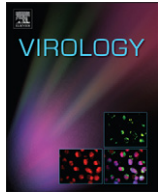




Contents lists available at SciVerse ScienceDirect

## Virology

journal homepage: [www.elsevier.com/locate/yviro](http://www.elsevier.com/locate/yviro)

## Molecular identification, cloning and characterization of transmitted/founder HIV-1 subtype A, D and A/D infectious molecular clones

Joshua Baalwa<sup>a</sup>, Shuyi Wang<sup>b,c</sup>, Nicholas F. Parrish<sup>b,c</sup>, Julie M. Decker<sup>d</sup>, Brandon F. Keele<sup>e</sup>, Gerald H. Learn<sup>b,c</sup>, Ling Yue<sup>f</sup>, Eugene Ruzagira<sup>g</sup>, Deogratus Ssemwanga<sup>g</sup>, Anatoli Kamali<sup>g</sup>, Pauli N. Amornkul<sup>h</sup>, Matt A. Price<sup>h</sup>, John C. Kappes<sup>d</sup>, Etienne Karita<sup>f</sup>, Pontiano Kaleebu<sup>g</sup>, Eduard Sanders<sup>h</sup>, Jill Gilmour<sup>h</sup>, Susan Allen<sup>i</sup>, Eric Hunter<sup>f</sup>, David C. Montefiori<sup>j</sup>, Barton F. Haynes<sup>j</sup>, Emmanuel Cormier<sup>h</sup>, Beatrice H. Hahn<sup>b,c</sup>, George M. Shaw<sup>b,c,\*</sup>

<sup>a</sup> Department of Pathology, University of Alabama at Birmingham, Birmingham, AL 35294, USA

<sup>b</sup> Department of Medicine, University of Pennsylvania, PA 19104, USA

<sup>c</sup> Department of Microbiology, University of Pennsylvania, PA 19104, USA

<sup>d</sup> Department of Medicine, University of Alabama at Birmingham, Birmingham, AL 35294, USA

<sup>e</sup> AIDS and Cancer Virus Program, SAIC-Frederick Inc., Frederick National Laboratory for Cancer Research, Frederick, MD 21702, USA

<sup>f</sup> Department of Pathology, Emory University, Atlanta, GA, 30322 USA

<sup>g</sup> Uganda Virus Research Institute/Medical Research Council Program on AIDS, Entebbe, Uganda

<sup>h</sup> International AIDS Vaccine Initiative, New York, NY 10004, USA

<sup>i</sup> Department of Global Health, Emory University, Atlanta, GA 30322, USA

<sup>j</sup> Duke Human Vaccine Institute, Duke University School of Medicine, Durham, NC 27710, USA

### ARTICLE INFO

#### Article history:

Received 22 August 2012

Returned to author for revisions

25 September 2012

Accepted 5 October 2012

Available online 1 November 2012

#### Keywords:

HIV-1

Transmitted/founder virus

Single genome sequencing

HIV-1 transmission

HIV-1 subtype A

HIV-1 subtype D

Neutralizing antibodies

### ABSTRACT

We report the molecular identification, cloning and initial biological characterization of 12 full-length HIV-1 subtype A, D and A/D recombinant transmitted/founder (T/F) genomes. T/F genomes contained intact canonical open reading frames and all T/F viruses were replication competent in primary human T-cells, although subtype D virus replication was more efficient ( $p < 0.05$ ). All 12 viruses utilized CCR5 but not CXCR4 as a co-receptor for entry and exhibited a neutralization profile typical of tier 2 primary virus strains, with significant differences observed between subtype A and D viruses with respect to sensitivity to monoclonal antibodies VRC01, PG9 and PG16 and polyclonal subtype C anti-HIV IgG ( $p < 0.05$  for each). The present report doubles the number of T/F HIV-1 clones available for pathogenesis and vaccine research and extends their representation to include subtypes A, B, C and D.

© 2012 Elsevier Inc. All rights reserved.

### Introduction

A globally effective human immunodeficiency virus type 1 (HIV-1) vaccine must prevent transmission of widely diverse virus strains including the less commonly studied subtypes A and D. Subtypes A and D account for approximately 12% and 2%, respectively, of all HIV-1 infections worldwide (Hemelaar et al., 2011), and together with their recombinant forms, are endemic in East and Central Africa where they co-circulate in the same regional and ethnic populations (Lihana et al., 2006; Ssemwanga

et al., 2012). Interestingly, there have been reports suggesting that distinct clinical outcomes may be associated with the two different virus subtypes. For example, compared to subtype A, subtype D infections have been associated with faster disease progression (Baeten et al., 2007; Kaleebu et al., 2002; Kiwanuka et al., 2008; Vasani et al., 2006) and a higher risk for developing HIV-1 associated neurocognitive disorders (Sacktor et al., 2009). Infection with subtype D has also been associated with preferential transmission of CXCR4 (X4) tropic and CXCR4/CCR5 (X4/R5) dual-tropic viruses (Church et al., 2008), an early co-receptor tropism switch from R5 to X4 (Kaleebu et al., 2007), a higher prevalence of X4 strains during chronic infection (Huang et al., 2007), and a faster rate of CD4+T lymphocyte decline (Kiwanuka et al., 2010) compared with subtype A infection. Phenotypic properties of subtype D and A virus strains that might contribute

\* Corresponding author at: Perelman School of Medicine, University of Pennsylvania, 409 Johnson Pavilion, 3610 Hamilton Walk, Philadelphia, PA 19104-6076, USA. Fax: +1 215 573 8976.

E-mail address: shawg@upenn.edu (G.M. Shaw).

to these clinical differences have not been identified. Thus, we reasoned that from both a vaccine and pathogenesis perspective, a precise molecular identification and biological characterization of molecularly cloned strains of transmitted/founder (T/F) subtype A, D and A/D viruses could be informative.

One limitation of previous studies aimed at characterizing genetic and biological properties of subtype A and D viruses has been the fact that subtype assignments, diversity measurements and biological analyses have frequently been conducted on subgenomic regions of these viruses rather than on full-length viral genes or genomes (Blish et al., 2007; Church et al., 2008; Haaland et al., 2009; Kaleebu et al., 2002; Provine et al., 2012; Redd et al., 2012). Such studies have been further confounded by the frequent occurrence of unique inter-subtype recombinants *in vivo* (Khoja et al., 2008; Lihana et al., 2006; Ssemwanga et al., 2012) as well as artifactual recombinants that may have been generated *in vitro* as a consequence of *Taq* polymerase DNA strand switching during the amplification of heterogeneous cDNA target sequences (Salazar-Gonzalez et al., 2008). There is the additional concern that viral sequences cloned from chronically infected patients might not reflect properties of viruses that result in virus transmission (Ochsenbauer et al., 2012; Parrish et al., 2012; Wilen et al., 2011) or could contain sporadic (and unidentifiable) mutations introduced by the most recent HIV-1 reverse transcriptase (RT) step, an RNA *Pol II* transcription error, or an MuLV RT error during the cDNA synthesis step prior to PCR amplification (Keele et al., 2008). To address these concerns, we sought to generate a panel of full-length molecularly cloned subtype A, D and A/D transmitted/founder genomes using single genome amplification (SGA)—direct amplicon sequencing (Keele et al., 2008; Lee et al., 2009; Salazar-Gonzalez et al., 2009), a strategy adapted from previously described single genome sequencing (SGS) methods for analyzing intact HIV-1 *pro*, *pol* and *env* genes (Palmer et al., 2005; Simmonds et al., 1990). Our laboratory's innovation was to use SGS in the context of acute HIV-1 and simian immunodeficiency virus (SIV) infection, together with a mathematical model of random virus evolution, to infer the exact nucleotide sequences of T/F virus genomes (Keele et al., 2008, 2009; Lee et al., 2009; Salazar-Gonzalez et al., 2009).

We define a T/F genome as a viral sequence that is transmitted and gives rise to productive clinical infection (Keele et al., 2008; Salazar-Gonzalez et al., 2009). We note that one or more viruses might breach the cervicovaginal or rectal mucosa or otherwise be introduced into a naïve individual and fail to replicate or be extinguished due to early innate immune responses or early stochastic events in the transmission process (Pearson et al., 2011). Such viruses are of no consequence to the present study since they do not lead to productive clinical infection. We also note that a 'transmitted' virus genome and a 'founder' virus genome from which subsequent genomes evolve may or may not be identical. For example, in the case of HIV-1 or SIV, the transmitted viral RNA genome must first undergo reverse transcription in order to productively infect the first cell in the naïve host. It is thus possible that the DNA provirus in this cell differs from the infecting viral RNA genome by either point mutation(s) or recombination event(s). The former are expected to occur with a frequency of about  $2.26 \times 10^{-5}$ , or about 0.2 mutations per 10 kb genome per infection event (Keele et al., 2008; Lee et al., 2009; Mansky and Temin, 1995). The latter are also common and potentially significant when the diploid viral RNA genome is heterozygous (Keele et al., 2008, 2009; Lee et al., 2009). RT mediated vRNA strand transfers (recombination) can occur multiple times with each reverse transcription event but are generally not apparent unless the virus genome is heterozygotic or polymerase slippage occurs resulting in sequence deletion or duplication (Levy et al., 2004). To account for these different

possibilities, we use the term 'transmitted/founder' genome to describe the viral genome that is transmitted and leads to productive clinical infection. We also make the important distinction between T/F virus genomes derived by single genome sequencing methods and that are *identical* to actual founding virus genomes at or near the moment of transmission, and "early" or "near-transmitted/founder" sequences derived by non-SGS techniques that are susceptible to *Taq* polymerase-mediated *in vitro* recombination, founder effects, and ambiguities arising from population sequencing, cloning-sequencing strategies, and in some studies, short-term virus culture effects (Aasa-Chapman et al., 2006; Blish et al., 2007; Coetzer et al., 2008; Derdeyn et al., 2004; Isaacman-Beck et al., 2009; Nedellec et al., 2009; Quakkelaar et al., 2007; Sagar et al., 2009, 2003; Seaman et al., 2010).

To pursue the identification and cloning of full-length subtype A, D and A/D T/F virus genomes, investigators representing the Center for HIV/AIDS Vaccine Immunology (CHAVI), the International AIDS Vaccine Initiative (IAVI) and the Uganda Virus Research Institute/Medical Research Council/Wellcome Trust collaborated in a molecular survey of incident HIV-1 infections in Uganda, Kenya and Rwanda with the goal of identifying acutely infected subjects at early Fiebig stages (Fiebig et al., 2003). HIV-1 negative high risk commercial sex workers, men who have sex with men and HIV-1 serodiscordant co-habiting couples were serially tested for antibody seroconversion. Testing was monthly or quarterly, and those individuals who became infected were offered enrollment into IAVI protocol C, which called for frequent blood draws for HIV-1 research. Because identification of T/F viral sequences depends on a phylogenetic analysis of very early viral sequences that undergo essentially random diversification (i.e., prior to the onset of adaptive immune responses that select for escape mutants) (Goonetilleke et al., 2009; Keele et al., 2008; Salazar-Gonzalez et al., 2009), we tested plasma specimens corresponding to the first ELISA HIV antibody positive time point by western immunoblot so as to determine the clinicopathological stage of acute HIV-1 infection as described by Fiebig et al. (2003). For subjects with later stages of acute infection (Fiebig stages V and VI), we tested the preceding antibody negative sample for HIV-1 vRNA, hoping to identify subjects in the earliest vRNA+/Ab- stages (Fiebig Stage I–II). The primary objectives of the present study were: (i) to determine if in the setting of observational field studies in Central and East Africa, we could identify patients and clinical samples amenable for a precise and unambiguous molecular identification of full-length T/F viral genomes; (ii) to molecularly clone and validate the replication competence of such T/F genomes and to assess their basic genetic and phenotypic characteristics; and (iii) to contribute subtype A and D clones and sequences to HIV/AIDS reagent and sequence repositories to encourage vaccine and pathogenesis research on these relatively neglected HIV-1 subtypes.

## Results

### Study subjects

Between October 2004 and February 2009, 165 subjects participating in IAVI vaccine preparedness studies, 14 subjects in the Uganda Virus Research Institute/Medical Research Council/Wellcome Trust HIV high risk cohort, and two subjects from an HIV discordant couple study seroconverted to HIV-1 antibody positivity were evaluated for inclusion in the current study. On the basis of specimen availability, serological data indicating very recent infection, virus load and specimen integrity, we selected 12 subjects for in-depth analysis. Six subjects were at

**Table 1**  
Socio-demographic, clinical and laboratory characteristics of study subjects.

Volunteer ID	Transmission type	Country sample	Sampling date	Clinical specimen	Viral load RNA copies/ml	ELISA	Western blot	Fiebig stage
R880F	MTF <sup>a</sup>	Rwanda	1/12/07	PBMC/Plasma	430,843	Reactive	Negative	III
R463F	MTF	Rwanda	3/9/07	PBMC/Plasma	150,000,000	Reactive	Indeterminate	IV
191084	MTF	Uganda	3/27/07	PBMC/Plasma	34,600	Reactive	Indeterminate	IV
191845	MTF	Uganda	4/5/07	Plasma	31,200	Reactive	Indeterminate	IV
9004SS	FTM	Uganda	5/17/07	Plasma	29,921	Reactive	Indeterminate	IV
191727	MTF	Uganda	1/24/07	Plasma	15,200	Reactive	Indeterminate	IV
190049	MTF	Uganda	6/21/05	Plasma	38,266,666	NR <sup>b</sup>	Negative	II
191647	FTM	Uganda	2/5/07	Plasma	242,222	NR	Negative	II
191859	MTF	Uganda	2/12/08	Plasma	53,334	NR	Negative	I
191882	FTM	Uganda	12/12/07	Plasma	15,378	NR	Negative	I
191947	FTM	Uganda	9/11/07	Plasma	5,466	NR	Negative	I
191982	MTF	Uganda	5/28/08	Plasma	7,156	NR	Negative	I

<sup>a</sup> MTF—Male to female, FTM—Female to male.

<sup>b</sup> NR—Non-reactive.

Fiebig Stage I–II (vRNA+/Ab–) and 6 had Fiebig Stage III–IV infection. Eight subjects were female and four were male. Two subjects were from Rwanda and ten were from Uganda. Other demographic and clinical information is as listed (Table 1).

#### Diversity analysis

SGA-direct amplicon sequencing was used to derive 160 5′ and 201 3′ overlapping half genome sequences from plasma viral RNA of 10 subjects. For two additional subjects, R880F and R463F, sequencing of twenty-two 9-kb single genome-derived amplicons was performed. Maximum likelihood trees of 5′ and 3′ half sequences from all 12 subjects showed that viral sequences from each subject formed monophyletic lineages with no interspersions of sequences (Fig. 1A and B). Sequences from five subjects clustered with reference subtype A viruses in both 5′ and 3′ trees, and sequences from five other subjects clustered with reference subtype D viruses in both genomic regions. Recombination identification program (RIP) (<http://www.hiv.lanl.gov/content/sequence/RIP/RIP.html>) and REGA HIV-1 subtyping (<http://dbpartners.stanford.edu/RegaSubtyping/>) analyses of these sequences compared to A and D reference sequences showed them to be non-recombinants. The 5′ and 3′ half genome sequences of two subjects, 191947 and 191982, clustered inconsistently with subtype A and D reference sequences suggesting that these T/F viruses likely represented inter-subtype A/D recombinants. A formal recombination analysis of full-length genomes from these subjects revealed them to be unique complex A/D recombinants (see below).

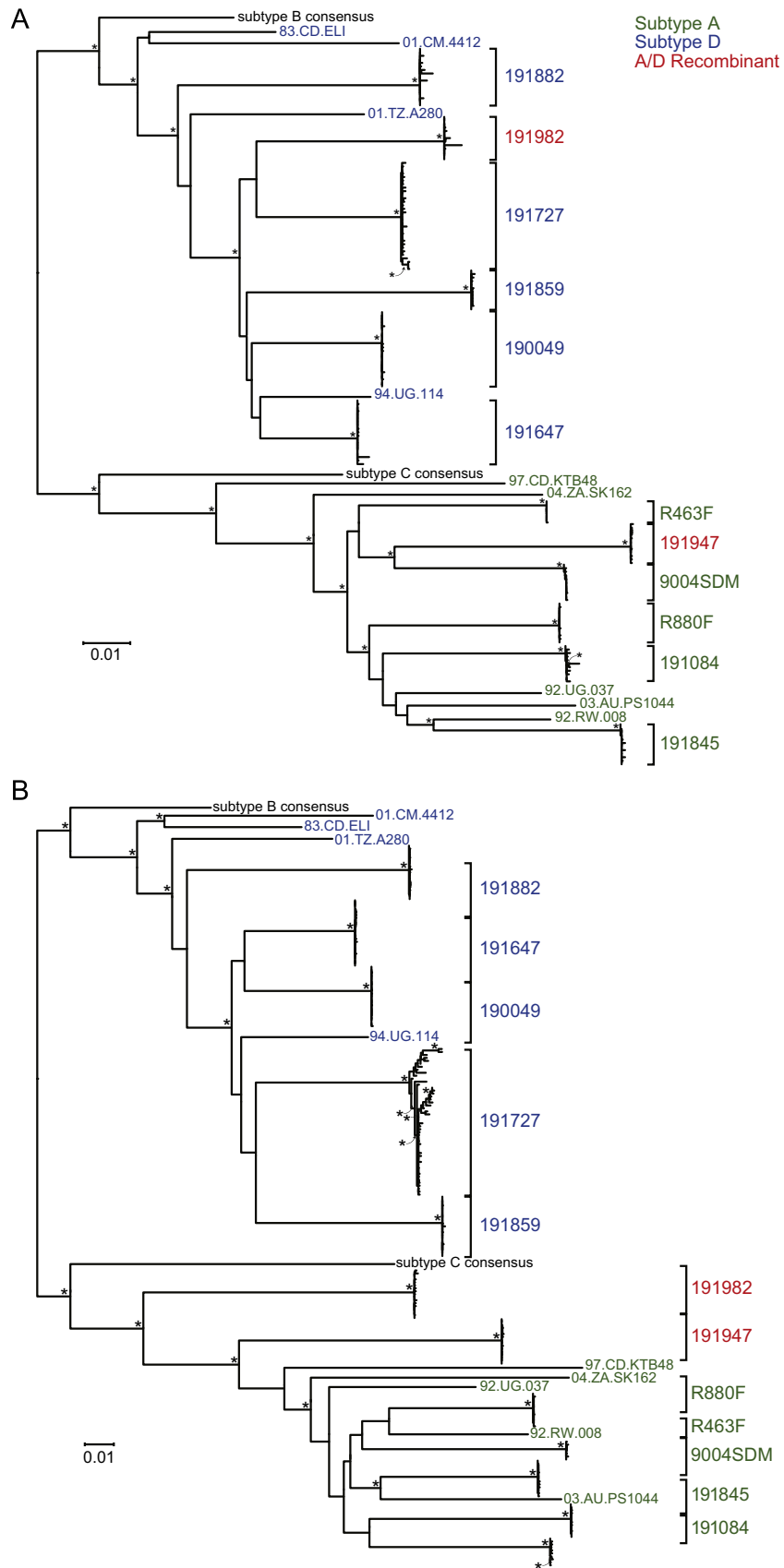
In 11 of 12 subjects, maximum within-patient sequence diversities were low ranging from 0.02% to 0.5% and 0.05% to 0.19% for the 5′ and 3′ half sequences, respectively (Table 2). Such low sequence diversity is generally indicative of single variant transmission based on model estimates and empirical observations of the rate of diversification of HIV-1 over the initial 100 days of infection (0.60%, 95% CI 0.54–0.68%) (Keele et al., 2008). However, transmission of two or more closely related viruses cannot be discriminated by this diversity cutoff value (Keele et al., 2008), and phylogenetic trees and *Highlighter* plots suggested that one of the 11 of subjects (191084) with relatively homogeneous sequences was actually productively infected by a minimum of two closely related viruses. A twelfth subject (191727) had a high maximum within-patient diversity of 1.54% and phylogenetic evidence indicated productive infection by at least three distinct T/F viral lineages (Fig. 1). We note that for the purposes of identifying examples of T/F full-length genomes, the numbers of sequences per subject that we analyzed (between 10 and 51 half-genomes) (Table 2) were sufficient to unambiguously infer consensus founder sequences. However, this

number is not sufficient for estimating with precision the total number of T/F virus genomes per subject, and thus the numbers of T/F viruses that we identified (median=1; mean=1.25; range=1–3) must be considered minimum estimates (Table 2).

#### Identification of T/F HIV-1 genomes

HIV-1 transmission at mucosal surfaces is characterized by a stringent population bottleneck that in most instances results in one or few variants out of many establishing productive clinical infection (Abrahams et al., 2009; Derdeyn et al., 2004; Haaland et al., 2009; Keele et al., 2008; Salazar-Gonzalez et al., 2009). The goal of the present study was to identify and clone one full-length T/F genome from each of the 12 subjects. Fig. 2A–C show *Highlighter* plots of 5′ and 3′ half genome sequences from subjects 190049 and 191845 and near full-length sequences for subject R463F. For these three subjects, the proportion of sequences identical to the respective consensus sequence ranged from 30% (R463F) to 67% (190049–3′ genome) and all variants differed from the respective consensus by just 1 to 3 randomly distributed mutations. Thus, for these three subjects, the consensus sequence is readily inferred to be the T/F sequence. Fig. 2D shows a similar pattern of early virus diversification in subject R880FPL but with the added feature of a shared nucleotide polymorphism at position ~8900. Sequences from other subjects sometimes had shared polymorphisms in just two sequences (R880FPB; 191982; 191882) and sometimes more (191882; 191859) (see Sfigure 1 in Supplementary material). Shared polymorphisms like these commonly arise due to early stochastic mutations that occur during the first one or few replication cycles following virus transmission (Lee et al., 2009) and it may not be possible to determine which variant was actually transmitted. Regardless, both sets of sequences are extremely closely related (generally differing by only one or few nucleotides per 10 kb), both are fit for replication, and both meet the operational definition of a T/F genome. For these subjects, we generally selected the overall consensus as the T/F virus genome, unless sequences were available from a sexual partner that indicated the likely transmitted allele. For some subjects, we had serial samples for analysis (191947; 191647). Sequential samples can reveal which sequences are best fit to replicate over time and oftentimes show sequence diversification suggesting immune selection (Goonetilleke et al., 2009; Salazar-Gonzalez et al., 2009). The latter changes can facilitate a distinction between a transmitted genome and mutant sequences arising from it.

Sequences from subject 9004SS were more complicated to analyze because there was a shared sequence polymorphism at



**Fig. 1.** Maximum likelihood phylogenetic trees of 5' (**A**) and 3' (**B**) half genome sequences from 12 acutely infected subjects. HIV-1 reference and T/F sequences are labeled in green for subtype A, blue for subtype D, and red for A/D recombinants. Subtype B and C reference sequences are labeled in black. The scale bars represent 1% genetic distance.

**Table 2**  
Diversity analysis of SGA sequences and genotypic characteristics of infectious molecular clones.

Volunteer ID	Clone ID	Subtype	No. of SGA <sup>a</sup>		Maximum nt length <sup>b</sup>		Maximum diversity (%)		No. of transmitted/founder viruses	IMC <sup>c</sup> V3 charge <sup>d</sup>	IMC gp41 length	IMC TAT exon 2	IMC genome length <sup>e</sup>	RE sites (5', 3') <sup>f</sup>	Cloning vector
			5'	3'	5'	3'	5'	3'							
R880F	pR880F	A	10	12	4448	4643	0.05%	0.10%	1	1.8	477	Full length	9729	SpeI, MluI	pCR XL TOPO
R463F	pR463F	A	10	10	4216	4629	0.09%	0.04%	1	1.8	477	Full length	9739	NotI, MluI	pCR XL TOPO
191084	p191084	A	11	10	4381	4612	0.35%	0.17%	2	3.7	456	Full length	9701	MluI, XhoI	pBR322
191845	p191845	A	13	13	4416	4720	0.15%	0.05%	1	3.8	477	Full length	9854	MluI, NotI	pCR XL TOPO
9004SS	p9004SDM	A	11	13	4454	4639	0.07%	0.10%	1	3.8	477	Full length	9725	MluI, NotI	pCR XL TOPO
191727	pSC191727	D	31	51	4359	4618	0.30%	1.54%	3	1.8	456	Truncated	9715	NotI, XhoI	pBR322
190049	p190049	D	22	21	4412	4564	0.10%	0.07%	1	2.8	456	Full length	9674	MluI, NotI	pCR XL TOPO
191647	p191647	D	19	23	4390	4567	0.37%	0.12%	1	3.8	456	Truncated	9653	MluI, NotI	pCR XL TOPO
191859	p191859	D	12	21	4403	4646	0.12%	0.14%	1	2.8	456	Truncated	9763	MluI, NotI	pCR XL TOPO
191882	p191882	D	17	18	4398	4570	0.42%	0.10%	1	1.8	456	Full length	9792	MluI, NotI	pCR XL TOPO
191947	p191947	A/D	12	16	4413	4665	0.07%	0.12%	1	3.8	477	Full length	9710	MluI, NotI	pCR XL TOPO
191982	p191982	A/D	13	17	4392	4617	0.50%	0.19%	1	0.8	456	Truncated	9778	MluI, NotI	pCR XL TOPO

<sup>a</sup> Single genome amplicons.

<sup>b</sup> Maximum 5' and 3' sequence nucleotide (nt) lengths span overlapping HIV-1 genomic segments corresponding to HXB2 nt positions 582–5039 and 4984–9606.

<sup>c</sup> Infectious molecular clone.

<sup>d</sup> V3 charge at pH 7.5.

<sup>e</sup> IMC genome lengths include all HIV-1 genes and complete 5' and 3' LTR elements.

<sup>f</sup> Restriction enzyme sites appended to 5' and 3' ends, respectively.

position ~3650 in a substantial fraction of 5' half genome sequences (Fig. 3). To determine which allele was transmitted and which was acquired, we examined sequences from an epidemiologically-linked sexual partner. We found that the minority sequence polymorphism present in three of eleven 5' sequences in the initial 9004SS sequence set was the only detectable allele present in the sexual partner, thus allowing for an unambiguous determination of the T/F sequence as shown (Fig. 3). For subject 191084, there were several sets of shared mutations in the 3' half-genome variant 1 sequences (boxed in Fig. 4A). These shared mutations are indicated in Fig. 4B where each set is observed to occur in one of two short stretches of 9 or fewer amino acids in Nef. Sequence toggles like these have been described previously and generally represent cytotoxic T-cell epitope escape mutations (Goonetilleke et al., 2009; Keele et al., 2008; Salazar-Gonzalez et al., 2009). By examining the timing and directionality of such mutations, it is possible to infer the likely T/F sequence and the evolving escape mutations derived from it. Finally, for subject SC191727 (Fig. 5), SGS allowed us to decipher at least three T/F lineages along with multiple unique inter-lineage recombinants, and from the predominant T/F lineage we could readily infer an unambiguous T/F sequence. However, this required that we amplify an additional ~3 kb fragment spanning the *pol* and *vpu* genes to confirm the linkage of 5' V1 and 3' V1 consensus genomes (Fig. 5).

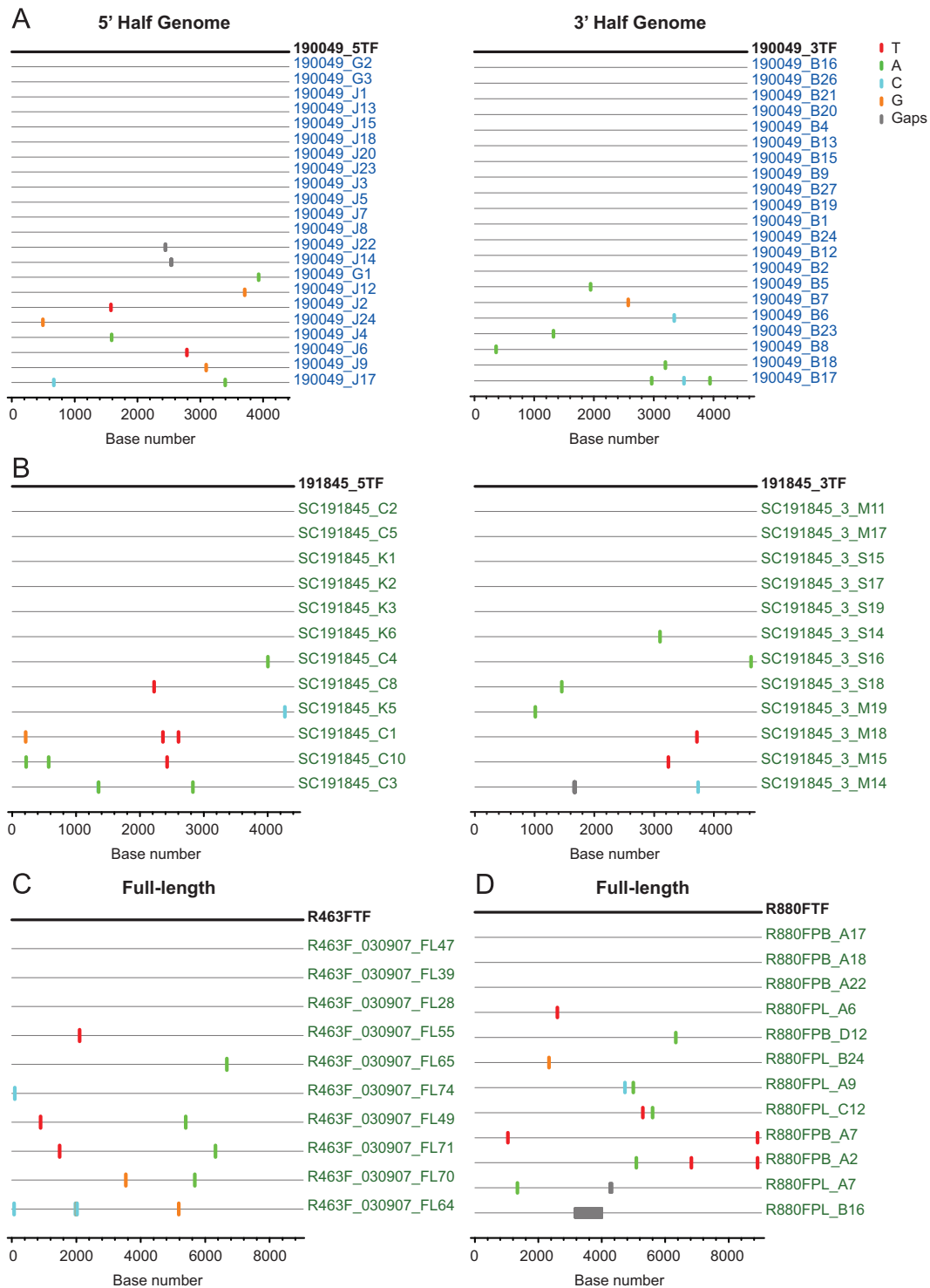
The 12 T/F genomes ranged in length from 9653 to 9854 nt (Table 2) and had intact open reading frames for the canonical HIV-1 genes *gag*, *pol*, *pro*, *int*, *vif*, *vpr*, *vpu*, *tat*, *rev*, *env* and *nef*. Five T/F genomes were entirely A subtype and five were entirely D subtype. Two genomes had evidence of recombination between A and D subtypes. We employed SimPlot and maximum likelihood phylogenetic analytical tools to characterize the recombination

breakpoint locations to distinguish unique from circulating recombinant form viruses. Using the 2010 subtype reference sequences, both SimPlot and maximum likelihood phylogenetic analyses concurred in different subtype assignments for the sub-genomic segments of 191947 and 191982 T/F genomes (Fig. 6). Both genomes were found to be complex unique subtype A and D recombinant forms as shown. The two viruses each contained subtype A gp120 Env and Vif sequences, but otherwise contained essentially antithetical distributions of A and D sequences.

We also noted salient subtype specific genetic polymorphisms in the envelope gp41 and *tat* exon 2 among the T/F subtype A and D genomes. All five subtype D and one A/D recombinant (191982) T/F genomes had a conserved seven amino acid deletion in the  $\alpha$ -helical amphipathic lentivirus lytic peptide (LLP) 2 domain (between g41 amino acid position 268–279; HXB2 numbering). One subtype A T/F genome, 191084, was distinct in having a seven amino acid truncation of the C-terminal LLP-2 domain at a position immediately adjacent and downstream of the subtype D deletion. Three of five subtype D (191727, 191647 and 191859) and one A/D recombinant (191982) genomes had a truncated *tat* exon 2 compared to none of five clade A genomes, resulting in a loss of 16 amino acids. V3 loop charges of subtype A, D and A/D T/F genomes V3 sequences at pH 7.5 were generally low, ranging from 0.8 to 3.8, typical of R5 tropic virus strains (Balasubramanian et al., 2012; De Wolf et al., 1994; Kaleebu et al., 2007).

#### Molecular cloning of subtype A, D and A/D T/F genomes

T/F genomes were cloned into either pCR XL TOPO or pBR322 plasmid vectors at the restriction sites indicated (Table 2). The nucleotide sequences of the T/F clones were deposited in Genbank



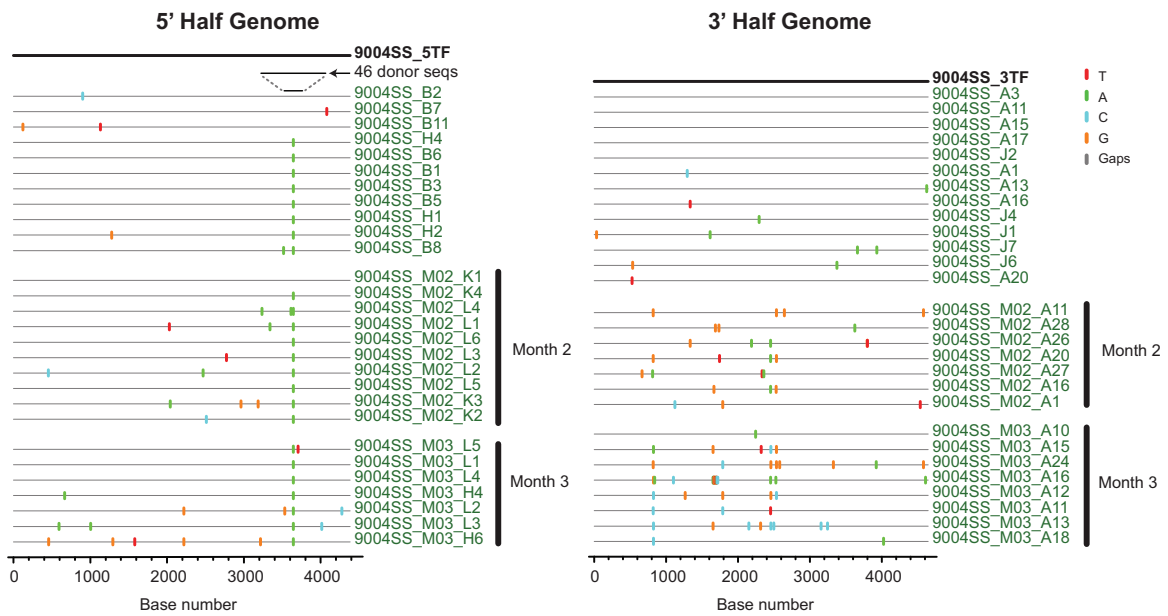
**Fig. 2.** Highlighter ([www.HIV.LANL.gov](http://www.HIV.LANL.gov)) plots of 5' and 3' half genomes or near full-length sequences for subjects 190049 (A), SC191845 (B), R463F (C) and R880FPB (D). Tic marks represent nucleotide substitutions as compared to the top-most T/F sequence in each plot. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(accession numbers JX236668–JX236679) and at [http://www.hiv.lanl.gov/content/sequence/HIV/USER\\_ALIGNMENTS/Baalwa/](http://www.hiv.lanl.gov/content/sequence/HIV/USER_ALIGNMENTS/Baalwa/).

#### Phenotypic characterization of molecularly cloned T/F viruses

Although T/F HIV-1 genomes are expected to be replication competent since their sequences correspond to founder viruses

that are responsible for productive clinical infection in humans, assaying virus replication is a key test of the T/F concept and the integrity of the T/F molecular virus clones. We thus tested the molecularly cloned viral genomic DNA corresponding to all 12 T/F viruses for replication competence in primary human CD4+ T-cells. T/F DNA was transfected and expressed in 293T cells and resulting virus was passaged cell-free onto activated primary



**Fig. 3.** Highlighter plots of 5' and 3' half genome sequences for subject 9004SS. A shared nucleotide polymorphism shown at position ~3650 in the 5' half genome Highlighter plot (HXB2 position 4300) was analyzed in 46 plasma vRNA/cDNA sequences from an epidemiologically-linked sexual partner of subject 9004SS. These sequences are illustrated in the figure, are accessible in Genbank JX877476–521), and allowed for an unambiguous determination of the T/F sequence at this position. Months 2 and 3 are post-seroconversion. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

human CD4+T lymphocytes from three different normal donors (Fig. 7A–C). All 12 viruses replicated relatively efficiently with kinetics generally comparable to the subtype B control virus SG3 and historical controls (Ochsenbauer et al., 2012; Salazar-Gonzalez et al., 2009) and with peak p24 antigen concentrations of 100–1000 ng/ml or higher. Interestingly, there was a statistically significant trend ( $p < 0.05$ ) for subtype D viruses to replicate more rapidly and to higher titers compared with subtype A viruses (Fig. 7D). This was a reproducible finding in different experiments, conducted on different days, at different multiplicities of infection (Fig. S2) and using CD4+T-cells from different normal human donors.

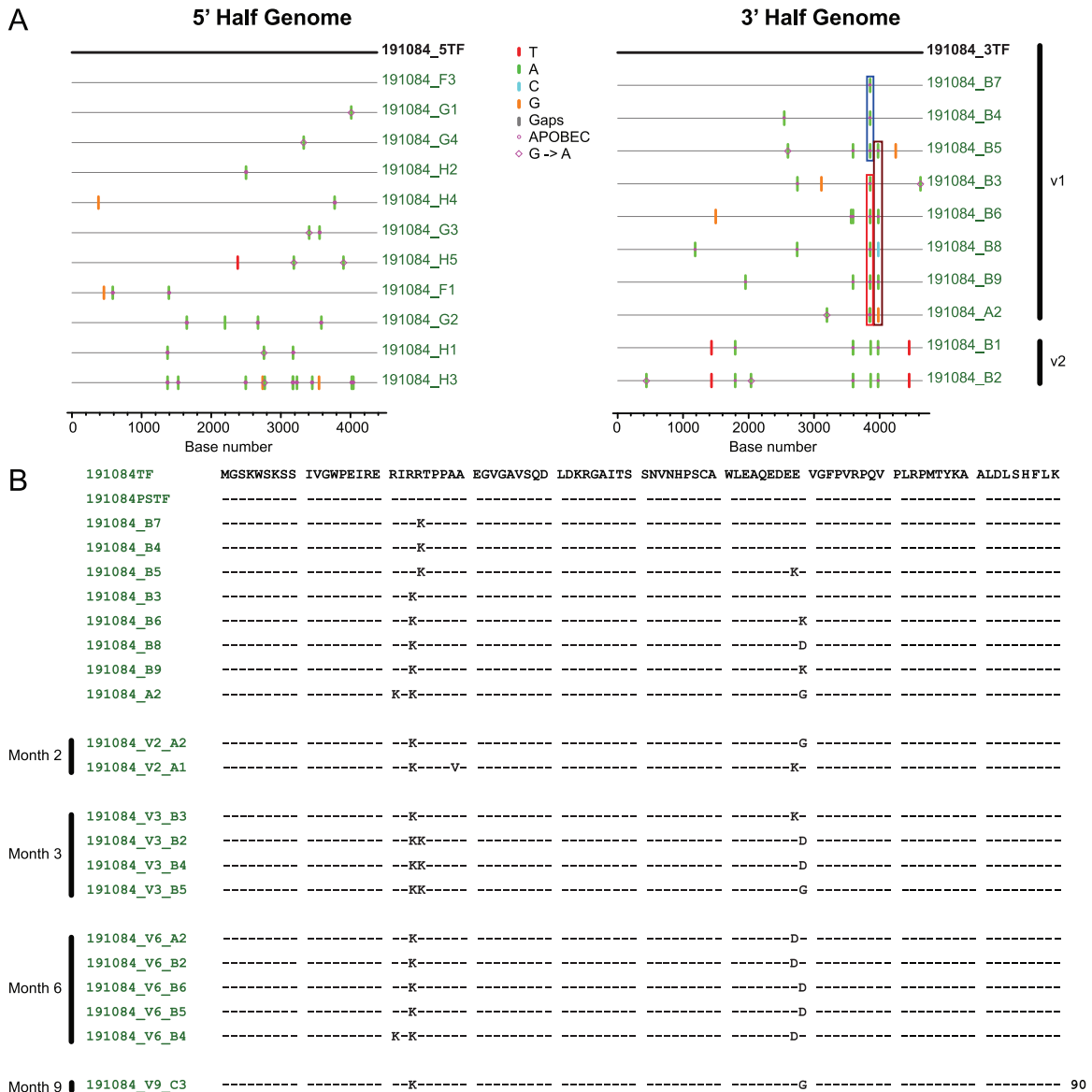
We next tested T/F subtype A, D and A/D viruses for use of the chemokine coreceptors CCR5 and CXCR4, anticipating possibly that there might be a predilection for T/F D subtype viruses to use CXCR4 based on previous literature reports of subtype D virus isolates (Huang et al., 2007; Kaleebu et al., 2007). Whereas AMD3100, a selective inhibitor of CXCR4 mediated entry, completely inhibited the entry of the control X4 tropic virus SG3, it had no effect on any of the 12 T/F viruses (Fig. 8A). Conversely, the CCR5 inhibitor TAK-779, inhibited the entry of the control R5 tropic virus YU2 and all T/F viruses. These findings suggested that CCR5 and not CXCR4 was the predominant co-receptor for virus entry by these T/F viruses. These findings were corroborated by analyses in NP2 cells that express CD4 and either CCR5 or CXCR4 but not both. Here again we found all 12 T/F viruses used CCR5 but not CXCR4 for cell entry (Fig. 8B).

Finally, we assayed the T/F viruses as IMCs or as Env-pseudotyped viruses for sensitivity to neutralization by pooled heterologous polyclonal HIV immunoglobulins from subtype B (HIVIG-B) or C (HIVIG-C) infected donors, by polyclonal IgG antibodies from each of three HIV-1 subtype C infected subjects previously shown to have broad and potent neutralizing antibodies, by the broadly neutralizing human monoclonal antibodies b12, VRC01, 2G12, 2F5, 4E10, PG9 and PG16, and by the fusion inhibitor T-1249 (Table 3). Interestingly, there was a trend for the subtype D virus Envs to be more resistant to HIVIG-C ( $P=0.045$ ) and HIVIG-B ( $P=0.059$ ) than were subtype A Envs. Similarly, four

out of five subtype D Envs were not neutralized by any of the 3 heterologous IgG polyclonal antibodies (SA-C62, SA-C72 and SA-C74) at IgG concentrations as high as 2500  $\mu\text{g/ml}$ , while all subtype A Envs were neutralized at concentrations ( $IC_{50}$ ) ranging from 226 to 1997  $\mu\text{g/ml}$  ( $p=0.139$ , 0.0007 and 0.052, respectively). The 12 T/F viruses were variably sensitive to the membrane proximal external region (MPER) specific mAb 4E10 and 2F5. Interestingly, all five T/F subtype A viruses were sensitive to neutralization by VRC01 (mean  $IC_{50}=1.75 \mu\text{g/ml}$ ) and extremely sensitive to PG9 (mean  $IC_{50}=0.05 \mu\text{g/ml}$ ) and to PG16 (mean  $IC_{50}=0.003 \mu\text{g/ml}$ ). This was not the case for subtype D T/F viruses, most of which exhibited resistance to these mAbs with  $IC_{50}$  values exceeding 10  $\mu\text{g/ml}$  ( $p=0.049$ , 0.004 and 0.0001, respectively). Conversely, subtype A viruses were pan-resistant to b12 ( $IC_{50} > 10 \mu\text{g/ml}$ ) whereas 3 of 5 subtype D viruses were sensitive (mean  $IC_{50}=5.69 \mu\text{g/ml}$ ;  $p=0.058$ ). The A/D recombinant viruses, which contain subtype A gp120 segments, exhibited neutralization patterns similar to subtype A T/F viruses. Subtype A, D and A/D viruses, like subtype B and C T/F viruses (Keele et al., 2008; Salazar-Gonzalez et al., 2009), were comparably sensitive to the gp41 reactive fusion inhibitor T-1249.

## Discussion

The concept of identifying full-length T/F virus genomes by SGS is a relatively new idea with numerous research applications in HIV/SIV research as well as for other viral systems. To our knowledge, there are only four reports in the scientific literature describing full-length HIV-1 T/F genomes and these are limited to subtypes B and C (Li et al., 2010; Ochsenbauer et al., 2012; Parrish et al., 2012; Salazar-Gonzalez et al., 2009). This approach has been widely adopted, since it makes possible a sensitive molecular probing and analysis of virus-host interactions operative at or near the moment of virus transmission, during the eclipse phase when virus replicates extensively but is not yet detectable in the plasma, and later when virus begins to diversify under host immune pressures (Bar et al., 2012; Goonetilleke et al., 2009).



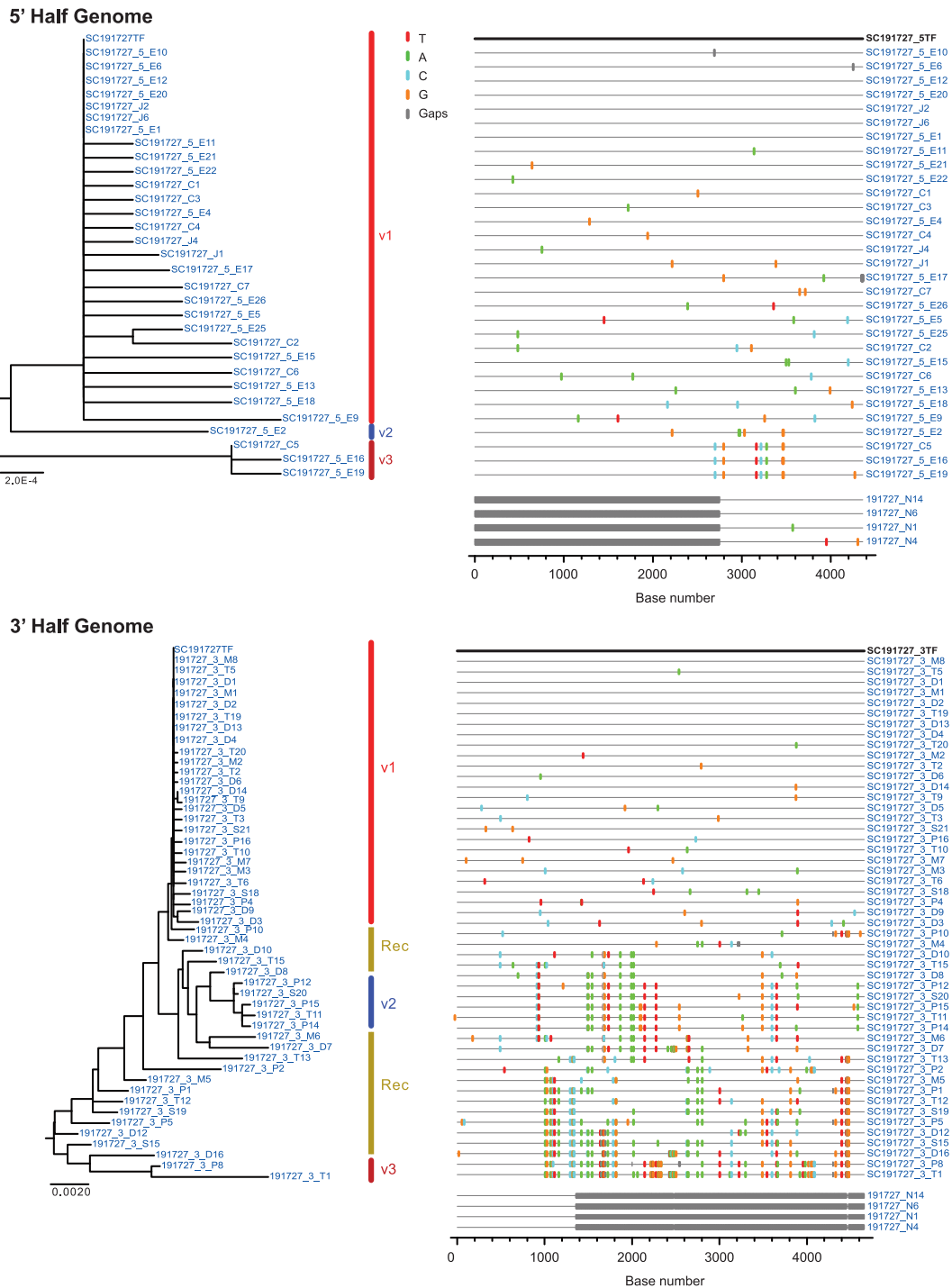
**Fig. 4. (A)** Highlighter plots of 5' and 3' half genome sequences for subject 191084 infected with two very closely related viruses evident in 3' sequences. Boxed mutations are represented in panel B below. **(B)** A partial Nef amino acid alignment of the inferred variant 1 T/F sequence and the month 2, 3, 6 and 9 post-seroconversion sequences is shown. The 191084PSTF sequence was derived from proviral DNA in peripheral blood mononuclear cells whereas all other sequences were derived from plasma viral RNA. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

New insights obtained by this approach include: (i) enumeration of T/F viral genomes responsible for productive clinical infection, which in heterosexuals is generally one, in homosexuals slightly higher, and in intravenous drug users higher still, sometimes involving 10 or more transmitted viruses (Bar et al., 2010; Keele et al., 2008; Li et al., 2010); (ii) biological characterization of T/F subtype B and C viruses resulting from mucosal transmission, which revealed CCR5 but not  $\alpha 4\beta 7$  tropism (Parrish et al., 2012; Wilen et al., 2011) as well as occasional use of alternative (non-CCR5/CXCR4) co-receptors (Jiang et al., 2011); (iii) antigenic characterization of T/F viruses revealing uniformly tier 2 or 3 level neutralization resistance typical of primary virus isolates (Keele et al., 2008; Li et al., 2010); (iv) failure of subtype B or C T/F viruses to replicate efficiently in macrophages compared with primary CD4+ T lymphocytes (Li et al., 2010; Ochsenbauer et al., 2012; Salazar-Gonzalez et al., 2009); (v) rapid evolution of CTL escape mutations (or reversions) away from the T/F proteome in the days and weeks following transmission followed weeks later by neutralizing antibody escape (Bar et al., 2012; Goonetilleke

et al., 2009); and (vi) immunologically-mediated sieving of viruses in phase III human vaccine trials (de Souza et al., 2012; Rolland et al., 2011). These findings highlight the enabling potential of T/F virus genome characterization and led us to pursue the analysis of T/F viruses for two less well studied HIV-1 subtypes A and D.

In our field survey of incident HIV-1 infections in Central and East Africa, we were able to identify discrete low diversity sequence lineages emanating from T/F viral genomes in each of 12 subjects in early stages of infection (Fiebig I–IV). Four of the earliest stage subjects (Fiebig I) had plasma viral loads between 7000 and 53,000 per milliliter representing early ramp-up viremia. But even in Fiebig Stage IV patients, we could identify T/F sequence lineages. Ten of 12 subjects had evidence of transmission of a minimum of one virus and two others had evidence of infection by at least two or three viruses. These are minimal estimates consistent with an extensive body of published data on hundreds of acutely infected subtype B and C subjects that suggest that the multiplicity of HIV-1 infection resulting from



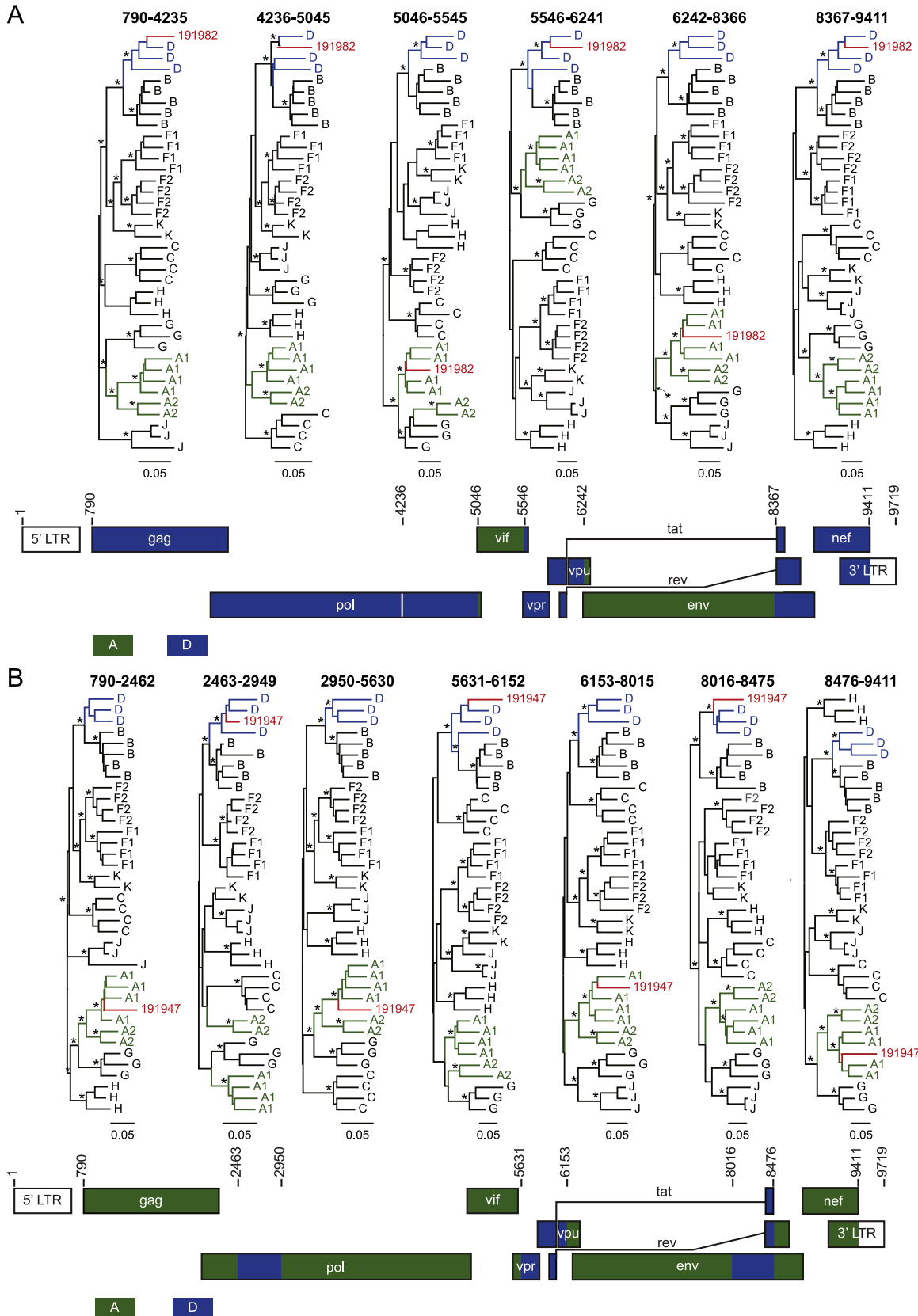


**Fig. 5.** Neighbor-joining tree phylogenies and *Highlighter* plots of 5' and 3' half genome sequences for subject 191727. T/F variants (v1-3) and recombinant (Rec) sequences are indicated. Amplicon sequences (N14, N6, N1 and N4) at the bottom of both panels that bridge the 5' and 3' half genome sequences show homology with lineage V1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

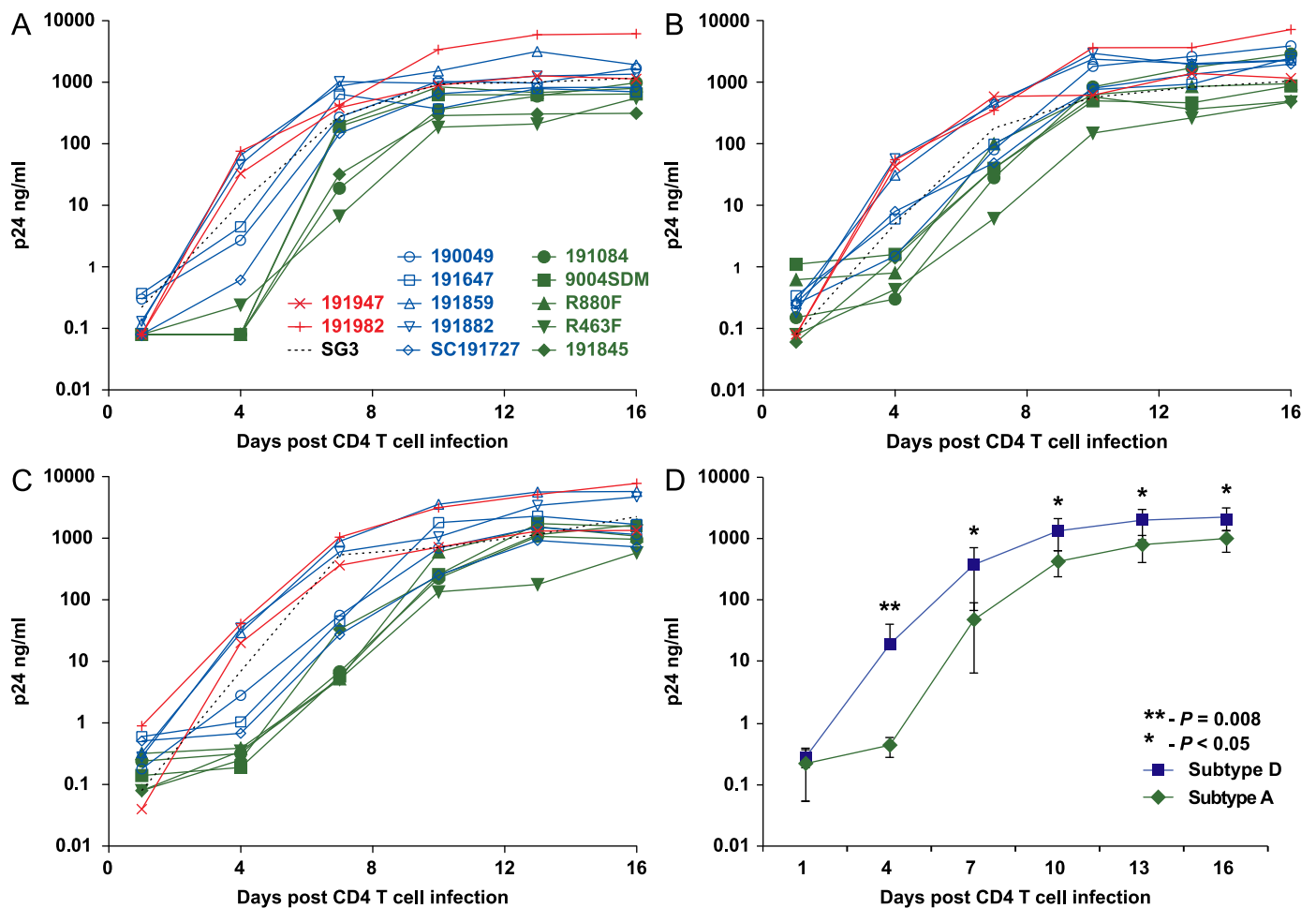
sexual transmission is generally one, although the range can reach as high as 10 or more (Abrahams et al., 2009; Haaland et al., 2009; Keele et al., 2008).

With the exception of shared stochastic mutations and one example of likely immune selection (Fig. 4; subject 191084), acute sequences emanating from discrete T/F genomes followed a pattern of essentially random variation with a near star-like phylogeny and a Poisson distribution of low frequency mutations indistinguishable from earlier findings with subtypes B and C (Li et al., 2010; Ochsenbauer et al., 2012; Salazar-Gonzalez et al.,

2009). This observation gives validation to the conclusion that the sequences identified in the present study as T/F genomes are genuine. Also consistent with this interpretation is the fact that all potential reading frames for all viruses were open, and each T/F virus was replication competent in activated human CD4+ T cells. Whether the modestly enhanced replicative properties of subtype D viruses compared with subtypes A viruses contributes to the clinical observations of enhanced subtype D pathogenicity described previously will require much more detailed analyses of additional viruses and comparative studies of virus biology.



**Fig. 6.** Recombination analysis of T/F viruses from 191982 and 191947. Different phylogenetic assignments for different genomic segments of 191982 (**A**) and 191947 (**B**) are shown in maximum likelihood phylogenetic trees corresponding to sequence regions mapped on the HIV-1 genome below. Numbers above trees and at breakpoints are HXB2 coordinates. Letters at branch tips indicate subtype reference sequences. Branches for subtypes A and D are shown in green and blue, respectively. Branches for recombinants are shown in red. Asterisks indicate branches for subtypes with  $\geq 70\%$  bootstrap support. Breakpoint locations are based on maximal  $X^2$  values for informative sites determined using SimPlot are shown by HXB2 coordinates. Boxes in the HIV-1 genome map represent the subtype assignments: subtype A, green; subtype D, blue. For segment 4236-5045, the 191982 sequence clusters significantly with another reference D sequence but the B sequence grouping nests within other D sequences, leading to phylogenetic ambiguity in this region.



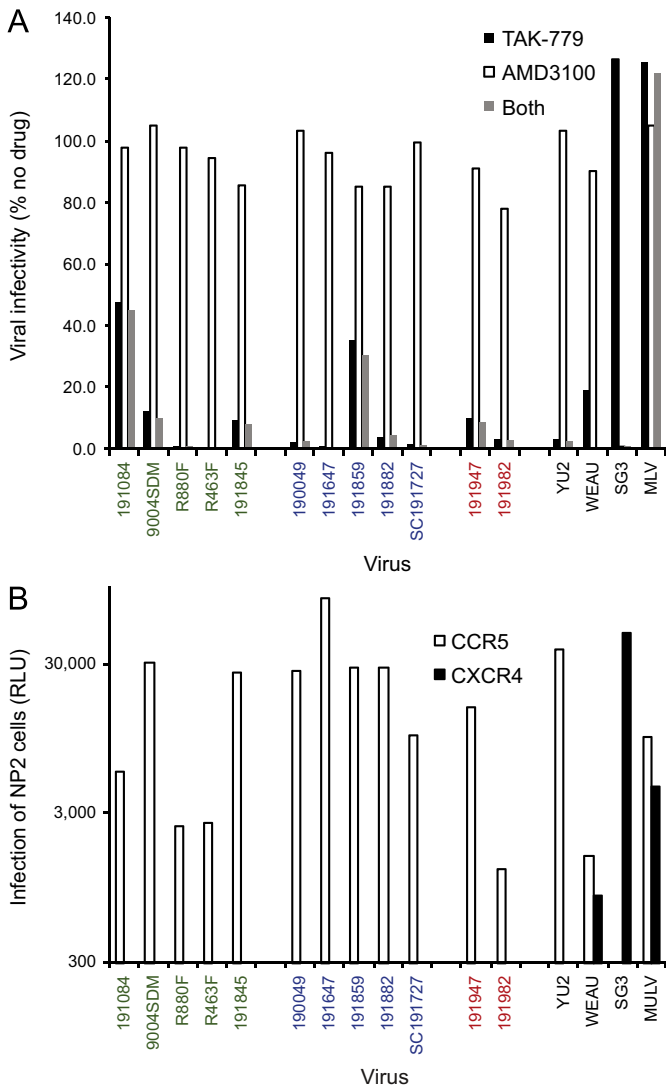
**Fig. 7.** Analysis of T/F virus replication in primary human CD4<sup>+</sup> T cells. CD4<sup>+</sup> T cells from three normal blood donors (A–C) were infected with T/F genome derived viruses at an M.O.I of 0.1. The replication kinetics of T/F viruses are shown relative to the prototypic viral clone SG3 (black dashed lines). Values on the y-axis denote ng per ml of p24 antigen and the x-axis days post-infection. (D) Comparison of replication kinetics of 5 subtype A with 5 subtype D T/F viruses is shown. Data points and error bars indicate mean and standard deviation. Statistical analysis was performed using the Mann–Whitney test and two-tailed *p*-values are shown. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

One important direction for future work will be on macrophage versus lymphocyte tropism of subtype D and A viruses and the ability of these viruses to replicate in cells bearing low levels of CD4 or alternative coreceptors in order to explore potential predilections of subtype D viruses for neurovirulence. This work is beyond the scope of the present study but is a potentially fruitful area of research.

One of the interesting findings of this study was the subtype-associated genetic polymorphisms between subtype A and D T/F genomes. We found that all five clade D genomes and one A/D recombinant genome had a highly conserved seven amino acid deletion within the amphipathic alpha-helical LLP-2 domain of the envelope intra-cytoplasmic tail compared to the five clade A genomes. A comparison of T/F genomes with subtype A, B, C and D HIV-1 reference sequences from the HIV Sequence Compendium 2011 (Kuiken et al., 2012), revealed that all subtype B and D sequences, but not subtype A or C sequences, shared this highly conserved deletion. The LLP 1, 2 and 3 domains have been shown to modulate HIV-1 envelope expression, incorporation, viral infectivity and virally-mediated cell-cell fusion (Bhakta et al., 2011; Kalia et al., 2003; Lambele et al., 2007; Lu et al., 2008; Steckbeck et al., 2011; Wyss et al., 2005). Interestingly, truncations that shorten the LLP2 domain have been shown to enhance fusion efficiency (Wyss et al., 2005). We also found that three of five subtype D genomes and one A/D recombinant genome had a

premature truncation of the second exon of *tat* resulting in the loss of sixteen amino acids. This *tat* polymorphism has been previously observed among some subtype B HIV-1 strains including HXB2, and there is conflicting evidence on the role of the second *tat* exon in viral replication (Guo et al., 2003; Mählknecht et al., 2008; Neuveut et al., 2003). The other notable finding was the similarity in V3 loop positive charge between subtype A (mean = +3.0, SD = 1.1) and D (mean = +2.6, SD = 0.8; *p* > 0.5) T/F genomes, which differs from previous studies that suggested that subtype D strains were likely to possess higher V3 loop positive charge than subtype A (De Wolf et al., 1994; Kaleebu et al., 2007). However, the sample size in our study is too low to extrapolate our findings; instead, our study provides naturally-occurring, clinically relevant T/F viral reagents for further biological analysis.

The antigenic properties of subtype A and D viruses are of interest since a broadly effective vaccine will need to protect against infection by these genetically diverse virus strains. A principal observation was that compared to subtype A, subtype D Env pseudoviruses as a group tended to be more resistant to neutralization by heterologous pooled immunoglobulins (HIVIG-B and HIVIG-C) and by broadly neutralizing polyclonal IgG antibodies isolated from individual HIV-1 subtype C infected patients. This finding was intriguing and suggests that vaccine immunogens based on subtype B and C Envs might induce less effective immunity against subtype D viruses. The 12 T/F viruses were



**Fig. 8.** Analysis of CCR5 and CXCR4 coreceptor usage. **(A)** TZM-bl cells were infected with replication competent T/F viruses in the presence of the CCR5 inhibitor TAK779 (10 μM), the X4 inhibitor AMD3100 (1.2 μM), or both inhibitors. The y-axis depicts percent infection in the presence of inhibitor relative to mock inhibitor control. **(B)** T/F *env* clones corresponding to the T/F full-length clones were pseudotyped by an HIV-1 backbone (NL4.3Δ*env*) expressing luciferase as a reporter and used to infect NP2 cells expressing CD4 and either CCR5 or CXCR4. Background values were determined by pseudovirus infection of NP2/CD4 cells expressing no co-receptor and were generally 100-fold less than the 300 RLU cut-off depicted in the figure. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

widely variable in their sensitivity to the MPER-specific mAbs 4E10 and 2F5, and this could not be simply explained by sequence variation in the canonical epitopes (Blish et al., 2007; Zwick et al., 2005). Interestingly, all five T/F subtype A viruses were extremely sensitive to PG9 and PG16, which was not the case for subtype D viruses, but again this could not be simply explained by sequence variation in these canonical epitopes (Walker et al., 2009). Conversely, subtype A viruses were pan-resistant to b12 whereas 3 of 5 subtype D viruses were sensitive. These neutralization profiles of subtype A, D, and A/D viruses may reflect the derivation of PG9 and PG16 mAbs from a subtype A donor and b12 from a subtype B donor, since subtypes B and D are more closely related phylogenetically than are subtypes B and A (Dwivedi and Sengupta, 2012).

Much remains to be explored concerning the transmission biology, immunopathogenesis, and prevention strategies for HIV-1 subtypes A and D. The molecular identification and cloning of

**Table 3**  
Phenotypic characteristics of molecularly cloned clade A, D and AD T/F viruses.

Clone	IMC derived virus assay										Tier	Env pseudovirus assay					
	VRC01 ug/ml	PG9 ug/ml	PG16 ug/ml	b12 ug/ml	2F5 ug/ml	4E10 ug/ml	2G12 ug/ml	T-1249 ug/ml	Co-receptor usage	HIVIG-B ug/ml		HIVIG-C ug/ml	SA-C62 ug/ml	SA-C72 ug/ml	SA-C7 ug/ml		
<b>Subtype A</b>	0.69	0.04	0.008	> 10	> 50	10.80	> 10	0.03	CCR5	971	142	1368	1442	308			
R880F	0.94	0.06	0.004	> 10	> 50	32.60	> 10	0.03	CCR5	964	167	1244	588	953			
9004SDM	2.18	0.07	0.002	> 10	> 50	20.70	> 10	0.52	CCR5	445	41	849	1185	1997			
191845	2.52	0.05	0.002	> 10	> 50	4.10	> 10	0.03	CCR5	495	30	492	772	876			
R463F	2.40	0.007	0.0005	> 10	> 50	> 40	> 10	0.01	CCR5	615	59	1982	226	358			
190049	3.28	> 10	> 10	0.37	15.10	> 40	> 10	0.02	CCR5	992	176	214	1626	407			
191647	> 10	0.03	> 10	> 10	15.70	21.60	> 10	0.03	CCR5	641	235	> 2500	> 2500	> 2500			
191859	8.80	> 10	5.12	5.65	> 50	21.90	> 10	0.09	CCR5	> 2500	621	> 2500	> 2500	> 2500			
191882	> 10	> 10	2.44	37.90	3.30	3.30	> 10	0.11	CCR5	> 2500	> 625	> 2500	> 2500	> 2500			
SC191727	0.21	> 10	> 10	0.058	0.071	0.941	0.667	0.555	CCR5	1289	129	> 2500	> 2500	> 2500			
191947	0.049	0.004	0.0001	0.058	0.071	0.941	0.667	0.555	0.059	0.045	0.139	0.0007	0.052				
191982	5.02	0.19	0.03	9.21	> 50	22.70	> 10	0.08	CCR5								
YU2	0.23	0.09	0.002	> 10	> 50	5.50	> 10	0.01	CCR5								
<b>Controls-IMCs</b>	0.28	1.37	0.05	> 10	> 50	7.00	> 10	0.13	CCR5								
SG3	0.08	> 10	> 10	0.03	4.50	12.40	4.12	0.02	CXCR4								
<b>Controls-pseudoviruses</b>	Q23.17																
MN.3																	
MW965.26																	
									1B	453	14	817	764	51			
									1A	8	10	33	12	8			
									1A	11	2	2	< 1	< 1			

<sup>\*</sup> P values are for group comparisons between subtype A and D viruses using an unpaired t-test (two-tailed).

full-length T/F HIV-1 genomes by SGS takes a step toward these goals by taking advantage of the natural virus population bottlenecks that occurs with virus transmission from one individual to the next, thus ensuring that the cloned genome represents a biologically and clinically relevant T/F virus unaltered by *in vitro* cultivation, selective amplification, or *Taq* polymerase, RNA *Pol II* or MuLV RT mutational errors. The 12 subtype A, D, and A/D T/F genomes reported here double the number of HIV-1 subtypes for which T/F genomes have been derived and double the number of T/F virus genomes available for analysis. This work can facilitate a systematic molecular analysis of genome-wide structure, function, antigenicity and immunogenicity of these clinically important viruses and virus subtypes. Beyond this, the identification and cloning of other highly variable single-stranded T/F RNA (or DNA) viruses including, for example, the Flaviviridae represent an important new research opportunity (Li et al., 2012).

## Methods and materials

### Study subjects

Subjects were enrolled in IAVI-sponsored prospective vaccine preparedness cohort studies of HIV-1 antibody negative heterosexuals or men who have sex with men (Price et al., 2012, 2011; Tang et al., 2011), in a Uganda Virus Research Institute/Medical Research Council/Wellcome Trust HIV-1 acquisition cohort study, and in a heterosexual sero-discordant couples cohort study in Rwanda (Derdeyn et al., 2004). Subjects were given HIV counseling, condom provision and regular HIV testing either monthly or quarterly. Those who seroconverted to HIV-1 were screened for stage of primary HIV-1 infection and those between Fiebig Stage I to IV were included in the present study if sufficient cryopreserved plasma and/or peripheral blood mononuclear cells (PBMCs) were available for analysis and if viral loads and vRNA/cDNA integrity were sufficient for SGA analysis. The guideline for laboratory staging of primary HIV-1 infection (Fiebig staging) has been described elsewhere (Fiebig et al., 2003; Keele et al., 2008). All subjects gave informed consent and blood specimen collections were undertaken with institutional review board and other regulatory approvals.

### Proviral DNA extraction, viral RNA isolation and cDNA synthesis

Plasma viral RNA (vRNA) and proviral DNA were extracted from blood specimens using the QIAamp RNA or DNA Mini Kits (Qiagen, Valencia, CA) as previously described (Keele et al., 2008; Salazar-Gonzalez et al., 2009). Reverse transcription was performed using Superscript III (Invitrogen) according to the manufacturer's instructions in two steps. First, 30  $\mu$ l of RNA were mixed with 0.5 mM of each deoxynucleoside triphosphate (dNTP) and 0.25 mM of reverse primer 1.R3B3R (5'-ACTACTTGAAGCACTCAAGGCAAGCTTTATTG-3'), reconstituted in a volume of 36  $\mu$ l and incubated at 65 °C to denature RNA secondary structure. Second, 24  $\mu$ l of an RT reaction mix consisting of first strand RT buffer (1X), 5 mM dithiothreitol, 2 U/ $\mu$ l of RNase inhibitor-RNase OUT and 5 U/ $\mu$ l Superscript III was added to the first mix, bringing the total reaction volume to 60  $\mu$ l, which was then incubated at 50 °C for 80 min. The RT reaction was then terminated by incubation at 85 °C for 5 min, followed by the addition of 3  $\mu$ l of RNase H and a final incubation at 37 °C for 20 min to denature the RNA strands. The reverse primer 1.R3B3R was used to synthesize cDNA templates for full-length and 3' half genome amplification. To synthesize cDNA templates for amplification of the 5' genomic half, the same RT protocol was followed but using B5R1 (5'-CTTGCCACACAATCATCACCTGCCAT-3'; nt 5052–5077) as the

reverse primer. The synthesized cDNA was either used immediately or stored at –80 °C.

### Single genome amplification

cDNA was serially diluted with each dilution replicated in 12 wells and amplified by nested PCR. cDNA dilutions found to result in 30% or fewer PCR positive wells were adopted to obtain additional amplicons using 96-well reaction plates. Two approaches were used to derive single genome amplicons. In one approach, ~9 kb genomes extending from 5'U5 to 3'U3-R were amplified as previously described (Keele et al., 2008; Salazar-Gonzalez et al., 2009). Briefly, first round PCR was carried out in the presence of 1  $\times$  Expand Long template buffer containing MgCl<sub>2</sub> at a final concentration of 1.75 mM, 0.35 mM of each dNTP, 0.3  $\mu$ M of each primer and 3.75 u/ $\mu$ l of expand long template enzyme mix (Roche) in a 50  $\mu$ l reaction. First round primers were: forward-U5Cc (5'-CCTTGAGTGTCTAAGTAGTG-TGTGCCCGTCTGT-3') and the reverse primer 1.R3B3R. The second round PCR primers were: forward-U5Cd (5'-AGTAGTGTGTGCCCGTCTGTGTGTGACTC-3') and reverse-2.3'3'pIcB (5'-TAGAGCACTCAAGGCAAGCTTTATTGAGGCTTA-3'). PCR conditions were: 94 °C for 2 min, followed by 10 cycles of 94 °C for 15 s, 55 °C for 30 s, and 68 °C for 8 min, followed by 25 cycles of 94 °C for 15 s, 55 °C for 30 s, and 68 °C for 8 min, with cumulative increments of 20 s at 68 °C with each successive cycle and a final extension period of 10 min at 68 °C. The first round PCR products were then utilized as templates for the second round PCR reactions under the same conditions. PCR products were examined on 1% agarose gels for amplicons > 8 kb. The other approach consisted of amplifying overlapping 5' (U5, gag-pol and vif) and 3' (pol, vif, vpr, rev, vpu tat, env, nef and U3-R) half genomes. Amplification reactions were carried out in the presence of 1  $\times$  High Fidelity Platinum Taq PCR buffer, 0.2  $\mu$ M of each primer, 2 mM Mg SO<sub>4</sub>, 0.2 mM each deoxynucleoside triphosphate and 0.025 u/ $\mu$ l of Platinum Taq High Fidelity polymerase in 20  $\mu$ l reactions (Invitrogen, Carlsbad, CA). The first round PCR conditions were 94 °C for 2 min, followed by 35 cycles of 94 °C for 20 s, 55 °C for 30 s, and 68 °C for 5.50 min, followed by a final extension of 68 °C for 5 min. For the second round of PCR, the total number of cycles was increased to 37 and the annealing temperature changed from 55 °C to 58 °C. The 5' (1.U5.B1F/B5R1 and 2.U5.B4F/B5R2) and 3' (B3F1/1.R3B3R and B3F3/2.R3B6R) half genome primers used are described elsewhere (Salazar-Gonzalez et al., 2009). The amplicons were examined on precast 1% agarose 96-well E-gel. For both approaches, the first 30 nucleotides on the 5' end of the ~9 kb/5'-half genomes and the last 30 nucleotides on the 3' end of the ~9 kb/3'-half genomes corresponded to internal nested primers U5Cd/U5.B4F and 2.3'3'pIcB/1.R3B6R, respectively, used during the second round of PCR amplification and were thus primer derived. A separate PCR reaction was therefore carried out to determine the precise viral nucleotides at these positions utilizing the same cDNA aliquots used to obtain 5' half amplicons and primers encompassing a ~0.4 kb region from the 5' R (5'-GGTCTCTGGTITAGACCAGAT-3'; nt 455–476) to gag (5'-TCCAGCTCCCTGCTTGCCCATACTA-3'; nt 914–890). This enabled the determination of the complete viral LTR sequence (U3-R-U5), and subsequently, the complete 5' half (U3-R-U5, gag-pol and vif) and 3' half (pol, vif, vpr, rev, vpu tat, env, nef and U3-R-U5) genomes. Thus, the full-length HIV-1 genomes (U3-R-U5-HIV genes–U3-R-U5) could be ascertained.

### DNA sequencing and sequence alignment

DNA sequencing and sequence alignments were performed as previously described (Bar et al., 2010; Keele et al., 2008, 2009;

Li et al., 2010; Salazar-Gonzalez et al., 2009). All sequences are available under GenBank accession nos. JX202785–JX203216, JX236668–JX236679 and JX877476–JX877521. Alignments are available at ([http://www.hiv.lanl.gov/content/sequence/HIV/USER\\_ALIGNMENTS/Baalwa/](http://www.hiv.lanl.gov/content/sequence/HIV/USER_ALIGNMENTS/Baalwa/)).

#### *Sequences, diversity analysis and phylogenetic inference of T/F genomes*

Sequences were aligned using Clustal W version 2.11 (Larkin et al., 2007). Phylogenetic trees were constructed by maximum-likelihood estimation using PhyML version 3.0 (Guindon et al., 2010) or by the neighbor joining method using ClustalW (Larkin et al., 2007). Maximum within patient sequence diversity was determined by phylogenetic analysis using parsimony (PAUP) version 4.0 (Rogers and Swofford, 1999). Subtype assignments were based on the REGA ([www.dbpartners.stanford.edu/RegaSubtyping/](http://www.dbpartners.stanford.edu/RegaSubtyping/)) and RIP ([www.hiv.lanl.gov](http://www.hiv.lanl.gov)) tools and maximum likelihood phylogenetic tree analyses. Recombination breakpoint locations were identified by SimPlot (Lole et al., 1999). *Highlighter* plots ([www.hiv.lanl.gov](http://www.hiv.lanl.gov)) were used to display nucleotide substitutions in the viral quasispecies as compared to the T/F sequence. T/F sequences were inferred based on a model of early random virus evolution as previously described (Keele et al., 2008; Lee et al., 2009; Ochsenbauer et al., 2012; Salazar-Gonzalez et al., 2009).

#### *Generation of infectious molecular clones*

To obtain IMCs corresponding to T/F viruses from subjects 191845, 90045S, 190049, 191647, 191859, 191882, 191727, 191947 and 191982, each inferred T/F full-length genome was chemically synthesized as three overlapping fragments containing unique restriction sites to allow concatenation (Blue Heron Biotechnology). Additional restriction enzyme cleavage sites were added to the 5' and 3' ends of the outermost fragments to facilitate insertion into an appropriate cloning vector (pCR XL TOPO or pBR322). This was followed by a one-step ligation of the three viral DNA fragments into a cloning vector using T4 DNA ligase (New England Biolabs Inc.).

To construct IMCs of T/F viruses from subjects R880F, R463F and 191084, overlapping 5' and 3' fragments containing complete LTR elements at both genomic ends were amplified using proviral DNA as the template. The antisense primer for the 5' fragment and the sense primer for the 3' fragment were designed to lie downstream and upstream, respectively, of a unique restriction enzyme site to allow for generation of overlapping fragments that encompass this site. The sense primer for the 5' fragment was designed to exactly anneal to the beginning of the U3-R-U5 5' LTR and the antisense primer of the 3' fragment to exactly begin at the end of the U3-R-U5 3' LTR. Fragments were amplified using bulk PCR as previously described (Salazar-Gonzalez et al., 2009). Amplification products were analyzed for correct size by 1% agarose gel and the desired amplicons subjected to another PCR reaction in 1 × buffer, 0.2 mM of each dNTP and Platinum Taq High Fidelity DNA polymerase (Invitrogen) at 94 °C for 2 min followed by a single extension at 68 °C for 5 min to add 3' adenine overhangs. Products were then purified by gel extraction (Qiagen) with each genomic half being eluted independently and T/A ligated into the pCR-XL TOPO vector (Invitrogen). The ligation reactions were used to transform TOP-10 cells, which were then plated on LB agar media containing 50 µg/ml kanamycin and cultured overnight at 30 °C. For each genomic fragment, 20 to 100 single colonies were selected and grown overnight at 30 °C with constant shaking at 225 rpm. Plasmid DNA with 5' and 3' fragments identical to the T/F sequence were excised and linearized to allow for a one step 3-piece DNA ligation with the appropriate vector.

#### *Phenotypic analyses*

Infectious molecular clones were assessed for replication competence, co-receptor usage and susceptibility to neutralization by monoclonal and polyclonal antibodies and fusion inhibitors. Viral stock generation, titrations, cell infections and neutralization assays were carried out using 293T cells, TZM-bl cells and activated primary human CD4+T-cells, as previously described (Decker et al., 2005; Keele et al., 2008; Salazar-Gonzalez et al., 2009). TZM-bl cells were obtained from the NIH AIDS Research and Reference Reagent Program (catalogue #8129). Coreceptor usage was determined using two different approaches. First, replication competent IMC derived viruses were used to infect TZM-bl cells in the presence of neither, either or both of the CCR5 and CXCR4 selective inhibitors, TAK-779 and AMD3100 (NIH AIDS Research and Reference Reagent Program catalogue #4983 and #8128). Second, use of CCR5 but not CXCR4 was confirmed by Env pseudovirus infection of NP2 cells expressing CD4 and neither or either CCR5 or CXCR4. Neutralizing antibody assays were conducted independently in two laboratories at the University of Alabama at Birmingham (J.B.) and Duke University (D.C.M.) using both IMCs and Env-pseudotyped HIV-1 virus stocks as previously described (Keele et al., 2008; Salazar-Gonzalez et al., 2009; Seaman et al., 2010; Wei et al., 2003). HIVIG-B is a pool of purified IgG from HIV-1 subtype B chronically infected subjects. HIVIG-C is a pool of purified IgG from six South African subjects chronically infected by HIV-1 subtype C. High-titer, broadly neutralizing serum IgG was also purified from three South African subtype C infected subjects (SA-C62, SA-C72 and SA-C74) not included in the HIVIG-C pool and these were tested individually.

#### **Competing interests**

The authors have no competing interests to declare.

#### **Acknowledgments**

We thank Frederic Bibollet-Ruche, Katharine Bar and Jesus Salazar-Gonzalez for critical readings of early versions of the manuscript; Maria Salazar for technical assistance; Lynn Morris and Advanced Bioscience Laboratories for providing HIVIG-C IgG; the clinical and immunology cores of IAVI and the Center for HIV/AIDS Vaccine Immunology; IAVI-affiliated investigators, clinicians and study participants; and Jamie C. White and Patricia Crystal for manuscript preparation. This work was supported by grants from the NIH (AI067854 and AI100645), the Bill and Melinda Gates Foundation (37874), and the United States Agency for International Development (GPO-A-00-06-00006-00). The study and report are the responsibility of the study authors and do not necessarily reflect the views of USAID or the United States Government.

#### **Appendix A. Supporting information**

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.virol.2012.10.009>.

#### **References**

- Aasa-Chapman, M.M., Seymour, C.R., Williams, I., McKnight, A., 2006. Novel envelope determinants for CCR3 use by human immunodeficiency virus. *J. Virol.* 80, 10884–10889.

- Abrahams, M.R., Anderson, J.A., Giorgi, E.E., Seoighe, C., Mlisana, K., Ping, L.H., Athreya, G.S., Treurnicht, F.K., Keele, B.F., Wood, N., Salazar-Gonzalez, J.F., Bhattacharya, T., Chu, H., Hoffman, I., Galvin, S., Mapanje, C., Kazembe, P., Thebus, R., Fiscus, S., Hide, W., Cohen, M.S., Karim, S.A., Haynes, B.F., Shaw, G.M., Hahn, B.H., Korber, B.T., Swanstrom, R., Williamson, C., 2009. Quantitating the multiplicity of infection with human immunodeficiency virus type 1 subtype C reveals a non-poisson distribution of transmitted variants. *J. Virol.* 83, 3556–3567.
- Baeten, J.M., Chohan, B., Lavreys, L., Chohan, V., McClelland, R.S., Certain, L., Mandaliya, K., Jaoko, W., Overbaugh, J., 2007. HIV-1 subtype D infection is associated with faster disease progression than subtype A in spite of similar plasma HIV-1 loads. *J. Infect. Dis.* 195, 1177–1180.
- Balasubramanian, C., Chillemi, G., Abbate, I., Capobianchi, M.R., Rozera, G., Desideri, A., 2012. Importance of V3 loop flexibility and net charge in the context of co-receptor recognition. A molecular dynamics study on HIV gp120. *J. Biomol. Struct. Dyn.* 29, 1–13.
- Bar, K.J., Li, H., Chamberland, A., Tremblay, C., Routy, J.P., Grayson, T., Sun, C., Wang, S., Learn, G.H., Morgan, C.J., Schumacher, J.E., Haynes, B.F., Keele, B.F., Hahn, B.H., Shaw, G.M., 2010. Wide variation in the multiplicity of HIV-1 infection among injection drug users. *J. Virol.* 84, 6241–6247.
- Bar, K.J., Tsao, C.Y., Iyer, S.S., Decker, J.M., Yang, Y., Bonsignori, M., Chen, X., Hwang, K.K., Montefiori, D.C., Liao, H.X., Hraber, P., Fischer, W., Li, H., Wang, S., Sterrett, S., Keele, B.F., Ganusov, V.V., Perelson, A.S., Korber, B.T., Georgiev, I., McLellan, J.S., Pavlicek, J.W., Gao, F., Haynes, B.F., Hahn, B.H., Kwong, P.D., Shaw, G.M., 2012. Early low-titer neutralizing antibodies impede HIV-1 replication and select for virus escape. *PLoS Pathog.* 8, e1002721.
- Bhakta, S.J., Shang, L., Prince, J.L., Claiborne, D.T., Hunter, E., 2011. Mutagenesis of tyrosine and di-leucine motifs in the HIV-1 envelope cytoplasmic domain results in a loss of Env-mediated fusion and infectivity. *Retrovirology* 8, 37.
- Blish, C.A., Nedellec, R., Mandaliya, K., Mosier, D.E., Overbaugh, J., 2007. HIV-1 subtype A envelope variants from early in infection have variable sensitivity to neutralization and to inhibitors of viral entry. *AIDS* 21, 693–702.
- Church, J.D., Huang, W., Mwatha, A., Toma, J., Stawiski, E., Donnell, D., Guay, L.A., Mmiro, F., Musoke, P., Jackson, J.B., Parkin, N., Eshleman, S.H., 2008. HIV-1 tropism and survival in vertically infected Ugandan infants. *J. Infect. Dis.* 197, 1382–1388.
- Coetzer, M., Nedellec, R., Salkowitz, J., McLaughlin, S., Liu, Y., Heath, L., Mullins, J.I., Mosier, D.E., 2008. Evolution of CCR5 use before and during coreceptor switching. *J. Virol.* 82, 11758–11766.
- de Souza, M.S., Ratto-Kim, S., Chuenarom, W., Schuetz, A., Chantakulki, S., Nuntapinit, B., Valencia-Micolta, A., Thelian, D., Nitayaphan, S., Pitisuttithum, P., Paris, R.M., Kaewkungwal, J., Michael, N.L., Rerks-Ngarm, S., Mathieson, B., Marovich, M., Currier, J.R., Kim, J.H., 2012. The Thai phase III trial (RV144) vaccine regimen induces T cell responses that preferentially target epitopes within the V2 region of HIV-1 envelope. *J. Immunol.* 188, 5166–5176.
- De Wolf, F., Hogervorst, E., Goudsmit, J., Fenyo, E.M., Rubsamen-Waigmann, H., Holmes, H., Galva-Castro, B., Karita, E., Wasi, C., Sempala, S.D., et al., 1994. Syncytium-inducing and non-syncytium-inducing capacity of human immunodeficiency virus type 1 subtypes other than B: phenotypic and genotypic characteristics. WHO network for HIV isolation and characterization. *AIDS Res. Hum. Retroviruses* 10, 1387–1400.
- Decker, J.M., Bibollet-Ruche, F., Wei, X., Wang, S., Levy, D.N., Wang, W., Delaporte, E., Peeters, M., Derdeyn, C.A., Allen, S., Hunter, E., Saag, M.S., Hoxie, J.A., Hahn, B.H., Kwong, P.D., Robinson, J.E., Shaw, G.M., 2005. Antigenic conservation and immunogenicity of the HIV coreceptor binding site. *J. Exp. Med.* 201, 1407–1419.
- Derdeyn, C.A., Decker, J.M., Bibollet-Ruche, F., Mokili, J.L., Muldoon, M., Denham, S.A., Heil, M.L., Kasolo, F., Musonda, R., Hahn, B.H., Shaw, G.M., Korber, B.T., Allen, S., Hunter, E., 2004. Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. *Science* 303, 2019–2022.
- Dwivedi, S.K., Sengupta, S., 2012. Classification of HIV-1 sequences using profile hidden markov models. *PLoS One* 7, e36566.
- Fiebig, E.W., Wright, D.J., Rawal, B.D., Garrett, P.E., Schumacher, R.T., Peddada, L., Heldebrant, C., Smith, R., Conrad, A., Kleinman, S.H., Busch, M.P., 2003. Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection. *AIDS* 17, 1871–1879.
- Goonetilleke, N., Liu, M.K., Salazar-Gonzalez, J.F., Ferrari, G., Giorgi, E., Ganusov, V.V., Keele, B.F., Learn, G.H., Turnbull, E.L., Salazar, M.G., Weinhold, K.J., Moore, S., Letvin, N., Haynes, B.F., Cohen, M.S., Hraber, P., Bhattacharya, T., Borrow, P., Perelson, A.S., Hahn, B.H., Shaw, G.M., Korber, B.T., McMichael, A.J., 2009. The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection. *J. Exp. Med.* 206, 1253–1272.
- Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321.
- Guo, X., Kameoka, M., Wei, X., Roques, B., Gotte, M., Liang, C., Wainberg, M.A., 2003. Suppression of an intrinsic strand transfer activity of HIV-1 Tat protein by its second-exon sequences. *Virology* 307, 154–163.
- Haaland, R.E., Hawkins, P.A., Salazar-Gonzalez, J., Johnson, A., Tichacek, A., Karita, E., Manigart, O., Mulenga, J., Keele, B.F., Shaw, G.M., Hahn, B.H., Allen, S.A., Derdeyn, C.A., Hunter, E., 2009. Inflammatory genital infections mitigate a severe genetic bottleneck in heterosexual transmission of subtype A and C HIV-1. *PLoS Pathog.* 5, e1000274.
- Hemelaar, J., Gouws, E., Ghys, P.D., Osmanov, S., 2011. Global trends in molecular epidemiology of HIV-1 during 2000–2007. *AIDS* 25, 679–689.
- Huang, W., Eshleman, S.H., Toma, J., Fransen, S., Stawiski, E., Paxinos, E.E., Whitcomb, J.M., Young, A.M., Donnell, D., Mmiro, F., Musoke, P., Guay, L.A., Jackson, J.B., Parkin, N.T., Petropoulos, C.J., 2007. Coreceptor tropism in human immunodeficiency virus type 1 subtype D: high prevalence of CXCR4 tropism and heterogeneous composition of viral populations. *J. Virol.* 81, 7885–7893.
- Isaacman-Beck, J., Hermann, E.A., Yi, Y., Ratcliffe, S.J., Mulenga, J., Allen, S., Hunter, E., Derdeyn, C.A., Collman, R.C., 2009. Heterosexual transmission of human immunodeficiency virus type 1 subtype C: macrophage tropism, alternative coreceptor use, and the molecular anatomy of CCR5 utilization. *J. Virol.* 83, 8208–8220.
- Jiang, C., Parrish, N.F., Wilen, C.B., Li, H., Chen, Y., Pavlicek, J.W., Berg, A., Lu, X., Song, H., Tilton, J.C., Pfaff, J.M., Henning, E.A., Decker, J.M., Moody, M.A., Drinker, M.S., Schutte, R., Freel, S., Tomaras, G.D., Nedellec, R., Mosier, D.E., Haynes, B.F., Shaw, G.M., Hahn, B.H., Doms, R.W., Gao, F., 2011. Primary infection by a human immunodeficiency virus with atypical coreceptor tropism. *J. Virol.* 85, 10669–10681.
- Kaleebu, P., French, N., Mahe, C., Yirrell, D., Watera, C., Lyagoba, F., Nakiyingi, J., Rutebemberwa, A., Morgan, D., Weber, J., Gilks, C., Whitworth, J., 2002. Effect of human immunodeficiency virus (HIV) type 1 envelope subtypes A and D on disease progression in a large cohort of HIV-1-positive persons in Uganda. *J. Infect. Dis.* 185, 1244–1250.
- Kaleebu, P., Nankya, I.L., Yirrell, D.L., Shafer, L.A., Kyosiimire-Lugemwa, J., Lule, D.B., Morgan, D., Beddows, S., Weber, J., Whitworth, J.A., 2007. Relation between chemokine receptor use, disease stage, and HIV-1 subtypes A and D: results from a rural Ugandan cohort. *J. Acquir. Immune Defic. Syndr.* 45, 28–33.
- Kalia, V., Sarkar, S., Gupta, P., Montelaro, R.C., 2003. Rational site-directed mutations of the LLP-1 and LLP-2 lentivirus lytic peptide domains in the intracytoplasmic tail of human immunodeficiency virus type 1 gp41 indicate common functions in cell-cell fusion but distinct roles in virion envelope incorporation. *J. Virol.* 77, 3634–3646.
- Keele, B.F., Giorgi, E.E., Salazar-Gonzalez, J.F., Decker, J.M., Pham, K.T., Salazar, M.G., Sun, C., Grayson, T., Wang, S., Li, H., Wei, X., Jiang, C., Kirchherr, J.L., Gao, F., Anderson, J.A., Ping, L.H., Swanstrom, R., Tomaras, G.D., Blattner, W.A., Goepfert, P.A., Kilby, J.M., Saag, M.S., Delwart, E.L., Busch, M.P., Cohen, M.S., Montefiori, D.C., Haynes, B.F., Gaschen, B., Athreya, G.S., Lee, H.Y., Wood, N., Seoighe, C., Perelson, A.S., Bhattacharya, T., Korber, B.T., Hahn, B.H., Shaw, G.M., 2008. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc. Nat. Acad. Sci. USA* 105, 7552–7557.
- Keele, B.F., Li, H., Learn, G.H., Hraber, P., Giorgi, E.E., Grayson, T., Sun, C., Chen, Y., Yeh, W.W., Letvin, N.L., Mascola, J.R., Nabel, G.J., Haynes, B.F., Bhattacharya, T., Perelson, A.S., Korber, B.T., Hahn, B.H., Shaw, G.M., 2009. Low-dose rectal inoculation of rhesus macaques by SIV<sub>me660</sub> or SIV<sub>mac251</sub> recapitulates human mucosal infection by HIV-1. *J. Exp. Med.* 206, 1117–1134.
- Khoja, S., Ojwang, P., Khan, S., Okinda, N., Harania, R., Ali, S., 2008. Genetic analysis of HIV-1 subtypes in Nairobi, Kenya. *PLoS One* 3, e3191.
- Kiwanuka, N., Laeyendecker, O., Robb, M., Kigozi, G., Arroyo, M., McCutchan, F., Eller, L.A., Eller, M., Makumbi, F., Bix, D., Wabwire-Mangen, F., Serwadda, D., Sewankambo, N.K., Quinn, T.C., Wawer, M., Gray, R., 2008. Effect of human immunodeficiency virus type 1 (HIV-1) subtype on disease progression in persons from Rakai, Uganda, with incident HIV-1 infection. *J. Infect. Dis.* 197, 707–713.
- Kiwanuka, N., Robb, M., Laeyendecker, O., Kigozi, G., Wabwire-Mangen, F., Makumbi, F.E., Nalugoda, F., Kagaayi, J., Eller, M., Eller, L.A., Serwadda, D., Sewankambo, N.K., Reynolds, S.J., Quinn, T.C., Gray, R.H., Wawer, M.J., Whalen, C.C., 2010. HIV-1 viral subtype differences in the rate of CD4+ T-cell decline among HIV seroincident antiretroviral naive persons in Rakai district, Uganda. *J. Acquir. Immune Defic. Syndr.* 54, 180–184.
- Kuiken, C., Foley, B., Leitner, T., Apetrei, C., Hahn, B., Mizrachi, I., Mullins, J., Rambaut, A., Wolinsky, S., Korber, B.T., 2012. HIV sequence compendium 2011. *Theor. Biol. Biophys. Group, I.A.N.L., NM, LA-U, 10-03684.* (Ed.).
- Lambele, M., Labrosse, B., Roch, E., Moreau, A., Verrier, B., Barin, F., Roingard, P., Mammano, F., Brand, D., 2007. Impact of natural polymorphism within the gp41 cytoplasmic tail of human immunodeficiency virus type 1 on the intracellular distribution of envelope glycoproteins and viral assembly. *J. Virol.* 81, 125–140.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Willm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948.
- Lee, H.Y., Giorgi, E.E., Keele, B.F., Gaschen, B., Athreya, G.S., Salazar-Gonzalez, J.F., Pham, K.T., Goepfert, P.A., Kilby, J.M., Saag, M.S., Delwart, E.L., Busch, M.P., Hahn, B.H., Shaw, G.M., Korber, B.T., Bhattacharya, T., Perelson, A.S., 2009. Modeling sequence evolution in acute HIV-1 infection. *J. Theor. Biol.* 261, 341–360.
- Levy, D.N., Aldrovandi, G.M., Kutsch, O., Shaw, G.M., 2004. Dynamics of HIV-1 recombination in its natural target cells. *Proc. Nat. Acad. Sci. USA* 101, 4204–4209.
- Li, H., Bar, K.J., Wang, S., Decker, J.M., Chen, Y., Sun, C., Salazar-Gonzalez, J.F., Salazar, M.G., Learn, G.H., Morgan, C.J., Schumacher, J.E., Hraber, P., Giorgi, E.E., Bhattacharya, T., Korber, B.T., Perelson, A.S., Eron, J.J., Cohen, M.S., Hicks, C.B., Haynes, B.F., Markowitz, M., Keele, B.F., Hahn, B.H., Shaw, G.M., 2010. High multiplicity infection by HIV-1 in men who have sex with men. *PLoS Pathog.* 6, e1000890.

- Li, H., Stoddard, M.B., Wang, S., Blair, L.M., Giorgi, E.E., Parrish, E.H., Learn, G.H., Hraber, P., Goepfert, P.A., Saag, M.S., Denny, T.N., Haynes, B.F., Hahn, B.H., Ribeiro, R.M., Perelson, A.S., Korber, B.T., Bhattacharya, T., Shaw, G.M., 2012. Elucidation of hepatitis C virus transmission and early diversification by single genome sequencing. *PLoS Pathog.* 8, e1002880.
- Lihana, R.W., Khamadi, S.A., Kiptoo, M.K., Kinyua, J.G., Lagat, N., Magoma, G.N., Mwaui, M.M., Makokha, E.P., Onyango, V., Osman, S., Okoth, F.A., Songok, E.M., 2006. HIV type 1 subtypes among STI patients in Nairobi: a genotypic study based on partial pol gene sequencing. *AIDS Res. Hum. Retroviruses* 22, 1172–1177.
- Lole, K.S., Bollinger, R.C., Paranjape, R.S., Gadkari, D., Kulkarni, S.S., Novak, N.G., Ingersoll, R., Sheppard, H.W., Ray, S.C., 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J. Virol.* 73, 152–160.
- Lu, L., Zhu, Y., Huang, J., Chen, X., Yang, H., Jiang, S., Chen, Y.H., 2008. Surface exposure of the HIV-1 env cytoplasmic tail LLP2 domain during the membrane fusion process: interaction with gp41 fusion core. *J. Biol. Chem.* 283, 16723–16731.
- Mahlknecht, U., Dichamp, I., Varin, A., Van Lint, C., Herbein, G., 2008. NF- $\kappa$ B-dependent control of HIV-1 transcription by the second coding exon of Tat in T cells. *J. Leukoc. Biol.* 83, 718–727.
- Mansky, L.M., Temin, H.M., 1995. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J. Virol.* 69, 5087–5094.
- Nedellec, R., Coetzer, M., Shimizu, N., Hoshino, H., Polonis, V.R., Morris, L., Martenson, U.E., Binley, J., Overbaugh, J., Mosier, D.E., 2009. Virus entry via the alternative coreceptors CCR3 and FPRL1 differs by human immunodeficiency virus type 1 subtype. *J. Virol.* 83, 8353–8363.
- Neuveut, C., Scoggins, R.M., Camerini, D., Markham, R.B., Jeang, K.T., 2003. Requirement for the second coding exon of Tat in the optimal replication of macrophage-tropic HIV-1. *J. Biomed. Sci.* 10, 651–660.
- Ochsenbauer, C., Edmonds, T.G., Ding, H., Keele, B.F., Decker, J., Salazar, M.G., Salazar-Gonzalez, J.F., Shattock, R., Haynes, B.F., Shaw, G.M., Hahn, B.H., Kappes, J.C., 2012. Generation of transmitted/founder HIV-1 infectious molecular clones and characterization of their replication capacity in CD4 T lymphocytes and monocyte-derived macrophages. *J. Virol.* 86, 2715–2728.
- Palmer, S., Kearney, M., Maldarelli, F., Halvas, E.K., Bixby, C.J., Bazmi, H., Rock, D., Falloon, J., Davey Jr., R.T., Dewar, R.L., Metcalf, J.A., Hammer, S., Mellors, J.W., Coffin, J.M., 2005. Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. *J. Clin. Microbiol.* 43, 406–413.
- Parrish, N.F., Wilen, C.B., Banks, L.B., Iyer, S.S., Pfaff, J.M., Salazar-Gonzalez, J.F., Salazar, M.G., Decker, J.M., Parrish, E.H., Berg, A., Hopper, J., Hora, B., Kumar, A., Mahlokozer, T., Yuan, S., Coleman, C., Vermeulen, M., Ding, H., Ochsenbauer, C., Tilton, J.C., Permar, S.R., Kappes, J.C., Betts, M.R., Busch, M.P., Gao, F., Montefiori, D., Haynes, B.F., Shaw, G.M., Hahn, B.H., Doms, R.W., 2012. Transmitted/founder and chronic subtype C HIV-1 use CD4 and CCR5 receptors with equal efficiency and are not inhibited by blocking the integrin  $\alpha$ 4 $\beta$ 7. *PLoS Pathog.* 8, e1002686.
- Pearson, J.E., Krapivsky, P., Perelson, A.S., 2011. Stochastic theory of early viral infection: continuous versus burst production of virions. *PLoS Comput. Biol.* 7, e1001058.
- Price, M.A., Rida, W., Mwangome, M., Mutua, G., Middelkoop, K., Roux, S., Okuku, H.S., Bekker, L.G., Anzala, O., Ngugi, E., Stevens, G., Chetty, P., Amornkul, P.N., Sanders, E.J., 2012. Identifying at-risk populations in Kenya and South Africa: HIV incidence in cohorts of men who report sex with men, sex workers, and youth. *J. Acquir. Immune. Defic. Syndr.* 59, 185–193.
- Price, M.A., Wallis, C.L., Lakhi, S., Karita, E., Kamali, A., Anzala, O., Sanders, E.J., Bekker, L.G., Twesigye, R., Hunter, E., Kaleebu, P., Kayitenkore, K., Allen, S., Ruzagira, E., Mwangome, M., Mutua, G., Amornkul, P.N., Stevens, G., Pond, S.L., Schaefer, M., Papathanasopoulos, M.A., Stevens, W., Gilmour, J., 2011. Transmitted HIV type 1 drug resistance among individuals with recent HIV infection in East and Southern Africa. *AIDS Res. Hum. Retroviruses* 27, 5–12.
- Provine, N.M., Cortez, V., Chohan, V., Overbaugh, J., 2012. The neutralization sensitivity of viruses representing human immunodeficiency virus type 1 variants of diverse subtypes from early in infection is dependent on producer cell, as well as characteristics of the specific antibody and envelope variant. *Virology* 427, 25–33.
- Quakkelaar, E.D., van Alphen, F.P., Boeser-Nunnink, B.D., van Nuenen, A.C., Pantophlet, R., Schuitemaker, H., 2007. Susceptibility of recently transmitted subtype B human immunodeficiency virus type 1 variants to broadly neutralizing antibodies. *J. Virol.* 81, 8533–8542.
- Redd, A.D., Mullis, C.E., Serwadda, D., Kong, X., Martens, C., Ricklefs, S.M., Tobian, A.A., Xiao, C., Grabowski, M.K., Nalugoda, F., Kigozi, G., Laeyendecker, O., Kagaayi, J., Sewankambo, N., Gray, R.H., Porcella, S.F., Wawer, M.J., Quinn, T.C., 2012. The rates of HIV superinfection and primary HIV incidence in a general population in Rakai, Uganda. *J. Infect. Dis.* 206, 267–274.
- Rogers, J.S., Swofford, D.L., 1999. Multiple local maxima for likelihoods of phylogenetic trees: a simulation study. *Mol. Biol. Evol.* 16, 1079–1085.
- Rolland, M., Tovanabutra, S., deCamp, A.C., Frahm, N., Gilbert, P.B., Sanders-Buell, E., Heath, L., Magaret, C.A., Bose, M., Bradfield, A., O'Sullivan, A., Crossler, J., Jones, B.S., Nau, M., Wong, K., Zhao, H., Raugi, D.N., Sorensen, S., Stoddard, J.N., Maust, B.S., Deng, W., Hural, J., Dubey, S., Michael, N.L., Shiver, J., Corey, L., Li, F., Self, S.G., Kim, J., Buchbinder, S., Casimiro, D.R., Robertson, M.N., Duerr, A., McElrath, M.J., McCutchan, F.E., Mullins, J.I., 2011. Genetic impact of vaccination on breakthrough HIV-1 sequences from the STEP trial. *Nat. Med.* 17, 366–371.
- Sacktor, N., Nakasujja, N., Skolasky, R.L., Rezapour, M., Robertson, K., Musisi, S., Katabira, E., Ronald, A., Clifford, D.B., Laeyendecker, O., Quinn, T.C., 2009. HIV subtype D is associated with dementia, compared with subtype A, in immunosuppressed individuals at risk of cognitive impairment in Kampala, Uganda. *Clin. Infect. Dis.* 49, 780–786.
- Sagar, M., Laeyendecker, O., Lee, S., Gamiel, J., Wawer, M.J., Gray, R.H., Serwadda, D., Sewankambo, N.K., Shepherd, J.C., Toma, J., Huang, W., Quinn, T.C., 2009. Selection of HIV variants with signature genotypic characteristics during heterosexual transmission. *J. Infect. Dis.* 199, 580–589.
- Sagar, M., Lavreys, L., Baeten, J.M., Richardson, B.A., Mandaliya, K., Chohan, B.H., Kreiss, J.K., Overbaugh, J., 2003. Infection with multiple human immunodeficiency virus type 1 variants is associated with faster disease progression. *J. Virol.* 77, 12921–12926.
- Salazar-Gonzalez, J.F., Bailes, E., Pham, K.T., Salazar, M.G., Guffey, M.B., Keele, B.F., Derdeyn, C.A., Farmer, P., Hunter, E., Allen, S., Manigart, O., Mulenga, J., Anderson, J.A., Swanstrom, R., Haynes, B.F., Athreya, G.S., Korber, B.T., Sharp, P.M., Shaw, G.M., Hahn, B.H., 2008. Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing. *J. Virol.* 82, 3952–3970.
- Salazar-Gonzalez, J.F., Salazar, M.G., Keele, B.F., Learn, G.H., Giorgi, E.E., Li, H., Decker, J.M., Wang, S., Baalwa, J., Kraus, M.H., Parrish, N.F., Shaw, K.S., Guffey, M.B., Bar, K.J., Davis, K.L., Ochsenbauer-Jambor, C., Kappes, J.C., Saag, M.S., Cohen, M.S., Mulenga, J., Derdeyn, C.A., Allen, S., Hunter, E., Markowitz, M., Hraber, P., Perelson, A.S., Bhattacharya, T., Haynes, B.F., Korber, B.T., Hahn, B.H., Shaw, G.M., 2009. Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. *J. Exp. Med.* 206, 1273–1289.
- Seaman, M.S., Janes, H., Hawkins, N., Grandpre, L.E., Devoy, C., Giri, A., Coffey, R.T., Harris, L., Wood, B., Daniels, M.G., Bhattacharya, T., Lapedes, A., Polonis, V.R., McCutchan, F.E., Gilbert, P.B., Self, S.G., Korber, B.T., Montefiori, D.C., Mascola, J.R., 2010. Tiered categorization of a diverse panel of HIV-1 Env pseudoviruses for assessment of neutralizing antibodies. *J. Virol.* 84, 1439–1452.
- Simmonds, P., Balfe, P., Ludlam, C.A., Bishop, J.O., Brown, A.J., 1990. Analysis of sequence diversity in hypervariable regions of the external glycoprotein of human immunodeficiency virus type 1. *J. Virol.* 64, 5840–5850.
- Ssemwanga, D., Ndemi, N., Lyagoba, F., Bukonya, J., Seeley, J., Vandepitte, J., Grosskurth, H., Kaleebu, P., 2012. HIV type 1 subtype distribution, multiple infections, sexual networks, and partnership histories in female sex workers in Kampala, Uganda. *AIDS Res. Hum. Retroviruses* 28, 357–365.
- Steckbeck, J.D., Craig, J.K., Barnes, C.O., Montelaro, R.C., 2011. Highly conserved structural properties of the C-terminal tail of HIV-1 gp41 protein despite substantial sequence variation among diverse clades: implications for functions in viral replication. *J. Biol. Chem.* 286, 27156–27166.
- Tang, J., Cormier, E., Gilmour, J., Price, M.A., Prentice, H.A., Song, W., Kamali, A., Karita, E., Lakhi, S., Sanders, E.J., Anzala, O., Amornkul, P.N., Allen, S., Hunter, E., Kaslow, R.A., 2011. Human leukocyte antigen variants B\*44 and B\*57 are consistently favorable during two distinct phases of primary HIV-1 infection in sub-Saharan Africans with several viral subtypes. *J. Virol.* 85, 85894–858902.
- Vasan, A., Renjifo, B., Hertzmark, E., Chaplin, B., Msamanga, G., Essex, M., Fawzi, W., Hunter, D., 2006. Different rates of disease progression of HIV type 1 infection in Tanzania based on infecting subtype. *Clin. Infect. Dis.* 42, 843–852.
- Walker, L.M., Phogat, S.K., Chan-Hui, P.Y., Wagner, D., Phung, P., Goss, J.L., Wrinn, T., Simek, M.D., Fling, S., Mitcham, J.L., Lehrman, J.K., Priddy, F.H., Olsen, O.A., Frey, S.M., Hammond, P.W., Kaminsky, S., Zamb, T., Moyle, M., Koff, W.C., Poignard, P., Burton, D.R., 2009. Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. *Science* 326, 285–289.
- Wei, X., Decker, J.M., Wang, S., Hui, H., Kappes, J.C., Wu, X., Salazar-Gonzalez, J.F., Salazar, M.G., Kilby, J.M., Saag, M.S., Komarova, N.L., Nowak, M.A., Hahn, B.H., Kwong, P.D., Shaw, G.M., 2003. Antibody neutralization and escape by HIV-1. *Nature* 422, 307–312.
- Wilen, C.B., Parrish, N.F., Pfaff, J.M., Decker, J.M., Henning, E.A., Haim, H., Petersen, J.E., Wojcechowskyj, J.A., Sodroski, J., Haynes, B.F., Montefiori, D.C., Tilton, J.C., Shaw, G.M., Hahn, B.H., Doms, R.W., 2011. Phenotypic and immunologic comparison of clade B transmitted/founder and chronic HIV-1 envelope glycoproteins. *J. Virol.* 85, 8514–8527.
- Wyss, S., Dimitrov, A.S., Baribaud, F., Edwards, T.G., Blumenthal, R., Hoxie, J.A., 2005. Regulation of human immunodeficiency virus type 1 envelope glycoprotein fusion by a membrane-interactive domain in the gp41 cytoplasmic tail. *J. Virol.* 79, 12231–12241.
- Zwick, M.B., Jensen, R., Church, S., Wang, M., Stiegler, G., Kunert, R., Katinger, H., Burton, D.R., 2005. Anti-human immunodeficiency virus type 1 (HIV-1) antibodies 2F5 and 4E10 require surprisingly few crucial residues in the membrane-proximal external region of glycoprotein gp41 to neutralize HIV-1. *J. Virol.* 79, 1252–1261.