Rapid, complex adaptation of transmitted HIV-1 full-length genomes in subtype C-infected individuals with differing disease progression

Melissa-Rose Abrahams[a], Florette K. Treurnicht[a], Nobubelo K. Ngandu[a], Sarah A. Goodier[a], Jinny C. Marais[a], Helba Bredell[a], Ruwayhida Thebus[a], Debra de Assis Rosa[b], Koleka Mlisana[c,d], Cathal Seoighe[e], Salim Abdool Karim[c], Clive M. Gray[b] and Carolyn Williamson[a]

Objective(s): There is limited information on full-length genome sequences and the early evolution of transmitted HIV-1 subtype C viruses, which constitute the majority of viruses spread in Africa. The purpose of this study was to characterize the earliest changes across the genome of subtype C viruses following transmission, to better understand early control of viremia.

Design: We derived the near full-length genome sequence responsible for clinical infection from five HIV subtype C-infected individuals with different disease progression profiles and tracked adaptation to immune responses in the first 6 months of infection.

Methods: Near full-length genomes were generated by single genome amplification and direct sequencing. Sequences were analyzed for amino acid mutations associated with cytotoxic T lymphocyte (CTL) or antibody-mediated immune pressure, and for reversion.

Results: Fifty-five sequence changes associated with adaptation to the new host were identified, with 38% attributed to CTL pressure, 35% to antibody pressure, 16% to reversions and the remainder were unclassified. Mutations in CTL epitopes were most frequent in the first 5 weeks of infection, with the frequency declining over time with the decline in viral load. CTL escape predominantly occurred in *nef*, followed by *pol* and *env*. Shuffling/toggling of mutations was identified in 81% of CTL epitopes, with only 7% reaching fixation within the 6-month period.

Conclusion: There was rapid virus adaptation following transmission, predominantly driven by CTL pressure, with most changes occurring during high viremia. Rapid escape and complex escape pathways provide further challenges for vaccine protection.

AIDS 2013, **27**:507–518

Keywords: acute infection, Africa, cytotoxic T-lymphocytes, genome, HIV-1, progression

[a]Division of Medical Virology, Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, [b]AIDS Research Unit: Immunology, National Institute of Communicable Diseases, Johannesburg, [c]Centre for the AIDS Program of Research in South Africa, [d]Department of Medical Microbiology, University of KwaZulu Natal, Durban, South Africa, and [e]School of Mathematics, Statistics and Applied Mathematics, National University of Ireland, Galway, Ireland.

Correspondence to Carolyn Williamson, Division of Medical Virology, Institute of Infectious Diseases and Molecular Medicine, Faculty of Health Sciences, University of Cape Town and National Health Laboratory Services, Observatory 7925, South Africa.
Tel: +27 21 406 6683; fax: +27 21 406 6682; e-mail: Carolyn.Williamson@uct.ac.za
Received: 17 July 2012; revised: 5 November 2012; accepted: 15 November 2012.

DOI:10.1097/QAD.0b013e32835cab64

Introduction

Early host selective pressures drive genetic diversification of the transmitted HIV and potentially influence the course of disease. In heterosexual infection, it is estimated that approximately 80% of individuals are infected with a single virus or virus-infected cell [1,2]. Vaccines that do not block the establishment of this initial infecting virus will need to target the early diversifying virus and thus an understanding of early viral evolution is important. The frequency and speed at which the transmitted virus changes in response to host immune pressures are of key interest to vaccine immunogen design as these provide insights into its strengths and vulnerabilities.

Recent methodologies using single genome amplification (SGA) applied to individuals with acute HIV-1 infection have enabled identification of the sequence of transmitted/founder (t/f) full-length viral genomes [1,3,4]. This approach uses mathematical modeling to derive the sequence of the virus(es) responsible for productive clinical infection, and was proven in the simian immunodeficiency virus (SIV) model wherein the derived t/f was identical, or differed by a few nucleotides, to the virus in the inocula [5].

Using this approach, early evolutionary changes following transmission in three subtype B infections from the USA have been mapped [3]. No longitudinal studies elucidating early evolution in full-length subtype C genomes have been reported despite subtype C being the dominant subtype both globally and in southern Africa where large vaccine and microbicide trials take place [6,7]. Studies in different population settings are essential given that differences in host genetics influence viral evolution [8].

The transmitted virus encounters immune selective pressures almost immediately following infection. Cytotoxic T lymphocyte (CTL) and neutralizing antibody pressure are the proven driving forces for viral diversification [9–16]. Evidence of CTL pressure on the viral genome has been identified in the very first weeks of subtype B and C infection [1–3,17–23], and CTL activity has been associated with control of viremia early in infection [24–26]. Most recently, a subtype B full-length genome study using mathematical modeling to determine the killing rate of CTL during acute viremia suggested that these cells have a role in controlling peak viremia [21].

In this study, we investigated changes observed across the genomes of t/f subtype C viruses from five heterosexually infected women with differing disease progression profiles. We extrapolated the near full-length genome sequence of the t/f viruses and quantified genetic mutations associated with positive selection from humoral and cellular immune pressures over the first 6 months following infection.

Methods

Study participants

Samples were obtained from the Centre for the AIDS Program of Research in South Africa (CAPRISA) 002 Acute Infection cohort (Durban, South Africa) [27]. Date of infection was estimated as the midpoint between last negative and first positive HIV antibody test and as 14 days prior for individuals who were RNA positive/antibody negative. Human leukocyte antigen (HLA) A, B and C types were determined using four-digit high-resolution HLA typing as described [28]. The study was approved by the Universities of Cape Town, Witwatersrand and KwaZulu Natal.

PCR amplification and sequencing

RNA was extracted from 200–400 μl plasma using the Qiagen Viral RNA Mini Kit (Qiagen, Valencia, California, USA) and reverse transcribed to cDNA using superscript III reverse transcriptase and Oligo$(dT)_{20}$ (Invitrogen, GmbH, Karlsruhe, Germany). SGA [4] and sequencing of near full-length genome amplicons was done using Expand Long Template Taq (Roche Diagnostics, Rotkreuz, Switzerland) as reported [4] with primers described by Rousseau *et al.* [29] optimized for subtype C. Sequences less than 6000 bp were excluded. Salazar-Gonzalez *et al.* [3] attributed one to five ambiguities within a genome obtained at less than 20% PCR positivity predominantly to PCR Taq error. We accepted amplicons obtained at less than 66% positivity with up to six ambiguities. Targeted epitope sequencing was done following gene-specific limiting dilution PCR. *Gag* and *nef* clones were generated using limiting dilution PCR and the pGEM-T Easy system as described [30]. *Env* SGA was described previously [4]. All products were directly sequenced using the ABI 3000 genetic analyzer (Applied Biosystems, Foster City, California, USA) and BigDye terminator reagents.

Sequence analysis

Analyses performed were sequence alignments, amino acid identity and frequency plots and consensus sequence derivation (BioEdit version 7.0.8.0 [31]); subtyping (REGA HIV Subtyping Tool; http://dbpartners.stanford.edu/RegaSubtyping/); phylogenetic and pairwise DNA distance analyses (Mega 4 [32]); Highlighter plots (http://www.hiv.lanl.gov/content/sequence/HIGHLIGHT/highlighter); CTL epitope prediction (Epitope Location Finder (ELF) (http://www.hiv.lanl.gov/content/sequence/ELF/epitope_analyzer) and NetMHCpan 2.2 (http://www.cbs.dtu.dk/services/NetMHCpan)[33]); Hypermut detection of APOBEC hypermutation (http://www.hiv.lanl.gov/content/sequence/HYPERMUT/hypermnut.html); Shannon entropies (Entropy One; http://www.hiv.lanl.gov/content/sequence/ENTROPY/entropy_one) and mapping of functionally/structurally relevant genome regions (http://www.hiv.lanl.gov/content/sequence/HIV/MAP/landmark). A

high average Shannon entropy score was taken as more than 0.25 as described by Bansal *et al.* [34] for variable HIV proteins. Where an optimal epitope (9–11-mer) has not yet been described, the entropy of an 18-mer peptide encompassing the mutating region was used.

Time to most recent common ancestor (MRCA) was determined using Bayesian Evolutionary Analysis Sampling Trees (BEAST) v1.4.7 [35,36] as described previously [1,2]. A relaxed (uncorrelated exponential) molecular clock with general time-reversible substitution model, mean of 2.16×10^{-5} substitutions per site per generation with rates unlinked across codon sites [37] and gamma distribution with four categories and a proportion of invariant sites was used.

Known HLA class I restricted epitopes or class I HLA-associated polymorphisms were identified using the Los Alamos HIV Molecular Immunology 2008 Compendium (http://www.hiv.lanl.gov/content/immunology/compendium.html) and Matthews *et al.* [38]. Subtype C database alignments were obtained from the Los Alamos HIV database (http://www.hiv.lanl.gov/components/sequence/HIV/).

Positive selection and statistical analysis

Nonsynonymous to synonymous substitution (dN/dS) rate ratios per codon site were estimated using the MG94xHKY85 codon model [39]. We allowed dS to vary across codon sites and employed the Dual model which takes into account that dS may vary independently of dN [40]. Models were implemented within HyPhy [41] and ensured that correct phylogenetic relationships were used for regions separated by recombination breakpoints [42].

Categorical statistical tests were carried out using the Fisher's exact two-tailed test (http://www.graphpad.com/quickcalcs).

Results

This study characterized virus evolution in five subtype C-infected women recruited 2–5 weeks following transmission. These women were selected based on infection with a single transmitted/founder virus [2] and clinical disease progression profile (Table 1). One woman was classified as a viremic controller (CAP45; viral loads consistently <2000 copies/ml and CD4 cell counts >350 cells/μl in the absence of antiretroviral therapy) [43], two as rapid progressors (CAP63 and CAP210; viral loads >100 000 copies/ml and CD4 cell counts <350 cells/μl on consecutive visits within the first year of infection) [23,44], and two as intermediate progressors (CAP85 and CAP239; not fitting either controller or rapid status) (Fig. 1). A total of 112 near full-length genomes were generated, with an average of nine at screening per enrolment (2–5 weeks after infection), six at 3 months (11–13 weeks after infection) and nine at 6 months (22–29 weeks after infection). No sequences could be generated at 6 months for viremic controller CAP45 due to low viral loads. Additional sequences (half-genome, SGA and clonal) were generated from various time points ranging from 2 to 117 weeks after infection (Table 1).

Derivation of transmitted/founder virus sequences

All sequences were classified as subtype C along the entire length of the genome. Mean intraparticipant DNA distances ranged from 0.008 to 0.25% (median 0.03%) at the first time point and mean number of days since MRCA ranged from 18 to 53 days (Table 1), indicating limited sequence diversification since transmission (see Fig. S1, Supplemental Digital Content 1, http://links.lww.com/QAD/A282, illustrating intraparticipant sequence diversity in a Neighbour-Joining tree). The t/f sequences were defined as the consensus of sequences from the earliest time point where all genes had an intact open reading frame and no ambiguities [3]. Although CAP63 and CAP85 were classified as infected with a single t/f variant based on *env* [2], we identified a minor early variant in each (not detected at later time points) which, in both cases, differed from the derived t/f at five nucleotide positions, suggesting that these individuals may each have been infected with two very closely related variants.

The majority of early genetic changes are due to cytotoxic T lymphocyte pressure or reversion

Using longitudinal near full-length genome and *env* screening/enrolment SGA sequences, viral evolution from the t/f was analyzed for evidence of immune escape or reversion. Substitutions from high-frequency/consensus to lower frequency/nonconsensus amino acids within or adjacent to known class I HLA-restricted epitopes, or corresponding to known HLA-associated polymorphisms, were classified as CTL pressure [16,19,45,46]. Mutations within the hypervariable loops and potential N-linked glycosylation sites (PNGSs) in *env* were classified as antibody pressure [9]. Mutations from low/nonconsensus to higher frequency/consensus subtype C amino acids within CTL epitopes not restricted by the participants HLA were classified as reversion of transmitted CTL escape mutations [16,45,46]. In addition, clustered mutations within amino acid 9-mers, previously reported to be associated with immune selection [18], or single amino acid mutations persisting to fixation and corresponding to sites under positive selection were identified as putative immune escape.

In viruses from the five women, immune pressure was identified in 55 genome regions (see Figs S2–S6, Supplemental Digital Content 2, http://links.lww.com/QAD/A282, which illustrate genome regions

Table 1. Demographic and sequencing information for five CAPRISA participants.

Participant ID	Age	Disease progression	Fiebig stage[a] at first sequenced time point	BEAST mean no. of days since MRCA (95% CI)[b]	HLA type	Sample date	Weeks postinfection	Whole (half) genome sequences	Subgenomic clone/SGA sequences
CAP45	41	Slow	I/II	18 (4–35)	A*23:01, 29:02	20 April 2005	2	3	16
					B*15:10, 45:01	11 May 2005	5	6	1
					Cw*06:02, 16:01	7 June 2005	9	3	8
						28 June 2005	12	1	7
						27 July 2005	16	3	
						4 April 2006	52		
						4 July 2006	65		
CAP63	32	Rapid	III	30 (10–53)	A*02:01, 23:01	6 January 2005	2	11	19
					B*45:01	26 January 2005	5	7	
					Cw*04:01, 16:01	9 March 2005	11	10	26
						13 July 2005	29	5	
						7 September 2005	37		
CAP85	24	Intermediate	V	53 (20–101)	A*30:02	22 June 2005	5	8	21
					B*08:01, 45:01	18 August 2013	13	9	
					Cw*07:01, 16:01	7 December 2005	29	7	11
						16 February 2006	39		1
						10 May 2006	51	2	9
						7 June 2006	55		9
						7 December 2006	81		1
						6 June 2007	107		7
CAP210	43	Rapid	I/II	11 (3–25)	A*68:02	3 May 2005	2	9	21
					B*15:10	13 June 2005	12	7	
					Cw*03:04	21 September 2005	22		8
						19 October 2005	26	11	3
						23 November 2005	31		3
CAP239	44	Intermediate	V[c]	34 (11–58)	A*01:23, 29:02	19 July 2005	2	2 (3)	21
					B*42:01, 58:01	10 August 2005	5	8	75
					Cw*06:02, 17:01	17 August 2005	6		19
						21 September 2005	11	2	
						7 December 2005	22	9	9
						14 June 2006	49	0 (4)	
						11 January 2007	79		9
						4 October 2007	117	3	

BEAST, Bayesian Evolutionary Analysis Sampling Trees; HLA, human leukocyte antigen; MRCA, most recent common ancestor; SGA, single genome amplification.
[a]Fiebig stage I/II [63] are HIV RNA positive but antibody negative, III is ELISA positive but nonreactive western blot and V is reactive western blot without p31 band.
[b]Bayesian Evolutionary Analysis Sampling Trees estimated period of infection to observed *env* diversity from a single infecting strain.
[c]Fiebig stage determined for 5 weeks postinfection

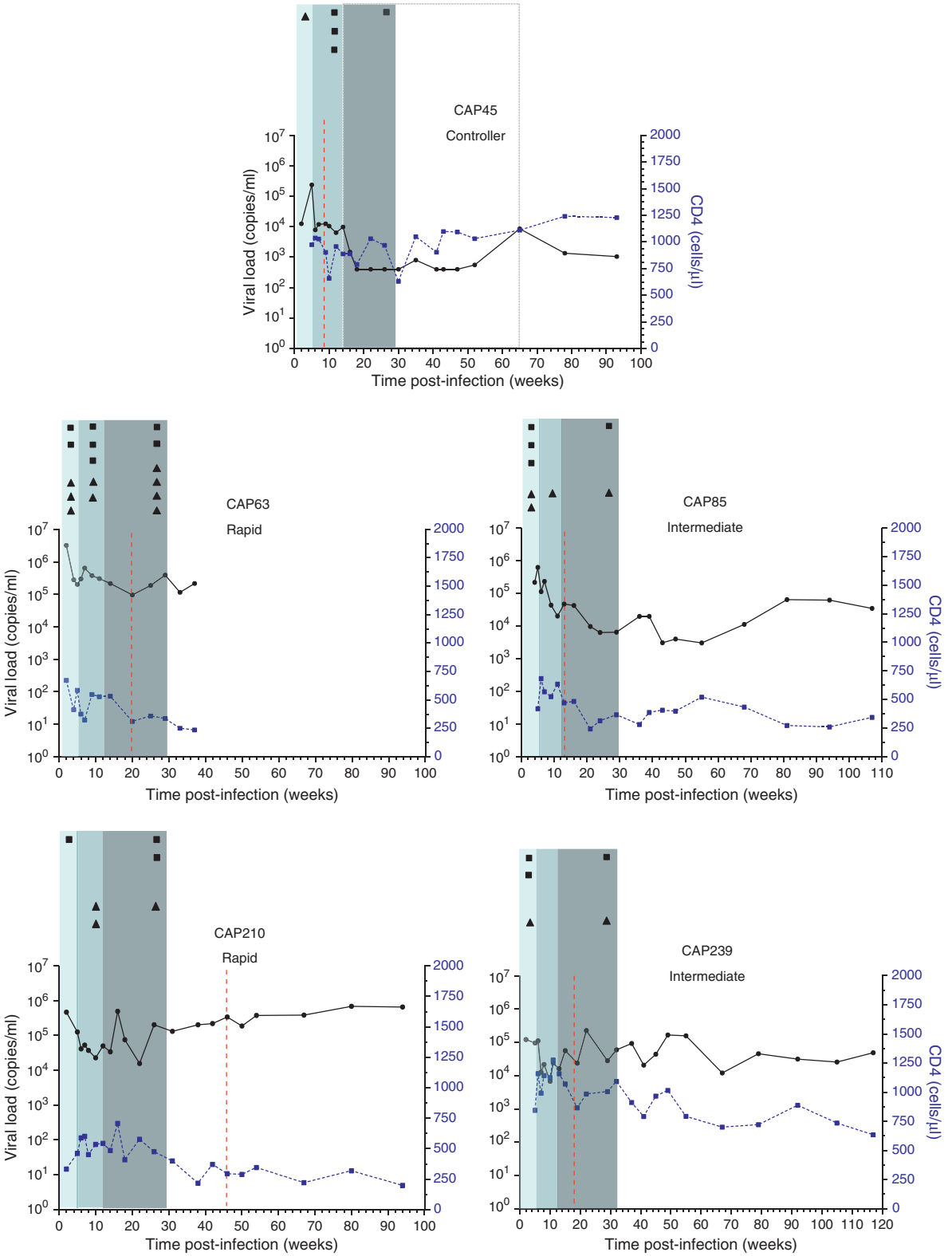

Fig. 1. Clinical and immune selection profiles for each of the five women illustrated with viral load and CD4 cell count graphs. The number of epitopes/genome regions identified as under putative cytotoxic T-lymphocyte pressure (black squares) and antibody pressure (black triangles) are illustrated for approximate periods after infection of 0–5 weeks (pale blue shaded region), 5–12 weeks (darker blue shaded region) and 12–29 weeks (gray shaded region). The dashed gray box is used for participant CAP45 as the black square represents an epitope which may have mutated anywhere between 12 and 65 weeks after infection. Time of first detection of autologous neutralizing antibodies is indicated by dashed red lines.

under immune pressure): 21 were classified as under CTL pressure (Table 2), 19 under antibody pressure (Table 3), nine as reversion (see Table S1, Supplemental Digital Content 3, http://links.lww.com/QAD/A282, which tabulates genome regions undergoing reversion) and further six regions contained mutations (clustered within amino acid 9-mers or single persisting to fixation in sites under positive selection) which did not conform to criteria described for CTL or antibody pressure or reversion (Table 3). Identification of regions under immune pressure was supported by selection analysis (for which *gag*, *env* and *nef* supplemental sequences were included), which found a total of 55 sites under positive selection (dN/dS >1) of which 84% (46/55) were situated within genome regions identified as under immune pressure (see Tables 2 and 3; Table S1, Supplemental Digital Content 3 and Figs S2–S6, Supplemental Digital Content 2, http://links.lww.com/QAD/A282, which illustrate positive selection sites corresponding to genome regions under immune pressure). The majority of positively selected sites were identified in *env* ($n = 30$), followed by *gag* ($n = 9$), *pol* ($n = 7$), *nef* ($n = 5$), *rev* ($n = 3$) and *tat* ($n = 1$). APOBEC-mediated G to A hypermutation was identified in 24% (13/55) of regions under immune selection.

Of the mutating genome regions associated with CTL pressure, 19 spanned known CTL epitopes, and two spanned predicted epitopes (Table 2). The highest frequency of CTL-driven mutations was located in *nef* (even when normalizing for amino acid length) ($n = 6$), followed by *pol* and *env* ($n = 4$), *gag* and *vif* ($n = 2$) and finally *rev*, *tat* and *vpr* ($n = 1$) (Fig. 2a). Reversion was identified most frequently in *vpu* ($n = 3$) followed by *gag*, *env* and *nef* ($n = 2$) (see Table S1, Supplemental Digital Content 3, http://links.lww.com/QAD/A282).

Cytotoxic T lymphocyte escape is most frequent in acute infection

We supplemented near full-length genomes with subgenomic (*gag*, *env*, *nef* and targeted epitope) sequence data to better elucidate timing of mutations associated with CTL pressure and escape. The majority of mutating epitopes was identified within the first 5 weeks of infection in structural genes *gag*, *pol*, *env* and *nef* (Figs 1 and 2b). The earliest of these was identified at 2 weeks postinfection in the *gag* HLA B*58 : 01 restricted TW10 epitope; and the *nef* HLA B*45 : 01 restricted EV11 epitope. The frequency of mutation associated with escape slowed over time with an initial 1.6 total escapes per week for the first 5 weeks of infection, to 0.9 escapes per week between 5 and 12 weeks postinfection, and 0.4 escapes per week between 12 and 29 weeks postinfection (Fig. 2b).

Of the 19 regions of *env* with changes in hypervariable loops and PNGSs, mutations in seven (37%) arose in the first 5 weeks of infection (Fig. 1 and Table 3). Since first detection of autologous neutralizing antibodies (nAb) for the women in this study ranged from 9 to 46 weeks [44,47] (Fig. 1), these early changes are unlikely to be the result of nAb pressure. However, in one instance, early mutating sites in V5 of participant CAP45 corresponded to sites where neutralization escape mutations (K460D and D462G) were identified later in infection (see Table 3) [44].

Finally, of the nine reversions identified, four occurred in the first 5 weeks of infection and the remaining between 13 and 29 weeks postinfection (data not shown).

Shuffling and toggling of amino acid mutations

Shifting of mutations between different positions within an epitope (shuffling) or between different amino acids at the same position (toggling) was observed in 81% (17/21) of mutating CTL epitopes (see Table 2 and Table S2, Supplemental Digital Content 4, http://links.lww.com/QAD/A282, illustrating shuffling and toggling in a Nef epitope). For four of the five participants, the number of mutant variants per CTL epitope increased over the 6 months of infection after which a plateau or decrease was seen owing in part to eventual fixation of escape mutants (Fig. 2c).

Twelve of the 17 CTL epitopes with shuffling/toggling mutations corresponded to subtype C database epitopes with high Shannon entropy (scores >0.25 [34]). Shuffling/toggling was however not more frequent in epitopes corresponding to high-entropy regions than in more conserved regions ($P = 0.25$). We investigated whether mutational shuffling/toggling was due to mutations arising in structurally or functionally essential sites, which when altered may abrogate efficient RNA folding or alter protein structure and function. We therefore compared the proportion of epitopes with shuffling/toggling in functionally or structurally relevant sites with the proportion in sites with no currently known functional or structural significance. We found no significant difference between the two categories ($P = 0.53$), even when including the six genome regions under unclassified immune pressure ($P = 0.28$).

Immune pressure, sequence diversification and disease progression

To determine whether rate of genetic diversification from t/f sequences differed with different disease progression profiles, we compared gene-specific tree lengths using sequences from the first 3 months of infection. No significant differences between rates of diversification were found (data not shown). Furthermore, no significant association between number of genome regions under immune pressure and viral load over time was identified (data not shown); possibly owing to low participant numbers.

Table 2. Putative cytotoxic T lymphocyte escape epitopes and polymorphisms.

Participant ID	ORF	Epitope/genome region sequence[a]	HXB2 position	Participant HLA association(s)[b]	Reference	Time of first AA change (range) (weeks)	High entropy epitope/peptide (LANL[c] subtype C database)	Shuffling/toggling of AA mutations
CAP45	Vif	**D**WHLGHGVSI	78–87	B*15:10	LANL database	12–65	No	No
	Rev	I**H**SISERI**L**	52–60	B*15:10	LANL database	5–12	Yes	Yes
	Tat	**N**CYCK**H**CSY	24–32	A*29:02	LANL database	5–12	Yes	Yes
	Nef	**EE**VGFPVRPQV	64–74	B*45:01	Matthews *et al.* [38]	5–9	No	Yes
CAP63	Pol	ALT**E**IC**EE**M	188–196	A*02:01	LANL database	5–11	Yes	Yes
	Pol	QLTE**AV**H**K**I	522–530	Predicted A*02:01		11–29	No	Yes
	Vpr	AL**IRI**LQQL	59–67	A*02:01	LANL database	5–11	Yes	Yes
	Gp41	SWSNKSE**E**DIW**G**NMTWMQ	102–119	A*23:01/Cw*04:01	LANL database	11–29	Yes	Yes
	Gp41	L**LD**SI**A**I**TV**	303–311	A*02:01	LANL database	2–4	Yes	Yes
	Nef	ALT**SS**NT**AA**	42–50	A*02:01	LANL database	5–11	Yes	Yes
	Nef	**EE**VGFPVRPQV	64–74	B*45:01	Matthews *et al.* [38]	0–2	No	Yes
CAP85	Pol	K**A**GYVTD**R**GRQKV**VS**LTE	609–626	B*08:01	Matthews *et al.* [38]	0–5	Yes	Yes
	Gp41	**RY**L**GS**L**V**Q**Y**	283–291	A*30:02	LANL database	0–5	Yes	Yes
	Nef	**KE**VGFP**V**RPQV	64–74	B*45:01	Matthews *et al.* [38]	0–5	No	Yes
	Nef	**Y**FPDWQNY	120–127	A*30:02	LANL database	13–29	No	No
CAP210	Gag	VHQ**A**ISPRTL	143–152	B*15:10	Matthews *et al.* [38]	12–16	No	No
	Vif	**D**WHLGHG**V**SI	78–87	B*15:10	LANL database	5–12	No	Yes
	Gp41	**EA**TDR**I**L**EL**	313–321	Predicted A*68:02		2–5	Yes	Yes
CAP239	Gag	TS**TL**QEQV**A**W	240–249	B*58:01	Matthews *et al.* [38]	0–2	No	Yes
	Pol	IVLP**EKE**SW	399–407	B*58:01	Matthews *et al.* [38]	2–5	Yes	Yes
	Nef	K**AA**V**V**DLSFF	82–90	B*58:01	Matthews *et al.* [38]	11–22	Yes	No

[a]Bold amino acids indicate sites undergoing mutation.
[b]Predicted epitopes obtained using Net MHCPan2.0 (www.cbs.dtu.dk/services/NetMHCPan).
[c]Los Alamos National Laboratory (LANL) database (www.hiv.lanl.gov) HIV Molecular Immunology 2008 compendium used.

Table 3. Genome regions under putative antibody-mediated or unclassified immune pressure.

Participant ID	ORF	Genome region sequence[a] (18–20 mer)	HXB2 position	Putative immune selective pressure	Time of first mutation in hypervariable loop or PNGS (range) (weeks)
CAP45	Gp120	LTRDGG**KTDR**N**DT**EIFRP	454–470	Antibody	0–2
CAP63	Gp120	QE**I**VLE**N**V**I**ENFNMWKND	82–99	Antibody	5–11
	Gp120	LTPLCVTLNCANA**N**I**T**KN	122–139	Antibody	11–29
	Gp120	**MI**GEI**K**NCSFN**ATTELR**D	147–167	Antibody	2–5
	Gp120	L**NNNRSNENS**YILINC**N**S	184–198	Antibody	0–2
	Gp120	IVH**F**NQSV**K**IVC**A**RPHNN	285–302	Antibody	11–29
	Gp120	IRQA**H**CNIS**KTQ**W**NTT**L**E**	326–343	Antibody	11–29
	Gp120	**FNSTYMPNGIHIPNGASEVIT**	396–415	Antibody	2–5
	Gp41	**L**WS**S**WFN**I**S**H**WL**W**YI**RI**FI	158–147	Antibody	11–29
	Gp41	IE**EE**GG**EQ**DN**S**RS**I**RLVS	222–239	Antibody	5–11
CAP85	Gp120	D**IVPLNNDIGNYSEY**RLI	180–194	Antibody	5–13
	Gp120	IVHLN**HS**V**K**IVCTRPGNN	285–302	Antibody	0–5
	Gp120	IRQAH**C**NIS**KA**EWN**NT**L**E**	326–343	Antibody	13–29
	Gp120	**GSSTT**T**NGSSP**ITLPCRI	404–420	Antibody	0–5
	Gp120	RPGGG**DM**K**D**NWRSELYKY	469–486	Unclassified	0–5
	Nef	GVGAASQDLGK**YG**ALT**SS**	29–46	Unclassified	0–5
CAP210	Pol	FFR**E**NLAFP**EGE**AR**ELPS**	1–18	Unclassified	12–16
	Gp120	ICSFNATTELRDKKKKEY	156–173	Antibody	5–12
	Gp120	FNSTHN**S**T**D**S**TVNSTDS**T	391–409	Antibody	5–12
	Gp120	ITCIS**NIT**GLLLTRDGGE	443–460	Antibody	22–26
	Nef	SLH**G**MEDTE**R**EVLQWKFD	169–186	Unclassified	5–12
CAP239	Gag	SNPS**GP**K**RPI**KC**F**NCGRE	382–399	Unclassified	2–5
	Rev	GRPA**EPV**P**F**QLPPIERLH	65–82	Unclassified	2–5
	Gp120	DIIIRSQ**NIL**D**N**TK**TIIV**	269–286	Antibody	2–5
	Gp120	GLLLT**W**DGGDS**KE**N**KT**R**H**	451–467	Antibody	11–22

PNGS, potential N-linked glycosylation site.
[a]Bold amino acids indicate sites undergoing mutation; underlined amino acids indicate sites evolving under positive selection; highlighted amino acids indicate sites mutating to result in a change within, or gain/loss of a potential N-linked glycosylation site gain/loss (NXS/Tx, where x is not Proline).

Discussion

Design of a globally relevant HIV-1 vaccine requires an understanding of transmitted viruses and their early immune adaptation in different populations. Here, we report the first study to comprehensively identify and classify the earliest changes to subtype C transmitted/founder full-length genome viruses. We provide a detailed analysis of the timing, frequency and patterns of these changes in five women with differing clinical disease progression. The predominant and earliest host selective pressure was from cytotoxic T lymphocytes. Despite increasing breadth of CTL responses over time [23], the frequency of mutations associated with CTL escape declined as viral load declined. We show that complex mutational pathways are used to escape in the majority of epitopes.

We predicted that CTL pressure, or reversion of transmitted escape, accounted for the majority (54%) of changes across the genome, reaffirming the importance of this response in early infection. Supporting our predictions, autologous peptide screening using IFN-γ ELISPOT in three of the five participants confirmed all nine epitopes predicted for these individuals (Liu *et al.*, in press). A further four of the 21 epitopes were likewise confirmed to be responsive (Liu *et al.*, unpublished data). The remaining eight were not confirmed either due to poor cell quality or screening after escape had already occurred. Mutations associated with CTL escape occurred earliest in *gag* and most frequently in *nef*, although this was likely influenced by the fact that three of the five individuals were HLA B*45 : 01 positive and targeted the same *nef* epitope. A further 35% of changes were associated with antibody pressure. We could not classify 11% of changes which may be associated with novel CTL epitopes (possibly in alternate reading frames [48,49]), compensation of escape, changes in antigen processing, $CD4^+$ T-cell pressure, natural killer (NK) cell pressure [50], nonneutralizing binding antibody activity, antibody-dependent cellular cytotoxicity (ADCC), viral fitness or evolutionary drift/hitchhiking. One unclassified mutating region in *pol* of participant CAP210 was subsequently found to be responsive by IFN-γ ELISPOT and may represent a novel CTL epitope (Liu *et al.*, in press). Another mutating region in *nef* of CAP210 corresponded to a known HLA-DR $CD4^+$ T-cell epitope; however, it was not restricted by the participants' HLA-DR (data not shown).

Three findings from this study illustrate the immense pressure the virus is under following infection. First, in these five women, escape was rapid and occurred at high frequency in acute infection (<5 weeks postinfection), with a four-fold reduction in the number of escapes per week over the 6-month period. Seventy percent of early

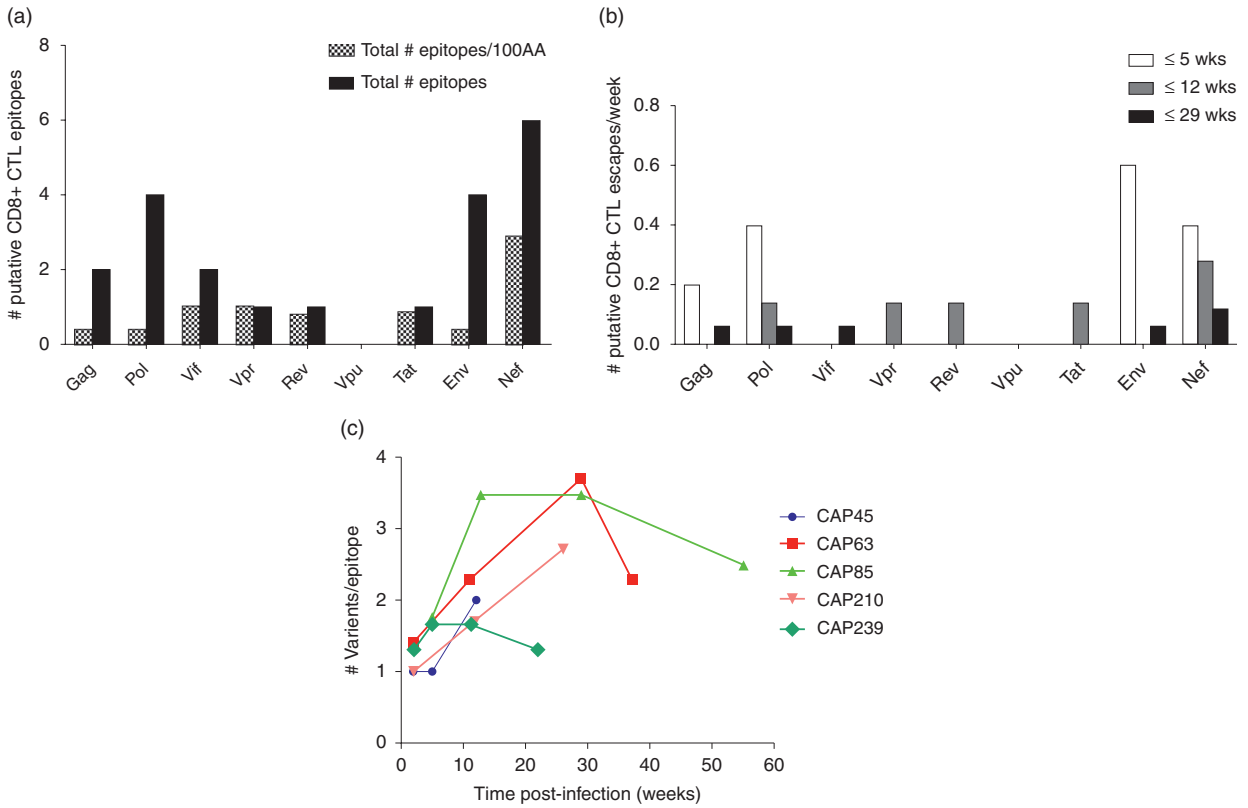

Fig. 2. Frequency, timing and complexity of putative cytotoxic T lymphocyte escape in the first 6 months of infection. (a) Total number of putative escapes per gene and per 100 amino acids (AAs). (b) Total number of putative escapes per week. (c) Total number of variants per CTL epitope over time for each of the five women. CTL, cytotoxic T lymphocyte.

escape (<12 weeks postinfection) occurred in high-entropy locations compared with only 33% after 12 weeks. These observations, although not statistically significant due to low sample size, consolidate reports that strong CTL pressure and rapid escape prevail early in infection [21,51] and that epitopes with higher entropy escape faster than more conserved epitopes [52]. Thus, the observed reduction in mutations associated with CTL escape later in infection seen in these women may be associated with targeting of epitopes that escape slowly.

Second, the pathway to CTL escape is complex, with this study showing an unprecedented level (81%) of genome regions under CTL pressure with mutations shifting between different amino acid sites or between different amino acids at the same site resulting in late or no fixation over the period investigated. A recent study proposed that initial mutations are replaced by secondary mutations which are less costly to the virus with respect to its survival [53]; however, we found no statistical support for more frequent shuffling/toggling in virus regions of structural or functional importance. It may be that this pattern of variation is seen simply because these regions of the genome are able to tolerate multiple changes (supported by the fact that 71% of the epitopes with shuffling/toggling corresponded to high-entropy database epitopes). Alternatively, fixation may only occur once mutations are introduced into sites facilitating complete escape as postulated in a detailed study of one participant by Henn *et al.* [54].

Third, and perhaps surprisingly since neutralizing antibody pressure typically emerges some weeks or months after infection [9,55,56], we identified escape patterns typically associated with antibody pressure as early as 2 weeks postinfection. Autologous nAb response data for these five participants indicate the earliest response to have arisen at 9 weeks postinfection [44]. These early changes could possibly be attributed to reversion of antibody escape mutations in the donor, early stochastic changes, nonneutralizing binding antibody activity, ADCC, *env* fitness or as very recently reported in a subtype B study by Bar *et al.* [57], may in fact be a result of early low level neutralizing antibody responses from which the virus escapes. We saw evidence for this in V5 of CAP45 wherein early changes corresponded to sites where nAb escape later in infection was confirmed in this participant [44].

When examining participant sequence data in the context of disease progression, we found no significant differences in early virus sequence diversification between individuals in our study. However, larger studies would be needed to evaluate the role of CTL escape in disease

progression. Recent studies revealed that clinical disease profiles are heritable in donor–recipient pairs [58,59] and that viral genotype can be a determinant of viral load set point [60]. Furthermore, transmission of a virus harboring a mutation with a fitness cost can also contribute to viral control [28,61]. We observed that viremic controller CAP45 was infected with a virus harboring mutations flanking the B*57/58 Gag TW10 and ISW9 epitopes previously reported to be associated with disease control [28]. It is worth mentioning that the two rapid progressors in this study were HLA-B (CAP63) and HLA-A, HLA-B and HLA-C (CAP210) homozygous. HLA homozygosity has previously been associated with poor clinical outcome [62].

Numerous CTL studies propose stable CTL epitopes that escape slowly or late in infection due to high fitness costs to be good vaccine targets [21,38,51]. This study suggests that the process of immune escape holds greater complexities which vaccine design strategies will need to take into account. Frequent and persistent shuffling and toggling of mutations within targeted epitopes may indicate high levels of epitope instability early in infection. Early APOBEC-mediated hypermutation, identified in more than a quarter of genome regions under putative immune pressure in this study, represents an efficient mechanism of CTL escape [22]. Not to be discounted is the unclassified 11% of genome changes which demonstrate that much more is at play very early in infection, possibly such as pressure from NK cells which has recently been emphasized [50]. These additional forces likely play a significant role in shaping the early virus.

Our findings provide novel data on the dynamic interplay between virus and host very early in infection and the complex pathways of escape in response to the earliest immune pressures acting on transmitted/founder subtype C viruses. These processes of immune adaptation are likely to pose further challenges that vaccines will need to overcome. Furthermore, this study demonstrates the highly sensitive nature of viral sequencing as a tool for the identification and characterization of early immune selective pressures that mould early HIV-1 evolution. Examining early host–virus interactions in the context of disease progression will enable us to identify those changes to the virus which are associated with better disease outcome and may therefore be incorporated into vaccine design.

Acknowledgements

This work is funded by the National Institute of Allergy and Infectious Disease (NIAID), National Institutes of Health (NIH) and the US Department of Health and Human Services (DHHS) (#AI51794, CAPRISA; #DK 49381 (MSC), CHAVI), as well as by the National Research Foundation (#67385), SA; the South African AIDS Vaccine Initiative; and amfAR grant 106997-43. We thank the clinical staff and participants from the CAPRISA 002 Acute Infection Study. We would like to thank Denis Chopera, Gama Bandawe and Roman Ntale for subgenomic sequence data and Ziyaad Valley-Omar for assistance with data analysis.

Author contributions: M.-R.A.: first author, manuscript writing, full-length genome and focused epitope DNA amplification and sequencing, data collation, sequence analyses, statistical analyses; F.K.T.: study design and coordination, full-length genome DNA amplification and sequencing, sequence analysis; N.K.N.: positive selection analyses and selection methods writing; S.A.G.: full-length genome and focused epitope DNA amplification and sequencing, statistical and sequence analyses; J.C.M.: full-length genome sequencing, focused epitope amplification and sequencing, sequence analysis; H.B.: half-genome DNA amplification and sequencing, sequence analysis; R.T.: full-length genome amplification and sequencing; D.d.A.R.: HLA typing of CAPRISA participants; K.M.: clinical site project director; C.S.: advising and guidance of clustered mutation and selection analyses; S.A.K.: CAPRISA study design; C.M.G.: immunology consultant and assistance with manuscript writing; and C.W.: principal investigator, corresponding author, manuscript writing.

Conflicts of interest

There are no conflicts of interest.

References

1. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, *et al*. **Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection.** *Proc Natl Acad Sci U S A* 2008; **105**:7552–7557.
2. Abrahams M-R, Anderson JA, Giorgi EE, Seoighe C, Mlisana K, Ping L-H, *et al*. **Quantitating the multiplicity of infection with human immunodeficiency virus type 1 subtype C reveals a nonpoisson distribution of transmitted variants.** *J Virol* 2009; **83**:3556–3567.
3. Salazar-Gonzalez JF, Salazar MG, Keele BF, Learn GH, Giorgi EE, Li H, *et al*. **Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection.** *J Exp Med* 2009; **206**:1273–1289.
4. Salazar-Gonzalez JF, Bailes E, Pham KT, Salazar MG, Guffey MB, Keele BF, *et al*. **Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing.** *J Virol* 2008; **82**:3952–3970.
5. Keele BF, Li H, Learn GH, Hraber P, Giorgi EE, Grayson T, *et al*. **Low-dose rectal inoculation of rhesus macaques by SIVsmE660 or SIVmac251 recapitulates human mucosal infection by HIV-1.** *J Exp Med* 2009; **206**:1117–1134.
6. Pantaleo G. **HIV-1 T-cell vaccines: evaluating the next step.** *Lancet Infect Dis* 2008; **8**:82–83.
7. Abdool Karim Q, Abdool Karim SS, Frohlich JA, Grobler AC, Baxter C, Mansoor LE, *et al*. **Effectiveness and safety of tenofovir gel, an antiretroviral microbicide, for the prevention of HIV infection in women.** *Science* 2010; **329**:1168–1174.

8. Yu XG, Addo MM, Perkins BA, Wej F, Rathod A, Geer SC, *et al*. **Differences in the expressed HLA class I alleles effect the differential clustering of HIV type 1-specific T cell responses in infected Chinese and caucasians.** *AIDS Res Human Retroviruses* 2004; **20**:557–564.
9. Wei X, Decker JM, Wang S, Hui H, Kappes JC, Wu X, *et al*. **Antibody neutralization and escape by HIV-1.** *Nature* 2003; **422**:307–312.
10. Treurnicht FK, Seoighe C, Martin DP, Wood N, Abrahams M-R, Rosa DDA, *et al*. **Adaptive changes in HIV-1 subtype C proteins during early infection are driven by changes in HLA-associated immune pressure.** *Virology* 2010; **396**:213–225.
11. Rybarczyk BJ, Montefiori D, Johnson PR, West A, Johnston RE, Swanstrom R. **Correlation between env V1/V2 region diversification and neutralizing antibodies during primary infection by simian immunodeficiency virus sm in rhesus macaques.** *J Virol* 2004; **78**:3561–3571.
12. Richman DD, Wrin T, Little SJ, Petropoulos CJ. **Rapid evolution of the neutralizing antibody response to HIV type 1 infection.** *Proc Natl Acad Sci U S A* 2003; **100**:4144–4149.
13. Moore CB, John M, James IR, Christiansen FT, Witt CS, Mallal SA. **Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level.** *Science* 2002; **296**:1439–1443.
14. Frost SDW, Wrin T, Smith DM, Kosakovsky Pond SL, Liu Y, Paxinos E, *et al*. **Neutralizing antibody responses drive the evolution of human immunodeficiency virus type 1 envelope during recent HIV infection.** *Proc Natl Acad Sci U S A* 2005; **102**:18514–18519.
15. Carlson JM, Brumme ZL. **HIV evolution in response to HLA-restricted CTL selection pressures: a population-based perspective.** *Microbes Infect* 2008; **10**:455–461.
16. Allen TM, Altfeld M, Geer SC, Kalife ET, Moore C, Sullivan KMO, *et al*. **Selective escape from CD8+ T-cell responses represents a major driving force of human immunodeficiency virus type 1 (HIV-1) sequence diversity and reveals constraints on HIV-1 evolution.** *J Virol* 2005; **79**:13239–13249.
17. Borrow P, Lewicki H, Wei X, Horwitz MS, Peffer N, Meyers H, *et al*. **Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus.** *Nature* 1997; **3**:205–211.
18. Jones NA. **Determinants of human immunodeficiency virus type 1 escape from the primary CD8+ cytotoxic T lymphocyte response.** *J Exp Med* 2004; **200**:1243–1256.
19. Liu Y, McNevin J, Cao J, Zhao H, Genowati I, Wong K, *et al*. **Selection on the human immunodeficiency virus type 1 proteome following primary infection.** *J Virol* 2006; **80**:9519–9529.
20. Gray CM, Mlotshwa M, Riou C, Mathebula T, de Assis Rosa D, Mashishi T, *et al*. **Human immunodeficiency virus-specific gamma interferon enzyme-linked immunospot assay responses targeting specific regions of the proteome during primary subtype C infection are poor predictors of the course of viremia and set point.** *J Virol* 2009; **83**:470–478.
21. Goonetilleke N, Liu MKP, Salazar-Gonzalez JF, Ferrari G, Giorgi E, Ganusov VV, *et al*. **The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection.** *J Exp Med* 2009; **206**:1253–1272.
22. Wood N, Bhattacharya T, Keele BF, Giorgi E, Liu M, Gaschen B, *et al*. **HIV evolution in early infection: selection pressures, patterns of insertion and deletion, and the impact of APOBEC.** *PLoS Pathog* 2009; **5**:e1000414.
23. Mlotshwa M, Riou C, Chopera D, de Assis Rosa D, Ntale R, Treunicht F, *et al*. **Fluidity of HIV-1-specific T-cell responses during acute and early subtype C HIV-1 infection and associations with early disease progression.** *J Virol* 2010; **84**:12018–12029.
24. Koup RA, Safrit JT, Cao Y, Andrews CA, McLeod G, Borkowsky W, *et al*. **Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome.** *J Virol* 1994; **68**:4650–4655.
25. Borrow P, Lewicki H, Hahn BH, Shaw GM, Oldstone MB. **Virus-specific CD8+ cytotoxic T-lymphocyte activity associated with control of viremia in primary human immunodeficiency virus type 1 infection.** *J Virol* 1994; **68**:6103–6110.
26. Ogg GS, Jin X, Bonhoeffer S, Dunbar PR, Nowak MA, Monard S, *et al*. **Quantitation of HIV-1-specific cytotoxic T lymphocytes and plasma load of viral RNA.** *Science* 1998; **279**:2103–2106.
27. van Loggerenberg F, Mlisana K, Williamson C, Auld SC, Morris L, Gray CM, *et al*. **Establishing a cohort at high risk of HIV infection in South Africa: challenges and experiences of the CAPRISA 002 acute infection study.** *PLoS One* 2008; **3**:e1954.
28. Chopera DR, Woodman Z, Mlisana K, Mlotshwa M, Martin DP, Seoighe C, *et al*. **Transmission of HIV-1 CTL escape variants provides HLA-mismatched recipients with a survival advantage.** *PLoS Pathog* 2008; **4**:e1000033.
29. Rousseau CM, Birditt BA, McKay AR, Stoddard JN, Lee TC, McLaughlin S, *et al*. **Large-scale amplification, cloning and sequencing of near full-length HIV-1 subtype C genomes.** *J Virol Methods* 2006; **136**:118–125.
30. Chopera DR, Mlotshwa M, Woodman Z, Mlisana K, de Assis Rosa D, Martin DP, *et al*. **Virological and immunological factors associated with HIV-1 differential disease progression in HLA-B 58:01-positive individuals.** *J Virol* 2011; **85**:7070–7080.
31. Hall TA. **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.** *Nucleic Acids Symp Ser* 1999; **41**:95–98.
32. Tamura K, Dudley J, Nei M, Kumar S. **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0.** *Mol Biol Evol* 2007; **24**:1596–1599.
33. Nielsen M, Lundegaard C, Blicher T, Lamberth K, Harndahl M, Justesen S, *et al*. **NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence.** *PLoS One* 2007; **2**:e796.
34. Bansal A, Gough E, Sabbaj S, Ritter D, Yusim K, Sfakianos G, *et al*. **CD8 T-cell responses in early HIV-1 infection are skewed towards high entropy peptides.** *AIDS* 2005; **19**:241–250.
35. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. **Relaxed phylogenetics and dating with confidence.** *PLoS Biol* 2006; **4**:e88.
36. Drummond AJ, Rambaut A. **BEAST: Bayesian evolutionary analysis by sampling trees.** *BMC Evol Biol* 2007; **7**:214.
37. Mansky LM, Temin HM. **Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase.** *J Virol* 1995; **69**:5087–5094.
38. Matthews PC, Prendergast A, Leslie A, Crawford H, Payne R, Rousseau C, *et al*. **Central role of reverting mutations in HLA associations with human immunodeficiency virus set point.** *J Virol* 2008; **82**:8548–8559.
39. Kosakovsky Pond SL, Frost SDW. **Not so different after all: a comparison of methods for detecting amino acid sites under selection.** *Mol Biol Evol* 2005; **22**:1208–1222.
40. Pond SK, Muse SV. **Site-to-site variation of synonymous substitution rates.** *Mol Biol Evol* 2005; **22**:2375–2385.
41. Pond SLK, Frost SDW, Muse SV. **HyPhy: hypothesis testing using phylogenies.** *Bioinformatics* 2005; **21**:676–679.
42. Scheffler K, Martin DP, Seoighe C. **Robust inference of positive selection from recombining coding sequences.** *Bioinformatics* 2006; **22**:2493–2499.
43. Pereyra F, Addo MM, Kaufmann DE, Liu Y, Miura T, Rathod A, *et al*. **Genetic and immunologic heterogeneity among persons who control HIV infection in the absence of therapy.** *Int J Infect Dis* 2008; **197**:563–571.
44. Moore PL, Ranchobe N, Lambson BE, Gray ES, Cave E, Abrahams M-R, *et al*. **Limited neutralizing antibody specificities drive neutralization escape in early HIV-1 subtype C infection.** *PLoS Pathog* 2009; **5**:e1000598.
45. Li B, Gladden AD, Altfeld M, Kaldor JM, Cooper DA, Kelleher AD, *et al*. **Rapid reversion of sequence polymorphisms dominates early human immunodeficiency virus type 1 evolution.** *J Virol* 2007; **81**:193–201.
46. Brumme ZL, Brumme CJ, Carlson J, Streeck H, John M, Eichbaum Q, *et al*. **Marked epitope- and allele-specific differences in rates of mutation in human immunodeficiency type 1 (HIV-1) Gag, Pol, and Nef cytotoxic T-lymphocyte epitopes in acute/early HIV-1 infection.** *J Virol* 2008; **82**:9216–9227.
47. Moore PL, Gray ES, Choge IA, Ranchobe N, Mlisana K, Abdool Karim SS, *et al*. **The C3-V4 region is a major target of autologous neutralizing antibodies in human immunodeficiency virus type 1 subtype C infection.** *J Virol* 2008; **82**:1860–1869.

48. Bansal A, Carlson J, Yan J, Akinsiku OT, Schaefer M, Sabbaj S, *et al*. **CD8 T cell response and evolutionary pressure to HIV-1 cryptic epitopes derived from antisense transcription.** *J Exp Med* 2010; **207**:51–59.
49. Berger CT, Carlson JM, Brumme CJ, Hartman KL, Brumme ZL, Henry LM, *et al*. **Viral adaptation to immune selection pressure by HLA class I-restricted CTL responses targeting epitopes in HIV frameshift sequences.** *J Exp Med* 2010; **207**:61–75.
50. Alter G, Heckerman D, Schneidewind A, Fadda L, Kadie CM, Carlson JM, *et al*. **HIV-1 adaptation to NK-cell-mediated immune pressure.** *Nature* 2011; **476**:96–100.
51. Davenport MP, Loh L, Petravic J, Kent SJ. **Rates of HIV immune escape and reversion: implications for vaccination.** *Trends Microbiol* 2008; **16**:561–566.
52. Ferrari G, Korber B, Goonetilleke N, Liu MKP, Turnbull EL, Salazar-Gonzalez JF, *et al*. **Relationship between functional profile of HIV-1 specific CD8 T cells and epitope variability with the selection of escape mutants in acute HIV-1 infection.** *PLoS Pathog* 2011; **7**:e1001273.
53. Herbeck JT, Rolland M, Liu Y, McLaughlin S, McNevin J, Zhao H, *et al*. **Demographic processes affect HIV-1 evolution in primary infection before the onset of selective processes.** *J Virol* 2011; **85**:7523–7534.
54. Henn MR, Boutwell CL, Charlebois P, Lennon NJ, Power KA, Macalalad AR, *et al*. **Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection.** *PLoS Pathog* 2012; **8**: e1002529.
55. Gray ES, Moore PL, Choge IA, Decker JM, Bibollet-Ruche F, Li H, *et al*. **Neutralizing antibody responses in acute human immunodeficiency virus type 1 subtype C infection.** *J Virol* 2007; **81**:6187–6196.
56. Yeh WW, Rahman I, Hraber P, Coffey RT, Nevidomskyte D, Giri A, *et al*. **Autologous neutralizing antibodies to the transmitted/ founder viruses emerge late after simian immunodeficiency virus SIVmac251 infection of rhesus monkeys.** *J Virol* 2010; **84**:6018–6032.
57. Bar KJ, Tsao C-Y, Iyer SS, Decker JM, Yang Y, Bonsignori M, *et al*. **Early low-titer neutralizing antibodies impede HIV-1 replication and select for virus escape.** *PLoS Pathog* 2012; **8**:e1002721.
58. Hecht FM, Hartogensis W, Bragg L, Bacchetti P, Grant R, Barbour J, *et al*. **HIV RNA level in early infection is predicted by viral load in the transmission source.** *AIDS* 2010; **24**:941–945.
59. Hollingsworth TD, Laeyendecker O, Shirreff G, Donnelly C a, Serwadda D, Wawer MJ, *et al*. **HIV-1 transmitting couples have similar viral load set-points in Rakai, Uganda.** *PLoS Pathog* 2010; **6**:e1000876.
60. Alizon S, von Wyl V, Stadler T, Kouyos RD, Yerly S, Hirschel B, *et al*. **Phylogenetic approach reveals that virus genotype largely determines HIV set-point viral load.** *PLoS Pathog* 2010; **6**:e1001123.
61. Miura T, Brumme ZL, Brockman M a, Rosato P, Sela J, Brumme CJ, *et al*. **Impaired replication capacity of acute/early viruses in persons who become HIV controllers.** *J Virol* 2010; **84**:7581–7591.
62. Tang J, Costello C, Keet IP, Rivers C, Leblanc S, Karita E, *et al*. **HLA class I homozygosity accelerates disease progression in human immunodeficiency virus type 1 infection.** *AIDS Res Human Retroviruses* 1999; **15**:317–324.
63. Fiebig EW, Wright DJ, Rawal BD, Garrett PE, Schumacher RT, Peddada L, *et al*. **Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection.** *AIDS* 2003; **17**:1871–1879.