International Journal of Advanced Smart Convergence Vol.9 No.1 90-97 (2020)
http://dx.doi.org/10.7236/IJASC.2020.9.1.90

IJASC 20-1-11

Construction of a Genetic Information Database for Analysis of Oncolytic Viruses

Myeongji Cho[1], Hyeon Seok Son[2], Hayeon Kim[3,*]

[1]Postdoctoral Researcher, Institute of Health and Environment, Seoul National University, Korea
[2]Professor, Laboratory of Computational Biology & Bioinformatics, Institute of Public Health and Environment, Graduate School of Public Health, Seoul National University, Korea
[3]Assistant Professor, Department of Biomedical Laboratory Science, Kyungdong University, Korea
e-mail: {[1]yummy0112, [2]hss2003}@snu.ac.kr, [3,]hykim1984@kduniv.ac.kr*

Abstract

Oncolytic viruses are characterized by their ability to selectively kill cancer cells, and thus they have potential for application as novel anticancer agents. Despite an increase in the number of studies on methodologies involving oncolytic viruses, bioinformatic studies generating useful data are lacking. We constructed a database for oncolytic virus research (the oncolytic virus database, OVDB) by integrating scattered genetic information on oncolytic viruses and proposed a systematic means of using the biological data in the database. Our database provides data on 14 oncolytic viral strains and other types of viruses for comparative analysis. We constructed the OVDB using the basic local alignment search tool, and therefore can provides genetic information on highly homologous oncolytic viruses. This study contributes to facilitate systematic bioinformatics research, providing valuable data for development of oncolytic virus-based anticancer therapies.

Keywords: *Oncolytic Virus, Database, Bioinformatics, BLAST*

1. Introduction

Cancer is a major cause of death worldwide, despite technological advancements. Many studies on effective means of treating and preventing cancer have been conducted, some of which focused on oncolytic viruses. These viruses are characterized by the ability to selectively infect cancer tissues, replicate and proliferate inside cancer cells, and ultimately induce cancer-cell apoptosis without harming normal tissues; hence, these viruses have potential as novel anticancer agents [1]. Furthermore, oncolytic viruses kill cancer cells via several mechanisms and exhibit genetic polymorphisms [2]. Should oncolytic viruses prove useful in treating cancer, they would significantly contribute to human health and welfare. To function as anticancer agents, oncolytic viruses require appropriate genetic manipulation for optimum selectivity and should exert a pro-apoptotic effect on appropriate target cells [3]. However, the pro-apoptotic effect of oncolytic viruses on cancer in actual

Manuscript Received: January. 10, 2020 / Revised: January. 13, 2020 / Accepted: January. 19, 2020
Corresponding Author: hykim1984@kduniv.ac.kr
Tel: +82-33-738-1374, Fax: +82-33-738-0243
Assistant Professor, Department of Biomedical Laboratory Science, Kyungdong University, Korea

patients, as well as in experimental models, needs to be investigated. Safety assessments of oncolytic viruses based on strict criteria derived from experimental models are also required [4]. Nevertheless, establishment of scientific standards for oncolytic virus safety a priori would considerably reduce costs and time frames and facilitate determination of the oncolytic abilities of as-yet-unidentified viral strains. The questions of what differentiates oncolytic and non-oncolytic viruses, how cancer specificity is acquired, and why an oncolytic virus is harmless to normal cells need to be answered [5]. Thereafter, the safety of oncolytic viruses should be determined by identifying the target tissue and mechanism of virulence, as well as assessing the risk associated with genetic manipulation and viral infectivity [6, 7].

With advancements in molecular biology, oncolytic virus processing and application have been investigated, whereas few bioinformatics studies have analyzed the data obtained. Thus, we developed a systematic method of applying biological data from a specialized database containing genetic information on oncolytic viruses based on our previous study using bioinformatics techniques [8, 9]. Processing and analysis of the data on oncolytic viruses in the database will facilitate discovery of an effective and safe oncolytic virus and in this way contribute to anticancer research.

2. Methods

We collected data primarily from National Center for Biotechnology Information (NCBI) GenBank (ftp://ftp.ncbi.nih.gov/genbank/) to construct a database of genetic information on oncolytic viruses [10]. Data were parsed using JAVA programming to extract only necessary information. Briefly, all genetic information and viral sequences in GenBank were extracted, and only those data with ACCESSION were reconstructed as the primary key in MySQL, although the design allowed selective searches for partial and complete sequences. Collection years were extracted from the LOCUS field, definitions regarding the genetic information of each virus were extracted from DEFINITION, and unique lot numbers for sequences were extracted from ACCESSION. Information on the protein products of genes and the deduced amino acid sequences were extracted from FEATURES, and gene sequences were extracted from ORIGIN. The FASTA format files required for basic local alignment search tool (BLAST) analysis were produced via data processing. Information on tumor genetic mutations and tumor-selective killing mechanisms, targets of additional literature searches, and viral and host genes reportedly implicated in anticancer mechanisms were checked and stored. We constructed the database using a server based on the high-performance computing (HPC) cluster system with the following specifications: 8-core AMD Opteron-6128 2.0 GHz central-processing unit, 8 GB random-access memory and a 500 Gb serial ATA 7200 rpm 3 Gbps hard disk drive. The operating system was LINUX v. 2.6.18, and the data management system used for storage in the Linux server environment was MySQL. The programming language for data parsing was JAVA, and those for web linking were JSP, HTML, and JAVAscript. The web server program was based on Apache, and Tomcat v. 7.0 served as the web container (Table 1). We installed the Web BLAST package from NCBI, wwwblast (ncbi-blast-2.2.26, for Linux), to construct a BLAST server [11]. The database to link to BLAST was based on sequence data produced in FASTA format by reprocessing the gene sequences from oncolytic viruses (Table 2).

Table 1. System development environment

Category	System development environment
System	HPC cluster system
Central processing unit (CPU)	8C AMD Opteron-6128 2.0GHz × 1 (8Core)
Memory	8Gb
Hard disk drive (HDD)	500Gb SATA 7200rpm 3Gbps × 1
Operating system	Linux
Web server	Apache
Database management system (DBMS)	MySQL
Programming language	JAVA, JSP, HTML, JAVA script

Table 2. Type and amount of data constructed on BLAST

BLAST database	Number of data
Virus	1,405,570
Oncolytic Virus	303,124

3. Results

3.1 Construction of an oncolytic virus database

We constructed an oncolytic virus database (OVDB) to investigate the specificity and safety of oncolytic viruses. Its design applies a system that allows web-based searches and analyses (Figures 1 and 2). The database contains information on definitions, accession numbers, collection years, products, and the nucleotide and amino acid sequences of viruses and oncolytic viruses. The database contains approximately 1,405,570 viruses, including 303,124 oncolytic viruses, belonging to 14 strains: adenovirus, coxsackievirus, herpes simplex, influenza, measles, myxoma, Newcastle disease, polio, parvovirus, reovirus, retrovirus, Seneca valley, vesicular stomatitis, and vaccinia (Table 3). To compare genomes or gene/protein sequences using the OVDB, genes of interest are selected, and their sequence data are stored in a downloaded FASTA file. The stored sequence data or the data of the user are subjected to gene sequence alignments using a multiple sequence alignment program linked to OVDB, and the results are output as '.aln' files. In addition, phylogenetic trees can be generated using a tool in the ClustalW program [12] linked to the OVDB.

3.2 Search function of the OVDB

A search set has been created in the OVDB to enable efficient searches of virus data using two methods. Overall virus data can be searched by entering a virus name, accession number, or keyword (Figure 3(a)). To search oncolytic virus data, the system allows a choice of 14 species and of complete versus partial sequences (Figure 3(b)). Searches can also be based on accession numbers or keywords. The keyword search facilitates exploration of viral strains or subtypes and antigens or conjugated proteins. To obtain gene or protein sequences from the tables produced by searches, data can be accessed or stored as FASTA files by clicking 'Download' for the Origin or Translation column.

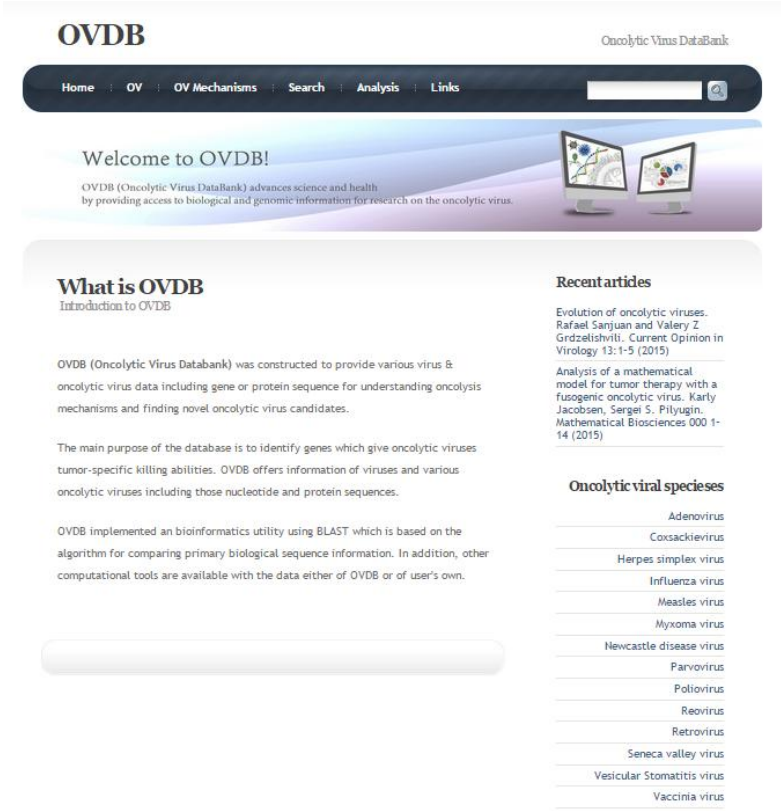


Figure 1. OVDB (oncolytic virus databank) main page

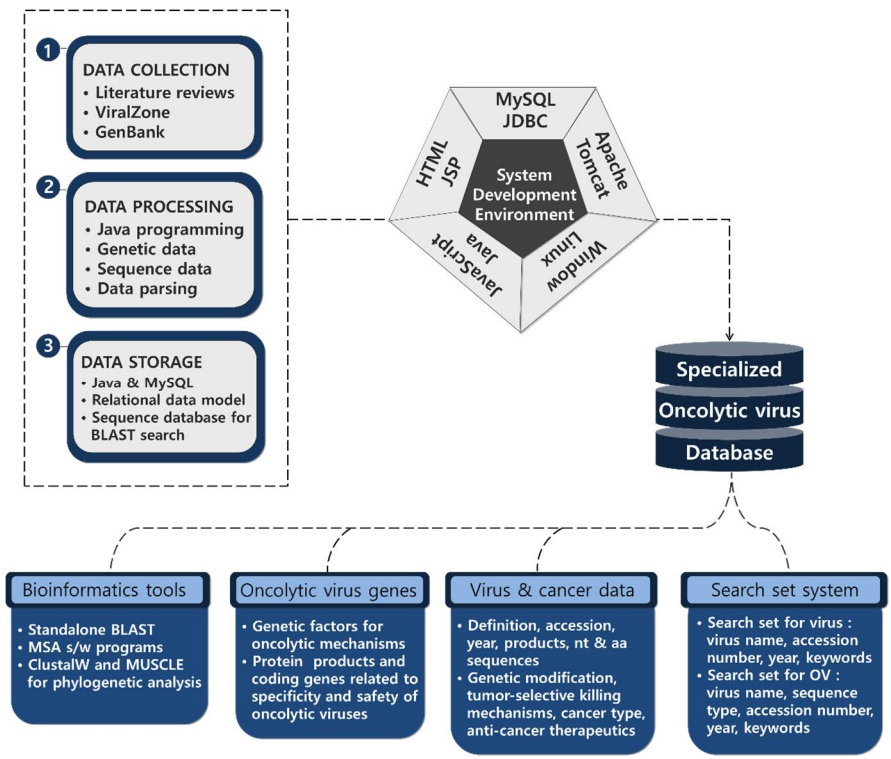
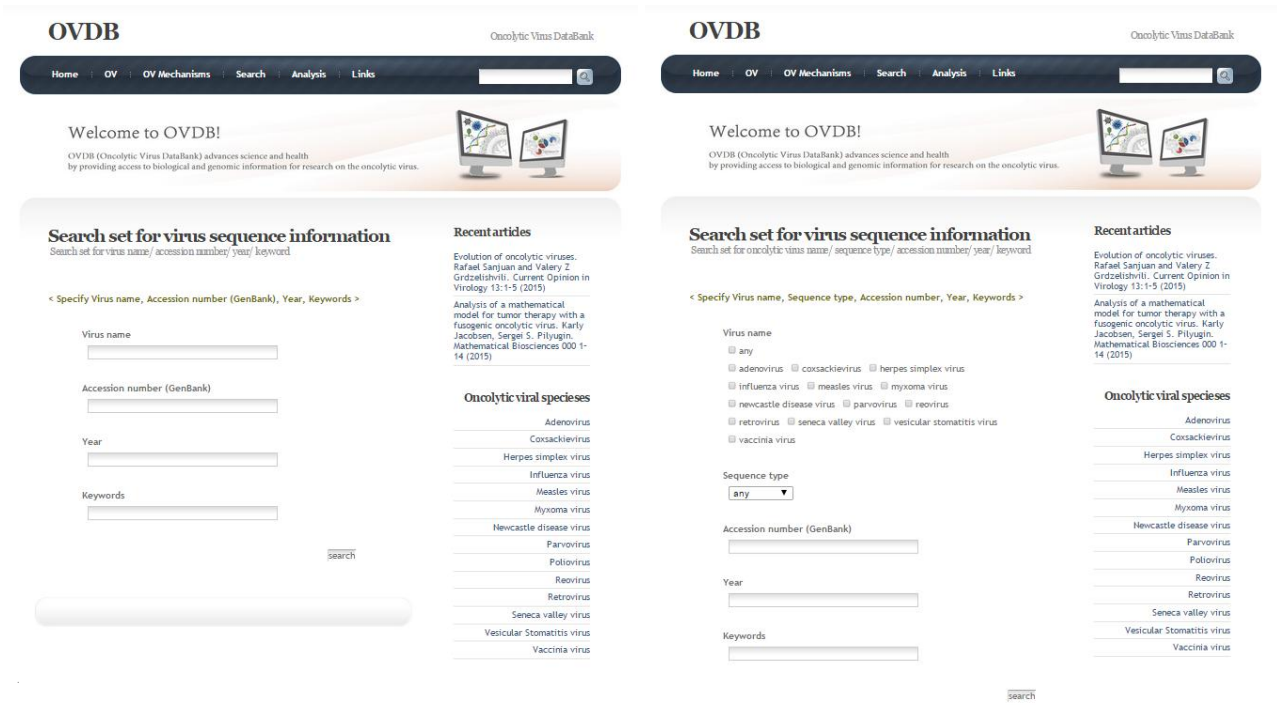


Figure 2. Data flow diagram

Table 3. The number of data in 'OVDB'

species	number	species	number
Adenovirus	5,970	Polio virus	4,391
Coxsackievirus	8,467	Parvo virus	3,538
Herpes simplex virus	201	Reovirus	1,875
Influenza virus	261,865	Retrovirus	1,771
Measles virus	7,356	Seneca valley virus	18
Myxoma virus	343	Vesicular stomatitis virus	834
Newcastle disease virus	5,938	Vaccinia virus	559
Oncolytic virus (total) 303,126			



(a) Search page for virus data **(b) Search page for oncolytic virus data**

Figure 3. Virus and oncolytic virus search set

3.3 Analysis function of the OVDB

A local BLAST system was built into the OVDB to enable analysis of the relationships and similarities among viruses based on an input query sequence. The user can use the BLAST tool to select the gene sequence database of a virus or oncolytic virus and then enter a query sequence (Figure 4). The system calculates bit-scores and e-values to produce ACCESSION numbers and DEFINITIONS for the top 100 viruses with the

highest homology. In addition to BLAST, links to the ClustalW [12] and MUSCLE [13] programs allow multiple sequence alignments for similarity analyses, and the phylogenetic analysis function provided by ClustalW2 [12] and MEGA [14] allows the generation of phylogenetic trees.

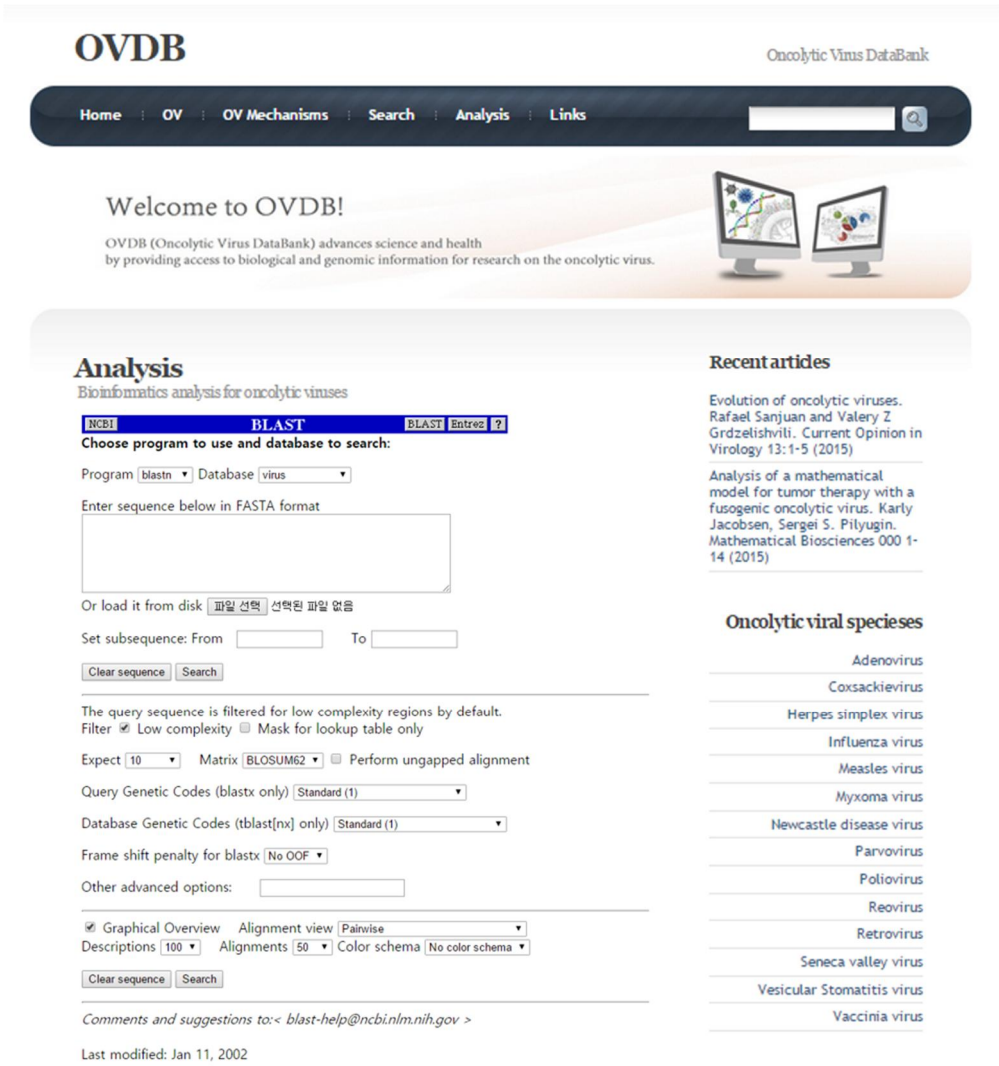


Figure 4. Local BLAST web page

Local BLAST is performed for similarity analysis. Users can select the database to use for identification of an unknown query sequence.

4. Discussion

We constructed a database by collecting and processing genetic information on oncolytic viruses to enable investigation of their selective tumor-killing mechanisms, specificity, and safety. Links to bioinformatics tools facilitate analysis of the data in the OVDB. In addition to data on 14 oncolytic viral strains, the OVDB contains information on other types of viruses for comparative analysis. Genetic information on oncolytic viruses with high homology can also be obtained because the database was constructed using the BLAST tool. Homology searches using BLAST were designed considering not only the sequences of genes encoding receptor-binding proteins but also the genes responsible for the specificity and safety of oncolytic viruses. The genetic

composition of oncolytic viruses critically influences their oncolytic activity. Thus, a method to compare gene sequences would facilitate discovery of novel candidate oncolytic viruses and analysis of sequence similarities among oncolytic viral strains. Our database, the OVDB, will assist the development of novel anticancer agents by supporting systematic bioinformatics research.

5. Conclusion

Oncolytic viruses are being actively investigated both within and outside South Korea. The potential practical applications of oncolytic viruses have been assessed in various clinical studies. Multiple aspects of the anticancer mechanisms of oncolytic viruses need to be investigated. We exclusively collected and stored data in the OVDB, for oncolytic viruses including sequence data. Bioinformatics tools will be applied to explore the genetic information in the OVDB pertaining to the specificity and safety of oncolytic viruses. Because the OVDB enables genetic investigation of the cancer cell-specific killing mechanisms of oncolytic viruses, it represents a considerable advance in the war on cancer. Our database will be continuously updated with new data and complemented with various bioinformatics-based analysis techniques. This will enable development of novel anticancer agents and applied research into oncolytic viruses.

Acknowledgement

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government (MOE and MSIT) (No. 2017R1D1A1B03033413 and No. 2019R1H1A1085901).

References

[1] E. Kelly, and S.J. Russell. “History of Oncolytic Viruses: Genesis to Genetic Engineering”, *Molecular Therapy*, Vol. 15, No. 4, pp. 651-659, 2007. DOI: http://doi.org/10.1038/sj.mt.6300108.

[2] L.K. Hawkins, N.R. Lemoine, and D. Kirn. “Oncolytic Biotherapy: A Novel Therapeutic Platform”, *The Lancet Oncology*, Vol. 3, No. 1, pp. 17–26, 2002. DOI: https://doi.org/10.1016/S1470-2045(01)00618-0.

[3] B. Everts, and H.G. van der Poel. “Replication-selective oncolytic viruses in the treatment of cancer”, *Cancer Gene Therapy*, Vol. 12, No. 2, pp. 141-161, 2005. DOI: https://doi.org/10.1038/sj.cgt.7700771.

[4] J.J. Davis, and B. Fang. “Oncolytic Virotherapy for Cancer Treatment: Challenges and Solutions”, *The Journal of Gene Medicine*, Vol. 7, No. 11, pp. 1380–1389, 2005. DOI: https://doi.org/10.1002/jgm.800.

[5] S.J. Russell, K.W. Peng, and J.C. Bell. “Oncolytic virotherapy”, *Nature Biotechnology*,Vol. 30, No. 7, pp. 658-670, 2012. DOI: https://doi.org/10.1038/nbt.2287.

[6] F. McCormick, “Future prospects for oncolytic therapy”, *Oncogene*, Vol. 24, No. 52, pp. 7817-7819, 2005. DOI: https://doi.org/10.1038/sj.onc.1209064.

[7] J. Bell, and G. McFadden. “Viruses for tumor therapy”, *Cell Host Microbe*, Vol. 15, No. 3, pp. 260-265, 2014. DOI: https://doi.org/10.1016/j.chom.2014.01.002.

[8] M.R. Kim, J.H. Lee, H.S. Son, and H. Kim, “HCoV-IMDB: Database for the Analysis of Interactions between HCoV and Host Immune Proteins”, *International Journal of Advanced Smart Convergence*, Vol. 8, No. 1, pp. 1-8, 2019. DOI: http://dx.doi.org/10.7236/IJASC.2019.8.1.1.

[9] M. Je, H.S. Son, and H. Kim, “Data-processing pipeline and database design for integrated analysis of mycoviruses”, *International Journal of Advanced Smart Convergence*, Vol. 8, No. 3, pp. 115-122, 2019. DOI: http://dx.doi.org/10.7236/IJASC.2019.8.3.115.

[10] D.A. Benson, M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, D.J. Lipman, J. Ostell, and E.W. Sayers. “GenBank”, *Nucleic Acids Research*, Vol. 41, No. D1, pp. D36-D42, 2013. DOI: http://doi.org/10.1093/nar/gks1195.

[11] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, and D.J. Lipman, "Basic local alignment search tool", *Journal of Molecular Biology*, Vol. 215, No. 3, pp. 403–410, 1990. DOI: https://doi.org/10.1016/S0022-2836(05)80360-2.

[12] M.A. Larkin, G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, and D.G. Higgins, "Clustal W and Clustal X version 2.0", *Bioinformatics*, Vol. 23, No. 21, pp. 2947-2948, 2007. DOI: http://doi.org/10.1093/bioinformatics/btm404.

[13] R.C. Edgar, "MUSCLE: multiple sequence alignment with high accuracy and high throughput", *Nucleic Acids Research*, Vol. 32, No. 5, pp. 1792-1797, 2004. DOI: https://doi.org/10.1093/nar/gkh340.

[14] S. Kumar, G. Stecher, and K. Tamura, "MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets", *Molecular Biology and Evolution*, Vol. 33, No. 7, pp. 1870–1874, 2016. DOI: https://doi.org/10.1093/molbev/msw054.