

02 | Barcoding basics



Equipment list

- Computer for:
 - questionnaire
 - bioinformatics and BLAST activity
 - sample information record sheet
- Questionnaire for completion (available in Google Forms): bit.ly/B4B-PreQ
- Declaration of consent form for completion by students and their parents
- Instructions for bioinformatics and BLAST activity
- Text file of Unknown barcodes
- Worksheet to record BLAST results
- Instructions for collecting and recording a sample for barcoding
- **Optional:** Appropriate equipment for invertebrate sample collection, ziplock bag, sharpie marker

Health and safety

When collecting invertebrate samples for barcoding, the usual precautions should be taken. Newly deceased invertebrates can be used. The humane way to manage fresh invertebrate samples is to place them in a ziplock bag in the -20°C freezer. After 24 hours they will be ready for DNA extraction.

Instructions | Bioinformatics and BLAST

Bioinformatics is the development of software and computing tools to organise and analyse raw biological data.

One way in which bioinformatics is used is for the comparison of DNA sequence. This can be:

- Comparison of known organisms to each other, to find out how similar their DNA sequence is
- Identification of an unknown organism, by comparison of its DNA sequence to a database of DNA sequences from other organisms

To compare DNA sequences, a BLAST is used. **BLAST** stands for **B**asic **L**ocal **A**lignment **S**earch **T**ool. In this task you will use a BLAST to identify invertebrates from their DNA barcodes.

DNA barcode is the name given to the DNA sequence of a gene found in the mitochondrial DNA of all animals. The mitochondrial cytochrome oxidase subunit 1 gene is used as the DNA barcode. It is a useful tool for identifying organisms as the gene sequence is constant within a species, but varies between species.

From this point onwards if you would prefer to use online instructions follow this link:

app.tango.us/app/workflow/02-Bioinformatics-and-BLAST-instructions-b91f7a52046c4cfb92881f92e30fc93a

1. Find the National Centre for Biological Information (NCBI) website with free software for DNA sequence comparison

- Type **NCBI BLAST** into an internet search engine.
- Click on the link for **BLAST: Basic Local Alignment Search Tool**.
- Select **Nucleotide BLAST**. This should take you onto the blastn tab.

The screenshot shows the NCBI BLAST web interface. At the top, there's a navigation bar with 'BLAST' and 'blastn suite' tabs. The main section is titled 'Standard Nucleotide BLAST'. It features a 'Query sub-range' section with 'From' and 'To' input fields. Below this is a 'Job Title' field and a 'Choose Search Set' section. The 'Choose Search Set' section has radio buttons for 'Standard databases (nr etc.)', 'Core nucleotide database', 'Genomic + transcript databases', 'Betacoronavirus', and 'Experimental databases'. The 'Standard databases' option is selected. There are also links for 'Add organism' and 'Create custom database'. The 'Exclude' section has checkboxes for 'Modes (XMAPS)', 'Uncultured/environmental sample sequences', and 'Sequences from type material'. The 'Limit to' section has a checkbox for 'Enter a query to limit search'. The 'Program Selection' section has radio buttons for 'Highly similar sequences (megablast)' and 'More dissimilar sequences (discontiguous megablast)'. The 'Highly similar sequences (megablast)' option is selected.

2. Match DNA barcodes from unknown invertebrates against a database of DNA sequences, to find which organisms the barcodes came from

- Use the text file **02_R_Unknown-barcodes**.

This file gives DNA sequence in FASTA format, with > indicating the name of the DNA sequence, and the sequence then starting on the next line. FASTA (Fast-All) format allows the sequence to be easily used in sequence alignment with existing databases.

- Copy the >, name and sequence for Barcode 1.
- On the blastn tab, in the white box in the **Enter Query Sequence** section, copy and paste the >, name and sequence for Barcode 1.
- In the **Choose Search Set** section, the **Standard databases (nr etc.)** should be checked.
- Select **Nucleotide collection (nr/nt)** from the drop down menu.
- In the **Program Selection** section, optimise for **Highly similar sequences (megablast)**.
- Click the blue **BLAST** button. Algorithms will try to find the best match for your barcode by comparing it to all of the DNA sequences stored in its database. Depending on how many searches are submitted at the same time as yours this may take a few minutes.

3. Understanding the results

- Scroll down until you see 4 tabs.
- On the **Descriptions** tab, you can see the scientific name (binomial classification) of the organism and the name of the DNA sequence that matches your query.

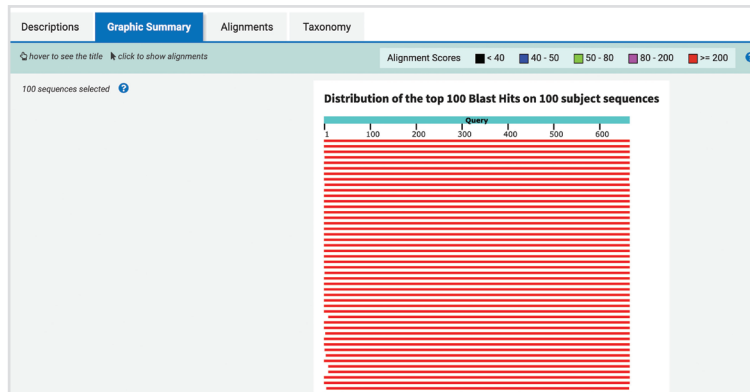
Sequences producing significant alignments									
Download Select columns Show 100									
	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	Coccinella septempunctata voucher NSMC-IN-150200623 cytochrome c oxidase subunit I (COX1) gene...	Coccinella sept...	1205	1205	100%	0.0	99.70%	683	OL663101.1
<input checked="" type="checkbox"/>	Coccinella septempunctata genome assembly, organelle, mitochondrion	Coccinella sept...	1205	1205	100%	0.0	99.70%	19413	OL015583.1
<input checked="" type="checkbox"/>	Coccinella septempunctata voucher CGAEC-070 cytochrome oxidase subunit 1 (COI) gene, partial cds...	Coccinella sept...	1205	1205	100%	0.0	99.70%	658	HM405545.1
<input checked="" type="checkbox"/>	Coccinella septempunctata voucher CHU06-COL-1-570 cytochrome oxidase subunit 1 (COI) gene, partial...	Coccinella sept...	1203	1203	99%	0.0	99.70%	657	KJ204183.1
<input checked="" type="checkbox"/>	Coccinella septempunctata voucher BBOI_Col_FK_1255 cytochrome oxidase subunit 1 (COI) gene, par...	Coccinella sept...	1199	1199	100%	0.0	99.54%	658	KM452668.1
<input checked="" type="checkbox"/>	Coccinella septempunctata voucher RFB_Col_FK_3694 cytochrome oxidase subunit 1 (COI) gene, par...	Coccinella sept...	1199	1199	100%	0.0	99.54%	658	KM439459.1
<input checked="" type="checkbox"/>	Coleoptera sp. BOLD:AAA8933 voucher NIBGE IMB-00072 cytochrome oxidase subunit 1 (COI) gene, p...	Coleoptera sp...	1199	1199	100%	0.0	99.54%	658	HM424025.1
<input checked="" type="checkbox"/>	Coccinella septempunctata voucher BIOUG02377-C12 cytochrome oxidase subunit 1 (COI) gene, partial...	Coccinella sept...	1197	1197	99%	0.0	99.54%	657	KY845819.1
<input checked="" type="checkbox"/>	Coccinella septempunctata voucher CHU06-COL-1-572 cytochrome oxidase subunit 1 (COI) gene, partial...	Coccinella sept...	1197	1197	99%	0.0	99.54%	657	KJ203556.1
<input checked="" type="checkbox"/>	Coccinella septempunctata voucher BBOI_Col_FK_0287 cytochrome oxidase subunit 1 (COI) gene, par...	Coccinella sept...	1195	1195	100%	0.0	99.39%	658	KM439318.1
<input checked="" type="checkbox"/>	Coccinella septempunctata voucher Lamar369 cytochrome oxidase subunit 1 (COI) gene, partial cds: mt...	Coccinella sept...	1195	1195	100%	0.0	99.39%	658	JF296194.1
<input checked="" type="checkbox"/>	Coccinella septempunctata voucher ZFMK-TIS-4023 cytochrome oxidase subunit 1 (COI) gene, partial c...	Coccinella sept...	1194	1194	100%	0.0	99.39%	658	KJ808178.1
<input checked="" type="checkbox"/>	Coccinella septempunctata voucher BIOUG02377-503 cytochrome oxidase subunit 1 (COI) gene, partial...	Coccinella sept...	1194	1194	100%	0.0	99.39%	658	KY845235.1
<input checked="" type="checkbox"/>	Coccinella septempunctata voucher BIOUG02377-803 cytochrome oxidase subunit 1 (COI) gene, partial...	Coccinella sept...	1194	1194	100%	0.0	99.39%	658	KY830676.1
<input checked="" type="checkbox"/>	Coccinella septempunctata voucher BIOUG02377-F11 cytochrome oxidase subunit 1 (COI) gene, partial...	Coccinella sept...	1194	1194	100%	0.0	99.39%	658	KY842295.1
<input checked="" type="checkbox"/>	Coccinella septempunctata voucher 088BCOL-3199 cytochrome oxidase subunit 1 (COI) gene, partial c...	Coccinella sept...	1194	1194	100%	0.0	99.39%	658	KM840653.1

- In the **Description** column, each line shows the species name, a sample reference made up of numbers and letters, then what the DNA sequence is. You should see the term **cytochrome c oxidase subunit I**, which is the gene used as a barcode for animal species, or **mitochondrion**, as this is where the DNA barcode gene is located.
- In the **Scientific name** column it gives the binomial classification of the organism. Click on this to see more information on classification and the common name, or type the binomical classification into an internet search engine to find out what the common name for the invertebrate is.

- In a column to the right, the **E value** or Expectation value, is the number of alignments with the query sequence that would be expected to occur by chance in the database.

Lower E values mean that the chance of the alignment occurring randomly is very low, so the probability that the sequence retrieved is related to the query sequence is high.

- On the **Graphic Summary** tab, you can see whether the sequence alignments are for the whole of the query sequence or just part of it.



Do you think the longer or shorter alignments would provide more precise results?

Precision = the closeness of agreement between independent measurements obtained under the same conditions.

- On the **Alignments** tab there is a detailed view of each sequence from the database aligned to the query sequence.

Score	Expect	Identities	Gaps	Strand
1205 bits (652)	0.0	657/659 (99%)	2/659 (0%)	Plus/Plus
Query 1	AACATTATATTTCTTATTCGGAATAGCGCGAATAATTGGACCTCTTTAAGAATTTT	60		
Sbjct 12	AACATTATATTTCTTATTCGGAATAGCGCGAATAATTGGACCTCTTTAAGAATTTT	71		
Query 61	AATTGCTCTTGAATTAGGAACCTAATAAGATTAAATGGAAATGACCAATTTATAATGT	120		
Sbjct 72	AATTGCTCTTGAATTAGGAACCTAATAAGATTAAATGGAAATGACCAATTTATAATGT	131		
Query 121	AATTGTAACAGCTCATGCCCTTCATTATAATATATAGTTATACCAATTAATATGG	180		
Sbjct 132	AATTGTAACAGCTCATGCCCTTCATTATAATATATAGTTATACCAATTAATATGG	191		
Query 181	AGGATTGGAAATGACTTGTCTTTAATAATGGAGACCTGACATAGCTTTCCCTCG	240		
Sbjct 192	AGGATTGGAAATGACTTGTCTTTAATAATGGAGACCTGACATAGCTTTCCCTCG	251		

- To present the results of your scientific identification of the invertebrate using a barcode, you should include this alignment. Click the **Download** button in the top left hand corner and select **Text (aligned sequences)**, then press **Continue**.
- Make sure that you save the alignment in a location and use a file name that you will be able to find again.

4. Writing up your results

Complete the worksheet **Identifying unknown organisms using DNA barcodes** entering the scientific name of the organism using binomial classification, the common name of the organism, the E value and the percentage of the query sequence that is identical to the DNA sequence from the database.

Instructions | Sampling for your investigation

DNA barcoding has been used for bioexploration and biomonitoring. The method you will use has been optimised to investigate the variety of invertebrates found in your local area. Your first task is to collect and record information on the specimens that you will use in your project.

☐ 1. Sample collection

To give the best chance of isolating DNA successfully, fresh samples of invertebrates should be used. When collecting your invertebrate for DNA barcoding you could find a newly deceased invertebrate (one that has died within the last 2 days). There are many other methods that can be used to collect invertebrate samples, including sweep nets, kick sampling, tree beating, pitfall traps, Tullgren funnels and pooter tubes.

- Choose a method appropriate to the type of invertebrate you wish to work with.
- Once you have your invertebrate, place it in a Ziplock bag.
- Label the bag with your initials and store it in the -20°C freezer.

If your invertebrate is caught alive, it can be humanely euthanised by placing in a ziplock bag in the -20°C freezer overnight.

☐ 2. Recording sample information

When you have your sample, you should record information about it carefully, so that another scientist could test the reproducibility of your results.

Reproducibility = precision obtained when measurement results are produced over a wider timescale by different people using equivalent equipment in different (but equivalent) places.

A sample information record sheet is provided for this purpose. It has space to include:

- Sample number
- School name
- Names of the collectors
- Date and time of collection
- Latitude and longitude of collection location (GPS or Google maps can help!)
- Description of habitat it was collected from
- Description of invertebrate (eg; colour, length, number of legs, any other identifying features)
- Provisional identification (if possible, using a field guide or taxonomic key)

You must also take a photo of the invertebrate and store it in a location and with a name that you will remember.

☐ 3. Sharing sample information

To allow comparisons to be made between the different samples collected for use in this project, the sample information record sheet should be stored somewhere accessible by the Barcoding for beginners group at your school. A template has been set up on Microsoft Word to allow this data to be collected for the group's samples.