

# Bioinformatics

## Alignment tools

MSc. Vicente Machaca Arceda

Universidad Nacional de San Agustín de Arequipa

June 4, 2021

# Table of Contents

- 1 Introduction
  - Objectives
  
- 2 Alignment tools
  - BioPython
  - Muscle
  - Clustal
  - BLAST
  - Mega

# Table of Contents

- 1 Introduction
  - Objectives
- 2 Alignment tools
  - BioPython
  - Muscle
  - Clustal
  - BLAST
  - Mega

# Introduction

## Objectives

- Learn the alignment tools and databases used in Bioinformatics.

# Table of Contents

- 1 Introduction
  - Objectives
  
- 2 Alignment tools
  - **BioPython**
  - Muscle
  - Clustal
  - BLAST
  - Mega

# BioPython

## Definition

Biopython is a set of freely available tools for biological computation written in Python by an international team of developers.

# BioPython

## Pairwise global alignment

### Documentation.

```
1 from Bio import pairwise2
  from Bio.pairwise2 import format_alignment
3
  alignments = pairwise2.align.globalxx("ACCGT", "ACG")
5 print(format_alignment(*alignments[0]))
```

### Results:

```
1 ACCGT
  |  |
3 A-CG-
  Score=3
```

# BioPython

## Pairwise local alignment

### Documentation.

```
1 from Bio import pairwise2
2 from Bio.pairwise2 import format_alignment

4 alignments = pairwise2.align.localxx("ACCGT", "ACG")
5 print(format_alignment(*alignments[0]))
```

### Results:

```
1 1 ACCG
  | | |
3 1 A-CG
  Score=3

5
```



# Table of Contents

- 1 Introduction
  - Objectives
  
- 2 Alignment tools
  - BioPython
  - **Muscle**
  - Clustal
  - BLAST
  - Mega

# Muscle

## Multiple sequence alignment

It is a tool for multiple sequence alignment. Download page: [link](#).

# Muscle

## Multiple sequence alignment

Samples: [seqs2.fasta](#), [seqs3.fasta](#)

1 Basic usage

2

```
muscle -in <inputfile> -out <outputfile>
```

# Muscle

## Multiple sequence alignment

seqs3.fasta after alignment.

```
1 >sp|P13786|HBAZ_CAPHI Hemoglobin subunit zeta OS=Capra  
    hircus GN=HBZ1 PE=3 SV=2  
2 MSLTRTERTIL---WSKISTQADVIGTETLERLFCSCYPQAKTYFPHFDLHSGSAQLRAHG  
    SKVVAAVGDAVKSIDNVTSALSKLSELHAYVLRVDPVNFKFLSHCLLVTLASHFPADFTA  
4 DAHAAWDKFLSIVSGVLTEKYR  
    >sp|P69905|HBA_HUMAN Hemoglobin subunit alpha OS=Homo  
        sapiens GN=HBA1 PE=1 SV=2  
6 MVLSPADKTNVKAAGKVGGAHAGEY-AEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHG  
    KKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLP AEFTP  
8 AVHASLDKFLASVSTVLTSKYR  
    >sp|P01942|HBA_MOUSE Hemoglobin subunit alpha OS=Mus  
        musculus GN=Hba PE=1 SV=2  
10 MVLSGEDKSNIAAGKIGGHGA EYGAEALERMFASFP TTKTYFPHFDVSHGSAQVKGHG  
    KKVADALASAAGHLDDLPGALSALSDLHAHKLRVD-VNFKLLSHCLLVTLASHHPADFTP  
12 AVHASLDKFLASVSTVLTSKYR
```

# Table of Contents

- 1 Introduction
  - Objectives
  
- 2 Alignment tools
  - BioPython
  - Muscle
  - **Clustal**
  - BLAST
  - Mega

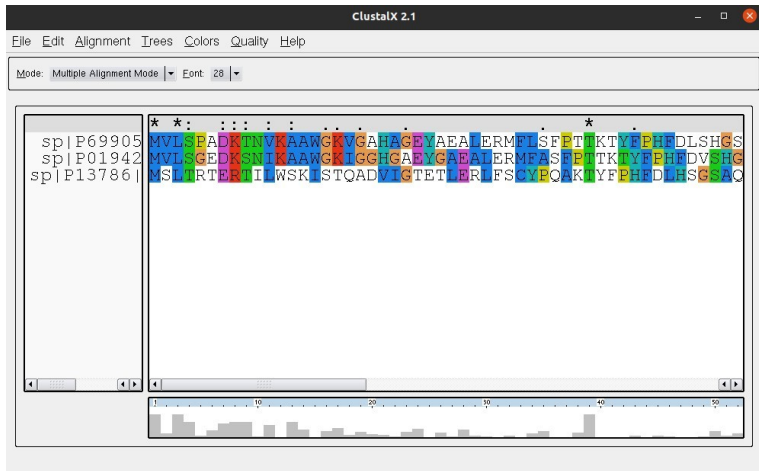
# Clustal

## Multiple sequence alignment

It is a tool for multiple sequence alignment like Muscle.  
Download page: [link](#).

# Clustal

Samples: [seqs2.fasta](#), [seqs3.fasta](#)



# Clustal

## Results

### Results:

#### CLUSTAL 2.1 multiple sequence alignment

```

sp|P69905|HBA_HUMAN      MVLSPADKTNVKAAWGKVGAHAGEY-
AEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHG
sp|P01942|HBA_MOUSE      MVLSGEDKSNIAAWGKIGGHGAIEYGAEALERMFASPTTKTYFPHFDVSHGSAQVKGHG
sp|P13786|HBAZ_CAPHI      MSLTRTERT---
ILWSKISTQADVIGTETLERLFSCYPQAKTYFPHFDLHSGSAQLRAHG
                                * *:  ::  *.*:  ..  :*:***:*  .:*  :*****:
****:..**

sp|P69905|HBA_HUMAN      KKVADALTNAVAHVDDMPNALSASDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTP
sp|P01942|HBA_MOUSE      KKVADALASAAGHLDDLPGALSASDLHAHKLRVD-
VNFKLLSHCLLVTLASHHPADFTP
sp|P13786|HBAZ_CAPHI      SKVVAAVGDAVKSIDNVTLSKLSLHAYVLRVDPVNFKLSHCLLVTLASHFPADFTA
                                **. *:  *.  :*:...**  **:***:  ****
****:*****:*  **:*.

sp|P69905|HBA_HUMAN      AVHASLDKFLASVSTVLTSKYR
sp|P01942|HBA_MOUSE      AVHASLDKFLASVSTVLTSKYR
sp|P13786|HBAZ_CAPHI      DAHAAWDKFLSIVSGVLTEKYR
                                .*:  *****:  **  ***.***

```

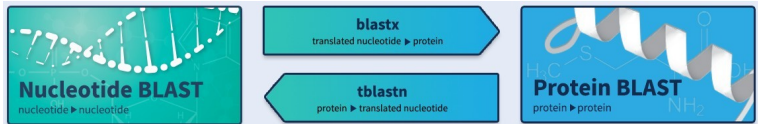


# Table of Contents

- 1 Introduction
  - Objectives
  
- 2 Alignment tools
  - BioPython
  - Muscle
  - Clustal
  - **BLAST**
  - Mega

# BLAST

Web page.



# BLAST

Evaluate BLAST with this sample: [HIVGRPCG.fasta](#).

Standard Nucleotide BLAST

blastn blastp blastx tblastn tblastx

BLASTN program search nucleotide databases using a nucleotide query. [more...](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#)

Query subrange

From

To

Or, upload file  HIV1\_M\_44.fasta [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

☐ Align two or more sequences [?](#)

Choose Search Set

Database ☒ Standard databases (nr etc.) ☐ rRNATS databases ☐ Genomic + transcript databases ☐ Betacoronavirus

Nucleotide collection (nr/nt) [?](#)

Organism

Optional ☐ exclude [Add organism](#)

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

Exclude ☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Optional

Limit to ☐ Sequences from type material

Optional

Entrez Query

Optional [View Entrez](#) [Create custom database](#)

Enter an Entrez query to limit search [?](#)

Program Selection

Optimize for

☒ Highly similar sequences (megablast)

☐ More dissimilar sequences (discontiguous megablast)

☐ Somewhat similar sequences (blastn)

Choose a BLAST algorithm [?](#)

Search database Nucleotide collection (nr/nt) using Megablast (Optimize for highly similar sequences)






☒ Show results in a new window

New columns added to the Description Table

Click "Select Columns" or "Manage Columns".

# BLAST

### Results:

Job Title	testing vicente		
RID	<a href="#">BNN6KF2T013</a>	Search expires on 06-06 09:35 am	<a href="#">Download All</a> 
Program	BLASTN 	<a href="#">Citation</a> 	
Database	nt	<a href="#">See details</a> 	
Query ID	lcl Query_44203		
Description	JQ403028		
Molecule type	dna		
Query Length	8750		
Other reports	<a href="#">Distance tree of results</a> <a href="#">MSA viewer</a> 		

### Filter Results

Organism ☐ exclude  
 Type common name, binomial, taxid or group name  
[+ Add organism](#)

Percent Identity  to  E value  to  Query Coverage  to

[Filter](#) [Reset](#)

[Descriptions](#)
[Graphic Summary](#)
[Alignments](#)
[Taxonomy](#)

Sequences producing significant alignments									
<div> <div>Download</div> <div>Select columns</div> <div>Show</div> <div>100</div> <div></div> </div>									
<div> <div>select all</div> <div>100 sequences selected</div> <div>GenBank</div> <div>Graphics</div> <div>Distance tree of results</div> <div>MSA Viewer</div> </div>									
	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	<a href="#">HIV-1 isolate HIVCH/BJD-V3538/2003 from Switzerland, partial genome</a>	Human immunod...	16159	16159	100%	0.0	100.00%	8750	<a href="#">JQ403028.1</a>
<input checked="" type="checkbox"/>	<a href="#">HIV-1 isolate P4039 from Democratic Republic of the Congo, complete genome</a>	Human immunod...	11762	12239	99%	0.0	90.92%	9686	<a href="#">MH705157.1</a>
<input checked="" type="checkbox"/>	<a href="#">HIV-1 isolate ML170-1985 from Kenya gag protein (gag) gene, complete cds; pol protein (pol) gene, partial cds;...</a>	Human immunod...	11605	11605	100%	0.0	90.67%	8995	<a href="#">AF539405.1</a>
<input checked="" type="checkbox"/>	<a href="#">HIV-1 isolate PBS888 from Democratic Republic of the Congo, complete genome</a>	Human immunod...	11559	12164	100%	0.0	90.54%	9744	<a href="#">MH705133.1</a>
<input checked="" type="checkbox"/>	<a href="#">Human immunodeficiency virus 1 proviral DNA, complete genome, clone: p52RW025A17</a>	Human immunod...	11376	11962	99%	0.0	90.18%	9638	<a href="#">AB287376.1</a>
<input checked="" type="checkbox"/>	<a href="#">HIV-1 isolate Q2K2P4V300497 from Kazakhstan gag protein (gag) and pol protein (pol) genes, partial cds; and vi...</a>	Human immunod...	11372	11372	99%	0.0	90.18%	8818	<a href="#">EF589039.1</a>
<input checked="" type="checkbox"/>	<a href="#">Human immunodeficiency virus 1 proviral DNA, complete genome, clone: p52RW025A46</a>	Human immunod...	11365	11957	99%	0.0	90.16%	9638	<a href="#">AB287377.1</a>
<input checked="" type="checkbox"/>	<a href="#">HIV-1 isolate Q2U20672 from Uzbekistan gag protein (gag) and pol protein (pol) genes, partial cds; and vi...</a>	Human immunod...	11348	11348	99%	0.0	90.14%	8816	<a href="#">AY829212.1</a>
<input checked="" type="checkbox"/>	<a href="#">HIV-1 isolate Q2K2P4V300502 from Kazakhstan nonfunctional gag protein (gag) gene, partial sequence; pol pro...</a>	Human immunod...	11326	11326	99%	0.0	90.10%	8810	<a href="#">EF589044.1</a>

# Table of Contents

- 1 Introduction
  - Objectives
- 2 Alignment tools
  - BioPython
  - Muscle
  - Clustal
  - BLAST
  - **Mega**

# Mega

Web page.

The objective of the MEGA software has been to provide tools for exploring, discovering, and analyzing DNA and protein sequences from an evolutionary perspective.

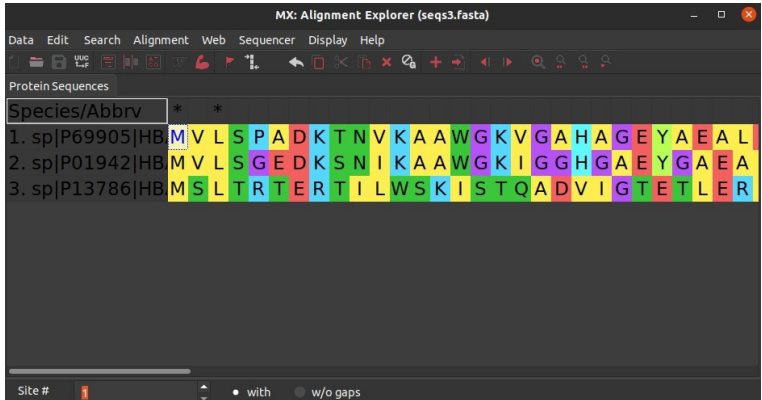
# Mega

Main window.



# Mega

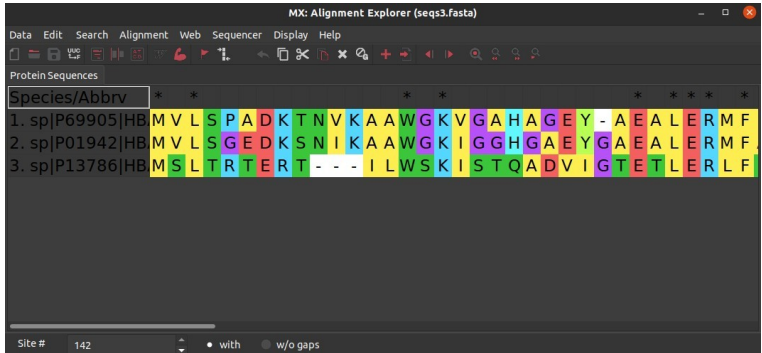
## Multiple sequence alignment.





# Mega

After alignment.



# Questions?



# References I



J. Xiong, *Essential bioinformatics*. Cambridge University Press, 2006.