Application of next generation sequencing in dual HIV infection studies

Ilya A. Lapovok✉, Pavel B. Baryshev, Daria V. Saleeva, Alina A. Kirichenko,
Anastasiia V. Shlykova (Murzakova), Dmitry E. Kireev

Central Research Institute of Epidemiology, Moscow, Russia

Abstract

Introduction. The aim of the study was to use comparative analysis for assessing efficiency of detection and confirmation of dual HIV infection, using conventional population sequencing (PS) and next generation sequencing (NGS) for an HIV-1 *pol* gene fragment, which encompasses protease and partially reverse transcriptase (positions 2253–3368).

Materials and methods. The study was performed on intersubtype dual HIV infection model samples containing viruses of HIV-1 subtype B, sub-subtype A6 and recombinant form CRF63_02A1. Viruses were mixed pairwise in proportions from 10 to 90% to obtain 3 groups of model samples: CRF63vsB, CRF63vsA6, and A6vsB. The nucleotide sequences obtained by using PS and NGS technologies having 5, 10, 15, and 20% sensitivity thresholds for minor virus variants (NGS5–NGS20, respectively) were used to estimate the number of degenerate nucleotides or the degenerate base (DB) count and the number of synonymous mutations (SM) or the SM count. The fragment of the studied region (positions 2725–2981) was used for the analysis of operational taxonomic units.

Results. The application of NGS5 proved highly efficient for detection of dual HIV infection in the model samples. The statistically significant ($p < 0.01$) increase in DB and SM counts was demonstrated by NGS5 compared to PS. As a result, NGS5 helped detect dual HIV infection in 25 out of 27 model samples, while with PS it was detected only in 15 samples. The analysis of operational taxonomic units confirmed dual HIV infection in all the groups of model samples.

Discussion. The efficiency of detection and confirmation of dual HIV infection depends both on the content of each virus in the sample and on genetic characteristics of these viruses.

Conclusion. Using NGS genetic testing in routine practice will be instrumental for efficient identification of genetic characteristics of infectious agents and for thorough analysis of the epidemiological situation.

Keywords: *dual HIV infection, model sample, next generation sequencing, nucleotide sequence, protease and reverse transcriptase region, degenerate base count, synonymous mutation count, operational taxonomic units*

Применение секвенирования следующего поколения для исследования двойной ВИЧ-инфекции

Лаповок И.А.✉, Барышев П.Б., Салеева Д.В., Кириченко А.А.,
Шлыкова (Мурзакова) А.В., Киреев Д.Е.

Центральный научно-исследовательский институт эпидемиологии Роспотребнадзора, Москва, Россия

Аннотация

Введение. Целью исследования был сравнительный анализ эффективности выявления и подтверждения двойной ВИЧ-инфекции с применением классического популяционного секвенирования (ПС) и секвениро-

вания следующего поколения (NGS) для фрагмента гена *pol* ВИЧ-1, кодирующего протеазу и часть обратной транскриптазы (позиции 2253–3368).

Материалы и методы. Исследованы модельные образцы межсубтиповой двойной ВИЧ-инфекции, содержащие вирусы ВИЧ-1 субтипа B, субсубтипа A6 и рекомбинантной формы CRF63_02A1. Вирусы смешивали попарно в соотношении от 10 до 90% для создания 3 групп модельных образцов: CRF63vsB, CRF63vsA6 и A6vsB. Нуклеотидные последовательности, полученные технологиями ПС и NGS, с порогами чувствительности к минорным вариантам вируса 5, 10, 15 и 20% (NGS5–NGS20 соответственно) использовали для определения величин индекса вырожденности (DB) и индекса синонимичности (SM). Фрагмент исследуемого региона (позиции 2725–2981) служил для анализа операционных таксономических единиц.

Результаты. Применение NGS5 оказалось наиболее эффективным для выявления двойной ВИЧ-инфекции в модельных образцах. Было обнаружено статистически достоверное ($p < 0,01$) увеличение DB- и SM-индексов для NGS5 по сравнению с ПС. Это позволило с помощью NGS5 выявить двойную ВИЧ-инфекцию в 25 модельных образцах из 27, в то время как ПС позволяло выявлять её лишь в 15 образцах. Анализ операционных таксономических единиц подтвердил двойную ВИЧ-инфекцию во всех группах модельных образцов.

Обсуждение. Эффективность выявления и подтверждения двойной ВИЧ-инфекции зависит как от доли каждого вируса в образце, так и от генетических особенностей данных вирусов.

Заключение. Внедрение в рутинную практику генетического анализа технологии NGS позволит не только более эффективно выявлять генетические особенности инфекционных агентов, но и проводить более глубокий анализ эпидемиологической ситуации.

Ключевые слова: *двойная ВИЧ-инфекция, модельный образец, секвенирование следующего поколения, нуклеотидная последовательность, область протеазы и обратной транскриптазы, индекс вырожденности, индекс синонимичности, операционные таксономические единицы*

Introduction

Dual HIV infection or the proven fact of presence of 2 and more HIV types (HIV-1 and/or HIV-2) in one patient [1] occurs at different frequency in different countries and among different vulnerable groups. There is direct relationship between the frequency of occurrence of this phenomenon, the level of poverty in the country and the level of marginalization of vulnerable groups. For example, in the Netherlands, the frequency of dual HIV infection in the group of men who have sex with men (MSM) is 1% [2], in Brazil (also among MSM), it is more than 12% [3], and in Uganda, among female sex workers, it is more than 16% [4]. In China, in the MSM group [5], this variable shows a wide range; based on different sources, from 13.3 to 28%, which is, most likely, associated with strong stigmatization of this vulnerable group in the country. In Russia, the frequency of dual HIV infection does not exceed 1.5% [6], which is commonly demonstrated by developed countries of Western Europe.

Since dual HIV infection is associated with generation of novel recombinant forms of the virus, progression of infection and even selection of drug resistance [1], its detection and confirmation are objectives of high priority.

The presence of dual HIV infection is most reliably confirmed by using single genome sequencing (SGS), which serves as a gold standard for diagnosis of this infection [7, 8]. The method offers sequencing genomes of individual viral particles after the limiting dilution of the sample [7], amplification of individual fragments of the viral genome and their subsequent cloning for accumulating clones containing fragments of each viral variant present in the sample [2]. Each clone goes through sequencing; the resulting nucleotide sequences (NSs) are put through phylogenetic analysis. Dual HIV infection is confirmed when there are statistically significant (with a bootstrap value of more than 80%) individual clusters on the tree diagram, which are typical of individual viral variants [2].

Studies are normally focused on short genomic fragments: V2–V3 loops of the *env* gene, fragment of the *gag* gene encoding p17 and/or p24 [9], fragment of the HIV-1 reverse transcriptase (RT) gene [8–10]. Besides, several of the above genomic fragments can be analyzed simultaneously [8].

SGS is too time-consuming and costly to be used on routine basis. For example, the study performed in 2010 [8] showed that the analysis of one sample in 2 genomic regions by using SGS took approximately 42 hours and its cost was more than $2,600. Compared to this, the analysis of the *pol* gene (protease and reverse transcriptase region, PR-RT), which employs conventional sequencing used for routine tests for drug resistance of HIV-1, takes only 3 hours spent on sample preparation and costs less than $280.

Researchers of dual HIV infection have been offered the next generation sequencing (NGS) technology that can be used for identification of minor variants of the virus accounting for less than 1% of the viral population in the sample [11]. High sensitivity makes it possible to use NGS technology as alternative to SGS. This approach is not only more advanced, but also more rewarding. NGS used for simultaneous analysis of 3 genomic regions takes only around 9.5 hours and costs slightly more than $1,000, being more than 4 times faster and more than 2.5 times less expensive than SGS [8].

The obtained NGS-based data on NSs in the sample, which are also referred to as "reads", go through subsequent mathematical analysis grouping them into operational taxonomic units (OTUs) – NSs formed by reads clustered at a certain level of similarity [9]. The further phylogenetic analysis of OTUs is performed in the way similar to the single-genome sequence assay resulting from SGS.

Despite the relative technological affordability of the analysis required for confirmation of dual HIV infection, the problem of initial identification of such samples remains unsolved. The presence of dual HIV infection is most frequently indicated by changes in the clinical picture: a sharp increase in HIV RNA blood concentration (viral load), a reduction of CD4 cell counts during long non-progression or slow progression of HIV infection [7, 12, 13]. However, these symptoms are observed only during HIV superinfection, when infection with a new viral variant occurs after the immune response to the initial infection with the first virus has been established [14]. Dual HIV infection may not be reported for patients with HIV coinfection (infection with 2 and more viral variants at the same time or within a brief period of time) [15] or in absence of regular check-ups of patients.

In the meantime, dual HIV infection can be detected by additional analysis of the test results for drug resistance of HIV-1. The drug resistance analysis, which is based on conventional Sanger sequencing (also known as population sequencing (PS)) [7, 14], makes it possible to obtain NSs of different HIV-1 genomic regions, including the most frequently studied PR-RT region encoding protease and reverse transcriptase of the virus. It has been found that the presence of large numbers of controversial or degenerate nucleotide positions (when the signal was indicative of the simultaneous presence of several nucleotides in this genomic position) within the RT region (positions 2550–3554) is a clear evidence of concurrent presence of different HIV-1 variants in the sample [2]. The larger the number of degenerate nucleotides (the DB count) is in the sequence, the higher the likelihood of dual HIV infection in the sample.

Another approach focuses on estimation of the ratio between the number of synonymous substitutions and the number of all potential sites of synonymous substitutions within the PR-RT region, positions 2253–3554 (SM count) [8]. Nonsynonymous mutations (causing changes in amino acid sequences) tend to be ignored. This method is based on the specific features of nonsynonymous mutations, which often result from changes in the virus affected by host body factors or by administered therapy, while synonymous mutations reflect natural diversity of virus population in dual HIV infection. Therefore, for dual HIV infection, the detected mutations will be primarily of the synonymous type.

It is understandable that the analysis of degeneracy and synonymy is highly efficient for dual HIV infection caused by different HIV subtypes (inter-subtype), different groups, for example, M and O (inter-group), or virus types (intertype infection or concurrent HIV-1 and HIV-2 infection) as compared to intra-subtype infection caused by different genetic variants of the same subtype [1].

Application of PS for detection of dual HIV infection may cause problems due to sensitivity of the above method: The above sequencing test can detect viral variants accounting for at least 20% of the virus population [16]. The solution is offered by the NGS technology, which proved to be successful in detection of drug resistance-associated mutations within the PR-RT region [17]. At the same time, special attention should be paid to the NGS sensitivity threshold for minor virus variants. For example, for the routine analysis of drug resistance the sensitivity threshold of 5% is most reliable and informative [17, 18].

Since the approach to the OUT analysis for a small RT fragment (positions 2708–3242) turned out to be successful [8], the application of NGS technology for sequencing the PR-RT region can help detect dual HIV infection (by estimation of DB and SM counts) and confirm it.

The **aim** of the study was to use comparative analysis for assessing efficiency of detection of dual HIV infection, using PS and NGS with different sensitivity thresholds for minor viral variants for the PR-RT region, as well as to evaluate the efficiency of the NGS-based results for confirmation of dual HIV infection using the OTU analysis.

Materials and methods

The model of inter-subtype dual HIV infection was created by pairwise mixing of plasma samples containing 3 variants of HIV-1: subtype B, sub-subtype A6, and recombinant form CRF63_02A1. The HIV-1 RNA concentration in each sample for correct mixing was estimated using an AmpliSens HIV-monitor-FRT kit (Central Research Institute of Epidemiology). A total of 9 variants of pairwise mixtures of viruses were prepared, each virus accounting for 10 to 90% in mixtures. In addition, tests included control samples containing only HIV-1 of subtype B, sub-subtype A6, and

CRF63_02A1. The viral load in each model sample was 5,000 cps/ml.

The conventional PS was performed using an AmpliSens HIV-Resist-Seq kit (Central Research Institute of Epidemiology) according to manufacturer's manual. Sequencing of the purified fragments was performed with an Applied Biosystems 3500 genetic analyzer (LifeTechnologies) according to manufacturer's manual. The sequencing data were processed using DEONA software (version 1.7.0). As a result, we obtained NSs of the HIV-1 PR-RT fragment (nucleotide positions 2253–3368 for HXB-2, GenBank accession number K03455).

When preparing samples for NGS, we used a two-step nested PCR protocol to obtain 4 overlapping specific HIV-1 DNA fragments that were 427–586 nucleotides long. During the first step, amplification was combined with reverse transcription using TaqF polymerase and MMLV reverse transcriptase (Central Research Institute of Epidemiology). The amplified fragments of the HIV-1 genome were purified using Sera-Mag Magnetic Speed Beads (GE Healthcare Biosciences). Concentrations of nucleic acids in the fragments were measured with a Qubit 2.0 fluorometer (Invitrogen). The purified and amplified HIV-1 fragment samples were mixed in equal proportions, and 50 ng of the final mixture were used for sequencing libraries with the Illumina platform. Libraries were prepared following the Nextera protocol (Illumina) modified as follows: amplification was performed using Q5 High-Fidelity DNA polymerase (NEB) containing intercalating EvaGreen dye (Biotium). Sequencing was performed using MiSeq (Illumina) and a MiSeq Reagent Kit V3 (Illumina).

The sequencing results were processed and the genome (positions 2253-3368) was assembled using HyDRA Web v1.6.1 software[1]. Consensus NSs were automatically assembled and had the 20, 15, 10, and 5% sensitivity threshold for minor populations (NGS20, NGS15, NGS10, and NGS5, respectively).

DB and SM counts were calculated for the received NSs in accordance with the established methods [2, 8].

The DB count was calculated as the total number of degenerate nucleotide bases in the RT region (nucleotide positions 2550–3368). Two values: 34 and 45 were verified as threshold values for the DB count indicative of dual HIV infection [2].

The SM count was calculated by the formula: $SM = X/372$, where SM was synonymous mutation count, X — the total number of synonymous substitutions in the studied genomic fragment, 372 — the total number of amino acids encoded by the PR-RT fragment of *pol* gene (nucleotide positions 2253–3368). Values of 0.05 and 0.08 were selected as possible thresholds for the SM count [8].

The statistical analysis of differences between values of DB and SM counts, which were obtained using different techniques and for different model samples, was based on Student's t-test. Qualitative variables (efficiency detection dual HIV infection) were analyzed using two-sided Fisher's exact test [19]. Differences were considered statistically significant at $p < 0.01$.

OTUs for NGS results were clustered for 1 of 4 fragments encoding part of the reverse transcriptase (nucleotide positions 2725–2981) in accordance with the recommendations[2]. The Mega 6.0 software was used for the phylogenetic analysis based on the maximum likelihood method and for searching the optimum model of nucleotide substitutions for the obtained OTUs. We used the following NSs from the international GenBank sequence database as reference NSs: K03455 and AY819715 (subtype B), AY500393 and EU861977 (sub-subtype A6), AY829204 and JN230353 (CRF63_02A1). Along with OTUs of model samples, we analyzed the OUT variants received for control samples, which were used to create the models of dual HIV infection.

Results

Analysis of DB and SM counts in models of dual HIV infection

We have found that the DB and SM counts depend not only on the viral variants present in the mixture, but also on their ratio (**Fig. 1, 2**). Both counts for BvsA6 model samples containing sub-subtype A6 as the minor variant were higher than those in samples containing minor amounts of the subtype B virus. The similar effect was observed in CRF63vsB and CRF63vsA6 models, where the recombinant virus was present in minor amounts. It is indicative of dominance of the sub-subtype A6 virus in BvsA6 models and dominance of CRF63 in CRF63vsB and CRF63vsA6 models, thus decreasing the degeneracy level in the obtained NSs.

As expected, NGS-derived DB and SM counts in NSs of the studied models were generally higher than in NSs obtained by using the PS method. The average DB count for all model samples obtained by using PS was 25.59 (95% CI 15.6–35.54), while for NGS20, NGS15, NGS10, NGS5, it was 27.48 (95% CI 18.27–36.69), 37.44 (95% CI 27.56–47.33), 49.19 (95% CI 38.17–60.20), and 68.19 (95% CI 55.81–80.56), respectively. Nevertheless, statistically significant differences ($p < 0.01$) in the DB counts were revealed only by comparison of PS or NGS20 with NGS10 and NGS5 as well as between NGS15 and NGS5.

The average SM count also increased significantly when the sensitivity of sequencing increased: For PS, it was 0.069 (95% CI 0.046–0.092), and for NGS20, NGS15, NGS10, NGS5 — 0.067 (95% CI 0.048–

ОРИГИНАЛЬНЫЕ ИССЛЕДОВАНИЯ

Fig. 1. DB counts obtained by using PS and NGS20-NGS5 technologies for CRF63vsB, CRF63vsA6, and BvsA6 model samples of dual HIV infection.

Horizontal hatched lines indicate the specified threshold values for the count.
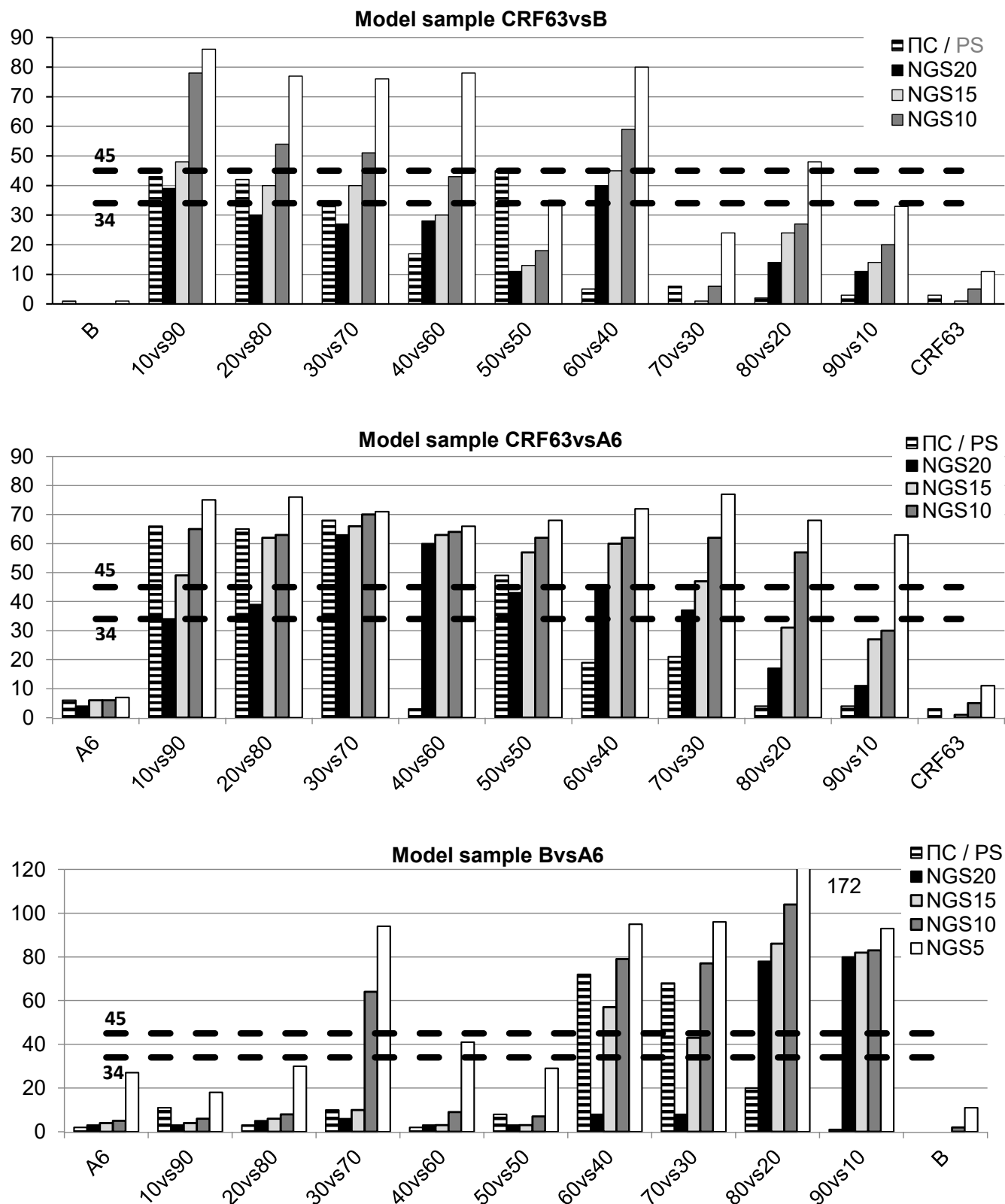
Fig. 2. SM counts obtained by using PS and NGS20-NGS5 technologies for CRF63vsB, CRF63vsA6, and BvsA6 model samples of dual HIV infection.

Horizontal hatched lines indicate the specified threshold values for the count.
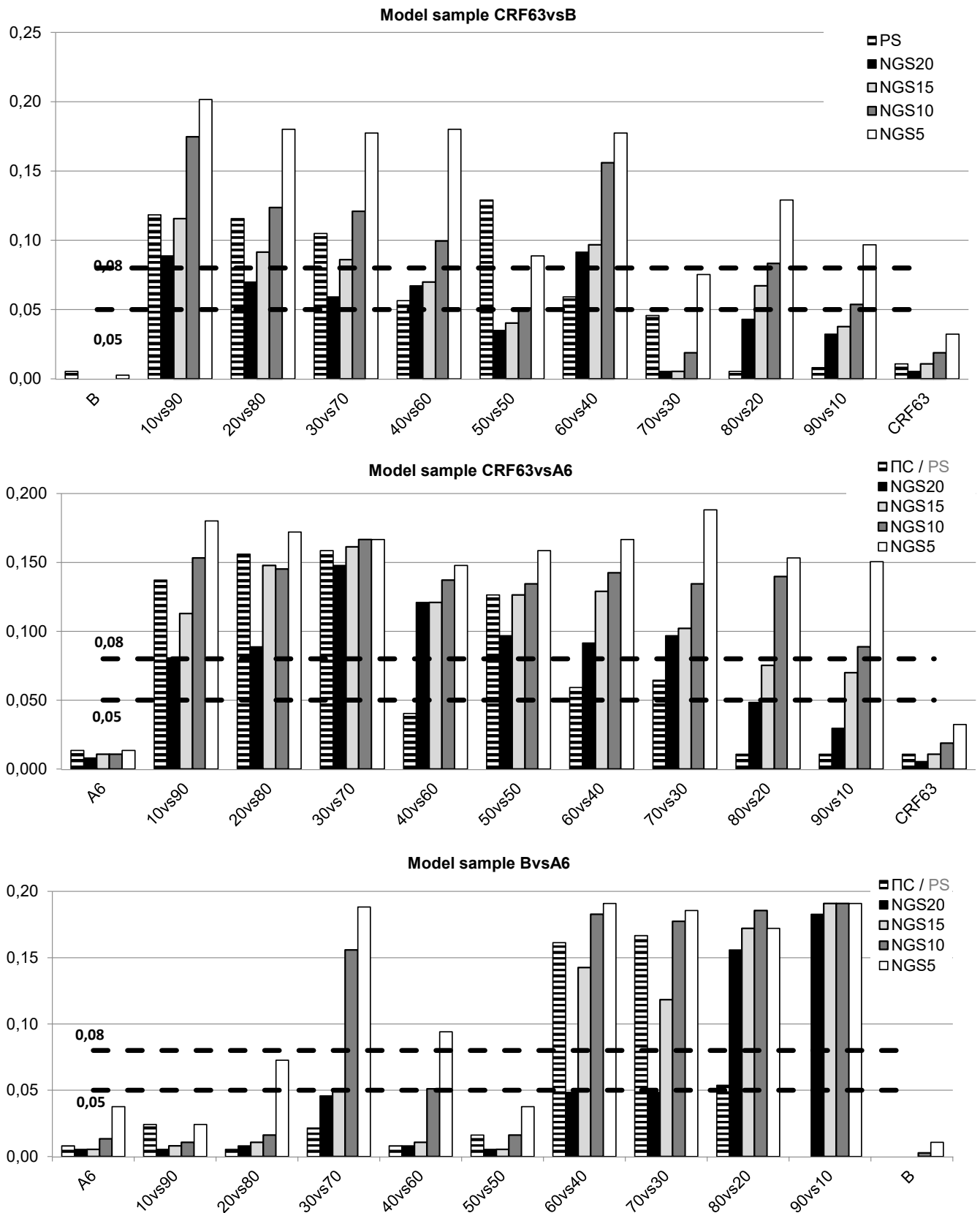
0.086), 0.088 (95% CI 0.066–0.109), 0.115 (95% CI 0.092–0.138), and 0.146 (95% CI 0.126–0.166), respectively. Similarly to the DB count, statistically significant differences were those between PS or NGS20 and NGS10–NGS5. However, even the NGS15 and NGS10 results demonstrated significant differences, thus giving evidence of high efficiency of the NGS technology used for the analysis of the SM count.

In the meantime, this significant increase in the degeneracy level in NGS5 compared to PS was demonstrated by virus mixtures rather than by control samples containing only A6, B, or CRF06_02A1. The average DB count estimated for the control samples by using the PS method was 2.50 (95% CI 1.68–3.32); for NGS20, NGS15, NGS10, and NGS5, it was 1.17 (95% CI 0.44–1.89), 2.00 (95% CI 1.03–2.97), 3.83 (95% CI 2.97–4.75), and 11.33 (95% CI 7.92–14.74), respectively. The average SM count for PS and all NGS types was 0.008 (95% CI 0.006–0.010), 0.004 (95% CI 0.003–0.005), 0.006 (95% CI 0.004–0.008), 0.011 (95% CI 0.008–0.014), and 0.022 (95% CI 0.016–0.027), respectively. These differences between the sequencing methods were statistically insignificant ($p > 0.01$). Neither the DB count nor the SM count estimated for control samples using any method exceeded the minimum threshold value specified in the literature [2, 8] for this analysis, i.e. 34 for the DB count and 0.05 for the SM count.

At the same time, the increased counts associated with the increased sensitivity of the sequencing method led to identification of a larger number of variants of dual HIV infection models. The threshold value of 34 for the DB count was exceeded only by 10 of 27 model variants analyzed by PS, while the NGS20-NGS5 methods increased their numbers to 11, 15, 18, and 22, respectively. When the threshold value of 0.05 for the SM

count was used, dual HIV infection was detected in 15, 14, 20, 23, and 25 variants for PS and NGS20–NGS5, respectively. The lowest counts (and, consequently, the likelihood of detection of dual HIV infection) were obtained for BvsA6 mixtures. With the DB count, only differences between PS and NGS5 were statistically significant; with the SM count, differences between PS or NGS20 and NGS10–NGS5 were statistically significant, thus giving another proof of high efficiency of the SM count for detection of dual HIV infection.

Although the NGS methods demonstrated higher sensitivity compared to PS, the analysis of the average SM count for each type of the models revealed an insignificant decrease in the above count in CRF63vsB mixtures by using NGS20 and NGS15 compared to PS (0.055 and 0.068 against 0.071, respectively) (**Fig. 3**).

NGS20 also decreased insignificantly the average DB count for BvsA6 models compared to PS — 21.56 against 21.67.

Analysis of OTUs

On average, 8, 9, and 11 OTU types were received for each of 9 variants of model samples of BvsA6, CRF63vsB, and CRF63vsA6 groups. The value of 90% was selected as the optimum similarity threshold for reads. Threshold values above 90% led to generation of multiple OTUs and dubious results of the subsequent phylogenetic analysis. Lower threshold values resulted in a sharp reduction of OTU variants, thus having an adverse impact on the accuracy of the subsequent analysis.

The results of OTU clustering correlated with the results of the DB and SM count analysis. Besides, the distribution of OTU variants demonstrated a high level of non-uniformity for different viral variants: In the BvsA6 models, 77.78% of OTUs belonged to A6 and
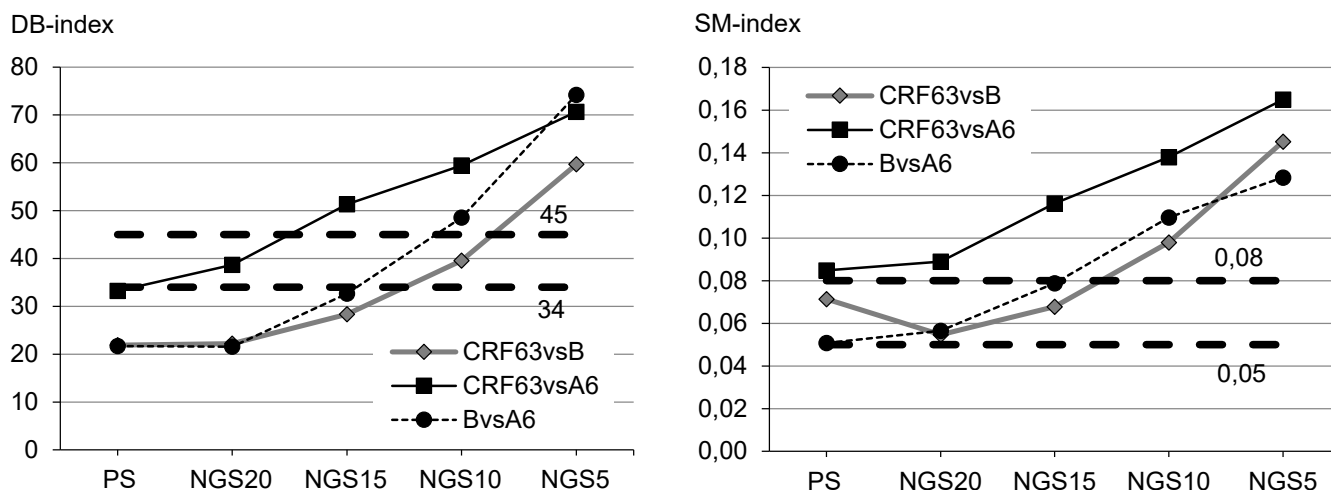
Fig. 3. The average DB and SM counts estimated with PS and NGS20–NGS5 methods for CRF63vsB, CRF63vsA6, and BvsA6 model samples of dual HIV infection.
Horizontal hatched lines indicate the specified threshold values for DB and SM counts.

only 22.22% belonged to subtype B. In the CRF63vsB models, 69.27% of OTUs belonged to recombinant, while 11.11% — to subtype B; 19.63% of OTU variants showed controversial results, forming statistically insignificant clusters with different reference NSs (mainly, with A6). In the CRF63vsA6 models, the recombinant, sub-subtype A6 accounted for 57.10% and 41.15% of OTUs, respectively. At the same time, 1.75% of OTUs demonstrated controversial results.

The results of the analysis of OTUs clustered for the models of dual HIV infection, in which each virus accounted for 30 and 70%, are presented in **Fig. 4**. We did not find any clear correlation between the results of genotyping of OTU variants and the virus content in the sample. For example, for CRF63vsA6(70vs30), 7 OTU variants were close to CRF63, and 4 variants – to A6. On the other hand, OTUs for CRF63vsA6(30vs70) were distributed between CRF63 and A6 almost evenly.

Similar results were demonstrated by the analysis of CRF63vsB. The distribution of OTUs between CRF63 and B for CRF63vsB(30vs70) was 9 to 4; for CRF63vsB(70vs30), it was 9 to 2. However, the analysis of the BvsA6 model samples demonstrated the opposite result. In the BvsA6(30vs70) sample, 6 OTUs belonged to sub-subtype A6 and 3 OTUs belonged to subtype B, while in the BvsA6(70vs30) sample, 7 and 2 OTUs belonged to A6 and B, respectively.

Nevertheless, the results of genotyping of OTUs obtained for models of dual HIV infection differed from those for A6, B, and CRF63_02A1 control samples, for which all the OTU variants truly belonged to the respective genetic variant.

Discussion

For the first time in Russia, we have validated the method of analysis of HIV-1 NSs obtained by using conventional sequencing and NGS for the analysis of the rare and critically important phenomenon — dual HIV infection.

In our study, we focused on model samples of dual HIV infection caused by concurrent presence of the viral variants in clinical samples, which are most important for Russia from the epidemiological perspective. In Russia, sub-subtype A6 HIV-1 prevails, though the CRF63_02A1 recombinant variant has been actively spreading in the country in the recent years [20–22]. We also studied the subtype B virus as a reference virus, which has high prevalence in countries of Western Europe and is used for validation of dual HIV infection evaluation methods [2, 8, 20].

The NS analysis used for detection of dual HIV infection is not only less time-consuming and more economically beneficial than the SGS technology, but also can be used for dual HIV infection screening similar to the analysis of drug resistance. In our study, the application of DB and SM counts in PR-RT sequences obtained with the PS method resulted in detection of dual HIV infection only in 37% and 55.56% of cases, respectively.

Our results were of lower quality than the results obtained by foreign authors. For example, DB counts may result in detection of 43.2% of dual HIV infection samples [2], while SM counts may increase the detection rate to 100% of samples [8]. However, foreign colleagues analyzed PR-RT fragments that were longer than those in our study: 335 [2] and 434 [8] amino acids for estimation of DB and SM counts, respectively, compared to 273 and 372 amino acids in our study. Shorter fragments contain smaller numbers of potentially degenerate positions. Furthermore, in the study by Pacold *et al.* [8], based on the calculation of SM counts, several monoinfection samples were mistakenly assigned to dual HIV infection samples, while no mistakes of this kind were recorded in our tests.

Our results demonstrate that application of NGS technology with the 5–10% sensitivity threshold for minor variants makes it possible to detect dual HIV infection more efficiently than when using conventional sequencing. The test becomes most efficient for detection of inter-subtype dual HIV infection in the *pol* gene region (positions 2253–3368) when NGS technology has a 5% sensitivity threshold for minor variants and the SM count has the threshold value of 0.05 for dual infection. These parameters are important for detection of the largest number of dual HIV infection cases (in 25 out of 27 model samples).

Although the sensitivity of conventional sequencing for minor viral variants in the sample is usually estimated at 20% [16], NGS with the same sensitivity threshold showed poorer performance in detecting degeneracy in some models of dual HIV infection. On the one hand, this property of NGS20 may minimize the number of errors and artifacts resulting from sequencing, being an important feature for obtaining whole-genome sequences of the virus [11, 17]. On the other hand, the NGS20 method may pose a risk of missing the degeneracy associated with dual HIV infection in the sample.

For most of the model samples, the application of the OTU analysis for the genomic region with positions 2725–2981 proved to be efficient. This is confirmed by statistical clusters (>80%) in diagrams, featuring respective control and reference sequences. Our results are similar to the data obtained during the study of dual HIV infection among MSM in Spain [9]. However, there was a discrepancy between the proportion of OTUs and the proportion of viruses in model samples. For example, in the 63vsB and BvsA6 mixtures, most of the OTUs were genotyped as CRF63_02A1 and A6, respectively. It can be explained by the divergence of sub-subtype A6 and CRF63_02A1 viruses, which is much higher than in subtype B. When the threshold value is 90%, the reads of subtype B are removed automatically. This obvious drawback inherent in this
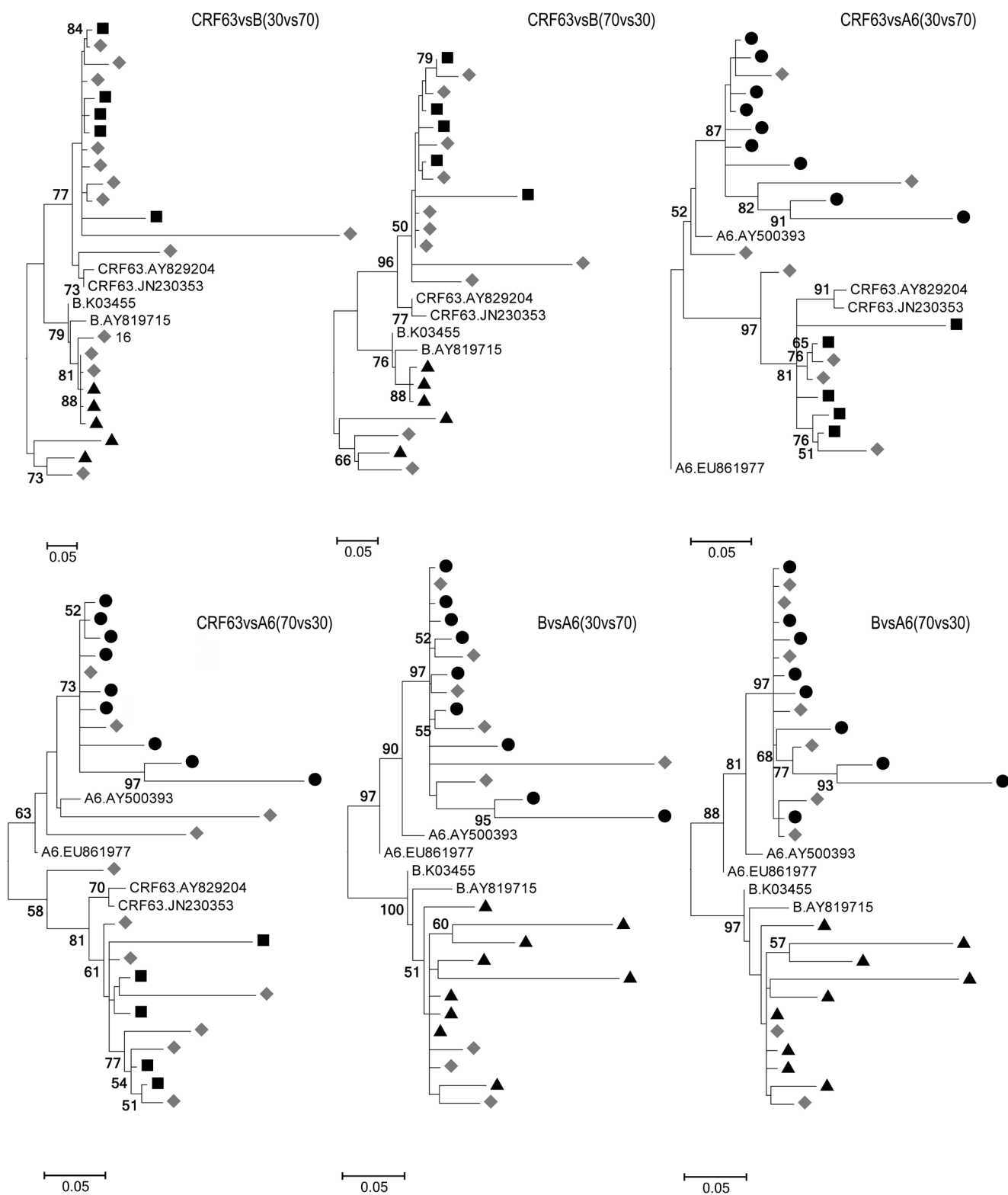
Fig. 4. Results of the phylogenetic analysis of the fragment from the RT region (positions 2725–2981) of dual HIV infection model samples and control samples; the analysis was based on the maximum likelihood method and used the GTR+G model of nucleotide substitutions in 1,000 bootstrap replicates.

The statistical significance of clusters ≥ 50% is shown next to them. Grey diamonds indicate OTU variants obtained for the model sample. The OTU sequences obtained for the control sample of sub-subtype A6 are shown by black circles; the subtype B sample is shown by black triangles; black squares represent CRF63_02A1.

method must be taken in consideration in studies of dual infection caused by different HIV types. In addition, such discrepancy of OTU results may imply that the systems used for amplification and/or sequencing amplify and sequence viruses with different efficiency. As a result, the signal from a more efficiently detected virus suppresses the signal from another variant. Such non-uniform sensitivity to different virus variants was pointed out previously during the comparative assessment of PCR kits for estimation of the HIV-1 viral load [23].

Nevertheless, despite all its downsides, the OTU analysis makes it possible to reliably differentiate dual HIV infection from monoinfection in most model samples and can be used in further studies of multiple infections with HIV ВИЧ.

Note that active application of conventional sequencing and NGS makes these technologies more affordable, reducing the testing time and cost compared to the SGS method. Our analysis of one sample by using conventional sequencing cost around 10,000 rubles, being almost twice as cheap as the similar study conducted in 2010 [8]. Thanks to its enormous throughput, NGS is very cost-effective: The cost of the analysis of the genomic fragment of each sample in our study was around 5,000 rubles, being 5 times lower than the cost of the similar NGS assay performed in 2010 [8].

Conclusion

Our results demonstrate that the NGS technology can be of great use for studying the phenomenon of dual HIV infection, providing high sensitivity of the test. Thus, this technology, when used in routine practice, will help not only identify genetic characteristics of infectious agents, but also obtain more data for more efficient assessment of the epidemiological situation.

R E F E R E N C E S

1. Lapovok I.A., Lopatukhin A.E., Kireev D.E. Dual HIV infection: epidemiology, clinical significance, and diagnosis. *Infektsionnye bolezni.* 2019; 17(2): 87–93. https://doi.org/10.20953/1729-9225-2019-2-87-93 (in Russian)

2. Cornelissen M., Jurriaans S., Kozaczynska K., Prins J.M., Hamidjaja R.A., Zorgdrager F., et al. Routine HIV-1 genotyping as a tool to identify dual infections. *AIDS.* 2007; 21(7): 807–11. https://doi.org/10.1097/QAD.0b013e3280f3c08a

3. Soares de Oliveira A.C., Pessoa de Farias R., da Costa A.C., Melillo Sauer M., Bassichetto K.C., Santos Oliveira S.M., et al. Frequency of subtype B and F1 dual infection in HIV-1 positive, Brazilian men who have sex with men. *Virol. J.* 2012; 9: 223. https://doi.org/10.1186/1743-422X-9-223

4. Redd A.D., Ssemwanga D., Vandepitte J., Wendel S.K., Ndembi N., Bukenya J., et al. The rates of HIV-1 superinfection and primary HIV-1 infection are similar in female sex workers in Uganda. *AIDS.* 2014; 28(14): 2147–52. https://doi.org/10.1097/QAD.0000000000000365

5. Luan H., Han X., Yu X., An M., Zhang H., Zhao B., et al. Dual infection contributes to rapid disease progression in men who have sex with men in China. *J. Acquir. Immune Defic. Syndr.* 2017; 75(4): 480–7. https://doi.org/10.1097/qai.0000000000001420

6. Lapovok I.A., Saleeva D.V., Kirichenko A.A., Murzakova A.V., Lopatukhin A.E., Kireev D.E. The study of dual HIV infection prevalence in Russia. *Infektsionnye bolezni.* 2020; 18(4): 138–48. https://doi.org/10.20953/1729-9225-2020-4-138-148 (in Russian)

7. Gao Y., Tian W., Han X., Gao F. Immunological and virological characteristics of human immunodeficiency virus type 1 superinfection: implications in vaccine design. *Front. Med.* 2017; 11(4): 480–9. https://doi.org/10.1007/s11684-017-0594-8

8. Pacold M., Smith D., Little S., Cheng P.M., Jordan P., Ignacio C., et al. Comparison of methods to detect HIV dual infection. *AIDS Res. Hum. Retroviruses.* 2010; 26(12): 1291–8. https://doi.org/10.1089/aid.2010.0042

9. Casado C., Pernas M., Rava M., Ayerdi O., Vera M., Alenda R., et al. High-risk sexual practices contribute to HIV-1 double infection among men who have sex with men in Madrid. *AIDS Res. Hum. Retroviruses.* 2020; 36(11): 896–904. https://doi.org/10.1089/AID.2020.0068

10. Popescu B., Banica L., Nicolae I., Radu E., Niculescu I., Abagiu A., et al. NGS combined with phylogenetic analysis to detect HIV-1 dual infection in Romanian people who inject drugs. *Microbes Infect.* 2018; 20(5): 308–11. https://doi.org/10.1016/j.micinf.2018.03.004

11. Kireev D.E., Lopatukhin A.E., Murzakova A.V., Pimkina E.V., Speranskaya A.S., Neverov A.D., et al. Evaluating the accuracy and sensitivity of detecting minority HIV-1 populations by Illumina next-generation sequencing. *J. Virol. Methods.* 2018; 261: 40–5. https://doi.org/10.1016/j.jviromet.2018.08.001

12. Yerly S., Jost S., Monnat M., Telenti A., Cavassini M., Chave J.P., et al. HIV-1 co/super-infection in intravenous drug users. *AIDS.* 2004; 18(10): 1413–21. https://doi.org/10.1097/01.aids.0000131330.28762.0c

13. Ring K., Muir D., Mackie N., Bailey A.C. HIV-1 superinfection in a patient with known HIV-2 — A case report. *Int. J. STD AIDS.* 2020; 31(3): 271–3. https://doi.org/10.1177/0956462420901969

14. Wagner G.A.., Pacold M.E., Kosakovsky Pond S.L., Caballero G., Chaillon A., Rudolph A.E., et al. Incidence and prevalence of intrasubtype HIV-1 dual infection in at-risk men in the United States. *J. Infect. Dis.* 2014; 209(7): 1032–8. https://doi.org/10.1093/infdis/jit633

15. Van der Kuyl A.C., Cornelissen M. Identifying HIV-1 dual infections. *Retrovirology.* 2007; 4: 67. https://doi.org/10.1186/1742-4690-4-67

16. Inzaule S.C., Hamers R.L., Noguera-Julian M., Casadellà M., Parera M., Kityo C., et al. Clinically relevant thresholds for ultrasensitive HIV drug resistance testing: a multi-country nested case-control study. *Lancet HIV.* 2018; 5(11): e638–46. https://doi.org/10.1016/S2352-3018(18)30177-2

17. Kirichenko A.A., Sviridova A.A., Lopatukhin A.E., Murzakova A.V., Lapovok I.A., Goptar' I.A., et al. Correlation of results analysis drug resistance of human immunodeficiency virus among patients with virological failure by next-generation sequencing and traditional population sequencing. *Infektsionnye bolezni.* 2019; 17(2): 12–9. https://doi.org/10.20953/1729-9225-2019-2-12-19 (in Russian)

18. Trabaud M.A., Icard V., Ramière C., Tardy J.C., Scholtes C., André P. Comparison of HIV-1 drug-resistance genotyping by ultra-deep sequencing and sanger sequencing using clinical samples. *J. Med. Virol.* 2017; 89(11): 1912–9. https://doi.org/10.1002/jmv.24872

19. Gerasimov A.N. *Medical Statistics: A Textbook [Meditsinskaya statistika: Uchebnoe posobie].* Moscow: MIA; 2007. (in Russian)

20. Foley B.T., Leitner T., Paraskevis D., Peeters M. Primate immunodeficiency virus classification and nomenclature: Review. *Infect. Genet. Evol.* 2016; 46: 150–8. https://doi.org/10.1016/j.meegid.2016.10.018

21. Lapovok I.A., Lopatukhin A.E., Kireev D.E., Kazennova E.V., Lebedev A.V., Bobkova M.R., et al. Molecular epidemiological

ОРИГИНАЛЬНЫЕ ИССЛЕДОВАНИЯ

analysis of HIV-1 variants circulating in Russia in 1987–2015. *Terapevticheskiy arkhiv.* 2017; 89(11): 44–9. https://doi.org/10.17116/terarkh2017891144-49 (in Russian)

22. Baryshev P., Bogachev V., Gashnikova N. HIV-1 genetic diversity in Russia: CRF63_02A1, a new HIV type 1 genetic variant spreading in Siberia. *AIDS Res. Hum. Retroviruses.* 2014; 30(6): 592–7. https://doi.org/10.1089/AID.2013.0196

23. Holmes H., Davis C., Heath A. Development of the 1st International Reference Panel for HIV-1 RNA genotypes for use in nucleic acid-based techniques. *J. Virol. Methods.* 2008; 154 (1–2): 86–91. https://doi.org/10.1016/j.jviromet.2008.08.014

СПИСОК ИСТОЧНИКОВ

1. Лаповок И.А., Лопатухин А.Э., Киреев Д.Е. Двойная ВИЧ-инфекция: эпидемиология, клиническая значимость и диагностика. *Инфекционные болезни.* 2019; 17(2): 87–93. https://doi.org/10.20953/1729-9225-2019-2-87-93

2. Cornelissen M., Jurriaans S., Kozaczynska K., Prins J.M., Hamidjaja R.A., Zorgdrager F., et al. Routine HIV-1 genotyping as a tool to identify dual infections. *AIDS.* 2007; 21(7): 807–11. https://doi.org/10.1097/QAD.0b013e3280f3c08a

3. Soares de Oliveira A.C., Pessoa de Farias R., da Costa A.C., Melillo Sauer M., Bassichetto K.C., Santos Oliveira S.M., et al. Frequency of subtype B and F1 dual infection in HIV-1 positive, Brazilian men who have sex with men. *Virol. J.* 2012; 9: 223. https://doi.org/10.1186/1743-422X-9-223

4. Redd A.D., Ssemwanga D., Vandepitte J., Wendel S.K., Ndembi N., Bukenya J., et al. The rates of HIV-1 superinfection and primary HIV-1 infection are similar in female sex workers in Uganda. *AIDS.* 2014; 28(14): 2147–52. https://doi.org/10.1097/QAD.0000000000000365

5. Luan H., Han X., Yu X., An M., Zhang H., Zhao B., et al. Dual infection contributes to rapid disease progression in men who have sex with men in China. *J. Acquir. Immune Defic. Syndr.* 2017; 75(4): 480–7. https://doi.org/10.1097/qai.0000000000001420

6. Лаповок И.А., Салеева Д.В., Кириченко А.А., Мурзакова А.В., Лопатухин А.Э., Киреев Д.Е. Исследование частоты встречаемости двойной ВИЧ-инфекции в России. *Инфекционные болезни.* 2020; 18(4): 138–48. https://doi.org/10.20953/1729-9225-2020-4-138-148

7. Gao Y., Tian W., Han X., Gao F. Immunological and virological characteristics of human immunodeficiency virus type 1 superinfection: implications in vaccine design. *Front. Med.* 2017; 11(4): 480–9. https://doi.org/10.1007/s11684-017-0594-8

8. Pacold M., Smith D., Little S., Cheng P.M., Jordan P., Ignacio C., et al. Comparison of methods to detect HIV dual infection. *AIDS Res. Hum. Retroviruses.* 2010; 26(12): 1291–8. https://doi.org/10.1089/aid.2010.0042

9. Casado C., Pernas M., Rava M., Ayerdi O., Vera M., Alenda R., et al. High-risk sexual practices contribute to HIV-1 double infection among men who have sex with men in Madrid. *AIDS Res. Hum. Retroviruses.* 2020; 36(11): 896–904. https://doi.org/10.1089/AID.2020.0068

10. Popescu B., Banica L., Nicolae I., Radu E., Niculescu I., Abagiu A., et al. NGS combined with phylogenetic analysis to detect HIV-1 dual infection in Romanian people who inject drugs. *Microbes Infect.* 2018; 20(5): 308–11. https://doi.org/10.1016/j.micinf.2018.03.004

11. Kireev D.E., Lopatukhin A.E., Murzakova A.V., Pimkina E.V., Speranskaya A.S., Neverov A.D., et al. Evaluating the accuracy and sensitivity of detecting minority HIV-1 populations by Illumina next-generation sequencing. *J. Virol. Methods.* 2018; 261: 40–5. https://doi.org/10.1016/j.jviromet.2018.08.001

12. Yerly S., Jost S., Monnat M., Telenti A., Cavassini M., Chave J.P., et al. HIV-1 co/super-infection in intravenous drug users. *AIDS.* 2004; 18(10): 1413–21. https://doi.org/10.1097/01.aids.0000131330.28762.0c

13. Ring K., Muir D., Mackie N., Bailey A.C. HIV-1 superinfection in a patient with known HIV-2 – A case report. *Int. J. STD AIDS.* 2020; 31(3): 271–3. https://doi.org/10.1177/0956462420901969

14. Wagner G.A.., Pacold M.E., Kosakovsky Pond S.L., Caballero G., Chaillon A., Rudolph A.E., et al. Incidence and prevalence of intrasubtype HIV-1 dual infection in at-risk men in the United States. *J. Infect. Dis.* 2014; 209(7): 1032–8. https://doi.org/10.1093/infdis/jit633

15. Van der Kuyl A.C., Cornelissen M. Identifying HIV-1 dual infections. *Retrovirology.* 2007; 4: 67. https://doi.org/10.1186/1742-4690-4-67

16. Inzaule S.C., Hamers R.L., Noguera-Julian M., Casadellà M., Parera M., Kityo C., et al. Clinically relevant thresholds for ultrasensitive HIV drug resistance testing: a multi-country nested case-control study. *Lancet HIV.* 2018; 5(11): e638–46. https://doi.org/10.1016/S2352-3018(18)30177-2

17. Кириченко А.А., Свиридова А.А., Лопатухин А.Э., Мурзакова А.В., Лаповок И.А., Гоптарь И.А. и др. Корреляция результатов высокопроизводительного и классического методов секвенирования при анализе лекарственной устойчивости вируса иммунодефицита человека у пациентов на фоне неэффективной антиретровирусной терапии. *Инфекционные болезни.* 2019; 17(2): 12–9. https://doi.org/10.20953/1729-9225-2019-2-12-19

18. Trabaud M.A., Icard V., Ramière C., Tardy J.C., Scholtes C., André P. Comparison of HIV-1 drug-resistance genotyping by ultra-deep sequencing and sanger sequencing using clinical samples. *J. Med. Virol.* 2017; 89(11): 1912–9. https://doi.org/10.1002/jmv.24872

19. Герасимов А.Н. *Медицинская статистика: учебное пособие.* М.: МИА; 2007.

20. Foley B.T., Leitner T., Paraskevis D., Peeters M. Primate immunodeficiency virus classification and nomenclature: Review. *Infect. Genet. Evol.* 2016; 46: 150–8. https://doi.org/10.1016/j.meegid.2016.10.018

21. Лаповок И.А., Лопатухин А.Э., Киреев Д.Е., Казеннова Е.В., Лебедев А.В., Бобкова М.Р. и соавт. Молекулярно-эпидемиологический анализ вариантов ВИЧ-1, циркулировавших в России в 1987–2015 гг. *Терапевтический архив.* 2017; 89(11): 44–9. https://doi.org/10.17116/terarkh2017891144-49

22. Baryshev P., Bogachev V., Gashnikova N. HIV-1 genetic diversity in Russia: CRF63_02A1, a new HIV type 1 genetic variant spreading in Siberia. *AIDS Res. Hum. Retroviruses.* 2014; 30(6): 592–7. https://doi.org/10.1089/AID.2013.0196

23. Holmes H., Davis C., Heath A. Development of the 1st International Reference Panel for HIV-1 RNA genotypes for use in nucleic acid-based techniques. *J. Virol. Methods.* 2008; 154(1-2): 86–91. https://doi.org/10.1016/j.jviromet.2008.08.014

Информация об авторах

Лаповок Илья Андреевич✉ — к.б.н., с.н.с. Научной группы разработки новых методов диагностики ВИЧ-инфекции и вирусных гепатитов ЦНИИ эпидемиологии, Москва, Россия, i_lapovok@mail.ru, https://orcid.org/0000-0002-6328-1415

Барышев Павел Борисович — к.б.н., биоинформатик Научной группы разработки новых методов диагностики ВИЧ-инфекции и вирусных гепатитов ЦНИИ Эпидемиологии Москва, Россия, https://orcid.org/0000-0003-3666-2201

Салеева Дарья Владиславовна — м.н.с., Научной группы разработки новых методов диагностики ВИЧ-инфекции и вирусных гепатитов ЦНИИ эпидемиологии, Москва, Россия, https://orcid.org/0000-0002-5870-5594

Кириченко Алина Алексеевна — н.с. Научной группы разработки новых методов диагностики ВИЧ-инфекции и вирусных гепатитов ЦНИИ эпидемиологии, Москва, Россия, https://orcid.org/0000-0002-7116-0138

Шлыкова (Мурзакова) Анастасия Вениаминовна — н.с. Научной группы разработки новых методов диагностики ВИЧ-инфекции и вирусных гепатитов ЦНИИ эпидемиологии, Москва, Россия, https://orcid.org/0000-0002-1390-8021

Киреев Дмитрий Евгеньевич — к.б.н., н.с. Научной группы разработки новых методов диагностики ВИЧ-инфекции и вирусных гепатитов ЦНИИ эпидемиологии, Москва, Россия, https://orcid.org/0000-0002-7896-2379

Участие авторов. Все авторы внесли существенный вклад в проведение поисково-аналитической работы и подготовку статьи, прочли и одобрили финальную версию до публикации.

Information about the authors

Ilya A. Lapovok✉ — Cand. Sci. (Biol.), senior researcher, HIV and viral hepatitis research group, Central Research Institute of Epidemiology, Moscow, Russia, i_lapovok@mail.ru, https://orcid.org/0000-0002-6328-1415

Pavel B. Baryshev — Cand. Sci. (Biol.), bioinformatist, HIV and viral hepatitis research group, Central Research Institute of Epidemiology, Moscow, Russia, https://orcid.org/0000-0003-3666-2201

Daria V. Saleeva — junior researcher, HIV and viral hepatitis research group, Central Research Institute of Epidemiology, Moscow, Russia, https://orcid.org/0000-0002-5870-5594

Alina A. Kirichenko — researcher, HIV and viral hepatitis research group, Central Research Institute of Epidemiology, Moscow, Russia, https://orcid.org/0000-0002-7116-0138

Anastasiia V. Shlykova (Murzakova) — researcher, HIV and viral hepatitis research group, Central Research Institute of Epidemiology, Moscow, Russia, https://orcid.org/0000-0002-1390-8021

Dmitry E. Kireev — Cand. Sci. (Biol.), researcher, HIV and viral hepatitis research group, Central Research Institute of Epidemiology, Moscow, Russia, https://orcid.org/0000-0002-7896-2379

Author contribution. All authors made a substantial contribution to the conception of the work, acquisition, analysis, interpretation of data for the work, drafting and revising the work, final approval of the version to be published.