International Journal of Science and Research Archive
eISSN: 2582-8185
Cross Ref DOI: 10.30574/ijsra
Journal homepage: https://ijsra.net/

(RESEARCH ARTICLE)

Check for updates

# Identification of genomes: Clustal Omega and BLAST: One introduction

Carlos Navarro Venegas *

*Faculty of Veterinary and Livestock Sciences (FAVET). University of Chile, Chile.*

International Journal of Science and Research Archive, 2022, 06(02), 026–029

Publication history: Received on 12 June 2022; revised on 18 July 2022; accepted on 20 July 2022

Article DOI: https://doi.org/10.30574/ijsra.2022.6.2.0154

## Abstract

Knowing the sequence of a pathogen isolated in our laboratory is currently an accessible task, and there are no excuses to start it.

The remarkable occurrence of Kary Mullis and the existence of computer programs has helped a lot in this task and in this introductory article we give an account of that and invite those who still do not dare to this adventure.

**Keywords:** Genomes; Nucleotide sequence; Clustal Omega; BLAST

## 1. Introduction

The genome is generally DNA except for one type of virus: the Ribovirus whose genome is RNA (existing single-stranded and double-stranded).

Knowing the foregoing and the existence of the Genbank® database (https://www.ncbi.nlm.nih.gov/genbank/ ), today there are no excuses for the detection of any pathogen or for identifying which genotype (for example) would belong [1, 2, 3, 4].

The identification of the genome begins with its detection, using the incredible occurrence of Kary Mullis (5), obtaining double-stranded DNA fragments by means of the appropriate PCR (RT-PCR in the case of most RNA viruses).

## 2. Material and Methods

Once these fragments have been obtained, it is necessary to know their nucleotide sequence and due to the inevitable errors inherent in the laboratory, it is preferred to send each sample to be sequenced in triplicate (minimum).

## 3. Results

Thus, for each sample three (at least) sequences are received. So, it is necessary to know the "average" sequence known as the consensus sequence and for that the CLUSTAL OMEGA software (https://www.ebi.ac.uk/Tools/msa/clustalo/ ) can be used. When entering the sequences into the program, an alignment is obtained, which will indicate by means of a "*" where there is 100% nucleotide identity for each nucleotide position, for example (Table 1).

* Corresponding author: Carlos Navarro Venegas
Faculty of Veterinary and Livestock Sciences (FAVET). University of Chile, Chile.

Copyright © 2022 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution Liscense 4.0.

**Table 1** Results of alignment from Clustal Omega use

```
S3 TTAAATATGCCTGAAAGTGAAAATTATTAAAGAGTAGTTGTAAATAATATGGATTAAACT    60
S1 TTAATTATGCCAGAAAGTGAAAATTTTAAAAGAGTTGTTGTAAATAATATGGATAAAACT    60
S2 TTAAATATGCCAGAAAGTGAAAATTATAAAAGAGTAGTTGTAAATAATATGGATAAAACT    60
   **** ****** ************* * ******* ***************** *****
```

Here, S1, S2 and S3 correspond to the sequences of the same sample and the number of asterisks show that there is "*a great resemblance*" between the sequences (obviously, asterisks should appear in all positions: it is the same sample). We must build the consensus sequence by assigning a nucleotide in the place where the "*" does not appear, considering the one that is repeated the most in that position, in this case it corresponds to:

Consensus sequence: TTAAATATGCCATGAAAGTGAAAATTATAAAAGAGTAGTTGTAAAATAATATGGATAAAACT

This program (CLUSTAL OMEGA) can align several sequences at once and each sequence can be of variable length.

The constructed consensus sequence represents the work done in the laboratory, but.... What does it correspond to? Does it correspond to the genome we want to detect? ... Let's see and let nothing be left up *in the air* (*minute 5.5*)!!!

The next stage contemplates the use of another free online software up to now: BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome ) a very powerful tool that indicates the percentage of nucleotide identity (PIN) of the entered sequence with respect to ALL the available database. That is, it tells us with which genome it has the greatest "coincidence". In this output, the program shows the first 100 alignments and each PIN.

For example, (Figure 1) inputting the newly constructed consensus sequence to BLAST yields the following output:

Sequences producing significant alignments — Manage Columns — Show 100

select all — 100 sequences selected — GenBank — Graphics — Distance tree of results

| Description | Max Score | Total Score | Query Cover | E value | Per. Ident | Accession |
|---|---|---|---|---|---|---|
| Feline panleukopenia virus isolate Jun, partial genome | 106 | 106 | 100% | 8e-20 | 98.36% | MN400979.1 |
| Feline panleukopenia virus isolate Gigucheon, partial genome | 106 | 106 | 100% | 8e-20 | 98.36% | MN400978.1 |
| Feline panleukopenia virus isolate MHS2019, complete genome | 106 | 106 | 100% | 8e-20 | 98.36% | MN908257.1 |
| Carnivore protoparvovirus 1 strain FPLV_CT1375/17, partial genome | 106 | 106 | 100% | 8e-20 | 98.36% | MK413738.1 |
| Carnivore protoparvovirus 1 strain FPLV_PA11334/17, partial genome | 106 | 106 | 100% | 8e-20 | 98.36% | MK413737.1 |
| Carnivore protoparvovirus 1 strain FPLV_58774/15, partial genome | 106 | 106 | 100% | 8e-20 | 98.36% | MK413731.1 |
| Carnivore protoparvovirus 1 strain FPLV_55611/15, partial genome | 106 | 106 | 100% | 8e-20 | 98.36% | MK413730.1 |
| Carnivore protoparvovirus 1 strain FPLV_52333eva/15, partial genome | 106 | 106 | 100% | 8e-20 | 98.36% | MK413729.1 |
| Carnivore protoparvovirus 1 strain FPLV_32369/15, partial genome | 106 | 106 | 100% | 8e-20 | 98.36% | MK413728.1 |
| Carnivore protoparvovirus 1 strain FPLV_149/15, partial genome | 106 | 106 | 100% | 8e-20 | 98.36% | MK413726.1 |

**Figure 1** Results from BLAST use

## 4. Discussion

That is, the consensus sequence has a PIN=98.36% with respect to the Feline Pan leukopenia virus, referring to several viral isolates and strains described in the Genbank ® database.

Hmmm... But only 10 results are shown...(Figure 1)

Only for reasons of space, the remaining 90 results are not shown, but the last 10 results indicate a PIN of 98.36%. Undoubtedly... it is the virus already mentioned...!!! (Figure 2)

| Description | | | | | | |
|---|---|---|---|---|---|---|
| Feline panleukopenia virus strain CMU-F01 capsid protein (VP2) gene, partial cds | 106 | 106 | 100% | 8e-20 | 98.36% | MK425497.1 |
| Feline parvovirus strain JL-3 capsid protein VP2 (vp2) gene, partial cds | 106 | 106 | 100% | 8e-20 | 98.36% | MK295775.1 |
| Feline panleukopenia virus strain FPLV-QDDX VP2 gene, partial cds | 106 | 106 | 100% | 8e-20 | 98.36% | MK301396.1 |
| Feline panleukopenia virus isolate TH091305 NS1 (NS1) and NS2 (NS2) genes, complete cds; and VP1 (VP1) and VP2 (VP2) genes, partial c | 106 | 106 | 100% | 8e-20 | 98.36% | KP019621.2 |
| Canine parvovirus isolate CN/HB1714 capsid protein (VP2) gene, partial cds | 106 | 106 | 100% | 8e-20 | 98.36% | MK518002.1 |
| Carnivore protoparvovirus 1 isolate H450 VP2 gene, partial cds | 106 | 106 | 100% | 8e-20 | 98.36% | MK251444.1 |
| Carnivore protoparvovirus 1 isolate H440 VP2 gene, partial cds | 106 | 106 | 100% | 8e-20 | 98.36% | MK251443.1 |
| Carnivore protoparvovirus 1 isolate H418 VP2 gene, partial cds | 106 | 106 | 100% | 8e-20 | 98.36% | MK251441.1 |
| Carnivore protoparvovirus 1 isolate H414 VP2 gene, partial cds | 106 | 106 | 100% | 8e-20 | 98.36% | MK251440.1 |
| Carnivore protoparvovirus 1 isolate H401 VP2 gene, partial cds | 106 | 106 | 100% | 8e-20 | 98.36% | MK251439.1 |

**Figure 2** Continuation of results from BLAST use

Thus, with the use of these biotools in our laboratory we have detected and identified several pathogens of veterinary interest [6, 7, 8, 9, 10].

## 5. Conclusion

Both programs can significantly help us to scrutinize the nucleotide sequence of any pathogen, in our case, and compare it with the official data stored in Genbank® (https://www.youtube.com/watch?v=LcvNpwS8hu8).

## Compliance with ethical standards

*Acknowledgments*

We thank the FAVET students (for achieving part of their dreams with us) and the Wolf Foundation, Illinois, USA (since 2020).

## References

[1] National Center for Biotechnology Information. Genbank®. NIH genetic sequence database. 2022. Available from: https://www.ncbi.nlm.nih.gov/genbank/

[2] Clustal Omega. Multiple Sequence Alignment. 2022. Available from: https://www.ebi.ac.uk/Tools/msa/clustalo/

[3] Vargas M, Jara MA, Navarro C. Use of primer design to detect the Glycoprotein C gene of Canine Herpes Virus by Polymerase Chain Reaction. Acta Scientific Veterinary Sciences 2022; 4 (3): 94-101.

[4] National Center for Biotechnology Information. BLAST. Basic Local Alignment Search Tool. 2022. Available from: https://blast.ncbi.nlm.nih.gov/Blast.cgi

[5] Mullis K, Faloona F. Specific synthesis of DNA in vitro via a polymerase catalyzed chain reaction. Methods Enzymol. 1987; 155: 335-350.

[6] Salas, V., Pizarro, J., Navarro, C. Phylogenetic analysis of canine distemper virus detected in Chile. 2018 International Journal of Current Research, 2018; 1(08): 72402-72407

[7] Cisternas F, Navarro C, Jara MA. Feline calicivirus. Molecular detection with primers design. EC Veterinary Science 2020; 5(7): 54-72.

[8] Carreño C, Navarro C, Jara MA. Design of primers in the molecular detection of Feline Panleukopenia Virus. World Journal of Biology Pharmacy and Health Sciences, 2021; 8(3): 019–029.

[9] Fuentes A, Jara MA, Navarro C. Detection of ul37 gene from Canine Herpes Virus by Polymerase Chain Reaction. International Journal of Multidisciplinary Research and Growth Evaluation. 2022; 3(1):401-407

[10] Vera C, Jara MA, Navarro C. A preliminary studio for the F gen of Canine Distemper Virus as target for phylogenetic analysis. GSC Advanced Research and Reviews. 2022, 10(01); 105–118