Virus Evolution, 2018, 4(1): vey015

doi: 10.1093/ve/vey015
Research article

Patterns of genomic site inheritance in HIV-1M inter-subtype recombinants delineate the most likely genomic sites of subtype-specific adaptation

Marcel Tongo,[1,2,*,†] Tulio de Oliveira,[1] and Darren P. Martin[3,‡]

[1]KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP), School of Laboratory Medicine and Medical Sciences, College of Health Sciences, Nelson R Mandela School of Medicine, University of KwaZulu-Natal (UKZN), 719 Umbilo Road, Durban 4001, South Africa, [2]Center of Research for Emerging and Re-Emerging Diseases (CREMER), Institute of Medical Research and Study of Medicinal Plants (IMPM), Yaoundé, Cameroon and [3]Division of Computational Biology, Department of Integrative Biomedical Sciences and Institute of Infectious Disease and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Cape Town 7925, South Africa

*Corresponding author: E-mail: marcel.tongo@gmail.com

†http://orcid.org/0000-0002-5262-892X

‡http://orcid.org/0000-0002-8785-0870

Abstract

Recombination between different HIV-1 group M (HIV-1M) subtypes is a major contributor to the ongoing genetic diversification of HIV-1M. However, it remains unclear whether the different genome regions of recombinants are randomly inherited from the different subtypes. To elucidate this, we analysed the distribution within 82 circulating and 201 unique recombinant forms (CRFs/URFs), of genome fragments derived from HIV-1M Subtypes A, B, C, D, F, and G and CRF01_AE. We found that viruses belonging to the analysed HIV-1M subtypes and CRF01_AE contributed certain genome fragments more frequently during recombination than other fragments. Furthermore, we identified statistically significant hot-spots of Subtype A sequence inheritance in genomic regions encoding portions of Gag and Nef, Subtype B in Pol, Tat and Env, Subtype C in Vif, Subtype D in Pol and Env, Subtype F in Gag, Subtype G in Vpu-Env and Nef, and CRF01_AE inheritance in Vpu and Env. The apparent non-randomness in the frequencies with which different subtypes have contributed specific genome regions to known HIV-1M recombinants is consistent with selection strongly impacting the survival of inter-subtype recombinants. We propose that hotspots of genomic region inheritance are likely to demarcate the locations of subtype-specific adaptive genetic variations.

Key words: HIV-1; recombination; CRF/URF; permutation-based test; selection

1. Introduction

One of the main characteristics of HIV-1 group M (HIV-1M) is its high genetic diversity, both at the level of single infected individuals, and at the level of the global epidemic. This genetic diversity is generated mostly by the high rates of both mutation and recombination events that occur during HIV-1M replication (Jetzt et al. 2000; Rhodes et al. 2003). Although recombination events likely occur during most HIV replication cycles, the recombinant progeny genomes that are generated will only be detectable as such if their two parental genomes were genetically non-identical. If the parental genomes belonged to different HIV-1M subtypes, the resultant chimeric genomes are called inter-subtype recombinants. When an HIV inter-subtype recombinant is found infecting at least three individuals who have no immediate epidemiological linkage with one another, it is called a circulating recombinant form (CRF); otherwise it is called a unique recombinant form (URF) (LANL 2015). CRFs and URFs have been primarily found in geographical regions where multiple subtypes are co-circulating. To date, at least 80 CRFs are known to be circulating in different parts of the world (LANL 2015), with two of these—CRFs 01_AE and 02_AG—accounting for ∼13% of all HIV-1M infections worldwide (Hemelaar et al. 2011).

Recombination is also a major mechanism for maintaining genetic diversity of other viruses. Many herpesviruses, and particularly Herpes Simplex Virus type 1 (HSV-1), undergo frequent recombination throughout their genomes; a process which increases the rate of adaptive evolution in response to changing environments and vaccine-induced immune responses (Bowden et al. 2004; Muylaert et al. 2011; Norberg et al. 2011; Szpara et al. 2014). Further, portions of HSV-1 genomes have been found in circulating HSV-2 strains, also suggesting the possibility of HSV inter-species recombination (Koelle et al. 2017). Previous findings have also suggested that influenza A strains are highly recombinogenic. Influenza viruses have segmented RNA genomes and a major consequence of this genome structure is the capacity for genome segments to be exchanged among different viral strains. The viral diversity pool generated through reassortment plays an important role in the evolution of these viruses (Ghedin et al. 2005; Nelson et al. 2006, 2008; Schweiger et al. 2006; Rambaut et al. 2008).

Despite the existence of many known HIV-1M recombinant forms, only a small fraction of these are known to be epidemiologically important (Rodgers et al. 2017). This suggests that most inter-subtype recombinants may have a low degree of fitness that prevents them from spreading to a degree where they are detectably circulating. Also noteworthy, is the absence of any known CRFs between co-circulating Subtypes A and F (which co-circulate in the Congo basin), Subtypes C and F (which co-circulate in Brazil), and Subtypes A and C (which co-circulate in east Africa). Assuming that coinfections with these subtypes do occasionally occur, this suggests that there may be selective or mechanistic barriers to viruses in these subtypes either recombining or spreading. Examples of mechanistic barriers to recombination could be either genetic incompatibilities that prevent the co-packaging of viruses belonging to particular subtypes, or when co-packaged genomes have degrees of nucleotide sequence conservation that are not high enough to facilitate template switching during reverse transcription (Magiorkinis et al. 2003; Baird et al. 2006; Archer et al. 2008). Alternatively, selective barriers to recombination would arise when recombination between certain subtype pairings tends to yield low-fitness progeny genomes (Fan et al. 2007; Golden et al. 2014).

Analyses of large numbers of recombinant HIV-1M CRFs and URFs for which full genome sequences have been determined have enabled the mapping of recombination breakpoint distributions across the HIV-1M genome (Fan et al. 2007; Minin et al. 2007; Simon-Loriere et al. 2009). These distributions appear non-random (Simon-Loriere et al. 2010; Golden et al. 2014; Woo et al. 2014) with notable recombination breakpoint hot-spots occurring near the 5' and 3' ends of *env* and cold-spots occurring within the *gp120* encoding region of this gene (Fan et al. 2007). Although such studies have illuminated why particular sites in known HIV-1M genomes might be more, or less, prone to recombination (Simon-Loriere et al. 2010; Golden et al. 2014; Woo et al. 2014), they have not determined the relative frequencies with which nucleotide sequences in different parts of recombinant genomes have been derived from parental genomes belonging to the different HIV-1M subtypes. Such patterns could yield useful insights into the relative fitness values of particular genomic components. If, for example, it is found that Subtype A parental viruses tended to more frequently contribute *gag* genes to recombinant offspring than could be accounted for by chance, this might indicate that the Subtype A *gag* gene is in some way superior to those of viruses in the other HIV-1 subtypes.

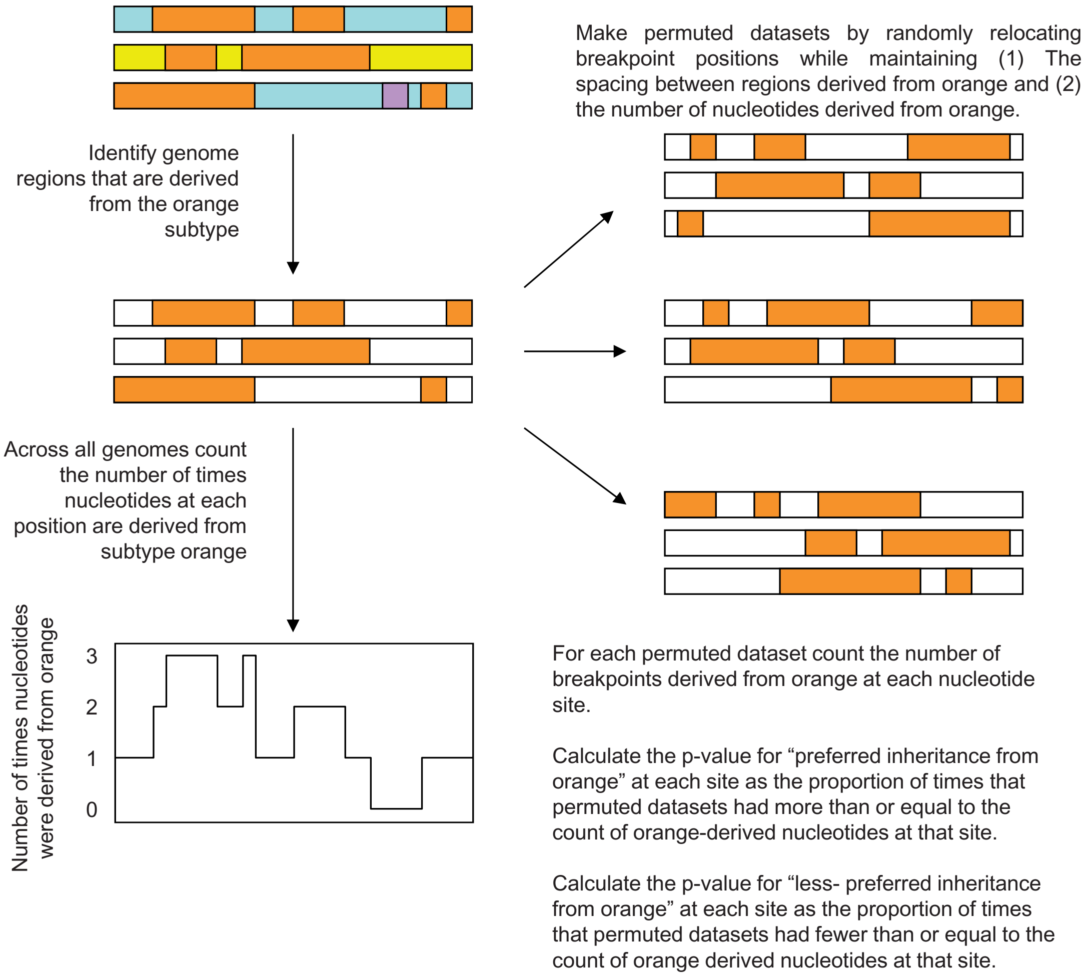
Several studies have suggested that recombination can play a role both in the enhancement of HIV-1M pathogenesis, and in the facilitation of immune evasion (Liu et al. 2002; Koulinska et al. 2006; Labrosse et al. 2006; Nora et al. 2007; Streeck et al. 2008; Shi et al. 2010; Nishimura et al. 2011). For example, in a Tanzanian study cohort, inter-subtype recombinants between Subtypes A, C, or D were apparently more transmissible through breast-feeding than were pure Subtypes A, C, and D viruses (Koulinska et al. 2006). In addition, under strong immune or drug pressures, detectable immune or drug escape variants are frequently recombinant (Nora et al. 2007; Streeck et al. 2008). Recombination and reassortment can also impact the pathogenesis of other viruses. Live-attenuated vaccines have been developed to prevent herpesvirus infections in humans and poultry (Takahashi et al. 1974; Bagust et al. 2000). Viruses in these preparations can still replicate and can therefore recombine to yield progeny genomes with elevated virulence (Lee et al. 2012). Likewise, the re-emergence of influenza A (H1N1) in humans in 2009, which was likely triggered by a reassortment event (Garten et al. 2009).

In this study, we therefore determined the distributions of Subtypes A–G and CRF 01_AE derived genome fragments within the genomic sequences of 201 URFs and 82 CRFs. We also tested whether viruses in some subtypes contribute particular genome fragments to recombinants more frequently than can be accounted for by chance under random recombination.

2. Materials and methods

2.1 Selection of sequences

We first retrieved all full-length sequences belonging to the CRFs that were available within the Los Alamos National Laboratory HIV sequence database (LANL) (LANL 2015) in October 2016. Because of the large numbers of sequences available for CRFs 01_AE and 02_AG, we selected representative sequences that included the broadest diversity within these clades (Tongo et al. 2015b). To identify the URFs, we searched in PubMed for all published papers that described any recombinant form of HIV-1M containing Subtypes A–G, or CRF 01_AE, using the terms ‘HIV-1 group M diversity’ or ‘HIV-1 group M (HIV-1M) recombination’. We then used the accession number of the respective recombinant viruses described in the literature to retrieve their genome sequences from Genbank. Finally, a representative selection of near full-length sequences from each of the other eight ‘pure’ HIV-1M subtypes that we had


Identify genome regions that are derived from the orange subtype

Make permuted datasets by randomly relocating breakpoint positions while maintaining (1) The spacing between regions derived from orange and (2) the number of nucleotides derived from orange.

Across all genomes count the number of times nucleotides at each position are derived from subtype orange

For each permuted dataset count the number of breakpoints derived from orange at each nucleotide site.

Calculate the p-value for “preferred inheritance from orange” at each site as the proportion of times that permuted datasets had more than or equal to the count of orange-derived nucleotides at that site.

Calculate the p-value for “less- preferred inheritance from orange” at each site as the proportion of times that permuted datasets had fewer than or equal to the count of orange derived nucleotides at that site.

Figure 1. The permutation test used to determine whether certain nucleotides within recombinant HIV genomes have been inherited from parental viruses from particular subtypes more or less frequently than can be accounted for by chance. The nucleotides derived from different subtypes are indicated by different colours. In this case we are interested in nucleotides derived from the orange subtype.

Table 1. Genes’ location of fragments within the CRFs and URFs that have been derived from the various subtypes and CRF01_AE

Clades	Gene with fragments most frequently contributed to recombinants	Gene with fragments less frequently contributed to recombinants
A	*gag*, *RT*, *int*, *env*, *nef*	*Prot*, *Rnase*, *tat*, *env*
B	*RT*, *int*, *rev*	*p24-gag*, *Rnase*, *int*, *nef*
C	*gag*, *pol*, vif, rev,	*env*, *nef*
D	*prot*, *RT*, tat, *gp41*	*prot*, *Rnase*, *int*, *gp120*
F	*gag*, *prot*, *RT*, *int*, *gp120*	*p17*, *tat*, *vpu*, *gp41*
G	*gag-prot*, *rev*, *nef*	*pol*, *gp120*, *gp41*
CRF01_AE	*gag*, *vpu*	*pol*, *gp41*

previously used (Tongo et al. 2016) were also added to this dataset. The full-length genome sequences of this dataset were aligned using MUSCLE (Edgar 2004) and manually edited. We constructed a maximum likelihood tree using Fastree (Price et al. 2009) implemented in RDP4 (Martin et al. 2015) to identify both duplicate sequences among the retrieved URFs, and URFs that clustered within known CRF sub-trees; these sequences were subsequently removed from the dataset. We included only one unique sequence representing a specific URF and one representative of each CRF for subsequent analyses.

2.2 Recombination analysis

Because the retrieved URFs were previously characterized using a variety of different phylogenetic methods, we repeated the recombination analyses for all of these using the bootscanning method (Salminen et al. 1995) implemented in Simplot (Lole et al. 1999). The previously described viruses were queried against representatives of isolates from Subtypes A–H, J, and K. They were also queried against CRF01_AE, and CRF02_AG when they were isolated in individuals originating from countries within the Congo basin; and against CRF01_AE when viruses were sampled in South East Asia (Tongo et al. 2015a,b). For each recombinant, the reliability of plot topologies was assessed by bootstrapping with 500 replicates, and a sliding window of 450 bp advancing with 50-bp increments. Genome segments within queried recombinants were assigned to a particular clade when peaks encompassing that segment suggested >70% bootstrap support for the segment clustering phylogenetically with that clade.

The positions of breakpoints bounding the recombinationally derived genome fragments of the CRF genomes that were used are those defined by the LANL HIV sequence database relative to HXB2.

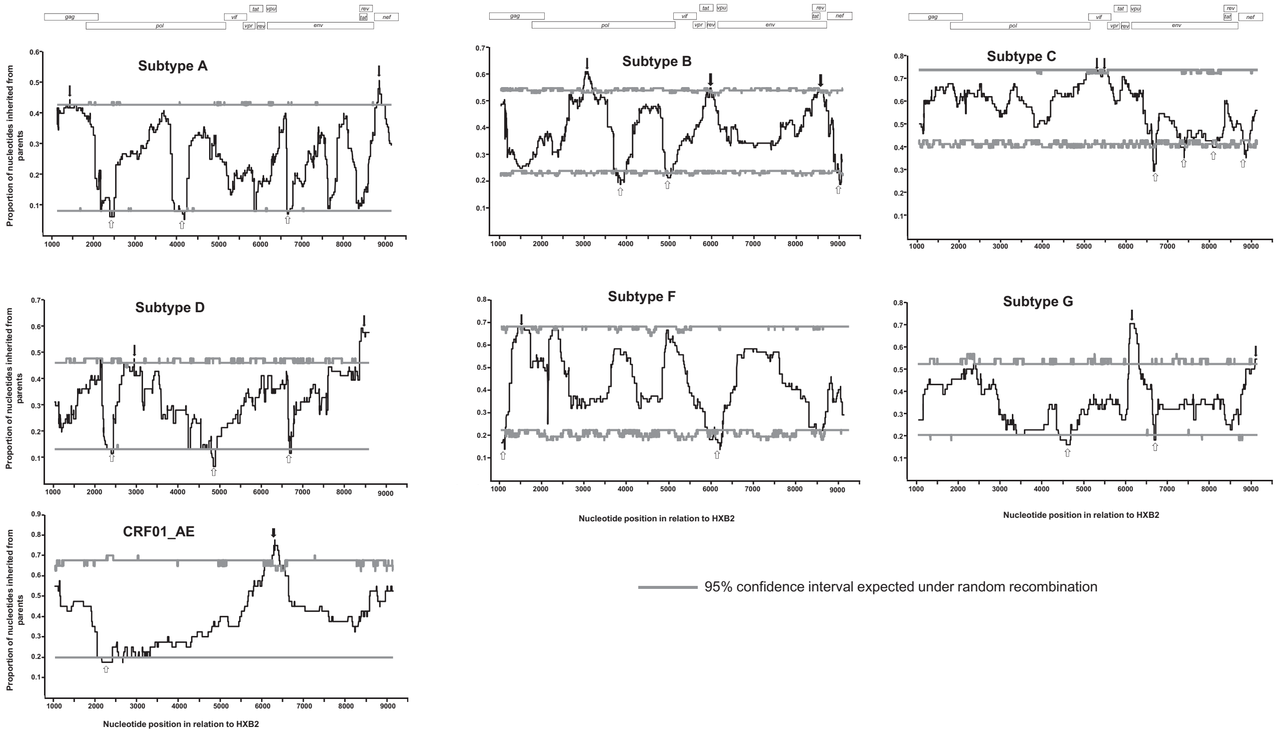


Figure 2. Distribution of subtypes- and CRF01_AE-derived genome fragments within 283 different circulating and URFs. For each nucleotide position along the genome, the proportion of recombinants that inherited a nucleotide from a specific subtype and CRF01_AE parental virus is plotted. The grey line indicates the 95th percentile bounds on analogous proportions calculated with 1,000,000 permuted datasets containing the same numbers and sizes of subtypes or CRF01_AE attributed fragments as the real datasets (but where the positions and orderings of breakpoints were randomized). Solid vertical arrows indicate hot-spots of recombinationally acquired subtypes and CRF01_AE genome regions, while unfilled vertical arrows indicate cold-spots of recombinationally acquired subtypes and CRF01_AE genome regions.

2.3 Patterns of recombinationally transferred sequences

For the recombinant fragment distribution analyses, counts were made along the alignment of the number of times individual nucleotides along the genomes of viruses in the different subtypes were inherited by recombinant genomes. To determine whether certain nucleotides were more frequently inherited by recombinant genomes from parents belonging to a particular subtype than could be accounted for by chance under random fragment exchanges, a permutation test was performed. In this test, for each of one million permutations, the positions of genome fragments derived from a given subtype within the actual recombinants were randomly shuffled while maintaining both the nucleotide spacing between the breakpoints bounding each individual fragment and, when recombinants were derived from three or more parental viruses all belonging to different subtypes, maintaining the spacing (but not the ordering) of the recombinationally derived fragments relative to one another. Instances of ‘significantly preferred inheritance’ of nucleotides at particular genome sites from a particular subtype were identified whenever the proportion of real recombinants possessing genome fragments from that subtype at those sites was higher than that observed for 95% of the permuted datasets, i.e. when the frequency of counts in the permuted datasets were larger than or equal to that of the actual dataset for any given nucleotide site along the genome. Conversely, instances of ‘significantly less preferred inheritance’ of nucleotides derived from a particular subtype were identified whenever the proportion of recombinants with nucleotides derived from that subtype at a particular genome site was lower than that determined for 95% of the permuted datasets (Fig. 1).

3. Results

We were interested in determining whether, based on presently sampled CRFs and URFs, any evidence exists of viruses belonging to particular HIV-1M lineages which contribute particular genome fragments more frequently during recombination than other genome fragments. Towards this end, we analysed the distribution of genome fragments derived from HIV-1M Subtypes A–D, F, and G and the CRF01_AE, within 82 CRF and 201 URF genomes. Genome fragments from these clades are found at high frequency in most of the recombinant sequences that are available in publically accessible sequence databases. The under-representation within recombinant genomes of sequences derived from viruses belonging to other subtypes may be due to the generally sparse sampling of viruses in the Congo Basin region (a region which, relative to anywhere else in the world, harbors a far more diverse HIV-1M epidemic in term of the numbers of circulating subtypes and recombinants). This under-representation may also be due to potential mechanistic barriers to recombination between certain subtype combinations and/or selective processes that disfavor the survival of the recombinant offspring of particular subtype pairings.

For Subtype A, only fragments belonging to the best sampled sub-subtype of this lineage, sub-subtype A1 (Tongo et al. 2018), were selected. There were 21 CRFs and 92 URFs containing Subtype A-derived genome fragments, 52 CRFs and 65 URFs containing Subtype B-derived fragments, 19 CRFs and 49 URFs containing Subtype C-derived fragments, eight CRFs and 53 URFs containing Subtype D-derived fragments, 14 CRFs and 58 URFs containing Subtype F-derived fragments, 17 CRFs and 27 URFs containing Subtype G-derived fragments, and 24 CRFs and

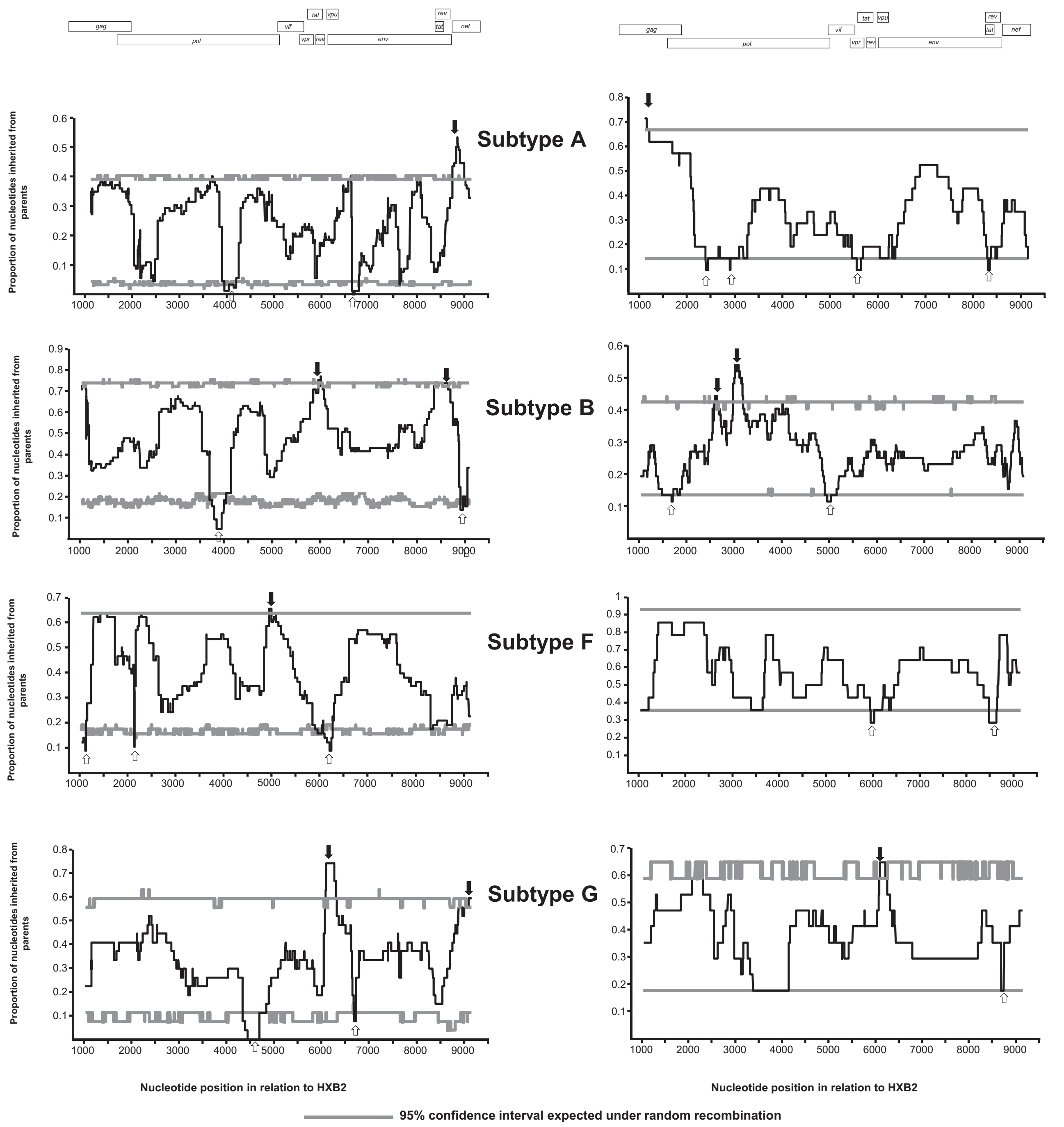


Figure 3. Distribution of subtypes-derived genome fragments within CRFs and URFs. The description is identical as in Figure 2. The left panel represent the URFs and right panel the CRFs. Shown here are data from Subtypes A, B, F, and G.

16 URFs containing CRF01_AE derived fragments. The representative plots illustrating the CRFs and URFs analysed here can be seen at the Los Alamos National Library web site (LANL 2015) and in the Supplementary Fig. S1, respectively.

Plots of the distribution of genome fragments within the CRFs and URFs that have been derived from the various subtypes and CRF01_AE revealed notable differences between these major HIV-1M lineages with respect to the genome regions that they have most or least frequently contributed to the analysed recombinants (Fig. 2 and Table 1). For instance, Subtype A-derived nucleotides were most frequently found in *gag*, *RT*, *int*, *env*, and *nef*; Subtype B-derived nucleotides in *RT*, *int*, and *rev*; Subtype C-derived nucleotides in *gag*, *pol* and accessory genes; Subtype D-derived nucleotides in *prot*, *RT*, accessory genes and *gp41*; Subtype F-derived nucleotides in *gag*, *prot*, *RT*, *int* and *gp120*; Subtype G-derived nucleotides in *gag-prot*, *rev*, and *nef*; and CRF01_AE-derived nucleotides in *gag* and *vpu* (Fig. 2 and Table 1).

Having demonstrated that there were striking differences in the locations of genomic sites that different HIV-1M subtypes and CRF01_AE contribute to recombinants, we next investigated whether there was a difference in the distribution of

Table 2. Hot- and cold-spots of HIV-1M subtypes and CRF01_AE inheritance in recombinant forms

Clades	Hot-spot		Cold-spot	
	Position[a]	Region	Position[a]	Region
A	1440	*gag p24*	2397–2495	*prot*
	8811–8916	*nef*	4147–4195	*Rnase*
			6659–6689	*env gp120*
B	2979–3192	*RT*	3747–3945	*RT-Rnase*
	5958–6008	*tat*	4927–5045	*int*
	8573–8578	*env gp41*	8949–8965	*nef*
	8606–8614		9000–9053	
C	5096–5107	*vif*	6647–6729	*env gp120*
	5112–5122		7301–7404	
	5127–5156		8069–8162	*env gp41*
	5168–5179		8818–8949	*nef*
	5189–5198			
	5204–5215			
	5293–5318			
	5456–5490			
D	2747–2768	*RT*	2397–2445	*prot*
	2771–2772		4797–4895	*int*
	2777–2781		6691–6729	*env gp120*
	8363–8586	*env gp41*		
F	1497–1550	*gag p24*	1042–1044	*gag p17*
			5948–6051	*tat*
			6079–6274	*vpu*
G	6078–6323	*vpu-env*	4447–4695	*int*
	9104–9153	*nef*	6691–6729	*env gp120*
CRF01_AE	6158–6185	*vpu*	2167–2418	*gag p6-prot*
	6234–6437	*env gp120*	2668–2673	*RT*
	6465–6472			
	6479–6496			
	6512–6545			

[a] Position relative to HXB2.

recombinationally derived genome fragments between CRFs and URFs. Although the distribution pattern of Subtypes F and G and CRF01_AE genome fragments within CRFs and URFs looked very similar, there was a marked difference between CRFs and URFs in the distribution of fragments that Subtypes A–D contributed to these (Fig. 3 and Supplementary Fig. S2).

To determine whether the peaks and troughs in the plots illustrated in Figs 1 and 2 were respectively significantly higher or significantly lower than could be accounted for by chance, we used a permutation test to identify the 95% CI bounds of the plots assuming random recombination (grey lines in Figs 2 and 3, Supplementary Fig. S2 and Table 2). This revealed statistically significant hot-spots of inheritance amongst the known recombinants for: (1) Subtype A-derived genome fragments from HXB2 genome Positions 2397 to 2495 (in *pol prot*), and from 8811 to 8916 (in *nef*); (2) Subtype B-derived genome fragments from genome Positions 2979 to 3192 (in *RT*), from 5958 to 6008 (in *tat*) and from 8573 to 8614 (in *env gp41*); (3) a Subtype C-derived genome fragment from 5096 to 5490 (in *vif*); (4) Subtype D-derived genome fragments from 2747 to 2781 (in *RT*) and from 8363 to 8586 (in *env gp41*); (5) a Subtype F-derived genome fragment from 1497 to 1550 (in *gag p24*); (6) Subtype G-derived genome fragments from 6078 to 6323 (in *vpu-env*) and from 9104 to 9153 (in *nef*); and (7) CRF01_AE derived genome fragments from 6158 to 6437 (in *vpu*) and 6234 to 6545 (in *env gp120*) (Fig. 2 and Table 2). As is shown in Table 2, regions found to contain a hot-spot of fragment exchange in one subtype were frequently identified in another subtype to contain a cold-spot of sequence exchange. For example, a statistically significant hot-spot of Subtype A sequence fragment inheritance within *nef* corresponded with a cold-spot of Subtype C sequence fragment inheritance (Fig. 2 and Table 2). When comparing statistically significant hotspots between the known CRFs and URFs, we found that hot-spots were only identified in the same position for both sets of recombinants for Subtype G and for CRF01_AE (Fig. 3, Supplementary Fig. S2 and Supplementary Table S1).

Taken together, these results indicate that there exist difference between the HIV-1M subtypes with respect to the genome regions that they tend to contribute to recombinants, and differences between CRFs and URFs with respect to the distributions of genome regions that they acquire from some subtypes.

4. Discussion

We have investigated the distribution of genome fragments that viruses in Subtypes A–D, F, and G and CRF01_AE have contributed to 283 recombinant sequences and found some evidence of non-random inheritance. It is unlikely that the genotype distribution in the HIV database accurately reflects the frequency distribution of HIV subtypes in nature. We nevertheless hypothesize that the genome regions that parental viruses in a particular subtype have tended to contribute most frequently to recombinants are most likely to be the genome regions that contain fitness determinants that have provided that subtype with selective advantages over other subtypes. Conversely, genome regions that are least frequently contributed to recombinants by members of a particular subtype might demarcate genome sites that are least likely to have provided that subtype with competitive advantages over other subtypes.

Several studies have tried to quantify differences in the relative functionality of individual genes from viruses belonging to different HIV-1M subtypes. The rationale behind such studies is that replicative capacity, which is often used as proxy of virulence, may also be a correlate of transmissibility or fitness. However, fitness can be very difficult to infer from focused functional assays in that a virus with a high replicative capacity (which is sometimes treated as synonymous with high virulence) could be less transmissible, and therefore less fit than a virus with a lower replicative capacity. An example of this comes from Uganda, where apparently less virulent Subtype A viruses have out-competed much more virulent Subtype D viruses (Blanquart et al. 2016). In addition, it has been proposed that the lower replicative capacity but increased transmissibility of Subtype C viruses, relative to those belonging to the other HIV-1M subtypes, has potentially contributed to the overwhelming predominance of Subtype C infections in the global epidemic (Renjifo et al. 2004; Abraha et al. 2009).

Nevertheless, it is likely that, based on functional studies of individual genes from viruses in different subtypes, there are subtype-specific differences in the relative fitness value of different HIV genes. For example, chimeric viruses containing a Subtype A-derived *gag-protease* exhibited lower replicative capacity than chimaeras containing a *gag-protease* derived from viruses belonging to other subtypes (Kiguoya et al. 2017). It may therefore seem counter-intuitive that we have detected a hot-spot of Subtype A derived *gag* inheritance within CRFs and URFs, unless one considers that it is optimal and not maximal replicative capacity that is a correlate of increased fitness. Our results therefore suggest that, if the Subtype A *gag-protease* is in general associated with lower replicative capacity than that of

Table 3. HIV-1 subtypes and CRF01_AE recombination inheritance associated with biological gene functions

Genes	Functional assay: effect	Subtype difference	Recombination inheritance
gag-protease	Replicative Capacity: lower replicative capacity correlate with increased fitness.	Subtype A had the lowest replicative capacity than Subtypes B and C (Kiguoya et al. 2017).	Hot-spot of Subtype A inheritance in *gag-protease*.
RT-RNase	Replicative Capacity: lower replicative capacity correlate with increased fitness.	Subtype B had a higher degree of replicative capacity than Subtype C (Iordanskiy et al. 2010).	Cold-spot of Subtype B inheritance in *RT-RNase*.
gp120 env	Cell entry efficiency.	Subtype B was superior to Subtype C (Marozsan et al. 2005).	Cold-spot of Subtype C inheritance in *gp120*.
vif	APOBEC3G degradation: this counteracts the innate antiretroviral effect of this molecule.	Subtype C had the highest activity compared with Subtypes A, B, and 01_AE and 02_AG (Iwabu et al. 2010).	Hot-spot of Subtype C inheritance in *vif*.
nef	Downregulation of CD4 and Class I HLA allele expression: this increases the pathogenesis of HIV-1M strains.	Subtype A had an intermediate degree of activity that might be optimal for fitness compared with Subtypes B and C (Mann et al. 2013).	Hotspot of Subtype A inheritance in *nef*.
rev	Efficiency of Rev–RRE dependent vector production: this decreases the translation of HIV-1 mRNAs.	Subtype G had the highest activity although non-significant (likely due to small sample size) compared with Subtypes A and CRF02_AG (Jackson et al. 2016).	Hotpot of Subtype G inheritance in *rev*.
prot	*Protease* inhibition by antiretroviral drugs	Antiretroviral drugs inhibit the A subtype proteases weaker than Subtype C (although not significant) (Velazquez-Campoy et al. 2001).	Cold-spot of Subtype A inheritance in *prot*.
vpu	Downregulation of CD4 and tetherin expression: this increases the pathogenesis of HIV-1M strains.	Subtype C had the highest activity compared with Subtypes B and C (Rahimi et al. 2017).	Hotpot of CRF01_AE inheritance in *vpu* but this clade was not included in the comparative study.

the other subtypes, then the optimal replicative capacity of HIV-1M is lower than that displayed by these other subtypes.

Further, laboratory constructed HIV-1M chimaeras containing the Subtype B *RT-RNase*, had a higher degree of overall replicative capacity than chimaeras containing the Subtype C *RT-RNase* (Iordanskiy et al. 2010). We have found a cold-spot of Subtype B inheritance in the *RT-RNase* gene, again suggesting that the increased replicative capacity afforded by this gene in Subtype B viruses may be above that required for optimal overall fitness.

In experiments focusing on cell entry efficiency instead of raw replicative capacity, chimeric genomes containing Subtype B *gp120* genes outcompeted those containing Subtype C *gp120* genes (Marozsan et al. 2005). Accordingly, we have identified a cold-spot of Subtype C inheritance in *gp120* which is consistent with the notion that Subtype C viruses may, in general, have *gp120* sequences that are less well adapted than those of viruses belonging to other HIV-1M subtypes.

Other functional assays have focused on comparing the pathogenic potential of viruses belonging to different subtypes by comparing the ability of HIV-1M genes derived from viruses in these different subtypes to impair the function of host cells. The notion that the hot-spot of Subtype C inheritance in the *vif* gene might indicate that the Subtype C *vif* gene confers this subtype with a selective advantage over other subtypes, squares well with functional comparisons of *vif* genes between the different subtypes. Such comparisons have revealed that Subtype C-derived *vif* genes are associated with more efficient APOBEC3G degradation than are Subtypes A, B, and 01_AE and 02_AG derived *vif* genes (Iwabu et al. 2010).

Similarly, the cold-spots of Subtypes C and B inheritance that we have detected in *nef* correspond with the results of functional assays. This indicates that, in comparison with *nef* genes derived from the other subtypes, the Subtypes C- and B-derived *nef* genes display the lowest and highest capacity to down-regulate CD4 and Class I HLA allele expression, respectively (Mann et al. 2013). The fact that we detected a hotspot of *nef* inheritance for Subtype A suggests that the intermediate degree of CD4 and Class I HLA down regulation that is afforded by Subtype A *nef* gene might be optimal for fitness. Table 3 summarises previously published studies characterizing the biological properties of HIV-1 genes drawn from different subtypes. These biological differences coincidentally support our hypothesis that selection may preferentially favor the survival of recombinants with particular genome fragments inherited from particular subtypes (Velazquez-Campoy et al. 2001; Marozsan et al. 2005; Iordanskiy et al. 2010; Iwabu et al. 2010; Mann et al. 2013; Jackson et al. 2016; Kiguoya et al. 2017; Rahimi et al. 2017). At the very least, these functional assays are consistent with our hypothesis that hot-spots of inheritance correspond with genome regions that provide viruses in particular subtypes with fitness advantages over viruses belonging to other subtypes.

It is probable that genome regions that have most and least often been transferred during past recombination events will continue to be the genome regions that are most and least often transferred during future recombination events. Therefore, information on recombination patterns in currently circulating viruses may be applied to the selection of viral epitopes for the generation of multi-CTL-epitope anti-HIV vaccines. For example, targeting viral epitopes within the inheritance hot-spots of Subtype A genomes will increase the probability that immune responses to these epitopes will also target the recombinant progeny of Subtype A viruses. This is important because the production of broadly protective vaccines that are suitable for use in parts of the world where multiple divergent HIV-1M lineages are circulating could prove to be the most difficult task that vaccinologists have ever encountered.

Supplementary data

Supplementary data are available at *Virus Evolution* online.

Funding

This work was supported through the Sub-Saharan African Network for TB/HIV Research Excellence (SANTHE), a DELTAS Africa Initiative [grant no. DEL-15-006]. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust [grant no. 107752/Z/15/Z] and the UK government. The views expressed in this publication are those of the author(s) and not necessarily those of AAS, NEPAD Agency, Wellcome Trust or the UK government'. This research was also supported by the Poliomyelitis Research Foundation (PRF) of South Africa. M.T. is a SANTHE Postdoctoral Fellow. D.P.M. is supported by the National Research Foundation of South Africa. T.d.O is supported a South African Medical Research Council (SA MRC) Flagship grant (MRC-RFA-UFSP-01- 2013/UKZN HIVEPI) as well by an UK Royal Society Newton Advanced Fellowship.

Conflict of interest: None declared.

References

Abraha, A. et al. (2009) 'CCR5- and CXCR4-Tropic Subtype C Human Immunodeficiency Virus Type 1 Isolates Have a Lower Level of Pathogenic Fitness than Other Dominant Group M Subtypes: Implications for the Epidemic', *Journal of Virology*, 83: 5592–605.

Archer, J. et al. (2008) 'Identifying the Important HIV-1 Recombination Breakpoints', *PLoS Computational Biology*, 4: e1000178.

Bagust, T. J. et al. (2000) 'Avian Infectious Laryngotracheitis', *Revue Scientifique et Technique (International Office of Epizootics)*, 19: 483–92.

Baird, H. A. et al. (2006) 'Sequence Determinants of Breakpoint Location during HIV-1 Intersubtype Recombination', *Nucleic Acids Research*, 34: 5203–16.

Blanquart, F. et al. (2016) 'A Transmission-Virulence Evolutionary Trade-off Explains Attenuation of HIV-1 in Uganda', *eLife*, doi: 10.7554/eLife.20492.

Bowden, R. et al. (2004) 'High Recombination Rate in Herpes Simplex Virus Type 1 Natural Populations Suggests Significant Co-Infection', *Infection, Genetics and Evolution : Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, 4: 115–23.

Edgar, R. C. (2004) 'MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput', *Nucleic Acids Research*, 32: 1792–7.

Fan, J. et al. (2007) 'The Distribution of HIV-1 Recombination Breakpoints', *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, 7: 717–23.

Garten, R. J. et al. (2009) 'Antigenic and Genetic Characteristics of Swine-Origin 2009 a(H1N1) Influenza Viruses Circulating in Humans', *Science*, 325: 197–201.

Ghedin, E. et al. (2005) 'Large-Scale Sequencing of Human Influenza Reveals the Dynamic Nature of Viral Genome Evolution', *Nature*, 437: 1162–6.

Golden, M. et al. (2014) 'Patterns of Recombination in HIV-1M Are Influenced by Selection Disfavouring the Survival of Recombinants with Disrupted Genomic RNA and Protein Structures', *PLoS One*, 9: e100400.

Hemelaar, J. et al. (2011) 'Global Trends in Molecular Epidemiology of HIV-1 during 2000-2007', *Aids (London, England)*, 25: 679–89.

Iordanskiy, S. et al. (2010) 'Subtype-Associated Differences in HIV-1 Reverse Transcription Affect the Viral Replication', *Retrovirology*, 7: 85.

Iwabu, Y. et al. (2010) 'Differential anti-APOBEC3G Activity of HIV-1 Vif Proteins Derived from Different Subtypes', *The Journal of Biological Chemistry*, 285: 35350–8.

Jackson, P. E. et al. (2016) 'Rev-RRE Functional Activity Differs Substantially among Primary HIV-1 Isolates', *AIDS Research and Human Retroviruses*, 32: 923–34.

Jetzt, A. E. et al. (2000) 'High Rate of Recombination throughout the Human Immunodeficiency Virus Type 1 Genome', *Journal of Virology*, 74: 1234–40.

Kiguoya, M. W. et al. (2017) 'Subtype-Specific Differences in Gag-Protease-Driven Replication Capacity Are Consistent with Intersubtype Differences in HIV-1 Disease Progression', *Journal of Virology*, 91: e00253-17.

Koelle, D. M. et al. (2017) 'Worldwide Circulation of HSV-2 x HSV-1 Recombinant Strains', *Scientific Reports*, 7: 44084.

Koulinska, I. N. et al. (2006) 'Risk of HIV-1 Transmission by Breastfeeding among Mothers Infected with Recombinant and Non-Recombinant HIV-1 Genotypes', *Virus Research*, 120: 191–8.

Labrosse, B. et al. (2006) 'Role of the Envelope Genetic Context in the Development of Enfuvirtide Resistance in Human Immunodeficiency Virus Type 1-Infected Patients', *Journal of Virology*, 80: 8807–19.

LANL. (2015) Los Alamos National Laboratories reference webpage for Circulating Recombinant Forms (CRFs) <http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html> cited: 19 Jan 2015.

Lee, S. W. et al. (2012) 'Attenuated Vaccines Can Recombine to Form Virulent Field Viruses', *Science (New York, N.Y.)*, 337: 188.

Liu, S. L. et al. (2002) 'Selection for Human Immunodeficiency Virus Type 1 Recombinants in a Patient with Rapid Progression to AIDS', *Journal of Virology*, 76: 10674–84.

Lole, K. S. et al. (1999) 'Full-Length Human Immunodeficiency Virus Type 1 Genomes from Subtype C-Infected Seroconverters in India, with Evidence of Intersubtype Recombination', *Journal of Virology*, 73: 152–60.

Magiorkinis, G. (2003) 'In Vivo Characteristics of Human Immunodeficiency Virus Type 1 Intersubtype Recombination: Determination of Hot Spots and Correlation with Sequence Similarity', *Journal of General Virology*, 84: 2715–22.

Mann, J. K. et al. (2013) 'Ability of HIV-1 Nef to Downregulate CD4 and HLA Class I Differs among Viral Subtypes', *Retrovirology*, 10: 100.

Marozsan, A. J. et al. (2005) 'Differences in the Fitness of Two Diverse Wild-Type Human Immunodeficiency Virus Type 1 Isolates Are Related to the Efficiency of Cell Binding and Entry', *Journal of Virology*, 79: 7121–34.

Martin, D. P. et al. (2015) 'RDP4: Detection and Analysis of Recombination Patterns in Virus Genomes', *Virus Evolution*, 1: vev003.

Minin, V. N. et al. (2007) 'Phylogenetic Mapping of Recombination Hotspots in Human Immunodeficiency Virus via Spatially Smoothed Change-Point Processes', *Genetics*, 175: 1773–85.

Muylaert, I. et al. (2011) 'Replication and Recombination of Herpes Simplex Virus DNA', *The Journal of Biological Chemistry*, 286: 15619–24.

Nelson, M. I. et al. (2006) 'Stochastic Processes Are Key Determinants of Short-Term Evolution in Influenza a Virus', *PLoS Pathogens*, 2: e125.

—— et al. (2008) 'Multiple Reassortment Events in the Evolutionary History of H1N1 Influenza a Virus since 1918', *PLoS Pathogens*, 4: e1000012.

Nishimura, Y. et al. (2011) 'Recombination-Mediated Changes in Coreceptor Usage Confer an Augmented Pathogenic Phenotype in a Nonhuman Primate Model of HIV-1-Induced AIDS', *Journal of Virology*, 85: 10617–26.

Nora, T. et al. (2007) 'Contribution of Recombination to the Evolution of Human Immunodeficiency Viruses Expressing Resistance to Antiretroviral Treatment', *Journal of Virology*, 81: 7620–8.

Norberg, P. et al. (2011) 'A Genome-Wide Comparative Evolutionary Analysis of Herpes Simplex Virus Type 1 and Varicella Zoster Virus', *PLoS One*, 6: e22527.

Price, M. N. et al. (2009) 'FastTree: Computing Large Minimum Evolution Trees with Profiles Instead of a Distance Matrix', *Molecular Biology and Evolution*, 26: 1641–50.

Rahimi, A. et al. (2017) 'In Vitro Functional Assessment of Natural HIV-1 Group M Vpu Sequences Using a Universal Priming Approach', *Journal of Virology Methods*, 240: 32–41.

Rambaut, A. et al. (2008) 'The Genomic and Epidemiological Dynamics of Human Influenza a Virus', *Nature*, 453: 615–9.

Renjifo, B. et al. (2004) 'Preferential in-Utero Transmission of HIV-1 Subtype C as Compared to HIV-1 Subtype a or D', *Aids (London, England)*, 18: 1629–36.

Rhodes, T. et al. (2003) 'High Rates of Human Immunodeficiency Virus Type 1 Recombination: Near-Random Segregation of Markers One Kilobase Apart in One round of Viral Replication', *Journal of Virology*, 77: 11193–200.

Rodgers, M. A. et al. (2017) 'Sensitive Next-Generation Sequencing Method Reveals Deep Genetic Diversity of HIV-1 in the Democratic Republic of the Congo', *Journal of Virology*, 91: e01841-16.

Salminen, M. O. et al. (1995) 'Identification of Breakpoints in Intergenotypic Recombinants of HIV Type 1 by Bootscanning', *AIDS Research and Human Retroviruses*, 11: 1423–5.

Schweiger, B. et al. (2006) 'Reassortment between Human a(H3N2) Viruses Is an Important Evolutionary Mechanism', *Vaccine*, 24: 6683–90.

Shi, B. et al. (2010) 'Evolution and Recombination of Genes Encoding HIV-1 Drug Resistance and Tropism during Antiretroviral Therapy', *Virology*, 404: 5–20.

Simon-Loriere, E. et al. (2009) 'Molecular Mechanisms of Recombination Restriction in the Envelope Gene of the Human Immunodeficiency Virus', *PLoS Pathogens*, 5: e1000418.

—— et al. (2010) 'RNA Structures Facilitate Recombination-Mediated Gene Swapping in HIV-1', *Journal of Virology*, 84: 12675–82.

Streeck, H. et al. (2008) 'Immune-Driven Recombination and Loss of Control after HIV Superinfection', *The Journal of Experimental Medicine*, 205: 1789–96.

Szpara, M. L. et al. (2014) 'Evolution and Diversity in Human Herpes Simplex Virus Genomes', *Journal of Virology*, 88: 1209–27.

Takahashi, M. et al. (1974) 'Live Vaccine Used to Prevent the Spread of Varicella in Children in Hospital', *Lancet (London, England)*, 2: 1288–90.

Tongo, M. et al. (2015a) 'Near Full-Length HIV Type 1M Genomic Sequences from Cameroon: Evidence of Early Diverging under-Sampled Lineages', *Evolution, Medicine, and Public Health*, 2015: 254–65.

—— et al. (2016) 'High Degree of HIV-1 Group M (HIV-1M) Genetic Diversity within Circulating Recombinant Forms: Insight into the Early Events of HIV-1M Evolution', *Journal of Virology*, 90: 2221–9.

—— et al. (2015b) 'Phylogenetics of HIV-1 Subtype G Env: Greater Complexity and Older Origins than Previously Reported', *Infections, Genetics and Evolution*, 35: 9–18.

—— et al. (2018) 'Unravelling the Complicated Evolutionary and Dissemination History of HIV-1M Subtype a Lineages', *Virus Evolution*, 4: vey003.

Velazquez-Campoy, A. et al. (2001) 'Catalytic Efficiency and Vitality of HIV-1 Proteases from African Viral Subtypes', *Proceedings of the National Academy of Sciences of the United States of America*, 98: 6062–7.

Woo, J. et al. (2014) 'Constraints from Protein Structure and Intra-Molecular Coevolution Influence the Fitness of HIV-1 Recombinants', *Virology*, 454-455: 34–9.