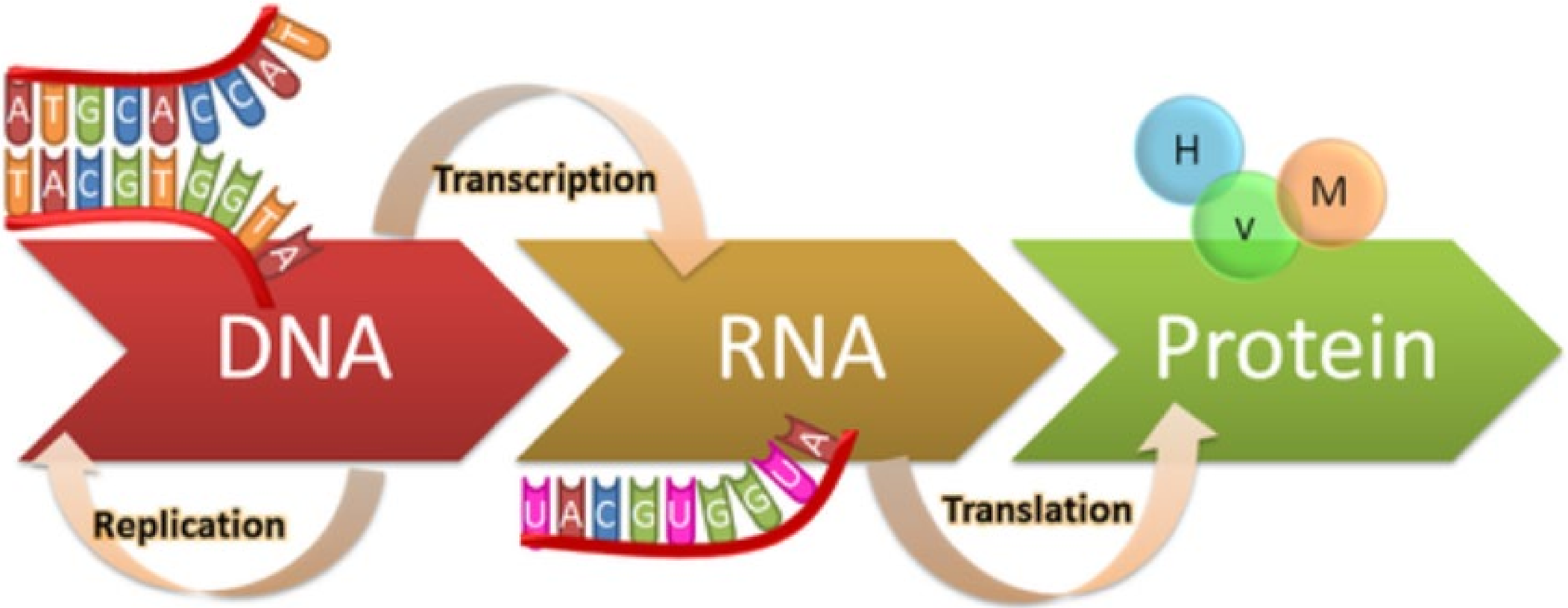
Water bioinformatics

NOURISH THE FUTURE

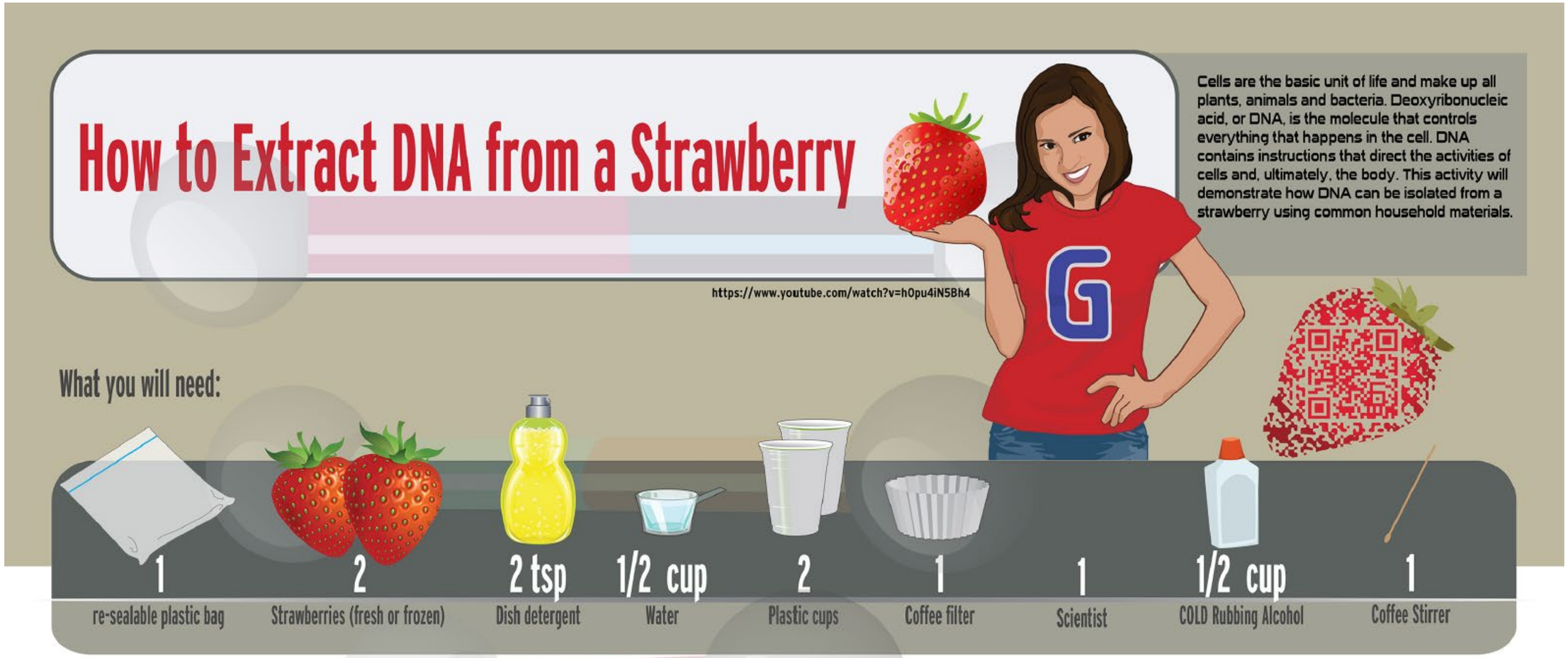
Tomorrow’s science is looking for leaders

Central dogma of molecular biology



genius.com/Biology-genius-the-central-dogma-annotated

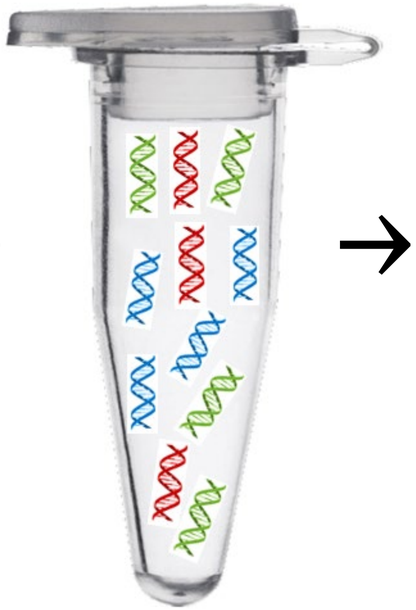
Extractions help students visualize DNA



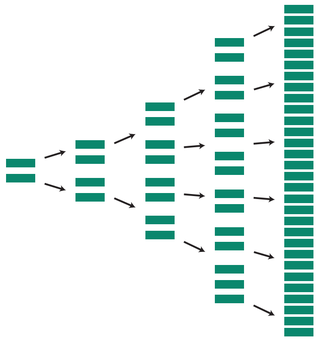
genome.gov/Pages/Education/Modules/StrawberryExtractionInstructions.pdf

The process: Sample to DNA analysis

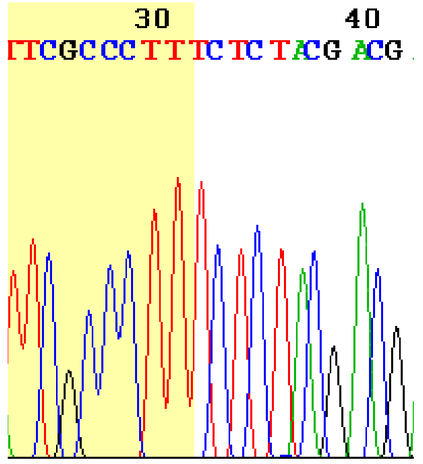
Sample collection



DNA extraction



PCR

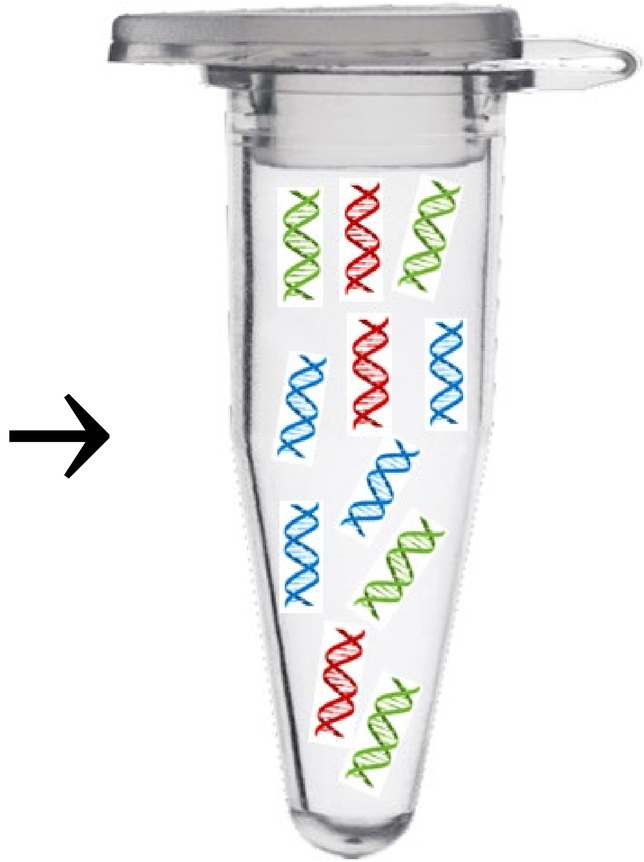


Sequencing

Bioinformatics

The process: Sample to DNA analysis

Soil sample



DNA extraction

Sample collection

↓

DNA extraction

↓

PCR

↓

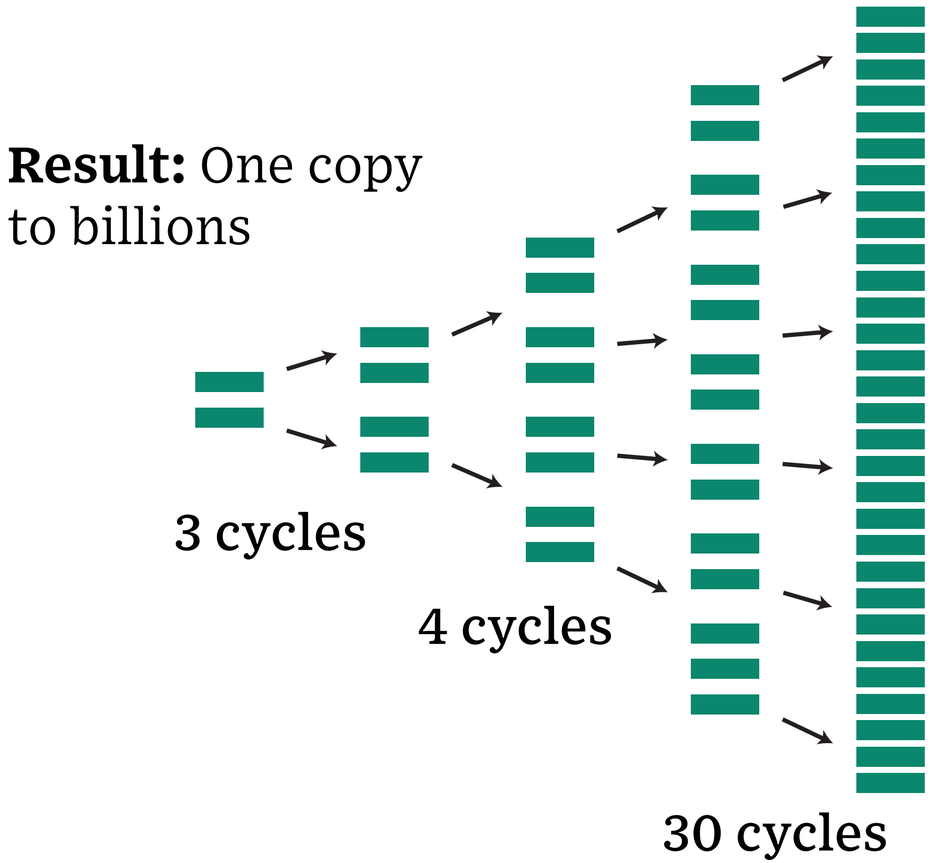
Sequencing

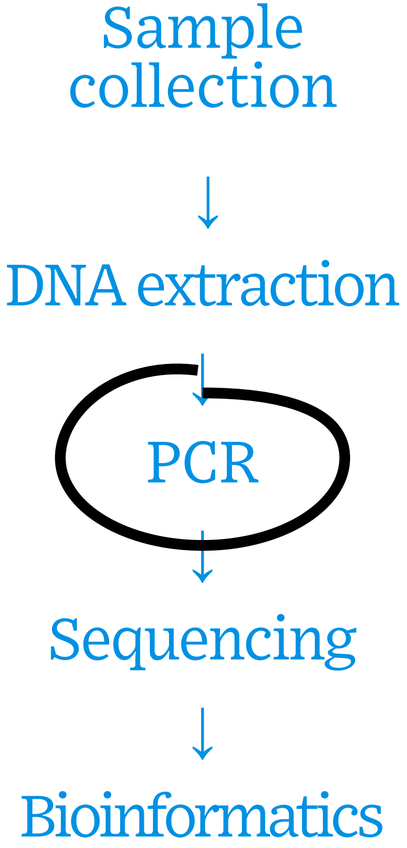
↓

Bioinformatics

The process: Sample to DNA analysis

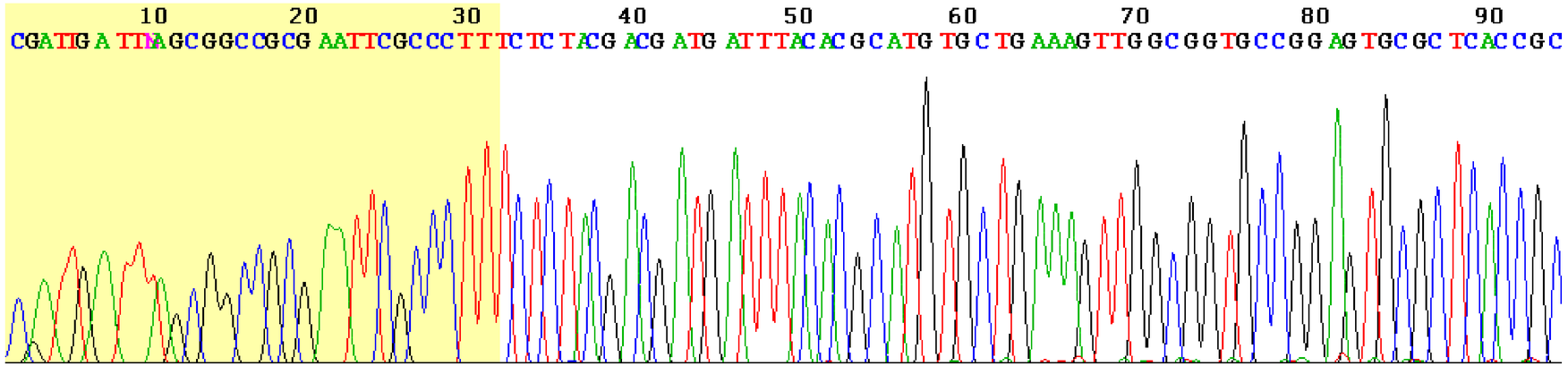
Polymerase chain reaction (PCR): Cycling process to artificially make copies (amplify) specific pieces of DNA in the genome





The process: Sample to DNA analysis

DNA sequencing: Amplified PCR products are digitized into their nucleotide sequence



Sample collection

↓

DNA extraction

↓

PCR

↓

Sequencing

↓

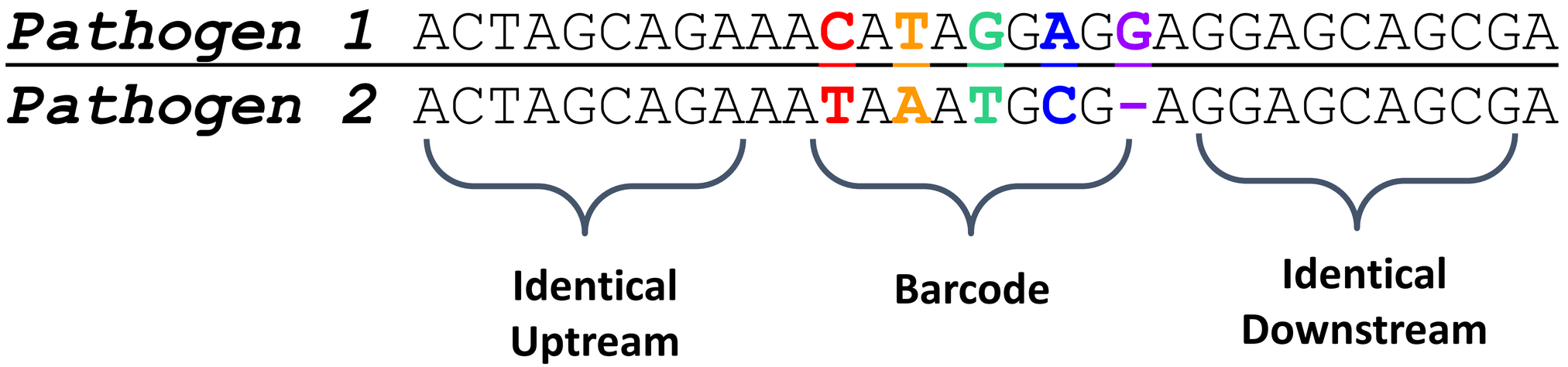
Bioinformatics

Barcodes

Linear lines that optical scanners (barcode readers) can determine a product based on lines of various widths.

Barcodes

Regions of DNA that differ between species.

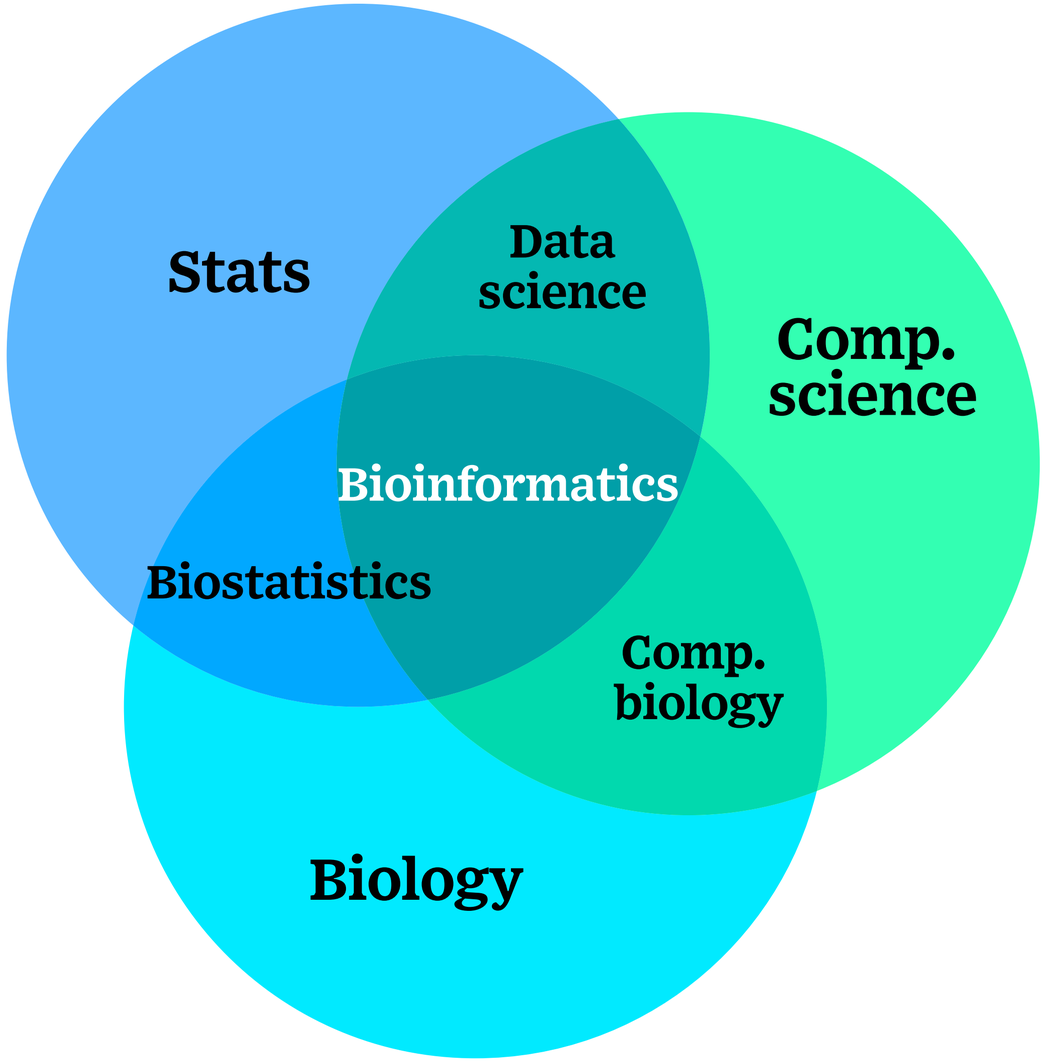


Barcodes

There are multiple barcoding regions or genes, which are used for different groups.

Organism	**Barcode**	**Function**	**Location**
Plants	RuBisCo (rbcL)	Carbon fixation from CO_2	Chloroplast
Plants	Maturase K (matK)	Protein that splices introns	Chloroplast
Plants and fungi	Ribosomal RNA including ITS1, ITS2	Essential for protein synthesis	Nucleus
Animals	Cytochrome C Oxidase Subunit I (COI)	Key enzyme in the electron transport chain	Mitochondria

The process: Sample to DNA analysis



Sample collection

↓

DNA extraction

↓

PCR

↓

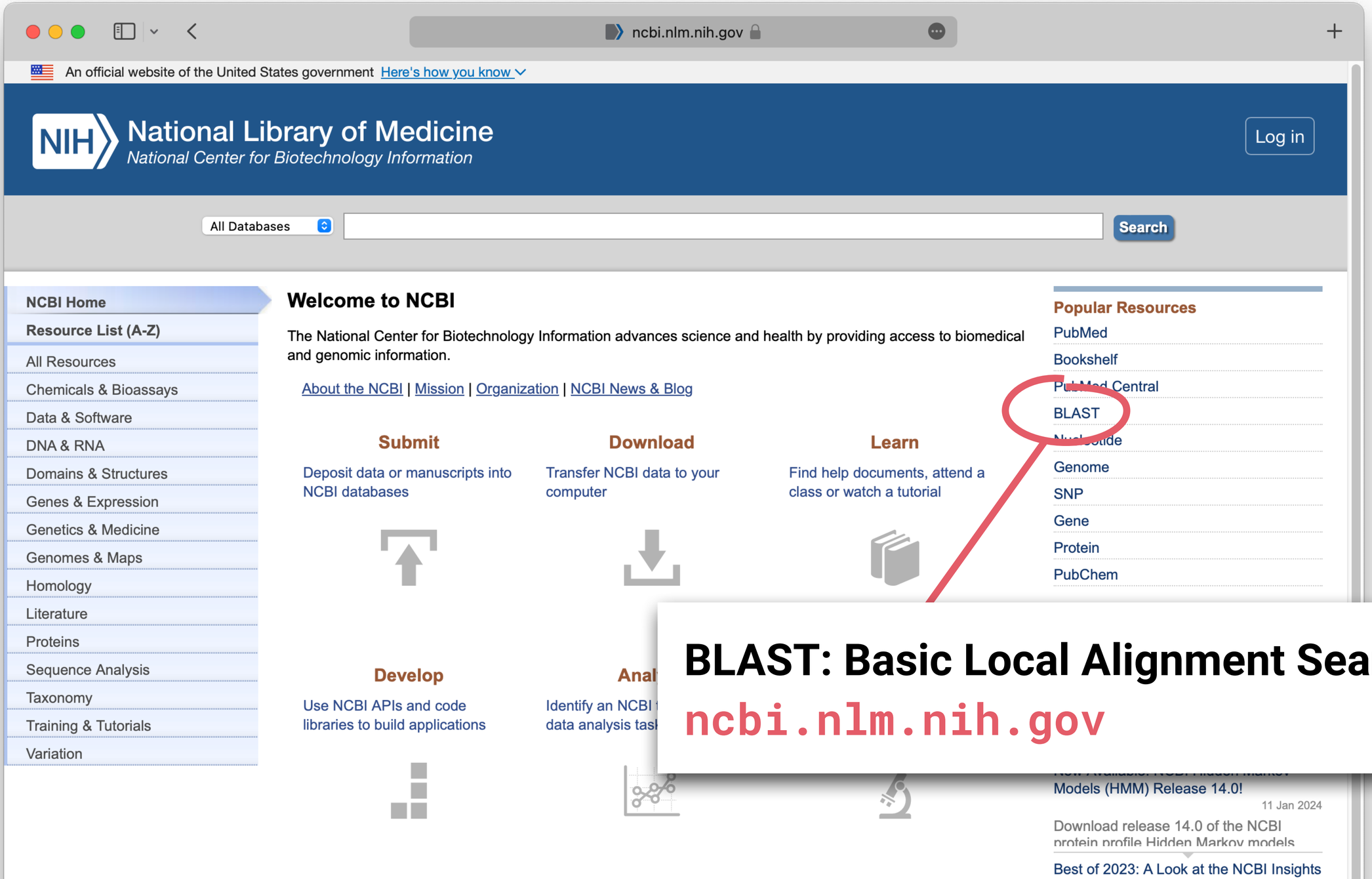
Sequencing

↓

Bioinformatics

Scenario

Ohioland Water Institute and Buckeye Genetic Laboratories have an ongoing water-monitoring project in Lake County. Last week, a summer field technician collected 100mL of water from Lake Erie for eDNA analysis. Buckeye Genetic Laboratories performed eDNA sequencing and released the report containing the top 10 sequences recovered from the water sample. Your job as a bioinformatician is to help identify the source of DNA by searching for similar sequences within the NCBI database using BLAST.


ncbi.nlm.nih.gov
An official website of the United States government Here's how you know
National Library of Medicine
National Center for Biotechnology Information
Log in
All Databases
Search
NCBI Home
Resource List (A-Z)
All Resources
Chemicals & Bioassays
Data & Software
DNA & RNA
Domains & Structures
Genes & Expression
Genetics & Medicine
Genomes & Maps
Homology
Literature
Proteins
Sequence Analysis
Taxonomy
Training & 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
Popular Resources
PubMed
Bookshelf
BLAST
Genome
SNP
Gene
Protein
PubChem
Models (HMM) Release 14.0!
11 Jan 2024
Download release 14.0 of the NCBI
Best of 2023: A Look at the NCBI Insights
BLAST: Basic Local Alignment Search
ncbi.nlm.nih.gov

blast.ncbi.nlm.nih.gov

An official website of the United States government Here's how you know

National Library of Medicine
National Center for Biotechnology Information

Log in

BLAST ®

Home Recent Results Saved Strategies Help

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

NEWS

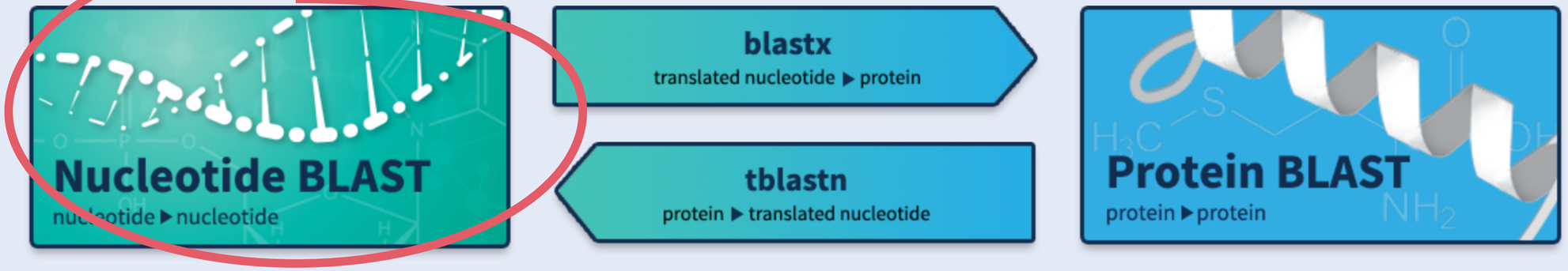
BLAST+ 2.15.0 is here!

We have included two exciting new features in the latest BLAST+ release

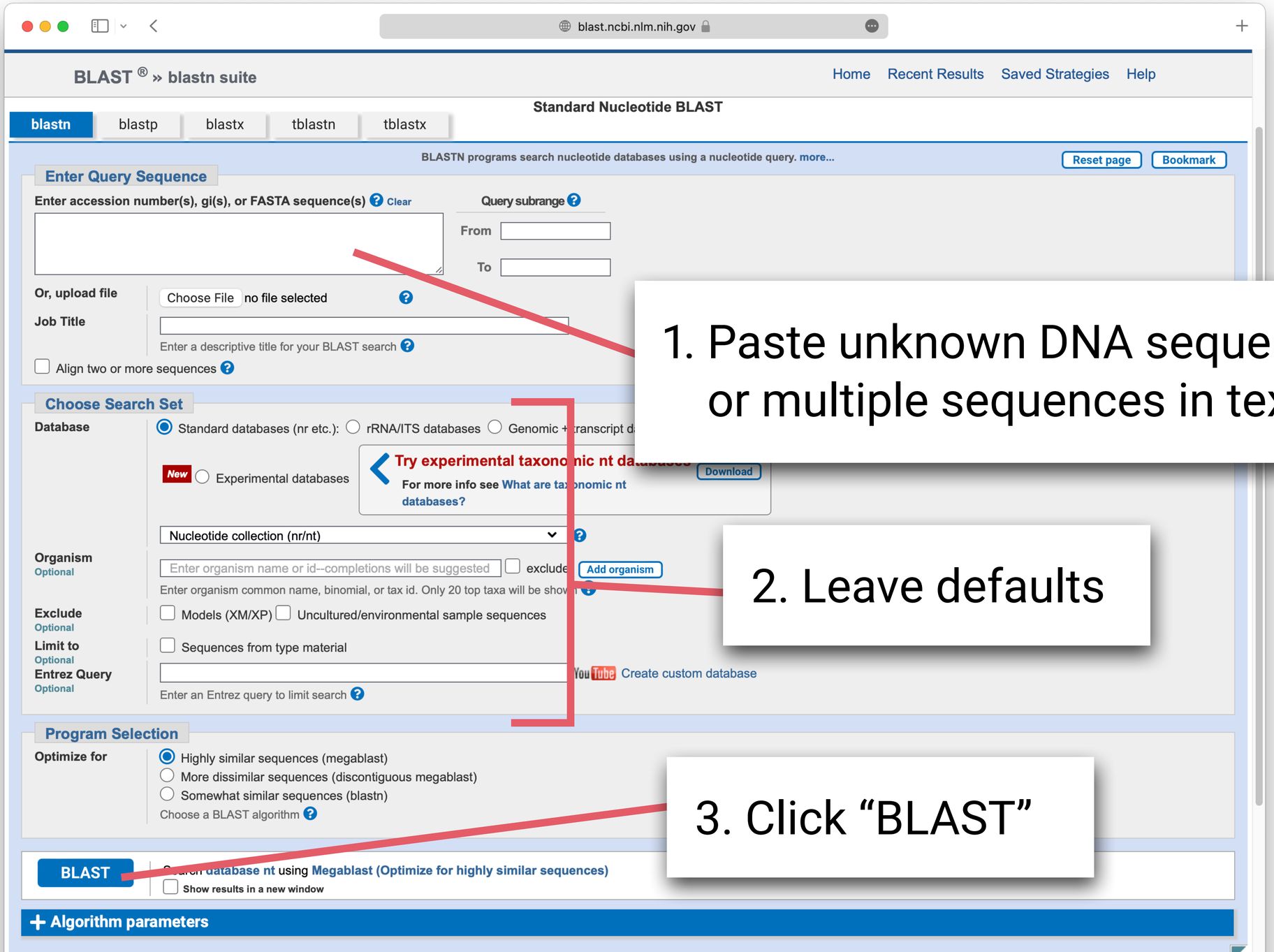
Tue, 28 Nov 2023

More BLAST news...

Web BLAST



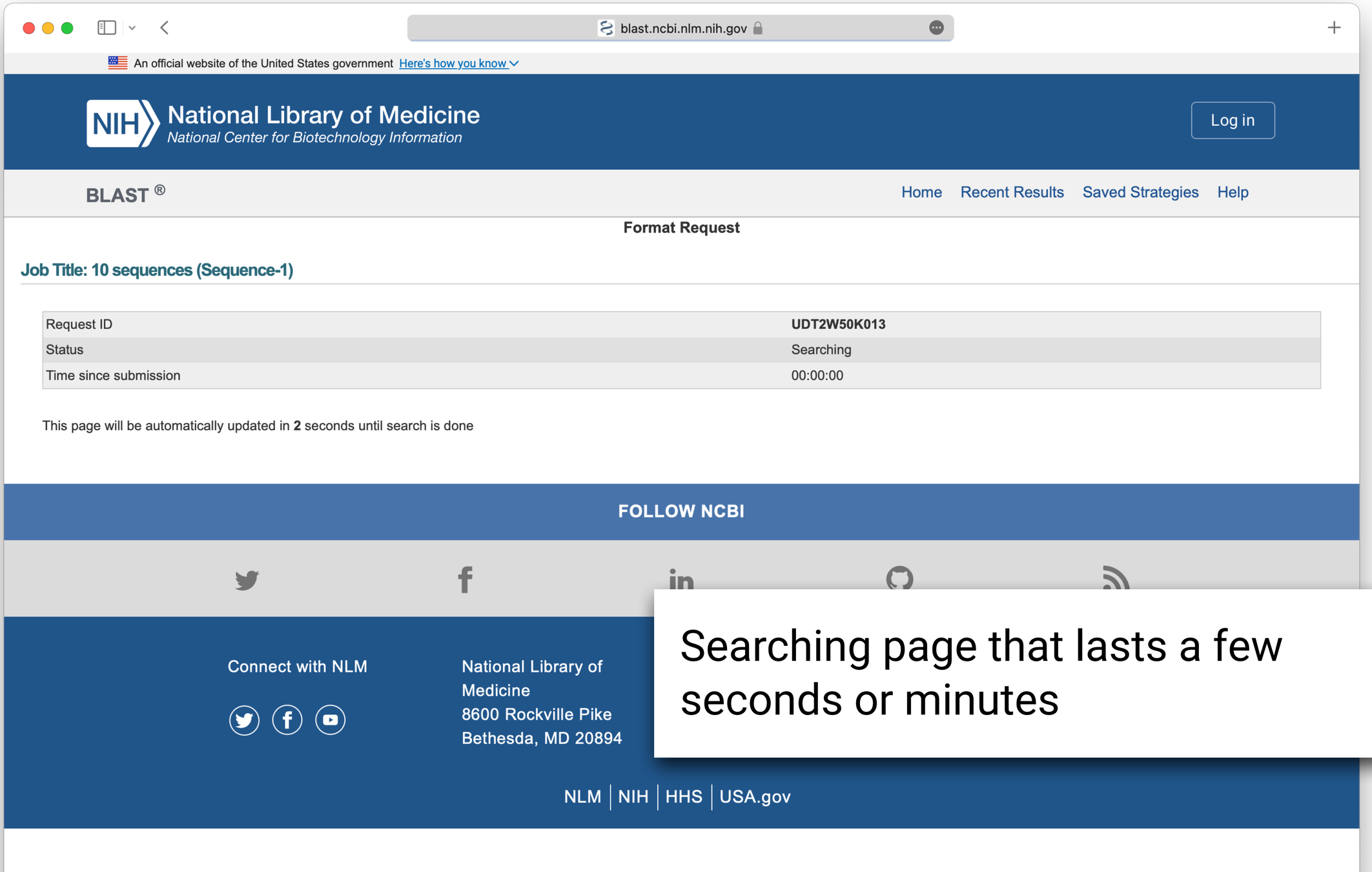
BLAST Genomes



1. Paste unknown DNA sequence or multiple sequences in text box

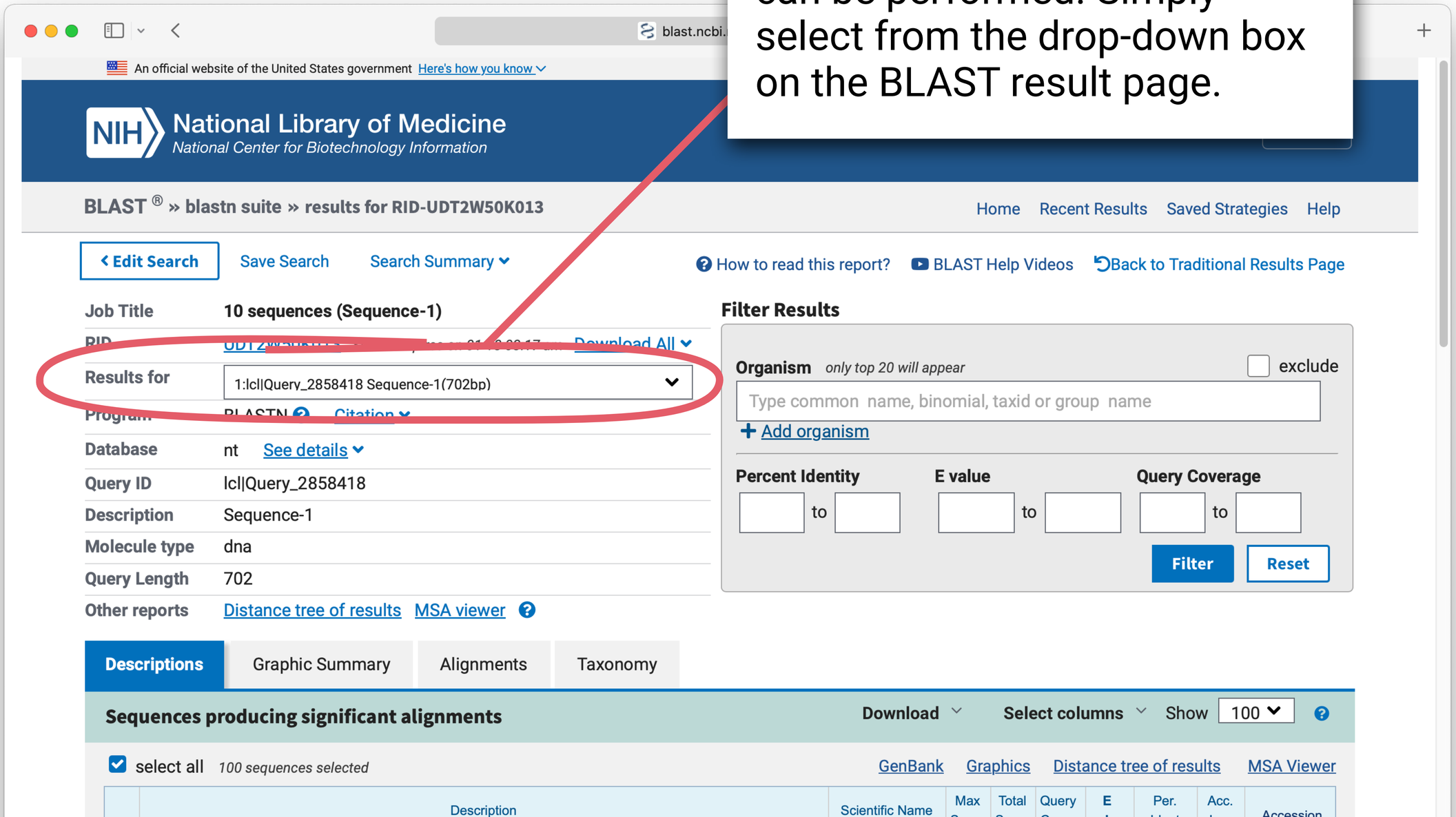
2. Leave defaults

3. Click “BLAST”


blast.ncbi.nlm.nih.gov
An official website of the United States government Here's how you know
NIH National Library of Medicine
National Center for Biotechnology Information
Log in
BLAST ®
Home Recent Results Saved Strategies Help
Format Request
Job Title: 10 sequences (Sequence-1)
Request ID UDT2W50K013
Status Searching
Time since submission 00:00:00
This page will be automatically updated in 2 seconds until search is done
FOLLOW NCBI
Connect with NLM
National Library of Medicine
8600 Rockville Pike
Bethesda, MD 20894
NLM | NIH | HHS | USA.gov
Searching page that lasts a few seconds or minutes

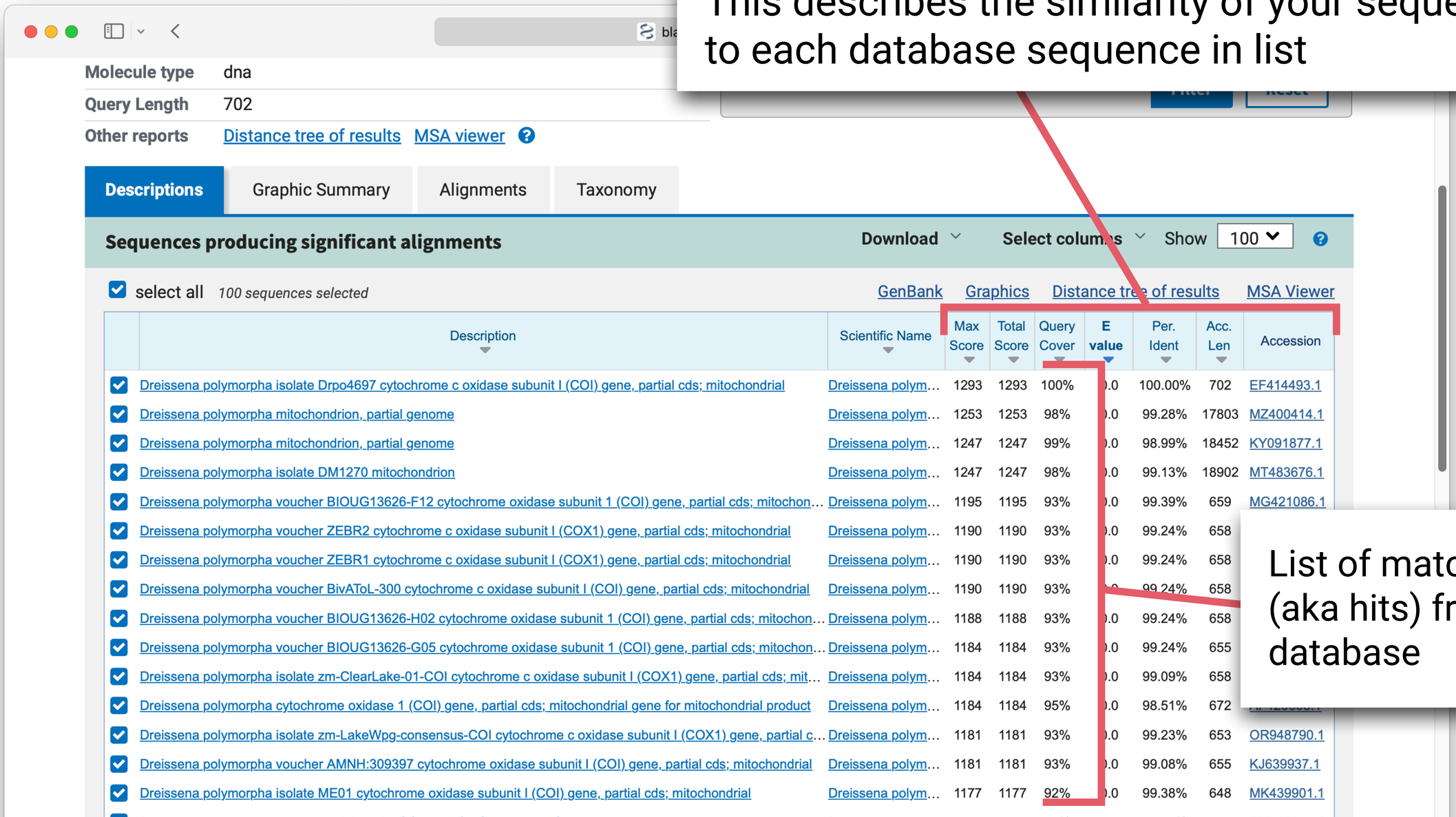
BLAST results

Multiple sequence searches can be performed. Simply select from the drop-down box on the BLAST result page.



BLAST results

This describes the similarity of your sequence to each database sequence in list



Molecule type dna
Query Length 702
Other reports Distance tree of results MSA viewer

Descriptions | Graphic Summary | Alignments | Taxonomy

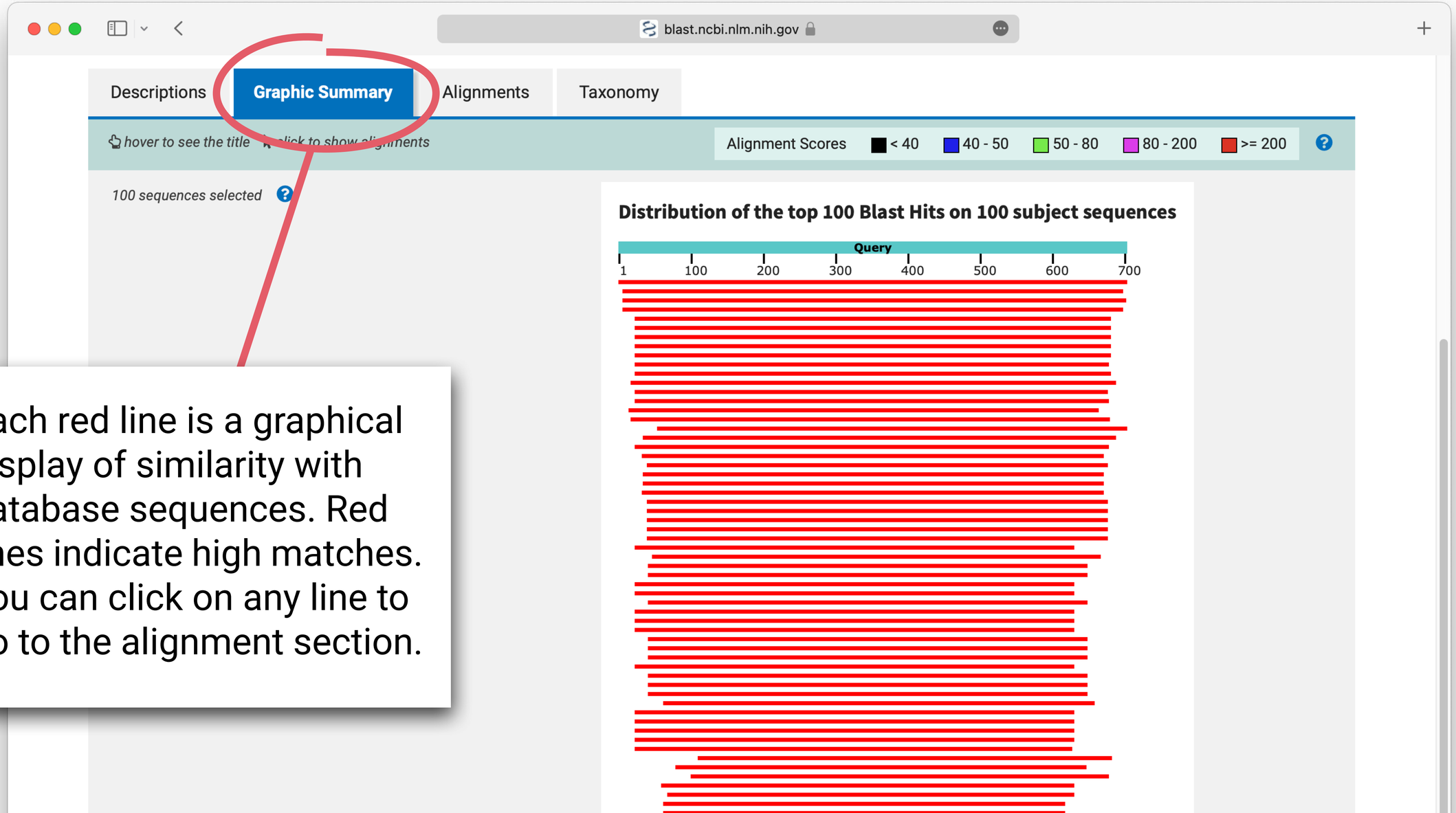
Sequences producing significant alignments Download Select columns Show 100

select all *100 sequences selected* GenBank Graphics Distance tree of results MSA Viewer

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Dreissena polymorpha isolate Drpo4697 cytochrome c oxidase subunit I (COI) gene, partial cds; mitochondrial	Dreissena polym…	1293	1293	100%	0.0	100.00%	702	EF414493.1
Dreissena polymorpha mitochondrion, partial genome	Dreissena polym…	1253	1253	98%	0.0	99.28%	17803	MZ400414.1
Dreissena polymorpha mitochondrion, partial genome	Dreissena polym…	1247	1247	99%	0.0	98.99%	18452	KY091877.1
Dreissena polymorpha isolate DM1270 mitochondrion	Dreissena polym…	1247	1247	98%	0.0	99.13%	18902	MT483676.1
Dreissena polymorpha voucher BIOUG13626-F12 cytochrome oxidase subunit 1 (COI) gene, partial cds; mitochon…	Dreissena polym…	1195	1195	93%	0.0	99.39%	659	MG421086.1
Dreissena polymorpha voucher ZEBR2 cytochrome c oxidase subunit I (COX1) gene, partial cds; mitochondrial	Dreissena polym…	1190	1190	93%	0.0	99.24%	658	
Dreissena polymorpha voucher ZEBR1 cytochrome c oxidase subunit I (COX1) gene, partial cds; mitochondrial	Dreissena polym…	1190	1190	93%	0.0	99.24%	658	
Dreissena polymorpha voucher BivAToL-300 cytochrome c oxidase subunit I (COI) gene, partial cds; mitochondrial	Dreissena polym…	1190	1190	93%	0.0	99.24%	658	
Dreissena polymorpha voucher BIOUG13626-H02 cytochrome oxidase subunit 1 (COI) gene, partial cds; mitochon…	Dreissena polym…	1188	1188	93%	0.0	99.24%	658	
Dreissena polymorpha voucher BIOUG13626-G05 cytochrome oxidase subunit 1 (COI) gene, partial cds; mitochon…	Dreissena polym…	1184	1184	93%	0.0	99.24%	655	
Dreissena polymorpha isolate zm-ClearLake-01-COI cytochrome c oxidase subunit I (COX1) gene, partial cds; mit…	Dreissena polym…	1184	1184	93%	0.0	99.09%	658	
Dreissena polymorpha cytochrome oxidase 1 (COI) gene, partial cds; mitochondrial gene for mitochondrial product	Dreissena polym…	1184	1184	95%	0.0	98.51%	672	
Dreissena polymorpha isolate zm-LakeWpg-consensus-COI cytochrome c oxidase subunit I (COX1) gene, partial c…	Dreissena polym…	1181	1181	93%	0.0	99.23%	653	OR948790.1
Dreissena polymorpha voucher AMNH:309397 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial	Dreissena polym…	1181	1181	93%	0.0	99.08%	655	KJ639937.1
Dreissena polymorpha isolate ME01 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial	Dreissena polym…	1177	1177	92%	0.0	99.38%	648	MK439901.1

List of matches (aka hits) from database

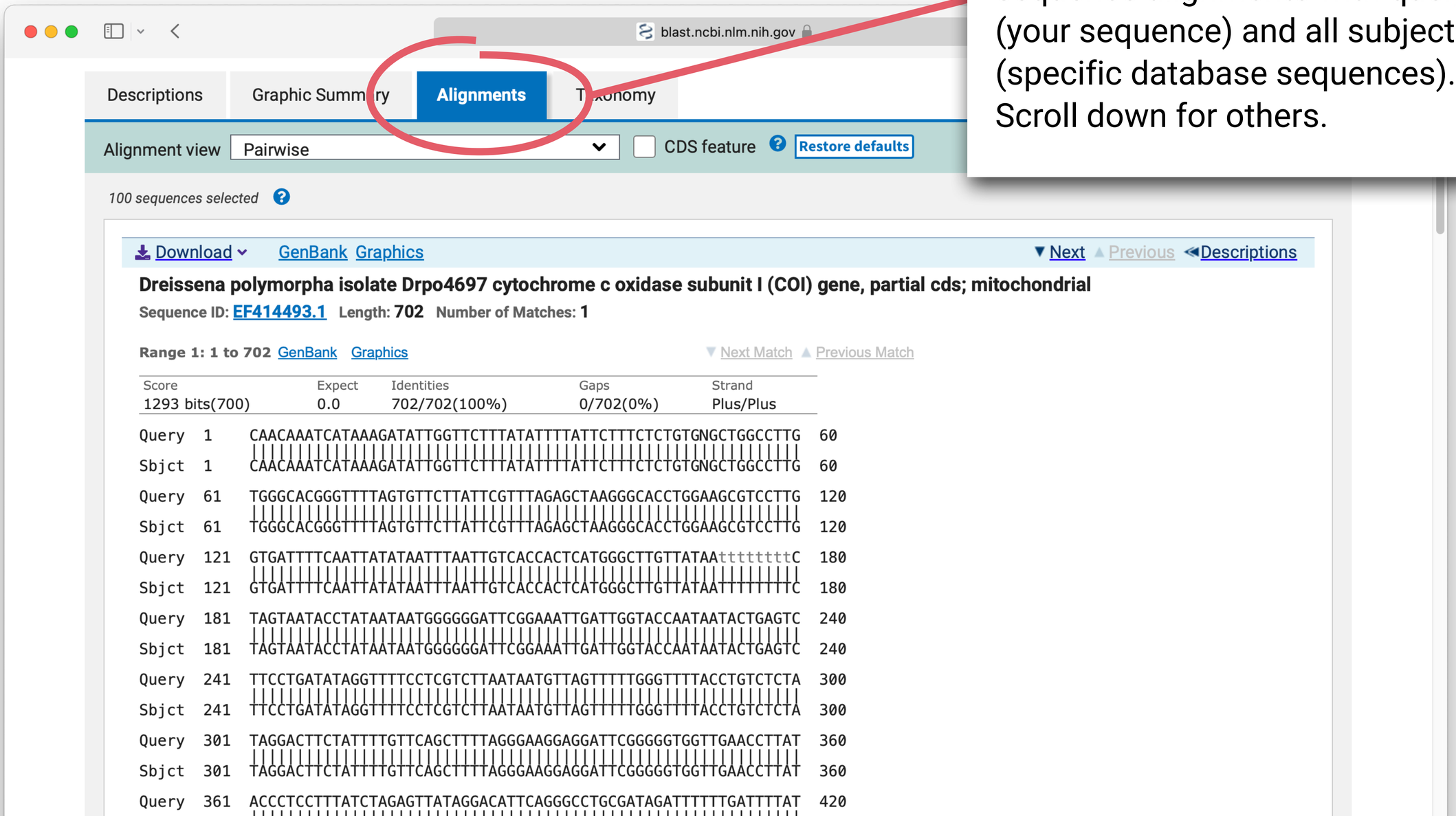
BLAST results: graphical summary



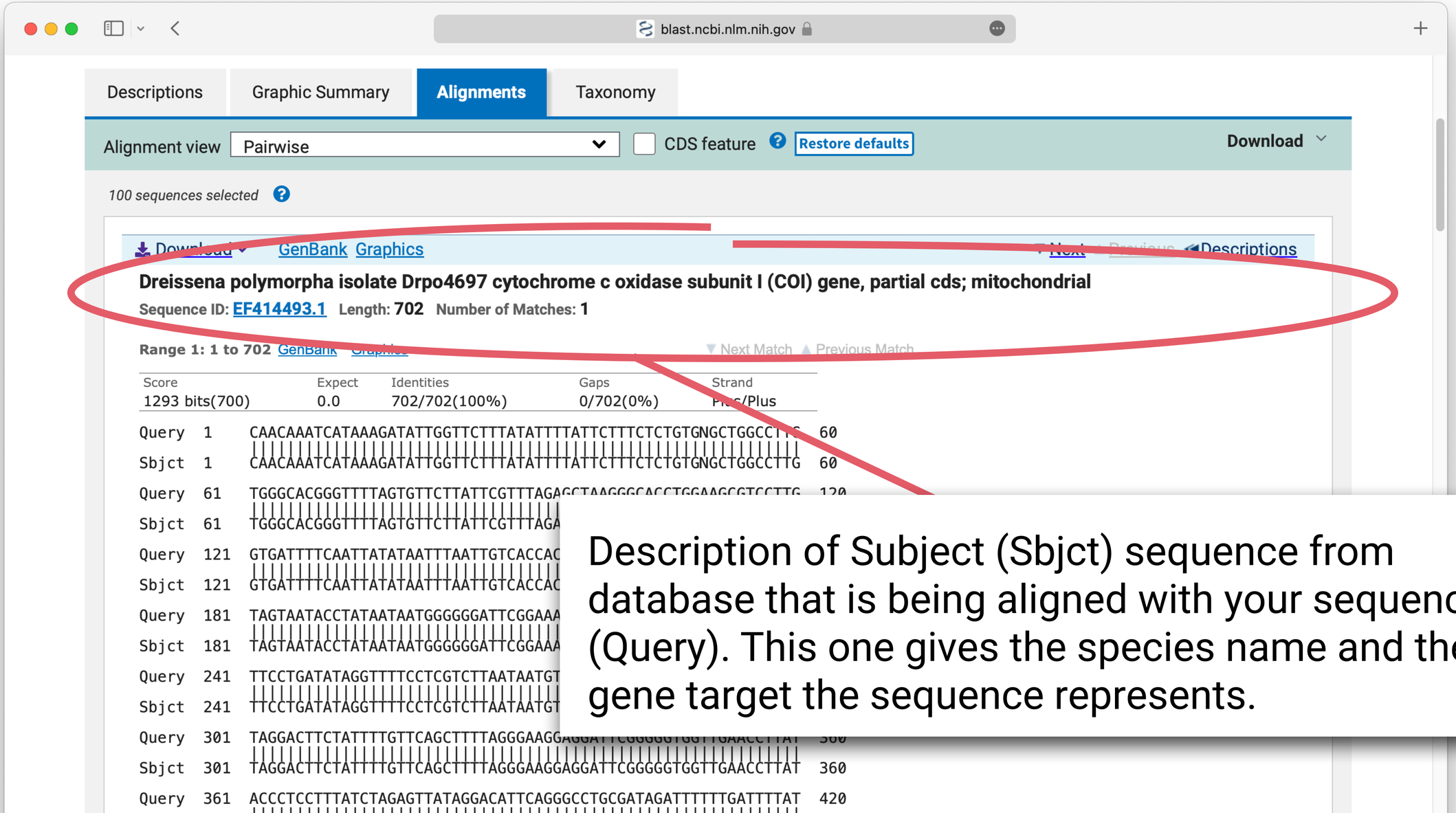
Each red line is a graphical display of similarity with database sequences. Red lines indicate high matches. You can click on any line to go to the alignment section.

BLAST results: alignment

Alignment tab displays individual sequence alignments with query (your sequence) and all subjects (specific database sequences). Scroll down for others.



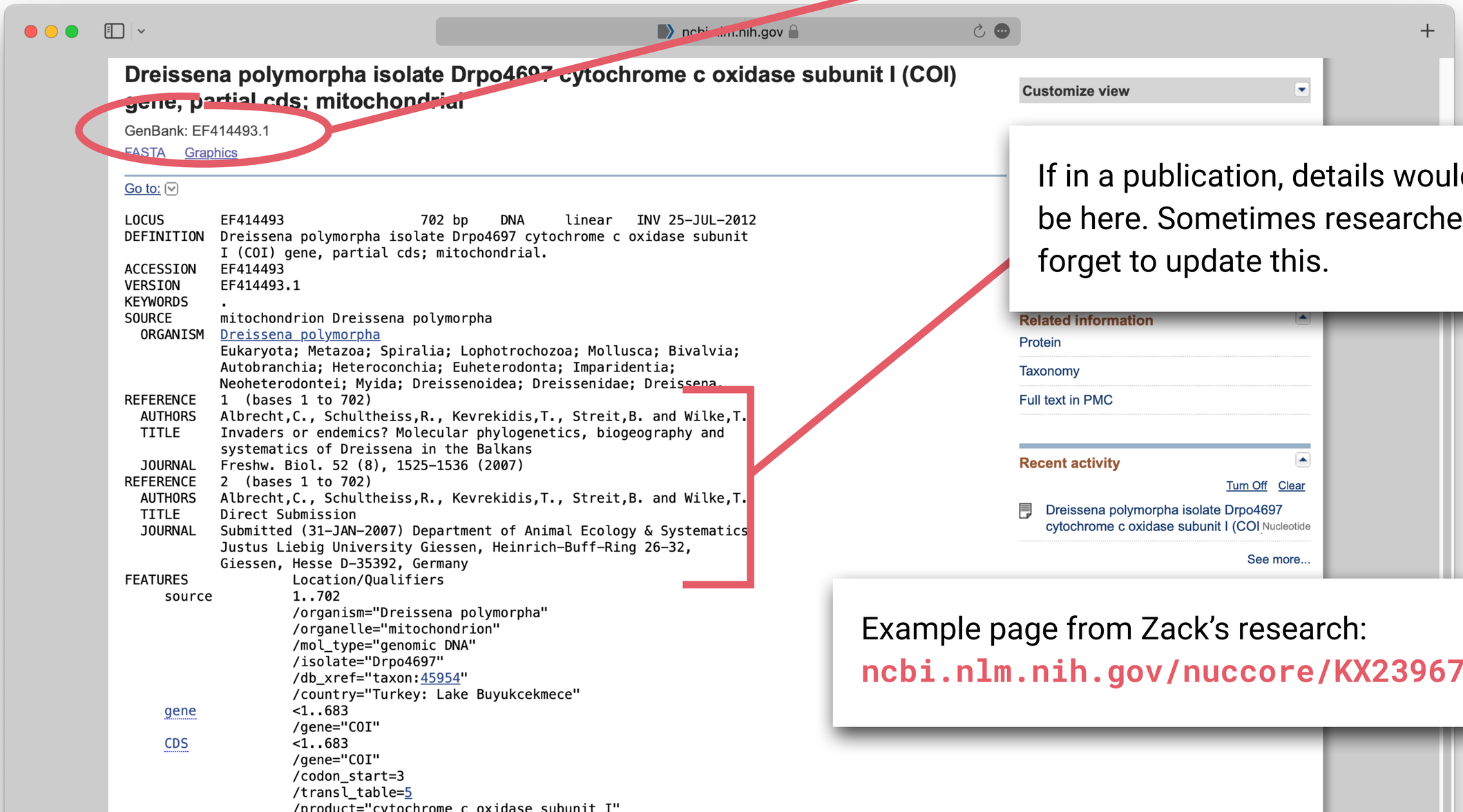
Sequence alignment (100% match)



Description of Subject (Sbjct) sequence from database that is being aligned with your sequence (Query). This one gives the species name and the gene target the sequence represents.

Subject sequence page

Sequence ID number



Dreissena polymorpha isolate Drpo4697 cytochrome c oxidase subunit I (COI) gene, partial cds; mitochondrial

GenBank: EF414493.1

FASTA Graphics

Go to:

```
LOCUS       EF414493                 702 bp    DNA     linear   INV 25-JUL-2012
DEFINITION  Dreissena polymorpha isolate Drpo4697 cytochrome c oxidase subunit
            I (COI) gene, partial cds; mitochondrial.
ACCESSION   EF414493
VERSION     EF414493.1
KEYWORDS    .
SOURCE      mitochondrion Dreissena polymorpha
  ORGANISM  Dreissena polymorpha
            Eukaryota; Metazoa; Spiralia; Lophotrochozoa; Mollusca; Bivalvia;
            Autobranchia; Heteroconchia; Euheterodonta; Imparidentia;
            Neoheterodontei; Myida; Dreissenoidea; Dreissenidae; Dreissena.
REFERENCE   1  (bases 1 to 702)
  AUTHORS   Albrecht,C., Schultheiss,R., Kevrekidis,T., Streit,B. and Wilke,T.
  TITLE     Invaders or endemics? Molecular phylogenetics, biogeography and
            systematics of Dreissena in the Balkans
  JOURNAL   Freshw. Biol. 52 (8), 1525-1536 (2007)
REFERENCE   2  (bases 1 to 702)
  AUTHORS   Albrecht,C., Schultheiss,R., Kevrekidis,T., Streit,B. and Wilke,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2007) Department of Animal Ecology & Systematics
            Justus Liebig University Giessen, Heinrich-Buff-Ring 26-32,
            Giessen, Hesse D-35392, Germany
FEATURES             Location/Qualifiers
     source          1..702
                     /organism="Dreissena polymorpha"
                     /organelle="mitochondrion"
                     /mol_type="genomic DNA"
                     /isolate="Drpo4697"
                     /db_xref="taxon:45954"
                     /country="Turkey: Lake Buyukcekmece"
     gene            <1..683
                     /gene="COI"
     CDS             <1..683
                     /gene="COI"
                     /codon_start=3
                     /transl_table=5
                     /product="cytochrome c oxidase subunit I"
```

Customize view

Related information

Protein

Taxonomy

Full text in PMC

Recent activity

Turn Off Clear

Dreissena polymorpha isolate Drpo4697 cytochrome c oxidase subunit I (COI Nucleotide

See more...

If in a publication, details would be here. Sometimes researchers forget to update this.

Example page from Zack's research:
ncbi.nlm.nih.gov/nuccore/KX239675.1