

BOOK OF ABSTRACTS



Organização



Apoio



TÍTULO | TITLE

Livro de Resumos do 18º Encontro de Investigação Jovem da U.Porto | *Book of Abstracts
Young Researchers Meeting of U.Porto*

UNIVERSIDADE DO PORTO

Professor Doutor Pedro Rodrigues

jjup@reit.up.pt

ISBN

978-989-746-418-8

DESIGN

Serviço de Comunicação e Imagem da U.Porto

23230 | ViruScopeDB: a comprehensive multi-omics database for highly infectious viruses

Ana Lima^{1,2}; João Carneiro^{3,4}; Sérgio Sousa⁵; Vítor Sá^{1,9}; Diogo Pratas^{6,7,8}

LabRP/CIR, School of Health, - Polytechnic of Porto, Porto, Portugal¹; CIIMAR – Interdisciplinary Centre of Marine and Environmental Research of the University of Porto, University of Porto, Matosinhos, Portugal²; Centre of Molecular and Environmental Biology (CBMA), Department of Biology, University of Minho, Braga, Portugal³; Institute of Science and Innovation for Bio-Sustainability (IB-S), University of Minho, 4710-057 Braga, Portugal⁴; LAQV/REQUIMTE - BioSIM, Department of Biomedicine, Faculty of Medicine, University of Porto, Porto, Portugal⁵; IEETA/LASI, Institute of Electronics and Informatics Engineering of Aveiro, University of Aveiro, Aveiro, Portugal⁶; DETI, Department of Electronics, Telecommunications and Informatics, University of Aveiro, Aveiro, Portugal⁷; DV, Department of Virology, University of Helsinki, Helsinki, Finland⁸; ALGORITMI/LASI, University of Minho, Braga, Portugal⁹

Background & Aim: Highly infectious viruses such as HIV, Ebola, and SARS-CoV-2 have presented ongoing challenges to global health. Consequently, the optimization of rapid detection tests, including PCR, and the identification of new therapeutic targets remain of paramount importance. The development of genomic and proteomic databases like the HIV Oligonucleotide Database (HIVoligoDB) [1], EbolaID [2], and CoV2ID [3] has facilitated the accumulation and accessibility of knowledge through comprehensive, user-friendly, open-access platforms. This study aims to update, expand, and integrate these databases into a single resource, ViruScopeDB, while conducting thorough analyses of informative genomic regions with the goal of enhancing viral detection methods and treatment strategies. **Methods:** Complete genomic sequence variants for each virus were compiled using NCBI Virus, followed by multiple sequence alignment via MAFFT within the Galaxy platform. The alignments were consequently uploaded to Geneious Prime for complete genome visualization and calculation of parameters such as percentage of pairwise identity. Primer data was extracted from open-access research articles available on PubMed using a newly-built custom pipeline for PDF to plain text conversion followed by data mining of oligonucleotide sequences. A fully automated script for primer validation, parameter scoring and calculation of best primer pairs for PCR is being constructed for subsequent upload into the database. **Results:** A total of 658 sequences with a mean length of 18,910 base pairs (bp) were collected for Ebolavirus, with percentage of pairwise identity (PPI) of 91.7%. 7,261 sequences with a mean length of 8,883 bp, with a PPI of 80.8% were identified for HIV-1. For HIV-2, 43 sequences with an average of 10,108 bp and PPI of 80.9% were analyzed.

For Ebola, a total of 709 primers were scraped from 257 articles, and for HIV articles this number rises to 10,290 primers collected from 2,579 articles. **Conclusions:** Using a combination of pre-existent and novel custom-built bioinformatics tools, it was possible to data mine key information related to each virus and their variants, as well as collect primer information for possible PCR optimizations. Further analysis will be conducted on the data collected, branching out into the realm of phylogenetics and 3D modelling/viral protein docking, in order to construct a database that is transversal to various omics.

Keywords: Bioinformatics, Virology, Database, Ddata mining.

Acknowledgments:

This research was supported by Portuguese national funds through the Foundation for Science and Technology (FCT) within the scope of UIDB/04423/2020 (CIIMAR) and UIDP/04423/2020 (CIIMAR), 10.54499/LA/P/0008/2020, 10.54499/UIDP/50006/2020, UIDB/04050/2020 (UMinho CBMA - <https://doi.org/10.54499/UIDB/04050/2020>), 10.54499/UIDB/50006/2020 (LAVQ), UIDB/00319/2020 (ALGORITMI/LASI) and UIDB/00127/2020 (IEETA/LASI - doi.org/10.54499/UIDB/00127/2020). JC acknowledges the FCT funding for his research contract established under the transitional rule of Decree Law 57/2016, amended by Law 57/2017. D.P. is funded by national funds through FCT—Fundação para a Ciência e a Tecnologia, I.P., under the Scientific Employment Stimulus—Institutional Call—reference CEECINST/00026/2018.

References:

- [1] Carneiro J, Resende A, Pereira F. The HIV oligonucleotide database (HIVoligoDB). Database. 2017 Jan 1;2017:bax005.
- [2] Carneiro J, Pereira F. EbolaID: An Online Database of Informative Genomic Regions for Ebola Identification and Treatment. PLoS Negl Trop Dis. 2016 Jul 21;10(7):e0004757.
- [3] Carneiro J, Gomes C, Couto C, Pereira F. CoV2ID: Detection and Therapeutics Oligo Database for SARS-CoV-2 [Internet]. bioRxiv; 2020 [cited 2024 Sep 30]. p. 2020.04.19.048991. Available from: <https://www.biorxiv.org/content/10.1101/2020.04.19.048991v2>.