

Research

Open Access

## Evidence of HIV-1 adaptation to host HLA alleles following chimp-to-human transmission

Nobubelo K Ngandu\*<sup>1</sup>, Cathal Seoighe<sup>2</sup> and Konrad Scheffler<sup>3</sup>

Address: <sup>1</sup>National Bioinformatics Node, Institute of Infectious Diseases and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Anzio Road, Observatory, 7925, South Africa, <sup>2</sup>School of Mathematics, Statistics and Applied Mathematics, National University of Ireland Galway, Ireland and <sup>3</sup>Computer Science Division, Dept of Mathematical Sciences, University of Stellenbosch, Private Bag X1, 7602 Matieland, South Africa

Email: Nobubelo K Ngandu\* - nobubelo@cbio.uct.ac.za; Cathal Seoighe - cathal.seoighe@nuigalway.ie;

Konrad Scheffler - kscheffler@cs.sun.ac.za

\* Corresponding author

Published: 10 October 2009

Received: 1 September 2009

*Virology Journal* 2009, **6**:164 doi:10.1186/1743-422X-6-164

Accepted: 10 October 2009

This article is available from: <http://www.virologyj.com/content/6/1/164>

© 2009 Ngandu et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** The cytotoxic T-lymphocyte immune response is important in controlling HIV-1 replication in infected humans. In this immune pathway, viral peptides within infected cells are presented to T-lymphocytes by the polymorphic human leukocyte antigens (HLA). HLA alleles exert selective pressure on the peptide regions and immune escape mutations that occur at some of the targeted sites can enable the virus to adapt to the infected host. The pattern of ongoing immune escape and reversion associated with several human HLA alleles has been studied extensively. Such mutations revert upon transmission to a host without the HLA allele because the escape mutation incurs a fitness cost. However, to-date there has been little attempt to study permanent loss of CTL epitopes due to escape mutations without an effect on fitness.

**Results:** Here, we set out to determine the extent of adaptation of HIV-1 to three well-characterized HLA alleles during the initial exposure of the virus to the human cytotoxic immune responses following transmission from chimpanzee. We generated a chimpanzee consensus sequence to approximate the virus sequence that was initially transmitted to the human host and used a method based on peptide binding affinity to HLA crystal structures to predict peptides that were potentially targeted by the HLA alleles on this sequence. Next, we used codon-based phylogenetic models to quantify the average selective pressure that acted on these regions during the period immediately following the zoonosis event, corresponding to the branch of the phylogenetic tree leading to the common ancestor of all of the HIV-1 sequences. Evidence for adaptive evolution during this period was observed at regions recognised by HLA A\*6801 and A\*0201, both of which are common in African populations. No evidence of adaptive evolution was observed at sites targeted by HLA-B\*2705, which is a rare allele in African populations.

**Conclusion:** Our results suggest that the ancestral HIV-1 virus experienced a period of positive selective pressure due to immune responses associated with HLA alleles that were common in the infected human population. We propose that this resulted in permanent escape from immune responses targeting unconstrained regions of the virus.

## Background

Phylogenetic analysis indicates that the human immunodeficiency virus type 1 (HIV-1) originated from simian immunodeficiency virus infecting chimpanzees (SIVcpz) through a chimpanzee-to-human zoonotic transmission [1-4]. Until recently [5], the natural hosts of the virus, the chimpanzee, have been thought to remain asymptomatic throughout infection despite high viral loads [6-8]. In humans, however, an increase in viral load is usually associated with progression to the acquired immunodeficiency syndrome (AIDS) and subsequently death [9-13]. The causes of the difference in disease progression may involve either differences in the host and/or between the HIV-1 and the SIVcpz viruses.

A zoonotic (*i.e.* cross-species) event is expected to be accompanied by mutations that enable the pathogen to adapt to the new host environment, (e.g. as observed in a study by Baric et al [14]). Indeed, sequence changes have been identified in HIV-1 that are evidence of selective pressure associated with the genetics of the human host [15-17]. In particular, the human cytotoxic T-lymphocyte (CTL) immune response directed against foreign antigens plays a major role in exerting selective pressure on antigenic proteins, including those of HIV-1. The activation and characteristics of the immune responses against the virus have been found to differ remarkably between human and chimpanzee [7,18-20]: an elevated anti-HIV immune response upon infection is characteristic in humans, but the chimpanzee generally maintains a low level of immune activation. The human immune response may therefore exert higher selective pressure on the virus sequence compared to immune responses of the natural host. However, the virus is capable of overcoming the immune response, leading to AIDS.

The CTL immune response is mediated by Human Leukocyte Antigen (HLA) molecules that bind to endogenous antigenic peptides known as epitopes, and transport them to the surface of the infected cell for recognition by CTLs resulting in killing of the infected cell [21]. The HLA gene is highly polymorphic and each HLA molecule binds to peptides that contain specific sequence motif patterns (known as anchor residue motifs) [22,23]. For binding to occur between a peptide and the HLA binding groove, only limited amino acid variation at the main anchor positions of the peptide is allowed [21,24,25]. Successful binding, efficient transport and presentation of a peptide to a CTL depend on the presence of the appropriate anchor residue motif and the overall affinity between the HLA binding groove and the epitope [26,27]. The strength of selective pressure varies between specific CTL immune responses directed by different HLA alleles [28]. Some HLA molecules have been associated with immune escape

mutations at anchor sites which enable the virus to adapt to the host, thus increasing viral load [8,29-31].

Investigation of the evolutionary dynamics of immune escape has focussed primarily on escape mutations that incur a fitness cost and consequently revert to wild type, upon transmission to a host that mounts different immune responses. This can result in a pattern of toggling between escape and wild-type amino acids that is detectable using evolutionary modelling [32]. In this study the focus is on escape mutations that do not incur a cost in terms of viral fitness. Such escape mutations do not experience selection pressure to revert to the wild-type state following transmission to a new host. Consequently, they are associated with episodic selection, rather than the ongoing rapid evolution associated with escape and reversion. Upon transmission to human, SIV is likely to have experienced selective pressure to escape from common human immune responses. Some of these escape mutations would not have had a significant effect on the fitness of the virus and thus would not have experienced strong selection to revert. Consequently, we hypothesized that the branch of the SIV-HIV-1 phylogenetic tree leading to the ancestor of the HIV-1 sequences would include evidence of episodic selection to escape from common HLA alleles.

To investigate the evidence of episodic selection for CTL escape along this branch, we predicted epitopes for HLA alleles, using the SIV consensus sequence to approximate the sequence that was transmitted to humans. We used a structure-based method that estimates the strength of binding between a viral amino acid sequence and an HLA molecule from amino acid pair-wise potentials for the epitope prediction. We selected regions where known anchor residue motifs were present and which had high binding affinity, limiting our analysis of selective pressure to these regions. Finally, we used models of codon sequence evolution to quantify the selective pressure, inferring positive selection from the ratio of nonsynonymous substitution rates ( $dN$ ) to synonymous substitution rates ( $dS$ ) for individual branches in a phylogeny. Branch-specific analysis of selective pressure enabled us to investigate selective pressure along the branch ancestral to the HIV sequences, and hence to study how HIV-1 adapted to the human host upon transmission from chimpanzee.

## Methods

### Sequence data

We downloaded an alignment of HIV-1 group M reference genome sequences and chimpanzee sequences from the Los Alamos database [33]. Previously, we found that some synonymous sites of the nucleotide sequence are highly conserved due to purifying selective pressure acting upon them, and that such conservation of synonymous sites can

cause errors in the prediction of positive selection [34]. Therefore, in this study we removed the conserved regions identified in that study from the alignment (see Additional File 1). The resulting alignment consisted of 9 chimpanzee sequences and 32 HIV-1 sequences starting from codon 1 of the *gag* gene and ending with the *nef* stop codon, but excluding regions listed in Additional File 1. We used HyPhy [35] to build a phylogenetic tree (Figure 1) from the alignment using a neighbour joining method and the pairwise distances calculated using maximum likelihood.

### Predicting HLA binding regions

We used PREDEP [36], a structure-based method for predicting HLA binding peptides to determine potential binding regions across the genome. We used the consensus chimpanzee sequence to predict the best HLA binding regions because it has not been exposed to selective pressure resulting from human HLA and approximates the sequence that was transmitted to human. Consequently, it may be possible to detect epitopes in the chimpanzee sequence that were eliminated from HIV-1 shortly after transmission to humans. PREDEP does not require knowledge of known HLA-binding peptides. The program requires solved crystal structures of the HLA molecules as well as knowledge of amino acid residues on the HLA binding groove that interact with each position of the antigenic peptide sequence. Amino acid pair-wise potentials between the peptide and the amino acids in the HLA binding groove are calculated based on backbone and side-chain interactions.

A score for each HLA-peptide interaction is calculated as the sum of amino acid pair-wise potentials between each peptide residue and the interacting residues of the HLA binding groove. The lower the score, the better the peptide binds to the HLA binding groove, i.e., the higher the binding affinity. Peptides with strong binding affinities to the HLA molecule are most likely to be successfully presented to CTLs *in-vivo*. A test of PREDEP performance showed that 80% of the top 15 percentile best binders were known optimal HLA binding peptides [36]. In this study, we determined the binding energy of all possible peptides across the SIVcpz consensus sequence to each of the six HLA alleles with known crystal structures available in PREDEP. For each available HLA allele, we first selected peptide regions that had binding scores in the best 5% (a conservative threshold chosen to ensure minimal false positives) of those obtained across the chimpanzee sequence for that particular HLA allele. Next, we discarded regions that did not contain the amino acid residues known to give optimal binding at the major binding pockets, i.e. peptides that matched the anchor residue motifs of the HLA allele. For each HLA allele, we generated a new alignment consisting of only the sites in the

potential binding regions identified for that allele for further analysis.

### Analysis of Selective pressure along the HIV-1 ancestral branch

We used the BranchAPriori [37] and GABranch [38] algorithms implemented in HyPhy [35] to analyse branch-specific selective pressure exerted by each HLA allele. We were specifically interested in selective pressure along the SIVcpz branch ancestral to the HIV-1 sequences (labelled 'B' in Figure 1, and referred to below as the HIV ancestral branch), because we expect that this reflects the evolution of the virus around the time of transmission from chimpanzees to human. We therefore investigated whether there is higher selective pressure on this branch at sites that are potential targets for each HLA allele under study. These two approaches calculate the average selective pressure acting upon all regions potentially targeted by an HLA allele, thus combining evidence from multiple sites. We expect that this should result in more powerful tests than can be obtained via site-specific analysis.

#### The BranchAPriori analysis

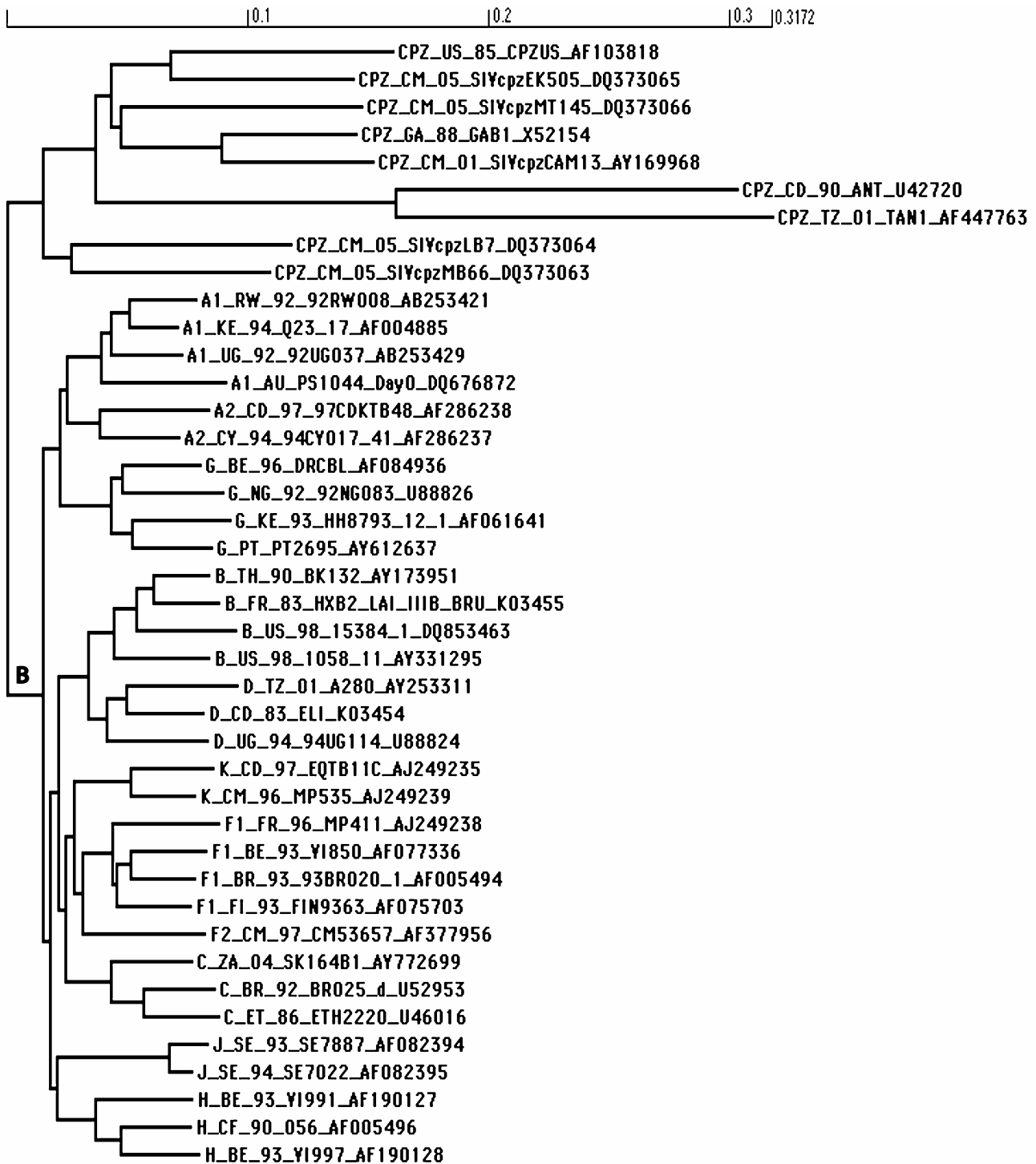
For each HLA-related alignment described in the previous section, we compared the selective pressure along the HIV ancestral branch to the rest of the branches in the tree using the BranchAPriori algorithm. The program outputs a p-value derived from the difference in the log likelihood between the null and the alternative models. The null model assumes a single global dN/dS ratio ( $\omega$ ) across the tree; in the alternative model,  $\omega$  is allowed to have a different value for the HIV ancestral branch.

The *a priori* analysis has the disadvantage of assuming that all the branches in the rest of the tree are under uniform selective pressure. This could result in the analysis having reduced power, for instance when there is strong among-branch heterogeneity of selective pressure in the rest of the tree [39].

#### The GABranch analysis

In order to construct a more realistic null model, we therefore considered models that allow selective pressure to vary across all branches of the tree. We used GABranch [38], a genetic algorithm implemented in HyPhy, to infer branch-specific selective pressure across the entire phylogeny of SIV and HIV-1 sequences and determine, for each of the potential HLA binding regions, whether it evolved under positive selection in the HIV ancestral branch.

As input, the GABranch analysis requires an underlying nucleotide model - we determined the best fitting nucleotide model using a maximum likelihood-based tool available in HyPhy [35]. GABranch then searches through a range of possible codon models with varying dN/dS rate



**Figure 1**

The phylogenetic tree of the 32 HIV-1 group M reference genome sequences and 9 SIVcpz sequences from the Los Alamos sequence database [33]. The chimpanzee sequence names start with 'CPZ' and the group M sequences start with the subtype name. The branch lengths are scaled in reference to the scale given at the top of the tree. The zoonosis event is located on the branch marked "B", referred to in the text as the HIV ancestral branch.

classes, starting with a single rate model, i.e. a model that assumes uniform selection across all the branches of the tree. It tests models with more than one rate class, with the evolutionary rate of each branch being assigned to the best fitting rate class. An Akaike Information Criterion (AIC) value is calculated for each model, based on its fit to the data compared to the single rate model. The model with the best fit to the data, as indicated by the lowest AIC score, is selected and each branch is assigned to a dN/dS rate class (indicated on the phylogeny in figures 2, 3 and 4). Additionally, GABranch provides the proportion of tested models that show support for dN > dS for each branch.

## Results

### **Prediction of HLA binding regions in the chimp sequence**

Only six HLA alleles with solved crystal structures (given in Table 1) were available for analysis within the PREDEP program. Of these six, only HLAs A\*0201, A\*6801 and B\*2705 showed strong binding to regions of the SIVcpz consensus sequence, that is, regions with scores that were within the top five percentile and also contained the preferred anchor residue motifs. The total length of the overlapping peptide regions predicted to be the best binders for each HLA molecule across the chimp genome are given in Table 2.

### **BranchAPriori analysis of differential selection between the HIV and SIV lineages**

We used the BranchAPriori analysis [39] to test, in the predicted binding region for each HLA allele, for evidence of differential selective pressure between the HIV ancestral lineage and the rest of the tree branches. For HLA A\*0201 and HLA A\*6801, the nonsynonymous-synonymous rate ratio  $\omega$  was higher in the HIV ancestral branch than in the rest of the tree (Table 3). The difference was significant only for HLA A\*6801, with a  $\omega$  of 3.5 in the HIV ancestral branch and 1.6 in the rest of the tree ( $p$  value = 0.04). Selective pressure acting along the HIV ancestral branch for sites associated with A\*0201 ( $\omega$  = 2.5) was also high, but failed to differ significantly from the rest of the tree ( $\omega$  = 1.2,  $p$  value = 0.08). For sites associated with B\*2705 there was no significant difference between the HIV ancestral branch and the rest of the tree, with no indication of strong selective pressure in either case ( $\omega$  = 1.0 in the HIV ancestral branch and 1.1 in the rest of the tree,  $p$  value = 0.65).

### **Branch-by-branch analysis of selective pressure using the GABranch algorithm**

We ran the GABranch analysis on the sequence alignments of predicted binding regions from each HLA allele. The  $\omega$  rate categories for the best fitting model as well as the number of branches assigned to each rate category are given in Table 4. Also shown, are the model-averaged val-

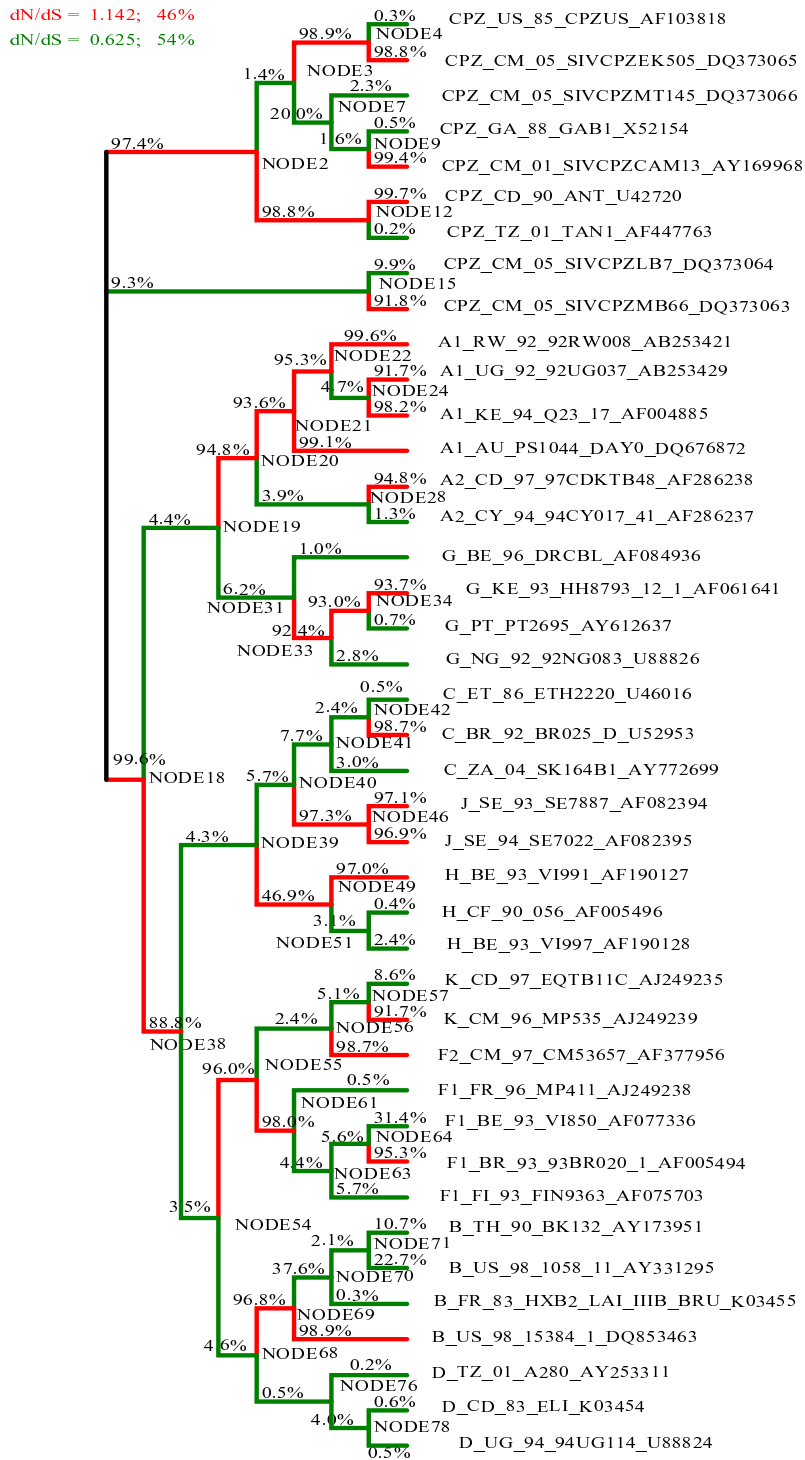
ues obtained for  $\omega$  and estimated probabilities that  $\omega > 1$  on the HIV ancestral branch. The proportion of models that have support for dN > dS for each branch is given in Figures 2, 3 and 4. The mean omega values and model support data for all 79 branches of the three trees are given in Additional files 2 (A\*0201), 3 (A\*6801) and 4 (B\*2705). The tree in Figure 1 was made from the HIV and SIV full length sequences before selecting binding regions for each HLA allele, while those of Figures 2, 3 and 4 were generated from the screened alignments of binding regions for individual HLA alleles. We therefore do not expect these trees to have exactly the same topology.

For HLA A\*6801 (Figure 3), positive selective pressure was inferred along the HIV ancestral branch ( $\omega$  = 1.15). A very high proportion of the tested models (0.996) supported dN > dS along this branch. We also found positive selective pressure along the HIV ancestral branch for HLA A\*0201 ( $\omega$  = 1.14, Figure 2) and again 0.996 of models supported for dN > dS. The best fitting model for the HLA B\*2705 predicted binding sites had only a single rate class under weak purifying selection, and support for dN > dS was not obtained on any branch of the phylogeny (Figure 4). These results are consistent with those of the BranchAPriori analysis and suggest that HLA A\*0201 and HLA A\*6801 exerted positive selective pressure on the HIV-1 sequence in the period immediately following zoonosis, whereas HLA B\*2705 did not exert strong selective pressure on the HIV-1 sequence at any point in the phylogeny.

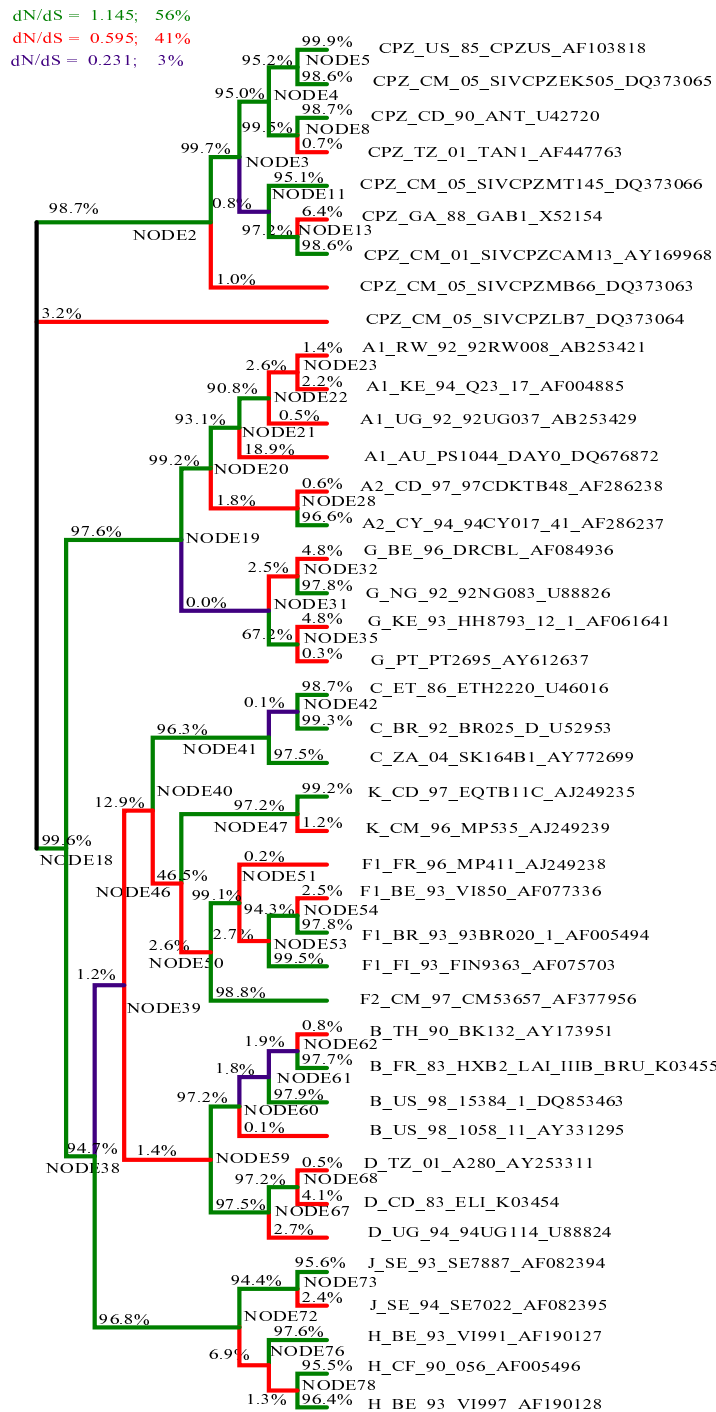
## Discussion

The PREDEP program provides binding predictions for a limited number of HLA molecules with solved crystal structures and preferred binding anchor residue motifs that were predicted from HLA-peptide structural conformations. Only six such HLA alleles known to mediate cytotoxic T-lymphocyte immune responses are available for analysis. Amongst these, we only observed HLAs A\*0201, A\*6801 and B\*2705 to bind strongly to some regions of the consensus SIVcpz genome. Our analysis was therefore restricted to selective pressure potentially exerted by each of these three alleles following the chimpanzee-to-human zoonosis event of HIV.

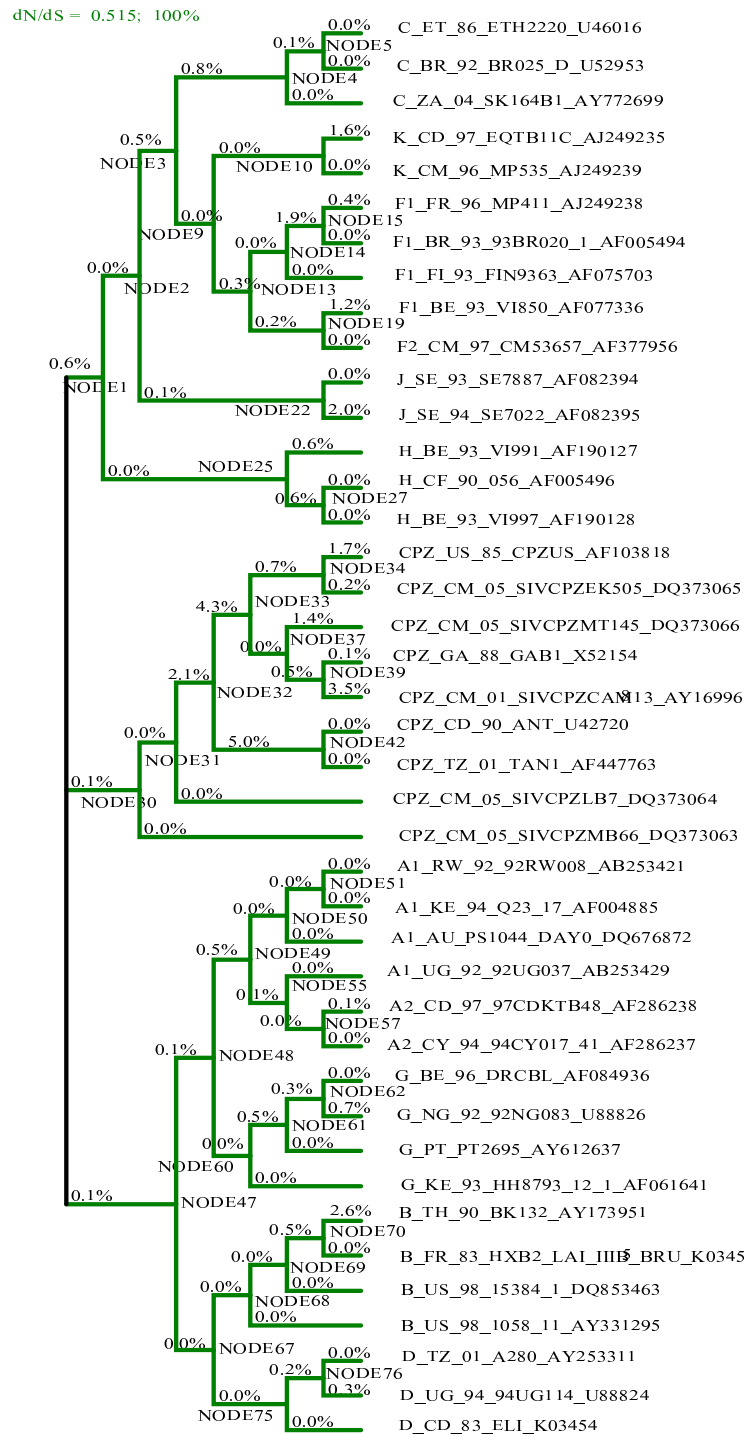
It is interesting that neither the *a priori* nor the GABranch analysis found evidence for positive selection in the HLA B\*2705 alignment, whether along the ancestral HIV-1 branch or any other branch. This is surprising because B27 alleles have been associated with delayed progression to AIDS in HIV-1 infected individuals [40], which in turn is associated with persistent strong positive selection at specific sites [41]. Also, delayed progression is a result of reduced viral replication - this indicates that these sites are important for the fitness of the virus. One possibility that could explain the observed HLA B\*2705 result is that it



**Figure 2**  
**Branch-by-branch selective pressure for regions predicted to be targeted by HLA-A\*0201.** The  $\omega$  classes for each branch are shown in colours given in the legend, along with the  $\omega$  value for each class and the percentage of branches falling in each category. The percentage of models that support dN>dS are written above each branch.



**Figure 3**  
**Branch-by-branch selective pressure for regions predicted to be targeted by HLA-A\*6801.** The ω classes for each branch are shown in colours given in the legend, along with the ω value for each class and the percentage of branches falling in each category. The percentage of models that support dN>dS are written above each branch.



**Figure 4**  
**Branch-by-branch selective pressure for regions predicted to be targeted by HLA-B\*2705.** The  $\omega$  classes for each branch are shown in colours given in the legend, along with the  $\omega$  value for each class and the percentage of branches falling in each category. The percentage of models that support  $dN>dS$  are written above each branch.



**Table 1: HLA class I alleles with available crystal structure**

HLA allele	Allele frequency in African population
A*0201	16%
A*6801	10%
B*2705	1%
B*3501	5%
B*5301	5%
B8	5%

HLA alleles with solved crystal structures that are available for analysis in the PREDEP program, with allele frequencies as estimated in [42].

may have caused positive selection on only a few sites. Such selection is hard to detect because  $\omega$  is averaged over all sites of the sequence. Selection may also have been weak due to the fact that this is a rare HLA allele (1%) [42].

In the HLA A\*0201 dataset, both the *a priori* and the GABranch analyses inferred positive selection on the HIV-1 ancestral branch, with very high support for  $dN>dS$ . Of the alleles available for analysis in this study, HLA A\*0201 is the most frequent in African populations (see Table 1). It is also the most frequent HLA allele in Caucasian populations and many studies have been carried out to determine its effect on HIV disease progression [42]. Even though the allele recognizes immunodominant peptide regions of the HIV-1 sequence, it fails to exert strong selective pressure on some virus peptides [43]. Some studies have also shown that the outcome of an immune response does not only depend on the HLA molecule but also on the specific peptide sequences that are targeted [44-48]. Our results suggest that immune escape mutations that occurred for HLA A\*0201 mediated CTL responses may have been selected for in the period immediately following zoonosis. If these adaptations subse-

quently became fixed in the viral population they would no longer be under diversifying selection today.

HLA A\*6801 (another common allele in African populations) appears to have exerted strong selective pressure on the HIV-1 ancestral branch compared to the rest of the tree. High support (99.6% of the tested models) for  $\omega > 1$  was observed at the ancestral HIV branch. This allele has anchor residue motif restrictions that are shared within the HLA A3 supertype, the second most frequent super-type in the human population [49]. The HLA A\*6801 allele itself targets the Tat protein, which is expressed in the early stages of the HIV-1 lifecycle, and CTL responses to this protein cause a significant reduction in disease progression rate [50]. Escape mutations from the CTL immune response have also been identified within Tat at the population level, causing reduced viral load [51,52]. The virus may have adapted well to the A\*6801 responses early after the cross-species transmission event at sites that do not affect the replication of the virus. The recently observed association with a reduction in viral load indicates that there were also functionally important sites contained in A\*6801 epitopes - this would have made it difficult for these regions to adapt to the immune response.

## Conclusion

This is the first study that analyses HLA-associated selective pressure following the transmission from chimpanzee to human across all potential target sites of the HIV-1 genome. We identified regions of the HIV-1 sequence that were initially targeted by the CTL immune response immediately after the cross-species transmission of HIV-1 from chimpanzee to human using the chimpanzee consensus sequence. Of the six HLA alleles with crystal structures available for analysis, we found strong binding regions; this could imply successful immune responses *in vivo*, for HLAs A\*0201, A\*6801 and B\*2705. We determined the average extent of selective pressure exerted by each HLA allele along the branch leading to HIV-1 sequences. This branch represents the sequences that first encountered human immune response-directed selective pressure immediately following the zoonosis event. Our results suggest that HIV-1 adapted to CTL responses

**Table 2: Sequence data for the regions predicted to have potential HLA binding peptides**

HLA allele	Anchor residue motif <sup>1</sup>	Predicted HLA binding regions
A*0201	.[AILTVM]..... [AILTVM]	395 codons
A*6801	[AILTVM]..... [RK].	345 codons
B*2705	.[RK]..... [LFYRKHMI].	148 codons

<sup>1</sup> The anchor residue motifs were predicted from HLA-peptide structural conformations in PREDEP, residues in square brackets are the most preferred at the specific anchor site and the dots represent any other amino acid.

**Table 4: The best fitting models and model-averaged results obtained by the GABranch analysis**

HLA allele	Best fitting model: $\omega$ rate classes (number of branches) <sup>1</sup>	Model-averaged $\omega$ for HIV ancestral branch <sup>2</sup>	Model-averaged Prob(dN>dS) <sup>3</sup> for HIV ancestral branch
A*0201	1.14 (33), 0.62 (46)	1.14	0.996
A*6801	1.15 (42), 0.60 (31), 0.23(6)	1.15	0.996
B*2705	0.52 (79)	0.52	0.06

<sup>1</sup>For each rate class, the value of  $\omega$  is indicated followed by the number of branches (out of a total of 79) in brackets. <sup>2</sup>The  $\omega$  values are means over results from all the models. <sup>3</sup>Prob (dN>dS) is the fraction of models that show support for dN>dS.

**Table 3: BranchApriori  $\omega$  values for the HIV ancestral branch and the rest of the tree**

Allele	HIV ancestral branch $\omega$	Rest of the tree $\omega$	p-value
A*0201	2.5	1.2	0.08
A*6801	3.5	1.6	0.04
B*2705	1.0	1.1	0.65

directed by HLAs A\*6801 and A\*0201, which are amongst the most common HLA genotypes in African populations (Table 1). It is therefore likely that the virus was frequently exposed to selective pressure exerted by common immune responses during initial exposure to the human host following transmission of the virus from chimpanzees. As observed from the results, we did not find evidence for strong selective pressure exerted by the HLA B\*2705, which has extremely low frequencies in the African populations (Table 1) [53,54].

In this study we focussed specifically on epitopes that we infer were likely to have been present in the viral sequence that first infected humans. We propose that the selection we observe at these positions along the branch of the phylogenetic tree leading to all of the HIV-1 sequences reflects episodic selection to evade human cytotoxic immune responses. Episodic selection has been proposed to be an important aspect of cross-species pathogen transmission and, in fact, observed in a laboratory setting previously [14]. However, this is the first time, to our knowledge, that evidence has been presented of transient positive selection associated with human immune responses against unconstrained regions of the virus shortly after transmission to human.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

NKN performed the analysis, interpreted the results and wrote the manuscript. CS conceived and supervised the study and edited the manuscript. KS supervised and co-wrote the manuscript. All authors read and approved the final manuscript.

### Additional material

#### Additional file 1

*Regions of the HIV-1 sequence excluded from analysis. Regions of the HIV-1 genome with highly conserved synonymous sites under purifying selective pressure reported in our previous study [34]. Co-ordinates are adapted from the HXB2 numbering system. No conserved synonymous sites were found along the vif gene region.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1743-422X-6-164-S1.DOC>]

#### Additional file 2

*Model Averaged Branch dN/dS for HLA A\*0201. The statistical distribution of dN/dS values for the HLA A\*0201 binding regions along each branch of the tree, obtained via AIC-based model averaging. Branches with high model-averaged support for dN>dS are shown in bold. The HIV ancestral branch is Node 18.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1743-422X-6-164-S2.DOC>]

**Additional file 3**

**Model Averaged Branch dN/dS for HLA A\*6801.** The statistical distribution of dN/dS values for the HLA A\*6801 binding regions along each branch of the tree, obtained via AIC-based model averaging. Branches with high model-averaged support for dN>dS are shown in bold. The HIV ancestral branch is Node 18.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1743-422X-6-164-S3.DOC>]

**Additional file 4**

**Model Averaged Branch dN/dS for HLA B\*2705.** The statistical distribution of dN/dS values for the HLA B\*2705 binding regions along each branch of the tree, obtained via AIC-based model averaging. No model-averaged support for dN>dS was observed in any of the branches.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1743-422X-6-164-S4.DOC>]

**Acknowledgements**

This study was funded by the South African National Bioinformatics Network. NKN was supported by a training grant under the Stanford-South Africa Biomedical Informatics Training Program, which is supported by the Fogarty International Center, part of the National Institutes of Health (grant no. 5D43 TW006993).

**References**

- Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, Michael SF, Cummins LB, Arthur LO, Peeters M, Shaw GM, Sharp PM, Hahn BH: **Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes.** *Nature* 1999, **397**:436-441.
- Zhu T, Korber BT, Nahmias AJ, Hooper E, Sharp PM, Ho DD: **An African HIV-1 sequence from 1959 and implications for the origin of the epidemic.** *Nature* 1998, **391**:594-597.
- Korber B, Muldoon M, Theiler J, Gao F, Gupta R, Lapedes A, Hahn BH, Wolinsky S, Bhattacharya T: **Timing the ancestor of the HIV-1 pandemic strains.** *Science* 2000, **288**:1789-1796.
- Sharp P, Bailes E, Robertson D, Gao F, Hahn B: **Origins and evolution of AIDS viruses.** *Biol Bull* 1999, **196**:338-342.
- Keele BF, Jones JH, Terio KA, Estes JD, Rudicell RS, Wilson ML, Li Y, Learn GH, Beasley TM, Schumacher-Stankey J, Wroblewski E, Mosser A, Raphael J, Kamanya S, Lonsdorf EV, Travis DA, Mlengeya T, Kinsel MJ, Else JG, Silvestri G, Goodall J, Sharp PM, Shaw GM, Pusey AE, Hahn BH: **Increased mortality and AIDS-like immunopathology in wild chimpanzees infected with SIVcpz.** *Nature* 2009, **460**:515-519.
- Silvestri G, Paiardini M, Pandrea I, Lederman M, Sodora D: **Understanding the benign nature of SIV infection in natural hosts.** *J Clin Invest* 2007, **117**:3148-3154.
- Pandrea I, Ribeiro RM, Gautam R, Gauffin T, Pattison M, Barnes M, Monjure C, Stoullig C, Dufour J, Cyprian W, Silvestri G, Miller MD, Perelson AS, Apetrei C: **Simian immunodeficiency virus SIVagm dynamics in African green monkeys.** *J Virol* 2008, **82**:3713-3724.
- Silvestri G: **Immunity in natural SIV infections.** *Journal of INTERNAL MEDICINE* 2008, **10**:1365-2796.
- Musey L, Hughes J, Schacker T, Shea T, Corey L, McElrath MJ: **Cytotoxic-T-cell responses, viral load, and disease progression in early human immunodeficiency virus type I infection.** *N Engl J Med* 1997, **337**:1267-1274.
- Lemey P, Kosakovsky Pond SL, Drummond AJ, Pybus OG, Shapiro B, Barroso H, Taveira N, Rambaut A: **Synonymous substitution rates predict HIV disease progression as a result of underlying replication dynamics.** *PLoS Comput Biol* 2007, **3**:e29.
- Shearer WT, Quinn TC, LaRussa P, Lew JF, Mofenson L, Almy S, Rich K, Handelsman E, Diaz C, Pagano M, Smeriglio V, Kalish LA: **Viral load and disease progression in infants infected with human immunodeficiency virus type I. Women and Infants Transmission Study Group.** *N Engl J Med* 1997, **336**:1337-1342.
- Saksela K, Stevens C, Rubinstein P, Baltimore D: **Human immunodeficiency virus type I mRNA expression in peripheral blood cells predicts disease progression independently of the numbers of CD4+ lymphocytes.** *Proc Natl Acad Sci USA* 1994, **91**:1104-1108.
- Michael NL, Mo T, Merzouki A, O'Shaughnessy M, Oster C, Burke DS, Redfield RR, Birx DL, Cassol SA: **Human immunodeficiency virus type I cellular RNA load and splicing patterns predict disease progression in a longitudinally studied cohort.** *J Virol* 1995, **69**:1868-1877.
- Baric RS, Yount B, Hensley L, Peel SA, Chen W: **Episodic evolution mediates interspecies transfer of a murine coronavirus.** *J Virol* 1997, **71**:1946-1955.
- Wain LV, Bailes E, Bibollet-Ruche F, Decker JM, Keele BF, Van Heuverswyn F, Li Y, Takehisa J, Ngole EM, Shaw GM, Peeters M, Hahn BH, Sharp PM: **Adaptation of HIV-1 to its human host.** *Mol Biol Evol* 2007, **24**:1853-1860.
- Soares AE, Soares MA, Schrago CG: **Positive selection on HIV accessory proteins and the analysis of molecular adaptation after interspecies transmission.** *J Mol Evol* 2008, **66**:598-604.
- Choisy M, Woelk CH, Guegan JF, Robertson DL: **Comparative study of adaptive molecular evolution in different human immunodeficiency virus groups and subtypes.** *J Virol* 2004, **78**:1962-1970.
- Muller V, De Boer RJ: **The integration hypothesis: an evolutionary pathway to benign SIV infection.** *PLoS Pathog* 2006, **2**:e15.
- Bibollet-Ruche F, McKinney BA, Duverger A, Wagner FH, Ansari AA, Kutsch O: **The quality of chimpanzee T-cell activation and simian immunodeficiency virus/human immunodeficiency virus susceptibility achieved via antibody-mediated T-cell receptor/CD3 stimulation is a function of the anti-CD3 antibody isotype.** *J Virol* 2008, **82**:10271-10278.
- Rutjens E, Balla-Jhaghoorsingh S, Verschoor E, Bogers W, Koopman G, Heeney J: **Lentivirus infections and mechanisms of disease resistance in chimpanzees.** *Front Biosci* 2003, **8**:d1134-d1145.
- Madden DR, Gorga JC, Strominger JL, Wiley DC: **The three-dimensional structure of HLA-B27 at 2.1 Å resolution suggests a general mechanism for tight peptide binding to MHC.** *Cell* 1992, **70**:1035-1048.
- Rammensee H, Friede T, Stevanović S: **MHC ligands and peptide motifs: first listing.** *Immunogenetics* 1995, **41**:178-228.
- Saper MA, Bjorkman PJ, Wiley DC: **Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 Å resolution.** *J Mol Biol* 1991, **219**:277-319.
- Ohno S: **How cytotoxic T cells manage to discriminate non-self from self at the nonapeptide level.** *Proc Natl Acad Sci USA* 1992, **89**:4643-4647.
- DiBriño M, Parker KC, Shiloach J, Turner RV, Tsuchida T, Garfield M, Biddison WE, Coligan JE: **Endogenous peptides with distinct amino acid anchor residue motifs bind to HLA-A1 and HLA-B8.** *J Immunol* 1994, **152**:620-631.
- Rovero P, Riganelli D, Fruci D, Viganò S, Pegoraro S, Revoltella R, Greco G, Butler R, Clementi S, Tanigaki N: **The importance of secondary anchor residue motifs of HLA class I proteins: a chemometric approach.** *Mol Immunol* 1994, **31**:549-554.
- Sette A, Vitiello A, Reheman B, Fowler P, Nayarsina R, Kast WM, Melief CJ, Oseroff C, Yuan L, Ruppert J, Sidney J, del Guercio MF, Southwood S, Kubo RT, Chesnut RW, Grey HM, Chisari FV: **The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes.** *J Immunol* 1994, **153**:5586-5592.
- Carrington M, O'Brien SJ: **The influence of HLA genotype on AIDS.** *Annu Rev Med* 2003, **54**:535-551.
- Hendel N, Caillat-Zucman S, Leubaneac H, Carrington M, O'Brien S, Andrieu JM, Schachter F, Zagury D, Rappaport J, Winkler C, Nelson GW, Zagury JF: **New class I and II HLA alleles strongly associated with opposite patterns of progression to AIDS.** *J Immunol* 1999, **162**:6942-6946.
- Frahm N, Kiepiela P, Adams S, Linde CH, Hewitt HS, Sango K, Feeney ME, Addo MM, Lichtenfeld M, Lahaie MP, Pae E, Wurcel AG, Roach T, St John MA, Altfield M, Marincola FM, Moore C, Mallal S, Carrington M, Heckerman D, Allen TM, Mullins JI, Korber BT, Goulder PJ, Walker BD, Brander C: **Control of human immunodeficiency**

- virus replication by cytotoxic T lymphocytes targeting sub-dominant epitopes.** *Nat Immunol* 2006, **7**:173-178.
31. Honeyborne I, Prendergast A, Pereyra F, Leslie A, Crawford H, Payne R, Reddy S, Bishop K, Moodley E, Nair K, van der SM, McCarthy N, Rousseau CM, Addo M, Mullins JI, Brander C, Kiepiela P, Walker BD, Goulder PJ: **Control of human immunodeficiency virus type I is associated with HLA-B\*13 and targeting of multiple gag-specific CD8+ T-cell epitopes.** *J Virol* 2007, **81**:3667-3672.
  32. Delpont W, Scheffler K, Seoighe C: **Frequent toggling between alternative amino acids is driven by selection in HIV-1.** *PLoS Pathog* 2008, **4**:e1000242.
  33. Leitner T, Foley B, Hahn B, Marx P, McCutchan F, Mellors J, Wolinsky S, Korber B: **HIV Sequence Compendium.** *Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, NM, LA-UR* 2005. 06-0680
  34. Ngandu NK, Scheffler K, Moore P, Woodman Z, Martin D, Seoighe C: **Extensive purifying selection acting on synonymous sites in HIV-1 Group M sequences.** *Virology* 2008, **5**:160.
  35. Kosakovsky Pond SL, Frost SD, Muse SV: **HyPhy: hypothesis testing using phylogenies.** *Bioinformatics* 2005, **21**:676-679.
  36. Altuvia Y, Margalit H: **A structure-based approach for prediction of MHC-binding peptides.** *Methods* 2004, **34**:454-459.
  37. Nielsen R, Yang Z: **Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene.** *Genetics* 1998, **148**:929-936.
  38. Kosakovsky Pond SL, Frost SD: **A genetic algorithm approach to detecting lineage-specific variation in selection pressure.** *Mol Biol Evol* 2005, **22**:478-485.
  39. Yang Z: **Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution.** *Mol Biol Evol* 1998, **15**:568-573.
  40. McNeil AJ, Yap PL, Gore SM, Brettell RP, McColl M, Wyld R, Davidson S, Weightman R, Richardson AM, Robertson JR: **Association of HLA types A1-B8-DR3 and B27 with rapid and slow progression of HIV disease.** *QJM* 1996, **89**:177-185.
  41. Ross HA, Rodrigo AG: **Immune-mediated positive selection drives human immunodeficiency virus type I molecular variation and predicts disease duration.** *J Virol* 2002, **76**:11715-11720.
  42. Marsh S, Parham P, Barber L: **The HLA Facts Book.** Academic Press; 2000:1.
  43. Brander C, Hartman KE, Trocha AK, Jones NG, Johnson RP, Korber B, Wentworth P, Buchbinder SP, Wolinsky S, Walker BD, Kalams SA: **Lack of strong immune selection pressure by the immunodominant, HLA-A\*0201-restricted cytotoxic T lymphocyte response in chronic human immunodeficiency virus-1 infection.** *J Clin Invest* 1998, **101**:2559-2566.
  44. Pereyra F, Addo MM, Kaufmann DE, Liu Y, Miura T, Rathod A, Baker B, Trocha A, Rosenberg R, Mackey E, Ueda P, Lu Z, Cohen D, Wrinn T, Petropoulos CJ, Rosenberg ES, Walker BD: **Genetic and immunologic heterogeneity among persons who control HIV infection in the absence of therapy.** *J Infect Dis* 2008, **197**:563-571.
  45. Borghans JA, Molgaard A, De Boer RJ, Kesmir C: **HLA alleles associated with slow progression to AIDS truly prefer to present HIV-1 p24.** *PLoS ONE* 2007, **2**:e920.
  46. Nelson GV, Kaslow R, Mann DL: **Frequency of HLA allele-specific peptide motifs in HIV-1 proteins correlates with the allele's association with relative rates of disease progression after HIV-1 infection.** *Proc Natl Acad Sci USA* 1997, **94**:9802-9807.
  47. Maurer K, Harrer EG, Goldwisch A, Eismann K, Bergmann S, Schmitt-Haendle M, Spriewald B, Mueller SM, Harrer T: **Role of cytotoxic T-lymphocyte-mediated immune selection in a dominant human leukocyte antigen-B8-restricted cytotoxic T-lymphocyte epitope in Nef.** *J Acquir Immune Defic Syndr* 2008, **48**:133-141.
  48. Miura T, Brockman MA, Schneidewind A, Lobritz M, Pereyra F, Rathod A, Block BL, Brumme ZL, Brumme CJ, Baker B, Rothchild AC, Li B, Trocha A, Cutrell E, Frahm N, Brander C, Toth I, Arts EJ, Allen TM, Walker BD: **HLA-B57/B\*5801 HIV-1 Elite Controllers Select for Rare Gag Variants Associated with Reduced Viral Replication Capacity and Strong CTL Recognition.** *J Virol* 2008, **10**:2265-2308.
  49. Sette A, Sidney J: **Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism.** *Immunogenetics* 1999, **50**:201-212.
  50. van Baalen CA, Pontesilli O, Huisman RC, Geretti AM, Klein MR, de Wolf F, Miedema F, Gruters RA, Osterhaus AD: **Human immunodeficiency virus type I Rev- and Tat-specific cytotoxic T lymphocyte frequencies inversely correlate with rapid progression to AIDS.** *J Gen Virol* 1997, **78**(Pt 8):1913-1918.
  51. Cao J, McNevin J, Malhotra U, McElrath MJ: **Evolution of CD8+ T cell immunity and viral escape following acute HIV-1 infection.** *J Immunol* 2003, **171**:3837-3846.
  52. Allen TM, O'Connor DH, Jing P, Dzuris JL, Mothe BR, Vogel TU, Dunphy E, Liebl ME, Emerson C, Wilson N, Kunstman KJ, Wang X, Allison DB, Hughes AL, Desrosiers RC, Altman JD, Wolinsky SM, Sette A, Watkins DI: **Tat-specific cytotoxic T lymphocytes select for SIV escape variants during resolution of primary viraemia.** *Nature* 2000, **407**:386-390.
  53. Trachtenberg E, Korber B, Sollars C, Kepler TB, Hraber PT, Hayes E, Funkhouser R, Fugate M, Theiler J, Hsu YS, Kunstman K, Wu S, Phair J, Erlich H, Wolinsky S: **Advantage of rare HLA supertype in HIV disease progression.** *Nat Med* 2003, **9**:928-935.
  54. Solberg OD, Mack SJ, Lancaster AK, Single RM, Tsai Y, Sanchez-Mazas A, Thomson G: **Balancing selection and heterogeneity across the classical human leukocyte antigen loci: a meta-analytic review of 497 population studies.** *Hum Immunol* 2008, **69**:443-464.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

