

# You've Caught Them All...Now What?

## Using Pokémon to Teach Bioinformatics

A visual step-by-step guide to gathering sequence data and  
building a phylogeny

Created by:  
**Janelle Nunez-Castilla**  
**Emily Warschefsky**  
**Laura Timm**

# **PART II: BIOINFORMATICS IS A BLAST**



## NCBI Home

## Resource List (A-Z)

All Resources

Chemicals &amp; Bioassays

Data &amp; Software

DNA &amp; RNA

Domains &amp; Structures

Genes &amp; Expression

Genetics &amp; Medicine

Genomes &amp; Maps

Homology

Literature

Proteins

Sequence Analysis

Taxonomy

Training &amp; Tutorials

Variation

## Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

[About the NCBI](#) | [Mission](#) | [Organization](#) | [NCBI News](#) | [Blog](#)

## Submit

Deposit data or manuscripts into NCBI databases



## Download

Transfer NCBI data to your computer



## Learn

Find help documents, attend a class or watch a tutorial



## Develop

Use NCBI APIs and code libraries to build applications



## Analyze

Identify an NCBI tool for your data analysis task



## Research

Explore NCBI research and collaborative projects



## Popular Resources

PubMed

Bookshelf

PubMed Central

PubMed Health

BLAST

Nucleotide

Genome

SNP

Gene

Protein

PubChem



## NCBI Announcements

Tree Viewer version 1.12 implements new API to markup trees

14 Feb 2017

Tree Viewer version 1.12 has several improvements, updates and bug fixes

Interim annotation updates for the human GRCh37.p13 and GRCh38.p10 assemblies

14 Feb 2017

Updates to the annotation of the human

February 22nd webinar: Introducing the

## Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance.

[Learn more](#)

### BLAST+ 2.6.0 released

A new version of the BLAST+ executables is available.

Tue, 10 Jan 2017 12:00:00 EST

[More BLAST news...](#)

### Web BLAST



### BLAST Genomes

Enter organism common name, scientific name, or tax id

Search

Nucleotide BLAST: Search X

Secure | [https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE\\_TYPE=BlastSearch&LINK\\_LOC=blasthome](https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome)

NIH U.S. National Library of Medicine NCBI National Center for Biotechnology Information Sign in to NCBI

BLAST® > blastn suite Home Recent Results Saved Strategies Help

blastn blastp blastx tblastn tblastx

Standard Nucleotide BLAST

BLASTN programs search nucleotide databases using a nucleotide query. [more...](#) Reset page Bookmark

Enter Query Sequence

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

[Clear](#) [Query subrange](#) [?](#)

From   
To

Or, upload file  Choose File No file chosen [?](#)

Job Title

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

Align two or more sequences [?](#)

You will end up at a page that looks like this

Choose Search Set

Database  Human genomic + transcript  Mouse genomic + transcript  Others (nr etc.):  
 Nucleotide collection (nr/nt) [?](#)

Organism Optional Enter organism name or id—completions will be suggested  Exclude [+](#)  
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

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

Limit to Optional  Sequences from type material  
 You Tube [Create custom database](#)

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

**DNA sequences:**

&gt;Pikachu

```
CGCCCGAGGGGGGCCACAGCCGCACCAGCTCCCCAACACCCCCCTTCTGCAGGTGGAGCTCACTGCCA  
CCCTGGCTAGCTGTTACACACAGTTCTGCTTTGTTCCCATTCTGTCTAGGCAGGATCATAGCAGA  
GAGACTTTGGGGTAGAAGGTCCATGCTAAGATGGGTATCGTGTCAAGGTCAGAGGTCACTCTAGTG  
AAATTCTTTGCCTTCCTCCAGCGGGTATCTAGAAGTTCTAGTACAGCTGCTTTCTGAGTGTGG  
GTATGTCCTCACTGATTGGTTGGTTTATAAGACGGGTCTCACTATGTGGTCCAGGGTGA  
CCTAAGAACTCACTATGCAGTCGAGGTGGTTGAACTCGGATCTCCTATCTCAGCTTCAGTACT  
GGGATTCCGGCATATACCAATGCCAGCTCGGAAGGTTTCTGCAACAAGAGGATTGTGTCTATCTGCCAACTCAGG  
TGTATCCCTGCAGAACTAGTTCACAGGGTTCTGCAACAAGAGGATTGTGTCTATCTGCCAACTCAGG  
CTCCCTGGTACTCTCCTAATTCTCTGTTCTCCAAGGCTGGCGGCTCTTCA
```



Open the Pokémon Sequence File. It should look similar to this.

▶ Export PDF

▶ Create PDF

▼ Edit PDF

Adobe Acrobat Pro

Easily edit text and images in PDF documents

[Start Now](#)

▶ Combine PDF

▶ Send Files

▶ Store Files



## DNA sequences:



&gt;Pikachu

CGCCCAGGGGGCCACAGCCGCACCAGCTCCCCAACACCCCCCTTCTGCAGGTGGAGCTCACTGCCA  
CCCTGGCTAGCTGTTACACACAGTTCTGCTTTGTTCCCGAGACTTTGGGGTAGAAGGTCCATGCTAAGATGGGTGA  
AAATTCTTTGCCTTCCTCCAGCGGGTATCTAGAAGTTGTTGGTTGGTTATAAGG  
GTATGTCTCCACTTGATTGGTTGGTTGGTTATAAGG  
CCTAAGAACTCACTATGCACTGAGGCTGGTTGAACCTG  
GGGATTCCGGCATATACCAATGCCAGCTCGGAAGGTTG  
TGTATCCCTGCAGAACTAGTTCACAGGGTTCTGCAACAAGAGGATTGTGTCTATCTGCCAACTCAGG  
CTCCCTGGTACTCTCCTAATTCTCTGTTCTCCAAGGCTGGCGGCTTTCA

Scroll through the document to find your assigned Pokémon. Highlight its corresponding DNA sequence. Copy *only* the sequence.

Sequence



Copy Ctrl+C

Highlight Text

Strikethrough Text

Add Note to Replace Text

Add Note to Text

Look Up "CTCCCTGGTACTCTCCTAATTCTCTGTTCTCCAAGGCTGGCGGCTTTCA"



Sign In

- ▶ Export PDF
- ▶ Create PDF
- ▼ Edit PDF

Adobe Acrobat Pro

Easily edit text and images in PDF documents

[Start Now](#)

- ▶ Combine PDF
- ▶ Send Files
- ▶ Store Files

Nucleotide BLAST: Search X

Secure | https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE\_TYPE=BlastSearch&LINK\_LOC=blasthome

NIH U.S. National Library of Medicine NCBI National Center for Biotechnology Information Sign in to NCBI

BLAST® > blastn suite Home Recent Results Saved Strategies Help

blastn blastp blastx tblastn tblastx

Enter Query Sequence

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

Reset page Bookmark

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

Clear Query subrange [?](#)

From  To

Or, upload file  Choose File No file chosen [?](#)

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

Align two or more sequences [?](#)

Pokémon sequence is pasted here

Choose Search Set

Database  Human genomic + transcript  Mouse genomic + transcript  Others (nr etc.)  
Nucleotide collection (nr/nt) [?](#)

Organism Optional Enter organism name or id—completions will be suggested  Exclude [+](#)  
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

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

Limit to Optional  Sequences from type material

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

You Tube Create custom database

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

Make sure the correct Database is selected

Nucleotide BLAST: Search X

Secure | https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE\_TYPE=BlastSearch&LINK\_LOC=blasthome

NIH U.S. National Library of Medicine NCBI National Center for Biotechnology Information Sign in to NCBI

BLAST® > blastn suite Home Recent Results Saved Strategies Help

Standard Nucleotide BLAST

blastn blastp blastx tblastn tblastx

Enter Query Sequence

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

Reset page Bookmark

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

CGCCCAGGGGGCCACAGCCGACCCACCCCCAACACCCCCCTTCTGCAGGTGGAGCTACTGCCACCC  
GGCTAGCTTACACACAGTTCTGCTTTGTTCCCATTTCCTGCTAGGCAGGATCATAGCAGAGAGACTT  
TGGGGTAGAGGTCCATGCTAAAGATGGGTGATCGTGCAGGGTCAGGGTCACTCTAGTGAATTCTTT  
GCCCTCCCTCAGGGGTATCTAGAAGTCTAGAACAGCTTGCTTTCTAGTGGGTATGCTCCACTTG  
ATTGGTTGGTTGGTTATAAGACGGGTCTCACTATGTTGCCAGGGTACCTAAGAACACTATGCA

Clear Query subrange [?](#)

From  To

Or, upload file Choose File No file chosen [?](#)

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

Align two or more sequences [?](#)

Choose Search Set

Database Human genomic + transcript  Mouse genomic + transcript  Others (nr etc.): Nucleotide collection (nr/nt) [?](#)

Organism Optional Enter organism name or id—completions will be suggested  Exclude [+](#)  
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

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

Limit to Optional Sequences from type material

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

YouTube Create custom database

It will look like this once the sequence is added

Now scroll down

Nucleotide BLAST: Search X

Secure | https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE\_TYPE=BlastSearch&LINK\_LOC=blasthome

Organism  
Optional

Enter organism name or id—completions will be suggested   Exclude

Exclude  
Optional

Models (XM/XP)  Uncultured/environmental sample sequences

Limit to  
Optional

Sequences from type material

Entrez Query  
Optional

Enter an Entrez query to limit search  [YouTube](#) [Create custom database](#)

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

[Algorithm parameters](#)

BLAST is a registered trademark of the National Library of Medicine

[Support center](#) [Mailing list](#) [YouTube](#)

NCBI

National Center for Biotechnology Information, U.S. National Library of Medicine  
8600 Rockville Pike, Bethesda MD, 20894 USA

[Policies and Guidelines](#) | [Contact](#)



NCBI Blast:Nucleotide Se X

Secure | https://blast.ncbi.nlm.nih.gov/Blast.cgi

U.S. National Library of Medicine NCBI National Center for Biotechnology Information Sign in to NCBI

BLAST® > blastn suite > RID-AE3HUTWD014 Home Recent Results Saved Strategies Help

Edit and Resubmit Save Search Strategies ▶ Formatting options ▶ Download YouTube How to read this page Blast report description

Job title: Nucleotide Sequence (615 letters)

RID AE3HUTWD014 (Expires on 02-18 21:52 pm)

Query ID Icl|Query\_160011

Description None

Molecule type nucleic acid

Query Length 615

Database Name nr

Description Nucleotide collection (nt)

Program BLASTN 2.6.1+ ▶ Citation

Other reports: ▶ Search Summary [Taxonomy reports] [Distance tree of results]

Graphic Summary

Your results should like similar to this

Distribution of the top 101 Blast Hits on 100 subject sequences ⓘ

Mouse over to see the title, click to show alignments

Color key for alignment scores

Score Range	Color
<40	Black
40-50	Blue
50-80	Green
80-200	Magenta
>=200	Red

Query

1 100 200 300 400 500 600

Now scroll down

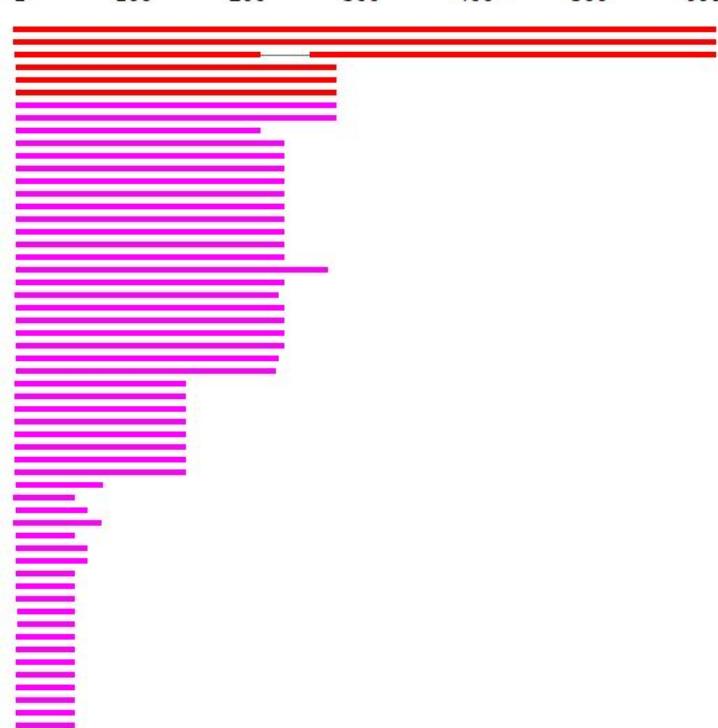
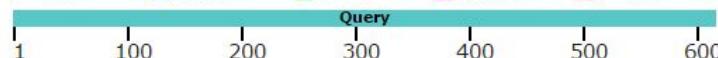


## Distribution of the top 101 Blast Hits on 100 subject sequences ⓘ

Mouse over to see the title, click to show alignments

## Color key for alignment scores

■ <40	■ 40-50	■ 50-80	■ 80-200	■ >=200
-------	---------	---------	----------	---------



This is a graphical representation of your results



Now scroll down

NCBI Blast:Nucleotide Se  Secure | <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

**Descriptions**

Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenBank Graphics Distance tree of results

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input checked="" type="checkbox"/>	Dipodomys californicus isolate 25-Dca SINE B1F sequence; and unknown gene	1136	1136	100%	0.0	100%	GQ506818.1
<input type="checkbox"/>	Dipodomys merriami isolate 25-Dme SINE B1F sequence; and unknown gene	1118	1118	100%	0.0	99%	GQ506819.1
<input type="checkbox"/>	Heteromys gaumeri isolate 25-Hqa SINE B1F sequence; and unknown gene	311	600	92%	1e-80	93%	GQ506820.1
<input type="checkbox"/>	Anomalurus sp. JJS-2010a unknown gene	237	237	45%	3e-58	82%	GQ506816.1
<input type="checkbox"/>	Pongo abelii BAC clone CH276-89I6 from chromosome 6, complete sequence	200	200	45%	3e-47	80%	AC206407.5
<input type="checkbox"/>	Pongo abelii BAC clone CH276-443O1 from chromosome 6, complete sequence	200	200	45%	3e-47	80%	AC206706.3
<input type="checkbox"/>	Chlorocebus aethiops BAC clone CH252-18K9 from chromosome 6, complete sequence	198	198	45%	1e-46	80%	AC242675.3
<input type="checkbox"/>	Chlorocebus aethiops BAC clone CH252-345L3 from chromosome 6, complete sequence	198	198	45%	1e-46	80%	AC241927.2
<input type="checkbox"/>	Felis catus FLA extended class II, class II, class III, proximal and central class I region genomic sequence	196	196	34%	4e-46	83%	EU153401.1
<input type="checkbox"/>	Homo sapiens advanced glycosylation end-product specific receptor (AGER), RefSeqGene on chromosome 6	189	189	38%	7e-44	81%	NG_029868.1
<input type="checkbox"/>	Pan troglodytes BAC clone CH251-607M18 from chromosome 6, complete sequence	189	189	38%	7e-44	81%	AC190421.3
<input type="checkbox"/>	Gorilla DNA sequence from clone CH255-56N15, complete sequence	189	189	38%	7e-44	81%	CU104674.1
<input type="checkbox"/>	Human DNA sequence from clone DASS-39F23 on chromosome 6, complete sequence	189	189	38%	7e-44	81%	BX284686.9
<input type="checkbox"/>	Human DNA sequence from clone DAQB-143M3 on chromosome 6, complete sequence	189	189	38%	7e-44	81%	AL845464.5
<input type="checkbox"/>	Human DNA sequence from clone CH501-300A18 on chromosome 6, complete sequence	189	189	38%	7e-44	81%	AL662884.11
<input type="checkbox"/>	Human DNA sequence from clone DAMC-108L3 on chromosome 6, complete sequence	189	189	38%	7e-44	81%	CR933878.7
<input type="checkbox"/>	Human DNA sequence from clone DAAP-218M18 on chromosome 6, complete sequence	189	189	38%	7e-44	81%	CR812478.6
<input type="checkbox"/>	H.sapiens G17 gene	189	189	38%	7e-44	81%	X80700.1

**Select 10**

Dipodomys californicus isolate 25-Dca SINE B1F sequence; and unknown gene

Dipodomys merriami isolate 25-Dme SINE B1F sequence; and unknown gene

Heteromys gaumeri isolate 25-Hqa SINE B1F sequence; and unknown gene

Anomalurus sp. JJS-2010a unknown gene

Pongo abelii BAC clone CH276-89I6 from chromosome 6, complete sequence

Pongo abelii BAC clone CH276-443O1 from chromosome 6, complete sequence

Chlorocebus aethiops BAC clone CH252-18K9 from chromosome 6, complete sequence

Chlorocebus aethiops BAC clone CH252-345L3 from chromosome 6, complete sequence

Felis catus FLA extended class II, class II, class III, proximal and central class I region genomic sequence

Homo sapiens advanced glycosylation end-product specific receptor (AGER), RefSeqGene on chromosome 6

Pan troglodytes BAC clone CH251-607M18 from chromosome 6, complete sequence

Gorilla DNA sequence from clone CH255-56N15, complete sequence

Human DNA sequence from clone DASS-39F23 on chromosome 6, complete sequence

Human DNA sequence from clone DAQB-143M3 on chromosome 6, complete sequence

Human DNA sequence from clone CH501-300A18 on chromosome 6, complete sequence

Human DNA sequence from clone DAMC-108L3 on chromosome 6, complete sequence

Human DNA sequence from clone DAAP-218M18 on chromosome 6, complete sequence

H.sapiens G17 gene

Selections should each have a different genus name



## Sequences producing significant alignments:

Select: All None Selected:10

Note the number 10

[All Alignments](#) [Download](#) [GenBank](#) [Graphics](#) [Distance tree of results](#)

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input checked="" type="checkbox"/>	<a href="#">Dipodomys californicus isolate 25-Dca SINE B1F sequence; and unknown gene</a>	1136	1136	100%	0.0	100%	GQ506818.1
<input type="checkbox"/>	<a href="#">Dipodomys merriami isolate 25-Dme SINE B1F sequence; and unknown gene</a>	1118	1118	100%	0.0	99%	GQ506819.1
<input checked="" type="checkbox"/>	<a href="#">Heteromys gaumeri isolate 25-Hqa SINE B1F sequence; and unknown gene</a>	311	600	92%	1e-80	93%	GQ506820.1
<input checked="" type="checkbox"/>	<a href="#">Anomalurus sp. JJS-2010a unknown gene</a>	237	237	45%	3e-58	82%	GQ506816.1
<input checked="" type="checkbox"/>	<a href="#">Pongo abelii BAC clone CH276-8916 from chromosome 6, complete sequence</a>	200	200	45%	3e-47	80%	AC206407.5
<input type="checkbox"/>	<a href="#">Pongo abelii BAC clone CH276-443O1 from chromosome 6, complete sequence</a>	200	200	45%	3e-47	80%	AC206706.3
<input checked="" type="checkbox"/>	<a href="#">Chlorocebus aethiops BAC clone CH252-18K9 from chromosome 6, complete sequence</a>	198	198	45%	1e-46	80%	AC242675.3
<input type="checkbox"/>	<a href="#">Chlorocebus aethiops BAC clone CH252-345L3 from chromosome 6, complete sequence</a>	198	198	45%	1e-46	80%	AC241927.2
<input checked="" type="checkbox"/>	<a href="#">Felis catus FLA extended class II, class II, class III, proximal and central class I region genomic sequence</a>	196	196	34%	4e-46	83%	EU153401.1
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens advanced glycosylation end-product specific receptor (AGER), RefSeqGene on chromosome 6</a>	189	189	38%	7e-44	81%	NG_029868.1
<input checked="" type="checkbox"/>	<a href="#">Pan troglodytes BAC clone CH251-607M18 from chromosome 6, complete sequence</a>	189	189	38%	7e-44	81%	AC190421.3
<input checked="" type="checkbox"/>	<a href="#">Gorilla DNA sequence from clone CH255-56N15, complete sequence</a>	189	189	38%	7e-44	81%	CU104674.1
<input type="checkbox"/>	<a href="#">Human DNA sequence from clone DASS-39F23 on chromosome 6, complete sequence</a>	189	189	38%	7e-44	81%	BX284686.9
<input type="checkbox"/>	<a href="#">Human DNA sequence from clone DAQB-143M3 on chromosome 6, complete sequence</a>	189	189	38%	7e-44	81%	AL845464.5
<input type="checkbox"/>	<a href="#">Human DNA sequence from clone CH501-300A18 on chromosome 6, complete sequence</a>	189	189	38%	7e-44	81%	AL662884.11
<input type="checkbox"/>	<a href="#">Human DNA sequence from clone DAMC-108L3 on chromosome 6, complete sequence</a>	189	189	38%	7e-44	81%	CR933878.7
<input type="checkbox"/>	<a href="#">Human DNA sequence from clone DAAP-218M18 on chromosome 6, complete sequence</a>	189	189	38%	7e-44	81%	CR812478.6
<input type="checkbox"/>	<a href="#">H.sapiens G17 gene</a>	189	189	38%	7e-44	81%	X80700.1
<input type="checkbox"/>	<a href="#">Homo sapiens pre-B-cell leukemia homeobox 2, mRNA (cDNA clone IMAGE:3501925), with apparent retained intron</a>	189	189	38%	7e-44	81%	BC003111.1
<input checked="" type="checkbox"/>	<a href="#">Macaca mulatta Major Histocompatibility Complex BAC MMU45P02, complete sequence</a>	185	185	44%	9e-43	79%	AC148714.1
<input type="checkbox"/>	<a href="#">Gorilla gorilla gorilla chromosome 6 clone CH277-51M21, complete sequence</a>	183	183	38%	3e-42	81%	AC270183.1

Genus names are shown underlined in red



## Sequences producing significant alignments:

Select: All None Selected: 10

[All Alignments](#) [Download](#) [GenBank](#) [Graphics](#) [Distance tree of results](#)

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input checked="" type="checkbox"/>	Dipodomys californicus isolate 25-Dca SINE B1F sequence; and unknown gene	1136	1136	100%	0.0	100%	GQ506818.1
<input type="checkbox"/>	Dipodomys merriami isolate 25-Dme SINE B1F sequence; and unknown gene	1118	1118	100%	0.0	99%	GQ506819.1
<input checked="" type="checkbox"/>	Heteromys gaumeri isolate 25-Hga SINE B1F sequence; and unknown gene	311	600	92%	1e-80	93%	GQ506820.1
<input checked="" type="checkbox"/>	Anomalurus sp. JJS-2010a unknown gene	237	237	45%	3e-58	82%	GQ506816.1
<input checked="" type="checkbox"/>	Pongo abelii BAC clone CH276-89I6 from chromosome 6, complete sequence	200	200	45%	3e-47	80%	AC206407.5
<input type="checkbox"/>	Pongo abelii BAC clone CH276-443O1 from chromosome 6, complete sequence	200	200	45%	3e-47	80%	AC206706.3
<input checked="" type="checkbox"/>	Chlorocebus aethiops BAC clone CH252-18K9 from chromosome 6, complete sequence	198	198	45%	1e-46	80%	AC242675.3
<input type="checkbox"/>	Chlorocebus aethiops BAC clone CH252-345L3 from chromosome 6, complete sequence	198	198	45%	1e-46	80%	AC241927.2
<input checked="" type="checkbox"/>	Felis catus FLA extended class II, class II, class III, proximal and central class I region genomic sequence	196	196	34%	4e-46	83%	EU153401.1
<input checked="" type="checkbox"/>	Homo sapiens advanced glycosylation end-product specific receptor (AGER), RefSeqGene on chromosome 6	189	189	38%	7e-44	81%	NG_029868.1
<input checked="" type="checkbox"/>	Pan troglodytes BAC clone CH251-607M18 from chromosome 6, complete sequence	189	189	38%	7e-44	81%	AC190421.3
<input checked="" type="checkbox"/>	Gorilla DNA sequence from clone CH255-56N15, complete sequence	189	189	38%	7e-44	81%	CU104674.1
<input type="checkbox"/>	Human DNA sequence from clone DASS-39F23 on chromosome 6, complete sequence	189	189	38%	7e-44	81%	BX284686.9
<input type="checkbox"/>	Human DNA sequence from clone DAQB-143M3 on chromosome 6, complete sequence	189	189	38%	7e-44	81%	AL845464.5
<input type="checkbox"/>	Human DNA sequence from clone CH501-300A18 on chromosome 6, complete sequence	189	189	38%	7e-44	81%	AL662884.11
<input type="checkbox"/>	Human DNA sequence from clone DAMC-108L3 on chromosome 6, complete sequence	189	189	38%	7e-44	81%	CR933878.7
<input type="checkbox"/>	Human DNA sequence from clone DAAP-218M18 on chromosome 6, complete sequence	189	189	38%	7e-44	81%	CR812478.6
<input type="checkbox"/>	H.sapiens G17 gene	189	189	38%	7e-44	81%	X80700.1
<input type="checkbox"/>	Homo sapiens pre-B-cell leukemia homeobox 2, mRNA (cDNA clone IMAGE:3501925) with apparent retained intron	189	189	38%	7e-44	81%	BC003111.1
<input checked="" type="checkbox"/>	Macaca mulatta Major Histocompatibility Complex BAC MMU45P02, complete sequence	185	185	44%	9e-43	79%	AC148714.1
<input type="checkbox"/>	Gorilla gorilla gorilla chromosome 6 clone CH277-51M21, complete sequence	183	183	38%	3e-42	81%	AC270183.1

NCBI Blast:Nucleotide Se MAFFT alignment and N

Secure | https://blast.ncbi.nlm.nih.gov/Blast.cgi

Sequences producing significant alignments:

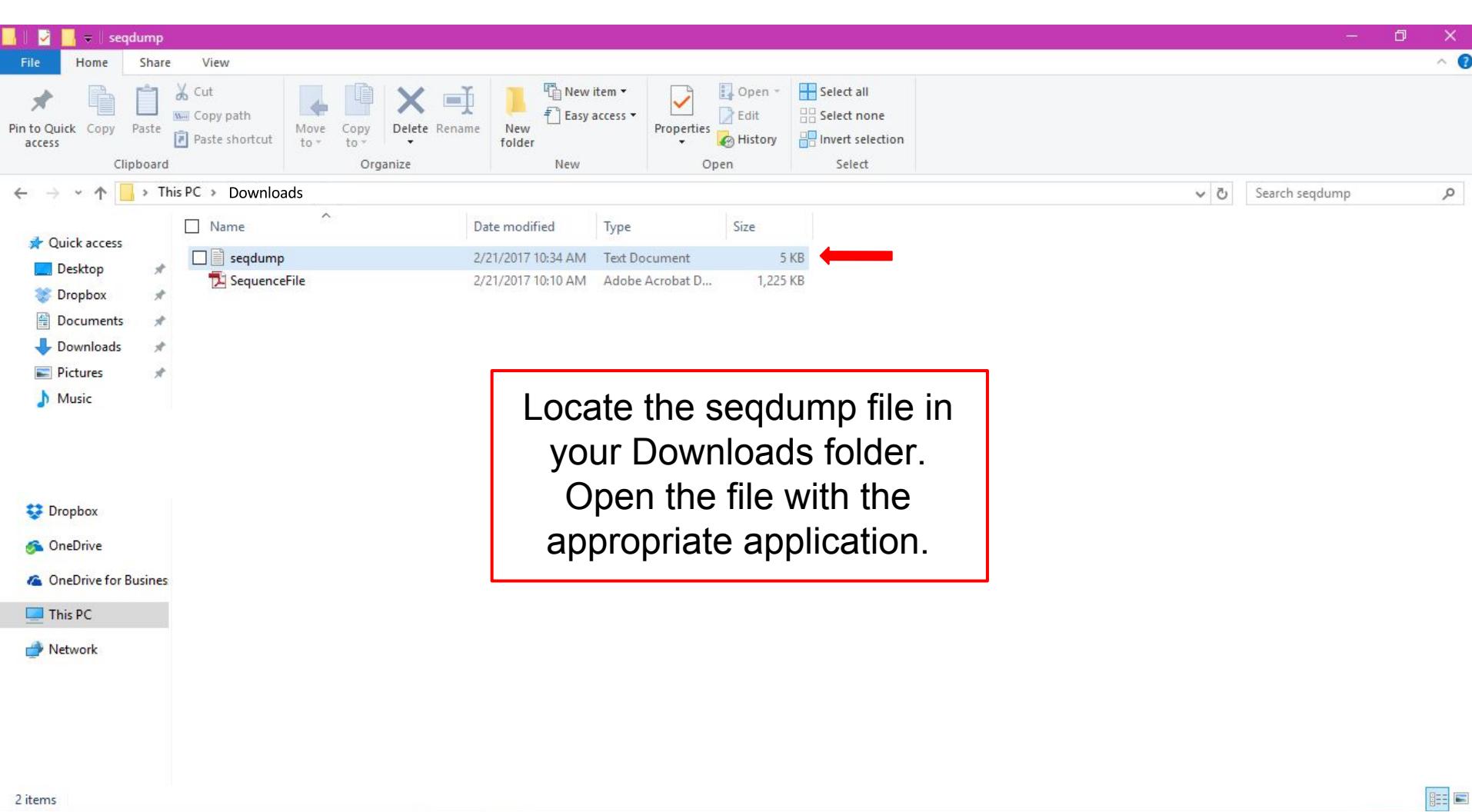
Select: All None Selected: 10

All Alignments Download GenBank Graphics Distance tree of results

Description

	Max score	Total score	Query cover	E value	Ident	Accession
FASTA (complete sequence)	1136	1136	100%	0.0	100%	GQ506818.1
<b>FASTA (aligned sequences)</b>	1118	1118	100%	0.0	99%	GQ506819.1
GenBank (complete sequence)	311	600	92%	1e-80	93%	GQ506820.1
Hit Table (text)	237	237	45%	3e-58	82%	GQ506816.1
Hit Table (CSV)	200	200	45%	3e-47	80%	AC206407.5
Text	200	200	45%	3e-47	80%	AC206706.3
XML	198	198	45%	1e-46	80%	AC242675.3
ASN.1	198	198	45%	1e-46	80%	AC241927.2
<b>Continue</b>	196	196	34%	4e-46	83%	EU153401.1
<b>2. Click here</b>	189	189	38%	7e-44	81%	NG_029868.1
<input checked="" type="checkbox"/> Homo sapiens advanced glycosylation end-product specific receptor (AGER), RefSeqGene on chromosome 6	189	189	38%	7e-44	81%	AC190421.3
<input checked="" type="checkbox"/> Pan troglodytes BAC clone CH251-607M18 from chromosome 6, complete sequence	189	189	38%	7e-44	81%	CU104674.1
<input checked="" type="checkbox"/> Gorilla DNA sequence from clone CH255-56N15, complete sequence	189	189	38%	7e-44	81%	BX284686.9
<input type="checkbox"/> Human DNA sequence from clone DASS-39F23 on chromosome 6, complete sequence	189	189	38%	7e-44	81%	AL845464.5
<input type="checkbox"/> Human DNA sequence from clone DAQB-143M3 on chromosome 6, complete sequence	189	189	38%	7e-44	81%	AL662884.11
<input type="checkbox"/> Human DNA sequence from clone CH501-300A18 on chromosome 6, complete sequence	189	189	38%	7e-44	81%	CR933878.7
<input type="checkbox"/> Human DNA sequence from clone DAMC-108L3 on chromosome 6, complete sequence	189	189	38%	7e-44	81%	CR812478.6
<input type="checkbox"/> Human DNA sequence from clone DAAP-218M18 on chromosome 6, complete sequence	189	189	38%	7e-44	81%	X80700.1
<input type="checkbox"/> H.sapiens G17 gene	189	189	38%	7e-44	81%	BC003111.1
<input type="checkbox"/> Homo sapiens pre-B-cell leukemia homeobox 2, mRNA (cDNA clone IMAGE:3501925) with apparent retained intron	189	189	38%	7e-44	81%	AC148714.1
<input checked="" type="checkbox"/> Macaca mulatta Major Histocompatibility Complex BAC MMU457P02, complete sequence	183	183	38%	9e-43	79%	AC270183.1

https://blast.ncbi.nlm.nih.gov/Blast.cgi#dsDownload



seqdump - WordPad

A screenshot of the Microsoft Word ribbon. The 'Home' tab is highlighted in blue, indicating it is the active tab. Below the ribbon, there are two groups of icons: 'Cut' and 'Copy'. Each group contains a small icon followed by the word 'Cut' or 'Copy'.

Clipboard

Your sequence file should look similar to this

board Font Paragraph

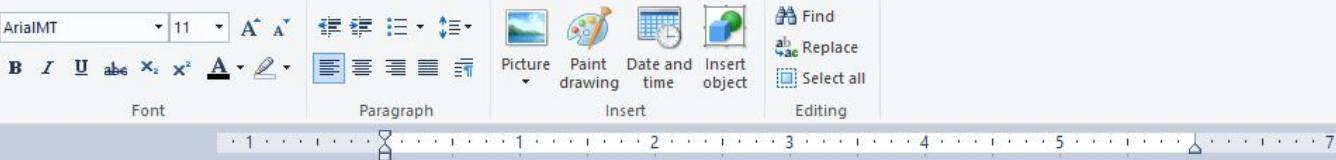
Your sequence file should look similar to this

```
>GQ506818.1 Dipodomys californicus isolate 25-Dca SINE B1F  
sequence; and unknown gene  
CGCCCAGGGGGGCCACAGCCGCACAGCTCCCCAACACCCCTTCTGCAGGTGGAGCTCA  
TGCCACCCCTGGCTAGC  
TGTTACACACAGTTCTGCTTTGTTCCCATTCTGTCTAGGCAGGATCATAGCAGAGAGACT  
TTTGGGGGTTAGAAGG  
TCCATGCTTAAGATGGGTGATCGTGTCAAGGGTCAGAGGTCACTCTAGTGAAATTCTTTGCCT  
CCTTCCAGGGGTATC  
TAGAAGTTCTAGAACAGCTGCTTTCTGAGTGTGGGTATGTCTCCACTTGATTGGTTGG  
TTGGTTTTATAAGACG  
GGGTCTCACTATGTGGTCCAGGTTGACCTAAGAACTCACTATGCAGTCAGGCTGGTTTGAA  
TCGGATCTCTATCT  
CAGCTCTCACTGACTGGGATTCCGGCATATACCACCATGCCAGCTGGGAAGGTTGTTCTA  
GGTGATTTGTGTATAA  
TCCACATGTATCCCTGAGAACTAGTTCACAGGGTTCTGCAACAAGAGGATTGTGTCTATCT  
GCCAACTCAGGCTCCC  
TGGTACTCTCTCTAATTCTCTGTTCTCCAAGGCTCTGGCGGCTCTTCA  
>GQ506820.1 Heteromys gaumeri isolate 25-Hga SINE B1F sequence;  
and unknown gene  
TGCCCAGGGGGGCCACAGCCGTACCAAGCTCTCCAACACCCCTTCTGCAGGTGGAGCTCA  
TGCCACCCCTGGCTAGC  
TGTTCTGCACATTCTGCTTTGTTCCCATTCTGTCTAGGCAGGATCATACCAGAGAGAGC  
ACTTTGGGGTTAGA  
AGGCCTGAGCTAAGATAGGTGATCTGTCAGGGTCAGAGGTCACTCTAGTGAAATTCTGGGA  
AACCACTCTTCTGA  
GTGTGGGTATGTCTTGGTTGAATTGGGTTGGTTTATAAGACGGGTCACTATG  
GGTCCAGGGTGA  
AGAACTCACTATGCAGTCAGGCTGGTTGAACCTGGATTCTACCTCAGCCT  
GCATACTGGGATTCA  
GGCTTTAACCATGGCTGGTGGAAAGGTTGTTGTTGGTAATTGTGTATAATCCACATAC  
AGTTGAGAGGGTTTC
```

# PART IV: COMPUTING RELATIONSHIPS

Return to the seqdump file. Add an empty line to the top of the file.

```
>GQ506818.1 Dipodomys californicus isolate 25-Dca SINE B1F
sequence; and unknown gene
CGCCCAGGGGGCCACAGCCGACCAGCTCCCAACACCCCCCTCTGCAGGTGGAGCTCAC
TGCCACCCCTGGCTAGC
TGTTCACACACAGTTCTGCTTTGTTCCATTCTGTCTAGGCAGGATCATAGCAGAGAGACT
TTTGGGGGTTAGAAGG
TCCATGCTAAAGATGGGTATCGTGTAGGGTCAGAGGTCACTCTAGTGAATTCTTTGCCTT
CCTTCCAGCGGGTATC
TAGAAGTTCTAGAACAGCTGCTTTCTGAGTGTGGGTATGTCTCCACTTGATTTGGTTGGT
TTGGTTTATAAAGACG
GGGTCTCACTATGTGGTCCAGGGTGACCTAACAGACTCACTATGCAGTCGAGGCTGGTTGAAC
TCGGATCTTCTATCT
CAGCTTCTCAGTACTGGGATTCCGGCATATACCACCATGCCAGCTGGAAAGGTTGTTCAATT
GGTGTGTGTATAA
TCCACATGTATCCCTGCGAACACTAGTTCACAGGGTTCTGCAACAAGAGGATTGTGTCTATCT
GCCAACCTCAGGCTCCCC
TGGTACTCTCTCCAATTCTCTCTGTTCTCAAGGGCTCTGGGGCTCTTCA
>GQ506820.1 Heteromys gaumeri isolate 25-Hga SINE B1F sequence;
and unknown gene
GCCCAAGGGGCCACAGCCGTACCAAGCTCTCCAAACACCCCCCTCTGCAGGTGGAGCTCACT
GCCACCCCTGGCTAGCT
GTTCTGCACATTCTCTGCTTTGTTCCATTCTGTCTAGGCAGGATCATACCAGAGAGAGCA
CTTTTGGGGTTAGAA
GGCCTGAGCTAAAGATAGGTATCGTGTAGGGTCAGAGGTCACTCTAGTGAATTGGTTGGTTATAAGACGGGG
>GQ506820.1 Heteromys gaumeri isolate 25-Hga SINE B1F sequence;
and unknown gene
GCTCTTCTGAGTGTGGGTATGTCTTGGTTGAATTGGTTGGTTATAAGACGGGG
TCTCACTATGTGGTCC
AGGTTGACCTAACAGACTCACTATGCAGTCGAGGCTGGTTGAACCTGGATTTTTTCT
ACCTCAGCCTCGCATA
GTTGGGTTTAAAGGCTTTAAATCAACATGGCTGCTGAAAGGTTGTTGGTTATAAGACGGGG
```



Empty line added

>GQ506818.1 Dipodomys californicus isolate 25-Dca SINE B1F sequence; and unknown gene  
CGCCCAGGGGGCCACAGCCGACCAGCTCCCCAACACCCCCCTTCTGCAGGTGGAGCTCAC  
TGCCACCCCTGGCTAGC  
TGTTACACACAGTTCTGCTTTGTTCCATTCTGTCTAGGCAGGATCATAGCAGAGAGACT  
TTGGGGGGTTAGAAGG  
TCCATGCTAAAGATGGGTATCGTGTCAAGGGTCAGAGGTCACTCTAGTGAATTTCTTTGCCTT  
CCTTCCAGCGGGTATC  
TAGAAGTTCTAGAACAGCTGCTTTCTGAGTGTGGGTATGTCTCCACTGATTGGTTGG  
TTGGTTTATAAAGACG  
GGGTCTCACTATGTGGTCAGGTTGACCTAAGAACTCACTATGCAGTCGAGGCTGGTTTGAAAC  
TCGGATCTCTCATCT  
CAGCTCTCAGTACTGGGATTCCGGCATATACCACCATGCCAGCTGGAAAGGGTTGTTCAT  
GGTGAATTGGTATCAA  
TCCACATGTAACCTGCAGAACTAGTTCACAGGGTTCTGCAACAAGAGGATTGTGTCTATCT  
GCCAACTCAGGCTCCC  
TGGTACTCTCCCTAATTCCCTCTGTGTTCTCCAAGGCTCTGGGGCTCTTCA  
>GQ506820.1 Heteromys gaumeri isolate 25-Hga SINE B1F sequence;  
and unknown gene  
GCCCAAGGGGGGCCACAGCGTACCAAGCTCTCCAAACACCCCCCTTCTGCAGGTGGAGCTCACT  
GCCACCCCTGGCTAGCT  
GTTCTGCACATTCTGCTTTGTTCCATTCTGTCTAGGCAGGATCATACCAGAGAGAGCA  
CTTTTGGGGGGTTAGAA  
GGCCTGAGCTAAAGATAGGTGATCCTGTCAAGGGTCAGAGGTCACTCTAGTGAATTTCT  
>GQ506820.1 Heteromys gaumeri isolate 25-Hga SINE B1F sequence;  
and unknown gene  
GCTCTTCTGAGTGTGGGTATGTCTTGTTGAATTGGGTTGGTTATAAGACGGGG  
TCTCACTATGTGGTCC  
AGGTTGACCTAAGAACTCACTATGCAGTCGAGGCTGGTTGAACCTGGATTTTTTCT  
AAGCTGAGCTCGGATTA

File Home View  
Cut Copy  
Paste  
Clipboard

CourierNewPSMT 11 A A B I U abe x x A A

Picture Paint drawing Date and time Insert object

Find Replace  
Select all

Font Paragraph Insert Editing

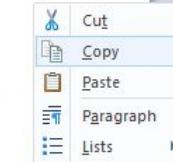
Paste your  
Pokémon name  
and sequence at  
the top of your  
seqdump file.

Do not worry about  
differences in  
formatting. This will  
not affect the next  
step.

```
>Pikachu
CGCCCAGGGGGGCCACAGCCGCACCAGCTCCCCAACACCCCCCTTCTGCAGGT
GGAGCTCACTGCCA
CCCTGGCTAGCTGTTACACAGTCCTGCTTTGTTCCATTCTGTCTAGGCAG
GATCATAGCAGA
GAGACTTTGGGGTAGAAGGTCCATGCTAAGATGGGTATCGTGTCAAGGTCAGAG
GTCACTCTAGTG
AAATTCTTTGCCTCCTCCAGCGGGTATCTAGAAGTTCTAGTACAGCTTGCTTT
TCTGAGTGTTG
GTATGTCCTCCACTTGATTGGTTGGTTATAAGACGGGTCTCACTATGTG
GTCCAGGTTGA
CCTAAGAACTCACTATGCAGTCGAGGCTGGTTGAECTCGGATCTCCTATCTCAG
CTTCTCAGTACT
GGGATTTCCGGCATACCAATGCCAGCTCGGAAGGTTGTTCAATTGGTATTGT
GTATAATCCACA
TGTATCCCTGCAGAACTAGTTCACAGGGTTCTGCAACAAGAGGATTGTGTCTATCT
GCCAACTCAGG
CTCCCTGGTACTCTCTCTTAATTCTCTGTTCTCAAGGCTCTGGCGGCTCTT
TCA
>Q506818.1 Dipodomys californicus isolate 25-Dca SINE B1F
sequence; and unknown gene
CGCCCAGGGGGGCCACAGCCGCACCAGCTCCCCAACACCCCCCTTCTGCAGGTGGAGCTCAC
TGCCACCTGGCTAGC
TGTTACACACAGTCCTGCTTTGTTCCATTCTGTCTAGGCAGGATCATAGCAGAGAGACT
TTGGGGGGTTAGAAGG
TCCATGCTAAGATGGGTATCGTGTCAAGGGTCAGAGGTCACTCTAGTGAATTCTTTGCCTT
CTTCTCAGCGGGTATC
TAGAAGTTCTAGAACAGCTGCTTTCTGAGTGTGGGTATGTCTCCACTTGATTGGTTGGT
TTGGGGTTATAAGAGC
GGGTCTCACTATGTGGTCAGGTTGACTTAAGAACTCACTATGCAGTCGAGGCTGGTTTGAAAC
TCCGGAATTCTCTATCT
```

Select all sequences and copy

```
CCCAGGGGGCCACAGCGCACCAAGCTCCCCGACACCCCTTCCTCTGCAGGTGGATCCCAC TG  
TCACCCCGGCTGACTG  
TTTGACACACTTCCTGCTTTGTTCCACTTCCTATCTAGGCAGGATCATAGCAGAGAGGGGC  
CTTTGGGGTGGAGAG  
GACCGAGCTGAGATAGGCTGGAGATGTCAGGGACAGAGGCCATTCCAGTGTCTTAGTTCTGC  
CTTTCTTCCCACGGGT  
>AC190421.3 Pan troglodytes BAC clone CH251-607M18 from  
chromosome 6, complete sequence  
CCCAGGGGGCCACAGCGCACCAAGCTCCCCGACACCCCTTCCTCTGCAGGTGGATCCCAC TG  
TCACCCCGGCTGACTG  
TTTGACACACTTCCTGCTTTGTTCCACTTCCTATCTAGGCAGGATCATAGCAGAGAGGGGC  
CTTTGGGGTGGAGAG  
GACCGAGCTGAGATAGGCTGAAGATGTCAGGGACAGAGGCCATTCCAGTGTCTTAGTTCTGC  
CTTTCTTCCCACGGGT  
>CU104674.1 Gorilla DNA sequence from clone CH255-56N15,  
complete sequence  
CCCAGGGGGCCACAGCGCACCAAGCTCCGACACCCCTTCCTCTGCAGGTGGATCCCAC TG  
TCACCCCGGCTGACTG  
TTTGACACACTTCCTGCTTTGTTCCACTTCCTATCTAGGCAGGATCATAGCAGAGAGGGGC  
CTTTGGGGTGGAGAG  
GACCGAGCTGAGATAGGCTGAAGATGTCAGGGACAGAGGCCATTCCAGTGTCTTAGTTCTGC  
CTTTCTTCCCACGGGT  
>AC148714.1 Macaca mulatta Major Histocompatibility Complex BAC  
MMU457P02, complete sequence  
ACCCGGGGAAAGGCAACTAAGTTCACTGGAGTGGCCTCTGCCCTTGACATCTCCAGC  
CTATCTCAGCTTGGTC  
CCTGTCACCCAAAAGGCCCTCTGCTATGATCTTGCTAGGTAGGAAGTGGAACAAAAG  
CAGGAAGTGTGCAAAA  
CAGTCAGCCGGGTGACAGTAGGATCCACCTGCAAGAGGAAGGGGGTGTGGGGAGCTGGTGC GG  
CTGTGGCCCCCTGGG
```



# MAFFT version 7

Multiple alignment program for amino acid or nucleotide sequences



## Download version

[Mac OS X](#)[Windows](#)[Linux](#)[Source](#)

## Online version

[Alignment](#)[mafft --add](#)[Merge](#)[Phylogeny](#)[Rough tree](#)

## Merits / limitations

[Algorithms](#)[Tips](#)[Benchmarks](#)[Feedback](#)[Follow](#)

For a large number of short sequences, try [an experimental service](#) (2016/Jul).

## Multiple sequence alignment and NJ / UPGMA phylogeny

### **Input:**

Paste protein or DNA sequences in fasta format. [Example](#)



Sequences are pasted here

or upload a **plain text** file:  [Choose File](#) No file chosen

- Use structural alignment(s)
- Allow unusual symbols (Selenocysteine "U", Inosine "I", non-alphabetical characters, etc.) [Help](#)

### UPPERCASE / lowercase:

- Same as input

# MAFFT version 7

Multiple alignment program for amino acid or nucleotide sequences



## Download version

[Mac OS X](#)[Windows](#)[Linux](#)[Source](#)

## Online version

[Alignment](#)[mafft --add](#)[Merge](#)[Phylogeny](#)[Rough tree](#)

## Merits / limitations

[Algorithms](#)[Tips](#)[Benchmarks](#)[Feedback](#)[Follow](#)

For a large number of short sequences, try [an experimental service](#) (2016/Jul).

## Multiple sequence alignment and NJ / UPGMA phylogeny

### Input:

Paste protein or DNA sequences in fasta format. [Example](#)

```
GACCGAGCTGAGATAGCTGGAGATGTCAAGGGGACAGAGGCCATTCCAGTGAITCTAGTCTGCC'▲  
>AC190421_3 Pan troglodytes BAC clone CH251-607M18 from chromosome 1  
CCTTGGGGGGCACACAGCCCACAGCTCCGACACCCCCCTTCCTCTGCAGGTGGATCCACTGT'  
TTTGACACACTCTGCTTTGTCTCCACTTCTTAITAGGCAGGATCATAGCAGAGAGGGGGCC'  
GACCGAGCTGAGATAGCTGGAGATGTCAAGGGGACAGAGGCCATTCCAGTGAITCTAGTCTGCC'  
>GU104674_1 Gorilla DNA sequence from clone CH255-56N15, complete  
CCCCAGGGGGGACACAGCCGACCCAGCTCTCCGACACCCCCCTTCCTCTGCAGGTGGATCCACTGT'  
TTTGACACACTCTGCTTTGTCTCCACTTCTTAITAGGCAGGATCATAGCAGAGAGGGGGCC'  
GACCGAGCTGAGATAGCTGGAGATGTCAAGGGGACAGAGGCCATTCCAGTGAITCTAGTCTGCC'  
>AC148714_1 Macaca mulatta Major Histocompatibility Complex BAC M  
ACACCGAGAAAGGAGCAGGTGTCTCTGCTCCATGGCCACCCGTGGCAAGAAAGGCAGAACT;  
GCCTCTGCCCCITGACATCTCCAGCTTAITCTCAGCTTGGTCCCTGTCAACCCAAAAGGCCCTC'  
CCTAGGTAGGAAGTGGGAACAAAGCAGGAAGTGTCAAACAGTCAGCGGGGTGACAGTAGGA'  
AGGGGGTGTICGGGGAGCTGGTICGGTICGGGCTGIGGCCCCCTGGG
```

or upload a [plain text](#) file:  No file chosen

Use structural alignment(s)

Allow unusual symbols (Selenocysteine "U", Inosine "I", non-alphabetical characters, etc.) [Help](#)

### UPPERCASE / lowercase:

Same as input

It will look like this once the sequences are added

Now scroll down

# MAFFT version 7

Multiple alignment program for amino acid or nucleotide sequences

[Download version](#)[Mac OS X](#)[Windows](#)[Linux](#)[Source](#)[Online version](#)[Alignment](#)[mafft --add](#)[Merge](#)[Phylogeny](#)[Rough tree](#)[Merits / limitations](#)[Algorithms](#)[Tips](#)[Benchmarks](#)[Feedback](#)[Follow](#)

or upload a **plain text** file:  No file chosen

- Use structural alignment(s)  
 Allow unusual symbols (Selenocysteine "U", Inosine "i", non-alphabetical characters, etc.) [Help](#)

**UPPERCASE / lowercase:**

- Same as input  
 Amino acid → UPPERCASE / Nucleotide → lowercase

**Direction of nucleotide sequences:** [Help](#)

- Same as input  
 Adjust direction according to the first sequence (accurate enough for most cases)  
 Adjust direction according to the first sequence (only for highly divergent data; **extremely slow**)

**Output order:**

- Same as input  
 Aligned

**Notify when finished** (optional; recommended when submitting large data):

Email address:



---

**Advanced settings****Strategy:**

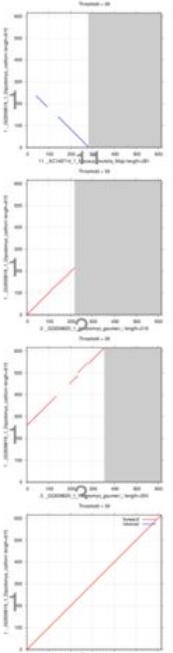
- Auto (FFT-NS-1, FFT-NS-2, FFT-NS-i or L-INS-i; depends on data size) [Updated](#)

NCBI Blast:Nucleotide S X Multiple sequence alignn X

mafft.cbrc.jp/alignment/server/spool/\_ho1702180029s248492XRMGrAeHhWCjtLNKo4wBZ.html

LAST hits (score>39) between the top sequence and the others.

[Open all plots](#)



LAST hits (score>39) between the top sequence and the others.

[Open all plots](#)

[Clustal format](#) | [Fasta format](#) | [MAFFT result](#) | [View](#) | [Tree](#) | [Refine dataset](#) | [Return to home](#)

[View](#)

[Reformat](#) to GCG, PHYLP, MSF, NEXUS, uppercase/lowercase, etc. with Readseq

[GUIDANCE2](#) computes the residue-wise confidence scores and extracts well-aligned residues.

[Refine dataset](#)

[Phylogenetic tree](#) *Visualization updated, 2016/Sep*

**MAFFT-L-INS-i Result**

CLUSTAL format alignment by MAFFT (v7.309)

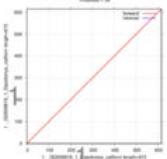
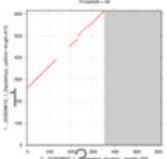
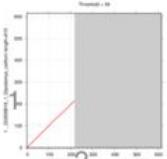
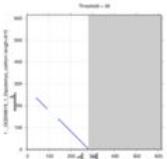
GQ506818.1	C-----
GQ506820.1	------
GQ506816.1	------
AC206407.5	------
NG_029868.1	------
AC190421.3	------
CU104674.1	------
AC242675.3	------
EU153401.1	------
GQ506820.1	------
AC148714.1	acaacgcagaaaagagcaggctgttccttagccctccatggccaccgcgtggaaagaaaggca
GQ506818.1	-gccccaggggggccacagccgcaccaggctccccaacacaccccttttgcagggtggac
GQ506820.1	-gccccaggggggccacagccgcaccaggctctccaaacacaccccttttgcagggtggac
GQ506816.1	--cccaggggggccacagccgcaccaggctccccaacacaccccttttgcagggtggac
AC206407.5	--cccaggggggccacagccgcaccaggctcccccggacacaccccttttgcagggtggac
NG_029868.1	--cccaggggggccacagccgcaccaggctcccccggacacaccccttttgcagggtggac
AC190421.3	--cccaggggggccacagccgcaccaggctcccccggacacaccccttttgcagggtggac
CU104674.1	--cccaggggggccacagccgcaccaggctcccccggacacaccccttttgcagggtggac
AC242675.3	--cccaggggggccacagccgcaccaggctcccccggacacaccccttttgcagggtggac
EU153401.1	--cccaggggggccacagccgcaccaggctcccccggacacaccccttttgcagggtggac
GQ506820.1	--cccaggggggccacagccgcaccaggctcccccggacacaccccttttgcagggtggac
AC148714.1	--cccaggggggccacagccgcaccaggctcccccggacacaccccttttgcagggtggac

Your results should like similar to this

Now scroll down

LAST hits (score > 39)  
between the top  
sequence and the  
others.

[Open all plots](#)



MAFFT-L-INS-i Result

CLUSTAL format alignment by MAFFT (v7.309)

GQ506818.1 C-----  
GQ506820.1 -----  
GQ506816.1 -----  
AC206407.5 -----  
NG\_029868.1 -----  
AC190421.3 -----  
CU104674.1 -----  
AC242675.3 -----  
EU153401.1 -----  
GQ506820.1 -----  
AC148714.1 acacccadaaaacaccaddctttcccttagccctcccatggccacccatggaaadaaaadcca

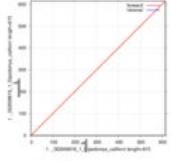
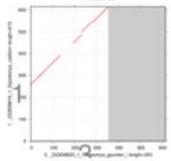
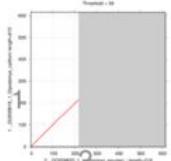
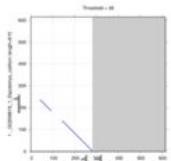
GQ506818.1 tcactggccaccctggctagctgttacacacagtctgttttgtccatattctgtct  
GQ506820.1 tcactggccaccctggctagctgttgcacattctgttttgtccatattctgtct  
GQ506816.1 ccgtggccaccctggctggctgtttgcacacttctgttttgtccacttctatct  
AC206407.5 ccactgtcaccggggctgactgtttgcacacttctgttttgtccacttctatct  
NG\_029868.1 ccactgtcaccggggctgactgtttgcacacttctgttttgtccacttctatct  
AC190421.3 ccactgtcaccggggctgactgtttgcacacttctgttttgtccacttctatct  
CU104674.1 ccactgtcaccggggctgactgtttgcacacttctgttttgtccacttctatct  
AC242675.3 ccactgtcaccggggctgactgtttgcacacttctgttttgtccacttctatct  
EU153401.1 ccactggccatctggctggctgtttgcacgttctgttttgtccacttctatct  
GQ506820.1 ---  
AC148714.1 ccactgtcaccggggatggctgtttgcacgttctgttttgtccacttctatct

This is one way  
that your results  
can be visualized

Now scroll back up

LAST hits (score>39)  
between the top  
sequence and the  
others.

[Open all plots](#)



Be careful if there are  
blue lines. By  
default, MAFFT

[Clustal format](#) | [Fasta format](#) | [MAFFT result](#) | [View](#) | [Tree](#) | [Refine dataset](#) | [Return to home](#)

[View](#)

[Reformat](#) to GCG, PHYLP, MSF, NEXUS, uppercase/lowercase, etc. with Readseq

[GUIDANCE2](#) computes the residue-wise confidence scores and extracts well-aligned residues.

[Refine dataset](#)

[Phylogenetic tree](#) *Visualization updated, 2016/Sep*

## MAFFT-L-INS-i Result

CLUSTAL format alignment by MAFFT (v7.309)

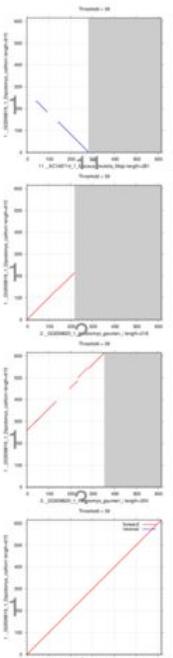
GQ506818.1	C-----
GQ506820.1	------
GQ506816.1	------
AC206407.5	------
NG_029868.1	------
AC190421.3	------
CU104674.1	------
AC242675.3	------
EU153401.1	------
GQ506820.1	------
AC148714.1	acaacgcagaaaagagcaggctgttccttagccctccccatggccaccgcgtggaaagaaaggca
GQ506818.1	-gccccaggggggccacagccgcaccaggctcccccacaccccccttttgcagggtggagc
GQ506820.1	-gccccaggggggccacagccgcaccaggctctccaaacaccccccttttgcagggtggagc
GQ506816.1	--ccccaggggggccacagccgcaccaggctcccccacaccccccttttgcagggtggagc
AC206407.5	--ccccaggggggccacagccgcaccaggctcccccacaccccccttttgcagggtggagc
NG_029868.1	--ccccaggggggccacagccgcaccaggctcccccacaccccccttttgcagggtggagc
AC190421.3	--ccccaggggggccacagccgcaccaggctcccccacaccccccttttgcagggtggagc
CU104674.1	--ccccaggggggccacagccgcaccaggctcccccacaccccccttttgcagggtggagc
AC242675.3	--ccccaggggggccacagccgcaccaggctcccccacaccccccttttgcagggtggagc
EU153401.1	--ccccaggggggccacagccgcaccaggctcccccacaccccccttttgcagggtggagc
GQ506820.1	--ccccaggggggccacagccgcaccaggctcccccacaccccccttttgcagggtggagc
AC148714.1	-----

NCBI Blast:Nucleotide S X Multiple sequence alignn X

mafft.cbrc.jp/alignment/server/spool/\_ho1702180029s248492XRMGrAeHhWCjtLNKo4wBZ.html

LAST hits (score>39) between the top sequence and the others.

[Open all plots](#)



MSAViewer *in Alpha testing, 2016/Jul*

[Start MSAViewer in this window](#) ←

[Start MSAViewer in new window](#)

MSAViewer runs on any modern browser.

Reference:  
[\*Yachdav et al. \(2016\)\*](#)

MSAViewer Home:  
<http://msa.biojs.net/>

Jalview

[Run Jalview using Java plugin](#)

Java plugin is not supported in Chrome and Edge. To run Jalview without Java plugin, [download the Desktop version](#).

Reference:  
[\*Waterhouse et al. \(2009\)\*](#)

Jalview home:  
<http://www.jalview.org/>

Be careful if there are **blue lines**. By default, MAFFT

NCBI Blast:Nucleotide S... X Multiple sequence alignn X

mafft.cbrc.jp/alignment/server/spool/\_ho1702180029s248492XRMGrAeHhWCjtLNKo4wBZ.html

LAST hits (score>39) between the top sequence and the others.

[Open all plots](#)

Clustal format | Fasta format | MAFFT result | View | Tree | Refine dataset | Return to home

MSAViewer

Increase height | Decrease height | Reload to fit to window width

Import | Sorting | Filter | Selection | Vis.elements | Color scheme | Extras | Export | Help

The scrollbar can be used to travel along the alignment

ID Label  
1 GQ508818\_1\_Dipodomys  
2 GQ508820\_1\_Heteromyss  
3 GQ508816\_1\_Anomalurus  
4 AC208407\_5\_Pongo\_abelii  
5 NG\_029883\_1\_Homo\_sapiens  
6 ACT90421\_3\_Pan\_troglodytes  
7 CUI104874\_1\_Gorilla\_bonobo  
8 AC242875\_3\_Chlorocebus\_aethiops  
9 EU153401\_1\_Felis\_catus  
10 GQ508820\_1\_Heteromyss  
11 AC148714\_1\_Macaca\_mulatta

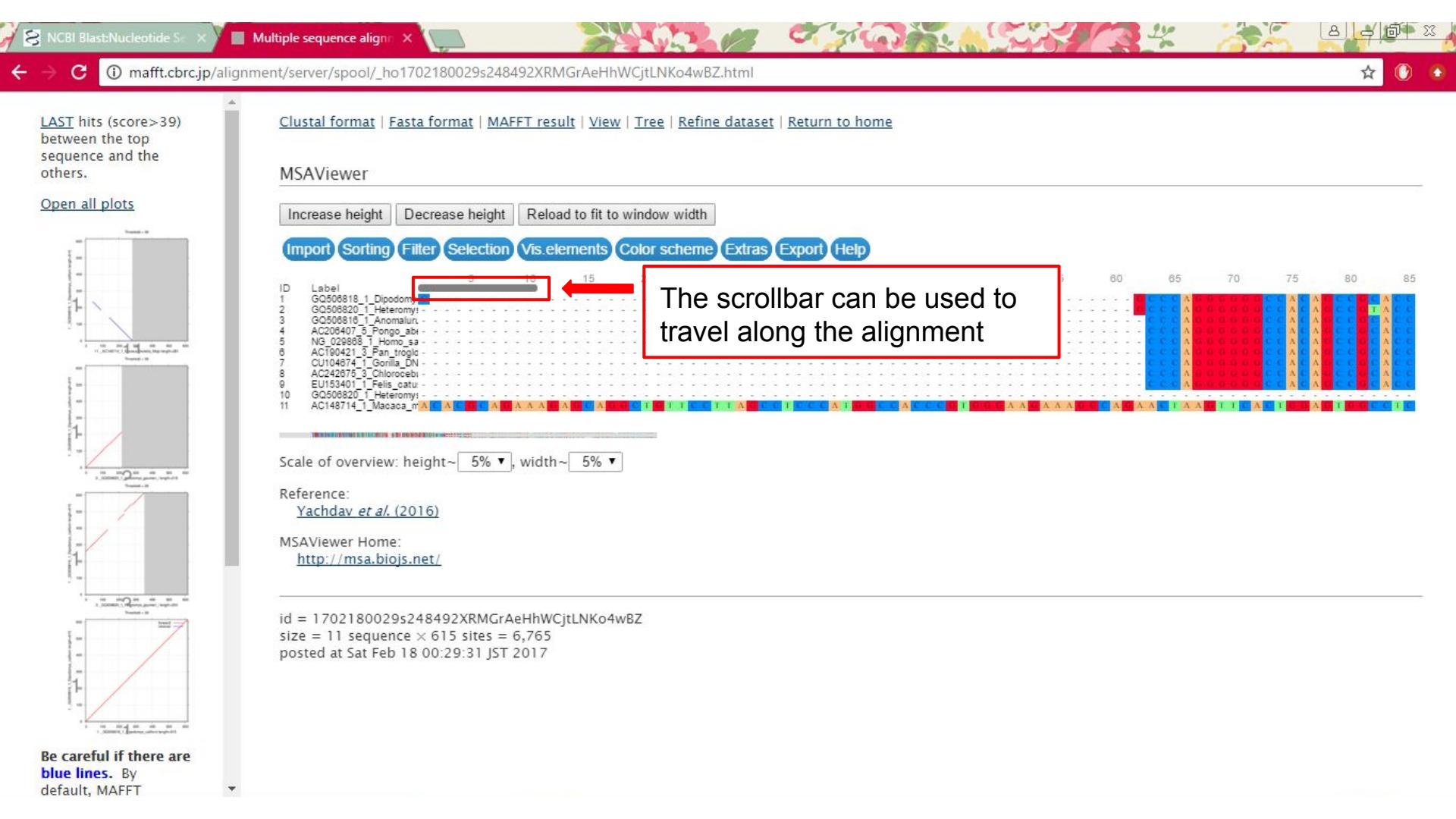
Scale of overview: height~ 5%, width~ 5%

Reference:  
[Yachdav et al. \(2016\)](#)

MSAViewer Home:  
<http://msa.biojs.net/>

id = 1702180029s248492XRMGrAeHhWCjtLNKo4wBZ  
size = 11 sequence × 615 sites = 6,765  
posted at Sat Feb 18 00:29:31 JST 2017

Be careful if there are blue lines. By default, MAFFT



NCBI Blast:Nucleotide S X Multiple sequence alignn X

mafft.cbrc.jp/alignment/server/spool/\_ho1702180029s248492XRMGrAeHhWCjtLNKo4wBZ.html

LAST hits (score>39) between the top sequence and the others.

[Open all plots](#)

MSAViewer

Clustal format | Fasta format | MAFFT result | View | **Tree** | Refine dataset | Return to home

Increase height | Decrease height | Reload to fit to window width

Import | Sorting | Filter | Selection | Vis.elements | Color scheme | Extras | Export | Help

ID	Label	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
1	GQ508818_1_Dipodomys																	
2	GQ508820_1_Heteromy																	
3	GQ508816_1_Anomalur																	
4	AC208407_5_Pongo_abr																	
5	NG_029863_1_Homo_sa																	
6	ACT90421_3_Pan_trogl																	
7	CU104874_1_Gorilla_DN																	
8	AC242875_3_Chlorocebi																	
9	EU153401_1_Felis_catu																	
10	GQ508820_1_Heteromy																	
11	AC148714_1_Macaca_m	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A

Scale of overview: height~ 5% ▾, width~ 5% ▾

Reference:  
[Yachdav et al. \(2016\)](#)

MSAViewer Home:  
<http://msa.biojs.net/>

id = 1702180029s248492XRMGrAeHhWCjtLNKo4wBZ  
size = 11 sequence × 615 sites = 6,765  
posted at Sat Feb 18 00:29:31 JST 2017

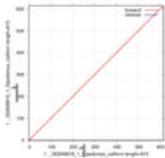
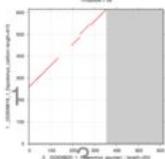
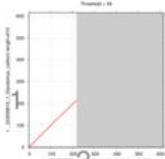
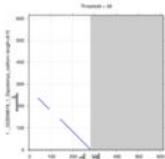
Be careful if there are blue lines. By default, MAFFT

A red arrow points to the "Tree" button in the navigation bar.



LAST hits (score>39)  
between the top  
sequence and the  
others.

[Open all plots](#)



Be careful if there are  
**blue lines**. By  
default, MAFFT

[Clustal format](#) | [Fasta format](#) | [MAFFT result](#) | [View](#) | [Tree](#) | [Refine dataset](#) | [Return to home](#)

### NJ or UPGMA tree ( $\beta$ )

11 sequences,

708 total sites,

0 gap-free sites,

0 conserved sites

**Go!** Reset

The number of gap-free sites, 0, is too small to produce a reliable tree. Possible workarounds:

- Split the groups by [clustering unaligned sequences](#).
- [Delete fr](#) unaligned sequences with MaxAlign or manually.
- A rough tree, used for constructing the MSA, is [available here](#).
- Try 'Average linkage' tree selected in the panel below, which is also a rough tree.

**2. Go!**

### Settings

**1. Make sure this option is selected**

#### Method:

- NJ — Conserved sites (0 bases)
- NJ — All gap-free sites (0 bases)
- Average linkage (UPGMA) — alignment scores (for up to 50,000 sequences)
- Minimum linkage — alignment scores (for up to 50,000 sequences)
- Memory-saving tree — alignment scores (for larger data)

**1. Make sure this option is selected**

#### Substitution model (valid when NJ is selected):

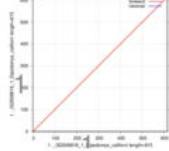
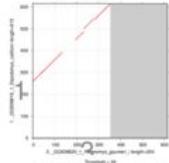
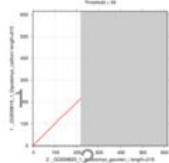
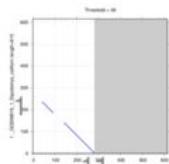
- Jukes-Cantor

#### Bootstrap (valid for NJ):

- On

LAST hits (score>39)  
between the top  
sequence and the  
others.

[Open all plots](#)



Be careful if there are  
blue lines. By  
default, MAFFT

[Clustal format](#) | [Fasta format](#) | [MAFFT result](#) | [View](#) | [Tree](#) | [Refine dataset](#) | [Return to home](#)

## Result (Phylo.io 1.0.0) *Updated, 2016/Sep*

Phylo.io runs on any modern browser.

[View tree on Phylo.io](#)



[Refine dataset on tree](#) | *Alpha testing, 2016/Aug*

## Result (Archaeopteryx with Java plugin)

Uses [Java plugin](#); no longer available on Chrome, Edge, etc.

[View tree on Archaeopteryx](#) (Signed; Forester 1.038)

[View tree on Archaeopteryx](#) (Unsigned; Forester 1.027; Try this if the signed version above does not work)

[Refine dataset on tree](#)

## Result (Archaeopteryx without Java plugin)

0. [Download and open forester.jar](#), which starts [Archaeopteryx](#) as a standalone Java program.

1. On the Archaeopteryx window, select **File** → **Read Tree from URL/Webservice...** to open a text field.

2. Copy and paste the following URL into the text field you opened in step 1.



[http://mafft.cbrc.jp/alignment/server/spool/\\_nj1702180038s24827435.on.nh.xml](http://mafft.cbrc.jp/alignment/server/spool/_nj1702180038s24827435.on.nh.xml)

## Result (for external tree viewers)

- Tree file ([Newick](#), [PhyloXML](#))
- Tree file without terminal node number ([Newick](#), [PhyloXML](#))

# Phylo.io

Version: 1.0.k

View      Compare

## Tree:

```
(((((1_GQ506818_1_Dipodomys_californicus_isolate_25-Dca_SINE_B1F_sequence_and_unknown_gene:0.14262,2_GQ506820_1_Heteromys_gaumeri_isolate_25-Hga_SINE_B1F_sequence_and_unknown_gene
```

Render ►

Settings

Share

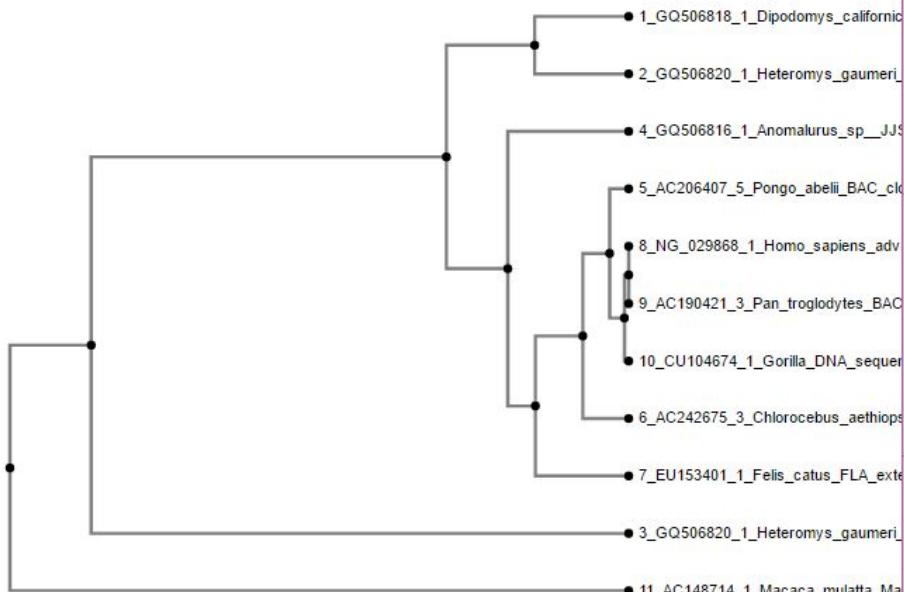
Help

SVG

Zoom:



Let's make things a little bigger



# Phylo.io

Version: 1.0.k

View

Compare

## Tree:

```
(((((1_GQ506818_1_Dipodomys_californicus_isolate_25-Dca_SINE_B1F_sequence_and_unknown_gene:0.14262,2_GQ506820_1_Heteromys_gaumeri_isolate_25-Hga_SINE_B1F_sequence_and_unknown_gene
```

Render ►

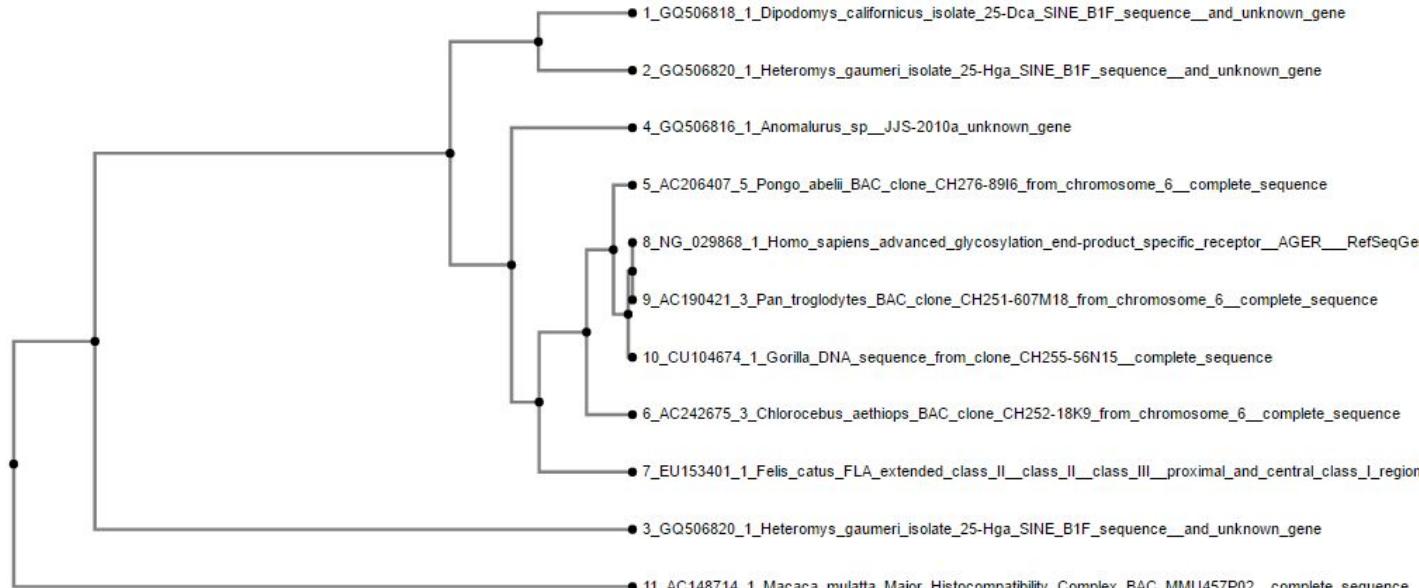
Settings

Share

Help

Download SVG

You've built a tree! Well done!



1.20