Short communication UDK 636.6.09:578.832
https://doi.org/10.46784/e-avm.v17i1.381

SEQUENCING PROTOCOL AND BIOINFORMATICS PIPELINE FOR THE AVIAN INFLUENZA VIRUS

Vladimir Gajdov[1*], Tamaš Petrović[1], Gospava Lazić[1], Biljana Đurđević[1], Marko Pajić[1], Slobodan Knežević[1], Jakov Nišavić[2]

[1] Scientific Veterinary Institute “Novi Sad”, Novi Sad, Republic of Serbia
[2] University of Belgrade, Faculty of Veterinary Medicine, Belgrade, Republic of Serbia

Abstract

The avian influenza virus (AIV), traditionally confined to avian hosts, has recently been detected in various mammal species, raising significant concerns for both animal and public health, necessitating efficient and accurate methods for virus detection and characterization. This study presents a sequencing protocol combined with a comprehensive bioinformatics pipeline designed for the sequencing and analysis of AIV genomes. The presented streamlined approach encompasses whole genome PCR-amplification of the viral genome, enabling the genome characterization and detection of viral mutations with high precision. An amplicon-based MiniSeq sequencing workflow based on a set of PCR primers targeting all genome segments was developed. Three samples from H5 high pathogenic avian influenza (HPAI) outbreaks in Serbia were sequenced using the MiniSeq platform. The protocol involves optimized sample preparation, tailored specifically for AIV, library preparation and sequencing. This is complemented by a robust bioinformatics pipeline that includes quality control, read mapping, consensus genome generation, subtyping and pathotyping, as well as statistical sequencing data. The pipeline efficiently processes raw sequencing data, ensuring high-quality genome assemblies and accurate identification of viral strains. The protocol was used on AIV samples from various avian species, demonstrating its applicability and reliability. The results highlight the protocol's capability to generate comprehensive genomic data, which is crucial for monitoring viral evolution and informing public health interventions. The described integrated approach offers a powerful

[1*] Corresponding Author: vladimir.g@niv.ns.ac.rs

tool for AIV surveillance and research, facilitating timely and informed decision-making in response to avian influenza outbreaks. This protocol can be readily adapted for use in various laboratory settings, contributing to global efforts in combating avian influenza and enhancing our understanding of its genomic landscape.

Key words: avian influenza, high-throughput sequencing, bioinformatics, viral genomics, molecular characterization

-

-

PROTOKOL ZA SEKVENCIRANJE I BIOINFORMATIČKU ANALIZU VIRUSA AVIJARNE INFLUENCE

Vladimir Gajdov[1*], Tamaš Petrović[1], Gospava Lazić[1], Biljana Đurđević[1], Marko Pajić[1], Slobodan Knežević[1], Jakov Nišavić[2]

[1] Naučni institut za veterinarstvo "Novi Sad", Novi Sad, Republika Srbija
[2] Univerzitet u Beogradu, Fakultet veterinarske medicine, Beograd, Republika Srbija

Kratak sadržaj

Virus avijarne influence (AIV), čiji su domaćini ptice, sve češće inficira različite vrsta sisara, što izaziva značajne zabrinutosti za zdravlje ljudi i životinja, što zahteva efikasne i precizne metode za detekciju i karakterizaciju virusa. U ovoj studiji predstavljen je protokol za sekvenciranje AIV sa kompletnom bioinformatičkom analizom, dizajniranim za sekvenciranje i analizu genoma AIV. Opisani protokol obuhvata PCR-amplifikaciju celog genoma virusa, omogućavajući karakterizaciju genoma i detekciju virusnih mutacija sa visokim stepenom preciznosti. Protokol obuhvata umnožavanje svih segmenata genoma AIV primenom jednog seta prajmera u jednoj PCR reakciji. Tri uzorka dobijena tokom epidemije H5 HPAIV u Srbiji su sekvencirana korišćenjem MiniSeq platforme. Takođe, protokol obuhvata optimizovanu pripremu uzoraka, posebno prilagođenu za AIV, pripremu biblioteke i sekvenciranje. Ovaj pristup je dopunjen robusnim bioinformatičkim analizama kojim je obuhvaćena kontrola kvaliteta, mapiranje dobijenih sekvenci, generisanje kompletnog genoma, tipizacija i patotipizaciju, kao i statističku analizu podataka sekvenciranja. Bioinformatičkom analizom se efikasno obrađuju podatke sekvenciranja, osiguravajući visokokvalitetne cele genoma i preciznu identifikaciju virusnih podtipova. Protokol je pri-

menjen i testiran na uzorcima AIV iz različitih vrsta ptica, čime je dokazana njegova primenljivost i pouzdanost. Rezultati ove studije ukazuju da razvijeni protokol generiše omogućava efikasno umnožavanje celog genoma AIV i generiše kvalitetne sekvence, što je ključno za karakterizaciju i praćenje evolucije virusa. Opisani integrisani pristup nudi moćan alat za nadzor i istraživanje AIV, olakšavajući pravovremeno i informisano donošenje odluka kao odgovor na epidemije avijarne influence. Ovaj protokol se može lako prilagoditi za upotrebu u različitim laboratorijskim okruženjima, doprinoseći globalnim naporima u borbi protiv avijarne influence i unapređenju razumevanje virusa na molekularnom nivou.

Ključne reči: avijarna influenca, sekvenciranje nove generacije, bioinformatika, genomika virsa, molekularna karakterizacija

INTRODUCTION

Avian Influenza Virus (AIV) is a highly pathogenic virus affecting both poultry and humans, with recent outbreaks causing significant economic and health impacts worldwide. The rapid evolution and genetic diversity of AIV necessitate robust and efficient methods for accurate detection, characterization, and surveillance. Advances in sequencing technologies have provided new opportunities, yet challenges remain in integrating these technologies into a streamlined workflow. Current literature highlights the need for comprehensive approaches combining sample amplification, high-throughput sequencing, and bioinformatics analysis (Imai et al., 2018; Puryear et al., 2023; Ip et al., 2023). Recent studies have demonstrated various methodologies for sequencing AIV, each contributing valuable insights into the virus's genetic makeup and evolution. However, there remains a gap in providing an all-encompassing protocol that can facilitate seamless transitions from sample collection to data analysis. This study presents a sequencing protocol with an accompanying bioinformatics pipeline specifically designed for AIV. This approach integrates the entire process, from sample amplification to sequencing and bioinformatics analysis, into a single, cohesive workflow. This protocol can be of use in various of AIV genomic studies, ultimately aiding in better understanding and management of AIV outbreaks, enabling more effective monitoring and control measures.

MATERIALS AND METHODS

Ethics statement

The animal procedures conducted in this study adhered strictly to Serbian and European regulations on animal experimentation and notifiable diseases. All samples were collected as part of the official surveillance of Avian Influenza, supervised by Serbian veterinary directorate and in accordance with Serbian veterinary legislation, as these procedures are routinely performed for official surveillance.

Sample collection and nucleic acid extraction

Carcasses of birds (both domestic and wild) with suspected AIV infection, collected during 2021-2022 were delivered to our laboratory for detection of AIV. Organs were harvested and prepared for nucleic acid extraction. Viral RNA extraction was performed with IndiMag Pathogen Kit (Indical Bioscience, Germany) as described by the manufacturer. The presence of the AIV genome in these samples was confirmed by targeting the detection of matrix gen (M gen), using RT-qPCR with TaqMan-based one-step RT-qPCR with oligonucleotide primers and probes and thermal profiles described by Spackman et al. (2002) and by using commercial kit RNA UltraSense™ One-Step Quantitative RT-PCR System (Invitrogen, ThermoFisher Scientific), according to the manufacturer's instructions. All samples were processed in a BSL2+ laboratory in strict compliance with biosafety procedures. Out of all the tested samples, three were selected for sequencing, based on the Ct values (below 28).

PCR amplification for whole genome sequencing

Since the beginning and the end of each genome segment are homologous, whole genome amplification was performed using only two primers (MBTuni-12 and MBTuni-13) described by Zhou et al. (2009). The PCR amplification was performed using SuperScript™ IV One-Step RT-PCR System (ThermoFisher Scientific, USA) and the PCR was performed in 25 µl reaction volumes with 1.5 µl forward and reverse primers (10 pmol/µl), 12.5 µl of SuperScript IV buffer mix, 0.5 µl of SuperScript enzyme mix, 5 µl of extracted RNA and DEPC water up to 25 µl. The PCR thermal profile was as follows: reverse transcription at 48°C for 45 minutes, DNA polymerase activation and reverse transcriptase inactivation at 98°C for 2 minutes, followed by 5 cycles

of denaturation at 94°C for 30 seconds, amplification at 45°C for 30 seconds, and elongation at 72°C for 3 minutes. This was further followed by 35 cycles of denaturation at 94°C for 30 seconds, amplification at 57°C for 30 seconds, and elongation at 72°C for 3 minutes, with a final elongation at 72°C for 10 minutes. The PCR amplification was performed using a PCR instrument -Thermocycler Gradient (Eppendorf, Germany).

Whole genome sequencing

To assess the PCR amplification, the concentration of dsDNA was measured using Qubit 4 (ThermoFisher Scientific, USA), following the manufacturers' instructions, followed by purification with the GeneJET PCR Purification Kit (ThermoFisher Scientific, USA). Before library preparation, samples were diluted to a working concentration of 1 ng/µl, instead of the recommended 100 ng/µl described in the library preparation protocol, because the low viral RNA concentration in the samples makes it challenging to achieve 100 ng/µl of DNA even after PCR amplification (although it should be noted that the protocol supports DNA concentration of 1 ng/µl to 500 ng/µl). Library preparation and sequencing was performed with the DNA Prep kit (Illumina, USA), following the manufacturers' instructions, with one modification (described in the previous sentence). The libraries were pooled and sequenced (2x150 paired-end) on the MiniSeq benchtop sequencer (Illumina, USA). The datasets generated in this study are available in NCBI GenBank under the accession numbers PP853088- PP853111.

Sequencing data analysis

Raw sequencing reads were initially assessed for quality using FastQC (v0.11.9) (Andrews, 2010). Statistical data on the raw reads were generated with Prinseq (v0.20.4) (Schmieder and Edwards, 2011). To ensure high-quality data, low-quality bases, primers, and adapters were trimmed using Trim Galore (v0.6.7) (Krueger, 2012). Reference-guided genome assembly was conducted using BWA (v0.6) (Li and Durbin, 2009), Samtools (v1.16.1) (Li et al., 2009), and iVar (v1.3.1) (Grubaugh et al., 2019). Coverage statistics were obtained using weeSAM (v1.6) (Aydemir and Baykal, 2018). A custom reference genome was constructed with the Vapor tool (v1.0.2) (Southgate et al., 2020) utilizing a custom-built database containing all Avian Influenza Virus (AIV) genomes from the NCBI GenBank. Subtyping was performed with the Blast command line tool with a custom-made database comprised of all available

AIV HA and NA genome segments from GISAID. The pathogenicity of the viruses was checked by identifying the presence of multi–basic hemagglutinin cleavage sites using the EMBOSS Transeq tool (Rice et al., 2000).

RESULTS and DISUSSION

Out of all the samples that tested positive for AIV presence by used RT-qPCR, three AIV positive sample materials were chosen for further analysis, PCR amplification and whole genome sequencing. A total of 2.6 Gbp of information was obtained from a sequencing run with a density of 321±3 K/mm^2, with 78.56±0.42% of the clusters passing quality control (QC) filters. Approximately 87.16% of Read1 and 92.43% of Read2 sequences contained high-quality bases, meeting the ≥ Q30 base calling QC standard, as generated by FastQC and Prinseq. BLAST analysis of the AIV hemagglutinin and neuraminidase sequences identified two viruses as belonging to the H5N1 subtype and one to the H5N8 subtype. All AIVs were categorized within clade 2.3.4.4b. The high pathogenicity of the viruses was confirmed by the presence of multi-basic hemagglutinin cleavage sites. Consensus genome was generated for all samples. The complete genomes of all three virus samples yielded equal or greater than 3000-fold sequence coverage depth. The results of this study demonstrate the efficacy of this comprehensive sequencing and bioinformatics pipeline in accurately identifying and characterizing AIV from field samples. The high-throughput sequencing yielded 2.6 Gbp of data, with a significant proportion of high-quality reads, underscoring the robustness of the initial quality control and data preprocessing steps. The initial quality assessment using FastQC and the subsequent data trimming with Trim Galore ensured that dataset maintained high standards, with over 87% of Read1 and 92% of Read2 sequences meeting the ≥ Q30 quality threshold. This high level of quality is essential for reliable downstream analyses. Our PCR protocol successfully amplified the whole genome of the viruses, enabling the simultaneous detection of two different subtypes. The reference-guided genome assembly using BWA, Samtools, and iVar was highly effective, as evidenced by the coverage depth exceeding 3000-fold for all samples. High depth is critical for ensuring the accuracy of variant calling and genome assembly, providing a solid foundation for subsequent analyses. Constructing a custom reference genome with the Vapor tool, utilizing a comprehensive database of AIV genomes from NCBI GenBank proved advantageous. This approach enhances the accuracy and relevance of genomic analyses by tailoring the reference to include the most recent and locally relevant strains, reducing bias, and improving assembly quality. It al-

lows for more precise detection of region-specific mutations and pathogenicity markers, providing detailed evolutionary insights and adaptability to rapidly evolving pathogens like AIV. BLAST analysis and phylogenetic classification identified three H5N1 and one H5N8 subtypes, all belonging to clade 2.3.4.4b. The ability to accurately subtype the viruses and place them within a specific clade demonstrates the effectiveness of this pipeline and the utility of using a custom reference genome. Most of the papers describing AIV sequencing use in-house developed protocols and pipelines (Puryear et al., 2023) or focus on the utilization of Oxford Nanopore technology (Imai et al., 2018) and that is why this paper presents an integrated protocol and bioinformatics pipeline designed to streamline the entire process of AIV. It combines sample amplification, sequencing, and data analysis into a cohesive workflow. By providing a comprehensive, all-in-one solution, the paper aims to facilitate scientists' efforts in efficiently obtaining and analyzing high-quality genomic data, enhancing the accuracy and speed of AIV research and surveillance.

CONCLUSION

In conclusion, the conducted study presents a simple yet effective sequencing and bioinformatics pipeline for the analysis of AIVs. The high quality of the sequencing data, combined with the use of a custom reference genome, allowed for accurate identification, assembly, and pathogenicity assessment of the viruses. These results contribute valuable data to the global understanding of AIV diversity and evolution and highlight the importance of continued surveillance and methodological advancements in the field. Future studies could build on this approach by incorporating additional bioinformatics tools and expanding the scope of bioinformatic analysis.

ACKNOWLEDGMENTS

This work was funded by Ministry of Science, Technological Development and Innovation of Republic of Serbia by the Contract of implementation and funding of research work of NIV-NS in 2024, Contract No: 451-03-66/2024-03/200031, and by the project "Exploring the Avian Microbial Frontier: Metagenomic Surveillance of Infectious Diseases and Antimicrobial Resistance in Wild Birds" – "MetAvian" funded by the Science Fund of the Republic of Serbia, #GRANT No 10945.

Author's Contribution

V.G. carried out literature research, designed the study, prepared samples for sequencing, sequenced the samples, analyzed the results, and drafted the manuscript. T.P. carried out the RT-qPCR testing of samples and, together with V.G., drafted the manuscript. G.L. was involved in sample collection and data analysis. B.Đ. collected the samples and prepared them for testing. D.V. and J.N. were involved in data analysis and the final polish of the manuscript. All authors read and approved the final manuscript.

Competing interest

The authors declare that they have no competing interests.

REFERENCES

1. Andrews S. 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data. [Online] Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 10 March 2024.
2. Aydemir O., Baykal A.T. 2018. weeSAM: A Web-Based Application for Classification of Single-Cell RNA-Seq Data. Bioinformatics, 34, 6, 1065-1067. doi:10.1093/bioinformatics/bty075.
3. Grubaugh N.D., Gangavarapu K., Quick J., Matteson N.L., De Jesus J.G., Main B.J., Tan A.L., Paul L.M., Brackney D.E., Grewal S., Gurfield N., Van Rompay K.K.A., Isern S., Michael S.F., Coffey L.L., Loman N.J., Andersen, K.G. 2019. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. Genome Biology, 20, 8. doi:10.1186/s13059-018-1618-7.
4. Imai K., Tamura K., Tanigaki T., Takizawa M., Nakayama E., Taniguchi T., Okamoto M., Nishiyama Y., Tarumoto N., Mitsutake K., Murakami T., Maesaki S., Maeda T. 2018. Whole genome sequencing of influenza A and B viruses with the MinION sequencer in the clinical setting: a pilot study. Frontiers in Microbiology, 9, 2748. doi: 10.3389/fmicb.2018.02748.
5. Ip H.S., Uhm S., Killian M.L., Torchetti M.K. 2023. An evaluation of Avian influenza virus Whole-Genome sequencing approaches using Nanopore technology.Microorganisms, 11, 2, 529. doi:10.3390/microorganisms11020529.
6. Krueger F. 2012. Trim Galore: A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files.

[Online] Available at: http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ Accessed 10 March 2024.

7. Li H., Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 25 14, 1754-1760. doi:10.1093/bioinformatics/btp324.
8. Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics, 25, 16, 2078-2079. doi:10.1093/bioinformatics/btp352.
9. Puryear W., Sawatzki K., Hill N., Foss A., Stone J. J., Doughty L., Runstadler J. 2023. Highly pathogenic Avian influenza A(H5N1) virus outbreak in New England Seals, United States. Emerging Infectious Diseases, 29, 4, 786–791. doi: 10.3201/eid2904.221538.
10. Rice P., Longden I., Bleasby A. 2000. EMBOSS: The European Molecular Biology Open Software Suite. Trends in Genetics, 16, 6, 276-277. doi:10.1016/s0168-9525(00)02024-2.
11. Schmieder R., Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics, 27, 6, 863-864. doi:10.1093/bioinformatics/btr026.
12. Southgate J., Bull M., Brown C., Watkins J., Corden S., Southgate B., Moore C., Connor T. 2020. Influenza classification from short reads with VAPOR facilitates robust mapping pipelines and zoonotic strain detection for routine surveillance applications. Bioinformatics, 36, 6, 1681–1688. doi: 10.1093/bioinformatics/btz814.
13. Spackman E., Senne D. A., Myers T.J., Bulaga L.L., Garber L.P., Perdue M.L., Suarez D.L. 2002. Development of a Real-Time Reverse Transcriptase PCR assay for Type A influenza virus and the Avian H5 and H7 hemagglutinin subtypes. Journal of Clinical Microbiology, 40, 9, 3256–3260. doi: 10.1128/jcm.40.9.3256-3260.2002.
14. Zhou B., Donnelly M.E., Scholes D.T., St George K., Hatta M., Kawaoka Y., Wentworth D.E. 2009. Single-Reaction genomic amplification accelerates sequencing and vaccine production for classical and swine origin human influenza A viruses. Journal of Virology, 83, 19, 10309–10313. doi: 10.1128/jvi.01109-09.

Received: 23.05.2024.
Accepted: 06.06.2024.