

DNA Variation & Bioinformatics

Analysing DNA to identify seabirds in fisheries bycatch

A resource developed by Imogen Foote in collaboration with Dr Wilda Laux



Toroa (*Diomedea antipodensis gibsoni*), Gibson's albatross
Photo: JJ Harrison, CC BY-SA 3.0

1. Introduction

Seabirds are birds that are highly specialised to catch their food from the ocean. Some seabirds spend most of their life at sea, only returning to land once a year to breed. Aotearoa New Zealand (NZ) is considered the seabird capital of the world. Around 85 seabird species breed here. Of these, 35 species breed nowhere else in the world. Many seabirds breed on offshore islands, including as far south as the New Zealand subantarctic region. Unfortunately, many seabird species are declining and highly threatened.

The most likely threats to seabirds are:

- **Incidental fisheries bycatch** – seabirds try to eat the bait on fishing hooks and get caught or they get tangled in nets underwater and drowned.
- **Climate change** – warming oceans mean their fish prey swim deeper to cooler waters, which makes it harder for seabirds to catch food. Storms also destroy coastal breeding habitat.
- **Predation** – predators such as rats, cats and mustelids prey upon eggs and chicks.
- **Pollution** – seabirds can mistake plastic for food, then die of malnourishment with a stomach full of plastic.

Want to know more about the seabirds of Aotearoa?

<https://teara.govt.nz/en/seabirds-overview>

<https://www.catchfishnotbirds.nz/nz-seabirds>

<https://www.doc.govt.nz/nature/native-animals/birds/sea-and-shore-birds/>

2. Task

Aim: To identify two seabirds captured in a NZ fishing vessel.

Method:

You will analyse DNA sequences obtained from two dead seabirds caught by a fishing boat to identify their species. This will help conservation managers know which seabirds are at risk of being caught and might need better protection.

A. Read the fictional scenario below from the Department of Fauna.

Each year, thousands of seabirds are captured in NZ fisheries as incidental bycatch. Some fishing vessels have onboard observers – these are people whose job is to monitor fishing activity and report catch levels including fish and unintended catch (bycatch) and make sure it matches what the fishing companies are reporting. A longline fishing boat (a boat that sets a long fishing line with baited hooks in the water behind the boat) off the West Coast of the North Island reported the catch of two seabirds that could not be identified by observers as they were recovered too badly damaged. Feathers were collected from the birds and sent to a laboratory to get DNA for analysis to allow identification of the species.

B. Identify the species

The Department of Fauna laboratory collected, processed and prepared DNA samples from feathers collected from the two seabirds, then sequenced a section of the Cytochrome Oxidase I (COI) gene from the DNA samples¹. COI is a gene in the mitochondrial genome that plays a role in cellular respiration. Respiration is the process by which glucose is turned into energy that is used to power many cell reactions. COI is useful for species identification because the mutation rate is high enough to create unique sequences allowing it to act like a genetic 'barcode' that can distinguish one species from another.

>unidentified_seabird_A

```
TGGGGGTTTATGTTGATGGCTGTTGTAATGAAGTTAATTGCCCTAAGATAGAGGATACACCT
GCTAGGTGAAGAGAGAAGATGGCCAAGTCAACCGAAGCTCCAGCATGGGCTAGATTACCGGCC
AGAGGGGGGTATACAGTTCATCCTGTGCCTGCTCCTGCTTCTACTGTAGATGAAGCTAGGAGGA
GAAGGAAGG
```

>unidentified_seabird_B

```
TATATTGATGGCAGTTGTGATGAAGTTAATTGCTCCTAGGATTGATGAAACACCTGCTAGGTGGAGGGA
GAAGATAGCCAGGTCTACTGAAGCCCCTGCGTGGGCAAGGTTGCCAGCTAGAGGCGGGTACACAGTTC
ATCCTGTACCTGCTCCTGCTTCTACTGTGGAGGATGCTAGTAGGAGGAGGAAGGATGGGGGTA
G
```

You are now going to analyse these sequences to identify the seabird species.

Your aim is to answer two questions:

- Which species do these DNA sequences belong to?
- Is this gene (COI) a suitable marker for identifying all species? If not, what are some limitations of this marker?

FOLLOW THE STEPS BELOW.

- 1) Open the link to [NCBI BLAST](#). The BLAST tool of NCBI performs a sequence similarity search across the hundreds of millions of DNA sequences stored in the GenBank database.
- 2) Copy the following DNA sequence (including the title):

```
>unidentified_seabird_A
TGGGGGTTTTATGTTGATGGCTGTTGTAATGAAGTTAATTGCCCTAAGATAGAGGATACACCT
GCTAGGTGAAGAGAGAAGATGGCCAAGTCAACCGAAGCTCCAGCATGGGCTAGATTACCGGCC
AGAGGGGGGTATACAGTTCATCCTGTGCCTGCTCCTGCTTCTACTGTAGATGAAGCTAGGAGGA
GAAGGAAGG
```

- 3) Paste the sequence into the “Enter Query Sequence Box” and make sure ‘blastn’ is selected at the top (the tab appears blue).

Standard Nucleotide BLAST

blastn
blastp
blastx
tblastn
tblastx

Enter Query Sequence

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

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

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

Query subrange [?](#)

From

To

- 4) Scroll to the bottom of the webpage and click the blue BLAST button.
- 5) Wait for the results to load (this may take a minute).
- 6) Scroll down the report page. The “descriptions” tab will be highlighted in Blue. You get ‘hits’ of sequences stored in the NCBI database that are most similar to your query sequence.

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"/> Procellaria parkinsoni voucher SP618 cytochrome oxidase subunit 1 (COI) gene, partial cds: mitochondrial	Procellaria parkinsoni	370	370	100%	4e-98	100.00%	648	MK261837.1
<input checked="" type="checkbox"/> Procellaria cinerea voucher SP615 cytochrome oxidase subunit 1 (COI) gene, partial cds: mitochondrial	Procellaria cinerea	337	337	100%	4e-88	97.00%	648	MK261801.1

- 7) Write down the names of the top 5 “hits” the DNA sequence matches with, and the percent identity (“Perc. Ident” column) of the match in the Seabird A column below

Seabird A		Seabird B	
Scientific name	Percent identity	Scientific name	Percent identity
<i>Procellaria parkinsoni</i>	100%	<i>Thalassarche melanophrys</i>	100%
<i>Procellaria cinerea</i>	97%	<i>Thalassarche melanophrys</i>	100%
<i>Procellaria cinerea</i>	97%	<i>Thalassarche melanophrys</i>	100%
<i>Procellaria aequinoctialis</i>	97%	<i>Thalassarche chrysostoma</i>	100%
<i>Procellaria aequinoctialis</i>	97%	<i>Thalassarche melanophrys</i>	100%

- 8) Open a new BLAST window and repeat steps 1-7 for seabird B (DNA sequence below) and fill out the Seabird B column in the table above. Keep your seabird A BLAST window open, we will

return to that later.

>unidentified_seabird_B

TATATTGATGGCAGTTGTGATGAAGTTAATTGCTCCTAGGATTGATGAAACACCTGCTAGGTGGAGGGA
GAAGATAGCCAGGTCTACTGAAGCCCCTGCGTGGGCAAGGTTGCCAGCTAGAGGCGGGTACACAGTTC
ATCCTGTACCTGCTCCTGCTTCTACTGTGGAGGATGCTAGTAGGAGGAGGAAGGATGGGGGTA

- 9) Review the 5 top hits from the table above. Based on the 'Percent Identity', which hit do you think is most likely to tell us about the species identity of seabird A and seabird B?
Hint: you will probably need to google the scientific names, and you may not be able to identify both specimens confidently.

Seabird A: Procellaria parkinsoni

Seabird B: Thalassarche (cannot identify species)

- 10) In the BLAST results for both seabird A and seabird B click MSA (Multiple Sequence Alignment) viewer to open a new window showing the BLAST results as a DNA sequence alignment.

The screenshot shows the BLAST results interface. At the top, there are tabs for 'Descriptions', 'Graphic Summary', 'Alignments', and 'Taxonomy'. Below these is a table titled 'Sequences producing significant alignments'. The table has columns for 'Description', 'Scientific Name', 'Max Score', 'Total Score', 'Query Cover', 'E value', 'Per. Ident', 'Acc. Len', and 'Accession'. Two sequences are listed: 'Procellaria parkinsoni voucher SP618 cytochrome oxidase subunit 1 (COI) gene, partial cds: mitochondrial' and 'Procellaria cinerea voucher SP615 cytochrome oxidase subunit 1 (COI) gene, partial cds: mitochondrial'. To the right of the table, there are links for 'GenBank', 'Graphics', 'Distance tree of results', and 'MSA Viewer'. The 'MSA Viewer' link is circled in red.

- 11) In the new window your query sequence is shown at the top (orange box below) as nucleotides (A, C, T and G) and all of the BLAST hits are shown in the following lines (green box below).

The screenshot shows the MSA Viewer interface. At the top, there is a 'Sequence ID' field with 'Query_7831521' and a 'Start' field with '1'. Below these is a sequence alignment view. The query sequence is shown in an orange box: 'TATATTGATGGCAGTTGTGATGAAGT'. Below the query sequence, there are two BLAST hits shown in a green box: 'MK261849.1 (-)456' and 'MK262312.1 (-)516'. The alignment view shows the query sequence aligned with the BLAST hits, with grey regions indicating matches and red regions indicating mismatches.

Read more about MSA here: <https://www.ncbi.nlm.nih.gov/tools/msaviewer/tutorial1/>

What do the grey and red regions of the MSA represent?

Grey: Sequence matches query

Red: Mismatch to query sequence

- 12) What do you notice about the difference in the amount of grey vs red regions in the top ~10 BLAST hits between seabird A and seabird B?

There is more red (indicating mismatches) between sequences for seabird A compared to seabird B

-
- 13) What does this difference mean in terms of genetic distance, and how do you think this might have influenced your ability to identify seabird A vs seabird B?

Hint: you might need to do some research about 'genetic distance'.

There is greater genetic distance between species in *Procellaria* i.e. different species have more differences in DNA sequence compared to *Thalassarche* where there aren't many differences i.e. species are more similar. This shows that species within *Thalassarche* are more closely related/recently diverged compared to *Procellaria*. This means we could easily distinguish seabird A from other species in the genus *Procellaria*, but for seabird B there are not enough nucleotide differences between species to confidently identify the species from COI sequence.

- 14) Here we were looking at a region of the DNA marker COI, but there are other DNA markers you can use e.g. cytochrome-*b*. Would we expect all DNA markers show the same genetic distance? What would be the next step we could take to try and identify seabird B with more confidence?

No, different DNA markers evolve at different rates, so some will have more differences between species while others will be mostly the same. Could investigate a more variable/less evolutionarily constrained marker OR look at multiple markers/whole genome for increased resolution.

¹DNA sequence data from Foote, I. (2024). *Population genomics of Antipodean and Gibson's albatross and use of genetic markers for threatened seabird species identification* [Doctoral thesis, Victoria University of Wellington].