

# Coreceptor Usage, Diversity, and Divergence in Drug-Naive and Drug-Exposed Individuals from Malawi, Infected with HIV-1 Subtype C for More Than 20 Years

Ishla Seager,<sup>1</sup> Simon A. Travers,<sup>1,2</sup> Michael D. Leeson,<sup>1</sup> Amelia C. Crampin,<sup>3,4</sup> Neil French,<sup>3-5</sup>  
Judith R. Glynn,<sup>4</sup> and Grace P. McCormack<sup>1</sup>

## Abstract

There are few cohorts of individuals who have survived infection with HIV-1 for more than 20 years, reported and followed in the literature, and even fewer from Africa. Here we present data on a cohort of subtype C-infected individuals from rural northern Malawi. By sequencing multiple clones from long-term survivors at different time points, and using multiple genotyping approaches, we show that 5 of the 11 individuals are predicted as CXCR4 using (by  $\geq 3/5$  predictors) but only one individual is predicted as CXCR4 using by all five algorithms. Using any one genotyping approach overestimates the number of predicted CXCR4 sequences. Patterns of diversity and divergence were variable between the HIV-1 long-term survivors with some individuals showing very small amounts of variation and change, and others showing a greater amount; both patterns are consistent with what has been described in the literature.

## Introduction

**H**IV/AIDS CONTINUES TO BE one of the most significant infectious diseases globally. The median time to death without treatment is 10 years, but progression is highly variable, being faster in older individuals.<sup>1,2</sup> Some patients can remain asymptomatic for over 15 years while maintaining a healthy immune system and remaining therapy naive. Those that maintain a high CD4<sup>+</sup> T cell count are known as long-term nonprogressors (LTNPs) while those that show a decline in CD4<sup>+</sup> T cell numbers over an extended period of time are known as long-term survivors (LTS) or slow progressors<sup>3</sup> and very little is known about what allows for natural control of the HIV-1 virus in these individuals.<sup>4</sup> Some evidence suggests that most LTNPs are infected with a fully pathogenic replication-competent virus<sup>5,6</sup> but several other studies have linked viral attenuation to mutations in particular viral genes such as *env*, *gag*, *nef*, *vpu*, *vif*, *rev*, and *tat*.<sup>7-13</sup>

People who survive for an extended period while infected with HIV-1 serve as important models for effective immunologic control of this virus, and can provide clues to natural therapeutics and therapeutic vaccines. There are few cohorts of survivors who have been followed for 20 years and those that have survived are

found in Europe, America, and Australia, with the majority focused on individuals infected with subtype B.<sup>14-18</sup> Very little information is available in the literature on nonprogression in Africa and even less information on subtype C [e.g., Tzitzivacos *et al.*<sup>19</sup> in children and Archary *et al.*<sup>20</sup> (in adults)] with the average follow-up being less than 2 years. Laeyendecker *et al.*<sup>21</sup> and Fang *et al.*<sup>22</sup> studied long-term survivors in Uganda and Nairobi, respectively; however, neither of these studies was based on individuals infected with subtype C. An LTS cohort in Karonga District Malawi provides a unique opportunity to investigate viral patterns in a group of primarily subtype C LTS in sub-Saharan Africa who have been followed for over 20 years.<sup>23-25</sup> The extensive period of follow-up on the LTS in Karonga has allowed us to examine changes in the subtype C viral population within some individuals over a 20-year period.

In the majority of HIV-1 infections with subtype B, CCR5 tropic viruses are found to predominate during early infection.<sup>26-28</sup> In approximately 50% of subtype B infections, CXCR4 usage subsequently emerges and is often associated with accelerated loss in CD4<sup>+</sup> T cells and progression to clinical AIDS.<sup>27, 29-31</sup> Studies on subtype C have reported that viral isolates almost exclusively use the CCR5 coreceptor with

<sup>1</sup>Molecular Evolution and Systematics Laboratory, Zoology, Ryan Institute, School of Natural Sciences, National University of Ireland Galway, Galway, Ireland.

<sup>2</sup>South African National Bioinformatics Institute, SA Medical Research Council Bioinformatics Unit, University of the Western Cape, Bellville, South Africa.

<sup>3</sup>Karonga Prevention Study, Chilumba, Malawi.

<sup>4</sup>Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom.

<sup>5</sup>Institute of Infection & Global Health, University of Liverpool, Liverpool, United Kingdom.

CXCR4 usage being observed only very rarely even within individuals with more advanced disease progression and who have advanced to AIDS. Despite the fact that subtype C represents over 50% of HIV-1 infections worldwide, relatively few CXCR4-utilizing subtype C viruses had been isolated up to 2008.<sup>32–39</sup> More recently, two studies found that approximately 30% of subtype C viral isolates retrieved from individuals with advanced disease could efficiently utilize CXCR4 *in vitro*.<sup>40,41</sup> Here, we used a genotyping approach to investigate coreceptor usage among the LTS in Karonga over multiple time periods, to identify any switch in coreceptor usage as some individuals began to exhibit signs of progression.

## Materials and Methods

### Patients and samples

LTS were visited four times between the 1980s and 2010. Of nearly 200 individuals identified as HIV positive during population surveys in Karonga District Malawi in the 1980s, 38 were alive and agreed to give a blood sample in the 1990s. In 2004, 17 were reported to be still alive but not all individuals were seen or consented to the study and therefore dried blood spot samples were collected only from 10. In 2010 13 individuals were sought: two had died, one had left, and one refused to participate, so nine whole blood samples

TABLE 1. SUMMARY OF SEQUENCE DATA, CORECEPTOR TROPISM, CD4<sup>+</sup> COUNTS, AND ANTIRETROVIRAL THERAPY STATUS FOR EACH LONG-TERM SURVIVOR

LTS number	Year	Consensus sequences R5/X4	Number of clones	Cloned sequences R5/X4	CD4 T cell count and ART status
LTS1	1999 2010	R5 R5	18 env, 17 gag	R5	2004–586 2010–449 No ART
LTS2	1999 2008 2010	R5 NA R5	20 env 22 env, 20 gag	X4 2R5, 20X4	ART–2009 2010–244
LTS5	1998 2010	R5 and X4? NA	22 env, 19 gag	R5	ART–2008 2010–789
LTS8	1989 1999 2004 2008 2010	R5 R5 R5 NA R5	20 env, 30 gag 39 env 23 env, 19 gag	R5 R5 R5	2004–265 2005–ART 2010–734
LTS9	1988 1999 2004	R5 R5 R5	20 env, 20 gag	R5	2004–56 Died 2008
LTS10	1989 1999 2004 2010	R5 R5 R5 X4?	20 env, 20 gag 23 env, 20 gag	R5 R5	2004–328 2010–138
LTS12	1988 1999 2004 2010	R5 R5 R5 R5	20 env, 20 gag 20 env, 19 gag	R5 R5	2004–390 2006–ART 2010–439
LTS17	1989 1998 2004	R5 R5 X4?	20 env, 20 gag	X4?	2004–452
LTS20	1988 2000 2004 2010	R5 R5 R5 R5	18 env, 19 gag	R5	ART–2006 2010–139
LTS21	1989 1999 2010	R5 R5 R5	21 env, 18 gag	R5	2004–47 2010–36
LTS22	1989 1998 2004	R5 R5 R5	19 env, 20 gag	18R5, 1X4?	2003–475 2005–Died
LTS30	2010	R5	19 env	R5	2004–675 2010–362

A different combination of three genotypic prediction tools predicted the X4?; only for LTS 2 were all five tools concordant. LTS, long-term survivor; ART, antiretroviral therapy.

were collected (Table 1). Two individuals (LTS2 and LTS8) were seen in unrelated studies in 2008 and whole blood cell pellets were available from that time also. LTS21 and LTS1 were suspected to be dual infected.<sup>23</sup> By 2010 five LTS had been placed on antiretroviral therapy (ART) (LTS2, 5, 8, 12, and 20) and two other individuals, LTS10 and LTS21 had been referred for ART due to low CD4<sup>+</sup> counts of 138 and 36 cells/mm<sup>3</sup>, respectively. Two LTS had CD4<sup>+</sup> counts greater than 200 cells/mm<sup>3</sup> (LTS1 subtype C and LTS30 unclassifiable) and have not been referred for ART (following national guidelines) (Table 1).

#### DNA extraction, PCR, and sequencing

Proviral DNA was extracted from the dried blood spot (DBS) using a QIAamp DNA Micro Kit (Qiagen) or from 200  $\mu$ l of cell pellets using a QIAamp DNA Blood Mini Kit (Qiagen). Nested polymerase chain reaction (PCR) of a 750-bp region of *gag* p17p24 and a 549-bp region of *env* C2V3 was carried out as previously described.<sup>42</sup> A consensus sequence was produced from gel-purified PCR products where possible and for some samples more than one consensus sequence was produced. Three secondary PCR products were pooled and TA cloned using the pCR2.1-TOPO Cloning Kit (Invitrogen) or using the StrataClone PCR Cloning Kit (Agilent Technologies). Approximately 20 individual clones were sequenced in one direction. Sequence chromatographs were examined for quality in Seqman 8.0.2 (DNASTAR). All clonal data were aligned to consensus sequences already available from the 1980s, 1990s, 2004, and 2010<sup>23,42</sup> in MacClade 4.0 (Sinauer Associates) and additional sequences randomly chosen from the overall Karonga dataset (20 sequences retrieved in 1990s and 20 retrieved in 2008).

#### Sequence analysis

Phylogenetic trees were reconstructed under the GTR + gamma model of DNA substitution implemented by RAxML 7.0.3<sup>43</sup> with all parameters optimized by RAxML. Confidence levels in the groupings in the phylogeny were assessed using 1,000 bootstrap replicates as part of the RAxML phylogeny reconstruction. The subtype C ancestral sequences derived by Travers *et al.*<sup>44</sup> were employed as out-groups for *env* and *gag* gene trees. In addition, multiple alignments of each gene region were assembled for each LTS individually and phylogenetic trees were reconstructed in a manner similar to that described above. Pairwise evolutionary genetic distances from nucleotide sequences were computed by PAUP\* 4.0 (D.L. Swofford, Sinauer Associates, Inc.) under Kimura's two-parameter model of evolution.<sup>45</sup> Inpatient genetic divergence was examined by estimating the genetic distance from the earliest sequence available and all subsequent time points. Inpatient genetic diversity at each available time point was also estimated. Coreceptor usage was determined for all *env* sequences for all of the LTS using five genotypic predictor tools: C-PSSM, Geno2pheno (5% FPR), CoRSeqV3-C, GPGQ, and Raymond.<sup>41,46-50</sup>

## Results

### Prediction of coreceptor tropism

LTS30, an unclassifiable subtype, was predicted as being R5 tropic. For subtype C sequences there was little congru-

ence (15%) between all five genotypic prediction approaches. CoRSeqV3-C was found to overpredict CXCR4 usage suggesting that 94% of the subtype C sequences here were X4 tropic. When this predictor was excluded, 286 R5 and 41X4 tropic sequences are predicted as such by all four remaining methods, a concordance of 80%. CPSSM also predicted a higher number of X4 tropic sequences compared to Geno2pheno, GPGQ, or Raymond methods. All *env* sequences from five of the subtype C LTS (LTS1, 8, 9, 12, and 20) were predicted to be R5 tropic at all time points for which sequence information was available (Table 1 and Supplementary Table S1; Supplementary Data are available online at [www.liebertpub.com/aid](http://www.liebertpub.com/aid)). Four individuals (LTS2, 10, 17, and 22) were indicated as having X4 tropic viruses at some point during the course of their infection by three or more of the predictors. Only one LTS (LTS2) showed CXCR4 usage by all five predictors. X4 tropic viruses constituted >90% of the sequenced viral population within two individuals (LTS2 and 17) at the time of the last sampling time point (Fig. 1).

The consensus sequences for LTS2 from 2010 and two of the cloned sequences from the same time point were predicted to use CCR5. The remaining clones from this time point were predicted to use CXCR4. The phylogenetic reconstruction (Fig. 1a) of the *env* C2-V3 fragment from LTS2 showed two viral lineages; one contained all CXCR4 tropic viruses from 2008 with the R5 consensus from 1999 as a distant sister, and the other showed a clade of CXCR4 sequences from 2010 having emerged from a CCR5 ancestral strain. LTS2 had been on ART since 2009.

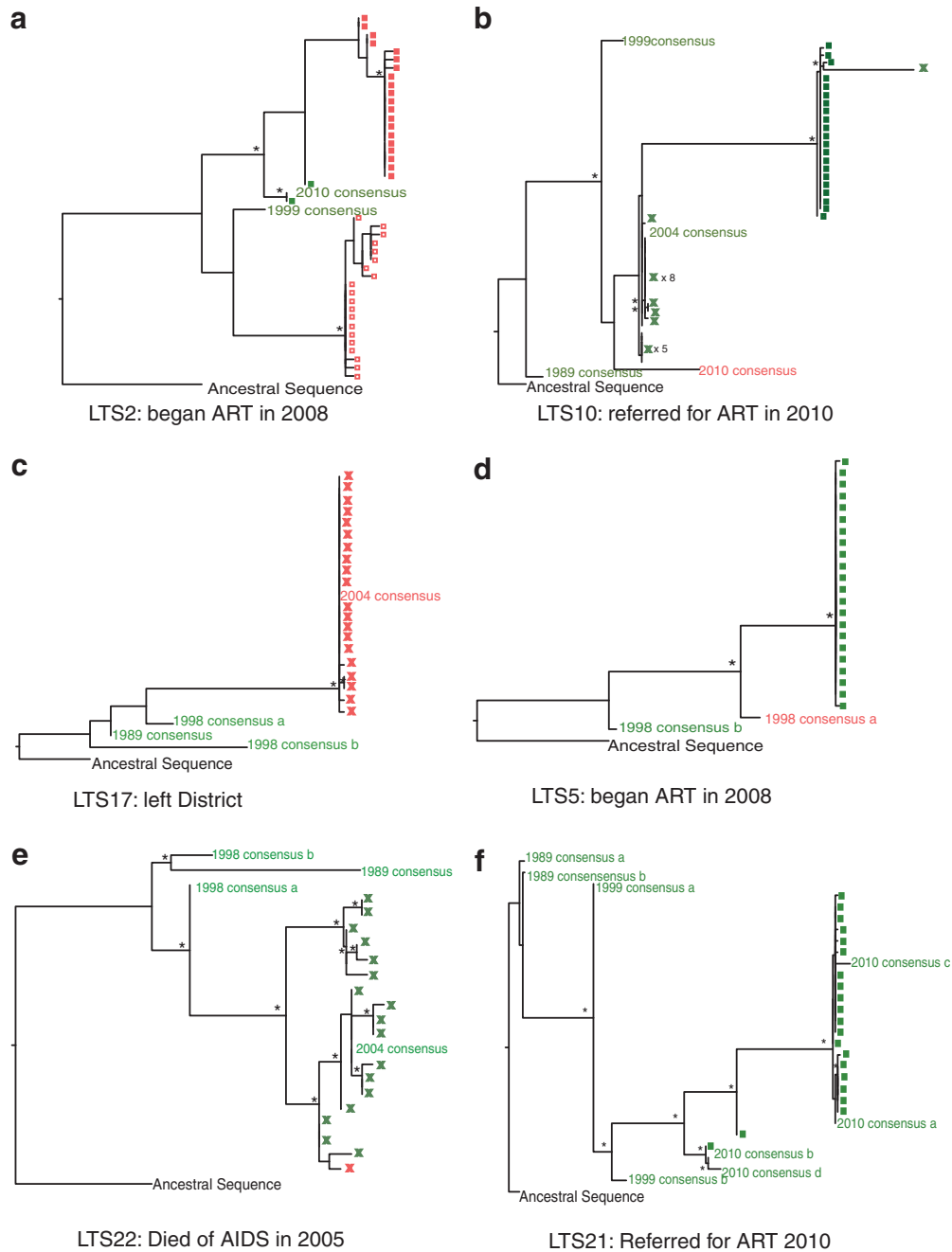
One of the *env* V3 consensus sequences retrieved from 1999 from LTS10 was predicted to be CXCR4 tropic by CPSSM-C and CoRSeqV3-C approaches as were all sequences from 2004 with the exception of one clone retrieved from the 2004 sample. This single 2004 CCR5 sequence shared a most recent common ancestor with clones from 2010 predicted as X4 by CPSSM only. The consensus sequence from 2010 was predicted as X4 also by geno2pheno and is the only sequence color coded as X4 tropic on Fig. 1b. LTS10 had a CD4<sup>+</sup> count of 138 cells/mm<sup>3</sup> in 2010 and was subsequently referred for ART.

V3 sequences obtained from LTS17 in 1998 were predicted to be CCR5 tropic (Fig. 1c) but by 2004 both the consensus sequence and 20 cloned sequences were predicted to use CXCR4 by CPSSM, CoRSeqV3-C, and the Raymond method. No further follow-up information was available for LTS17, as the individual had left the Karonga region by 2010.

For LTS5 two consensus sequences were available from 1998. Of these one sequence was predicted to use CCR5 while another was predicted to use CXCR4 (by CPSSM-C, CoRSeqV3-C, and GPGQ). Cloned sequences dating from 2010 were CCR5 tropic. V3 sequence data were available from LTS22 from 1989, 1998, and 2004 (Fig. 1e); all were predicted to utilize CCR5 except one of 19 clones from 2004 (X4 by Geno2Pheno, Raymond, and CPSSM-C approaches). The patient died of AIDS in 2005.

### Genetic diversity and divergence

Clonal sequences from within a time point from four of the nine LTS (LTS2, 5, 9, and 20) were homogenous (i.e., very similar if not identical) in both genes, and the consensus sequence from each time period grouped with the clones from that time period. The average *gag* pairwise genetic



**FIG. 1.** Individual maximum likelihood trees generated from all *env* sequences for individual long-term survivors (LTS); (a) LTS2, (b) LTS10, (c) LTS17, (d) LTS5, (e) LTS22, (f) LTS21. x represents clones dating from 2004, □ represents clones dating from the 2008, and ■ represents clones dating from 2010. Green = sequences predicted to use CCR5 and red = sequences predicted to use CXCR4. Bootstrap values of over 70 are marked with an asterisk (\*) on the relevant branches.

distances between all the clones from one time point within these individuals varied from 0% in LTS5 to 0.4% in LTS9 and 0.1–0.2% between all the clonal sequences from one time point in *env* (Tables 2 and 3). In other LTS (LTS17, 12, and 22) genetic diversity was more variable but sequences within a time period still grouped together. For LTS17 the genetic distance between *gag* clones was 1.6% while the *env* clones were homogeneous, while in LTS12 *gag* clones in each time period, and the *env* clones in 2010, were homogeneous but there was a good deal of diversity in the *env* clones in 2004.

LTS22 contained a number of variant clones in 2004 for both genes with an average pairwise genetic distance of 1.6% in *gag* and 2% in *env* (Fig. 1, Fig. 2, Tables 2 and 3).

Within LTS8, by comparison, *gag* and *env* clones did not always group with the consensus sequence of the same time period (e.g., Fig. 2d). Within LTS10 the *gag* clones from 2010 exhibited a high level of genetic diversity and the 1999, 2004, and 2010 consensus sequences (Fig. 2e) all grouped within the same clade as these clones. The clones from 2004 formed a separate but homogeneous sister clade to this with

TABLE 2. AVERAGE PAIRWISE GENETIC DIVERSITY IN *GAG* AND AVERAGE PAIRWISE DIVERGENCE (BETWEEN THE SEQUENCE FROM THE EARLIEST TIME POINT AND ALL SEQUENCES AVAILABLE FROM EACH SUBSEQUENT TIME POINT) FROM LONG-TERM SURVIVOR WHERE DATA WERE AVAILABLE

LTS number	Genetic diversity in <i>gag</i>		Genetic divergence in <i>gag</i>		
	2004	2010	1990s	2004	2010
LTS2		0.2%			2.1%
LTS5		0.0%	1.5%	1.2%	3.3%
LTS8	0.3%	1.2%	1.5%	2.2%	2.5%
LTS9	0.4%		3.3%	5.0%	
LTS10	0.2%	2.5%	3.1%	3.4%	5.5%
LTS12	0.2%	0.1%		3.1%	1.1%
LTS17	0.9%		0.7%	2.6%	
LTS20		0.1%	1.9%	3.4%	2.4%
LTS22	1.6%		2.0%	3.4%	

both lineages supported by high bootstrap values. However, the *env* 2004 and 2010 cloned sequences from LTS10 (Fig. 1b) showed very little variation within each time point with the exception of one clone from 2004 that grouped within the clones from 2010. A large amount of change had occurred between 2004 and 2010 resulting in a long branch separating the two clades with an average pairwise genetic distance of 12.2% between the two sets of clones. Seven of the nine individuals included showed an overall increase in pairwise genetic divergence between the first sampling time point in 1988/89 and the last sampling time point of either 2004 (LTS9, 17, and 22) or 2010 (LTS5, 8, 9, and 10) (Tables 2 and 3). However, in LTS12 a decrease in divergence was seen between the first sequence information from 1999 and the last sampling time point in 2010 in both genes (Tables 2 and 3).

**Discussion**

Early studies of subtype C tropism reported little if any usage of the CXCR4 coreceptor.<sup>32,34,39,51</sup> More recently,

Connell *et al.*<sup>40</sup> found that 30% of isolates from 20 subtype C-infected individuals with AIDS from South Africa were able to use CXCR4 for cell entry. Raymond *et al.*<sup>41</sup> detected CXCR4 tropic virus in 29% of 52 subtype C-infected individuals who had started ART in Malawi, while Jakobsen *et al.*<sup>52</sup> describe CXCR4 tropic virus detected in one of 21 ART-naïve individuals in Zimbabwe who had progressed to advanced stages of HIV-1 subtype C infection. In this study while CXCR4 tropic viruses were predicted in 5 out of 11 individuals by multiple genotypic prediction tools, in only one individual was X4 usage predicted by all five approaches.

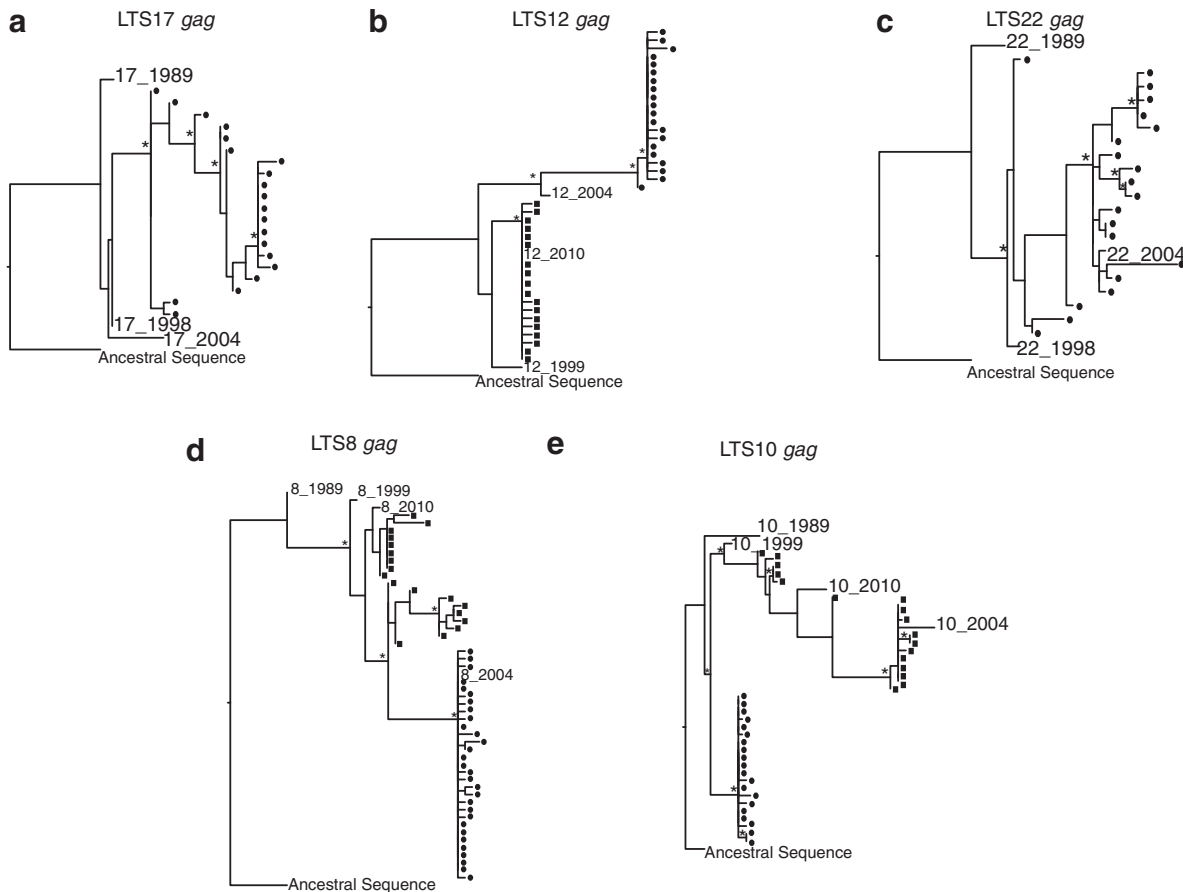
Studies on subtype C samples where both phenotypic methods and C-PSSM were used showed good concordance between the results (C-PSSM displayed a specificity of over 93% in detecting CXCR4 tropism).<sup>40,41,47</sup> In addition, the European Consensus Group on clinical management of HIV-1 tropism testing has recently indicated that genotyping approaches can be sufficient for determining resistance to CCR5 antagonists.<sup>53</sup> Due to a lack of appropriate samples at each time point phenotypic approaches could not be pursued for these individuals. This work clearly shows, however, that additional work needs to be done to elucidate the relationship between these two approaches for successful development of a genotypic predictor tool.

LTS10 was referred for ART in 2010 as the CD4<sup>+</sup> cell count had dropped to 138 cells/mm<sup>3</sup> indicating progression, which was 6 years after viruses able to use CXCR4 are predicted (but only by two genotypic predictor tools). This may suggest that a switch to CXCR4 tropism within subtype C is not as strongly correlated with disease progression, consistent with the suggestion by Meehan *et al.*<sup>54</sup> that CXCR4 usage can be transitory during disease progression. It may be that genotypic prediction by some but not all predictor tools indicates an intermediate phase or a dual phase for coreceptor usage or indeed that genotypic predictor tools are not yet sensitive enough to accurately predict coreceptor usage in subtype C. Of the six LTS who have been placed on ART, suggesting disease progression, four were CCR5 tropic, and of the two who had died one (LTS9) was at no time predicted to utilize CXCR4. However, as all of these individuals were classified as LTS, further exploration of the emergence of CXCR4 usage in nonprogressors as well as normal

TABLE 3. AVERAGE PAIRWISE GENETIC DIVERSITY IN *ENV* AND AVERAGE PAIRWISE DIVERGENCE IN *ENV* (BETWEEN THE SEQUENCE FROM THE EARLIEST TIME POINT AND ALL SEQUENCES AVAILABLE FROM EACH SUBSEQUENT TIME POINT)

LTS number	Genetic diversity in <i>env</i>			Genetic divergence in <i>env</i>			
	2004	2008	2010	1990s	2004	2008	2010
LTS2		0.6%	0.9%			4.8%	6.5%
LTS5			0.1%				5.7%, 8.5%
LTS8	2.9%	3.4%	0.0%	6.3%	10.1%	10.2%	10.5%
LTS9	0.2%			7.4%, 2.1%	7.4%		
LTS10	0.5%		0.1%	8.8%	10.5%		14.6%
LTS12	1.9%		0.1%	7.1%	7.7%		4.6%
LTS17	0.1%			3.7%, 9.1%	11.9%		
LTS20			0.2%	2.8%, 8.1%	6.2%		5.6%
LTS22	2.0%			5.9%, 5.5%	7.5%		

The two genetic divergences for LTS8, 17, 20, and 22 in the 1990s represent two different consensus sequences for that time point. The two genetic divergences for LTS5 in 2010 represent two consensus sequences available from the 1990s, which were both used to calculate the amount of genetic divergence by 2010.



**FIG. 2.** Individual maximum likelihood trees generated from all *gag* sequences for individual long-term survivors (LTS); (a) LTS17, (b) LTS12, (c) LTS22, (d) LTS8, (e) LTS10. Branches containing consensus sequences are labeled with the LTS number followed by the year of sampling. ● represents clones dating from 2004, and ■ represents clones dating from 2010.

progressors is required to fully understand the implications within subtype C-infected individuals.

Of the nine individuals infected with subtype C, who were not suspected to be dual infected, clonal sequences generated of both *env* and *gag* from four (LTS2, 5, 9, and 20) of the LTS appeared highly homogeneous with a diversity of below 1.5% in *env* and 0.5% for *gag*. There are no data available for diversity patterns within LTS; however, low levels of viral diversity have previously been recorded within LTNPS.<sup>55–58</sup> Bello *et al.*<sup>56</sup> reported a mean heterogeneity in *env* of < 1% in seven LTNPS who were in year 8 to 15 of infection. Braibant *et al.*<sup>58</sup> noted that viral diversity did not exceed 1% in three LTNPS who had been HIV positive for over 8 years with stable CD4<sup>+</sup> cell counts. Low levels of diversity have been suggested to be indicative of variants of lower fitness, which in turn has been associated with nonprogression.<sup>59,60</sup> However, the four individuals studied here who apparently harbor viral populations of extremely limited diversity have all shown signs of progression, i.e., low CD4<sup>+</sup> cell counts and placement on ART, so the low diversity detected may instead be a result of homogenization of the viral population associated with advanced disease.

It is also possible that the PCR and cloning approach may have contributed to the limited viral diversity in some individuals included in this study. This method may be influenced by reduced template detection in the early stages of PCR.

Currently it is impossible to tell to what extent experimental design may have impacted the results. Single genome analysis has been cited as a more efficient method in the detection of viral diversity.<sup>60</sup> However, Jordan *et al.*<sup>61</sup> comparing standard PCR/cloning to single genome sequencing determined that both methods are likely to provide a similar measure of viral population diversity within a given sample. Clonal sequences generated for LTS5 were almost identical in *env* and *gag*, indicating a possible lack of viral evolution and restricted viral diversity rather than PCR bottlenecking. Twenty identical clonal sequences were identified in an LTNPS 9 years after infection in a study by Bello *et al.*<sup>56</sup> with a second LTNPS producing 39 almost identical clones 17 years after infection in the same study. In LTS8, the high levels of viral diversity in *env* in 2008 (4%) were followed by a sharp decline in diversity in 2010 (0.6%) with identical clonal sequences generated. This restricted clonal diversity in LTS8 2010 was not mirrored in *gag* clonal sequences (1.2%), which points to a possible PCR bottlenecking in the 2010 *env* sample.

Higher levels of genetic diversity in *env* and *gag* were seen in LTS10 and 22, and in *env* for LTS12 and *gag* for LTS17. Across these three individuals the intraindividual genetic distance in *gag* ranged from 0.6% to 2.6%, and in *env* ranged from 1.8% to 4%, which are also comparable to diversity values described in the literature. Very few viral diversity

studies have focused on *gag*. Huang *et al.*<sup>12</sup> described an average of 1.7% diversity in full *gag* (range 0.03–3%) in eight LTNPs 12–15 years after infection, while Braibant *et al.*<sup>58</sup> described a range of viral diversity from <1% to 5.6% in *env* V1-V5 in nine LTNPs, both similar to what was recorded here.

In normal progressors, as HIV-1 infection progresses, the virus accumulates new mutations resulting in an increase in genetic divergence over time from the original infecting strain. This increase in divergence was also observed here in eight of the LTS from Karonga (LTS2, 5, 8, 9, 10, 17, 20, and 22) for whom data are available. The observed increase in divergence within these individuals indicates that evolution and replication have not been arrested in spite of the observed slow disease progression. Different rates of divergence were seen between the LTS and may be the result of the different rates of evolution of the HIV-1 virus within each individual, making it impossible to identify a pattern unique to LTS. Within two individuals (LTS12 and LTS20) a decrease in viral divergence was seen after an initial increase in divergence in 2004. LTS12 began ART in 2005 and LTS20 began ART in 2006, thus altering the environment within which the HIV-1 virus replicates and evolves. This decrease in divergence, however, may also be due to the sampling of an ancestral lineage that had existed previously as a minor variant and had remained latent, as has been described in other individuals. A similar pattern was shown by Bello *et al.*<sup>56</sup> where 7 of 16 LTNPs were seen to show slow or arrested viral divergence within the C2-V5 *env* region in conjunction with reduced viral diversity and viral loads.

Out of the 12 LTS studied, two individuals (LTS1 and LTS21) were identified as having possible dual infections.<sup>23</sup> LTS1 and LTS21 are of specific interest because both of them are still therapy free after 21 years (although LTS21 had a CD4 count of <200 cells/mm<sup>3</sup> in 2010, Table 1) and both were identified as putative dually infected individuals via phylogenetic analyses of consensus sequences.<sup>23</sup> Braibant *et al.*<sup>58</sup> reported that in a cohort of nine LTNPs three individuals, i.e., 30% ( $n=9$ ) were identified as being either co-infected or superinfected with two different strains of HIV-1. Bello *et al.*<sup>56</sup> reported that one individual within a cohort of 15 LTNPs, i.e., 6.25% ( $n=15$ ) was infected with two separate viral strains. The high level of superinfection within non-progressors may be due to the extended period of infection, increasing the possibility of reinfection occurring. Phylogenetic reconstruction of all sequences (consensus and clonal) from each LTS and 40 control sequences, for both *env* and *gag*, showed all sequences from each individual formed monophyletic clusters with the exception of these two LTS (data not shown). Further analyses (SH tests) suggests that one of the LTS (LTS21) is likely to be dually infected, but data from the other (LTS1) are also consistent with the sampling of a single diverse population. Both are females; the former individual (LTS21) was 31 in 1999 and had already had three children from two fathers. The latter individual was older (born circa 1950), had one husband, and had a low risk of superinfection. Ultradeep pyrosequencing is likely to be the best way forward for detecting multiple infections.

Studies on viral diversity of long-term survivors and non-progressors are often based on more regular sampling intervals but over shorter periods of time than we have presented here. Many are focused on small subsets of patients

using different gene regions, sampling times, frequency of sampling, and number of sequences, making it very difficult to draw any direct comparisons or detect patterns between the data described here and that described elsewhere. However, this work is consistent with both types of patterns described in the literature for diversity within LTS, i.e., some individuals showing very reduced diversity, with others showing a greater amount of diversity.<sup>55–58</sup> The inability to differentiate between methodological bottlenecks and detection of highly homogeneous populations makes it difficult to draw definitive conclusions from diversity data. Only studies employing ultradeep sequencing with the primer ID approach will determine the extent to which both the cloning and single genome amplification approaches have affected measures of diversity within hosts. This study, however, suggests a much higher proportion of CXCR4 variants within subtype C-infected individuals than previously thought, and at an earlier stage of infection. Additional unpublished data from the general population in this district also suggest a higher proportion of CXCR4-using virus than expected (Seager *et al.*, unpublished data). These observations may have implications for treatment regimes in the future in this cohort.

#### Sequence Data

All sequences have been deposited into GenBank; accession numbers (KJ738309–KJ738330, KJ809622–KJ809918, and KJ809919–KJ810193) and alignments are available from the corresponding author.

#### Acknowledgments

This material is based upon works supported by Science Foundation Ireland under grant 07/RFP/EEEE0BF424 and by an IRCSET EMBARK Scholarship to Ishla Seager (RS/2007/203). The Karonga Prevention Study is funded primarily by the Wellcome Trust, with contributions from LEPR. Permission for the study was received from the National Health Sciences Research Committee, Malawi, and the Ethics Committee of the London School of Hygiene and Tropical Medicine, UK.

#### Author Disclosure Statement

No competing financial interests exist.

#### References

1. Todd J, Glynn JR, Marston M, *et al.*: Time from HIV seroconversion to death: A collaborative analysis of eight studies in six low and middle-income countries before highly active antiretroviral therapy. *AIDS* 2007;21(Suppl. 6):S55–63.
2. Collaborative Group on AIDS Incubation and HIV Survival including the CASCADE EU Concerted Action: Time from HIV-1 seroconversion to AIDS and death before widespread use of highly-active antiretroviral therapy: A collaborative re-analysis. *Lancet* 2000;355:1131–1137.
3. Learmont J, Tindall B, Evans L, *et al.*: Long-term symptomless HIV-1 infection in recipients of blood products from a single donor. *Lancet* 1992;340:863–867.
4. Saksena NK, Wang B, and Dryer WB: Biological and molecular mechanisms in progression and non-progression of HIV disease. *AIDS Rev* 2001;3.

5. Bailey JR, Lassen KG, Yang HC, *et al.*: Neutralizing antibodies do not mediate suppression of human immunodeficiency virus type 1 in elite suppressors or selection of plasma virus variants in patients on highly active antiretroviral therapy. *J Virol* 2006;80:4758–4770.
6. Blankson J, Bailey J, Thayil S, *et al.*: Isolation and characterisation of replication-competent human immunodeficiency virus type 1 from a subset of elite suppressors. *J Virol* 2007;81:2508–2518.
7. Alexander L, Aquino-DeJesus MJ, Chan M, and Andiman WA: Inhibition of human immunodeficiency virus type 1 (HIV-1) replication by a two-amino-acid insertion in HIV-1 Vif from a nonprogressing mother and child. *J Virol* 2002;76:10533–10539.
8. Yamada T and Iwamoto A: Comparison of proviral accessory genes between long-term nonprogressors and progressors of human immunodeficiency virus type 1 infection. *Arch Virol* 2000;145:1021–1027.
9. Wang B, Ge Y, Palasanthiran P, *et al.*: Gene defects clustered at the C-terminus of the vpr gene of HIV-1 in long-term nonprogressing mother and child pair: In vivo evolution of vpr quasispecies in blood and plasma. *Virology* 1996;312:135–150.
10. Mologni D, Citterio P, Menzaghi B, *et al.*: Vpr and HIV-1 disease progression: R77Q mutation is associated with long-term control of HIV-1 infection in different groups of patients. *AIDS* 2006;20:567–574.
11. Iversen AK, Shpaer EG, Rodrigo AG, *et al.*: Persistence of attenuated rev genes in a human immunodeficiency virus type 1-infected asymptomatic individual. *J Virol* 1995;69:5743–5753.
12. Huang Y, Zhang L, and Ho DD: Characterization of gag and pol sequences from long-term survivors of human immunodeficiency virus type 1 infection. *Virology* 1998;240:36–49.
13. Cho YK and Jung YS: High frequency of gross deletions in the 5' LTR and gag regions in HIV type 1-infected long-term survivors treated with Korean red ginseng. *AIDS Res Hum Retroviruses* 2008;24:181–193.
14. Migueles SA, Osborne CM, Royce C, *et al.*: Lytic granule loading of CD8+ T cells is required for HIV-infected cell elimination associated with immune control. *Immunity* 2008;29:1009–1021.
15. Lopez M, Soriano V, Lozano S, *et al.*: No major differences in the functional profile of HIV Gag and Nef-specific CD8+ responses between long-term nonprogressors and typical progressors. *AIDS Res Hum Retroviruses* 2008;24:1185–1195.
16. Lambotte O, Boufassa F, Madec Y, *et al.*: HIV controllers: A homogeneous group of HIV-1-infected patients with spontaneous control of viral replication. *Clin Infect Dis* 2005;41:1053–1056.
17. Kloosterboer N, Groeneveld PH, Jansen CA, *et al.*: Natural controlled HIV infection: Preserved HIV-specific immunity despite undetectable replication competent virus. *Virology* 2005;339:70–80.
18. Birch MR, Learmont JC, Dyer WB, *et al.*: An examination of signs of disease progression in survivors of the Sydney Blood Bank Cohort (SBBC). *J Clin Virol* 2001;22:263–270.
19. Tzitzivacos DB, Tiemessen CT, Stevens WS, and Pathanapoulos MA: Viral genetic determinants of non-progressive HIV type 1 subtype C infection in antiretroviral drug-naïve children. *AIDS Res Hum Retroviruses* 2009;25:1141–1148.
20. Archary D, Gordon ML, Green TN, *et al.*: HIV-1 subtype C envelope characteristics associated with divergent rates of chronic disease progression. *Retrovirology* 2010;7:92.
21. Laeyendecker O, Redd AD, Lutalo T, *et al.*: Frequency of long-term nonprogressors in HIV-1 seroconverters from Rakai Uganda. *J Acquir Immune Defic Syndr* 2009;52:316–319.
22. Fang G, Kuiken C, Weiser B, *et al.*: Long-term survivors in Nairobi: Complete HIV-1 RNA sequences and immunogenetic associations. *J Infect Dis* 2004;190:697–701.
23. Seager I, Leeson MD, Crampin AC, *et al.*: HIV type 1 mutational patterns in HIV type 1 subtype C-infected long-term survivors in Karonga District Malawi: Further analysis and correction. *AIDS Res Hum Retroviruses* 2012;28:308–313.
24. McCormack GP, Glynn JR, Clewley JP, *et al.*: Emergence of a three codon deletion in gag p17 in HIV type 1 subtype C long-term survivors, and general population spread. *AIDS Res Hum Retroviruses* 2006;22:195–201.
25. Crampin AC, Floyd S, Glynn JR, *et al.*: Long-term follow-up of HIV-positive and HIV-negative individuals in rural Malawi. *AIDS* 2002;16:1545–1550.
26. van't Wout AB, Kootstra NA, Mulder-Kampinga GA, *et al.*: Macrophage-tropic variants initiate human immunodeficiency virus type 1 infection after sexual, parenteral, and vertical transmission. *J Clin Invest* 1994;94:2060–2067.
27. Xiao L, Owen SM, Goldman I, *et al.*: CCR5 coreceptor usage of non-syncytium-inducing primary HIV-1 is independent of phylogenetically distinct global HIV-1 isolates: Delineation of consensus motif in the V3 domain that predicts CCR-5 usage. *Virology* 1998;240:83–92.
28. Zhang LQ, MacKenzie P, Cleland A, *et al.*: Selection for specific sequences in the external envelope protein of human immunodeficiency virus type 1 upon primary infection. *J Virol* 1993;67:3345–3356.
29. Jekle A, Keppler OT, De Clercq E, *et al.*: In vivo evolution of human immunodeficiency virus type 1 toward increased pathogenicity through CXCR4-mediated killing of uninfected CD4 T cells. *J Virol* 2003;77:5846–5854.
30. Richman DD and Bozzette SA: The impact of the syncytium-inducing phenotype of human immunodeficiency virus on disease progression. *J Infect Dis* 1994;169:968–974.
31. Schuitemaker H, Koot M, Kootstra NA, *et al.*: Biological phenotype of human immunodeficiency virus type 1 clones at different stages of infection: Progression of disease is associated with a shift from monocytotropic to T-cell-tropic virus population. *J Virol* 1992;66:1354–1360.
32. Cecilia D, Kulkarni SS, Tripathy SP, *et al.*: Absence of coreceptor switch with disease progression in human immunodeficiency virus infections in India. *Virology* 2000;271:253–258.
33. Choge I, Cilliers T, Walker P, *et al.*: Genotypic and phenotypic characterization of viral isolates from HIV-1 subtype C-infected children with slow and rapid disease progression. *AIDS Res Hum Retroviruses* 2006;22:458–465.
34. Cilliers T, Nhlapo J, Coetzer M, *et al.*: The CCR5 and CXCR4 coreceptors are both used by human immunodeficiency virus type 1 primary isolates from subtype C. *J Virol* 2003;77:4449–4456.
35. Engelbrecht S, de Villiers T, Sampson CC, *et al.*: Genetic analysis of the complete gag and env genes of HIV type 1 subtype C primary isolates from South Africa. *AIDS Res Hum Retroviruses* 2001;17:1533–1547.



36. Michler K, Connell BJ, Venter WD, *et al.*: Genotypic characterization and comparison of full-length envelope glycoproteins from South African HIV type 1 subtype C primary isolates that utilize CCR5 and/or CXCR4. *AIDS Res Hum Retroviruses* 2008;24:743–751.
37. Morris L, Cilliers T, Bredell H, *et al.*: CCR5 is the major coreceptor used by HIV-1 subtype C isolates from patients with active tuberculosis. *AIDS Res Hum Retroviruses* 2001;17:697–701.
38. Papathanasopoulos MA, Patience T, Meyers TM, *et al.*: Full-length genome characterization of HIV type 1 subtype C isolates from two slow-progressing perinatally infected siblings in South Africa. *AIDS Res Hum Retroviruses* 2003;19:1033–1037.
39. Ping LH, Nelson JA, Hoffman IF, *et al.*: Characterization of V3 sequence heterogeneity in subtype C human immunodeficiency virus type 1 isolates from Malawi: Underrepresentation of X4 variants. *J Virol* 1999;73:6271–6281.
40. Connell BJ, Michler K, Capovilla A, *et al.*: Emergence of X4 usage among HIV-1 subtype C: Evidence for an evolving epidemic in South Africa. *AIDS* 2008;22:896–899.
41. Raymond S, Delobel P, Mavigner M, *et al.*: Prediction of HIV type 1 subtype C tropism by genotypic algorithms built from subtype B viruses. *J Acquir Immune Defic Syndr* 2010;53:167–175.
42. McCormack GP, Glynn JR, Crampin AC, *et al.*: Early evolution of the human immunodeficiency virus type 1 subtype C epidemic in rural Malawi. *J Virol* 2002;76:12890–12899.
43. Stamatakis A: RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 2006;22:2688–2690.
44. Travers SA, Clewley JP, Glynn JR, *et al.*: Timing and reconstruction of the most recent common ancestor of the subtype C clade of human immunodeficiency virus type 1. *J Virol* 2004;78:10501–10506.
45. Kimura M: A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 1980;16:111–120.
46. Jensen MA, Li FS, van 't Wout AB, *et al.*: Improved coreceptor usage prediction and genotypic monitoring of R5-to-X4 transition by motif analysis of human immunodeficiency virus type 1 env V3 loop sequences. *J Virol* 2003;77:13376–13388.
47. Jensen MA, Coetzer M, van 't Wout AB, *et al.*: A reliable phenotype predictor for human immunodeficiency virus type 1 subtype C based on envelope V3 sequences. *J Virol* 2006;80:4698–4704.
48. Lengauer T, Sander O, Sierra S, Thielen A, and Kaiser R: Bioinformatics prediction of HIV coreceptor usage. *Nat Biotechnol* 2007;25:1407–1410.
49. Cashin K, Gray LR, Jakobsen MR, *et al.*: CoRSeqv3-C: A novel HIV-1 subtype C specific V3 sequence based coreceptor usage prediction algorithm. *Retrovirology* 2013;10:24.
50. Lin N, Becerri C, Giguel F, *et al.*: Env sequence determinants in CXCR4-using human immunodeficiency virus type 1 subtype C. *Virology* 2012;433(20):296–307.
51. Bjorndal A, Sonnerborg A, Tscherning C, Albert J, and Fenyo EM: Phenotypic characteristics of human immunodeficiency virus type 1 subtype C isolates of Ethiopian AIDS patients. *AIDS Res Hum Retroviruses* 1999;15:647–653.
52. Jakobsen MR, Cashin K., Roche M, *et al.*: Longitudinal analysis of CCR5 and CXCR4 usage in a cohort of anti-retroviral therapy-naïve subjects with progressive HIV-1 subtype C infection. *PLoS One* 2013;8(6):65950.
53. Vandekerckhove LP, Wensing AM, Kaiser R, *et al.*: European guidelines on the clinical management of HIV-1 tropism testing. *Lancet Infect Dis* 2011;11:394–407.
54. Meehan CJ, Hedge JA, Robertson DL, *et al.*: Emergence, dominance, and possible decline of CXCR4 chemokine receptor usage during the course of HIV infection. *J Med Virol* 2010;82:2004–2012.
55. Wang B, Mikhail M, Dyer WB, *et al.*: First demonstration of a lack of viral sequence evolution in a nonprogressor, defining replication-incompetent HIV-1 infection. *Virology* 2003;312:135–150.
56. Bello G, Casado C, Sandonis V, *et al.*: A subset of human immunodeficiency virus type 1 long-term non-progressors is characterized by the unique presence of ancestral sequences in the viral population. *J Gen Virol* 2005;86:355–364.
57. Bello G, Casado C, Sandonis V, *et al.*: Plasma viral load threshold for sustaining intrahost HIV type 1 evolution. *AIDS Res Hum Retroviruses* 2007;23:1242–1250.
58. Braibant M, Agut H, Rouzioux C, *et al.*: Characteristics of the env genes of HIV type 1 quasispecies in long-term nonprogressors with broadly neutralizing antibodies. *J Acquir Immune Defic Syndr* 2008;47:274–284.
59. Joos B, Trkola A, Fischer M, *et al.*: Low human immunodeficiency virus envelope diversity correlates with low in vitro replication capacity and predicts spontaneous control of plasma viremia after treatment interruptions. *J Virol* 2005;79:9026–9037.
60. Mens H, Kearney M, Wiegand A, *et al.*: HIV-1 continues to replicate and evolve in patients with natural control of HIV infection. *J Virol* 2010;84:12971–12981.
61. Jordan MR, Kearney M, Palmer S, *et al.*: Comparison of standard PCR/cloning to single genome sequencing for analysis of HIV-1 populations. *J Virol Methods* 2010;168:114–120.

Address correspondence to:

Grace McCormack

Molecular Evolution and Systematics Laboratory

Zoology

Ryan Institute

School of Natural Sciences

National University of Ireland Galway

Galway

Ireland

E-mail: grace.mccormack@nuigalway.ie