SWFSbase User Guide

Table of Contents

Introduction *2*

Accessing SWFSbase: *2*

Logging in to SWFSbase *3*

Running a BLAST with SWFSbase *4*

Included Databases **4**

Search Strategies **5**

Viewing results *5*

Metadata **6**

Submitting sequences to SWFSbase *8*

Metadata entry **8**

Linking metadata to your FASTA **8**

Introduction

SWFSbase is a secure BLAST server developed for the wildlife forensic community. It centralises curated reference databases contributed by laboratories and links them to essential forensic metadata, while also incorporating selected publicly available reference data. SWFSbase is hosted on SequenceServer™ Cloud. More information about the platform is available on the SWFSbase webpage: https://www.wildlifeforensicscience.org/swfsbase/

This User Guide provides an overview of how to use SWFSbase, including the different interfaces of SWFSbase, how to run BLAST searches, and how to submit sequences.

If you have any questions or issues, please contact: SWFSbase@wildlifeforensicscience.org.

Accessing SWFSbase:

SWFSbase can only be used for wildlife forensic casework. It cannot be used for research unless the relevant sequences are 1) owned by you, 2) are publicly available, or 3) you have obtained permission from the data owner to use them for research.

To access SWFSbase you must:

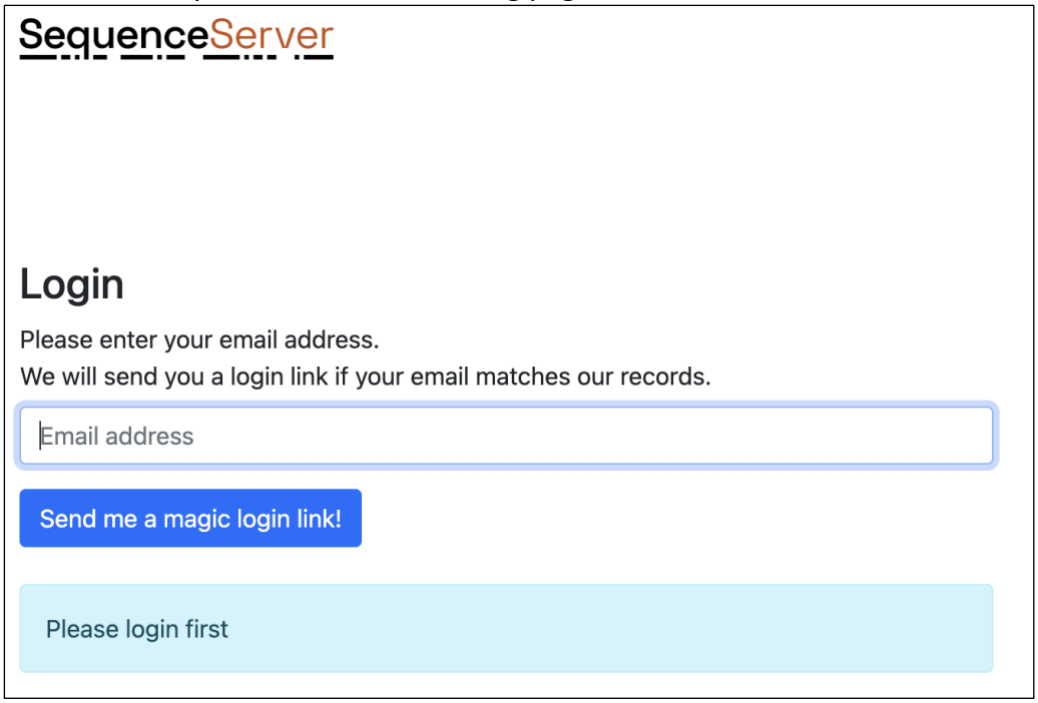
1. Be a member of the Society for Wildlife Forensic Science (SWFS).
2. Review and sign the SWFSbase Terms of Use.

The Terms of Use can be downloaded from the SWFSbase webpage. Signed agreements should be emailed to SWFSbase@wildlifeforensicscience.org. Please also include your full name, institution, and the email address you would like associated with your login.

Once these requirements have been completed, you will be added as a SWFSbase user. You can then access the platform using the following link:
https://wildlifeforensics.sequenceserver.com

Logging in to SWFSbase

SWFSbase uses a secure login system (two-step verification). When accessing the platform using the link above, you will see the following page:



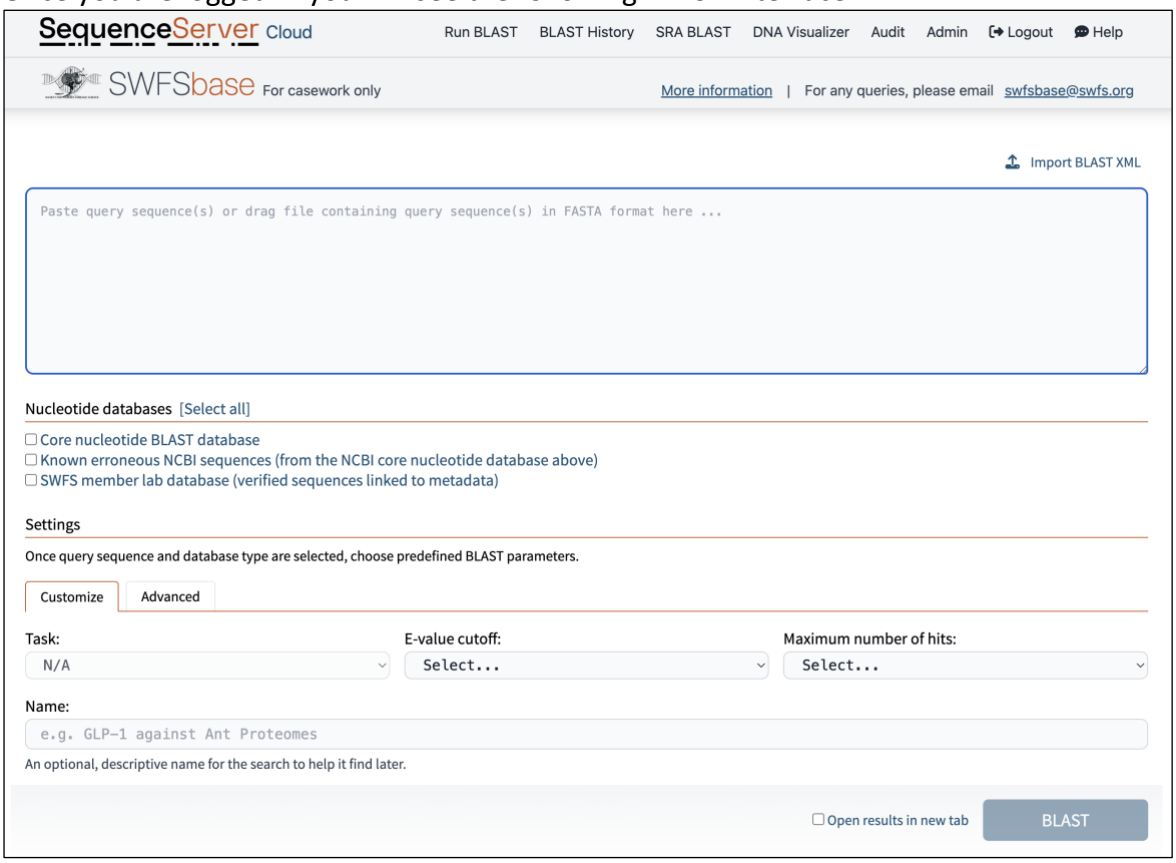
To log in:

1. Enter your registered email address.
2. You will receive an email containing a login link.
3. Click the link (or copy and paste it into your browser).

The login link remains valid for up to 30 days, or until you click Logout within SWFSbase. After this, you will need to repeat the login process. For security reasons, it is recommended that you log out after each session. Additional information about platform security can be found here.

Running a BLAST with SWFSbase

Once you are logged in you will see the following BLAST interface:



This interface functions similarly to an NCBI BLAST search. You can paste your sequence/s directly into the query box or drag and drop a FASTA file. Multiple sequences can be submitted in a single search.

Included Databases

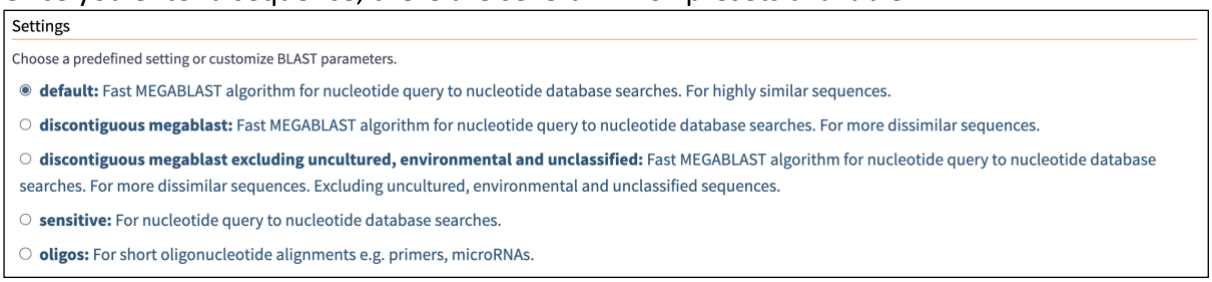
- ***SWFS member lab database (verified sequences linked to metadata):*** This is the key database within SWFSbase. It comprises vetted data contributed by multiple wildlife forensic laboratories, including data not available in other public repositories. Each sequence is linked to associated metadata to support assessment of specimen identification, sequence quality, and reliability (see image below). Laboratories may contribute sequences by following the submission instructions provided below.
- ***Core nucleotide BLAST database:*** This is the entire NCBI core nucleotide database ("core_nt"), updated following NCBI's frequency (typically daily to fortnightly). These sequences are not curated within SWFSbase and are not linked to forensic metadata. The database is provided for convenience, allowing users to conduct searches without leaving the platform.
- ***Known erroneous NCBI sequences (from the NCBI core nucleotide database above):*** This is a small database that contains sequences from the NCBI core nucleotide database that have previously been identified as erroneous. After conducting a search against the standard NCBI database, users may wish to query this database to determine whether top matches correspond to known problematic records.

Search Strategies

A useful strategy is to first search the '*SWFS member lab database'*, which contains highly reliable and transparent reference data. You may then wish to search the '*NCBI nucleotide database*', which contains a much broader set of sequences. Finally, if your top matches come from NCBI, you can search the *'Known erroneous NCBI sequences'* database to determine whether any of those matches correspond to previously identified problematic records.

Note that the larger the database used, the longer the BLAST search will take. The '*NCBI core nucleotide database*' is by far the largest, and searches will therefore typically run slower than the other two databases. However, repeated searches against this database may run faster after the first search due to caching.

Once you enter a sequence, there are several BLAST presets available:



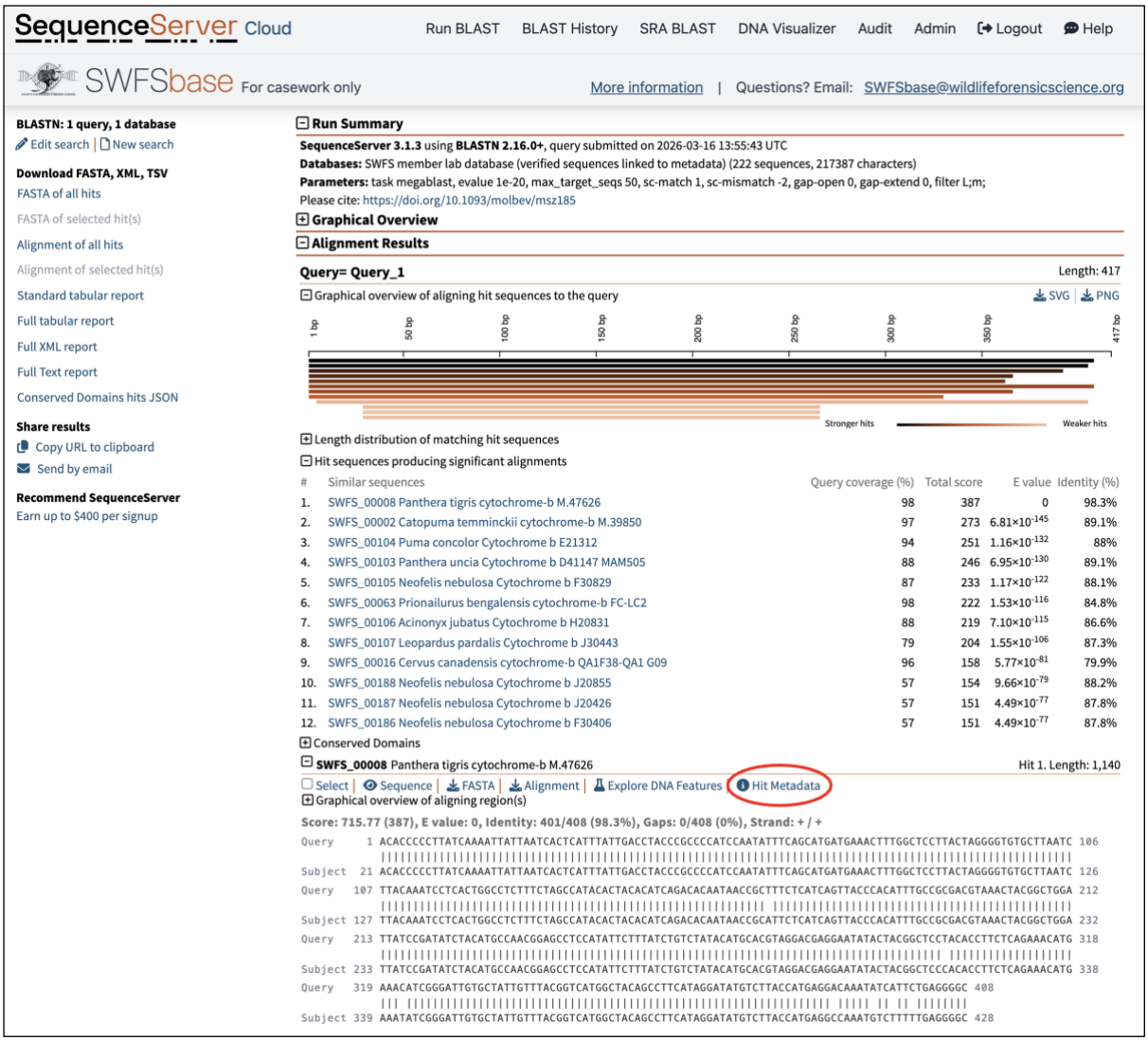
The ***'default'*** settings are appropriate for most searches, and are designed to quickly find highly similar sequences. If BLASTing against the '*NCBI nucleotide database*', it is useful to use the ***'discontiguous megablast excluding uncultured, environmental and unclassified'*** setting; this allows mismatches in the initial alignment seed, making the search more sensitive to distant sequence similarity, while excluding many low-quality or irrelevant sequences (searches may run slightly slower than the default preset).

Users may also click the 'Customize' and 'Advanced' settings to manually adjust BLAST parameters and filters.

Viewing results

The BLAST results interface in SWFSbase is largely similar to NCBI BLAST. Several result fields are displayed by default; some fields may be irrelevant for your analysis and can be folded away using the minus signs. Your browser cookies should remember which fields you have folded, so they will remain folded in subsequent searches.

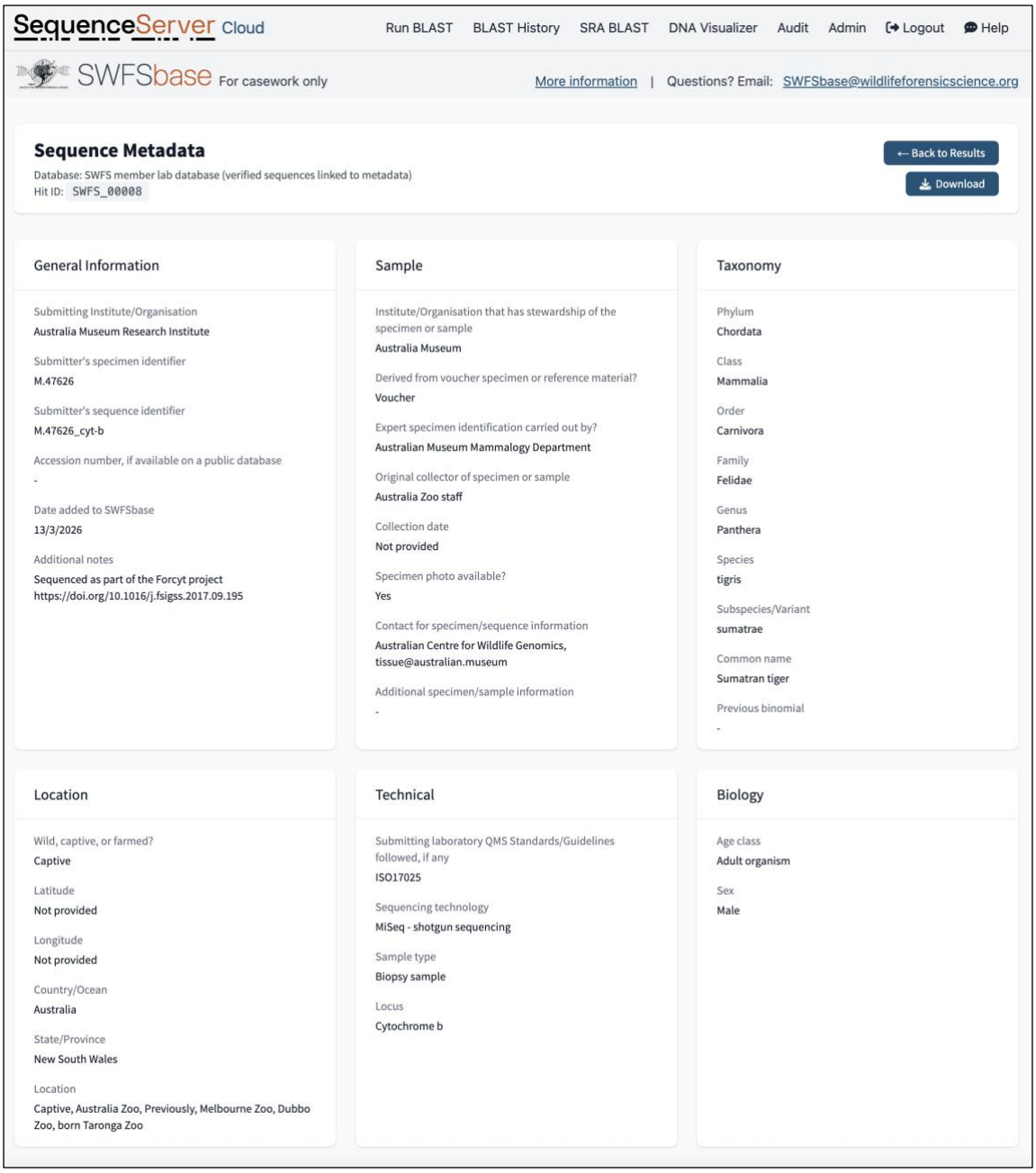
Below is an example of a BLAST result (the query is a tiger cytochrome-b sequence, searched against the '*SWFS member lab database*'), with some result fields already folded away:



The bottom of the screenshot displays the sequence alignment for the first sequence hit; scrolling down will show additional hits.

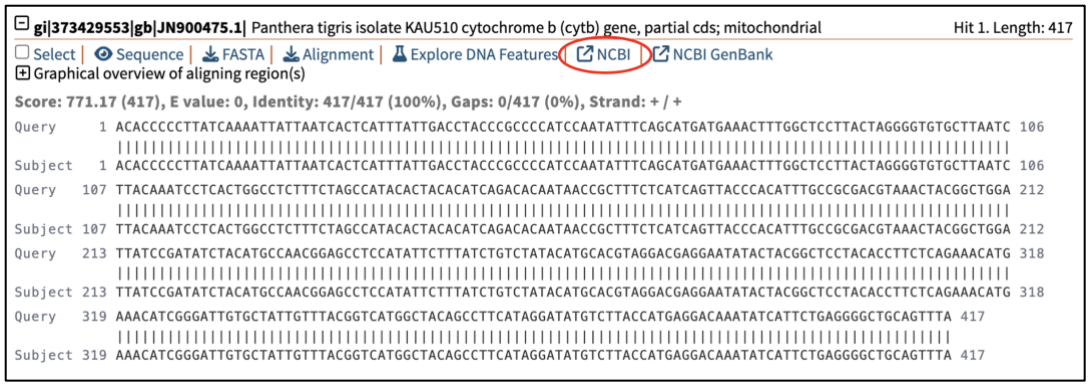
Metadata

If a sequence has metadata (as do all sequences in the '*SWFS member lab database*'), it will display a "Hit Metadata" link (circled in red above). Clicking this link opens the metadata interface, where you can view detailed information about the sequence. Below is the metadata for the top hit from the example search:



These metadata are intended to provide information about the reliability and provenance of the sequence hit, supporting interpretation of your results.

Sequences from the NCBI core nucleotide database are not linked to metadata within SWFSbase. To access additional information for these hits, you can click the NCBI link provided to view the sequence details on the NCBI website:



Submitting sequences to SWFSbase

Laboratories that contribute data to SWFSbase play a vital role in enhancing the platform and supporting its mission.

To contribute to the '*Known erroneous NCBI sequence*' database, please report any identified erroneous NCBI sequences to SWFSbase@wildlifeforensicscience.org.

To contribute your laboratory's sequence data and associated metadata suitable for forensic analysis, please follow the instructions below.

Each submission consists of two files: (1) a FASTA file containing sequences (.fasta, .fa, .fas, or .fsa), and (2) a metadata file prepared in Excel using the metadata template available on the SWFSbase webpage.

Metadata entry

Information about each metadata field is provided on the first tab of the template. Please contact the SWFSbase team if you are unsure about any field.

Linking metadata to your FASTA

A key field in the metadata is '***SWFSbase submission ID***', which links each metadata row to its corresponding sequence. Each row of metadata in the Excel sheet must have a unique 'SWFSbase submission ID' that matches the ***sequence ID*** in the FASTA file.

Note: *In a FASTA file, the sequence ID is the text immediately following the* **>** *symbol in the header, up to the first space. Anything after the first space is ignored.*

For example, the top row of this metadata excel template screenshot has a 'SWFSbase submission ID' **`ARS_0001.1`**:

	Submission details				
Required	Required	Required	Optional	Optional	Optional
SWFSbase submission ID	Submitting Institute/Organisation	Submitter's specimen identifier	Submitter's sequence identifier	Accession number, if available on a public database	Additional notes
ARS_00001.1	Australia Museum Resear	M.39850	M.39850_CO1		Sequenced as part of the Forcyt project https://doi.org/10.1016/j.fsigss.
ARS_00001.2	Australia Museum Resear	M.39850	M.39850_cyt-b		Sequenced as part of the Forcyt project https://doi.org/10.1016/j.fsigss.

This should match the associated sequence in the FASTA file (the full sequence isn't displayed here):

```
>ARS_0001.1 M.39850 CO1 Catopuma temmincki AustMus
ATGTTCATAAACCGCTGACTATTTTCAACCAATCATAAAGATATTGGTACTCTTTA...
```

Note, any labelling system can be used for your 'SWFSbase submission IDs', as long as they are unique within that submission. You may use existing sequence IDs from your database (in this case, the 'SWFSbase submission ID' might be the same as the 'Submitter's specimen identifier' and/or 'Submitter's sequence identifier').

Alternatively, you can assign sequential IDs for your sequences. In the example below, the sequences are both from specimen M.39850, but for different loci, so `seq01`, and `seq02` have been used as the 'SWFSbase submission IDs' in both the FASTA and metadata files:

```
>seq01 M.39850 CO1 Catopuma temmincki AustMus
ATGTTCATAAACCGCTGACTATTTTCAACCAATCATAAAGATATTGGTACTCTTTA...

>seq02 M.39850 cyt-b Catopuma temmincki AustMus
ATGACCAACATTCGAAAATCACACCCCCTTATTAAAATTATTAACCACTCATTCAT...
```

Important:

- Ensure that all 'SWFSbase submission IDs' are unique and match exactly between the metadata and sequence files.
- Contact the SWFSbase team if you are unsure about formatting or any field in the template.

Once your sequences and metadata are prepared:

- Open the submission form here: https://docs.google.com/forms/d/e/1FAIpQLSfKaROuBi3QilkBdS3xsBclujdTqFhFb304JcMC5LAVPeFIWA/viewform
- Fill in the required fields.
- Attach your FASTA sequence file and metadata Excel file to the form.
- Submit the form.

The SWFSbase team will review each submission and contact you if there are any issues