UDC 619:616.98:578:636.5:616-076

IMPLEMENTATION OF 454 LIFE SCIENCES

TECHNOLOGY LABORATORY PRACTICES

N.G. Zinyakov[1], Ye.V. Ovchinnikova[2], S.P. Lazareva[3], A.A. Kozlov[4], I.A. Chvala[5]

[1] Researcher, Candidate of sciences (Biology), FGBI «ARRIAH», Vladimir, e-mail: zinyakov@arriah.ru

[2] Junior Researcher, Candidate of sciences (Biology), FGBI «ARRIAH», Vladimir

[3] Leading Biologist, FGBI «ARRIAH», Vladimir

[4] PhD student, FGBI «ARRIAH», Vladimir

[5] Head of the Laboratory, Candidate of Science (Veterinary Medicine), FGBI «ARRIA», Vladimir

SUMMARY

The article demonstrates possibility of using highly productive pyrophosphate sequencing, also known as 454 Life Sciences technology for investigation of highly variable and understudied infectious viral agents. Genome-wide nucleotide sequences of infectious bronchitis virus, infectious laryngotracheitis virus, fowl adenovirus C, and avian adenoassociated virus were determined. Recombinations in infectious bronchitis genome were identified, structural and point mutations in fowl adenovirus genome were detected. Infectious laryngotracheitis virus and avian adeno-associated virus genome structure was characterized.

Key words: infectious bronchitis, infectious laryngotracheitis, fowl adenovirus, avian adeno-associated virus, genome, nucleic acids, sequencing.

INTRODUCTION

In the present time highly productive sequencing techniques are actively used for solving a wide range of scientific and research tasks. Combination of new methods is also called sequencing of the second generation or Next Generation Sequencing (NGS). NGS application field includes genome-wide sequencing, target sequencing, trancriptome analysis and metagenomic tests [7, 8]. NGS-techniques enable sequencing without designed primers due to random DNA fragmentation and sequencing of obtained library and have considerable advantages for investigation of variable and poorly studied agents. The first technique which was the basis of highly productive sequencing was pyrophosphate sequencing. The technique was based on detection of a fluorescent signal as a result of pyrophosphate destruction (byproduct of DNA synthesis) and was a sequencing technique by synthesis. Later the technique was actively modified and optimized and as a result 454 Life Sciences technology, also known as «454 sequencing», was created. The distinctive feature of this technology differentiating it from other technological solutions in the field of highly productive sequencing was the longest read size that could be used for tests involving both already known and poorly investigated organisms as due to longer reads it is possible to design a DNA sequence assembly at a lesser data volume and get reliable results at a lower coverage. However, using NGS technologies for decoding viral genomes implies some difficulties. The greatest challenge is virus purification from cell matrix. If purification was performed improperly obtained data can be insufficient for a comprehensive analysis. In order to increase volume of information about the investigated object in a generated database additional measures for target object purification and concentration are taken, such as immunoprecipitation of viral particles, graded-index centrifuging, extraction of nucleic acid using biochips with special probes, as well as additional amplification with specific primers. Considering the subsequent complicated data processing, presence of different repeats in viral genome, and a higher variability in case of analysis of already known viruses, long reads simplify further bio-information processing.

The purpose of the paper was to study the possibility of using 454 Life Sciences technology for decoding agent genome causing viral avian diseases.

MATERIALS AND METHODS

Isolation of infectious bronchitis virus and infectious laryngotracheitis is virus was performed using chicken SPF-embryos by inoculating pathological material suspension into an allantoic cavity and a chorioallantoic membrane.

Chicken adenovirus was isolated using passaged cell culture of chicken hepatocellular carcinoma (LMH). Supporting medium (pH 7,8), containing growth medium DMEM (Sigma, USA) with fetal serum (Bioclot, Brazil), 2% of total volume, medium and L-glutamine solution (Sigma, USA) with 1% final concentration of the active ingredient was used for culturing.

Viral purification from cell matrix was performed using ultracentrifuging in sucrose gradient.

For RNA and DNA isolation from purified material RNeasy Mini Kit (Qiagen, Germany), for nucleic acid isolation was used according to the manufacturer's instruction.

Total nucleic acid of IB was converted in a double-stranded cDNA using cDNA Synthesis System (Roche, Germany) kit according to the manufacturer's instruction.

Whole-genome sequencing was performed using 454 Life Sciences perphosphate sequencing by automatic sequencer GS Junior (Roche, Germany) according to the manufacturer's protocols. For viral genome assembly GS De Novo Assembler (454 Life Science Corp.) software and GS Reference Mapper (454 Life Science Corp.) were used. For detection of recombinations Recombination Detection Program (RDP), 3.44 Version, was used.

RESULTS AND DISCUSSION

Several different viruses having specific peculiarities in genome organization were used in experiments for implementation of 454 Life Sciences technology for decoding viral genomes. The first experiment was performed using IB agent. IBV belongs to *Gammacoronavirus* genus, *Coronavirinae* subfamily, *Coronaviridae* family, *Nidovirales* order. This agent is characterized by genetic diversity and high variability. The level of nucleotide differences between isolates of different genetic groups is more than 20%. Besides, there are IBV variant isolates which differ from all reference strains and don't belong to any known genotype. Changes in IBV genome result from point mutation accumulation, insertions, deletions and recombination. Recombination results from infecting one cell with different virus strains. Herewith, "daughter" virus can have fragments of parent virus genomes of different genotypes including vaccine strains [3].

IBV27-11 isolate of IBV was chosen for full genome sequencing as due to results of nucleotide sequencing analysis a recombination was detected in S1 gene site [3]. 4542 reads were used in full genome analysis from which a 27 577 b.p. sequence was assembled. Analysis of obtained nucleotide sequence using RDP programme made it possible to detect a genome mosaic structure with five recombination events (Fig. 1).

Attempts to assemble IBV genome according to mapping algorithm against a known referent sequence using GS Reference Mapper programme (454 Life Science Corp.) were unsuccessful and the number of obtained contigs varied from 5 to 14 depending on the used reference sequence. The result demonstrates necessity to use assembly algorithms *de novo* for objects with expected intra- and intergenomic recombinations.

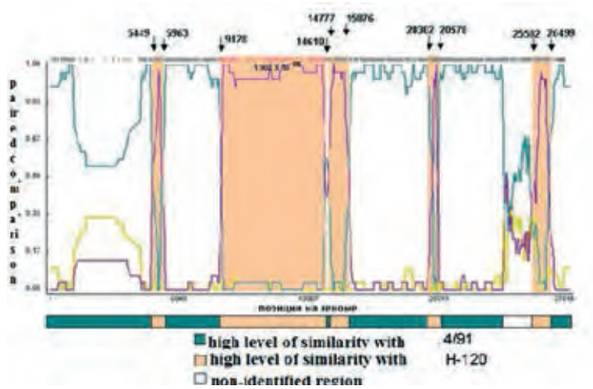

Fig. 1. IBV mosaic structure obtained using RDP programme

To study the possibility of using 454 Life Sciences technology for analysis of viruses having minor repeating motifs a field isolate of avian adenovirus was studied.

Avian adenovirus is a widely spread infectious agent belonging to *Adenoviridae* family, *Aviadenovirus* genus. «IBH»-inclusion body hepatitis and «HHS»-hepatitis-hydropericardium syndrome are the most popular avian diseases caused by adenoviruses and characterized by specific pathological signs [4]. It should be noted that despite adenoviruses are widely spread infectious agents today only some proteins of avian adenoviruses have been studied and characterized. «Krasnodar 2009» isolate was used in this investigation. The choice was conditioned by the results of previous work [1]. Analysis of fiber-1 and fiber-2 nucleotide sequences demonstrated difference between the recovered isolate from known strains including KR95 reference strain. Genetic differences are conditioned by point mutations (1,4% difference) as well as structural mutations (1,2%). 30% of point mutations are meaningful (0,36% of the general difference level). Number of amino acid substitutions detected in «Krasnodar 2009» isolate, in comparison with KR95 strain, in all open reading frames as a result of point mutations was 168. Among structural mutations slight deletions and insertions as well as large structural mutations 147 base pairs were detected. Also two insertion duplications were detected. Genetic mutations of inversion and recombination types were not detected. Structural mutations were noticed both in decoding and non-decoding parts of a genome.

Slight structural mutations (insertions and deletions), involving supposed frames were detected both among hypothetical, previously suggested reading frames as well as

Fig. 2. Region of deduced amino acid sequence of fowl adenovirus 100-kDa protein

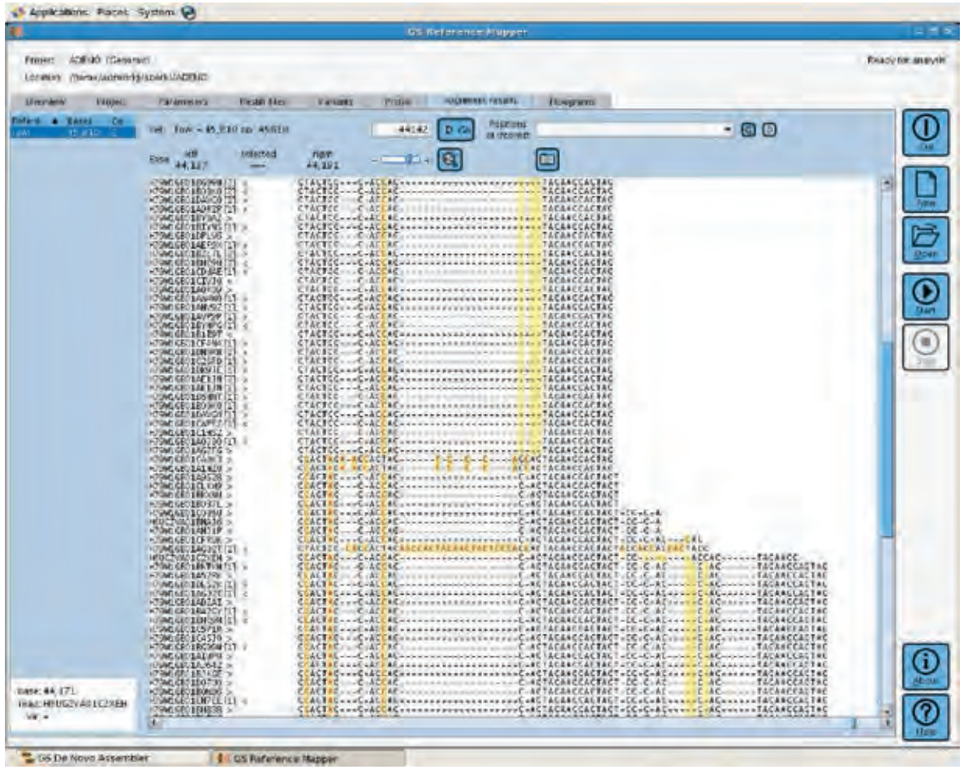

Fig. 3. Screen shot of data volume analysis results using GS Reference Mapper software

among studied proteins (genes of proteins fiber-1, fiber-2, penton, DNA-polymerase).

During analysis of obtained data for a genome region containing repeated genetic motifs discrepancies in obtained results were detected. In particular, in case of mapping to a reference sequence in primary nucleotide and deduced amino acid sequence different single substitutions were detected. Also no 60 b.p. insertion was observed. (Fig. 2).

Evidently, these discrepancies are conditioned by another algorithm of data processing used by GS Reference Mapper software. It should be noted that visualization of obtained results is a useful function of programs used for data analysis as detected heterogeneity of the investigated region is statistically enforced and could be assessed as genuine heterogeneity of the tested sample. However, specific "cuts" of analyzed readings at one site of all data volume can be observed due to visualization of results (Fig. 3). Detection of such assembly "cuts" is indicative of potential structural changes of the tested object (presence of insertions, deletions, recombinations) and necessity to perform additional analysis.

Besides, when using assembly algorithm *de novo* a fragment which was identified as adeno-associated virus on the analysis results was detected among assembled contigs [2]. As a result of comparative analysis of assembled genome specific genes of adeno--associated viruses were detected (ORF of non-structural protein «rep protein» and structural proteins «cap protein») [5]. «Rep protein» open reading frame (ORF) size of was 1992 b.p. Similarity percentage «rep protein» ORF nucleotide sequence was 92, 95, 94% with strains VR-865 (AY629582), DA-1 (AY629583), YZ-1 (GQ368252), respectively. «Cap protein» ZN-1 isolate ORF size was 2232 b.p. «Cap protein» ORF nucleotide sequence similarity percentage was 87, 95, 95% with strains VR-865, DA-1, YZ-1, respectively. Assembled genome of ZN-1 isolate was deposited to GenBank and number KF937794 was assigned to it. Therefore, implementation of assembly algorithm *de novo* is more advisable as it enables to investigate viruses with structural mutations such as recombinations, deletions, insertions. It also enables to detect virus-satellites.

Performed investigations demonstrated possibility to use 454 Life Sciences technology for analysis of viruses the genome of which contains inverted repeats. ILV causes acute respiratory disease in chicken. The agent is DNA-containing virus of *Herpesviridae* family, *Alphaherpesvirinae* subfamily. Viral particles have nuclear capsids with 80–100 nm diameter with icosahedral symmetry, whole virus particle of 195–250 nm diameter. Double-stranded linear DNA of the virus, 155 thousand b.p., contains unique long and short regions neighboring inverted repeats.

ILV vaccine strain «O» was used for testing. The size of obtained genome was 153 636 b.p. Unique low (UL) and unique short (US) segments of DNA were 113921 b.p. и 13093 b.p. respectively. Unique short segment is surrounded by inner (113922–126732 b.p.) and terminal (139826–153635 b.p.) inverted repeats in avian ILV DNA sequence (Fig. 4).

Inverted repeat regions contain genes, encoding regulatory protein ICP4, virion protein US10, as well as oriS segments, which have a palendromic structure and contain sites of viral DNA replication [6]. Analysis of ILV sequencing results demonstrated that most acceptable was algorithm of mapping obtained reads to reference sequence. Performed investigations show that it is necessary to properly select algorithms for subsequent analysis of data obtained using 454 Life Sciences technology when performing investigations related to decoding virus genomes having their specific peculiarities.

CONCLUSIONS

Obtained results demonstrated the possibility to use 454 Life Sciences technology for investigating genetically highly variable infectious agents, investigation of viruses, which are considerably different from reference strains and which have intergenomic recombonations. It was shown that for data volume analysis it is advisable to use assembly algorithm *de novo* in investigation of viruses genome of which can contain structural mutations, repeated motifs, genomic recombinations while for analysis of viruses with long inverted genomic repeats more suitable is algorithm of assembly of mapping to reference sequence.

Therefore, new high-technology equipment using NGS technologies gives additional opportunities for studying genetic properties of emerging and understudied microorganisms as well as agents of highly dangerous diseases.

BIBLIOGRAPHY

1. Genetic analysis of fowl adenovirus isolate «Krasnodar 2009» / O. S. Osipova, N. G. Zinyakov, I. A.Chvala, V. V. Drygin // Vestnik Veterinarii. — 2014. — № 3. — P. 33–37.

2. Zinyakov N. G. Detection of adeno-associated viruses using NGS-techniques // Molecular Diagnostics. — 2014. — Vol. 2. — P. 448–449.

3. Ovchinnikova Ye. V. Molecular and biological properties of IBV isolates detected in Russia in 2005–2011: thesis. … candidate of science (Biology). — Vladimir, 2012. — 116 p.

4. Adair B. M., Fitzgerald S. D. Group 1 adenovirus infections // Diseases of Poultry / ed. Y. M. Saif, A. M. Fadly, J. R. Glisson [et al.]. — 12th ed. — Ames, Iowa, — 2008. — P. 252–266.

5. Bossis I., Chiorini J. A. Cloning of an Avian Adeno-Associated Virus (AAAV) and Generation of Recombinant AAAV Particles / J. Virol. — 2003. — Vol. 77. — P. 6799–6810.

6. Knopf C. W., Spies B., Kaerner H. C. The DNA replication origins of herpes simplex virus type 1 strain Angelotti // Nucleic Acids Research. — 1986. — Vol. 14, № 21. — P. 8655–8667.

7. Metagenomic analysis of the viromes of three North American bat species: viral diversity among different bat species that share a common habitat / E. F. Donaldson, A. N. Haskew, J. E. Gates [et al.] // J. Virol. — 2010. — Vol. 84. — P. 13004–13018.

8. Sequencing viral genomes from a single isolated plaque / J. DePew, B. Zhou, J. M. McCorrison [et al.] // J. Virol. — 2013. — Vol. 10 (181). — doi: 10.1186/1743-422X-10-181.

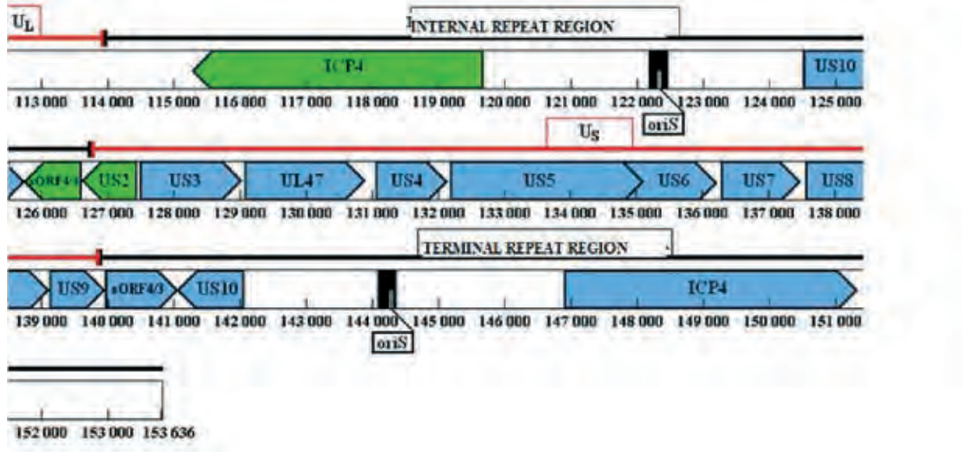

Fig. 4. Unique short (US) segment of avian strain "O" ILV genome neighboring inverted repeat regions