HIV-1 evolution and adaptation to the host during the course of infection
Gijsbers, E.F.

Citation for published version (APA):
Gijsbers, E. F. (2013). HIV-1 evolution and adaptation to the host during the course of infection
HIV-1 ENVELOPE DIVERSITY ONE YEAR AFTER SEROCONVERSION PREDICTS SUBSEQUENT DISEASE PROGRESSION

Andrea Rachinger\textsuperscript{1#}, Neeltje A. Kootstra\textsuperscript{1#}, Esther F. Gijsbers\textsuperscript{1}, Tom L.G.M. van den Kerkhof\textsuperscript{1}, Hanneke Schuitemaker\textsuperscript{1,2} and Angélique B. van ‘t Wout\textsuperscript{1,2}

\textsuperscript{1} Department of Experimental Immunology, Sanquin Research, Landsteiner Laboratory, and Center for Infectious Diseases and Immunity Amsterdam (CINIMA), Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands
\textsuperscript{2} Present address: Crucell Holland BV, Leiden, the Netherlands

\# equal contribution

AIDS, 2012, 26:1517-1522
ABSTRACT

Objective: Recent studies have suggested that the dynamics of HIV-1 evolutionary rate reflect the rate of disease progression. We wished to determine whether viral diversity early in infection is predictive of the subsequent disease course.

Design: HIV-1 envelope diversity at seroconversion and one year thereafter from 89 homosexual participants of the Amsterdam Cohort Studies on HIV infection and AIDS was correlated with clinical endpoints and markers of disease progression.

Methods: Heteroduplex mobility assay (HMA) and sequencing followed by calculation of pairwise genetic distances were applied to determine HIV-1 envelope diversity. The HMA pattern (presence or absence of heteroduplexes) and sequence diversity were each tested for correlation with the clinical course of infection.

Results: HMA pattern at 1-year post-seroconversion was significantly associated with progression to AIDS and AIDS-related death, with presence of heteroduplexes associated with accelerated disease progression. Moreover, not only this dichotomous measure of viral diversity (absence or presence of heteroduplexes), but also genetic diversity itself was associated with disease course. HMA pattern was an independent predictor of accelerated disease progression, also when CCR5 genotype, human leukocyte antigen (HLA)-type, viral load, CD4+ T cell counts, and coreceptor use at viral load set point were included in the analysis.

Conclusions: Viral diversity early in HIV-1 infection is predictive of the subsequent disease progression. It remains to be established whether viral diversity itself plays a causal role in the increased damage to the immune system or whether it is a reflection of immune pressure or other selective forces.
INTRODUCTION
During primary HIV-1 infection specific adaptive immune responses develop, which exert varying pressure on the viral genome. Viral escape variants may emerge depending on the humoral or cellular immune pressure on the wild type variant and the relative fitness of the escape variant. In addition, other selective forces such as specific cellular or anatomical niches may shape the viral quasispecies. Studies of viral diversity and divergence in HIV-infected patients with different disease progression rates have had conflicting results. Reports suggested an inverse relationship between viral evolution and disease progression [1,2] and between viral evolution and CD4+ T cell decline [3-6]. However, others have reported a positive relationship, suggesting that immune selection suppresses viral evolution in nonprogressors [7,8]. Moreover, not only overall evolution, but also positively selected sites and adaptation rates in env have been positively associated with disease progression [9,10]. In addition, it has been suggested that the level of genetic diversity that can be controlled by the host immune system is limited, and that exceeding a diversity threshold may be a key factor for disease progression [11]. More recently, Lee et al. [12] tried to resolve the conflicting findings by showing that the rate of intrahost HIV-1 evolution does not remain constant within a single infected individual, but rather slowed down at a rate correlated with the rate of CD4+ T cell decline. Thus, rather than using average rates, the dynamics of the evolutionary rate reflects the rate of disease progression.

What remains to be resolved is whether viral diversity, especially early in infection, is predictive of the subsequent disease course. To address this question we evaluated env viral diversity by heteroduplex mobility assay (HMA) and sequencing in serum samples from HIV-1 seroconverters taken around the moment of seroconversion and approximately 12 months thereafter.

METHODS
Study patients and selected samples
HIV-positive men who have sex with men (MSM), participants in the prospective Amsterdam Cohort Studies on HIV Infection and AIDS (ACS; for details see supplemental information), with a documented date of seroconversion between 1985 and 1997 were included in this study (n=141). Seroconversion was calculated as the mid-point between first positive and last negative sample, with sampling at approximately 3-monthly intervals. Patients with sera of detectable viral load (>1000 copies/ml plasma) available at seroconversion and at 1-year post-seroconversion were analyzed (n=95) [13].

The first and second time point was at a median of 2 and 14 months after seroconversion, respectively (range: 4 months before to 8 months after seroconversion and 8 to 19 months after seroconversion, respectively). Median difference between the two time points was 12 months (range: 6 to 21 months). Longitudinal CD4+ T cell count and viral load data were available for all 89 patients studied by HMA (see below).

RNA isolation, RT-PCR, PCR amplification, molecular cloning, and sequencing of HIV-1 envelope C2-C4 and HMA
Viral RNA was isolated from 140µl serum (Qiamp Viral Mini Kit, Qiagen, Hilden, Germany) and eluted in 50µl as described previously [13]. Briefly, viral RNA was reverse transcribed
HMA PATTERN PREDICTS HIV DISEASE PROGRESSION

(SuperScript First-strand synthesis system, Invitrogen, Carlsbad, California, USA), followed by PCR amplification (GoTaq Flexi DNA Polymerase, Promega, Madison, Wisconsin, USA), molecular cloning of multiple PCR products (pGEM T Easy Vector System, Promega, Madison, Wisconsin, USA) and sequencing (Big Dye Terminator Cycle Sequencing kit, v1.1 Applied Biosystems, Foster City, California, USA) [13] (primers see supplemental information). PCR fragments of env C2-C4 from both time points could be generated for 89 out of 95 patients and were studied for heterogeneity using HMA: When HIV-1 gene sequences are amplified from patient samples, related DNA products co-amplified from divergent templates can randomly re-anneal to form heteroduplexes that migrate with reduced mobility in neutral polyacrylamide gel as compared to the homoduplexes of re-annealed identical/highly similar templates [14]. Thus, HMA allows rapid quantitation of genetic relationships and sensitive detection of minor variants in complex quasispecies. Homoduplexes and heteroduplexes were generated in 5µl second-round PCR product of each time point separately from each patient, transferred to wet ice and resolved on a 5% non-denaturing polyacrylamide gel [14]. Of the 89 patients with positive PCR reaction at both time points, 44 samples from the first time point showed only homoduplexes and 45 samples showed both homoduplexes and heteroduplexes. For the second time point, 37 samples showed only homoduplexes and 52 samples also showed heteroduplexes.

Second-round PCR products from both time points for 74 of these 89 patients were cloned, and 3–21 colonies (average 9) per time point were sequenced as described in [13] (additional details in supplemental information). Pairwise genetic distance between clonal sequences within each sample was calculated using Mega applying the Kimura 2-Parameter (http://www.megasoftware.nex). Samples were also ranked based on this diversity and divided into three groups (low, medium or high diversity) of ~25 patients each, as a measure of relative envelope diversity.

HIV-1 envelope nucleotide sequences are available from GenBank (accession numbers JQ902168 to JQ903531).

**Determination of CD4+ T cell counts, plasma HIV viral load, HIV coreceptor use, CCR5 genotype and HLA-type**

Enumeration of CD4+ T cells and subsets was done using standard flow cytometry [15], serum viral load was measured using a quantitative HIV-1 RNA nucleic acid-based sequence amplification (Organon Teknika, Boxtel, the Netherlands) [16]. Viral load data were analyzed after log_{10} transformation. HIV-1 coreceptor usage was determined in real-time using the MT-2 assay [17]. CCR5 genotype was determined by PCR restriction fragment length analysis [18]. Typing of HLA class I loci was performed using serology [19] or sequence-specific primer PCR [20].

**Clinical endpoints**

When AIDS according to the CDC 1993 definition [21] was used as an endpoint in Kaplan-Meier survival analysis, 55 of 89 individuals had an event, 12 were censored due to loss of follow-up, and 22 were censored because of initiation of HAART. When AIDS-related death was used as an endpoint, 41 individuals had an event, 16 were censored due to loss of follow-up, and 32 were censored at the moment HAART was initiated. For the survival analysis in which a CD4+ T cell count of 400 cells/µl was used as an endpoint, 78 individuals had an event, six were censored due
to loss of follow-up, and five were censored because of initiation of HAART. For survival analysis after AIDS diagnosis, 55 individuals were included of whom 37 individuals had an event, eight were censored due to loss of follow-up, and 10 were censored because of initiation of HAART.

RESULTS

Association between HMA pattern or relative sequence diversity and subsequent disease progression

The association of viral envelope HMA pattern or relative sequence diversity with the clinical course of HIV-1 infection was tested in Kaplan-Meier survival analysis primarily using CD4+ T cell count below 400 cells/µl blood, AIDS according to the CDC 1993 definition [21], and AIDS-related death as endpoints.

HMA patterns of the first time point samples (seroconversion) were not associated with any of the clinical endpoints or markers of disease progression analyzed (Table S1A). However, heteroduplexes in the HMA of the second time point (12 months post-seroconversion) were associated with accelerated progression to AIDS according to the CDC 1993 definition [21] (Table S1A, Figure 1A), as well as with accelerated progression to AIDS according to the CDC 1987 definition [22] or using CD4+ T cell count less than 200 cells/µl as endpoint (data not shown). Interestingly, second time point heteroduplexes were associated with CD4+ T cell counts at

Figure 1. HIV-1 envelope diversity 1 year after seroconversion predicts subsequent disease progression. A. Kaplan-Meier survival analysis for the time from 1 year after seroconversion to AIDS (1993 CDC definition) for presence (black) or absence (grey) of heteroduplexes in the HMA pattern 1 year after seroconversion. Hashes indicates censors due to HAART initiation or loss of follow up. P value from log rank test, as implemented in SPSS software, is given. B. Kaplan-Meier survival analysis for the time from 1 year after seroconversion to AIDS (1993 CDC definition) for high (black), medium (dotted), or low pairwise sequence diversity (grey) 1 year after seroconversion. Hashes indicates censors due to HAART initiation or loss of follow up. P value from log rank test, as implemented in SPSS software, is given.
viral load set point, but not with viral load set point itself (Figure S1). Moreover, HMA patterns were inversely correlated with the percent CD45RO+ (memory) CD4+ T cells and positively correlated with the percent CD38+ (activated) CD4+ T cells (Figure S1).

Both absolute and relative envelope genetic diversity at 1-year post-seroconversion was associated with accelerated progression to AIDS CDC 1993 (p=0.0003 and p=0.0009, respectively), with medium diversity being associated with an intermediate rate of disease progression as compared to low and high diversity (Table S1B, Figure 1B). Interestingly, although the group of patients for which viral genetic diversity could be analyzed by sequencing was smaller (n=74), the p values were actually stronger, suggesting that the association between HMA pattern and disease progression is driven by the underlying viral sequence diversity rather than only deletions and insertions, as these were not included in the pairwise genetic distance analysis. A stronger association was also observed using time to AIDS-related death, but no association was found using time to CD4+ T cell count below 400 cells/µl (Table S1B).

Predictive value of HMA pattern for HIV-1 disease course

Applying univariate Cox proportional-hazard analysis, we observed a relative hazard for progression to AIDS CDC 1993 from 1 year after seroconversion of 2.11 [95% confidence interval (CI) 1.20-3.70; p=0.00914] for patients with heteroduplexes in the HMA at 1-year post-seroconversion (Table 1).

Only ten individuals were heterozygous for the CCR5 Δ32 genotype, five of which had only homoduplexes 1 year after seroconversion and five also had heteroduplexes. Of the five HLA-B57 individuals, however, all had only homoduplexes in the HMA of the first time point and four had only homoduplexes at the second time point. Hence, in the selective population analyzed here, the CCR5 Δ32 genotype and HLA-B57 had no significant effect on disease progression (Table 1).

Subsequent multivariate analysis including other established markers of disease progression in the model analyzing time to AIDS from 2 years after seroconversion to AIDS

<table>
<thead>
<tr>
<th>Factor</th>
<th>Comparison</th>
<th>n</th>
<th>p value</th>
<th>RH (95% CI)</th>
<th>n</th>
<th>p value</th>
<th>RH (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMA</td>
<td>HE vs HO</td>
<td>89</td>
<td>0.0158</td>
<td>2.01 (1.14-3.56)</td>
<td>81</td>
<td>0.0229</td>
<td>1.96 (1.10-3.51)</td>
</tr>
<tr>
<td>CCR5</td>
<td>WT vs Δ32</td>
<td>86</td>
<td>ns</td>
<td>0.95 (0.43-2.12)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HLA</td>
<td>X vs B57</td>
<td>85</td>
<td>ns</td>
<td>5.09 (0.07-36.9)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SP CD4</td>
<td>&lt;500 cells/µl</td>
<td>85</td>
<td>0.0011</td>
<td>2.62 (1.47-4.69)</td>
<td>81</td>
<td>0.0063</td>
<td>2.34 (1.27-4.31)</td>
</tr>
<tr>
<td>SP VL</td>
<td>&gt;1 x 10⁴⁵ cp/ml</td>
<td>85</td>
<td>ns</td>
<td>1.45 (0.84-2.51)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SP MT-2</td>
<td>POS vs NEG</td>
<td>81</td>
<td>&lt;0.00001</td>
<td>11.9 (4.60-31.2)</td>
<td>81</td>
<td>&lt;0.00001</td>
<td>8.83 (3.39-23.0)</td>
</tr>
</tbody>
</table>

HMA, heteroduplex mobility assay; MSM, men who have sex with men; n, number of individuals included in analysis; RH, relative hazard; CI, confidence interval; HE, heteroduplexes present; HO, homoduplexes only; Δ32, 32 base pair deletion; WT, wild type; X, all other HLA types; SP, set point; POS, positive MT-2 assay; NEG, negative MT-2 assay. Univariate and multivariate analyses in a model analyzing the time from 2 years after seroconversion until AIDS (CDC 1993) diagnosis. P values from univariate and multivariate Cox proportional-hazard analyses.
CDC 1993 indicated that HMA pattern, CXCR4-use and CD4+ T cell count at set point were all independently predictive of HIV-1 disease progression (Table 1). The same was true for relative envelope diversity (Table S2).

DISCUSSION

HIV has an unparalleled capacity to generate genetic diversity. Single genome amplification and sequencing of viral populations isolated from patients during primary HIV-1 infection indicate low multiplicity infection and limited viral evolution preceding peak viremia [23,24]. Throughout infection several different forces shape the viral quasispecies, therefore over time this population diverges from the infecting strain and increases in diversity [25]. We analyzed the diversity of the viral quasispecies at seroconversion and at 1 year after seroconversion using HMA. This showed a clear association between increased viral diversity at 1 year after seroconversion (but not at seroconversion) and clinical disease progression. This indicates that the changes in viral diversity as measured by HMA between seroconversion and 1 year thereafter are relevant for the subsequent clinical course. Analysis of viral envelope genetic diversity by sequencing showed even stronger associations with disease progression, indicating that differences in actual sequence diversity are indeed underlying the observed association between HMA pattern and disease progression. It is likely that both the humoral and cellular immune response as well as target cell availability restrict viral diversity, which may relate to the HMA pattern at 1 year after seroconversion. The fact that all but one of the samples from HLA-B57 typed individuals carried only homoduplexes suggest that indeed the immune response can be involved.

Surprisingly, the HMA pattern at 1 year after seroconversion was not related to viral load set point, but it was related to the CD4+ T cell count at viral load set point. Although we did not analyze samples with very low viral load (less than 1000 copies/ml), viral load in the samples studied here spanned almost 3 logs (ranging from 3.30 to 6.23 log10 copies/ml), a range expected to be sufficient to observe an association with HMA pattern. The observations that HMA patterns were inversely correlated with the percentage CD45RO+ (memory) CD4+ T cells and positively correlated with the percentage CD38+ (activated) CD4+ T cells might indicate that rather than reflecting overall levels of virus replication, HMA patterns - and hence viral diversity - are associated with immune activation. Supporting this hypothesis, a recent study showed that disease progression is predicted by synonymous substitution rates in the envelope gene, which is thought to reflect different levels of persistent immune activation [26].

It remains to be established whether viral diversity itself plays a causal role in the increased damage to the immune system or whether it is a reflection of viral adaptation to selective pressures such as host immune responses or target cell availability. In this context, HMA pattern could be applied in patient monitoring as a parameter for patients who should start antiretroviral therapy early.

ACKNOWLEDGEMENTS

The Amsterdam Cohort Studies on HIV infection and AIDS, a collaboration between the Amsterdam Health Service, the Academic Medical Center of the University of Amsterdam, Sanquin Blood Supply
Foundation, the University Medical Center Utrecht, and the Jan van Goyen Clinic are part of the Netherlands HIV Monitoring Foundation and financially supported by the Center for Infectious Disease Control of the Netherlands National Institute for Public Health and the Environment.

The authors are indebted to the participants of the ACS. The study could not be performed without their dedication and help.

Financial supports for this study were obtained from the Academic Medical Center at the University of Amsterdam and from the Aids Fonds Netherlands [grant 2008023].
REFERENCES


22. Centers for disease control. Revision of the CDC surveillance case definition for acquired immunodeficiency syndrome. Council of State and Territorial Epidemiologists; AIDS Program, Center for Infectious Diseases. MMWR; 1987, 36 Suppl 1:1S-15S.

23. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, et al. Identification and characterization of transmitted and early founder virus envelopes in


SUPPLEMENTARY METHODS AND RESULTS

RT-PCR, PCR (outer and inner), sequencing and vector primer sequences

C2-C4 env primers
Seq2: 5’-TCCTCCATATCCTCCTCCAGTC-3’ (RT-PCR, outer PCR)
Seq3: 5’-TATGGGATCAAAGCCCTAAAGCATG-3’ (outer PCR)
Seq5: 5’-GTCAACTCACTGCTGTTAAATGGC-3’ (inner PCR, seq)
Seq6: 5’-ATCTAATTGCTCCACTGATGGGAGG-3’ (inner PCR, seq)

Vector primers
T7: 5’-TAATACGACTCACTATAGGG-3’
SP6: 5’-GATTTAGGTGACACTATAG-3’

Patient population
The prospective Amsterdam Cohort Studies on HIV Infection and AIDS (ACS) is a well-documented cohort study among HIV-infected and HIV-uninfected individuals who are at risk for acquiring HIV infection. Enrolment of men who have sex with men (MSM) started in October 1984 and up to December 31, 2008, 2,383 MSM have had at least one visit, with 1,588 testing HIV seronegative, 585 testing HIV seropositive and 210 HIV seroconverters. Clinical and epidemiological data is collected, laboratory markers such as CD4+ T cell numbers and plasma viral load are determined, serum or plasma is stored at -70°C and peripheral blood mononuclear cells are cryopreserved at 3-monthly intervals.

The ACS are conducted in accordance with the ethical principles set out in the Declaration of Helsinki and are approved by the Medical Ethics Committee of the Academic Medical Center in Amsterdam. Written informed consent is obtained for every participant.

Molecular cloning and sequencing of HIV-1 env C2–C4

Second-round PCR products were cloned into the pGEM-T Easy Vector system (Promega), transformed into competent DH5α Escherichia coli (Invitrogen), and plated on Luria-Bertani agar with blue-white screening. White colonies were picked at random (4–32 colonies per reaction) and cloned PCR products were amplified. After purification of PCR products (ExoSAP-IT; USB), the T7 and SP6 primers were used for sequencing (Big Dye Terminator Cycle Sequencing kit, version 1.1; Applied Biosystems). Sequences were determined using an automated DNA sequencer (Applied Biosystems).

When sequence diversity did not reflect the heteroduplex pattern obtained by heteroduplex mobility assay (HMA), an additional 4 nested PCRs were performed, and products were cloned and plated. Two clones per plate were picked, PCR with the SP6 and T7 primers was performed, and PCR products were sequenced. Clonal sequences were aligned for each patient by means of the ClustalW algorithm, and alignments were manually edited using BioEdit software (BioEdit, version 7.0.5.3; Ibis Biosciences). Alignments were visually inspected for the presence of mismatches, insertions, and deletions.
Statistical analysis

For the seroconversion time point HMA pattern, we analyzed time from seroconversion to each end point; for the 1-year post-seroconversion time point HMA pattern or sequence diversity, we analyzed time from 1 year after seroconversion to each end point. In addition, we studied time to AIDS-related death from the moment of AIDS diagnosis onwards. Log Rank p value was used to determine significant differences in the clinical course of infection between both groups. Univariate and multivariate analyses in a model that included HMA heteroduplexes or sequence diversity, CCR5 Δ32 heterozygous genotype, HLA-B57, positive MT-2 cultures, set point CD4+ T-cells below 500 cells/µl blood and set point viral load above $10^{4.5}$ copies/ml plasma were performed from 2 years after seroconversion to AIDS (SPSS 16.0 software package, SPSS Inc., Chicago, IL, USA).

The association of HMA pattern or sequence diversity with viral load or CD4+ T cell count at set point was tested using Student’s T test as implemented in GraphPad Prism 5 (GraphPad Software, La Jolla, CA, USA).

Predictive value of HMA pattern for HIV-1 disease course

We tested the predictive value of HMA pattern in the group of 89 MSM seroconverters versus several established markers of disease progression, namely CCR5 Δ32 genotype, HLA-type, detection of CXCR4-using virus by MT-2 assay (CXCR4 use), viral load and CD4+ T cell count at viral load set point in a model analyzing the time from 2 years after seroconversion to AIDS CDC 1993. Univariate analysis showed significant effects only for HMA pattern, CXCR4 use and CD4+ T cell count at viral load set point in this group (Table 1). Subsequent multivariate analysis with only these three predictors in the model analyzing time to AIDS from 2 years after seroconversion to AIDS CDC 1993 indicated that all three were independently predictive of HIV-1 disease progression (Table 1).

As expected, univariate and multivariate analyses also confirmed the independent predictive value of relative envelope sequence diversity for HIV-1 disease progression (Table S2).
Supplementary figure S1. Set point viral load, set point CD4+ T cell count, and CD4+ T cell subsets by HMA pattern. Heteroduplexes in HMA at 1 year post-seroconversion are correlated with set point CD4+ T cell count (A) but not with set point viral load (B). HMA patterns showed a trend for correlation with the percent CD38+ (activated) CD4+ T cells (C) and were inversely correlated with the percent CD45RO+ (memory) CD4+ T cells (D). Comparisons were analyzed for statistical significance using Student’s T test as implemented in GraphPad Prism 5 software. Mean and SEM (error bars) are represented for the respective groups. P values for significance are shown above each panel.
### SUPPLEMENTARY TABLES

#### Table S1A. Kaplan-Meier survival analysis of association between presence of heteroduplexes in HMA and time to clinical endpoint.

<table>
<thead>
<tr>
<th>Endpoint</th>
<th>First time point (seroconversion)</th>
<th>Heteroduplex present</th>
<th>p value</th>
<th>Second time point (1 year post-seroconversion)</th>
<th>Heteroduplex present</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
<td></td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>CD4&lt;400&lt;sup&gt;c&lt;/sup&gt;</td>
<td>37/45&lt;sup&gt;d&lt;/sup&gt;</td>
<td>41/44</td>
<td>ns</td>
<td>31/39</td>
<td>27/29</td>
<td>ns</td>
</tr>
<tr>
<td>AIDS93&lt;sup&gt;e&lt;/sup&gt;</td>
<td>29/45</td>
<td>26/44</td>
<td>0.0916</td>
<td>33/49</td>
<td>21/37</td>
<td>0.0078</td>
</tr>
<tr>
<td>Death</td>
<td>23/45</td>
<td>18/44</td>
<td>ns</td>
<td>26/50</td>
<td>15/37</td>
<td>0.0530</td>
</tr>
</tbody>
</table>

HMA, heteroduplex mobility assay; ns, not significant (p value > 0.05). No associations were found with time to first emergence of CXCR4-using HIV-1 or time between AIDS diagnosis and AIDS-related death (data not shown). <sup>a</sup> Survival analysis for time from seroconversion to each end point. <sup>b</sup> Survival analysis for time from 1 year post-seroconversion to each end point. <sup>c</sup> CD4<sup>+</sup> T cell count less than 400 cells/µl. <sup>d</sup> Number of individuals with events/total number of individuals. <sup>e</sup> AIDS according to the 1993 CDC definition.

#### Table S1B. Kaplan-Meier survival analysis of association between relative viral envelope sequence diversity and time to clinical endpoint.

<table>
<thead>
<tr>
<th>Endpoint</th>
<th>Second time point (1 year post-seroconversion)</th>
<th>Relative sequence diversity</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
</tr>
<tr>
<td>CD4&lt;400&lt;sup&gt;b&lt;/sup&gt;</td>
<td>19/20&lt;sup&gt;e&lt;/sup&gt;</td>
<td>15/20</td>
<td>13/15</td>
</tr>
<tr>
<td>AIDS93&lt;sup&gt;d&lt;/sup&gt;</td>
<td>15/24</td>
<td>14/23</td>
<td>18/23</td>
</tr>
<tr>
<td>Death</td>
<td>8/24</td>
<td>12/23</td>
<td>14/24</td>
</tr>
</tbody>
</table>

ns, not significant (p value > 0.05). No associations were found with time to first emergence of CXCR4-using HIV-1 or time between AIDS diagnosis and AIDS-related death (data not shown). <sup>a</sup> Survival analysis for time from 1 year post-seroconversion to each end point, except AIDS to death. <sup>b</sup> CD4<sup>+</sup> T cell count less than 400 cells/µl. <sup>c</sup> Number of individuals with events/total number of individuals. <sup>e</sup> AIDS according to the 1993 CDC definition.
Table S2. Predictive value of relative viral envelope sequence diversity 1 year post-seroconversion, CCR5 genotype, HLA-type, set point CD4+ T cell count, set point HIV-1 viral load and set point MT-2 assay result for progression to AIDS (CDC 1993) in 74 MSM seroconverters.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Comparison</th>
<th>Univariate</th>
<th>Multivariate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>p value</td>
<td>RH (95% CI)</td>
</tr>
<tr>
<td>Diversity</td>
<td>low vs medium vs high</td>
<td>63</td>
<td>0.001</td>
</tr>
<tr>
<td>CCR5</td>
<td>WT vs Δ32</td>
<td>63</td>
<td>ns</td>
</tr>
<tr>
<td>HLA</td>
<td>X vs B57</td>
<td>63</td>
<td>ns</td>
</tr>
<tr>
<td>SP CD4</td>
<td>&lt;500 cells/µl</td>
<td>63</td>
<td>0.009</td>
</tr>
<tr>
<td>SP viral load</td>
<td>&gt;1 x 10^4 cp/ml</td>
<td>63</td>
<td>ns</td>
</tr>
<tr>
<td>SP MT-2</td>
<td>POS vs NEG</td>
<td>63</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

MSM, men who have sex with men; n, number of individuals included in analysis; RH, relative hazard; CI, confidence interval; HE, heteroduplexes present; HO, homoduplexes only; Δ32, 32 base pair deletion; WT, wild type; X, all other HLA types; SP, set point; POS, positive MT-2 assay; NEG, negative MT-2 assay. Univariate and multivariate analyses in a model analyzing the time from 2 years after seroconversion until AIDS (CDC 1993) diagnosis. P values from univariate and multivariate Cox proportional hazard analyses.