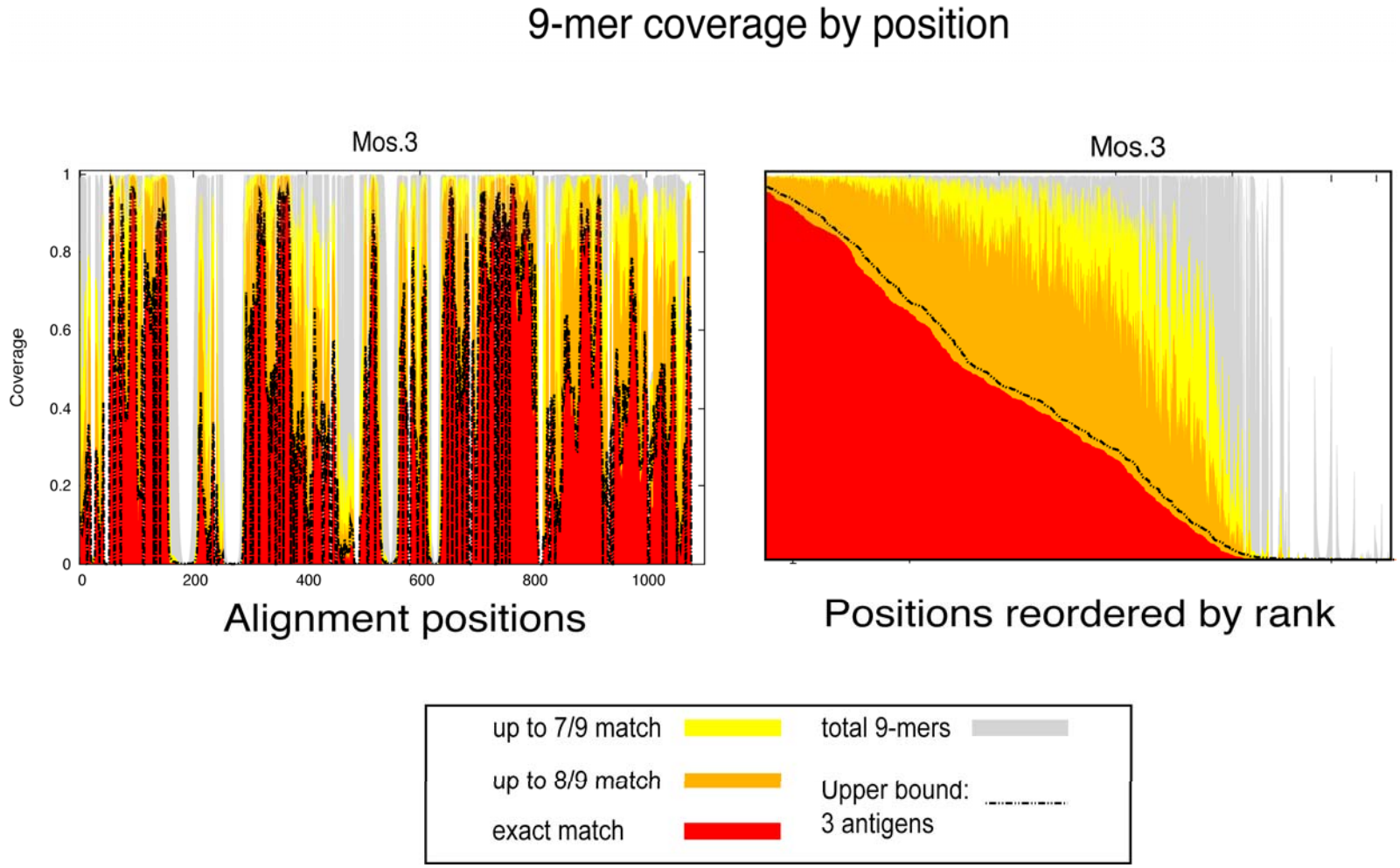


Genetic Algorithm for a *k=4* protein mosaic cocktail (Fischer et al, Nat Med 2007)

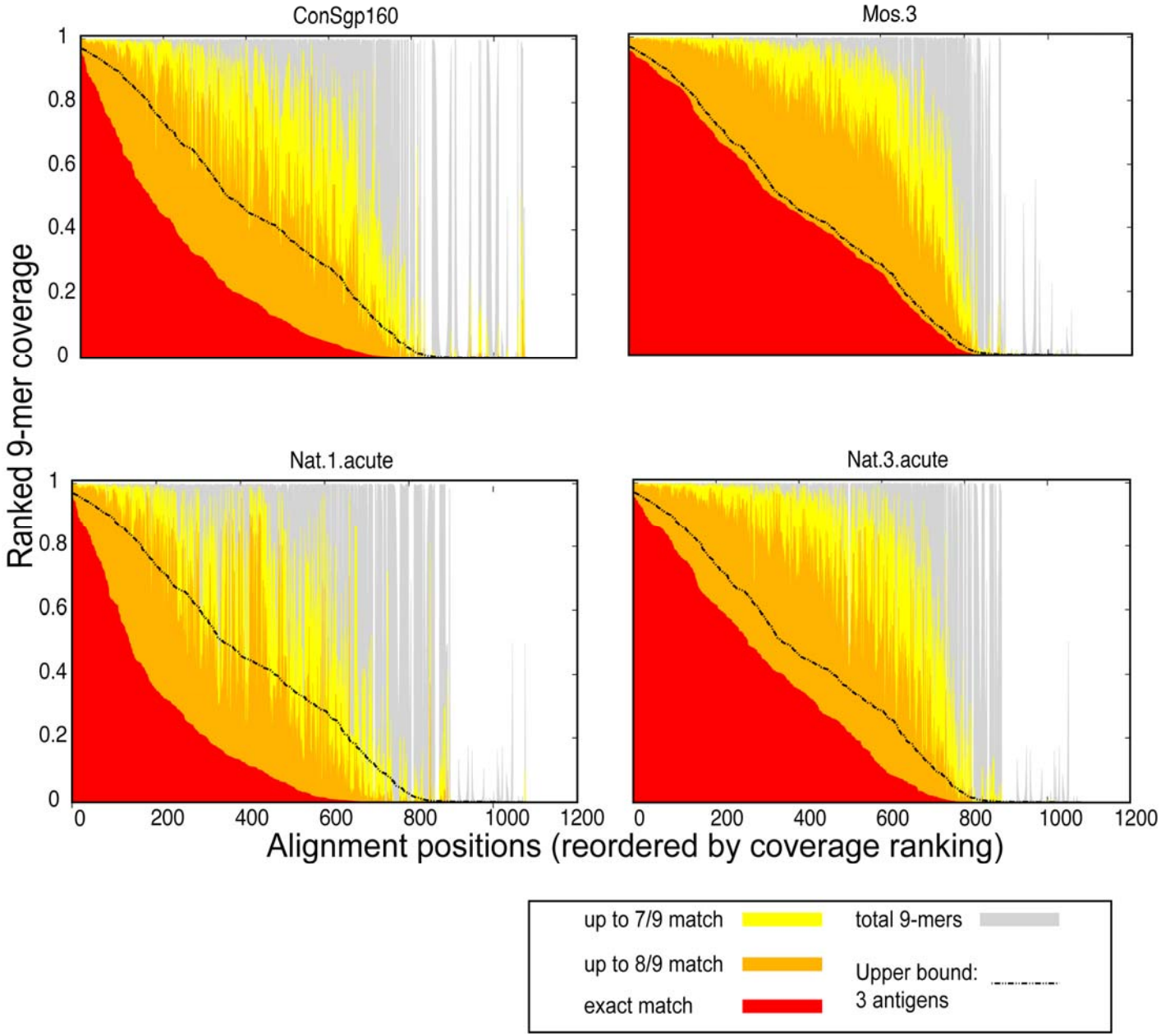


Iterations improve the populations, improve the cocktail

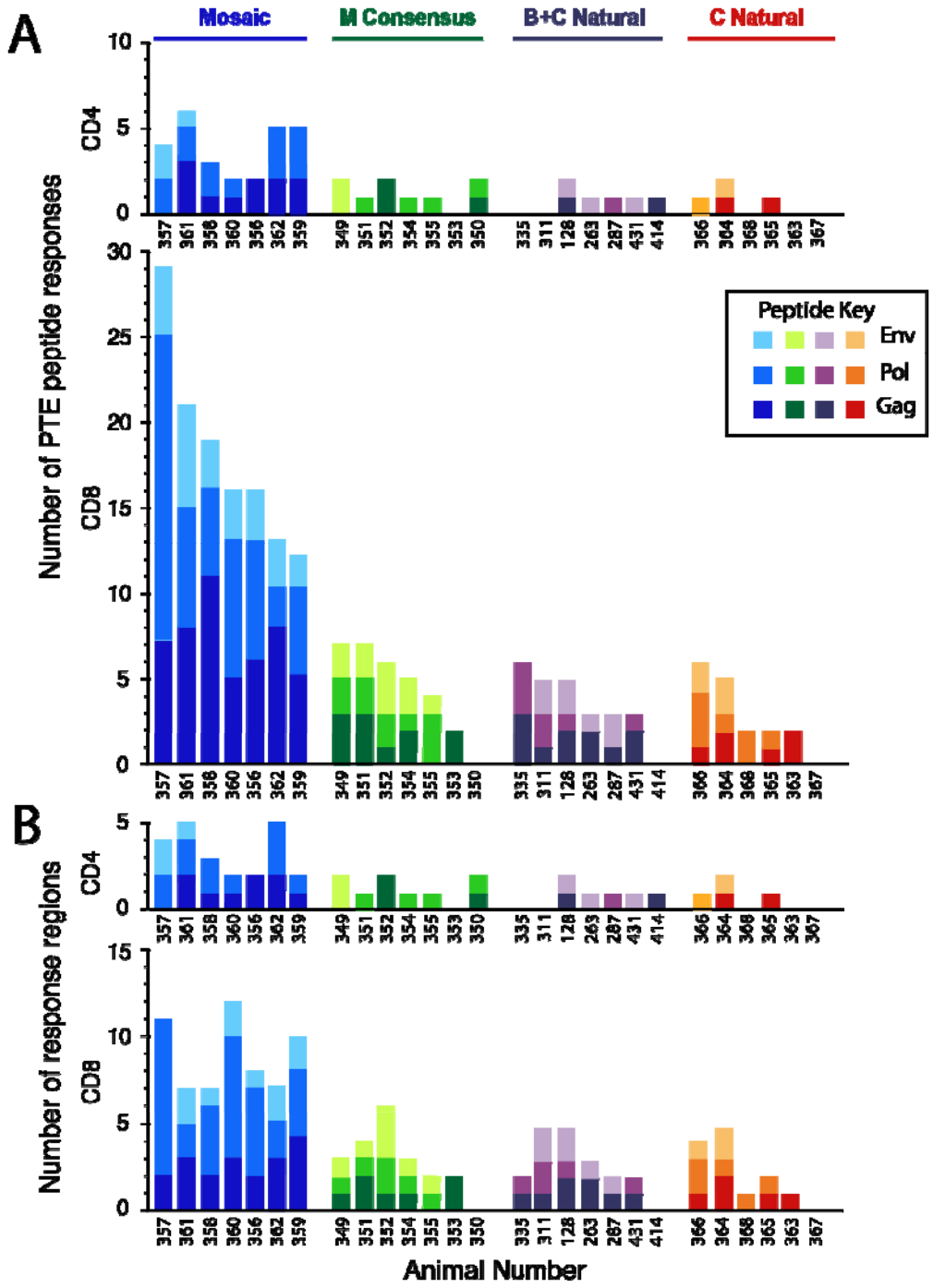
A view of potential epitope coverage -- 3 envelopes



Coverage by different cocktails



Mosaics generate more PTE peptide responses than the Mcon by a factor of 3.7, Poisson regression $p = 2 \times 10^{-12}$



Barouch et al, submitted, Nature Immunology 2009

Example of typical reactions to ConM or the optimal natural vaccine

OptC 366-07:

5 CD8 responses:

OptC	I**VQQQSNLLRAIEAQ**Q			
E54	VVQQQSNLLRAIEAQ	Env	548	562
E72	-VQQQNNLLRAIEAQH	Env	549	563
OptC	A**VFIHNFKRKGGIGG**Y			
P22	AVFIHNFKRKGGIGG	Pol	894	908
P236	-VLIHNFKRKGGIGGY	Pol	895	909
OptC	**MAICEEMEKEGKIT**K			
P224	TAICEEMEKEGKITK	Pol	190	204
OptC	**CTHGIKPVVSTQLLL**			
E15	CTHGIKPVVSTQLLL	Env	247	261
OptC	**GGPSHKARVLAEAMS**			
G76	GGPSHKARVLAEAMG	Gag	354	368

1 CD4 response:

OptC	**IIGQVRDQAEHLKTA**			
P86	LIGQVRDQAEHLKTA	Pol	876	890

Good matches with solid stretches of identity between vaccine and target PTE peptide

Mosaics had many responses that recognized multiple variant overlapping peptides:

No apparent antigenic competition, broad local responses

```
Mos1  ICTTTVPWNASWSNKSL
Mos2  ICTTAVPWNTSWSNKSQ
E334  ICTTTVPWNASWSNR
E214  -CTTTVPWNSSWSNKT
E158  --TTAVPWNASWSNKSL
E290  --TTAVPWNTSWSNKSL
```

T…A
T…S
A…A
A…T

1) Four variable PTE peptides were recognized
2) In the region of overlap both mosaic forms were recognized, as well a combination of the two
3) A new form (S) was recognized.

Typical pattern of CD8+ PTE peptide responses in a mosaic vaccinated animal (361-07).

22 PTE peptides

8 responsive regions

5 regions included variable peptides that match amino acids *in one or the other* of the mosaics.

(Fisher's p-value = 0.001975)

8 CD8 responses:

Mos1	EQLIKK**ERVYLSWVP**AHKGIG			
Mos2	EQLIKK**EKVYLAWVP**AHKGIG			
P221	EELIKKEKVYLAWVP	Pol	678	692
P250	EPLIKKEKVYLSWVP	Pol	678	692
P316	-KLIEKDKVYLSWVPA	Pol	679	693
P223	--LIKKERVYLSWVPAH	Pol	680	694
P169	------EKVYLAWVPAHKGIG	Pol	684	698
P83	------EKVYLSWVPAHKGIG	Pol	684	698
Mos1	**CTHGIRPVVSTQLLL**			
Mos2	**CTHGIKPVVSTQLLL**			
E15	CTHGIKPVVSTQLLL	Env	247	261
E47	CTHGIRPVVSTQLLL	Env	247	261
Mos1	IC**TTTVPWNASWSNK**SL			
Mos2	IC**TTAVPWNTSWSNK**SQ			
E334	ICTTTVPWNASWSN**R**	Env	603	617
E214	-CTTTVPWN**S**SWSNKT	Env	604	618
E158	--TTAVPWNASWSNKSL	Env	605	619
E290	--TTAVPWNTSWSNKSL	Env	605	619
Mos1	ACQGV**GGPGHKARVL**AEAMS			
Mos2	ACQGV**GGPSHKARVL**AEAMS			
G166	ACQEVGGPGHKARVL	Gag	349	363
G76	-----GGPSHKARVLAEAMG	Gag	354	368
Mos1	AAEW**DRVHPVHAGPIA**PGQ			
Mos2	AAEW**DRLHPVHAGPVA**PGQ			
G319	AAE-DRLHPVHAGPIP	Gag	209	225
G242	-ADWDRLHPVHAGPVA	Gag	210	224
G44	-AEWDRLHPVHAGPIA	Gag	210	224
G277	---WDRVHPVHAGPNPPG	Gag	212	226
G102	----DRVHPVHAGPIPPGQ	Gag	212	226
Mos1	**HSNWRAMASDFNLPP**			
Mos2	**HSNWRAMASEFNLPP**			
P93	HSNWRAMASDFNLPP	Pol	731	745
Mos1	**KGRPGNFLQNRPEPT**			
Mos2	**KGRPGNFLQSRPEPT**			
G86	KGRPGNFLQNRPEPT	Gag	442	456
Mos1	**SRELERFAVNPGLLE**			
Mos2	**SRELERFALNPGLLE**			
G39	SRELERFALNPGLLE	Gag	38	52

Conclusion I:

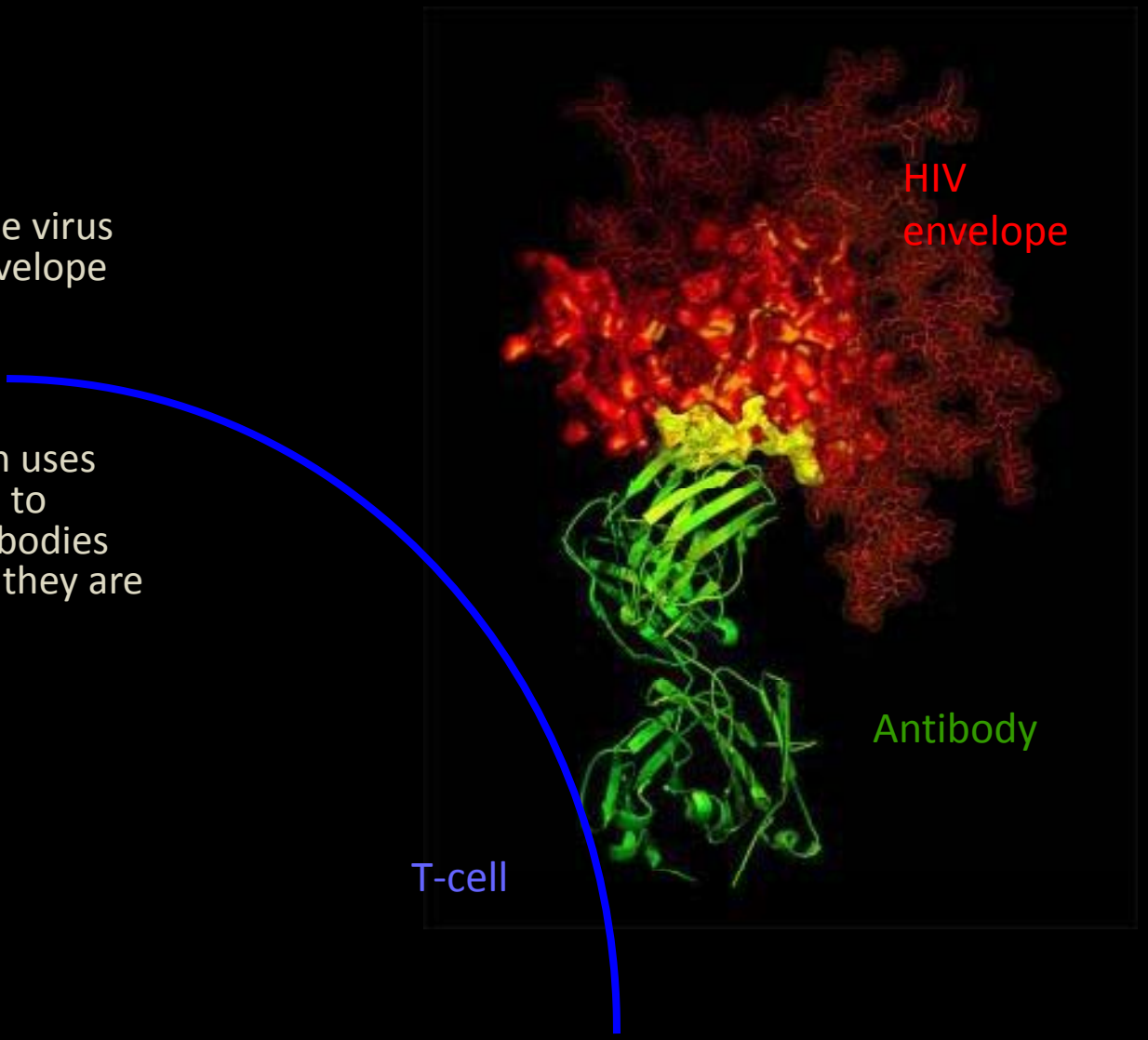
HIV mosaic vaccine elicit immune responses with greater breadth and depth.

1st clinical Phase 1 trial is underway

Cross-reactive neutralizing antibody responses are important for an HIV vaccine

Antibodies we make to the virus can bind directly to its envelope and keep
it from entering our cells.

Another vaccine approach uses the HIV envelope protein to simulate HIV-specific antibodies that will block the virus if they are exposed.



We hope to understand, predict and manipulate mutational patterns that correlate with:

- The potential of viral envelopes to elicit potent neutralizing antibodies for vaccine design
- Acute infection sequences, that are a consequence of the transmission bottleneck

Phylogenetics inform our understanding of the phenotypic consequences of mutation

Validation test: Computational analysis of the b12 epitope

Computational tools for clustering like-patterns of neutralization potency were designed based on K-means clustering, factoring in the uncertainty that results from limited sampling (bootstrap) and assay variability (noise).

Phylogenetically corrected methods were used to identify associations between genetic mutations and distinct neutralization clusters.

Analysis of data from 251 clonal Env-pseudotyped viruses assayed with b12 (25-50 μg/ml highest dose)

- 6 signatures identified in gp120
- 1 signature identified in gp41

ML Phylogeny

251 sequenced Envs with known b12 susceptibility

163 were resistant, 65%

Why a phylogenetic correction?

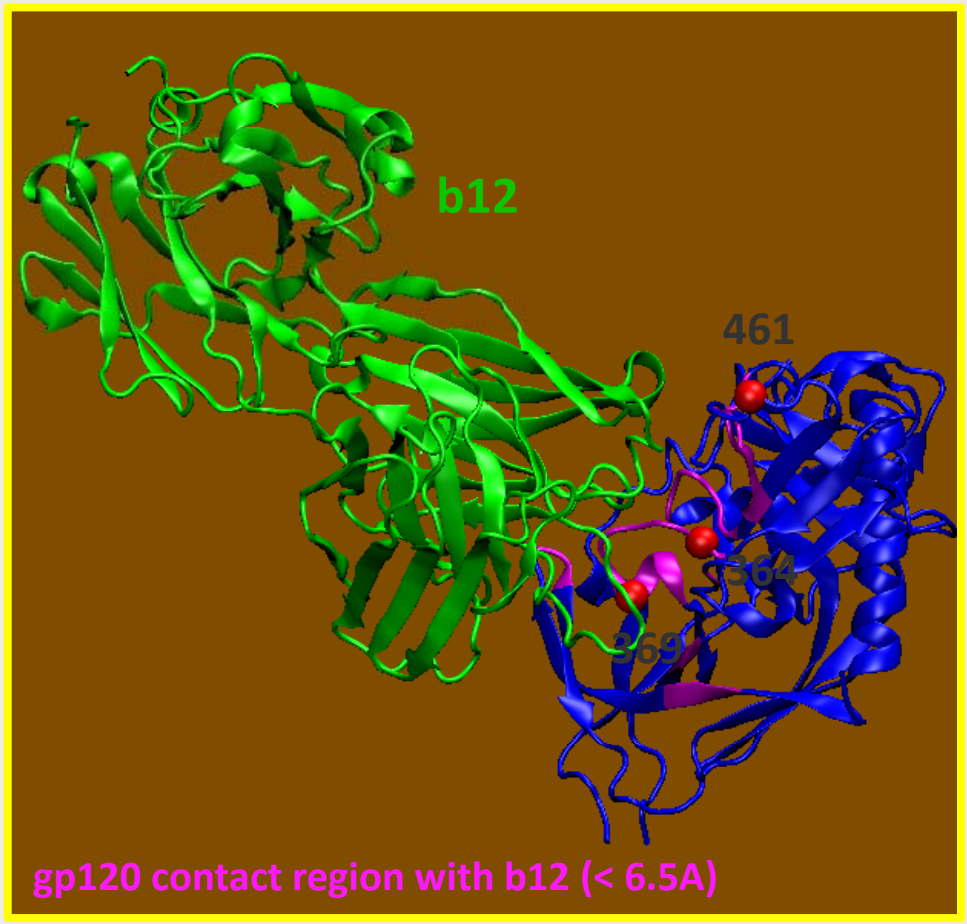
A few of the tested Envelopes were from isolates from the same individual.

In one case, 3301, they differed in susceptibility

Six of the seven b12 signatures are in gp120

Sites 364, 369 and 461 are at the b12 binding surface.

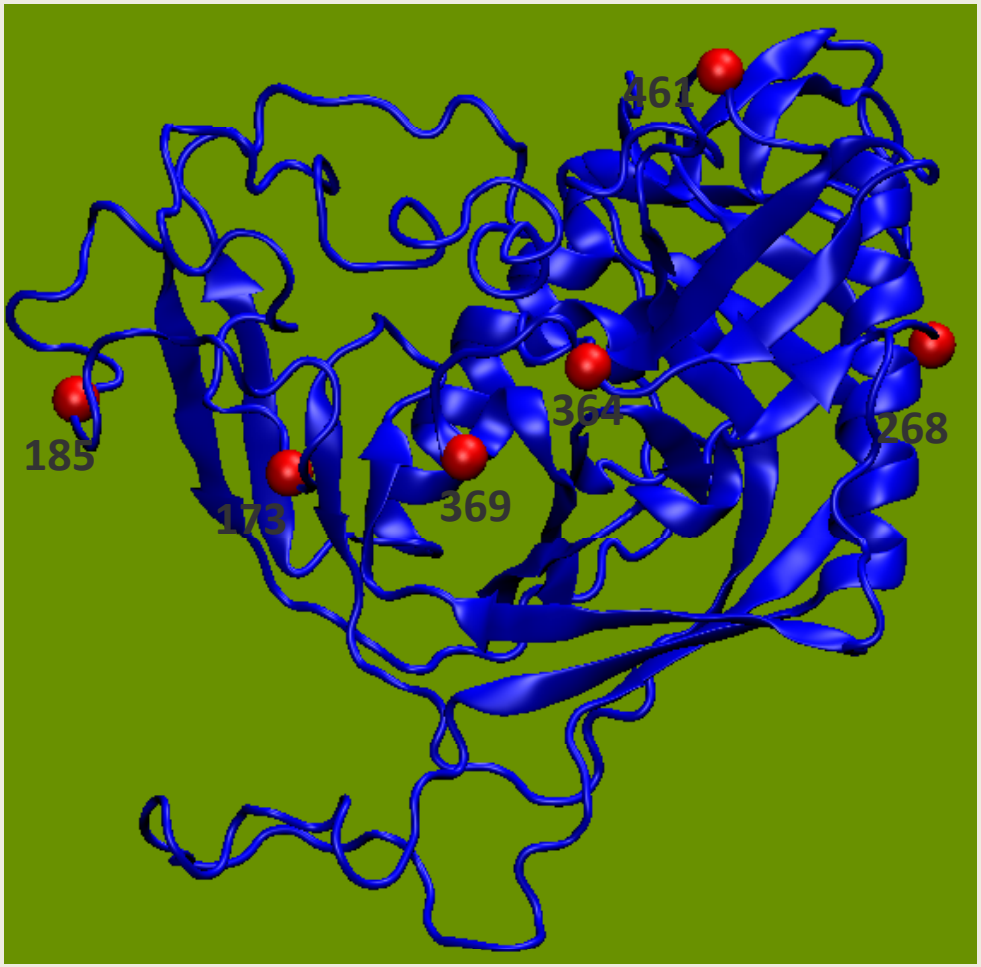
Violet marks the gp120 region <6.5 A from the b12 binding Fab.



Six of the seven b12 signatures are in gp120

- Sites 173 and 185 are in the C-terminus of V2. Mutation at 185 leads to b12 escape in JR-CSF.
- Site 268 is spatially far from the b12 binding site

A V2 modeled loop is shown for visualization



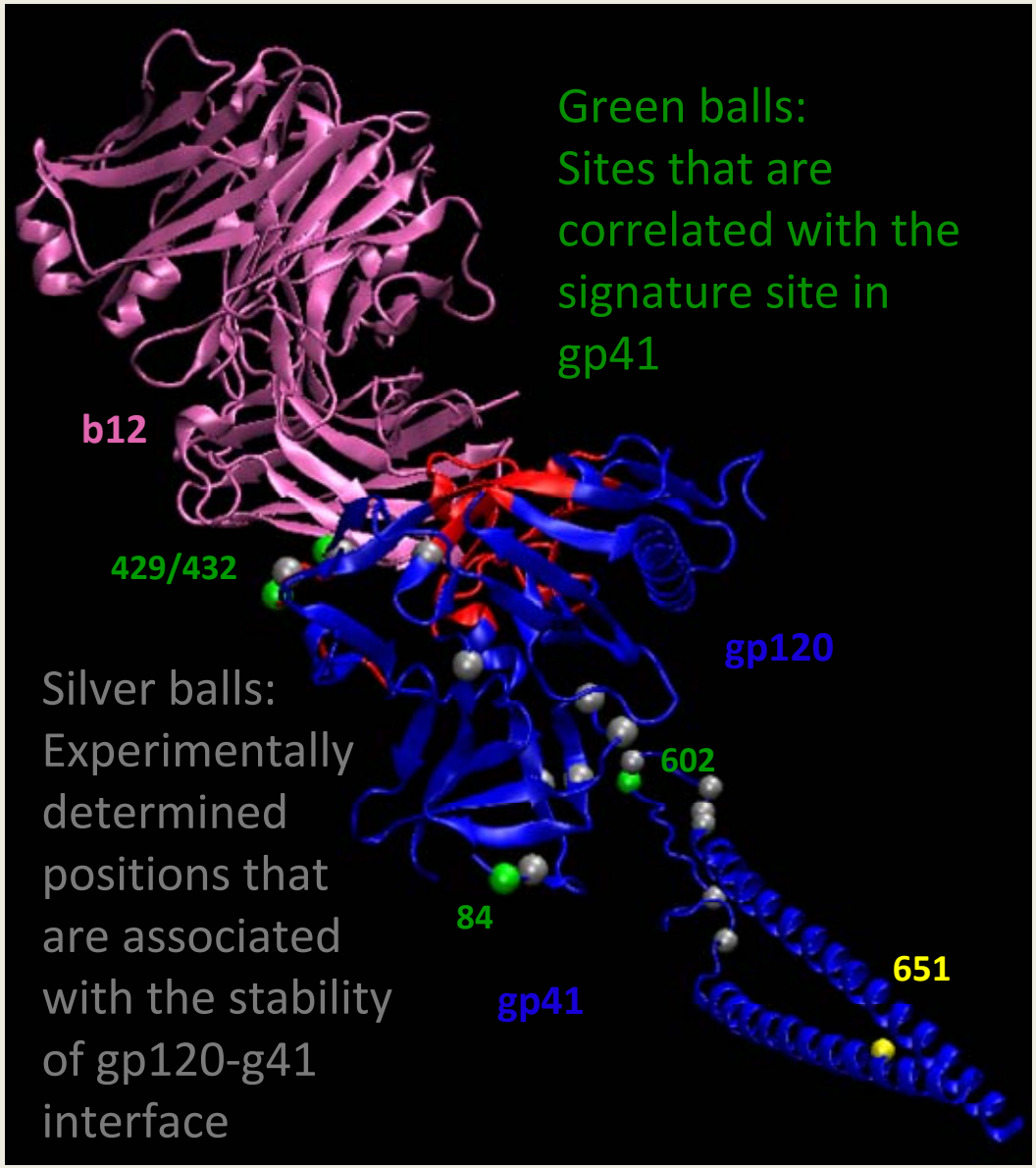
Site 651 was the only site in gp41 associated with b12. Proximal to the N-heptad repeat region associated with T-20.

Gp41 is modeled as a post-fusion conformation, placed proximal to gp120 in the expected orientation.

Green sites covary with 651.

Silver sites impact gp120/gp41 assembly.

•



Outcome of the validation test based on b12 epitope

❖ Correctly identified amino acid positions that are known to occur in the b12 binding surface of gp120.

❖ Able to identify other signatures that are consistent with an impact of quaternary structure on the b12 epitope.

❖ Computational analysis appears useful for delineating molecular determinants of complex neutralization epitopes on HIV-1 Env.

Signature Analysis of Broadly Cross-Reactive NAb Responses in HIV-1-Infected Individuals

Rationale: **Many Env epitopes are poor immunogens despite their strong antigenicity (e.g., b12 epitope).**

Goal: **Identify amino acid signatures in serum-derived gp160 genes that associate with the magnitude and breadth of neutralizing antibody responses in HIV-1-infected individuals.**

Approach: **Analyze checkerboard-style neutralization data comprising a multi-subtype panel of 25 clonal Env-pseudotyped viruses and sera from 68 HIV-1-infected individuals from whom a serum gp160 sequence was derived by single genome amplification (SGA).**

Finding: **Four amino acid signatures in serum gp160 sequences strongly associated with broadly NAb responses. All four signatures were in the CD4-induced coreceptor binding site of gp120. Fewer PNLG in V2 and shorter V4 and V5 regions also associated with strong NAb responses.**

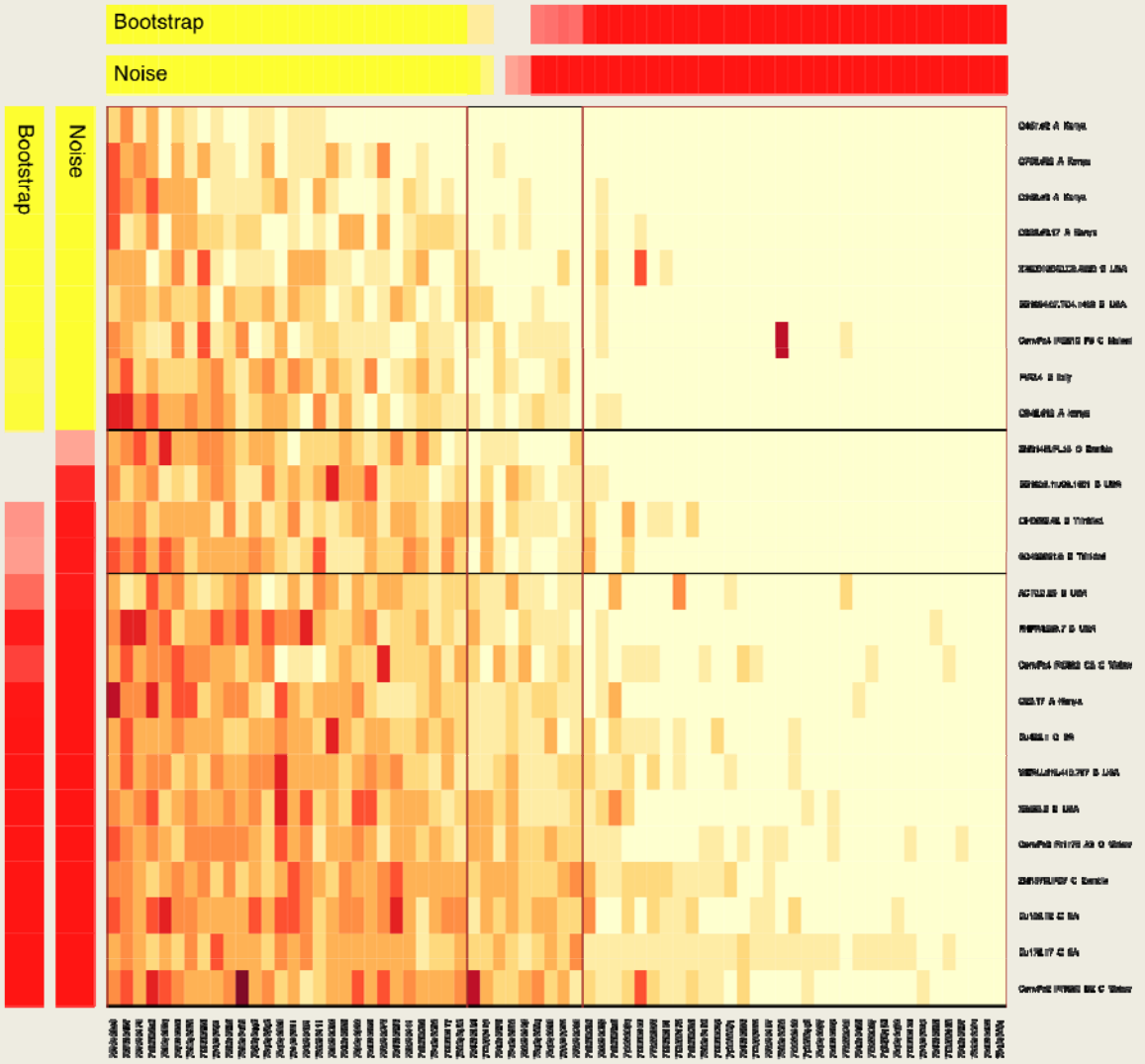
Clustering strategy

K = 2 heat map of the pilot study data

This map was used for the primary signature analysis

90% cutoff for plasma, 75% for test Envs

We also explored this option For defining cluster signatures, comparing the high (left) and low (right) extremes

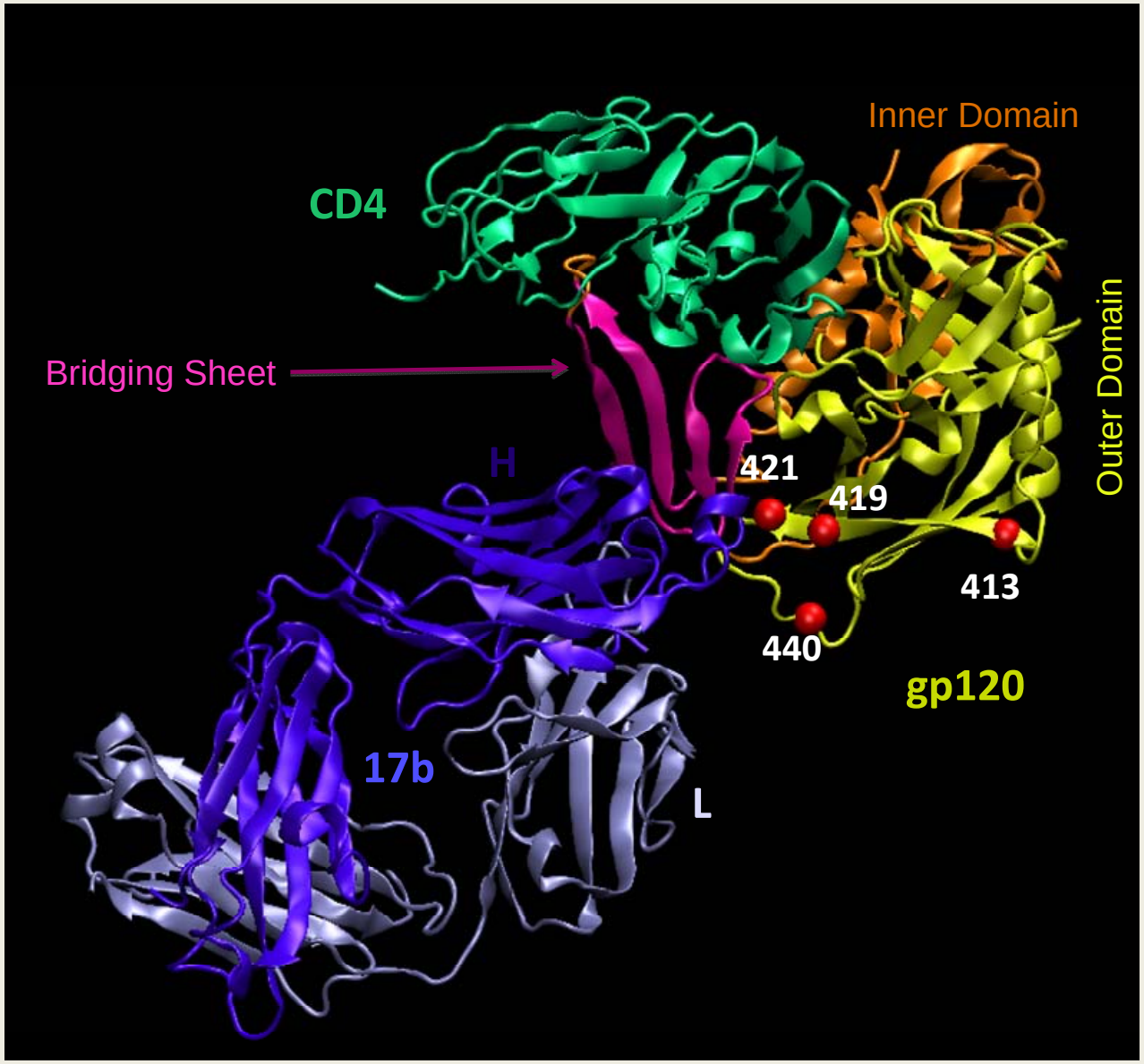


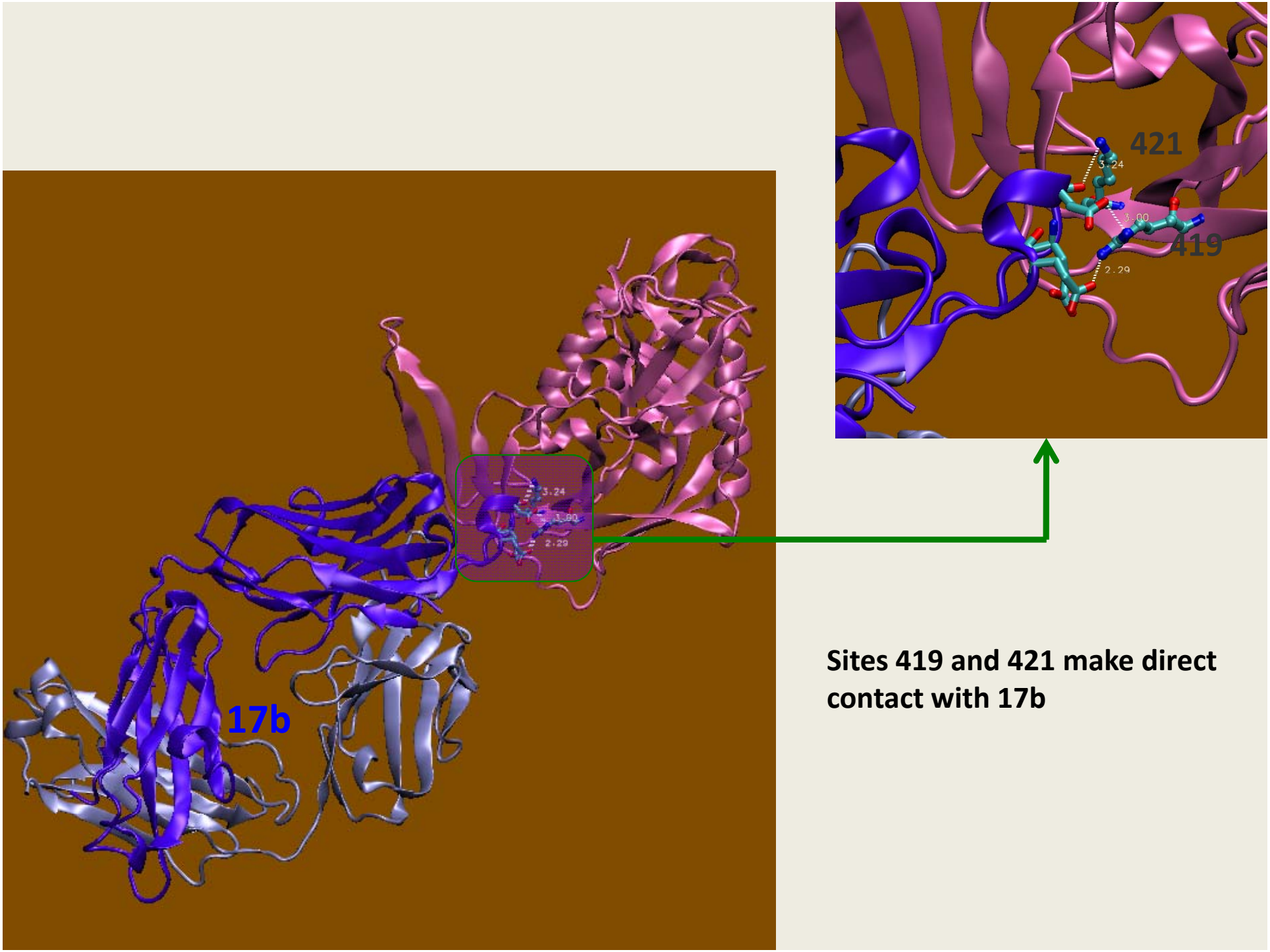
Four signatures identified.

All four sites are in the CD4i region of gp120.

Mutagenesis studies have shown that 419/421 (V4) and 440 (C5) are critical for CCR5 binding.

413 is often a PNLG in V4.





Sites 419 and 421 make direct contact with 17b

Conclusions II:

We have successfully completed a proof of concept signature study with a very well characterized neutralizing antibody, b12.

We have identified a functional domain on the HIV Env protein that has a highly significant association with the infected individuals ability to produce broadly neutralizing antibodies.

ML tree for acute signature analysis and B cell vaccine design

10,026 within-patient sequences

7,063 clade B, 2,963 clade C

Matched acute and chronic cases from 404 subjects

Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection.

Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, Sun C, Grayson T, Wang S, Li H, Wei X, Jiang C, Kirchherr JL, Gao F, Anderson JA, Ping LH, Swanstrom R, Tomaras GD, Blattner WA, Goepfert PA, Kilby JM, Saag MS, Delwart EL, Busch MP, Cohen MS, Montefiori DC, Haynes BF, Gaschen B, Athreya GS, Lee HY, Wood N, Seoighe C, Perelson AS, Bhattacharya T, Korber BT, Hahn BH, Shaw GM.
Proc Natl Acad Sci U S A. 2008 May 27;105(21):7552-7

Quantitating the multiplicity of infection with human immunodeficiency virus type 1 subtype C reveals a non-poisson distribution of transmitted variants.

Abrahams MR, Anderson JA, Giorgi EE, Seoighe C, Mlisana K, Ping LH, Athreya GS, Treurnicht FK, Keele BF, Wood N, Salazar-Gonzalez JF, Bhattacharya T, Chu H, Hoffman I, Galvin S, Mapanje C, Kazembe P, Thebus R, Fiscus S, Hide W, Cohen MS, Karim SA, Haynes BF, Shaw GM, Hahn BH, Korber BT, Swanstrom R, Williamson C; CAPRISA Acute Infection Study Team; Center for HIV-AIDS Vaccine Immunology Consortium.
J Virol. 2009 Apr;83(8):3556-67.

Los Alamos NATIONAL LABORATORY EST.1943

CHAVI CENTER FOR HIV/AIDS VACCINE IMMUNOLOGY

Unrooted bifurcating trees have double factorial possible representations for *n* leaf nodes:

(2n – 5)!! = 1x3x5x…(2n-5)

N = 10	2.0×10^6 population of New Mexico
N = 22	5×10^{22} stars in the observable universe
N = 56	10^{80} fundamental particles in the observable universe
N = 63	8×10^{100}, 8 googles
N = 1,000	2×10^{2860}

N = 10,000 8×10^{38658}

We need to consider >10,000 sequences… and an approximate answer will be useful.
In particular, ML trees allow reconstruction of ancestral states

Supercomputing effort to realize CHAVI's data as a phylogenetic tree

- ML phylogenic inference code
  - Parallelized by Tanmoy Bhattacharya, (Korber et al. 288:1789 Science 2000)
  - Adapted for HLA signature analysis by Marcus Daniels (Bhattacharya et al. Science. 2007 315:1583)

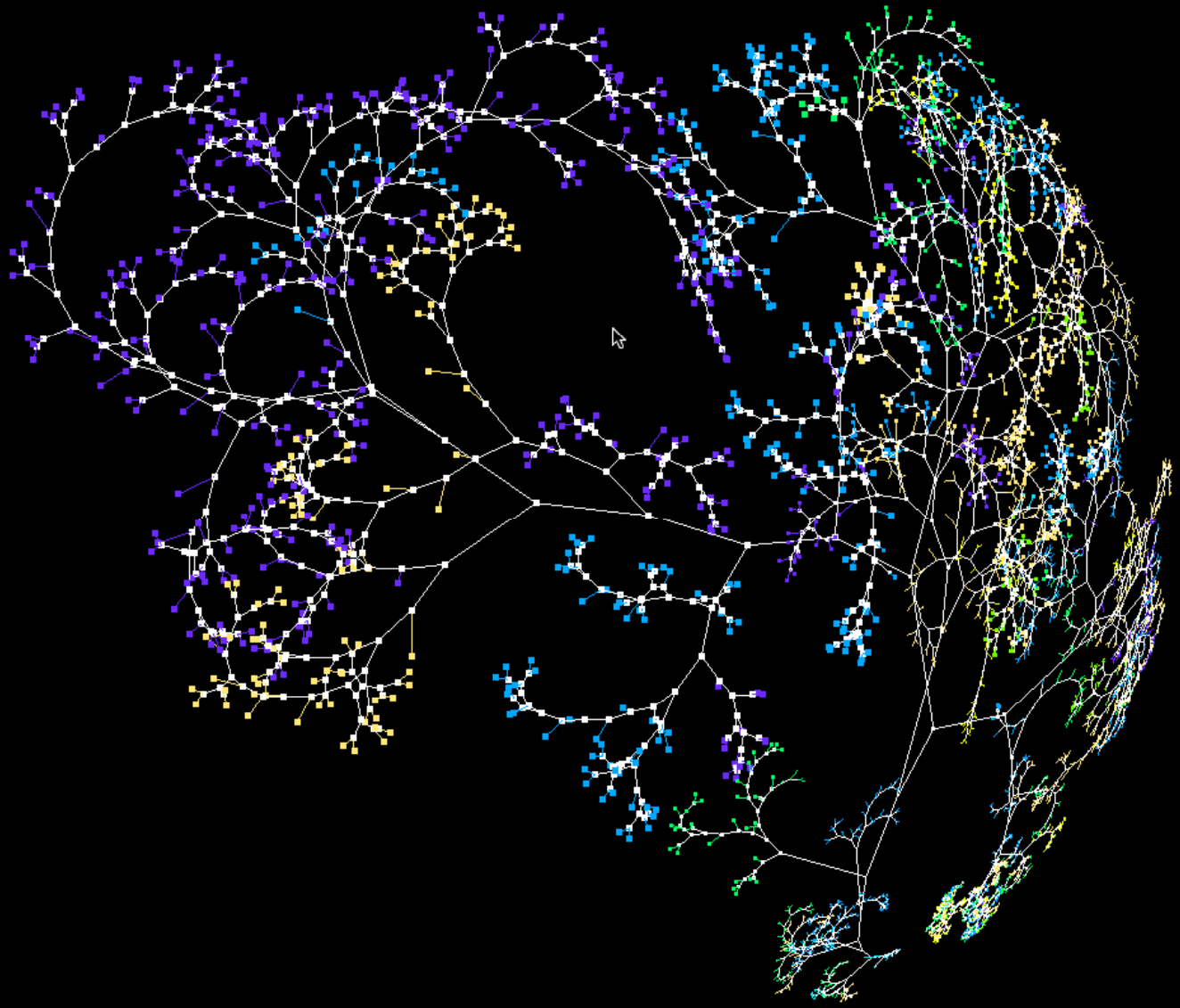
- Accelerated for the Cell processor and run on Roadrunner courtesy of the Los Alamos Supercomputing Center: Marcus Daniels

- Graph construction: using Phylo3D
  - http://digitised.info/content/view/15/42/

- Hyperbolic geometry layout
  - Tamara Muzner at Stanford developed the layout techniques in her PhD thesis
  - http://graphics.stanford.edu/papers/munzner_thesis/

- Vizualization program:
  - Walrus by Young Hyun at the San Diego Supercomputing Center
  - http://www.caida.org/tools/visualization/walrus/

Branches colored by study subject



Select the 4 transmitted virus strains that in combination give the maximum coverage of potential B cell epitopes based on structural proximity within Envelope

B Clade
Fiebig 4

Sequence Set:

04013242
ADARC

Acknowledgements

Analyses contributions: LANL
Marcus Daniels
Gnana Gnanakaran
Tanmoy Bhattacharya
Alan Lapedes
James Theiler
Peter Hraber
Will Fischer

Harvard
Mike Seaman
Norman Letvin
Dan Barouch

Duke
Feng Gao
Barton Haynes
David Montefiori
Kelly Greene
Hongmei Gao

UAB
Beatrice Hahn
George Shaw
Brandon Keele

Los Alamos NATIONAL LABORATORY EST. 1943

NIAID

CHAVI CENTER FOR HIV/AIDS VACCINE IMMUNOLOGY

Collaborating Networks: CHAVI, CAVD, LANL

CAVIMC Comprehensive Antibody - Vaccine Immune Monitoring Consortium