

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

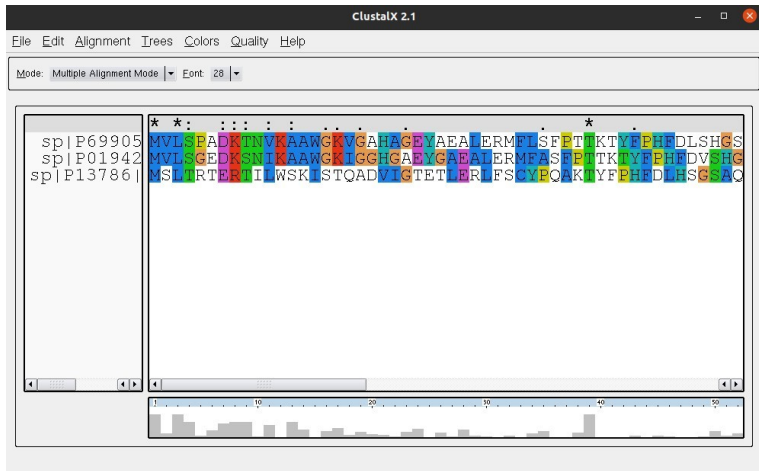
Clustal

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      MVLSGEDKSNIAAWGKIGGHGAIEYGAEALERMFAFPTTKTYFPHFDVSHGSAQVKGHG
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
                                .**:*:****:* ***.***

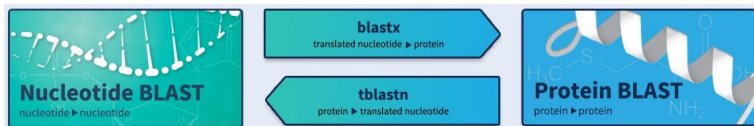
```


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

Choose 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

☐ 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

Limit to

☐ Sequences from type material

Entrez Query

[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 [?](#)

[BLAST](#)

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	BNN6KF2T013 Search expires on 06-06 09:35 am Download All ▼
Program	BLASTN Citation ▼
Database	nt See details ▼
Query ID	IcI/Query_44203
Description	JQ403028
Molecule type	dna
Query Length	8750
Other reports	Distance tree of results MSA viewer ?

Filter Results

Organism ☐ exclude
 only top 20 will appear

[+ Add organism](#)

Percent Identity to E value to Query Coverage to

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

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

Sequences producing significant alignments

[Download](#) ▼
 [New](#)
[Select columns](#) ▼
 [Show](#)

[?](#)

☒ select all 100 sequences selected

[GenBank](#)
[Graphics](#)
[Distance tree of results](#)
[New](#)
[MSA Viewer](#)

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	HIV-1 isolate HIV/CH/BID-V3538/2003 from Switzerland, partial genome	Human immunod...	16159	16159	100%	0.0	100.00%	8750	JQ403028.1
<input checked="" type="checkbox"/>	HIV-1 isolate P4039 from Democratic Republic of the Congo, complete genome	Human immunod...	11762	12239	99%	0.0	90.92%	9686	MH705157.1
<input checked="" type="checkbox"/>	HIV-1 isolate ML170-1986 from Kenya gag, protein (gag), gene, complete cds; pol protein (pol) gene, partial cds;...	Human immunod...	11605	11605	100%	0.0	90.67%	8995	AF539405.1
<input checked="" type="checkbox"/>	HIV-1 isolate PBS588 from Democratic Republic of the Congo, complete genome	Human immunod...	11559	12164	100%	0.0	90.54%	9744	MH705133.1
<input checked="" type="checkbox"/>	Human immunodeficiency virus 1 proviral DNA, complete genome, clone: p92RW025A17	Human immunod...	11376	11962	99%	0.0	90.18%	9638	AB287376.1
<input checked="" type="checkbox"/>	HIV-1 isolate 02KZPAV300497 from Kazakhstan gag, protein (gag), and pol protein (pol) genes, partial cds; and vt...	Human immunod...	11372	11372	99%	0.0	90.18%	8818	EF589039.1
<input checked="" type="checkbox"/>	Human immunodeficiency virus 1 proviral DNA, complete genome, clone: p92RW025A46	Human immunod...	11365	11957	99%	0.0	90.16%	9638	AB287377.1
<input checked="" type="checkbox"/>	HIV-1 isolate 02U20672 from Uzbekistan gag, protein (gag), and pol protein (pol) genes, partial cds; and vt...	Human immunod...	11348	11348	99%	0.0	90.14%	8816	AY829212.1
<input checked="" type="checkbox"/>	HIV-1 isolate 02KZPAV300502 from Kazakhstan nonfunctional gag, protein (gag), gene, partial sequence; pol prot...	Human immunod...	11326	11326	99%	0.0	90.10%	8810	EF589044.1

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.

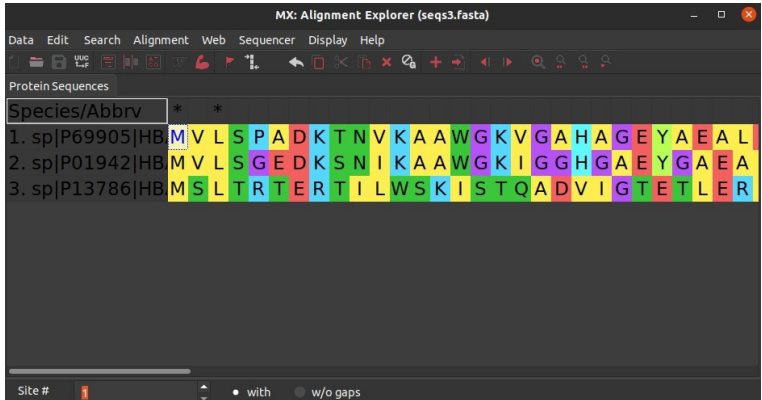
Mega

Main window.



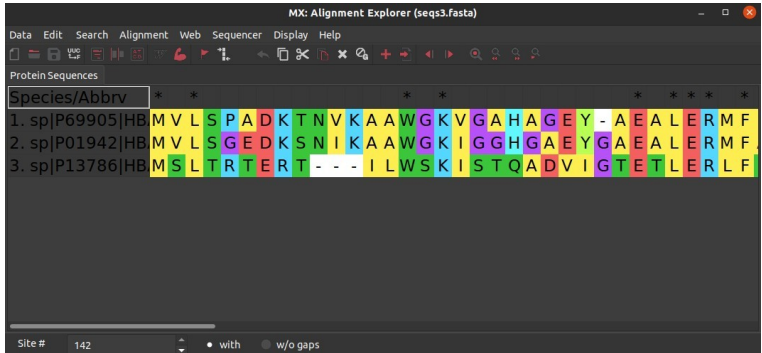
Mega

Multiple sequence alignment.



Mega

After alignment.



Questions?



References I