



Application of the genetic diversity analysis for determination of HIV infection duration

Murzakova Anastasia, Lapovok Ilya, Saleeva Daria, Kirichenko Alina, Kireev Dmitry

Central Research Institute of Epidemiology, Moscow, Russia

Background

The phenomenon of HIV-1 genome nucleotide sequence variability is an extremely important object for research. It was shown that HIV-1 genetic diversity may reflect the duration of infection in ART-naïve patients. It has already been established that genetic diversity grows over time of infection, reflecting the natural evolution of the virus in the patient.

To estimate the duration of HIV infection possible by counting ambiguous nucleotide rate (ANR) – the quota of ambiguous nucleotides in sequences studied in relation to a full number of nucleotides.

Ambiguous (degenerate) nucleotides are positions where sequencing results in the detection of several nucleotides. Some researches have determined ANR of 0.11% - 0.5% (for different genome regions) as a threshold for detecting an infection lasting less than 12 months.

Also, it should be noted that genetic variability in a sample may be due to dual HIV infection when two or more viral variants are present in patients. The synonymous index (SM-index) may be used to detect dual HIV infection and is calculated as the number of synonymous base pair mixtures divided by the number of all synonymous sites in sequences studied. The SM-index threshold 0.05 in PR-RT sequences (2253-3368 bp) may detect dual HIV infection.

The aim of this work was to estimate the rate of genetic variability in samples from patients with different duration of infection in order to establish a relationship between ANR and the infection duration, as well as to identify a possible dual HIV infection.

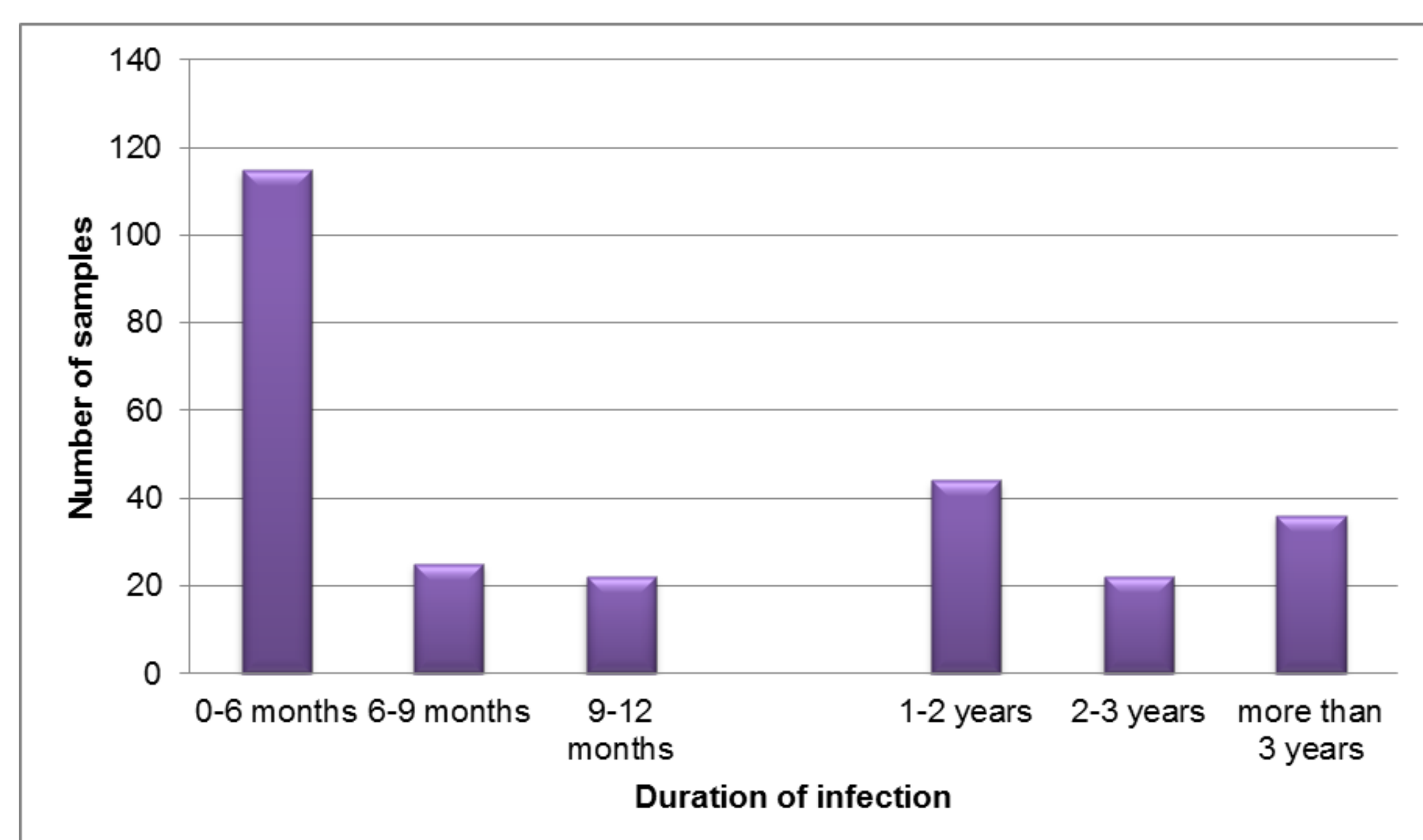


Fig.1. Distribution of samples by the duration of HIV infection.

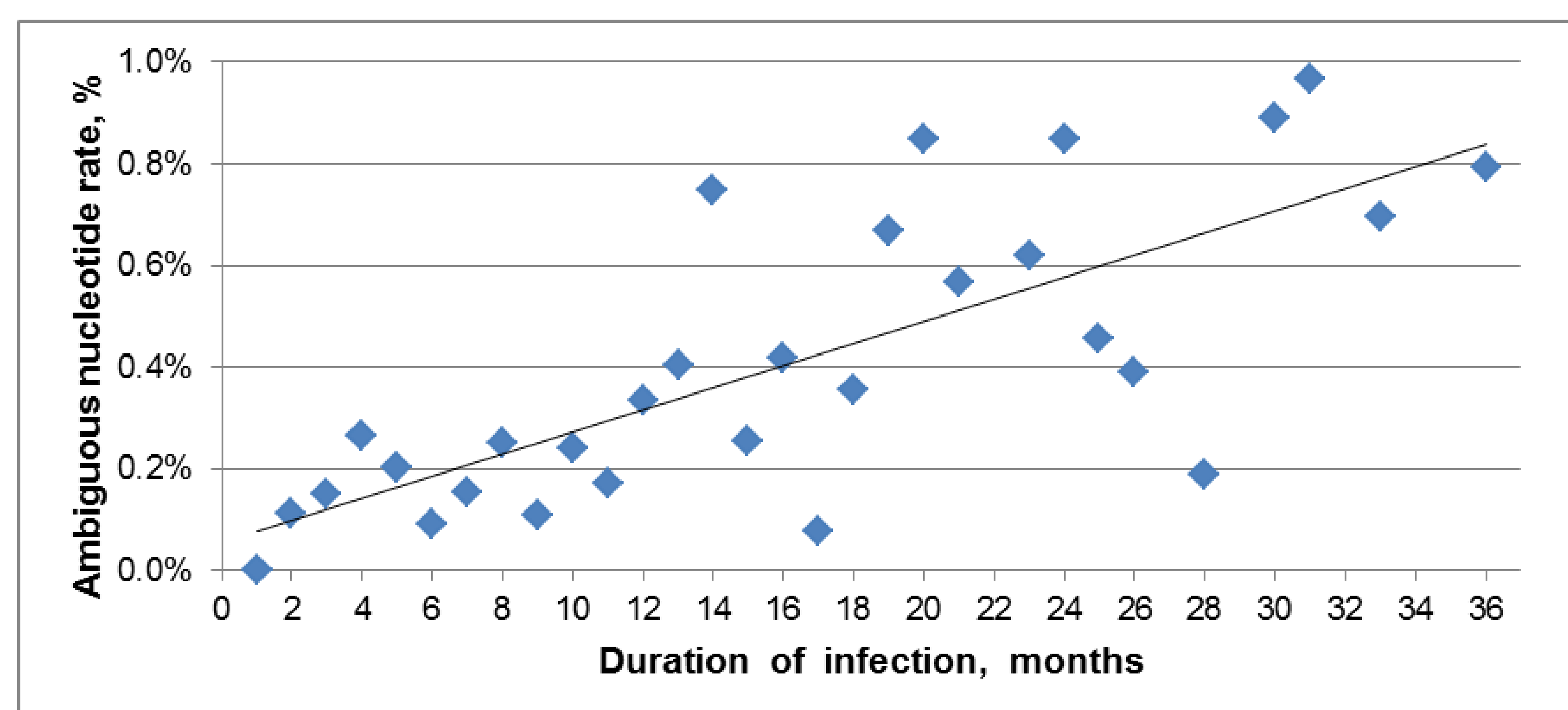


Fig. 3. Relationship between the duration of HIV infection and the ambiguous nucleotide rate in the sequences of the HIV-1 *pol* region fragment.

Materials & Methods

Plasma samples (n=262) were obtained from ARV-naïve patients: 162 samples from patients with an infection duration up to 12 months (recent samples) and 100 samples from patients with infection duration more than 12 months (established samples) (Fig. 1). The duration of infection was determined according to data of the last negative and first positive ELISA and immunoblot tests. Nucleotide sequences of PR-RT fragment including protease gene and a fragment of reverse transcriptase gene (positions 2253-3369 according to HXB2) were obtained using AmpliSens HIV-Resist-Seq kit (CRIE, Russia) by Sanger sequencing. Alignment and nucleotide analysis was carried out in BioEdit v7.2.5. Subtyping was carried out by REGA v3.0 and subsequently confirmed by phylogenetic analysis.

Results

The total 201 (76.7%) samples studied harbored HIV-1 sub-subtype A6, subtype B was detected in 27 (10.3%) samples, other subtypes and CRFs – in 34 (13.0%) samples (Fig. 2). We found that for HIV-1 sequences the fraction of ambiguous nucleotides increases significantly with the duration of infection. This relationship is shown in Figure 3. Also, we found reliable differences in ANR between recent and established samples (0.16% vs 0.71%, $p < 0.01$). The ANR value of 0.33% allowed discrimination 82.1% of recent and 62.7% of established samples and may be used as the optimal threshold.

The values of the SM-index in the two sample groups studied were 0.003 and 0.013 ($p < 0.01$). Only one sample had the SM-index of 0.056. But this sample was obtained from a patient with duration of infection more than 3 years and this result may be explained by natural polymorphism of the virus but not by dual HIV infection.

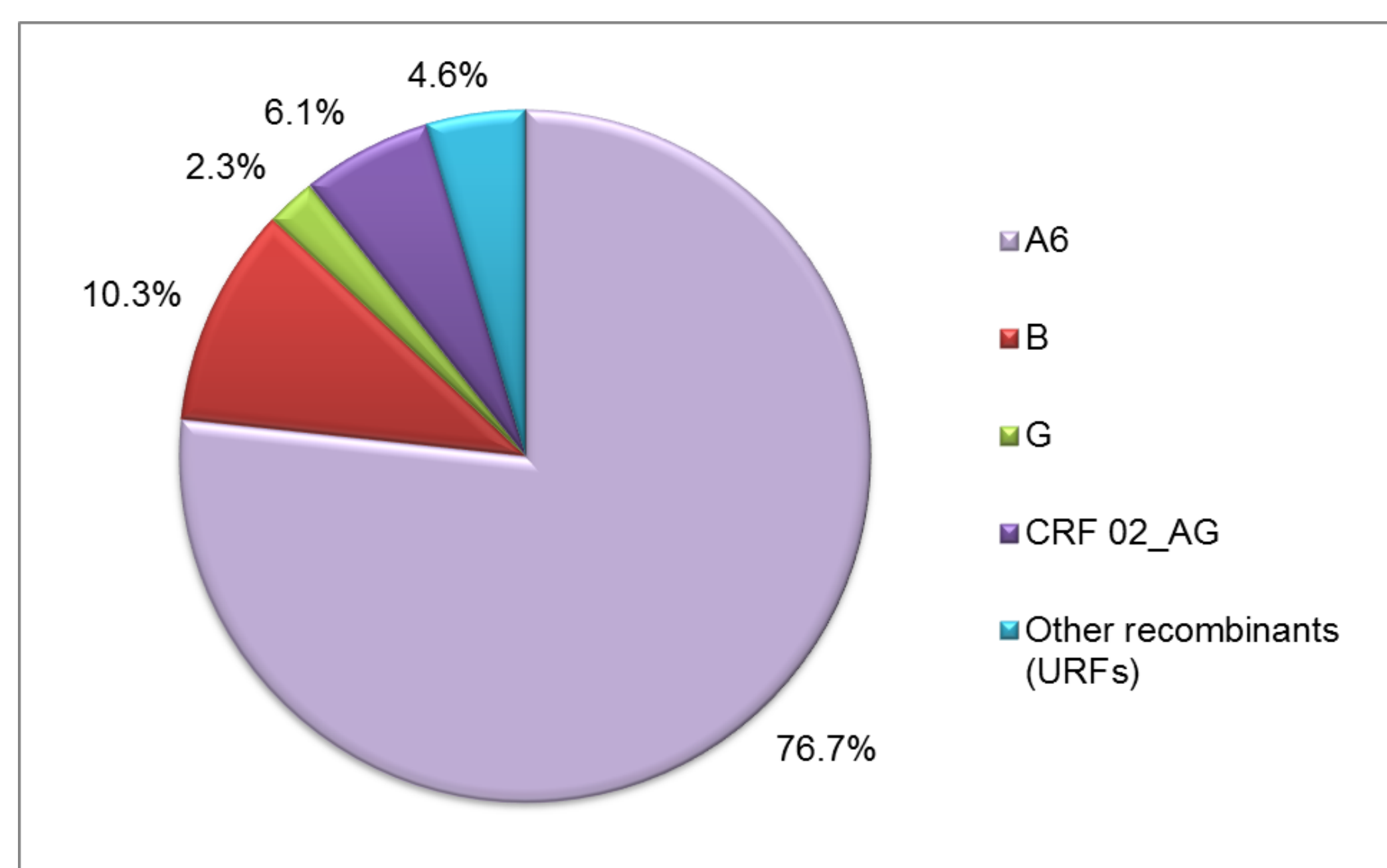


Fig.2. Distribution of samples by HIV subtype.

Conclusions

It was found that cohorts of patients with recent and established HIV infection differ with genetic variability by ambiguous nucleotide rate calculating. There were no samples with dual HIV infection in the samples studied.