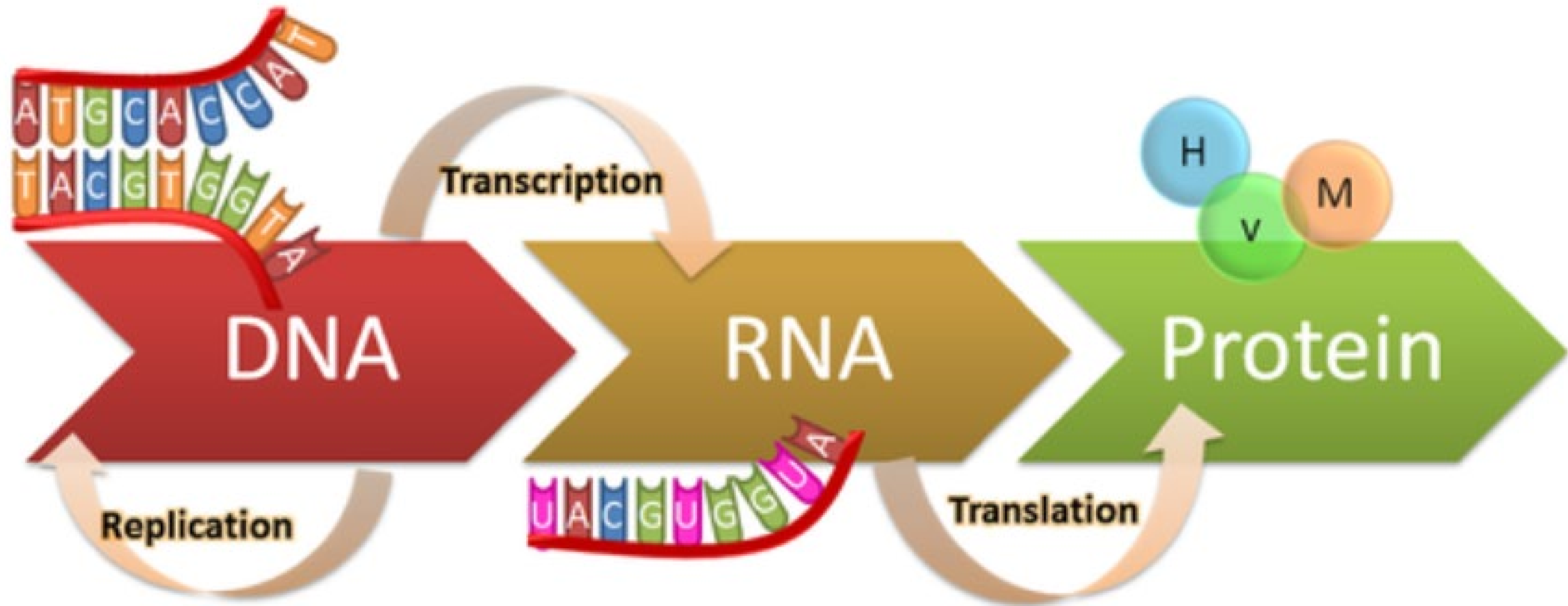


Water bioinformatics

NOURISH THE **FUTURE**

Tomorrow's science is looking for leaders

Central dogma of molecular biology



Extractions help students visualize DNA

How to Extract DNA from a Strawberry



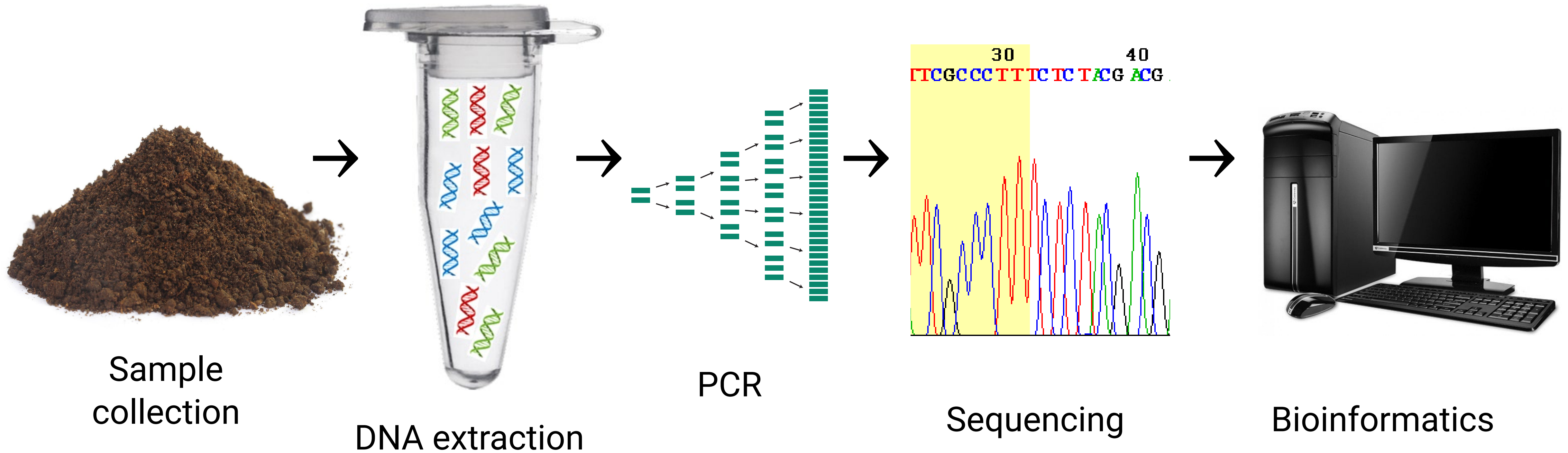
Cells are the basic unit of life and make up all plants, animals and bacteria. Deoxyribonucleic acid, or DNA, is the molecule that controls everything that happens in the cell. DNA contains instructions that direct the activities of cells and, ultimately, the body. This activity will demonstrate how DNA can be isolated from a strawberry using common household materials.

<https://www.youtube.com/watch?v=h0pu4iN5Bh4>

What you will need:



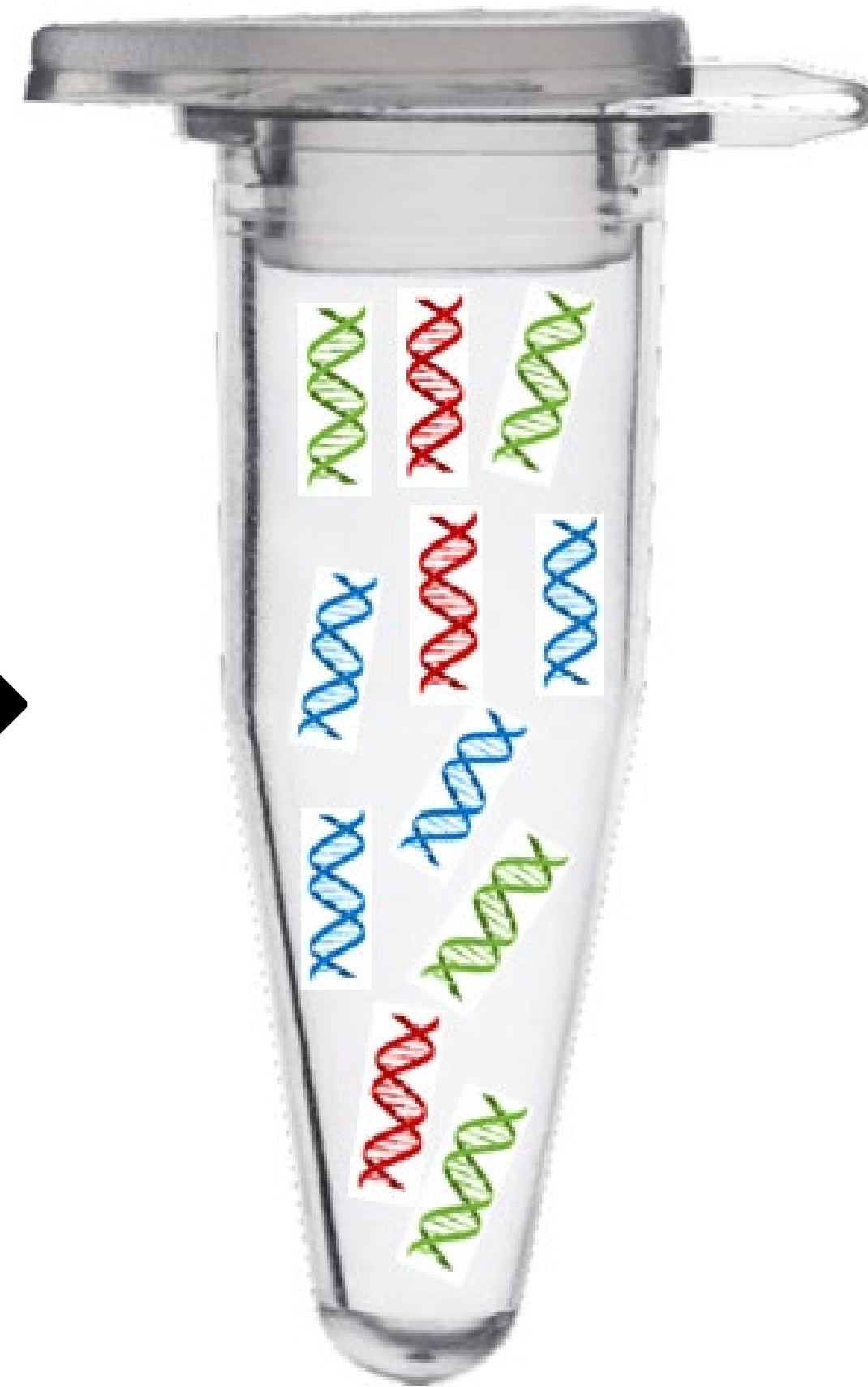
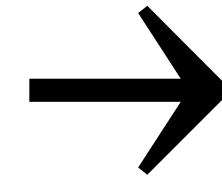
The process: Sample to DNA analysis



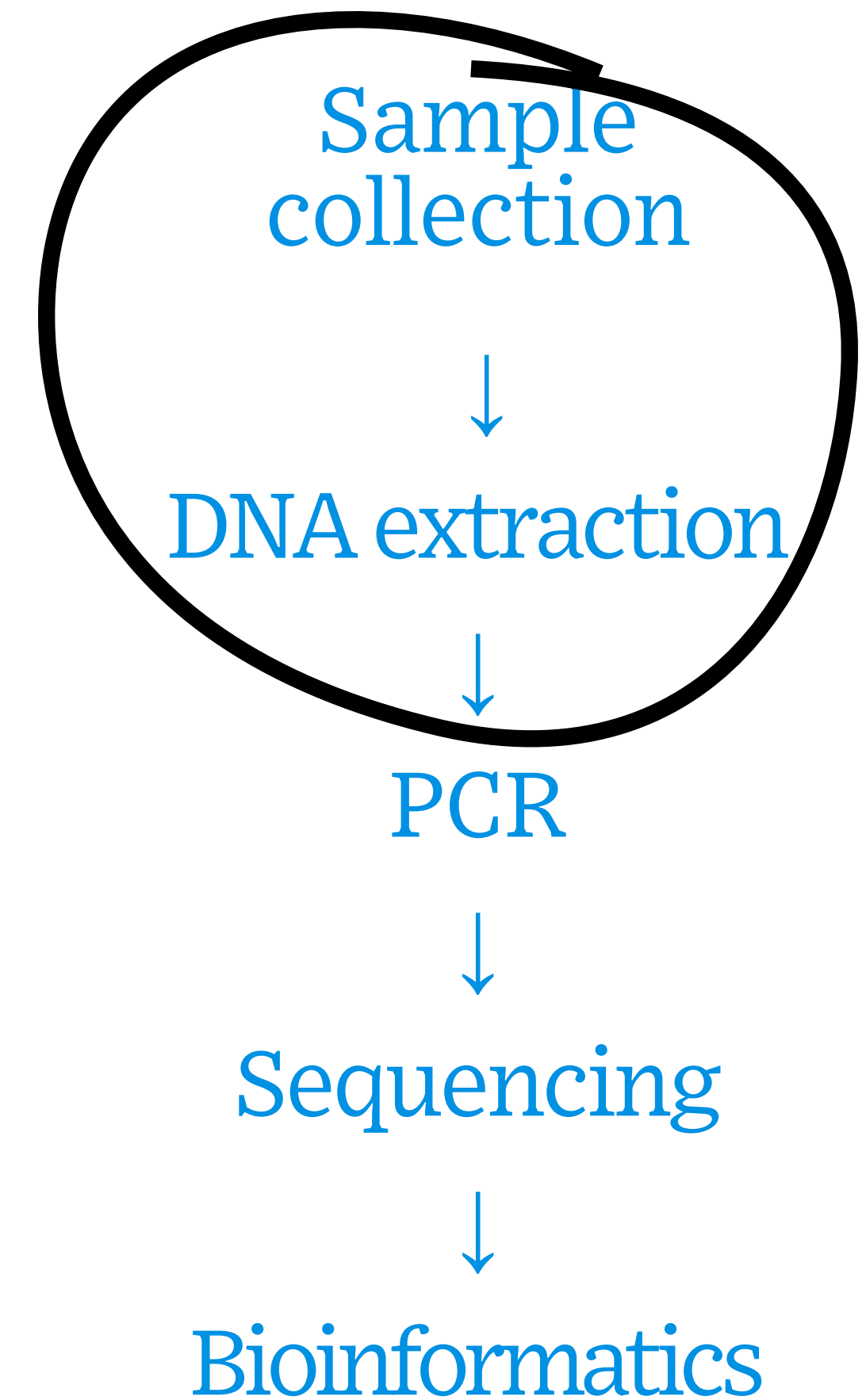
The process: Sample to DNA analysis



Soil sample

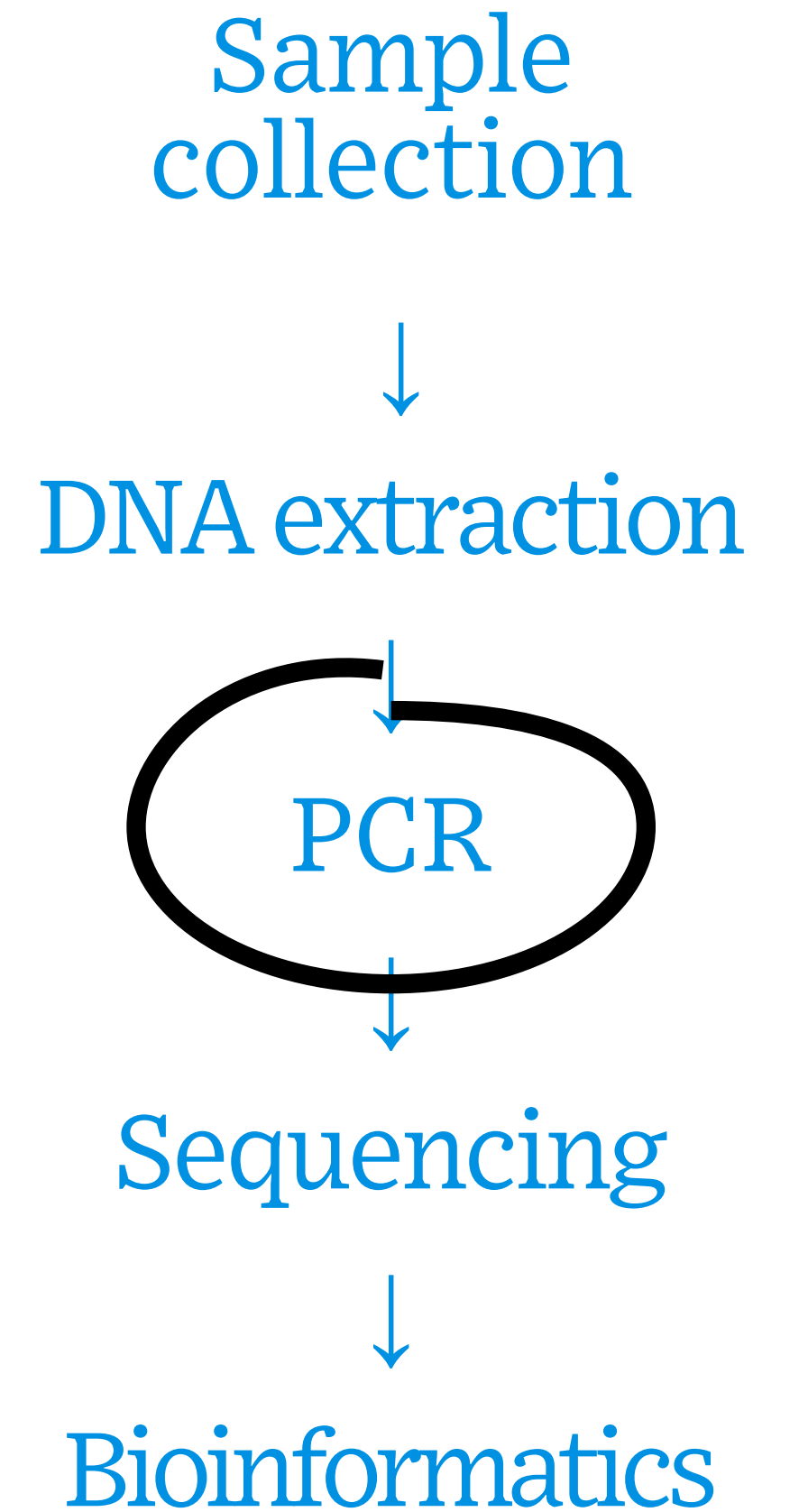
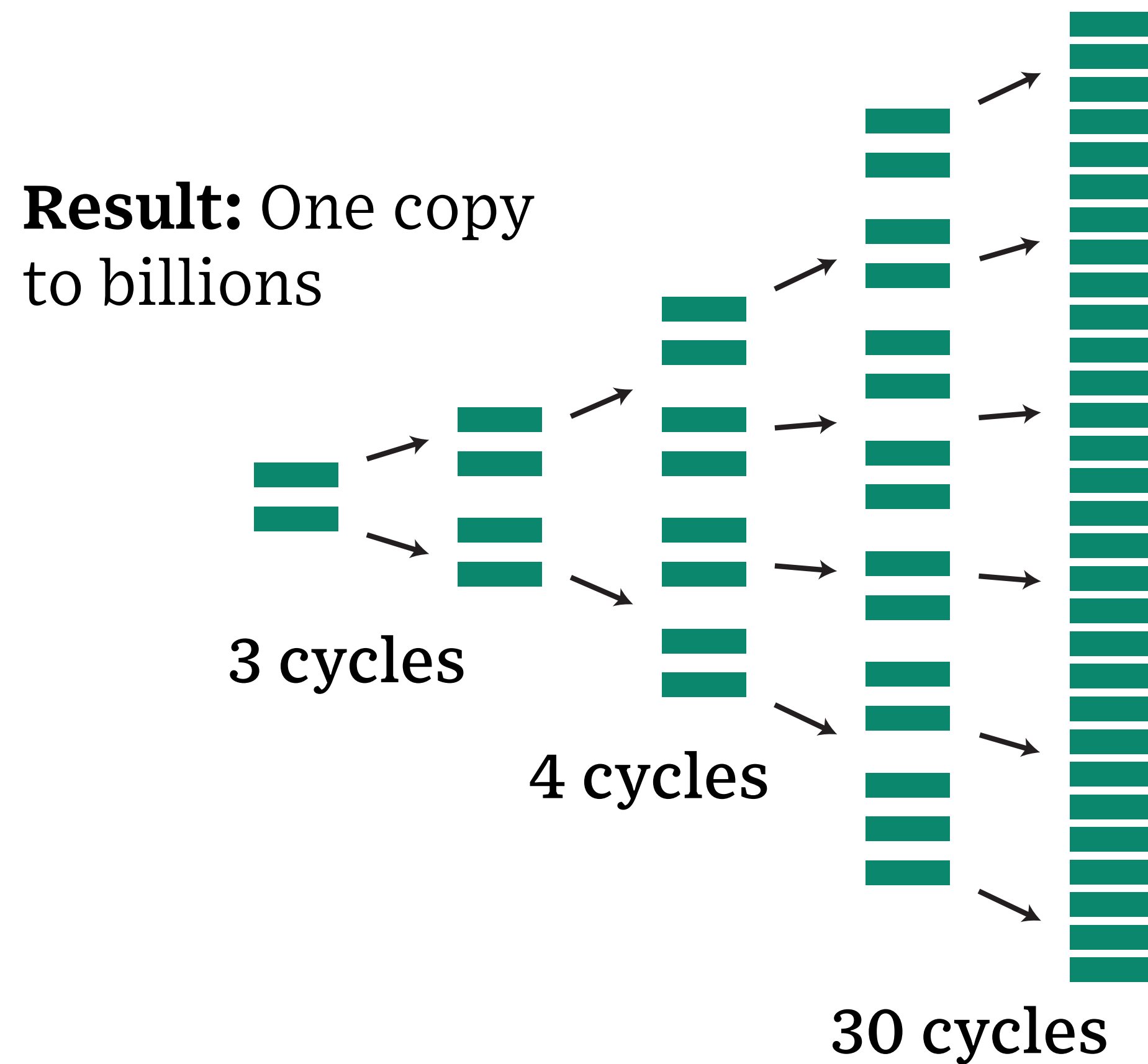


DNA extraction



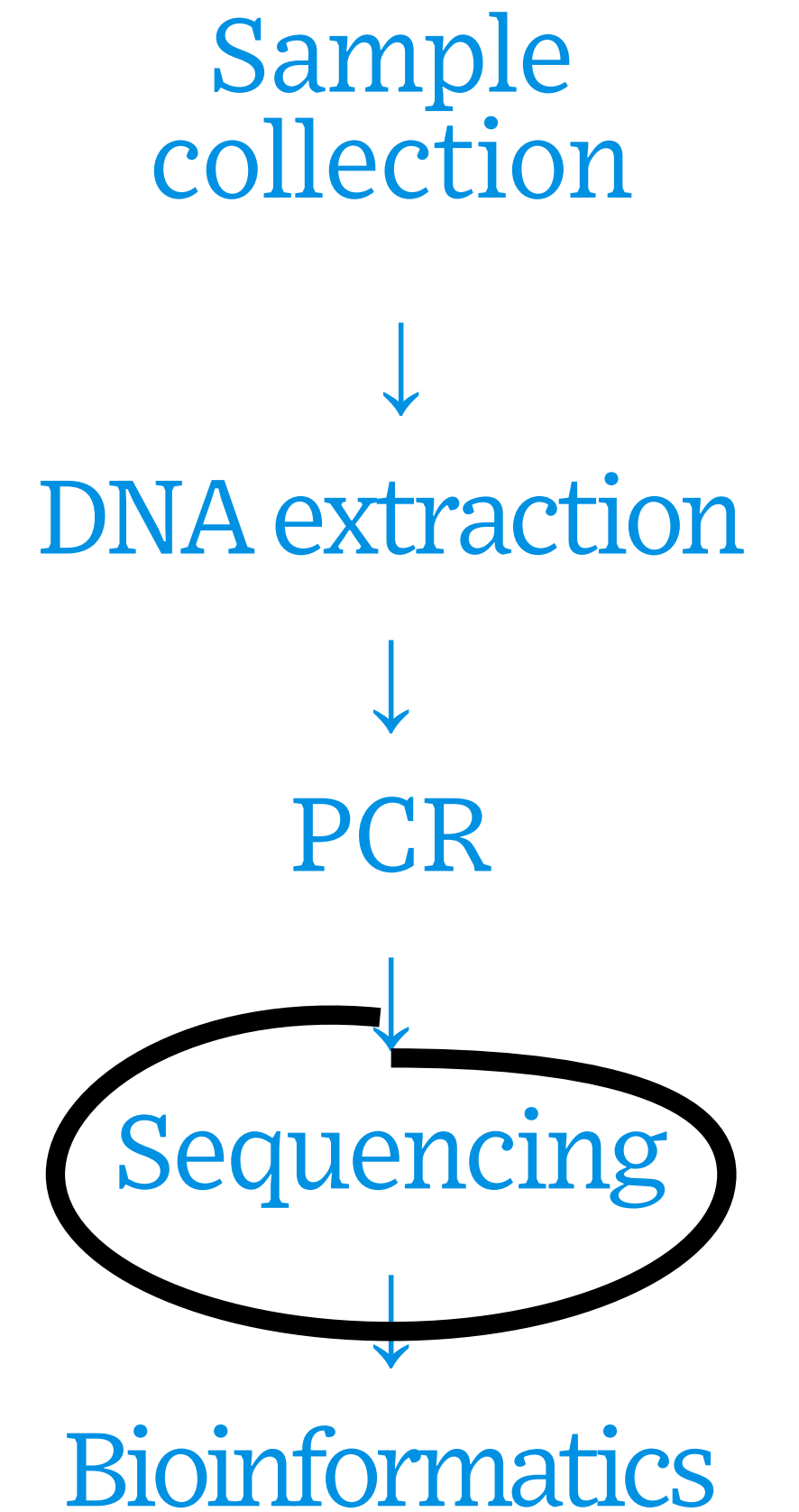
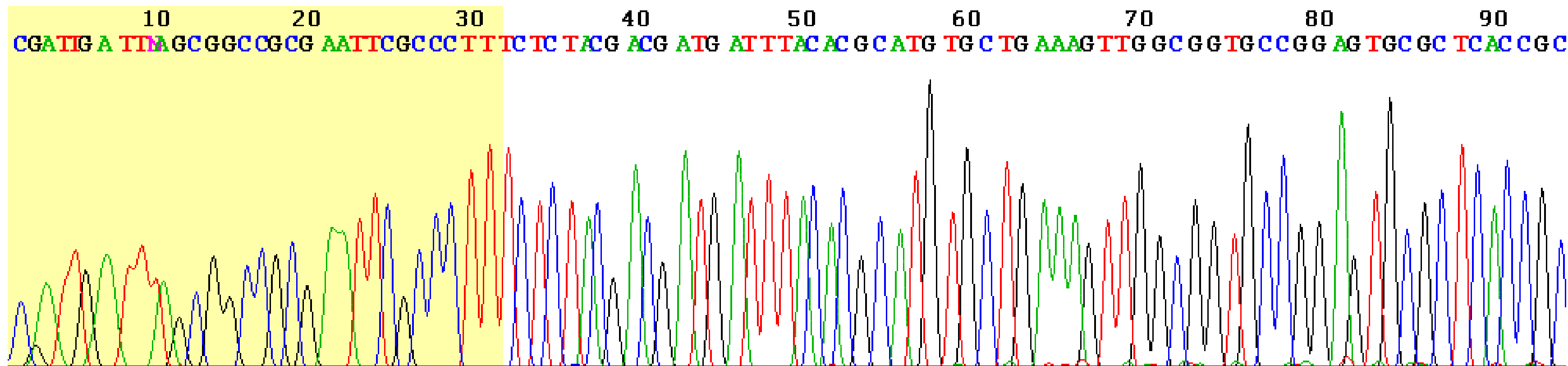
The process: Sample to DNA analysis

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



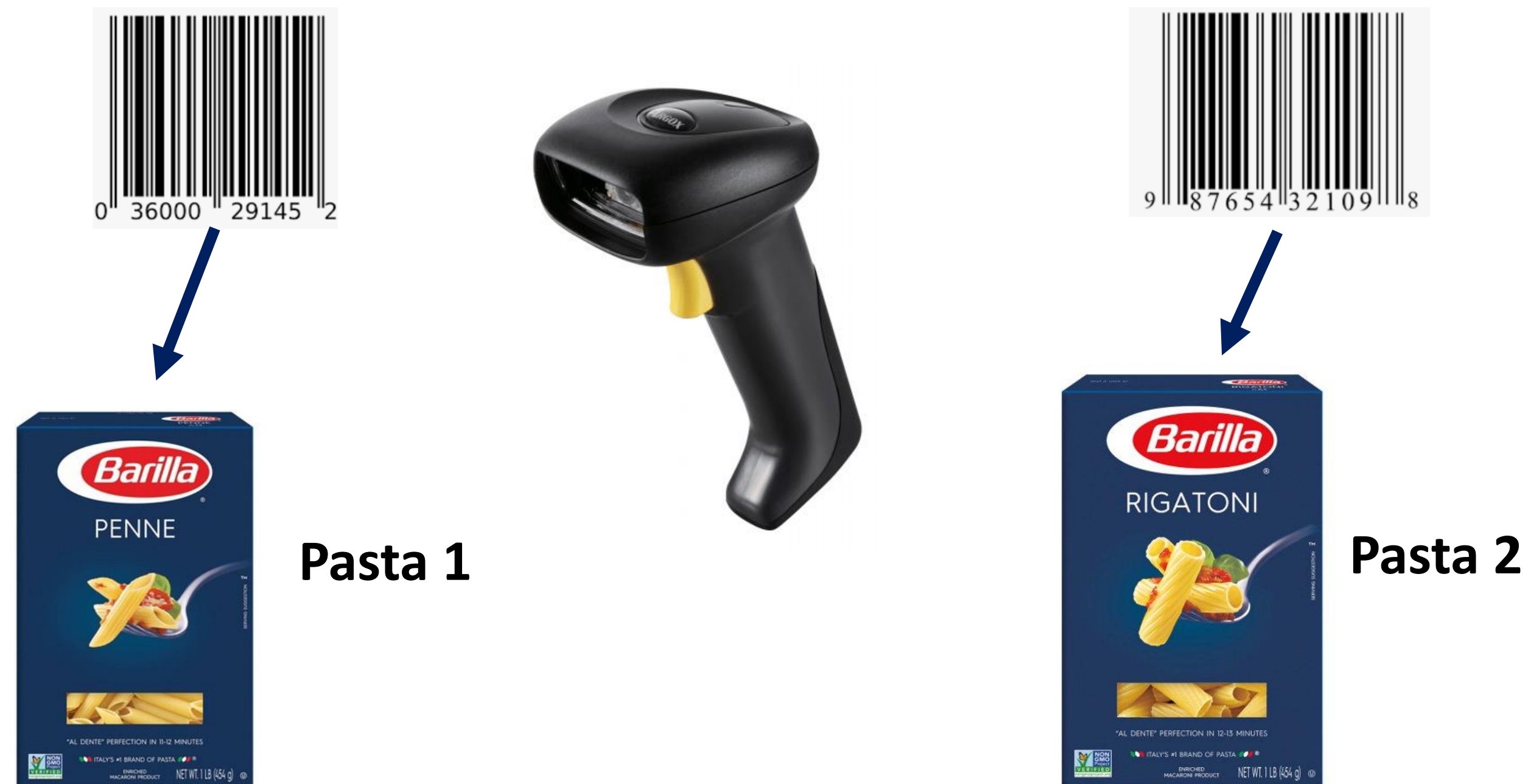
The process: Sample to DNA analysis

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



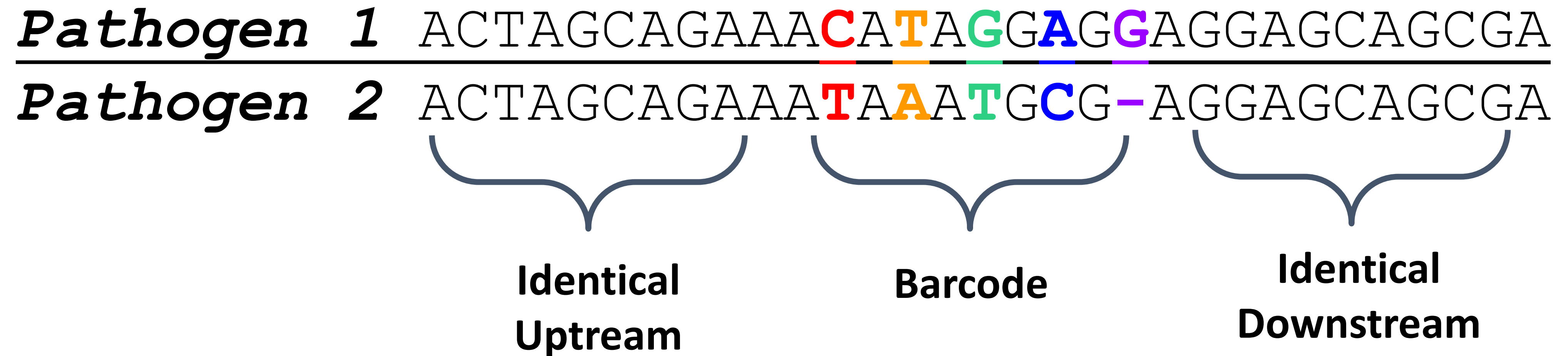
Barcodes

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.

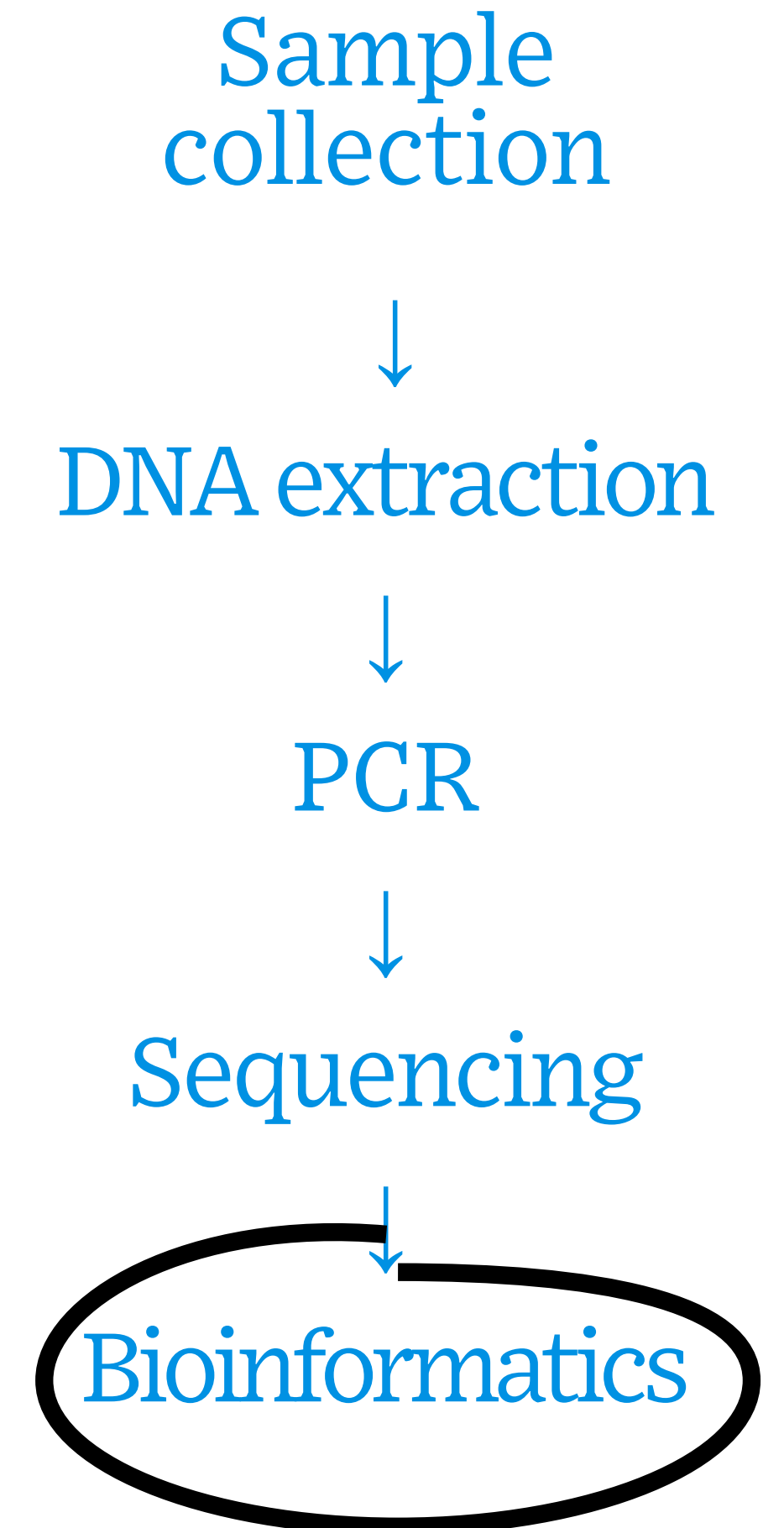
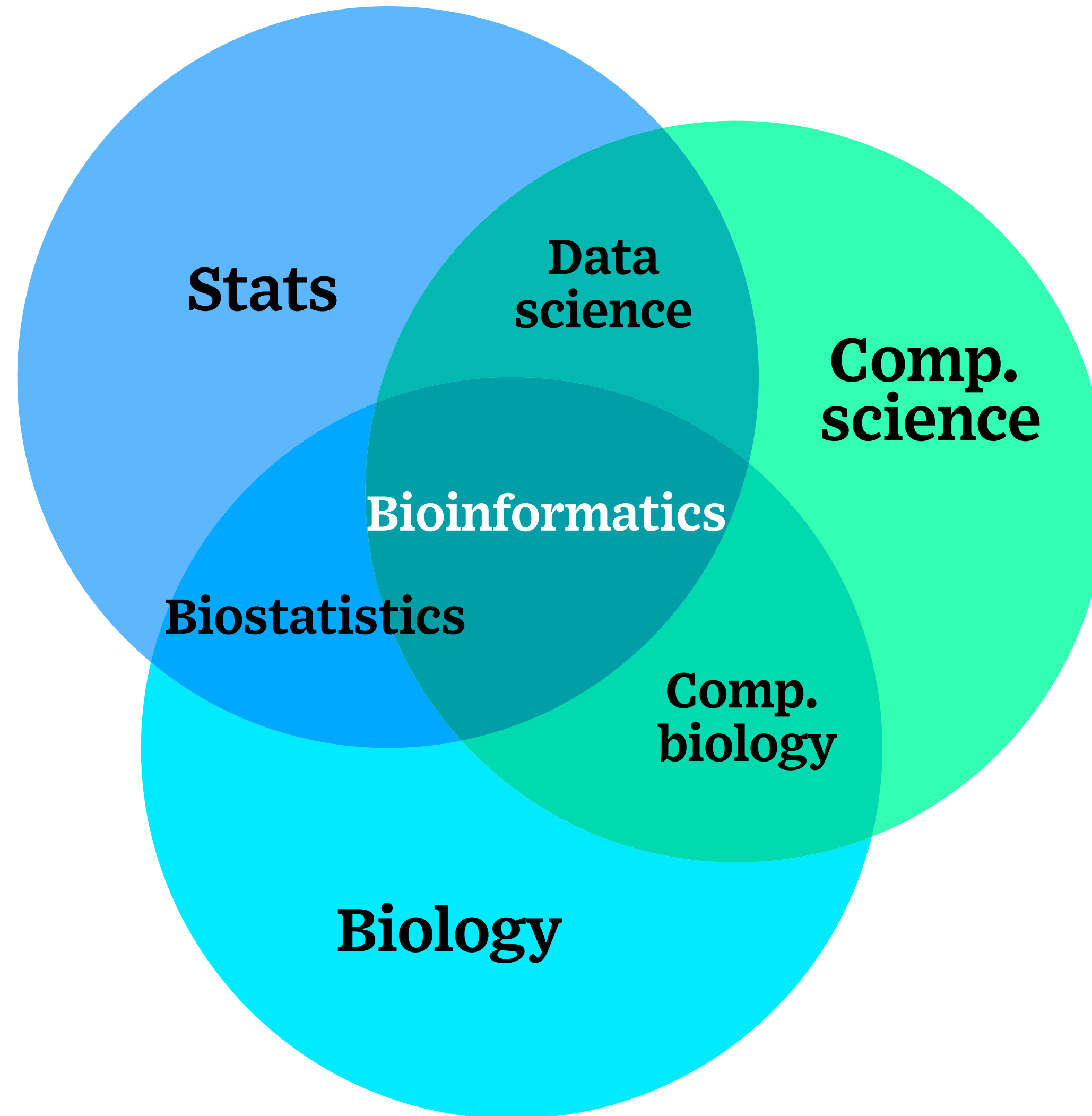


Barcodes

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



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



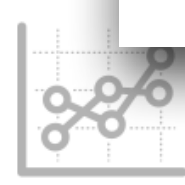
Develop

Use NCBI APIs and code libraries to build applications



Analysis

Identify an NCBI data analysis task



Popular Resources

PubMed

Bookshelf

PubMed Central

BLAST

Nucleotide

Genome

SNP

Gene

Protein

PubChem

BLAST: Basic Local Alignment Search

ncbi.nlm.nih.gov

Models (HMM) Release 14.0!

11 Jan 2024

Download release 14.0 of the NCBI protein profile Hidden Markov models

Best of 2023: A Look at the NCBI Insights


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...](#)

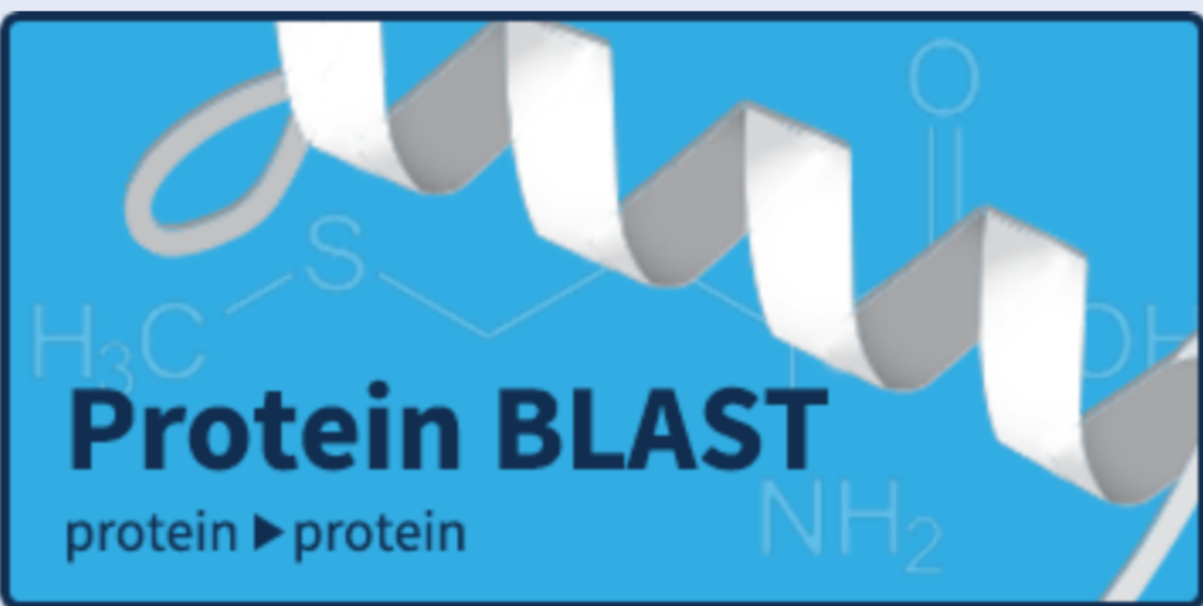
Web BLAST



Nucleotide BLAST
nucleotide ► nucleotide

blastx
translated nucleotide ► protein

tblastn
protein ► translated nucleotide



Protein BLAST
protein ► protein

BLAST® » blastn suite Home Recent Results Saved Strategies Help

Standard Nucleotide BLAST

blastn | blastp | blastx | tblastn | tblastx

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 no file selected [?](#)

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

Align two or more sequences [?](#)

Choose Search Set

Database Standard databases (nr etc.): rRNA/ITS databases Genomic + transcript d

New Experimental databases [Try experimental taxonomic nt databases](#) [Download](#)
For more info see [What are taxonomic nt databases?](#)

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

Limit to Optional Sequences from type material

Entrez Query Optional [YouTube](#) [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 nt using Megablast (Optimize for highly similar sequences)
 Show results in a new window

[+ Algorithm parameters](#)

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

2. Leave defaults

3. Click "BLAST"

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

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® » blastn suite » results for RID-UDT2W50K013

Home Recent Results Saved Strategies Help

[< Edit Search](#) Save Search Search Summary ▾

How to read this report? BLAST Help Videos Back to Traditional Results Page

Job Title 10 sequences (Sequence-1)

RID [UDT2W50K013](#) [Download All](#) ▾

Results for 1:lcl|Query_2858418 Sequence-1(702bp) ▾

Program BLASTN Citation ▾

Database nt [See details](#) ▾

Query ID lcl|Query_2858418

Description Sequence-1

Molecule type dna

Query Length 702

Other reports [Distance tree of results](#) [MSA viewer](#) ?

Filter Results

Organism *only top 20 will appear* 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 Download ▾ Select columns ▾ Show 100 ▾ ?

select all 100 sequences selected [GenBank](#) [Graphics](#) [Distance tree of results](#) [MSA Viewer](#)

Description	Scientific Name	Max	Total	Query	E	Per.	Acc.	Accession
-------------	-----------------	-----	-------	-------	---	------	------	-----------

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

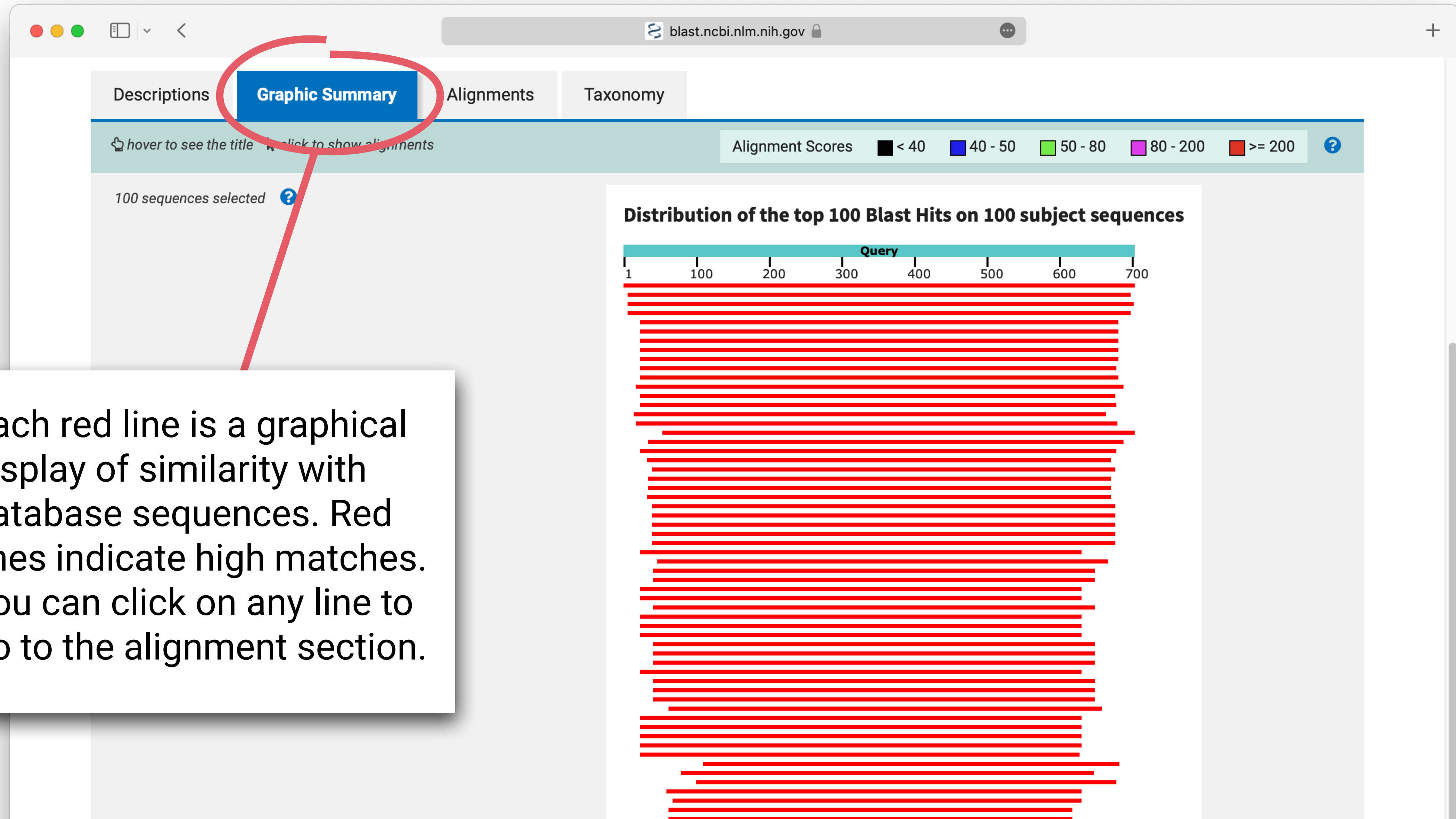
Sequences producing significant alignments Download Select columns Show 100 ?

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
<input checked="" type="checkbox"/>	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
<input checked="" type="checkbox"/>	Dreissena polymorpha mitochondrion, partial genome	Dreissena polym...	1253	1253	98%	0.0	99.28%	17803	MZ400414.1
<input checked="" type="checkbox"/>	Dreissena polymorpha mitochondrion, partial genome	Dreissena polym...	1247	1247	99%	0.0	98.99%	18452	KY091877.1
<input checked="" type="checkbox"/>	Dreissena polymorpha isolate DM1270 mitochondrion	Dreissena polym...	1247	1247	98%	0.0	99.13%	18902	MT483676.1
<input checked="" type="checkbox"/>	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
<input checked="" type="checkbox"/>	Dreissena polymorpha voucher ZEBR2 cytochrome c oxidase subunit I (COX1) gene, partial cds; mitochondrial	Dreissena polym...	1190	1190	93%	0.0	99.24%	658	
<input checked="" type="checkbox"/>	Dreissena polymorpha voucher ZEBR1 cytochrome c oxidase subunit I (COX1) gene, partial cds; mitochondrial	Dreissena polym...	1190	1190	93%	0.0	99.24%	658	
<input checked="" type="checkbox"/>	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	
<input checked="" type="checkbox"/>	Dreissena polymorpha voucher BIOUG13626-H02 cytochrome oxidase subunit 1 (COI) gene, partial cds; mitochon...	Dreissena polym...	1188	1188	93%	0.0	99.24%	658	
<input checked="" type="checkbox"/>	Dreissena polymorpha voucher BIOUG13626-G05 cytochrome oxidase subunit 1 (COI) gene, partial cds; mitochon...	Dreissena polym...	1184	1184	93%	0.0	99.24%	655	
<input checked="" type="checkbox"/>	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	
<input checked="" type="checkbox"/>	Dreissena polymorpha cytochrome oxidase 1 (COI) gene, partial cds; mitochondrial gene for mitochondrial product	Dreissena polym...	1184	1184	95%	0.0	98.51%	672	
<input checked="" type="checkbox"/>	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
<input checked="" type="checkbox"/>	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
<input checked="" type="checkbox"/>	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.

blast.ncbi.nlm.nih.gov

Descriptions Graphic Summary **Alignments** Taxonomy

Alignment view Pairwise CDS feature [Restore defaults](#)

100 sequences selected

[Download](#) [GenBank](#) [Graphics](#) [Next](#) [Previous](#) [Descriptions](#)

Dreissena polymorpha isolate Drpo4697 cytochrome c oxidase subunit I (COI) gene, partial cds; mitochondrial
Sequence ID: [EF414493.1](#) Length: 702 Number of Matches: 1

Range 1: 1 to 702 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Identities	Gaps	Strand
1293 bits(700)	0.0	702/702(100%)	0/702(0%)	Plus/Plus
Query 1	CAACAAATCATAAAGATATTGGTTCTTTATATTTTATTCTTTCTCTGTGNGCTGGCCTTG	60		
Sbjct 1	CAACAAATCATAAAGATATTGGTTCTTTATATTTTATTCTTTCTCTGTGNGCTGGCCTTG	60		
Query 61	TGGGCACGGGTTTTAGTGTTCTTATTCGTTTAGAGCTAAGGGCACCTGGAAGCGTCCTTG	120		
Sbjct 61	TGGGCACGGGTTTTAGTGTTCTTATTCGTTTAGAGCTAAGGGCACCTGGAAGCGTCCTTG	120		
Query 121	GTGATTTTCAATTATATAATTTAATTGTCACCACTCATGGGCTTGTTATAA+++++C	180		
Sbjct 121	GTGATTTTCAATTATATAATTTAATTGTCACCACTCATGGGCTTGTTATAA+++++C	180		
Query 181	TAGTAATACCTATAATAATGGGGGATTTCGAAATTGATTGGTACCAATAAATACTGAGTC	240		
Sbjct 181	TAGTAATACCTATAATAATGGGGGATTTCGAAATTGATTGGTACCAATAAATACTGAGTC	240		
Query 241	TTCCTGATATAGGTTTTCTCGTCTTAATAATGTTAGTTTTGGGTTTTACCTGTCTCTA	300		
Sbjct 241	TTCCTGATATAGGTTTTCTCGTCTTAATAATGTTAGTTTTGGGTTTTACCTGTCTCTA	300		
Query 301	TAGGACTTCTATTTTGTTCAAGTTTTAGGGAAGGAGGATTCGGGGTGGTTGAACCTTAT	360		
Sbjct 301	TAGGACTTCTATTTTGTTCAAGTTTTAGGGAAGGAGGATTCGGGGTGGTTGAACCTTAT	360		
Query 361	ACCCTCCTTTATCTAGAGTTATAGGACATTCAGGGCCTGCGATAGATTTTTGATTTTAT	420		

Sequence alignment (100% match)

The screenshot shows the BLAST interface with the 'Alignments' tab selected. The alignment view is set to 'Pairwise'. A red circle highlights the subject sequence description: **Dreissena polymorpha isolate Drpo4697 cytochrome c oxidase subunit I (COI) gene, partial cds; mitochondrial**. Below this, the sequence ID [EF414493.1](#) and length 702 are shown. The alignment table shows a perfect match (100% identities) between the query and subject sequences.

Score	Expect	Identities	Gaps	Strand
1293 bits(700)	0.0	702/702(100%)	0/702(0%)	Plus/Plus

Query	1	CAACAAATCATAAAGATATTGGTTCTTTATATTTTATTCTTTCTCTGTGNGCTGGCCTTG	60
Sbjct	1	CAACAAATCATAAAGATATTGGTTCTTTATATTTTATTCTTTCTCTGTGNGCTGGCCTTG	60
Query	61	TGGGCACGGGTTTTAGTGTTCTTATTCGTTTAGAGCTAAGGGCACCTGGAAGCGTCTTG	120
Sbjct	61	TGGGCACGGGTTTTAGTGTTCTTATTCGTTTAGA	
Query	121	GTGATTTTCAATTATATAATTTAATTGTCACCAC	
Sbjct	121	GTGATTTTCAATTATATAATTTAATTGTCACCAC	
Query	181	TAGTAATACCTATAATAATGGGGGGATTTCGGAAA	
Sbjct	181	TAGTAATACCTATAATAATGGGGGGATTTCGGAAA	
Query	241	TTCCTGATATAGGTTTTCTCGTCTTAATAATGT	
Sbjct	241	TTCCTGATATAGGTTTTCTCGTCTTAATAATGT	
Query	301	TAGGACTTCTATTTTGTTTCAGCTTTTAGGGGAAGGAGGATTCGGGGTGGTTGAACCTTAT	360
Sbjct	301	TAGGACTTCTATTTTGTTTCAGCTTTTAGGGGAAGGAGGATTCGGGGTGGTTGAACCTTAT	360
Query	361	ACCCTCCTTTATCTAGAGTTATAGGACATTCAGGGCCTGCGATAGATTTTTGATTTTAT	420

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; Neoheterodonte; Myida; Dreissenoida; 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