Chapter 6

**Evaluation of pre-screening methods for the identification of HIV-1 superinfection**

*Andrea Rachinger¹, Tom Derks van de Ven¹, Judith A. Burger¹, Hanneke Schuitemaker¹, Angélique B. van ’t Wout¹*

¹Department of Experimental Immunology, Sanquin Research, Landsteiner Laboratory, and Center for Infection and Immunity Amsterdam (CINIMA) at the Academic Medical Center of the University of Amsterdam, Amsterdam, The Netherlands.

*Journal of Virological Methods, in press*
Abstract

The aim of this study was to compare sensitivity thresholds of two pre-screening methods – the heteroduplex mobility assay (HMA) and the presence of ambiguity codes in population-based sequences – applied for detection of HIV-1 superinfection. HIV-1 env C2-C4 PCR products generated from 48 serum samples isolated from 24 HIV-1 positive therapy-naïve homosexual men at seroconversion and at approximately 1 year thereafter were subjected to HMA and population sequencing. Clonal sequence analysis was used to determine the sensitivity of each method to detect sequence variability. Results from HMA were compared to pairwise genetic distance of clonal sequences; heteroduplexes resulted from as little as 1.4% pairwise distance between 2 sequences and were detected even when only 1.5% of the pairwise distance comparisons exceeded this distance threshold. By contrast, the ambiguity code approach using population-based sequencing detected only 20.1% of existing sequence variation and was less sensitive to minority populations ≤20%, resulting in an underestimation of HIV-1 diversity. Thus, HMA was found to be more sensitive to detect sequence variation than the ambiguity code approach, suggesting that HMA would be a more appropriate method to pre-screen for HIV-1 superinfection.

It is unclear as yet, what the impact of HIV-1 superinfection is on HIV-1 disease and at what rates HIV-1 superinfection occurs. Studies have yielded contradictory results with superinfection incidence rates reported by some as high as the rate of initial infection [1-4] while others report the absence of HIV-1 superinfection [5, 6]. To date, almost 50 well-documented cases of HIV-1 superinfection have been published [1-4, 7-25]. In several cases, initial and superinfecting viral strains were of different subtypes [2-4, 7, 9, 11, 15, 18, 23], including initial CRF01_AE infection and superinfection with a subtype B variant [9], as well as superinfection with a non-B subtype variant after initial subtype B infection [9, 11, 18, 23]. However, in many other cases superinfecting strains had the same subtypes as the initial strain.

HIV-1 superinfection case reports were prompted by retroviral syndrome, occurrence of drug resistance, and sudden increases in viral load [7, 8, 10, 17, 23]. However, screening cohorts of female commercial sex workers, injecting drug users or homosexual men has identified most superinfection cases [1-4, 9, 11, 13-16, 18]. In these studies many different detection methods were used, such as restriction fragment length polymorphism (RFLP), multiregion hybridization assay (MHA), heteroduplex tracking assay (HTA), heteroduplex mobility assay (HMA), and sequencing of single copy PCR amplicons, population-based sequencing, clonal sequencing followed by phylogenetic analysis. The lack of systemic cohort screenings is apparent, leaving incidence rates of HIV-1 superinfection most certainly biased.
In some superinfection cases, an increased plasma viral load or a decline in CD4 T cell count was described [1, 7, 9-12, 14, 15, 17, 23] leading to the speculation that superinfection in general might be associated with faster disease progression. However, this does not hold true for all cases and often information on viral load and CD4 count were unavailable, not allowing firm conclusions on clinical consequences of HIV-1 superinfection. In most documented superinfection cases, the superinfecting strain was detected when representing the major or the only variant within the viral quasispecies and thus easily detectable by even less sensitive detection methods. However, with detection methods of low sensitivity, a significant number of superinfections might have been missed resulting in a significant underestimation of the incidence of HIV-1 superinfection. A more realistic incidence rate of HIV-1 superinfection is however also desirable to be able to accurately and truthfully counsel communities at risk for HIV-1 superinfection. Finally, to identify individuals that are able to contain a superinfecting virus at a low level might assist in identifying mechanisms that could contribute to the creation of an efficient, cross-protective vaccine.

In this comparative study, the sensitivity of 2 pre-screening methods - heteroduplex mobility assay (HMA) and ambiguity codes present in population-based sequences - was investigated. HMA has been applied in the pre-screening of cohorts to detect HIV-1 superinfection [5, 14, 15, 26] and allows for the detection of diversity within a viral quasispecies with a sensitivity threshold of 1-2% sequence variation [27]. Recently, a novel application was suggested for population-based HIV-1 polymerase (protease and reverse transcriptase) sequences generated for routine determination of resistance-associated mutations [28]. Within population-based sequences multiple nucleotides are possible at a single position and following the IUPAC-IUB nomenclature, this diversity can be described by single-letter symbols called ambiguity codes, like R for an A and a G present at one position [29]. Sequences containing more than 34 ambiguous sites in the viral reverse transcriptase were previously selected for clonal analysis in env and gag followed by phylogenetic analysis revealing new cases of dual infections [28]. However, clonal sequence analysis of drug-resistance associated mutations in pol revealed that minor variants could not be detected by population-based sequencing, since minor variants need to comprise 20 to 30% of a quasispecies to be detectable within the sequence chromatogram as an underlying nucleotide peak resulting in an ambiguity code [30-34].

To compare the performance of HMA and of IUB ambiguity codes in population-based sequences as pre-screening methods, we used serum samples that were isolated from therapy-naïve homosexual male participants of the Amsterdam Cohort Studies on HIV infection and AIDS (ACS). Forty-eight serum samples with a
detectable viral load (>1,000 copies/mL plasma) at seroconversion (SC) (median: 2 months after SC, range: 2 months before SC to 3 months after SC) and at 1 year after SC (median: 13 months after SC, range: 8-19 months after SC) from twenty-four HIV-1 positive homosexual men, all infected with subtype B HIV-1 variants, were included in this study (Figure 1).

Figure 1: Experimental approach for determination of sensitivity thresholds to compare 2 superinfection pre-screening methods using serum samples obtained from participants of the Amsterdam Cohort Studies on HIV Infection and AIDS (ACS). 48 serum samples with detectable VL using HMA (>1000 copies/mL) of 24 seroconverters of the ACS were investigated to compare sensitivity of the 2 pre-screening methods heteroduplex mobility assay (HMA) and population sequencing. To this end, env C2-C4 PCR products were subjected to HMA, population sequencing and clonal sequencing. Genetic pairwise distance of clonal sequences was compared to the presence of HMA heteroduplexes. Sequence alignments containing clonal and population sequences were investigated via single nucleotide site analysis.

Following isolation of viral RNA from 140μl serum (Qiamp Viral Mini Kit, Qiagen, Hilden, Germany), 10μl RNA was reverse transcribed into cDNA (SuperScript™ First strand synthesis system, Invitrogen, Carlsbad, CA) using the HIV-specific primer Seq2 (5’-TCCTCCATATCTCCTCCTCCAGGTC-3’). The C2-V3-C3-V4-C4 region of the HIV-1 env gene was amplified with primer pairs Seq2 and Seq3 (5’-TATGGGATCAAAGCCTAAAGCATG-3’) in the first-round PCR and Seq5 (5’-GTCAACTCAACTGCTGTTAAATGGC-3’) and Seq6 (5’-
ATCTAATTGTCCACTGATGGGAGG-3’) in the second-round PCR (GoTaq Flexi DNA Polymerase, Promega, Madison, WI) with 5μl cDNA input in the first-round PCR and 2μl first-round PCR product as input in the second-round PCR. Both PCR reactions were carried out in a total volume of 25μl using the following PCR cycles: 1 cycle at 94°C for 5 minutes, and 35 cycles at 94, 50 and 70°C of respectively 45, 30 and 90 sec, followed by a final elongation step of 10 min at 70°C resulting in the final PCR product of 549 nucleotides of env C2-V3-C3-V4-C4 (HXB2 positions 7012-7560). After purification (ExoSAP-IT, USB, Cleveland, OH), second-round PCR products were directly subjected to population sequencing using primers Seq5 and Seq6 with the Big Dye terminator v1.1 cycle sequencing kit (Applied Biosystems, Foster City, CA). For clonal analysis, second-round PCR products were cloned into the pGEM T Easy Vector System (Promega, Madison, WI), transformed into competent DH5α E.coli (Invitrogen, Carlsbad, CA) and plated on LB agar using blue/white screening. The primers Seq5 and Seq6 were used to amplify the cloned PCR products as described above. After purification (ExoSAP-IT, USB, Cleveland, OH), PCR products were sequenced with primers Seq5 and Seq6.

In total, population-based sequencing was performed on C2-C4 env fragments of all 48 samples and a median of 11 C2-C4 env clonal sequences was generated per sample (range: 4 to 20). Population-based and clonal sequences were aligned per patient (Figure 1) using the ClustalW algorithm [35] and alignments were manually edited in BioEdit (BioEdit, version 7.0.5.3; Ibis Biosciences) to maximize homology at each position.

Single nucleotide site variation was analyzed in patient’s alignments: population-based sequences were visually inspected for the presence of ambiguity codes and clonal sequences for nucleotide variation (Figure 2). Within all alignments, a total of 237 600 single nucleotide sites were analyzed of which 608 (0.3%) were categorized as such sequence variation. These were grouped in one of the following 4 categories: Category (I) contained 71 concordant sites (12%) with ambiguity codes present in the population sequence and the corresponding nucleotides in the clonal sequence; Category (II) contained 49 discordant sites (8%) that displayed ambiguity codes in the population sequence while no variation was detected in the clonal sequences; Category (III) contained 476 discordant sites (78%) with no ambiguity codes present in the population sequence while variation was detected in the clonal sequence; Category (IV) contained 12 discordant sites (2%) with no ambiguity codes present in the population sequence and a different nucleotide in the clonal sequence (Figures 2, 3a). Taken together, 120 ambiguity codes in population-based sequences were detected in categories (I) and (II), and 547 single nucleotides sites showed variation in clonal sequences (categories (I) and (III)).
Thus the IUB approach detected only 20.1% of the variation found in population-based and clonal sequences.

Figure 2: Single nucleotide site analysis. 237,600 nucleotide sites from all alignments containing population and clonal sequences were investigated for diversity. At 608 sites diversity was detected and categorized into Cat (I) through (IV) based on ambiguity codes present/absent in population-based sequences and variation detected/not detected in clonal sequences. Numbers in each category are shown.

We next investigated variable sites of categories (I) and (III) in more detail. For category (I), we analyzed the percentages at which the minor nucleotide variant was represented in the clonal sequences. Each minor nucleotide variant in a double peak in the chromatogram of population-based sequences was represented in a mean of 30.6% of the clones (Figure 3b).
Figure 3a: Single site category analysis. Variable site categories (I) through (IV) identified through single nucleotide site analysis per time point per patient A through X. Each column represents results of one time point from each patient, with variable sites of all 4 categories represented.

For category (III), we found that the nucleotide variation in the clonal sequence analysis that was not detected by ambiguity code analysis of the population-based sequence, was represented only in a mean of 15.3% of the clones (Figure 3b), which was significantly lower as compared to category (I) minor variants (Student’s t test, \( P < 0.0001 \)). This corresponds well to the reported threshold of 20-30% for a second nucleotide peak to be detectable within the sequence chromatogram [30-34]. To illustrate this further, the IUB approach detected only 8 of the 277 (2.9%) variants present in \( \leq 10\% \) of the clones and only 20 of the 396 (5.1%) variants present in \( \leq 20\% \) of the clones.

Figure 3b: Single site category analysis. Category (I) and (III) variable sites – representation of minor variants in molecular clones. For category (I): Each nucleotide that was represented in the ambiguity code and the minor variant at this site, was present in a mean of 30.6% of the clones. For category (III): Each nucleotide that was not represented in the population-based sequence was present in a mean of 15.3% of the clones. Representation was significantly different between both categories (Student’s t test, \( P < 0.0001 \)).
The discordant sites of categories (II) and (IV), in which ambiguity codes were detected but no variation in clonal sequences or different nucleotides between population and clonal sequences were found, were probably due to the fact that neither the population-based sequencing nor the clonal analysis fully captured all variability present in the quasispecies. In addition, DNA polymerase errors occurring during the PCR may have added to the number of discordant sites in our study. Both categories, however, constituted only a small fraction of the total nucleotide variation (8% and 2%, respectively).

Homo- and/or heteroduplexes of C2-C4 env fragments were generated from the viral quasispecies of the 48 serum samples (Figure 1). In brief, 5μl of the final PCR product of each sample was denatured at 95°C for 2 min and immediately transferred to wet ice; homo-/heteroduplexes were resolved on a 5% non-denaturing polyacrylamide gel [27]. Only homoduplexes were observed in 25 (14 at 1st time point, 11 at 2nd time point) of the 48 samples and homo- and heteroduplexes were observed in 23 (10 at 1st time point, 13 at 2nd time point) of the 48 samples (Figure 4a, b).

Figure 4: Comparison of HMA homo-/heteroduplex patterns to pairwise genetic distances within time points and presence of deletions between pairwise compared clonal sequences. (A) Maximum genetic pairwise distance of clonal sequences plotted against presence or absence of HMA heteroduplexes, with maximum genetic distances significantly different between PCRs of time points that generated homo- respectively homo- and heteroduplexes (Student’s t test, P=0.0002). (B) Percent of pairwise distance (p-dist) comparisons greater than 1.3% pairwise distance, with percentages significantly different between PCRs of time points that generated homo- respectively homo- and heteroduplexes (Student’s t test, P<0.0001). Closed circles represent samples without deletions between pairwise compared sequences, open circles represent samples with deletions.
Pairwise genetic distance between clonal sequences within each sample was calculated using Mega applying the Kimura 2-Parameter (http://www.megasoftware.net). The maximum pairwise distance in homoduplex-only samples was 1.4%, which is in complete concordance with an earlier publication on HMA describing the absence of heteroduplexes below a pairwise genetic difference of 1.4% [27]. It is likely that superinfections with heterologous HIV-1 strains exceed the pairwise distance threshold for generating heteroduplexes of 1.4%. The mean of the maximum pairwise difference in homoduplex-only samples was significantly different from that from samples with heteroduplexes (0.68% versus 1.58%, Student’s t test, P=0.0002, Figure 4a). Deletions were not present in pairwise compared clonal sequences of homoduplex-only samples, but were detected in PCR fragments resulting in heteroduplexes: 14 out of 23 of these samples contained deletions, with the size of the deletion between 1 to 28 nucleotides (Figure 4a). Indeed, all of the samples giving rise to heteroduplexes, but with a maximum pairwise distance (which is calculated excluding gaps) less than 1.4%, contained deletions, explaining the heteroduplex pattern.

We next calculated for each sample the number of pairwise comparisons that exceeded the threshold distance of 1.4%. The vast majority (25/26) of samples with only homoduplexes had no pairwise comparisons exceeding this threshold (median 0%, range 0-2.2%), while the majority (15/23) of samples with heteroduplexes did have pairwise comparisons exceeding the threshold (median 13.3%, range 0-80.0%) (Student’s t test, P<0.0001, Figure 4b), deletions again explaining the heteroduplex pattern in 8 samples with no pairwise comparisons exceeding the threshold (Figure 4b). Using the threshold distance of 1.4%, heteroduplexes were detected in 5 of 6 (83.3%) samples where ≤20% of the pairwise comparisons exceeded the threshold. Moreover, in one case heteroduplexes were detected even when only 1 in 66 (1.5%) of the pairwise distance comparisons exceeded the threshold.

To summarize, the IUB approach (5.1%) had less sensitivity than HMA (83.3%) had to detect minority species constituting less than 20%. This indicates that HMA is more likely to detect sequence variability - and by extension dual infections - at lower levels when used in pre-screens to determine the incidence of HIV-1 superinfection.

Interestingly, one patient (patient X) appeared to be HIV-1 co-infected based on visual inspection of the sequence alignment, which indicated the presence of 2 different virus strains at the first time point (confirmed by phylogenetic analysis, Rachinger et al, CID in press). Thus, this patient harbored 2 distinct viral strains (clonal strains A and B) at seroconversion, of which virus B was not detectable at
the 2nd time point analyzed. At the start of our study, we generated one population-based sequence from serum that was obtained at the first time point, close to seroconversion and one of the second time point that was about one year later. PCR products were cloned and sequenced and revealed the presence of 2 distinct virus strains at seroconversion, hence an HIV-1 co-infection. We performed a second PCR of the seroconversion time point and from this also a population-based sequence. Interestingly, the population sequence for the first PCR was highly similar to the clonal strain A, while the population sequence for the second PCR was highly similar to the clonal strain B with no ambiguity codes present in any of the population-based sequences (data not shown). Hence, the co-infection would have been missed if only 1 population-based sequence from this time point would have been performed and analyzed for the presence of unresolved IUB codes. HMA showed heteroduplexes present in the first PCR of the seroconversion time point, but not in the second PCR of this time point (data not shown). Hence, HMA can also give false negative results, most likely due to inconsistent PCR amplification of both co-infecting strains. In this case, the dual infection was indicated by both methods when the two PCR products were compared, also illustrating the importance of testing multiple PCR products from multiple time points.

Detection of HIV-1 superinfection is highly labor-intensive and compared to detection of initial HIV-1 infection it requires a different set of more sophisticated methods. While close to 50 cases of HIV-1 superinfections have been published to date, a reliable incidence of HIV-1 superinfection has not yet been established and as a result the question whether or not HIV-1 superinfection causes an accelerated disease progression has remained unsolved. While the rate at which superinfection occurs is most certainly influenced by other factors (e.g. sexual behavior) as well, the detection of HIV-1 superinfection depends primarily on the sensitivity of the applied detection methods. PCR-based methods such as the IUB approach and HMA are influenced by the high sequence variability of the worldwide HIV-1 epidemic [36], as selective amplification of certain HIV-1 subtypes may lead to underestimation of HIV-1 superinfection. As the early HIV-1 epidemic in the Netherlands was exclusively caused by HIV-1 subtype B, the current study was performed using HIV-1 subtype B-based primers. Given the increasing level of strain heterogeneity in Europe and for use in non-European settings, primer sets will need to be modified to allow sensitive detection of all HIV-1 subtypes.

Our results indicate that although the IUB approach might be a tool for fast pre-screening of large cohorts for which pol sequences are available from drug-resistance testing, HMA seems to be a more appropriate method to pre-screen for sequence variability and HIV-1 superinfection. However, signals detected by HMA also require clonal sequencing and additional follow-up studies to be able to confirm
superinfection. Methods that are even more sensitive in detection of sequence diversity than HMA are probably warranted to further increase accuracy of superinfection incidences. Novel high-throughput sequencing technologies may allow the detection of virus strains present at extremely low levels and thus a more reliable estimate of the rate of HIV-1 superinfection and its consequences for HIV-1 disease pathogenesis.

Acknowledgements
The Amsterdam Cohort Studies on HIV infection and AIDS, a collaboration between the Public Health Service of Amsterdam, the Academic Medical Center of the University of Amsterdam, the Sanquin Blood Supply Foundation, the University Medical Center Utrecht, and the Jan van Goyen Medical Center, 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. Financial support for this study was obtained from the Academic Medical Center at the University of Amsterdam. The authors are indebted to the participants of the ACS. We thank Peter van Swieten for excellent technical assistance, Diana Edo Matas for helpful discussions and Neeltje Kootstra for critical reading of the manuscript.

References