ATLANTIC AREA
Transnational Programme
INVESTING IN OUR COMMON FUTURE

European Union
European Regional Development Fund

STANDARD OPERATING PROCEDURE FOR THE GENETIC IDENTIFICATION OF FISH SPECIES USING DNA BARCODING (MITOCHONDRIAL CYTOCHROME-C-OXIDASE I SEQUENCING)

Prepared by the Labelfish Consortium

Labelfish®

December 2014

CSIC Consejo Superior de Investigaciones Científicas
ipma instituto português do mar e da atmosfera
Ifremer
Indigo Rock Marine Research Station
University of Salford Manchester
MRI Max Rubner-Institut Federal Research Institute of Nutrition and Food
BIM Bord Iascaigh Mhara Irish Sea Fisheries Board
SCL
Gobierno de España Ministerio de Agricultura, Alimentación y Medio Ambiente
defra Department for Environment Food and Rural Affairs
Gobierno de España Ministerio de Sanidad, Servicios Sociales e Igualdad

CONTENTS

1. BACKGROUND

The Labelfish project is an EU InterReg funded network of laboratories in the "Atlantic Area" of Europe, aiming to develop harmonised & standardise methods for the authentication of seafood products (www.labelfish.eu).

2. PURPOSE

The purpose of this SOP is to provide a genetic method for the identification of fish species, in order to support the implementation of food labelling/authenticity testing.

3. SCOPE

This method is suitable for the qualitative identification of DNA (deoxyribonucleic acid) in fish products. It has been tested against a very broad taxonomic range of fish species (but has failed in a small minority of cases, <5% of species tested; Ivanova *et al.*, 2007). The assay is designed to work with fresh, smoked, salted and frozen samples. It is also successful with cooked products, but success is dependent on the intensity of cooking. It is not suitable for highly processed foods e.g. tins of tuna. It is also unsuitable for the identification of complex fish products containing DNA from multiple species. For some species of relatively recent evolutionary origin, this method may only be able to identify the sample down to the genus level (e.g. some tunas of the genus *Thunnus*, or redfish of the *Sebastes* genus). In these cases, additional tests might be required for species level identification.

4. DEFINITIONS & ABBREVIATIONS

DNA: Deoxyribonucleic acid

PCR: Polymerase Chain Reaction

SOP: Standard Operating Procedure

UV: Ultraviolet

CO1/COI: Mitochondrial cytochrome c oxidase 1 gene

5. PRINCIPLE OF THE METHOD

The following is taken from the international Barcode of Life Project (http://www.barcodeoflife.org/);

"*Barcoding uses a very short genetic sequence from a standard part of the genome the way a supermarket scanner distinguishes products using the black stripes of the Universal Product Code (UPC). Two items may look very similar to the untrained eye, but in both cases the barcodes are distinct. The gene region that is being used as the standard barcode for almost all animal groups is a 648 base-pair region in the mitochondrial cytochrome c oxidase 1 gene ("CO1"). COI is proving highly effective in identifying many animal groups*".

6. MATERIALS & EQUIPMENT

The sections below report all the equipment and materials required to apply this protocol.

N.B. Batch numbers of kits used must be recorded.

6.1 Water

General use: Distilled or de-ionised water

PCR procedures: Sterile, DNase-, RNase- and Protease-free water e.g. Fisher Scientific DNA free water, product code: BPE2470-1

6.2 Solutions, standards and reference materials

The present SOP was validated using a ring trial based on 13 “blind” reference tissues (list of voucher specimens is held by the LABELFISH consortium). Details on the ring-trial procedure and results are available upon request to the LABELFISH consortium.

6.3 Commercial kits

DNA Extraction: The method has been validated using the ‘DNeasy Blood & Tissue Kit’ supplied by Qiagen (Product code 69504). DNA extraction kits from other suppliers must be shown to be appropriate before use.

6.4 Plastic ware

N.B. It is essential that all plastic-ware is sterile before use.

Item	Detail	Example Supplier	Product code
Pipette tips (filtered)	10, 20, 200 & 1000µl	Starlabs	S1120
PCR tubes	single, strip or 96-well	Starlabs	I1402
1.5ml tubes	1.5 ml	Starlabs	S1615

6.5 Equipment

The following items of equipment are required to undertake the analysis. Several alternative suppliers/models are available for each item. These must be shown to be appropriate before use.

Item	Detail	Example supplier	Product code
Precision pipettes	1-1000µl	Starlabs	G8900
Bench top vortex		Labnet	VX-100
Thermocycler	ABI Vereti 96 well		
Thermal mixer	to hold 1.5 ml tubes	Eppendorf	5355
DNA quantifier	Accurate to +/- 1 ng		ND1000
Microcentrifuge	to hold 1.5 ml tubes	Eppendorf	5452

Optional – laminar flow hood

6.6 Other materials

Disposable plastic gloves, sterile dissection equipment.

6.7 Electronic files / computer software

A computer with a text editor e.g. notepad.

Freely available sequence editing software e.g. Bioedit, FinchTV, ProSeq.

Internet access is required to utilise the Barcode of Life System: http://www.boldsystems.org/

7. PROCEDURES

It is essential to wear disposable plastic gloves during all laboratory procedures and to use pipette tips that are sterile and fitted with filters.

7.1 Sample preparation

All samples should be stored frozen at -20°C until processed. Samples can be stored frozen indefinitely.

N.B. In the ring trial ethanol-preserved samples were utilised.

The external surfaces of samples submitted for analysis may have been affected through preservation treatments or bacterial breakdown. Where possible, obtain subsamples for DNA extraction from the least degraded area of tissue in order to minimise contaminant DNA and DNA degradation. This will typically mean removing outer layers of tissue in contact with the environment before taking a subsample. Use sterile dissection equipment where appropriate.

7.2 DNA Extraction

Materials:

The extraction should be carried out with the Qiagen DNeasy Blood & Tissue Kit, following the manufacturer's protocol. It is recommended that the manufacturer's guidelines are checked each time kits are ordered to ensure any updates/changes made since development of this SOP are incorporated.

Procedure:

1. Cut up approx. 25 mg tissue into small pieces and place into a 1.5 ml tube.
2. Include an empty 1.5 ml tube as an extraction control. This is treated following the same procedure and carried through to the PCR stage (7.3).
3. Add 180 µl Buffer ATL (tissue lyser).
4. Add 20 µl proteinase K and vortex for 15 seconds.
5. Incubate in a thermal mixer at 56°C for 2 hours.
6. Vortex for 15 seconds.
7. Add 200 µl Buffer AL (cell lyser) to the sample and vortex for 15 seconds.
8. Add 200 µl 100% ethanol and vortex for 15 seconds.
9. Pipette the mixture into a DNeasy Mini spin column placed in a 2 ml collection tube.
10. Centrifuge at 8000 rpm for 1 minute
11. Discard the eluate and replace the collection tube.
12. Add 500 µl Buffer AW1 (wash 1).
13. Centrifuge at 8000 rpm for 1 minute
14. Discard the eluate and replace the collection tube.

15. Add 500 µl Buffer AW2 (wash 2).

16. Centrifuge at 13,000 rpm for 3 minutes

17. Discard the eluate and collection tube. Place the spin column in a 1.5 ml tube.

18. Pipette 100 µl Buffer AE (elution) directly onto the spin column membrane.

19. Incubate at room temperature for 1 minute

20. Centrifuge at 8000 rpm for 1 minute to elute DNA.

21. Discard the spin column, close the tube and store the eluate containing DNA at 4°C for up to one week or in a freezer (-20°C) long term.

22. DNA extract quantification. Extracted DNA must be quantified to assess the extraction process and enable normalisation of DNA concentration. One common method is to use a Nanodrop ND 1000 Spectrophotometer. DNA should be diluted to 10-50ng/µl using DNA-free water. Negative controls should read ~0 ng/µl.

Controls:

A negative extraction control (with no tissue) should be run in parallel with all batches of sample extraction and quantified alongside all tissue extractions.

7.3 PCR Amplification

Materials:

BIOTAQ DNA polymerase 500Units (Bioline Catalogue number BIO-21040, also contains reaction buffer & $MgCl_2$)

dNTP mix 10mM final concentration (Bioline Catalogue number BIO-39053, each dNTP at 2.5mM concentration)

Procedure:

1. Create a sample plan (ideally in Excel) describing the DNA being analysed and it's locations in the rack/plate.

2. Organise your DNA extractions (i.e. defrost, if necessary) according to the plan.

3. Alongside every set of reactions ensure a negative control (i.e. ultra pure water) and a positive control (*this can be determined internally in each lab, but the DNA must have originated from a fish for which the species has been accurately identified, or previously experimentally determined via COI sequencing; i.e. it needs to have successfully been PCR amplified previously*) are included.

4. Make up the primers to a 0.01 mM (i.e. 10 pM/µL) concentration.

Primers:

Primer Name	**Primer sequence (5'-3')**	**References**
VF2_t1	TGTAAAACGACGGCCAGTCAACCAACCACAAAGACATTGGCAC	Ward *et al.* 2005
FishF2_t1	TGTAAAACGACGGCCAGTCGACTAATCATAAAGATATCGGCAC	Ward *et al.* 2005
FishR2_t1	CAGGAAACAGCTATGACACTTCAGGGTGACCGAAGAATCAGAA	Ward *et al.* 2005

FR1d_t1	CAGGAAACAGCTATGACACCTCAGGGTGTCCGAARAAYCARAA	Ivanova *et al.* 2007

5. Prepare the PCR reactions as follows (this following recipe is enough for 1 reaction and requires multiplication for the number of samples being analysed, in order to account for pipetting error it is also recommended to add 10% to the total volume of each of the reagents utilised);

PCR master mix, per reaction with a total volume 20 µl;

10 µL of 10% trehalose (e.g. Sigma-Aldrich, catalogue number T-5251)

2.7 µL of ultra pure water

2 µL 10×reaction buffer

1 µL $MgCl_2$ (50 mM)

0.2 µL of each primer (0.01 mM)

0.4 µL of the Bioline 10mM dNTP mix

0.1 µL of BIOTAQ *Taq* DNA Polymerase

6. Vortex master mix thoroughly.

7. Place 17 µL of the master mix into every tube/well (can use the same pipette tip during this step).

8. Aliquot 3 µl of DNA template to each tube/well following your sample plan.

Reagent	**Per Reaction**
10% trehalose	10
ddH2O	2.7
10X buffer	2
50mM MgCl2	1
Primer VF2_t1	0.2
Primer FishF2_t1	0.2
Primer FishR2_t1	0.2
Primer FR1d_t1	0.2
dNTPs 10mM total mix	0.4
Taq	0.1
TOTAL	**17**

9. Thermal conditions for the PCR reaction are; 94°C for 2 min, 35 cycles of 94°C for 30 sec, 52°C for 40 sec, and 72°C for 1 min, with a final extension at 72°C for 10 min (the “hot lid” option should also be selected).

10. Place the tubes/plate in the PCR machine and run the PCR programme.

11. Once completed the PCR reactions can be stored in the fridge at 4°C. But for long term storage (i.e. great than a week) freezing at -20°C is recommended.

7.4 PCR Product Check

Gel electrophoresis of DNA in an agarose gel is a standard technique in molecular biology, but equipment, reagents, staining and visualisation varies considerably between laboratories, and according to local health & safety controls. Therefore, this SOP suggests general conditions that need to be adapted to each laboratory.

1. Make a 1-2% agarose gel

2. Once set, load 4 µL of the PCR product into the well (the addition of loading buffer/dye may be necessary).

3. Include appropriate size standard in one lane (e.g. 5 µl Bioline hyperladder 100, catalogue number BIO-33056).

4. Run at 100V for approximately 1 hr (depending on size of gel), ensuring the DNA does not run off the gel.

5. Visualise your DNA fragments in UV light (with appropriate safety precautions); if the PCR reaction has been successful the positive control will have a single bright band of approximately 700 base pairs in length. Your negative controls should not contain bands. A band in the lanes corresponding to your samples indicates successful amplification.

6. Keep a permanent record of your gel (electronic and/or hard copy) as proof that the PCR amplification was successful and contaminant free.

7.5 DNA Sequencing

For this SOP it is assumed that the majority of laboratories do not have access to Sanger sequencing equipment in-house, therefore it is recommended that the PCR products are sent to an external company for PCR clean up and sequencing reaction. The requirements for the sequencing services vary, especially in terms of the volume and concentration of PCR product and sequencing primers required. This needs to be checked specifically with your preferred service provider.

1. Estimate concentration of your PCR product. This can be done from the record you made of your PCR products when run on the agarose gel, by comparing the brightness of the bands to the size standard that was run (that has a standard concentration of DNA). This information is usually required by the sequencing service.

2. When placing an order for sequencing it is important to make clear that for each PCR product two sequencing reactions are required; one utilising the forward primer and a second utilising the reverse primer (so for each sample two complementary sequences will be obtained).

3. Ensure the PCR products are cleaned before the sequencing reaction is attempted. This can usually be completed by the external sequencing company (but there are protocols/kits to do this in-house e.g. ExoSAP-IT- USB Corporation; Cleveland, OH Cat. No. 78201).

4. Send your carefully labelled PCR products and sequencing reaction primers to the sequencing service, according to their instructions. The sequencing primers differ from those used in the PCR amplification and are detailed below;

Primer Name	**Primer sequence (5'-3')**
M13F (−21)	TGTAAAACGACGGCCAGT
M13R (−27)	CAGGAAACAGCTATGAC

Additional Resources;

A protocol developed by the consortium for the barcode of life is available below and deals with procedures 7.1 – 7.5 in greater detail, providing some useful background information and potential troubleshooting:

http://www.barcodeoflife.org/sites/default/files/Protocols_for_High_Volume_DNA_Barcode_Analysis.pdf

7.6 Raw Data Processing

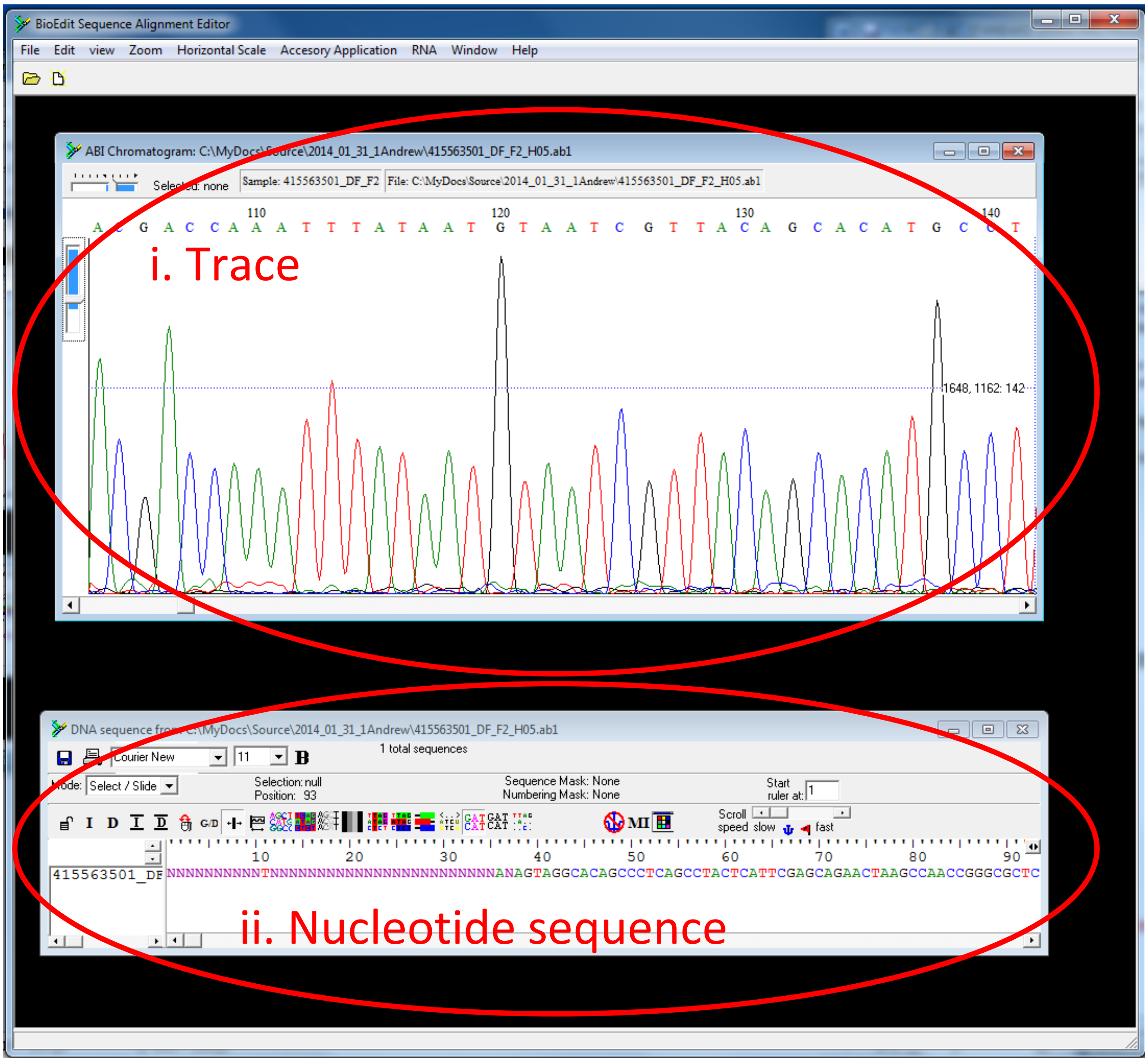
Sequencing services usually supply the results in a range of files, but it is the ABI data file (.abi) required in the SOP (it is important to ensure the sequencing company will supply these before making an order, but it is usually standard). The raw data needs to be checked and edited before it can be used.

The ABI files can be viewed and edited with a number of freely available software packages (mentioned in section 6.7). This SOP has been tested using BioEdit, which can be downloaded from the following webpage: http://www.mbio.ncsu.edu/bioedit/bioedit.html

1. Open the BioEdit software by clicking on the BioEdit.exe icon

2. Open the ABI file from your sample by selecting the file menu and the open option. Select the ABI sample from your sample

3. This will open up two windows within the software; (i.) The chromatogram, i.e. the sequence trace or peaks corresponding to the signal from each of the nucleotides in the DNA sequence; (ii.) A long string of letters, predominantly made up of A, T, C, & G, which correspond to the software's interpretation of the peaks and conversion into a representative nucleotide sequence;



ABI Chromatogram: C:\MyDocs\Source\2014_01_31_1Andrew\415563501_DF_F2_H05.ab1

Selected: none

Sample: 415563501_DF_F2

File: C:\MyDocs\Source\2014_01_31_1Andrew\415563501_DF_F2_H05.ab1

i. Trace

DNA sequence from C:\MyDocs\Source\2014_01_31_1Andrew\415563501_DF_F2_H05.ab1

1 total sequences

Mode: Select / Slide

Selection: null
Position: 93

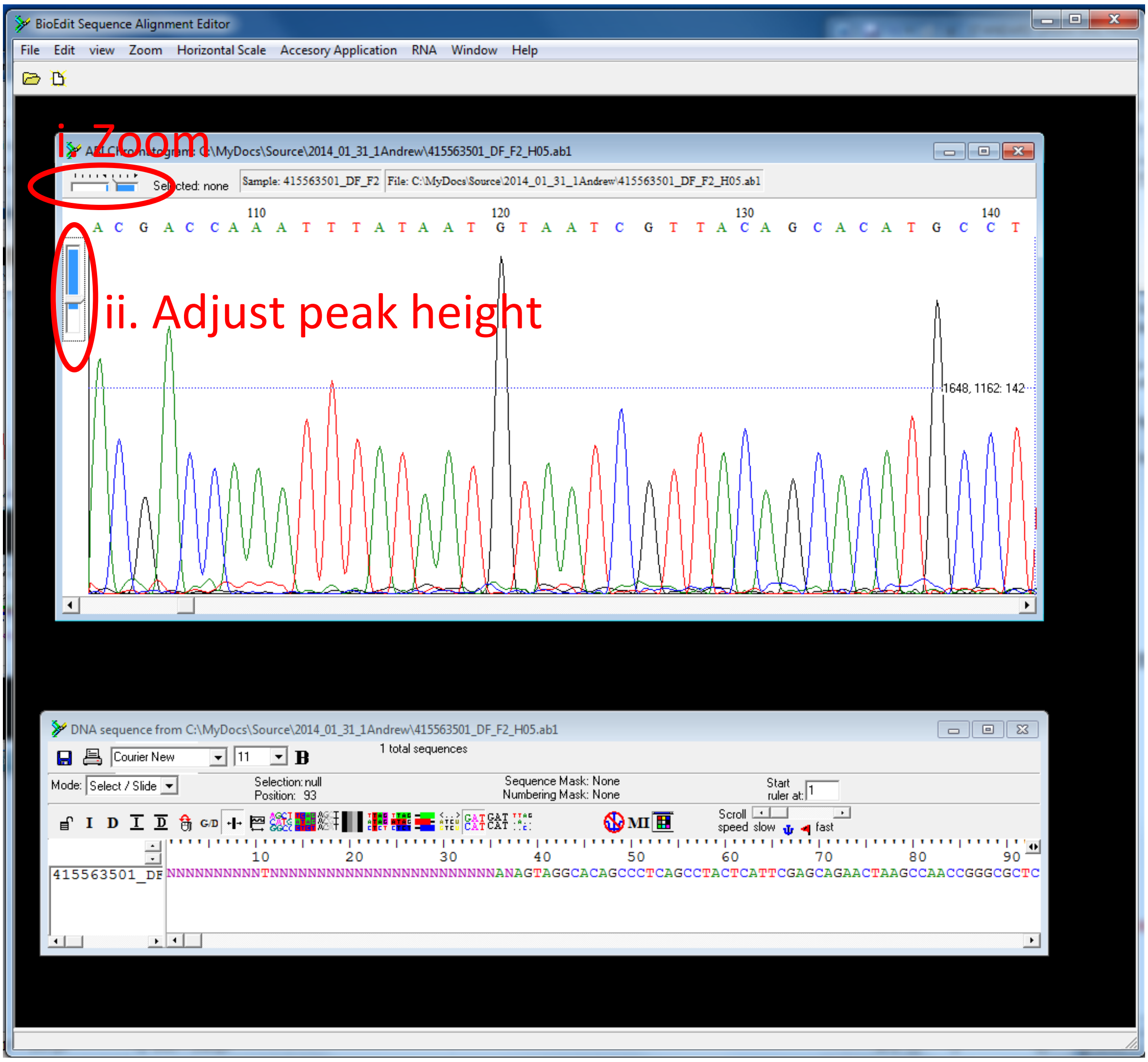
Sequence Mask: None
Numbering Mask: None

Start ruler at: 1

```
415563501_DF NNNNNNNNNNTNNNNNNNNNNNNNNNNNNNNNNNNANAGTAGGCACAGCCCTCAGCCTACTCATTCGAGCAGAACTAAGCCAACCGGGCGCTC
```

ii. Nucleotide sequence

4. In order to optimise the view of the trace within the software, both (i.) zoom and (ii.) relative peak height function are present that can be adjusted to your preference;



i. Zoom

ii. Adjust peak height

BioEdit Sequence Alignment Editor

ABI Chromatogram: C:\MyDocs\Source\2014_01_31_1Andrew\415563501_DF_F2_H05.abl

Selected: none

Sample: 415563501_DF_F2

File: C:\MyDocs\Source\2014_01_31_1Andrew\415563501_DF_F2_H05.abl

1648, 1162: 142

DNA sequence from C:\MyDocs\Source\2014_01_31_1Andrew\415563501_DF_F2_H05.ab1

1 total sequences

Mode: Select / Slide

Selection: null
Position: 93

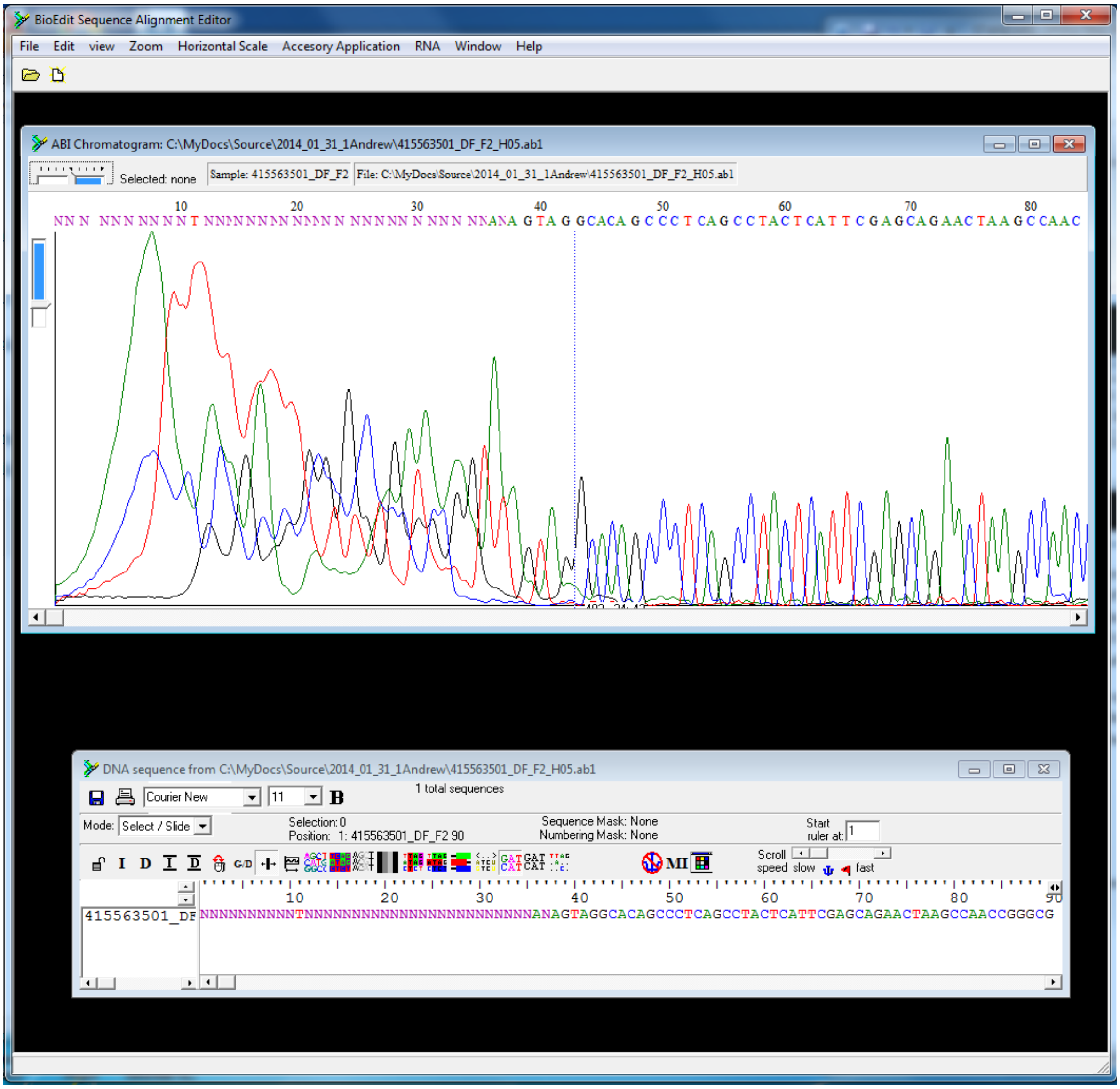
Sequence Mask: None
Numbering Mask: None

Start ruler at: 1

```
415563501_DF NNNNNNNNNNTNNNNNNNNNNNNNNNNNNNNNNNNANAGTAGGCACAGCCCTCAGCCTACTCATTCGAGCAGAACTAAGCCAACCGGGCGCTC
```

5. The sequence must be checked by eye to ensure the sequence reaction has worked successfully and the trace is of high quality. The majority of the trace should consist of a series of clear peaks (as in the figure above.) If the reaction has failed, or contamination is present, the peaks will look weak and/or it will be impossible to clearly resolve a single peak at each nucleotide position. If this is the case the results are not high quality enough for use.

6. Often the quality of the sequencing reaction is poor at either ends of the trace (as below). In this case the ambiguous region at either end can simply be deleted, just leaving the high quality sequence (i.e. delete the flanking sequence at each end until you are confident that you can easily call each peak). *In the example below unambiguous peaks appear approximately after nucleotide position 42.*



ABI Chromatogram: C:\MyDocs\Source\2014_01_31_1Andrew\415563501_DF_F2_H05.abl

Selected: none

Sample: 415563501_DF_F2

File: C:\MyDocs\Source\2014_01_31_1Andrew\415563501_DF_F2_H05.ab1

```
NN N  NNN NN N N T NNNNNNNNNNNNN N NNNNN N NNN NNANA G TA G GCACA G C CC T CAG C C TAC T CAT T C GAGCAGAAC TAA G C CAAC
```

DNA sequence from C:\MyDocs\Source\2014_01_31_1Andrew\415563501_DF_F2_H05.ab1

Courier New 11

1 total sequences

Mode: Select / Slide

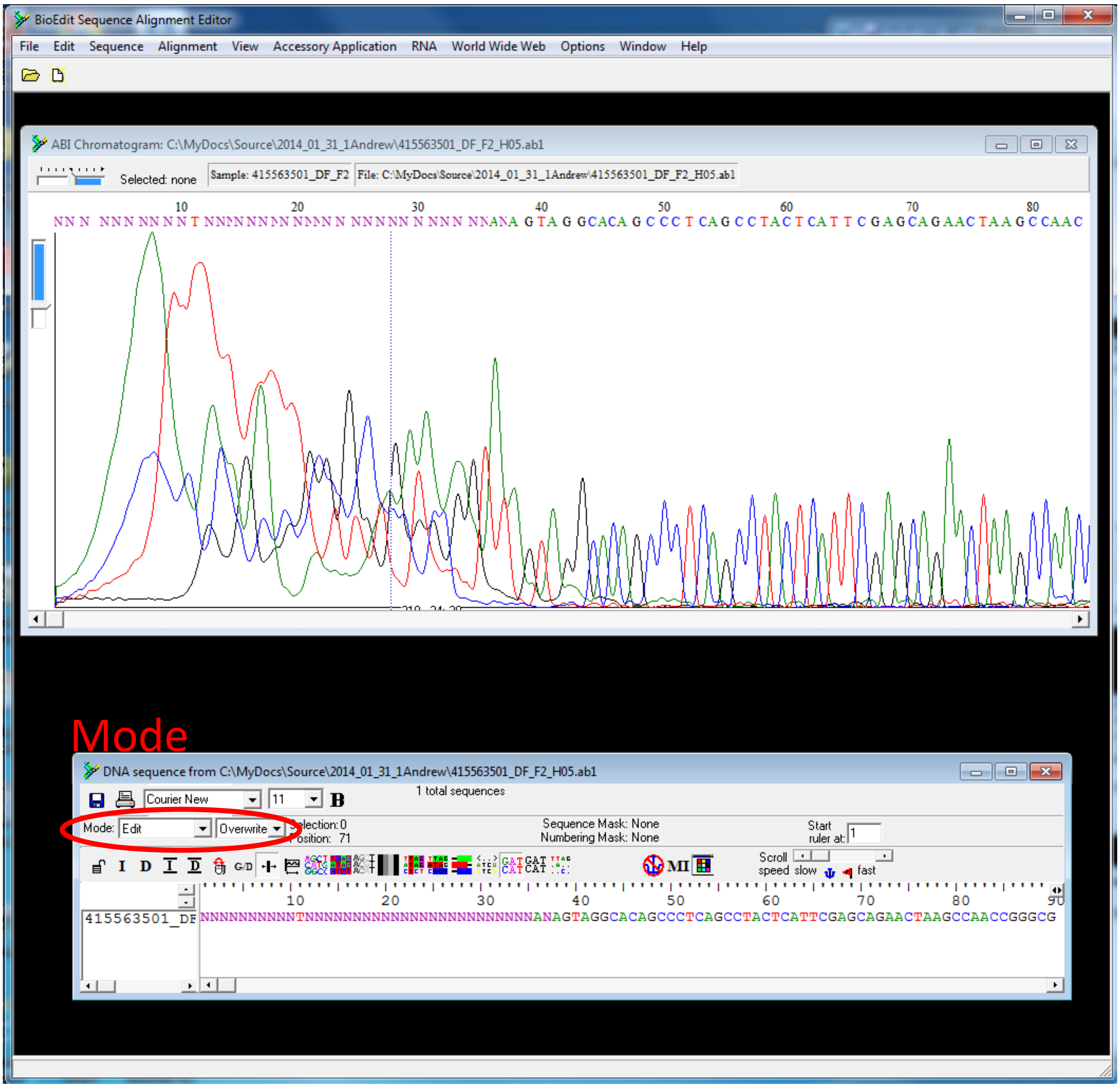
Selection: 0
Position: 1: 415563501_DF_F2 90

Sequence Mask: None
Numbering Mask: None

Start ruler at: 1

```
415563501_DF NNNNNNNNNNTNNNNNNNNNNNNNNNNNNNNNNNNNANAGTAGGCACAGCCCTCAGCCTACTCATTCGAGCAGAACTAAGCCAACCGGGCG
```

7. These low quality portions of the sequence at either end can be removed in the nucleotide sequence window. First it is necessary to switch the mode to edit, as indicated below, and then the sequence can be edited like any other text file. However, it is important to remember that the trace and sequence windows within BioEdit operate independently, so alterations to the sequence are not reflected in the trace window (meaning any edits to the sequence text will mean the nucleotide positions in the two windows will no longer match).



ABI Chromatogram: C:\MyDocs\Source\2014_01_31_1Andrew\415563501_DF_F2_H05.ab1

Selected: none

Sample: 415563501_DF_F2

File: C:\MyDocs\Source\2014_01_31_1Andrew\415563501_DF_F2_H05.ab1

```
                  10                  20                  30                  40                  50                  60                  70                  80
NN N  NNN NN N N T NNNNNNNNNN NNNN N NNNNN N NNN NNANA G TA G GCACA G C CC T CAG C C TAC T CAT T C GAGCAGAAC TAA G C CAAC
```

DNA sequence from C:\MyDocs\Source\2014_01_31_1Andrew\415563501_DF_F2_H05.ab1

Courier New 11

1 total sequences

Mode: Edit Overwrite

Selection: 0
Position: 71

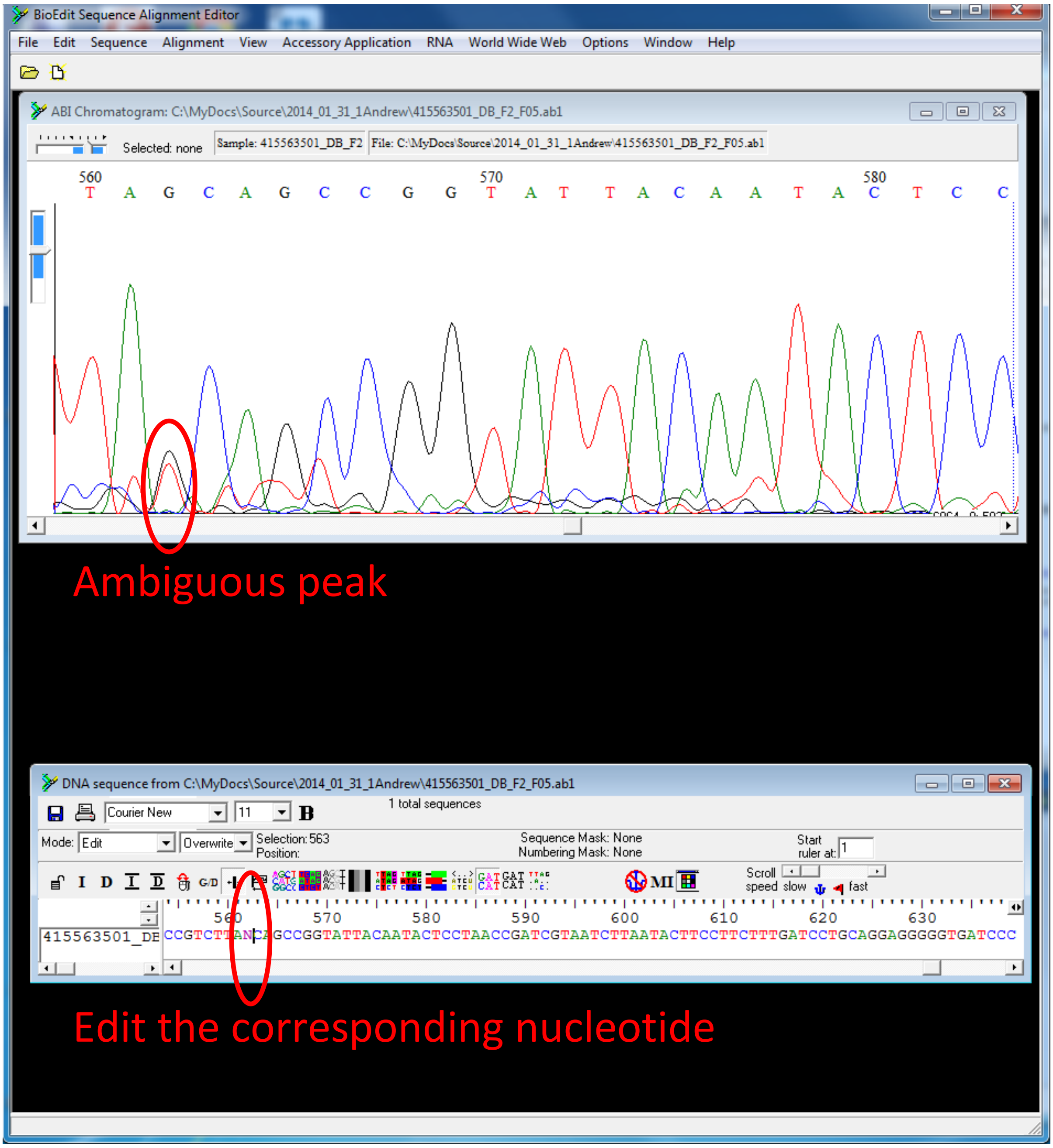
Sequence Mask: None
Numbering Mask: None

Start ruler at: 1

Scroll speed slow fast

```
                         10        20        30        40        50        60        70        80        90
415563501_DF NNNNNNNNNNTNNNNNNNNNNNNNNNNNNNNNNNNANAGTAGGCACAGCCCTCAGCCTACTCATTCGAGCAGAACTAAGCCAACCGGGCG
```

8. Ambiguous nucleotides may also arise within the middle of the sequences (see below). Provided the sequence window is in the editing mode, these can simply be overwritten with the letter “N” (indicating uncertainty about the call). **Do not delete them**.


BioEdit Sequence Alignment Editor
File Edit Sequence Alignment View Accessory Application RNA World Wide Web Options Window Help
ABI Chromatogram: C:\MyDocs\Source\2014_01_31_1Andrew\415563501_DB_F2_F05.ab1
560
570
580
T A G C A G C C G G T A T T A C A A T A C T C C
DNA sequence from C:\MyDocs\Source\2014_01_31_1Andrew\415563501_DB_F2_F05.ab1
Courier New
11
B
Scroll speed slow fast
560 570 580 590 600 610 620 630

Selected: none

Sample: 415563501_DB_F2

File: C:\MyDocs\Source\2014_01_31_1Andrew\415563501_DB_F2_F05.ab1

1 total sequences

Mode: Edit

Overwrite

Selection: 563

Position:

Sequence Mask: None

Numbering Mask: None

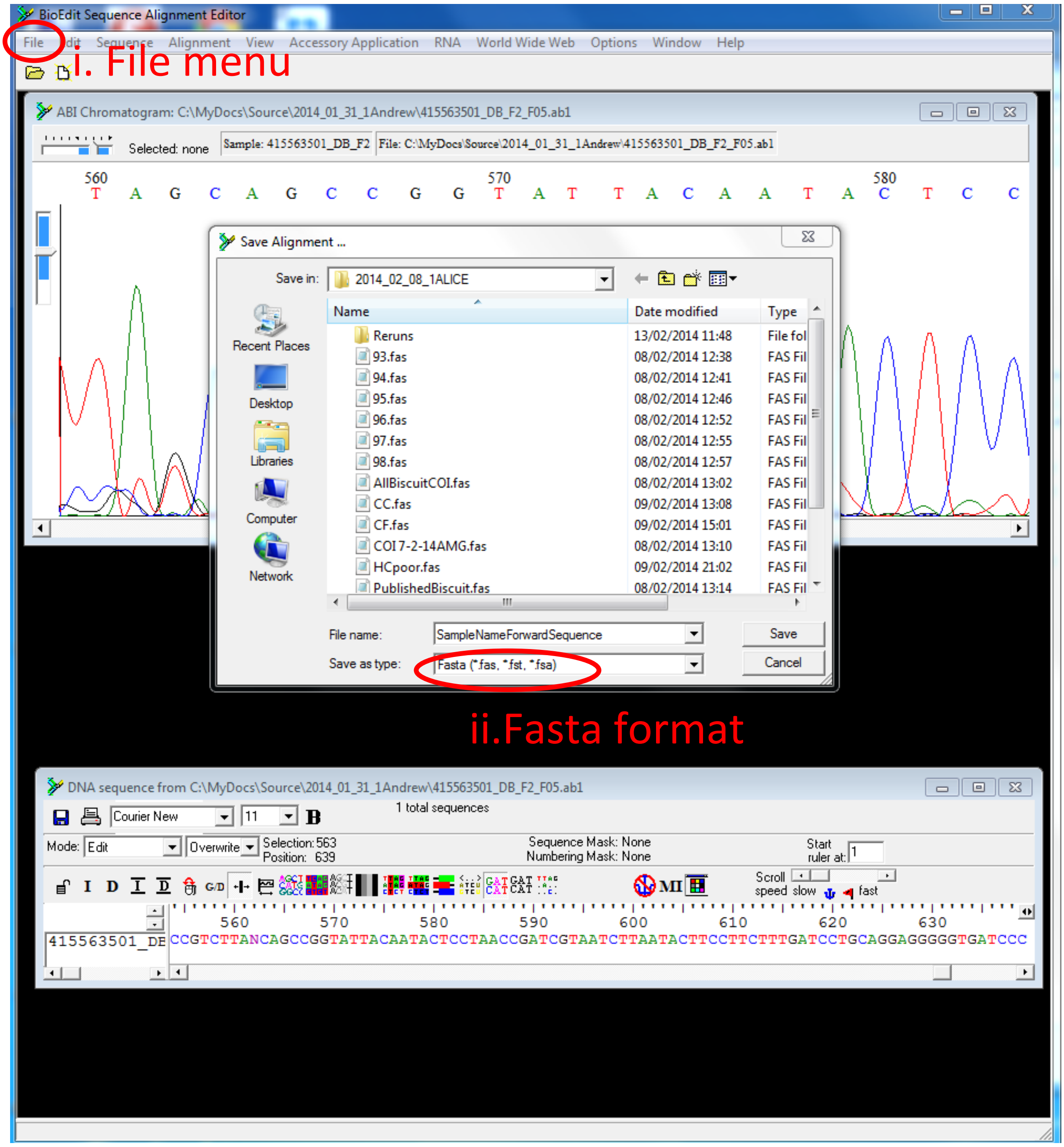
Start ruler at: 1

```
415563501_DE CCGTCTTANCAGCCGGTATTACAATACTCCTAACCGATCGTAATCTTAATACTTCCTTCTTTGATCCTGCAGGAGGGGGTGATCCC
```

Ambiguous peak

Edit the corresponding nucleotide

9. Once only high quality sequence remains, it is necessary to click on the nucleotide sequence window (so it is selected as the active window) and save the sequence. This is done in the file menu (i.) in the uppermost toolbar and selecting “Save As”. The file can be renamed (e.g. with the name of the original sample with indication as to whether the sequence was generated with the forward or reverse primer) and must be saved in fasta format (as ii. below):


BioEdit Sequence Alignment Editor
File
Edit
Sequence
Alignment
View
Accessory Application
RNA
World Wide Web
Options
Window
Help
ABI Chromatogram: C:\MyDocs\Source\2014_01_31_1Andrew\415563501_DB_F2_F05.ab1
560
570
580
T A G C A G C C G G T A T T A C A A T A C T C C
Save Alignment ...
Recent Places
Desktop
Libraries
Computer
Network
Save
Cancel
DNA sequence from C:\MyDocs\Source\2014_01_31_1Andrew\415563501_DB_F2_F05.ab1
Scroll speed slow fast
560
570
580
590
600
610
620
630

i. File menu

Selected: none

Sample: 415563501_DB_F2

File: C:\MyDocs\Source\2014_01_31_1Andrew\415563501_DB_F2_F05.ab1

Save in: 2014_02_08_1ALICE

Name	Date modified	Type
Reruns	13/02/2014 11:48	File fol
93.fas	08/02/2014 12:38	FAS Fil
94.fas	08/02/2014 12:41	FAS Fil
95.fas	08/02/2014 12:46	FAS Fil
96.fas	08/02/2014 12:52	FAS Fil
97.fas	08/02/2014 12:55	FAS Fil
98.fas	08/02/2014 12:57	FAS Fil
AllBiscuitCOI.fas	08/02/2014 13:02	FAS Fil
CC.fas	09/02/2014 13:08	FAS Fil
CF.fas	09/02/2014 15:01	FAS Fil
COI7-2-14AMG.fas	08/02/2014 13:10	FAS Fil
HCpoor.fas	09/02/2014 21:02	FAS Fil
PublishedBiscuit.fas	08/02/2014 13:14	FAS Fil

File name: SampleNameForwardSequence

Save as type: Fasta (*.fas, *.fst, *.fsa)

ii.Fasta format

Courier New 11

1 total sequences

Mode: Edit Overwrite

Selection: 563
Position: 639

Sequence Mask: None
Numbering Mask: None

Start ruler at: 1

```
415563501_DE CCGTCTTANCAGCCGGTATTACAATACTCCTAACCGATCGTAATCTTAATACTTCCTTCTTTGATCCTGCAGGAGGGGGTGATCCC
```

Additional Resources;

The Barcode of life systems page has an excellent description for assembling and editing sequences and the common errors that can arise:

http://www.boldsystems.org/index.php/resources/handbook?chapter=7_validation.html

Although not part of the SOP, it is also possible to get a free BOLD Systems account and upload ABI trace files onto the workspace, where the system can make an automated check of the quality of your sequence – see trace submission in the BOLDsystems handbook;

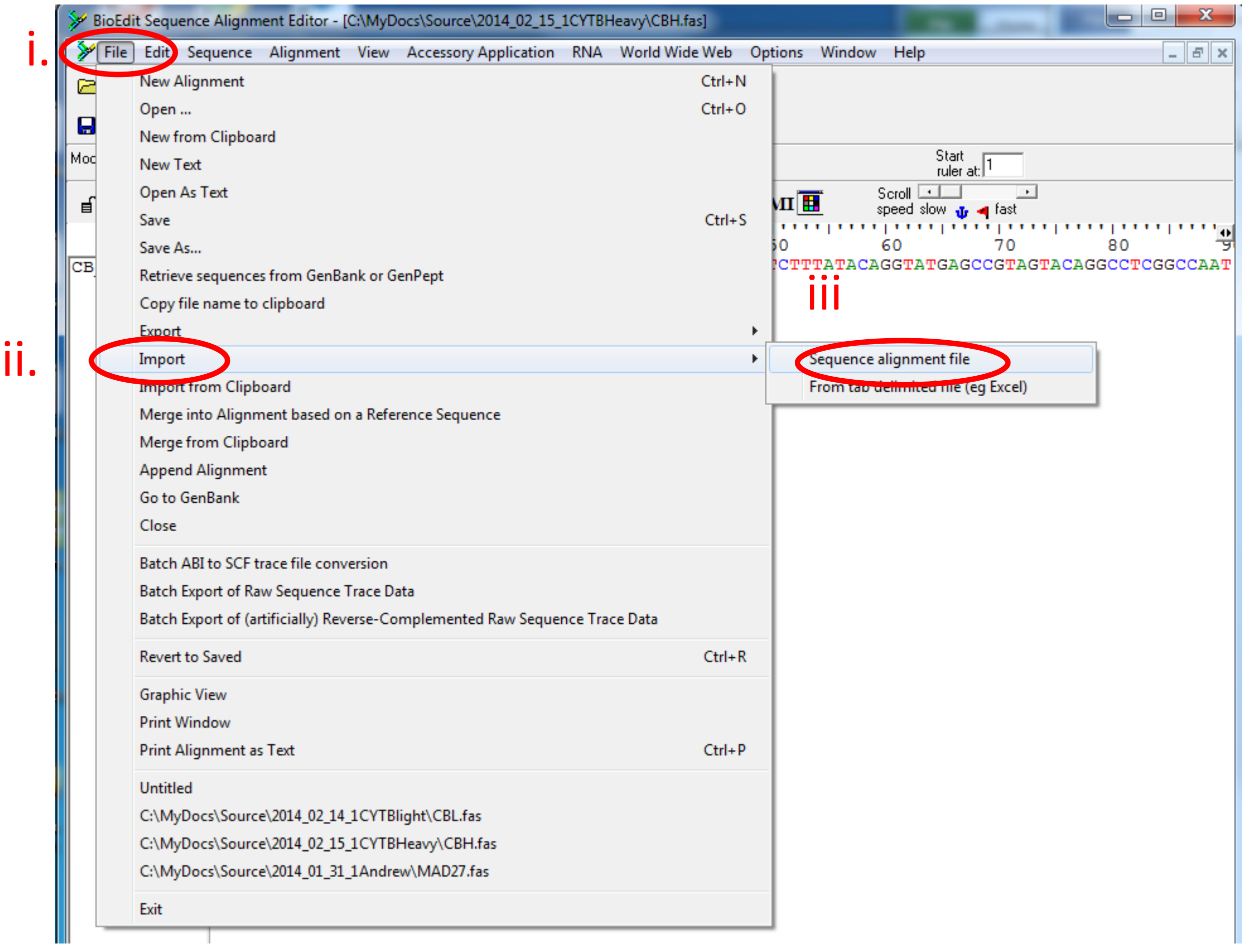
http://www.boldsystems.org/index.php/resources/handbook?chapter=3_submissions.html&section=trace_submissions

7.7 Generating a consensus sequence

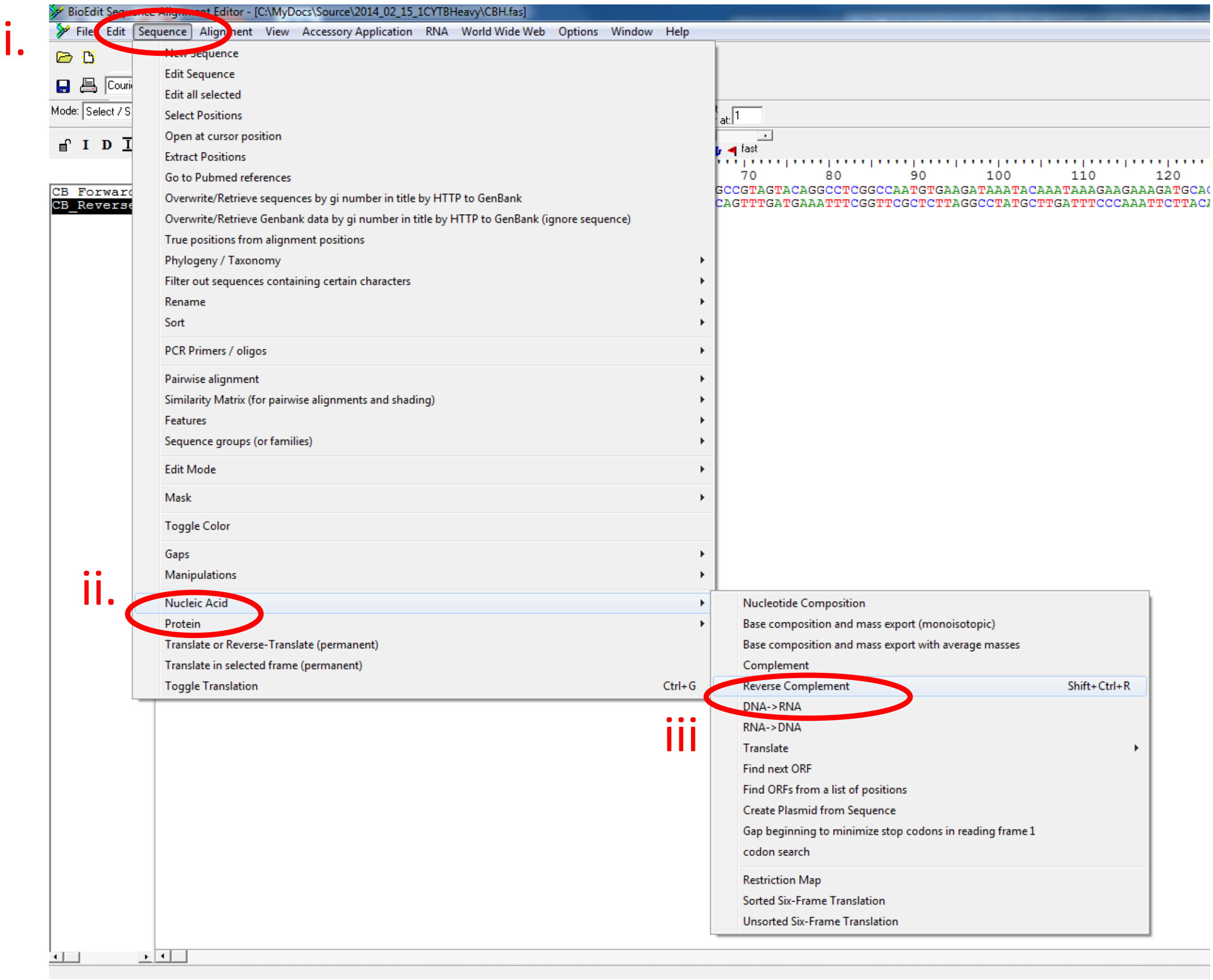
Each of the samples should have been sequenced in both the forward and reverse directions, meaning these complementary/overlapping sequences can be combined into a consensus. This serves as an important way of checking the accuracy of the sequence, and can help remove any ambiguous bases and generate a longer total sequence.

1. Start the BioEdit software by clicking on the BioEdit.exe icon and open the edited forward fasta files generated from the sample (as in 7.6). This will only open a nucleotide sequence window (there will be no trace window).

2. It is then necessary to import the reverse fasta file into the software. This is done in the file menu in the uppermost toolbar and selecting import, then sequence alignment file and locating your complementary reverse sequence fasta file.

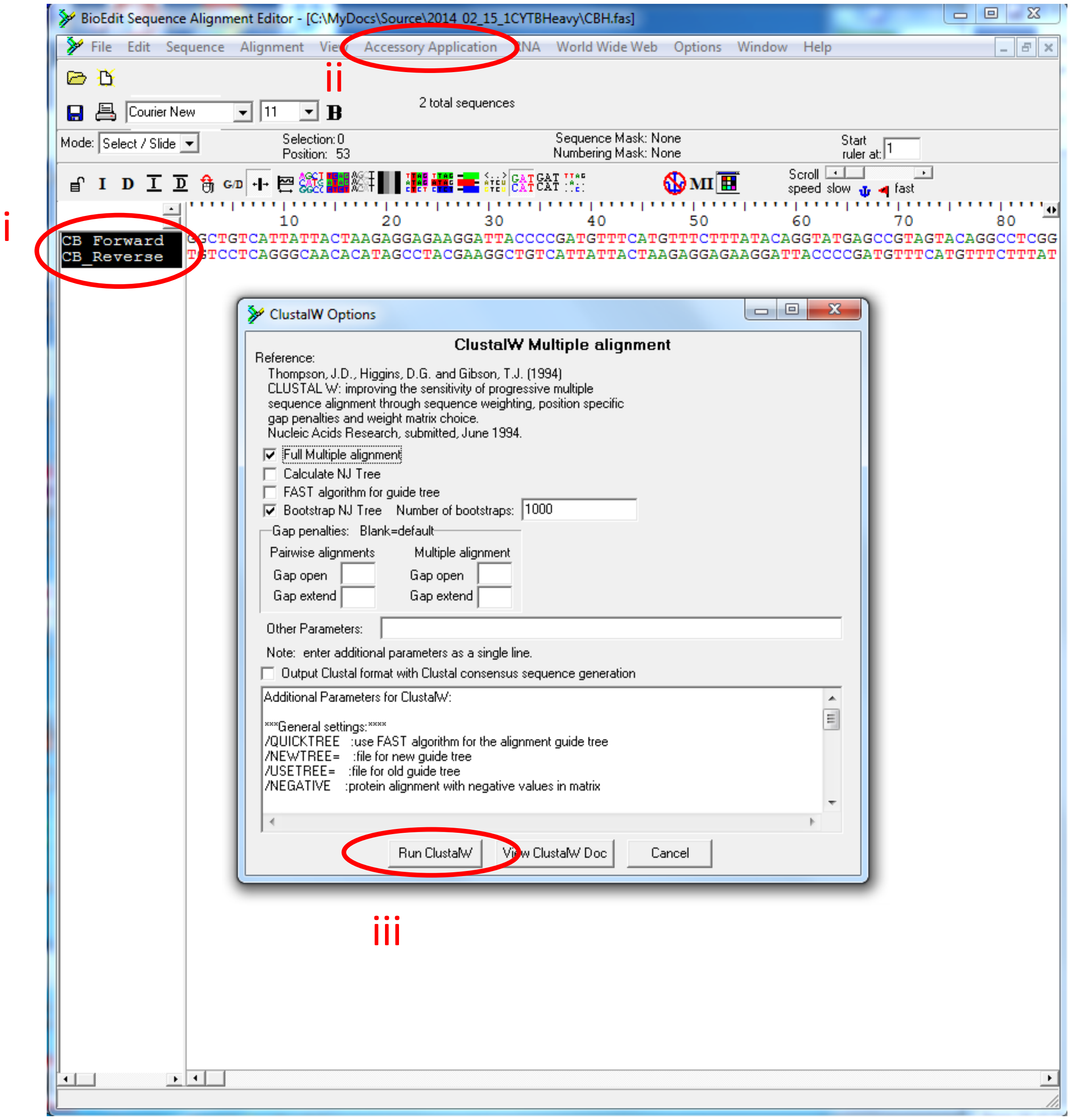

i.
ii.
iii
BioEdit Sequence Alignment Editor - [C:\MyDocs\Source\2014_02_15_1CYTBHeavy\CBH.fas]
File Edit Sequence Alignment View Accessory Application RNA World Wide Web Options Window Help
New Alignment Ctrl+N
Open ... Ctrl+O
New from Clipboard
New Text
Open As Text
Save Ctrl+S
Save As...
Retrieve sequences from GenBank or GenPept
Copy file name to clipboard
Export
Import
Import from Clipboard
Merge into Alignment based on a Reference Sequence
Merge from Clipboard
Append Alignment
Go to GenBank
Close
Batch ABI to SCF trace file conversion
Batch Export of Raw Sequence Trace Data
Batch Export of (artificially) Reverse-Complemented Raw Sequence Trace Data
Revert to Saved Ctrl+R
Graphic View
Print Window
Print Alignment as Text Ctrl+P
Untitled
C:\MyDocs\Source\2014_02_14_1CYTBlight\CBL.fas
C:\MyDocs\Source\2014_02_15_1CYTBHeavy\CBH.fas
C:\MyDocs\Source\2014_01_31_1Andrew\MAD27.fas
Exit
Sequence alignment file
From tab delimited file (eg Excel)
Start ruler at: 1
Scroll speed slow fast
60 70 80
CTTTATACAGGTATGAGCCGTAGTACAGGCCTCGGCCAAT

3. Select the reverse sequence within the nucleotide sequence window, just by clicking on its name on the far left. Then, in the sequence menu in the uppermost toolbar, select Nuleic Acid, followed by Reverse Complement.


i.
ii.
iii
BioEdit Sequence Alignment Editor - [C:\MyDocs\Source\2014_02_15_1CYTBHeavy\CBH.fas]
File
Edit
Sequence
Alignment
View
Accessory Application
RNA
World Wide Web
Options
Window
Help
Mode: Select / S
CB_Forward
CB_Reverse
fast
70
80
90
100
110
120
New Sequence
Edit Sequence
Edit all selected
Select Positions
Open at cursor position
Extract Positions
Go to Pubmed references
Overwrite/Retrieve sequences by gi number in title by HTTP to GenBank
Overwrite/Retrieve Genbank data by gi number in title by HTTP to GenBank (ignore sequence)
True positions from alignment positions
Phylogeny / Taxonomy
Filter out sequences containing certain characters
Rename
Sort
PCR Primers / oligos
Pairwise alignment
Similarity Matrix (for pairwise alignments and shading)
Features
Sequence groups (or families)
Edit Mode
Mask
Toggle Color
Gaps
Manipulations
Nucleic Acid
Protein
Translate or Reverse-Translate (permanent)
Translate in selected frame (permanent)
Toggle Translation Ctrl+G
Nucleotide Composition
Base composition and mass export (monoisotopic)
Base composition and mass export with average masses
Complement
Reverse Complement Shift+Ctrl+R
DNA->RNA
RNA->DNA
Translate
Find next ORF
Find ORFs from a list of positions
Create Plasmid from Sequence
Gap beginning to minimize stop codons in reading frame 1
codon search
Restriction Map
Sorted Six-Frame Translation
Unsorted Six-Frame Translation

```
GCCGTAGTACAGGCCTCGGCCAATGTGAAGATAAATACAAATAAAGAAGAAAGATGCAC
CAGTTTGATGAAATTTCGGTTCGCTCTTAGGCCTATGCTTGATTTCCCAAATTCTTAC
```

4. Select both the forward and reverse sequence within the nucleotide sequence window, using shift and select (i). Then, in the sequence menu in the uppermost toolbar, select Accessory Application (ii), followed by ClustalW Multiple Alignment. Leave the settings as defaults and click on the Run ClustalW tab (iii).



```
CB_Forward  GGCTGTCATTATTACTAAGAGGAGAAGGATTACCCCGATGTTTCATGTTTCTTTATACAGGTATGAGCCGTAGTACAGGCCTCGG
CB_Reverse  TGTCCTCAGGGCAACACATAGCCTACGAAGGCTGTCATTATTACTAAGAGGAGAAGGATTACCCCGATGTTTCATGTTTCTTTAT
```

ClustalW Options

ClustalW Multiple alignment

Reference:
Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994)
CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice.
Nucleic Acids Research, submitted, June 1994.

☑ Full Multiple alignment
☐ Calculate NJ Tree
☐ FAST algorithm for guide tree
☑ Bootstrap NJ Tree Number of bootstraps: 1000

Gap penalties: Blank=default

Pairwise alignments		Multiple alignment	
Gap open		Gap open	
Gap extend		Gap extend	

Other Parameters:

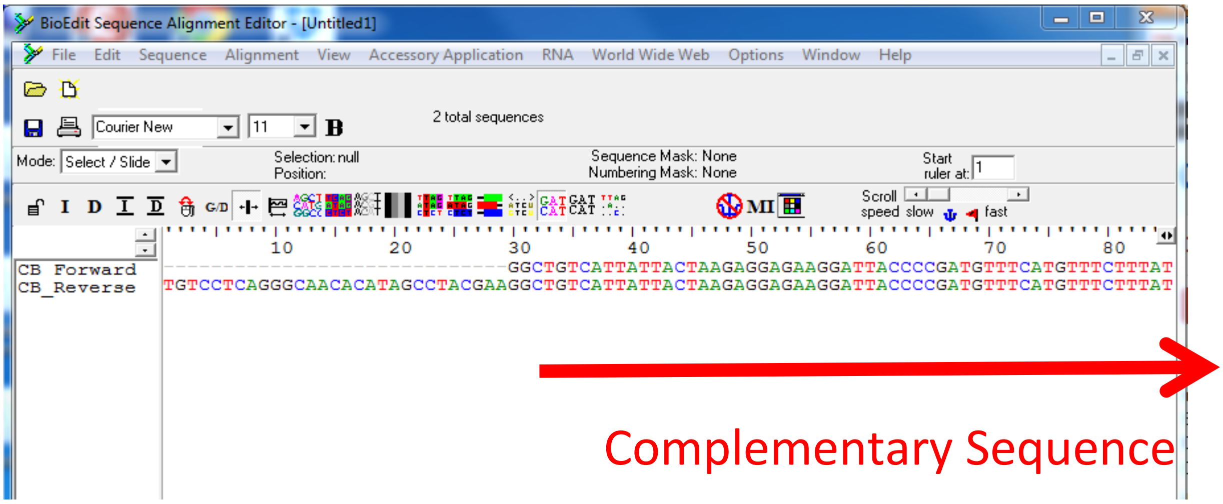
Note: enter additional parameters as a single line.

☐ Output Clustal format with Clustal consensus sequence generation

```
Additional Parameters for ClustalW:

***General settings:****
/QUICKTREE   :use FAST algorithm for the alignment guide tree
/NEWTREE=    :file for new guide tree
/USETREE=    :file for old guide tree
/NEGATIVE    :protein alignment with negative values in matrix
```

5. The software will take a few seconds to align these complementary sequences and the result is a large region of overlapping sequence. As these two sequences come from the same sample they should match perfectly with no mismatching nucleotides. However, any ambiguous nucleotides (i.e. "N") can now be resolved from the complementary sequence. Any mismatches also need to be resolved by consulting the original trace files and deciding which nucleotide call is correct (if this is not possible an "N" can be used at the position where the sequences mismatch, as section 7.6).


BioEdit Sequence Alignment Editor - [Untitled1]
File Edit Sequence Alignment View Accessory Application RNA World Wide Web Options Window Help
Courier New
11
B
Scroll speed slow fast
10 20 30 40 50 60 70 80

2 total sequences

Mode: Select / Slide

Selection: null
Position:

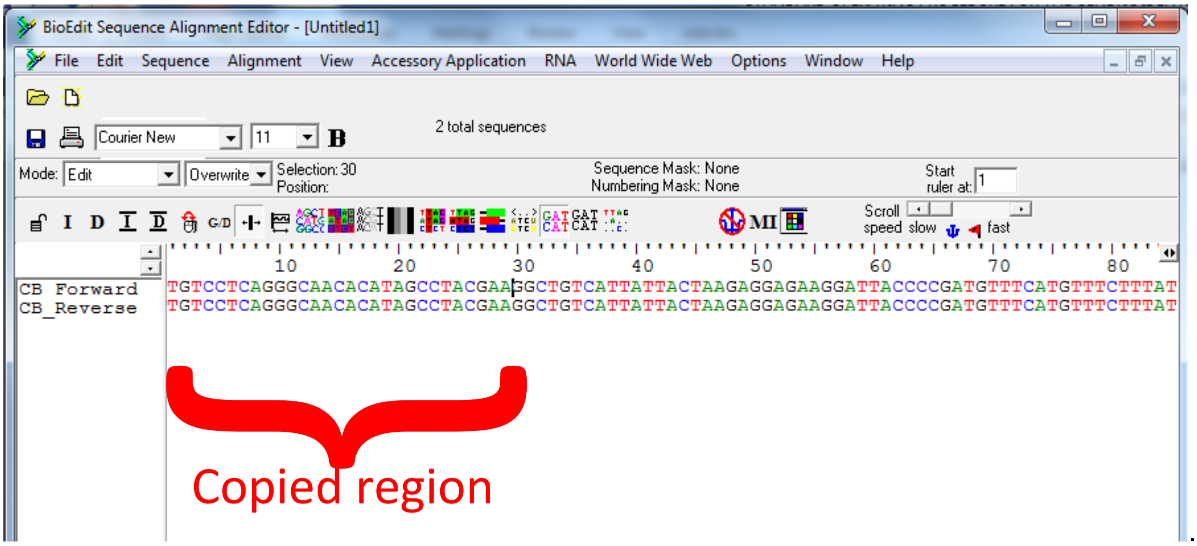
Sequence Mask: None
Numbering Mask: None

Start ruler at: 1

```
CB Forward   -----------------------------GGCTGTCATTATTACTAAGAGGAGAAGGATTACCCCGATGTTTCATGTTTCTTTAT
CB_Reverse   TGTCCTCAGGGCAACACATAGCCTACGAAGGCTGTCATTATTACTAAGAGGAGAAGGATTACCCCGATGTTTCATGTTTCTTTAT
```

Complementary Sequence

6. The region where the complementary sequences do not overlap on the reverse sequence needs to be copied and pasted (using the ctrl+c and ctrl+v keyboard shortcuts) onto the end forward sequence, creating a full length barcode.


BioEdit Sequence Alignment Editor - [Untitled1]
File Edit Sequence Alignment View Accessory Application RNA World Wide Web Options Window Help
Courier New
11
B
Scroll speed slow fast
10 20 30 40 50 60 70 80

2 total sequences

Mode: Edit Overwrite

Selection: 30
Position:

Sequence Mask: None
Numbering Mask: None

Start ruler at: 1

```
CB Forward   TGTCCTCAGGGCAACACATAGCCTACGAAGGCTGTCATTATTACTAAGAGGAGAAGGATTACCCCGATGTTTCATGTTTCTTTAT
CB_Reverse   TGTCCTCAGGGCAACACATAGCCTACGAAGGCTGTCATTATTACTAAGAGGAGAAGGATTACCCCGATGTTTCATGTTTCTTTAT
```

Copied region

7. The reverse sequence can now be deleted and this full length barcode sequence can be saved (as in 7.6), by renaming it "*sampleName*Complementary" and saving it in fasta format.

8. The final step in generating a DNA barcode is to remove the primers. This can be done by referring to the primer sequences 7.3 & 7.5 and removing them from both ends of your sequence. It can perhaps more easily be done by aligning the consensus sequence with a full length barcode downloaded from BOLD. The standard barcode length for most animal species

is 648bp, so your edited sequence should be approximately this long. Below is a full length barcode for Atlantic cod (*Gadus morhua*), in text format, obtained through the application of the steps illustrated above.

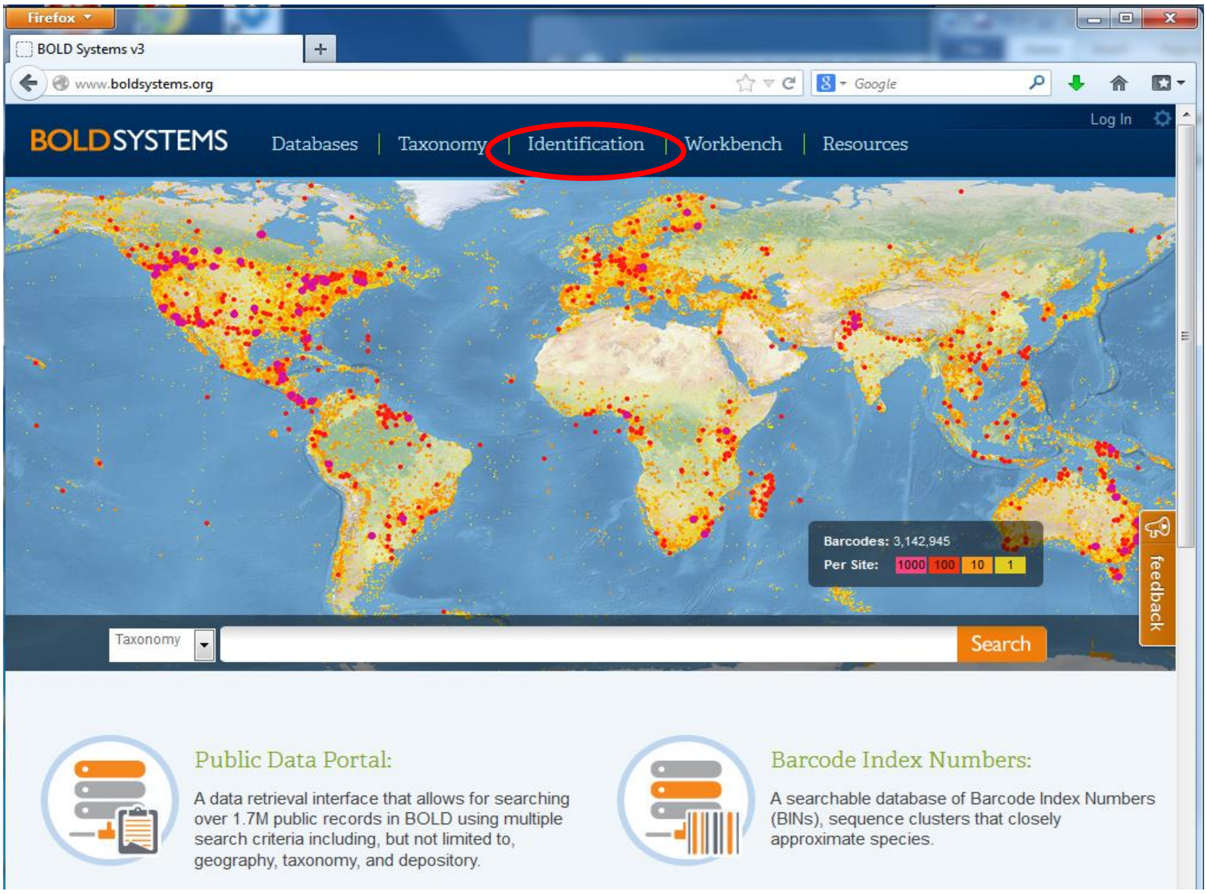
```
>GadusMorhuaSCFAC839-06

CCTTTATCTCGTATTTGGTGCCTGAGCCGGCATAGTCGGAACAGCCCTAAGCCTGCTCA
TTCGAGCAGAGCTAAGTCAACCTGGTGCACTTCTTGGTGATGATCAAATTTATAATGTG
ATCGTTACAGCGCACGCTTTCGTAATAATTTTCTTTATAGTAATACCACTAATAATTGGA
GGCTTTGGGAACTGACTCATTCCTCTAATGATCGGTGCACCAGATATAGCTTTCCCTCG
AATAAATAACATAAGCTTCTGACTTCTTCCTCCATCTTTCCTGCTCCTTTTAGCATCCTCT
GGTGTAGAAGCTGGGGCTGGAACAGGCTGAACTGTCTATCCACCTTTAGCCGGAAACC
TCGCTCATGCTGGGGCATCTGTTGATCTCACTATTTTTTCTCTTCATCTAGCAGGGATTT
CATCAATTCTTGGGGCAATTAATTTTATTACCACAATTATTAATATGAAACCTCCGGCAAT
TTCACAGTACCAAACACCCCTATTTGTTTGAGCAGTACTAATTACAGCTGTGCTTCTACT
ATTATCTCTCCCCGTCTTAGCAGCTGGTATCACAATACTTCTAACTGACCGTAATCTTAA
TACTTCTTTCTTTGACCCTGCTGGAGGAGGTGATCCCATTTTATACCAACA
```

7.8 Identifying the species on the Barcode of Life Database

In order to identify what species your consensus, full-length COI sequence originates from it is necessary to utilise freely available data that has been submitted to the Barcode of Life (BOLD) project. This includes a comprehensive database of COI sequence data that has been collected by individuals and organisation across the globe and is constantly being updated with new data.

1. Start by navigating to the **BOLD Systems** webpage (http://www.boldsystems.org/) and select the “Identification” tab at the top of the webpage.



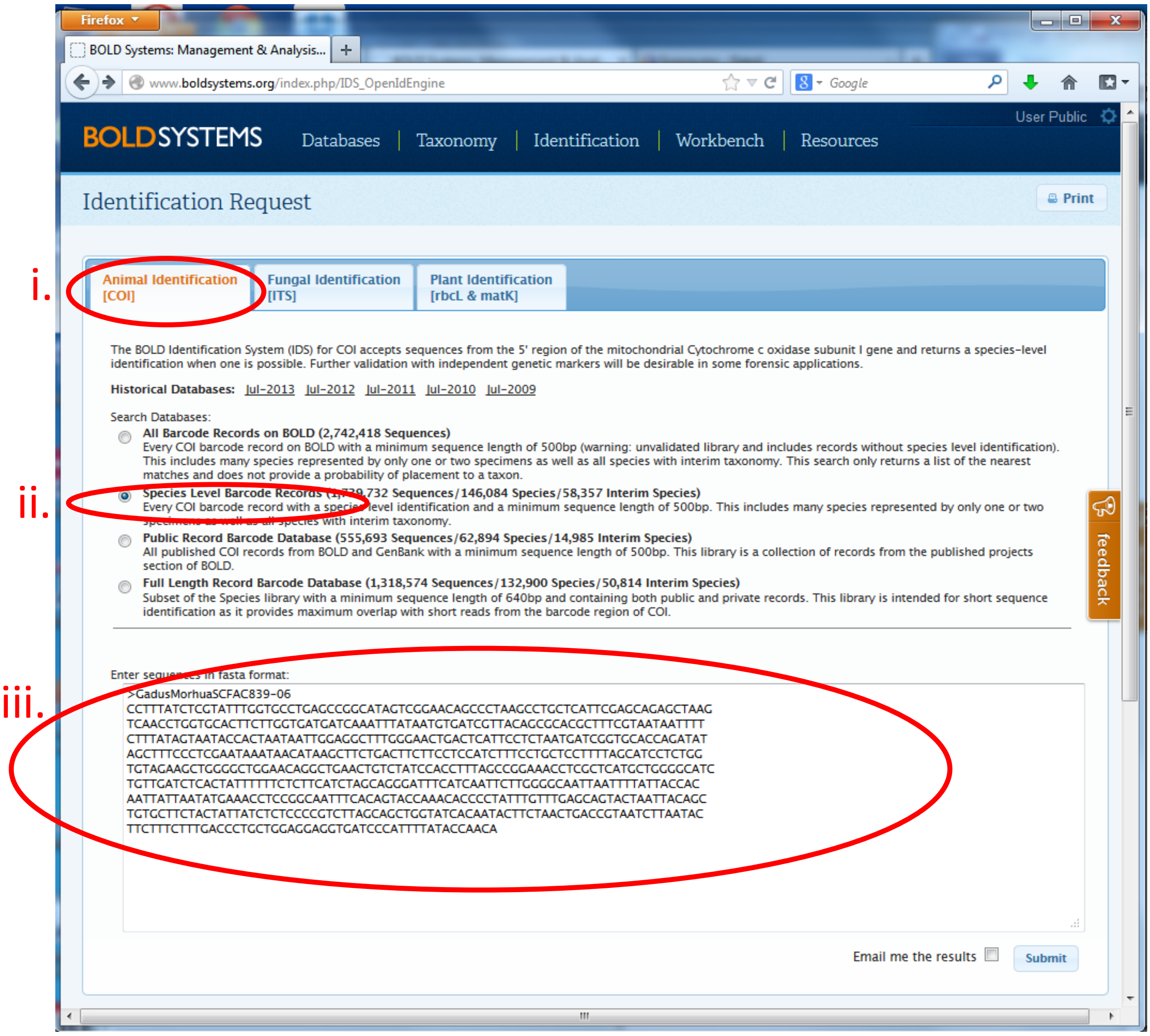
Public Data Portal:

A data retrieval interface that allows for searching over 1.7M public records in BOLD using multiple search criteria including, but not limited to, geography, taxonomy, and depository.

Barcode Index Numbers:

A searchable database of Barcode Index Numbers (BINs), sequence clusters that closely approximate species.

2. This page acts as a portal allowing the consensus sequence generated in the laboratory to be referenced against the entire BOLD database of reference data, i.e. from known species. **Various search options** are possible that relate to different collections of reference data, but the **default settings** provide an excellent initial step at identifying the species. However, it is important to ensure that the “Animal Identification (COI)” tab (i) and the “Species Level Barcode Records” database (ii) are both selected. The consensus sequence obtained from the sample can then be cut & pasted into the empty box at the bottom of the page (iii); in this example the published sequence from *Gadus morhua* included in the previous section has been utilised. The easiest way to copy the consensus sequence in your fasta file is to force windows to open the .fas file in Notepad. Alternatively, make a copy of the .fas file and edit the file extension to .txt allowing it to be opened in Notepad. Once the sequence has been entered, hit the submit button at the bottom of the page.



Identification Request

Animal Identification [COI] | Fungal Identification [ITS] | Plant Identification [rbcL & matK]

The BOLD Identification System (IDS) for COI accepts sequences from the 5' region of the mitochondrial Cytochrome c oxidase subunit I gene and returns a species-level identification when one is possible. Further validation with independent genetic markers will be desirable in some forensic applications.

Historical Databases: Jul-2013 Jul-2012 Jul-2011 Jul-2010 Jul-2009

Search Databases:

○ **All Barcode Records on BOLD (2,742,418 Sequences)**
Every COI barcode record on BOLD with a minimum sequence length of 500bp (warning: unvalidated library and includes records without species level identification). This includes many species represented by only one or two specimens as well as all species with interim taxonomy. This search only returns a list of the nearest matches and does not provide a probability of placement to a taxon.

● **Species Level Barcode Records (1,739,732 Sequences/146,084 Species/58,357 Interim Species)**
Every COI barcode record with a species level identification and a minimum sequence length of 500bp. This includes many species represented by only one or two specimens as well as all species with interim taxonomy.

○ **Public Record Barcode Database (555,693 Sequences/62,894 Species/14,985 Interim Species)**
All published COI records from BOLD and GenBank with a minimum sequence length of 500bp. This library is a collection of records from the published projects section of BOLD.

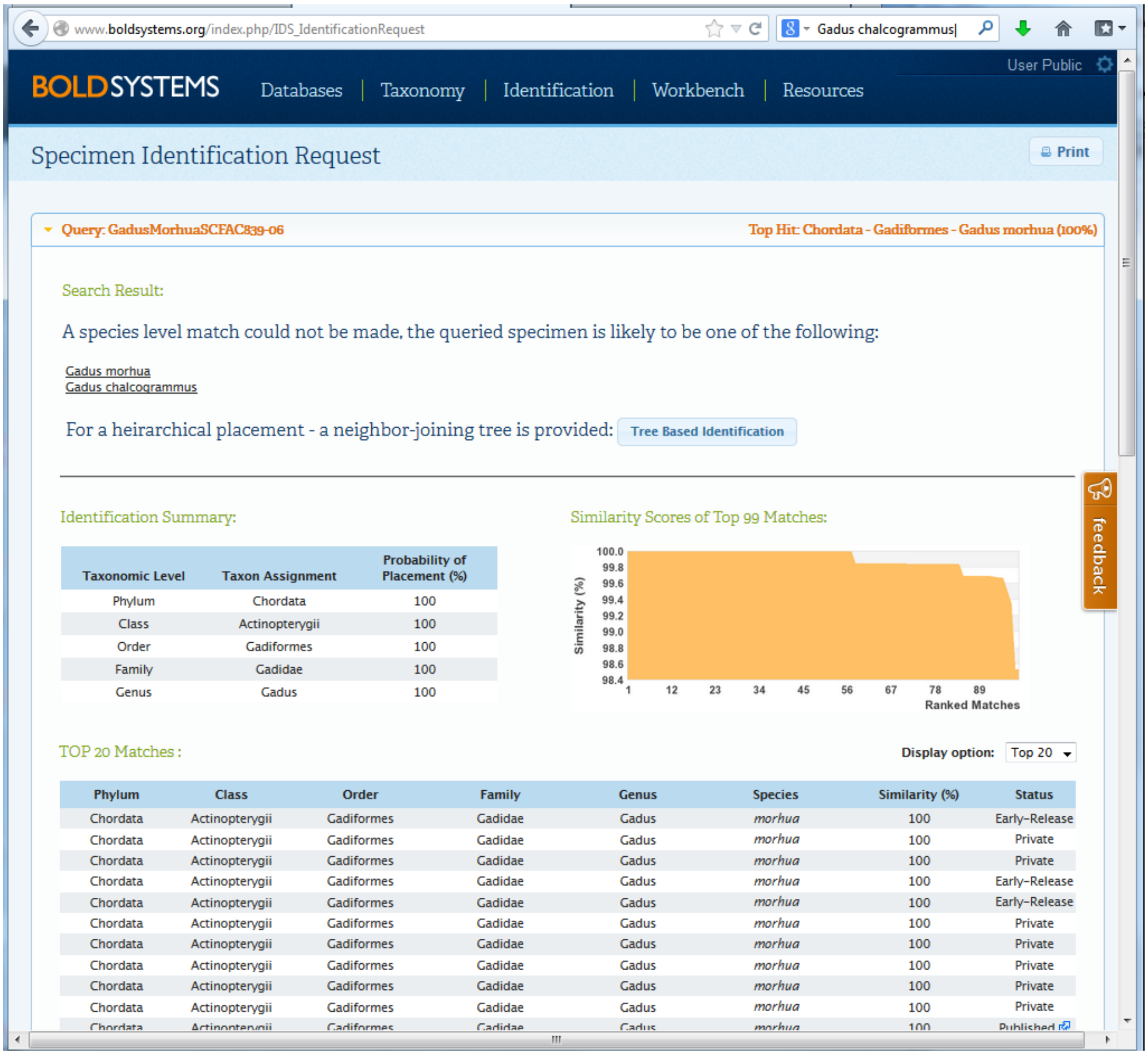
○ **Full Length Record Barcode Database (1,318,574 Sequences/132,900 Species/50,814 Interim Species)**
Subset of the Species library with a minimum sequence length of 640bp and containing both public and private records. This library is intended for short sequence identification as it provides maximum overlap with short reads from the barcode region of COI.

Enter sequences in fasta format:

```
>GadusMorhuaSCFAC839-06
CCTTTATCTCGTATTTGGTGCCTGAGCCGGCATAGTCGGAACAGCCCTAAGCCTGCTCATTCGAGCAGAGCTAAG
TCAACCTGGTGCACTTCTTGGTGATGATCAAATTTATAATGTGATCGTTACAGCGCACGCTTTCGTAATAATTTT
CTTTATAGTAATACCACTAATAATTGGAGGCTTTGGGAACTGACTCATTCCTCTAATGATCGGTGCACCAGATAT
AGCTTTCCCTCGAATAAATAACATAAGCTTCTGACTTCTTCCTCCATCTTTCCTGCTCCTTTTAGCATCCTCTGG
TGTAGAAGCTGGGGCTGGAACAGGCTGAACTGTCTATCCACCTTTAGCCGGAAACCTCGCTCATGCTGGGGCATC
TGTTGATCTCACTATTTTTTCTCTTCATCTAGCAGGGATTTCATCAATTCTTGGGGCAATTAATTTTATTACCAC
AATTATTAATATGAAACCTCCGGCAATTTCACAGTACCAAACACCCCTATTTGTTTGAGCAGTACTAATTACAGC
TGTGCTTCTACTATTATCTCTCCCCGTCTTAGCAGCTGGTATCACAATACTTCTAACTGACCGTAATCTTAATAC
TTCTTTCTTTGACCCTGCTGGAGGAGGTGATCCCATTTTATACCAACA
```

Email me the results □ Submit

3. In a few seconds the browser will update and give you the results of the search, revealing the records contained in the database that **yields the closest match** in terms of sequence similarity. First, it is important to save a screen grab of the results as proof of the result, something similar to the picture below (this can be done using the print screen option, pasting directly into Paint or a Microsoft Office software and saving as a picture).



www.boldsystems.org/index.php/IDS_IdentificationRequest
Gadus chalcogrammus|
BOLDSYSTEMS
Databases | Taxonomy | Identification | Workbench | Resources
User Public
Print
feedback

Specimen Identification Request

Query: GadusMorhuaSCFAC839-06

Top Hit: Chordata - Gadiformes - Gadus morhua (100%)

Search Result:

A species level match could not be made, the queried specimen is likely to be one of the following:

Gadus morhua
Gadus chalcogrammus

For a heirarchical placement - a neighbor-joining tree is provided: Tree Based Identification

Identification Summary:

Taxonomic Level	Taxon Assignment	Probability of Placement (%)
Phylum	Chordata	100
Class	Actinopterygii	100
Order	Gadiformes	100
Family	Gadidae	100
Genus	Gadus	100

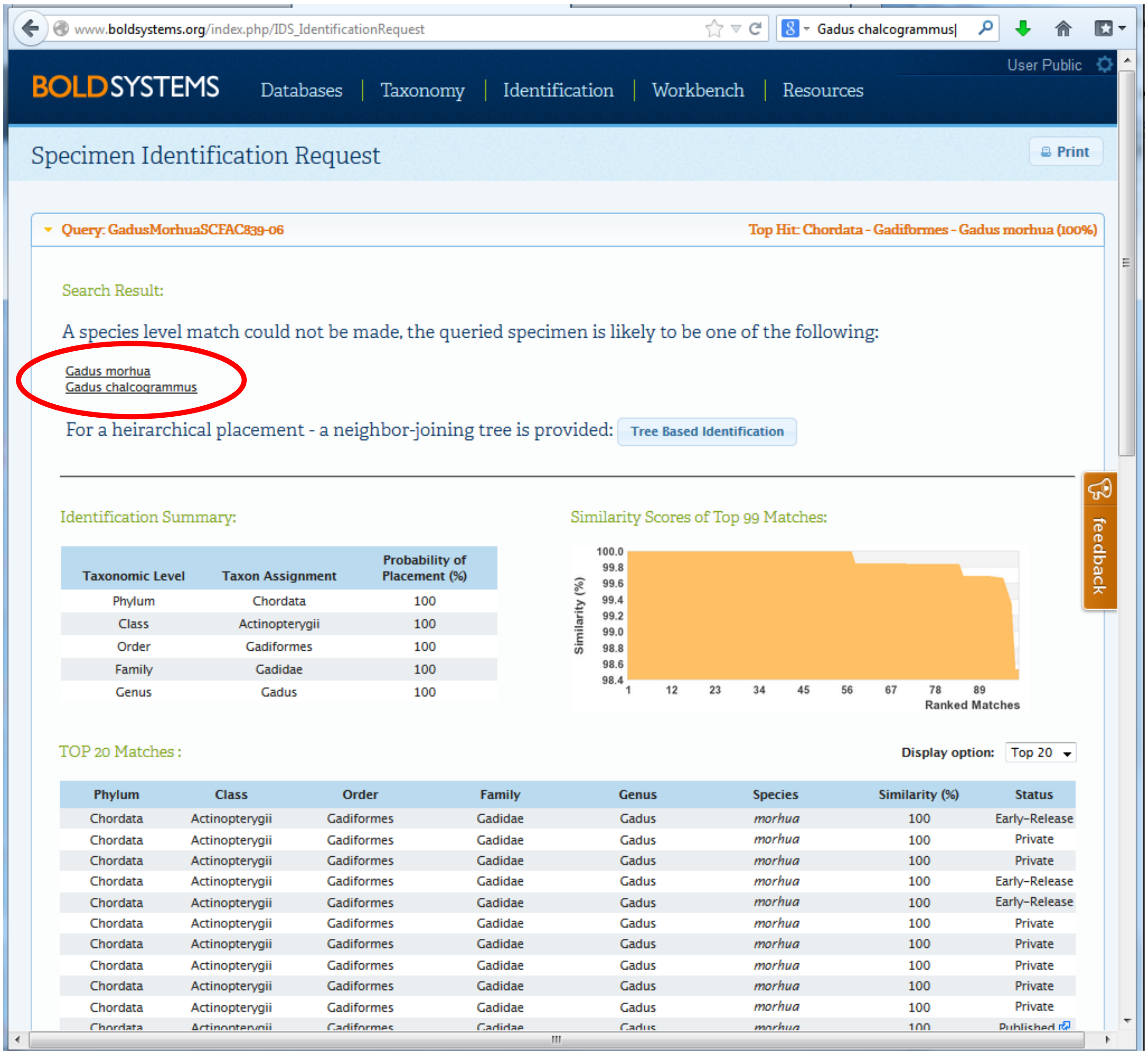
Similarity Scores of Top 99 Matches:

TOP 20 Matches :

Display option: Top 20

Phylum	Class	Order	Family	Genus	Species	Similarity (%)	Status
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	*morhua*	100	Early-Release
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	*morhua*	100	Private
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	*morhua*	100	Private
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	*morhua*	100	Early-Release
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	*morhua*	100	Early-Release
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	*morhua*	100	Private
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	*morhua*	100	Private
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	*morhua*	100	Private
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	*morhua*	100	Private
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	*morhua*	100	Private
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	*morhua*	100	Published

4. This screen also contains a lot of information that will allow a **confident identification** to be made from your sequence. At the top the search result is returned; in BOLD this generally means any species that has a sequence record that is 98% similar (or more) will be returned. Often this will just be a single species allowing an unambiguous identification to be made for the sample. **However**, in the example below, two species have been returned (highlighted in red), prompting BOLD to display the message "**A species match could not be made**, the queried specimen is likely to be one of the following". It is possible to interrogate the results further and still make a robust identification.



Specimen Identification Request

Query: GadusMorhuaSCFAC839-06

Top Hit: Chordata - Gadiformes - Gadus morhua (100%)

Search Result:

A species level match could not be made, the queried specimen is likely to be one of the following:

Gadus morhua
Gadus chalcogrammus

For a heirarchical placement - a neighbor-joining tree is provided: Tree Based Identification

Identification Summary:

Taxonomic Level	Taxon Assignment	Probability of Placement (%)
Phylum	Chordata	100
Class	Actinopterygii	100
Order	Gadiformes	100
Family	Gadidae	100
Genus	Gadus	100

Similarity Scores of Top 99 Matches:

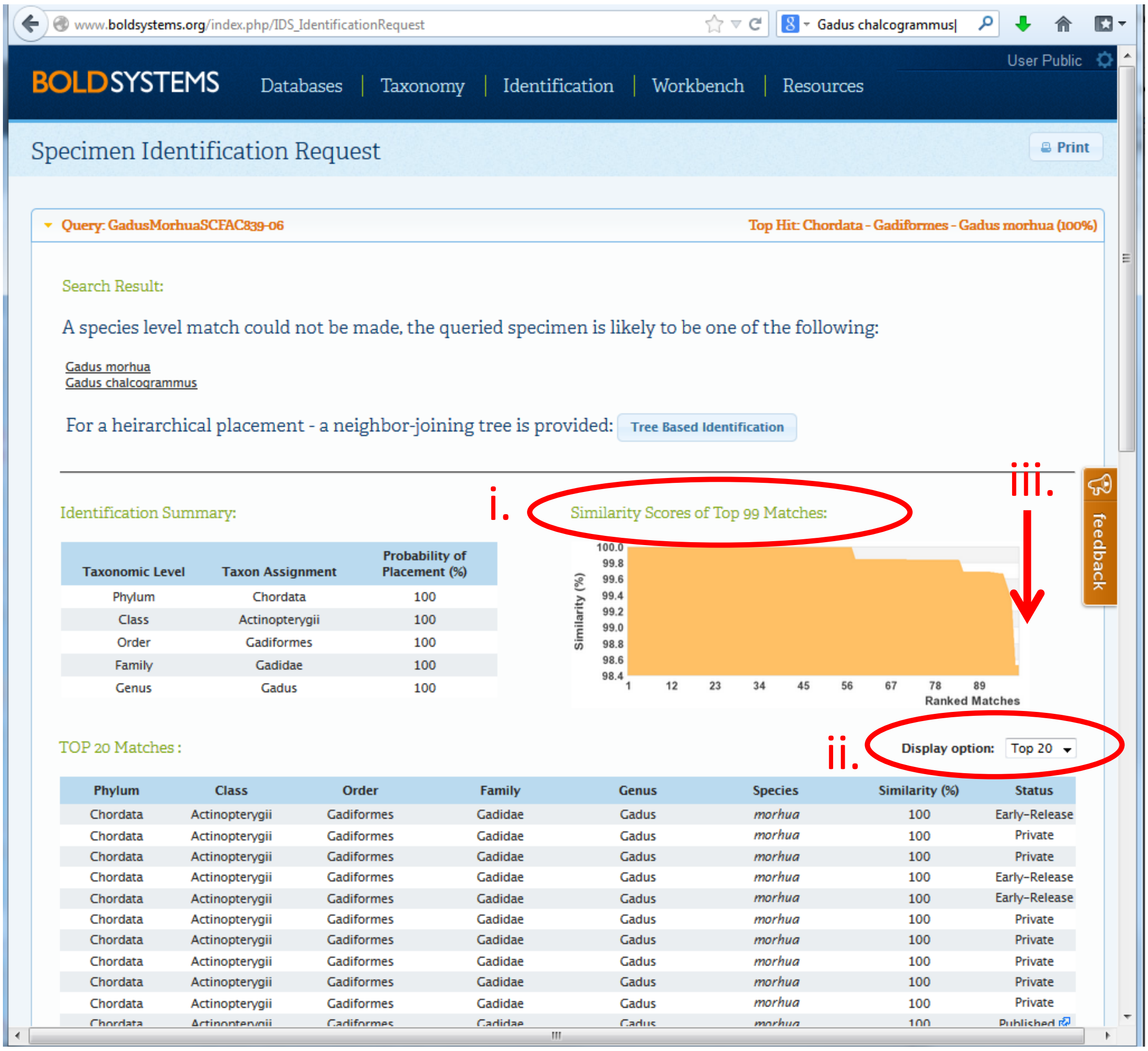
TOP 20 Matches :

Display option: Top 20

Phylum	Class	Order	Family	Genus	Species	Similarity (%)	Status
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	*morhua*	100	Early-Release
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	*morhua*	100	Private
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	*morhua*	100	Private
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	*morhua*	100	Early-Release
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	*morhua*	100	Early-Release
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	*morhua*	100	Private
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	*morhua*	100	Private
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	*morhua*	100	Private
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	*morhua*	100	Private
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	*morhua*	100	Private
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	*morhua*	100	Published

5. Next examine the graph entitled “Similarity Scores of Top 99 Matches” that shows the percent similarity for each of 99 top matching records in the database against your consensus sequences (i.). Also alter the display options in the drop down menu (ii.), to make BOLD show the full records for these corresponding top 99 matches.

In the example below, it is clearly illustrated that there is **100% sequence similarity** between our example consensus sequence and the *Gadus morhua* records. It is also clear that there is a sudden reduction in the level of similarity observed between the consensus sequence and the records originating from *Gadus morhua* (which are 100-99.35% similar) and those from *Gadus chalcogrammus* (whose highest similarity is 98.53%), as indicated by the red arrow (iii.). The 100% sequence match criterion alongside the reduced similarity between our consensus sequences and any other matching species record, are both strong indicators that the sequence originated from *Gadus morhua*.



Specimen Identification Request

Query: GadusMorhuaSCFAC839-06

Top Hit: Chordata - Gadiformes - Gadus morhua (100%)

Search Result:

A species level match could not be made, the queried specimen is likely to be one of the following:

Gadus morhua
Gadus chalcogrammus

For a heirarchical placement - a neighbor-joining tree is provided: Tree Based Identification

Identification Summary:

Taxonomic Level	Taxon Assignment	Probability of Placement (%)
Phylum	Chordata	100
Class	Actinopterygii	100
Order	Gadiformes	100
Family	Gadidae	100
Genus	Gadus	100

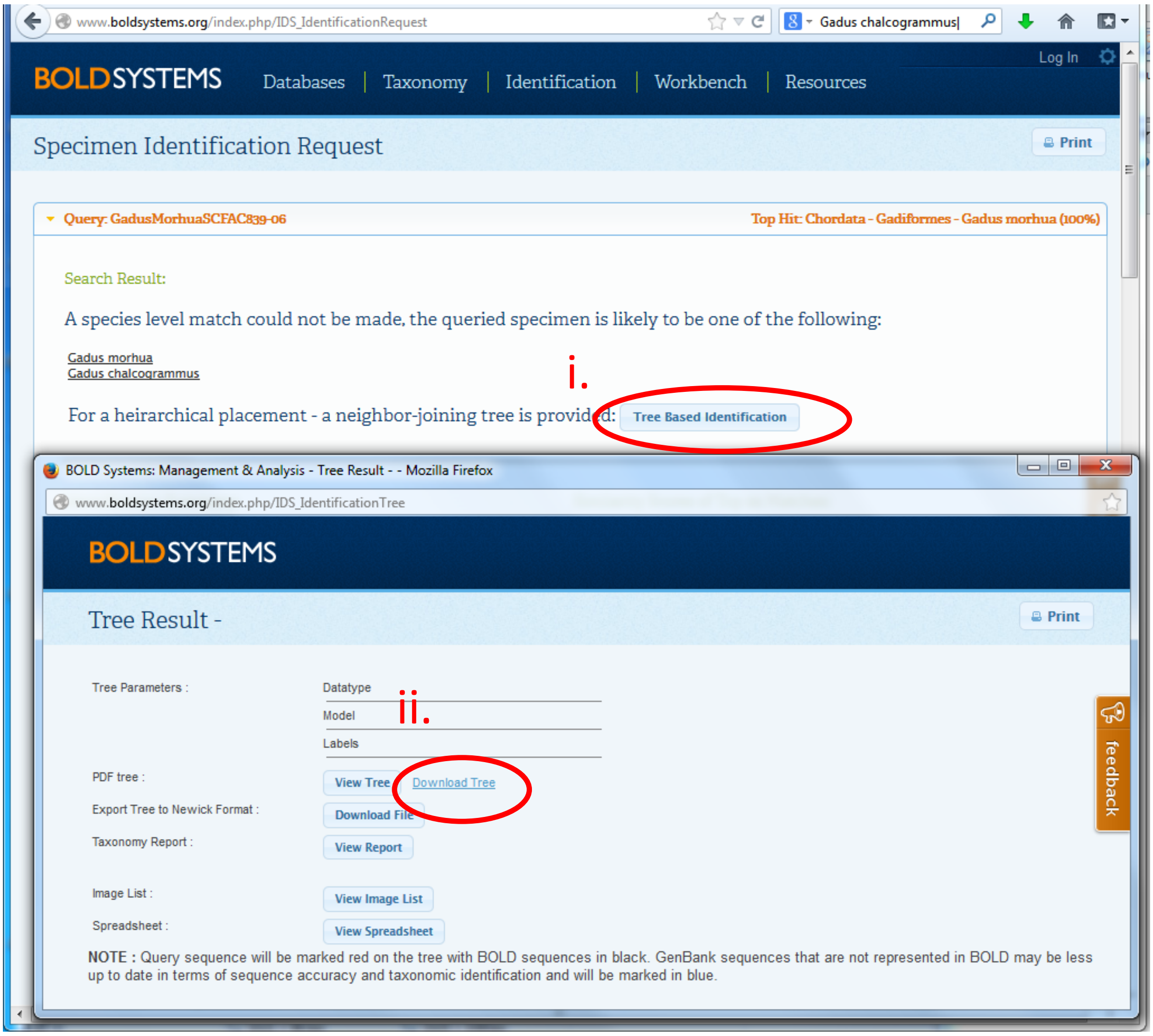
Similarity Scores of Top 99 Matches:

TOP 20 Matches :

Display option: Top 20

Phylum	Class	Order	Family	Genus	Species	Similarity (%)	Status
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	*morhua*	100	Early-Release
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	*morhua*	100	Private
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	*morhua*	100	Private
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	*morhua*	100	Early-Release
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	*morhua*	100	Early-Release
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	*morhua*	100	Private
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	*morhua*	100	Private
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	*morhua*	100	Private
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	*morhua*	100	Private
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	*morhua*	100	Private
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	*morhua*	100	Published

6. Besides referencing your sequence against the BOLD reference database, it is **also important to produce a simple tree** to graphically display the results of the homology search (although this is not a highly robust phylogenetic reconstruction). First click on the "Tree Based Identification" tab (i.), then a new window will pop up and the tree can be saved as a pdf by selecting the "Download Tree" option. This can then be saved as a permanent record of the results, to be kept alongside the previous screen-grab.



www.boldsystems.org/index.php/IDS_IdentificationRequest

Specimen Identification Request

Query: GadusMorhuaSCFAC839-06

Top Hit: Chordata - Gadiformes - Gadus morhua (100%)

Search Result:

A species level match could not be made, the queried specimen is likely to be one of the following:

Gadus morhua
Gadus chalcogrammus

For a heirarchical placement - a neighbor-joining tree is provided: Tree Based Identification

BOLD Systems: Management & Analysis - Tree Result - - Mozilla Firefox

www.boldsystems.org/index.php/IDS_IdentificationTree

Tree Result -

Tree Parameters : Datatype, Model, Labels

PDF tree : View Tree | Download Tree

Export Tree to Newick Format : Download File

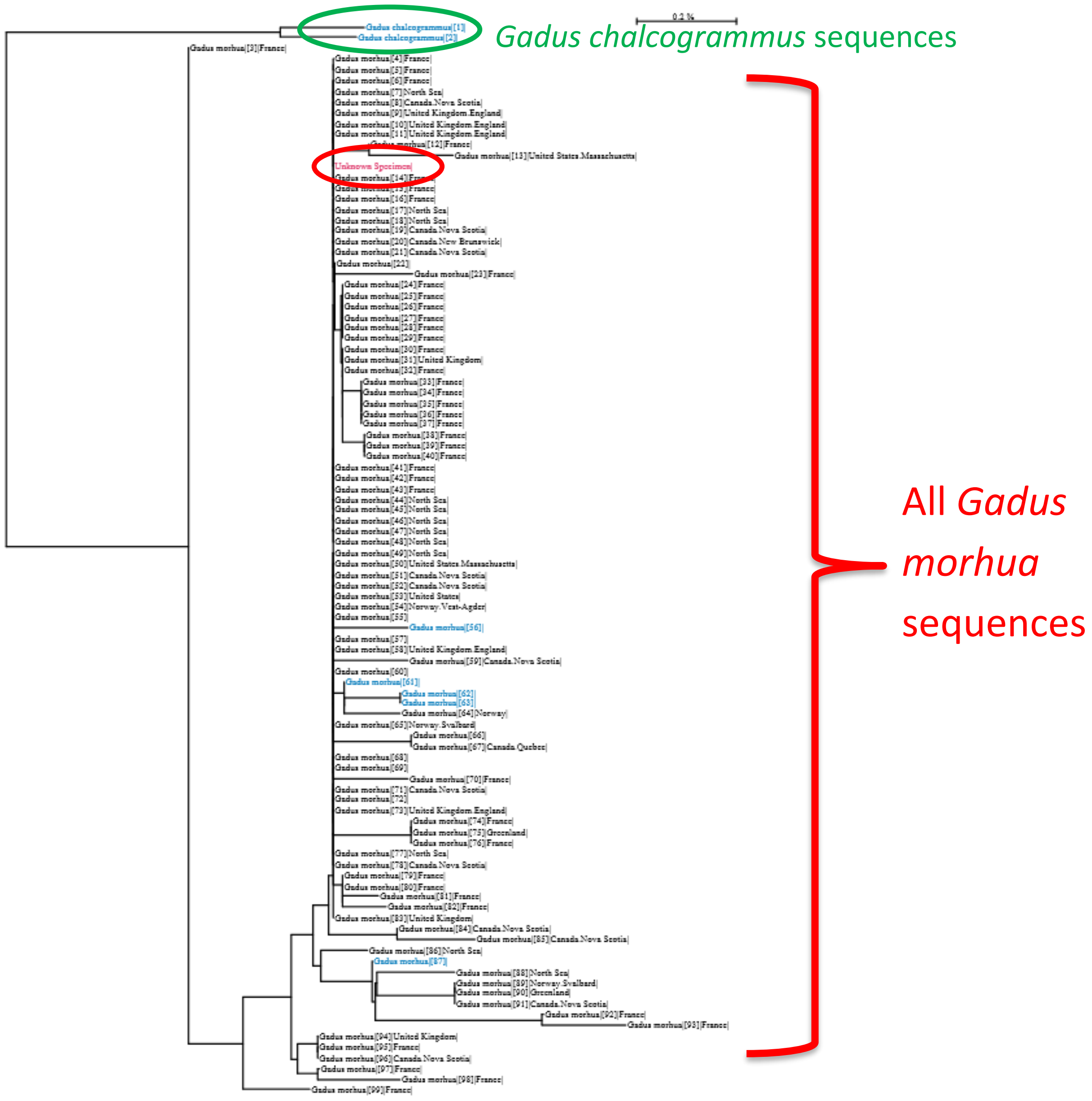
Taxonomy Report : View Report

Image List : View Image List

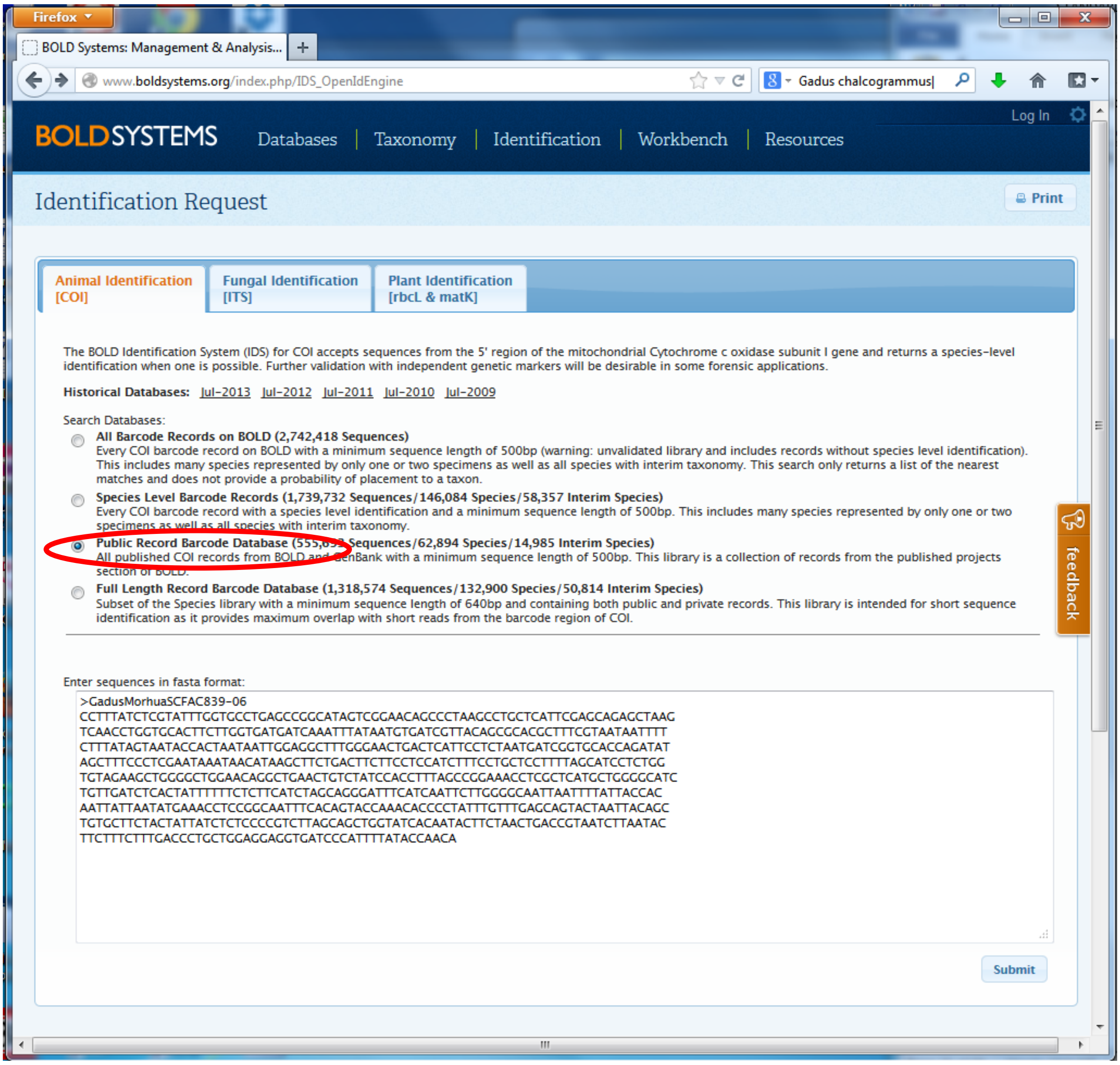
Spreadsheet : View Spreadsheet

NOTE : Query sequence will be marked red on the tree with BOLD sequences in black. GenBank sequences that are not represented in BOLD may be less up to date in terms of sequence accuracy and taxonomic identification and will be marked in blue.

7. In the tree diagram the uploaded sequence is highlighted in red. In order to make a clear identification, this “unknown specimen” should **only cluster with sequences originating from a single species** (i.e. from a *monophyletic* group). The tree generated from our example sequence is below; our uploaded sequence is clearly shown (highlighted in red) nested within sequences exclusively originating from *Gadus morhua,* with *Gadus chalcogrammus* (highlighted in green) forming a separate branch some distance from our unknown specimen. This is further evidence that this sequence originated from *Gadus morhua.*


0.2 %
Gadus chalcogrammus|[1]|
Gadus chalcogrammus|[2]|
Gadus chalcogrammus sequences
Gadus morhua|[3]|France|
Gadus morhua|[4]|France|
Gadus morhua|[5]|France|
Gadus morhua|[6]|France|
Gadus morhua|[7]|North Sea|
Gadus morhua|[8]|Canada.Nova Scotia|
Gadus morhua|[9]|United Kingdom.England|
Gadus morhua|[10]|United Kingdom.England|
Gadus morhua|[11]|United Kingdom.England|
Gadus morhua|[12]|France|
Gadus morhua|[13]|United States.Massachusetts|
Unknown Specimen|
Gadus morhua|[14]|France|
Gadus morhua|[15]|France|
Gadus morhua|[16]|France|
Gadus morhua|[17]|North Sea|
Gadus morhua|[18]|North Sea|
Gadus morhua|[19]|Canada.Nova Scotia|
Gadus morhua|[20]|Canada.New Brunswick|
Gadus morhua|[21]|Canada.Nova Scotia|
Gadus morhua|[22]|
Gadus morhua|[23]|France|
Gadus morhua|[24]|France|
Gadus morhua|[25]|France|
Gadus morhua|[26]|France|
Gadus morhua|[27]|France|
Gadus morhua|[28]|France|
Gadus morhua|[29]|France|
Gadus morhua|[30]|France|
Gadus morhua|[31]|United Kingdom|
Gadus morhua|[32]|France|
Gadus morhua|[33]|France|
Gadus morhua|[34]|France|
Gadus morhua|[35]|France|
Gadus morhua|[36]|France|
Gadus morhua|[37]|France|
Gadus morhua|[38]|France|
Gadus morhua|[39]|France|
Gadus morhua|[40]|France|
Gadus morhua|[41]|France|
Gadus morhua|[42]|France|
Gadus morhua|[43]|France|
Gadus morhua|[44]|North Sea|
Gadus morhua|[45]|North Sea|
Gadus morhua|[46]|North Sea|
Gadus morhua|[47]|North Sea|
Gadus morhua|[48]|North Sea|
Gadus morhua|[49]|North Sea|
Gadus morhua|[50]|United States.Massachusetts|
Gadus morhua|[51]|Canada.Nova Scotia|
Gadus morhua|[52]|Canada.Nova Scotia|
Gadus morhua|[53]|United States|
Gadus morhua|[54]|Norway.Vest-Agder|
Gadus morhua|[55]|
Gadus morhua|[56]|
Gadus morhua|[57]|
Gadus morhua|[58]|United Kingdom.England|
Gadus morhua|[59]|Canada.Nova Scotia|
Gadus morhua|[60]|
Gadus morhua|[61]|
Gadus morhua|[62]|
Gadus morhua|[63]|
Gadus morhua|[64]|Norway|
Gadus morhua|[65]|Norway.Svalbard|
Gadus morhua|[66]|
Gadus morhua|[67]|Canada.Quebec|
Gadus morhua|[68]|
Gadus morhua|[69]|
Gadus morhua|[70]|France|
Gadus morhua|[71]|Canada.Nova Scotia|
Gadus morhua|[72]|
Gadus morhua|[73]|United Kingdom.England|
Gadus morhua|[74]|France|
Gadus morhua|[75]|Greenland|
Gadus morhua|[76]|France|
Gadus morhua|[77]|North Sea|
Gadus morhua|[78]|Canada.Nova Scotia|
Gadus morhua|[79]|France|
Gadus morhua|[80]|France|
Gadus morhua|[81]|France|
Gadus morhua|[82]|France|
Gadus morhua|[83]|United Kingdom|
Gadus morhua|[84]|Canada.Nova Scotia|
Gadus morhua|[85]|Canada.Nova Scotia|
Gadus morhua|[86]|North Sea|
Gadus morhua|[87]|
Gadus morhua|[88]|North Sea|
Gadus morhua|[89]|Norway.Svalbard|
Gadus morhua|[90]|Greenland|
Gadus morhua|[91]|Canada.Nova Scotia|
Gadus morhua|[92]|France|
Gadus morhua|[93]|France|
Gadus morhua|[94]|United Kingdom|
Gadus morhua|[95]|France|
Gadus morhua|[96]|Canada.Nova Scotia|
Gadus morhua|[97]|France|
Gadus morhua|[98]|France|
Gadus morhua|[99]|France|
All *Gadus morhua* sequences

8. **In cases where BOLD returns more than one species** and displays the message: "A species match could not be made, the queried specimen is likely to be one of the following", an additional search can also be made, **utilising a different set of reference data**. Return to the identification request portal and upload the sequence, but select the "Public Record Barcode Database" (this restricts the search to sequences that have been published). In some instances this may help provide an unambiguous identification and the results can be generated and saved as above (with a screen-grab and tree, relating to this search).



Identification Request

Animal Identification [COI] | Fungal Identification [ITS] | Plant Identification [rbcL & matK]

The BOLD Identification System (IDS) for COI accepts sequences from the 5' region of the mitochondrial Cytochrome c oxidase subunit I gene and returns a species-level identification when one is possible. Further validation with independent genetic markers will be desirable in some forensic applications.

Historical Databases: Jul-2013 Jul-2012 Jul-2011 Jul-2010 Jul-2009

Search Databases:

○ **All Barcode Records on BOLD (2,742,418 Sequences)**
Every COI barcode record on BOLD with a minimum sequence length of 500bp (warning: unvalidated library and includes records without species level identification). This includes many species represented by only one or two specimens as well as all species with interim taxonomy. This search only returns a list of the nearest matches and does not provide a probability of placement to a taxon.

○ **Species Level Barcode Records (1,739,732 Sequences/146,084 Species/58,357 Interim Species)**
Every COI barcode record with a species level identification and a minimum sequence length of 500bp. This includes many species represented by only one or two specimens as well as all species with interim taxonomy.

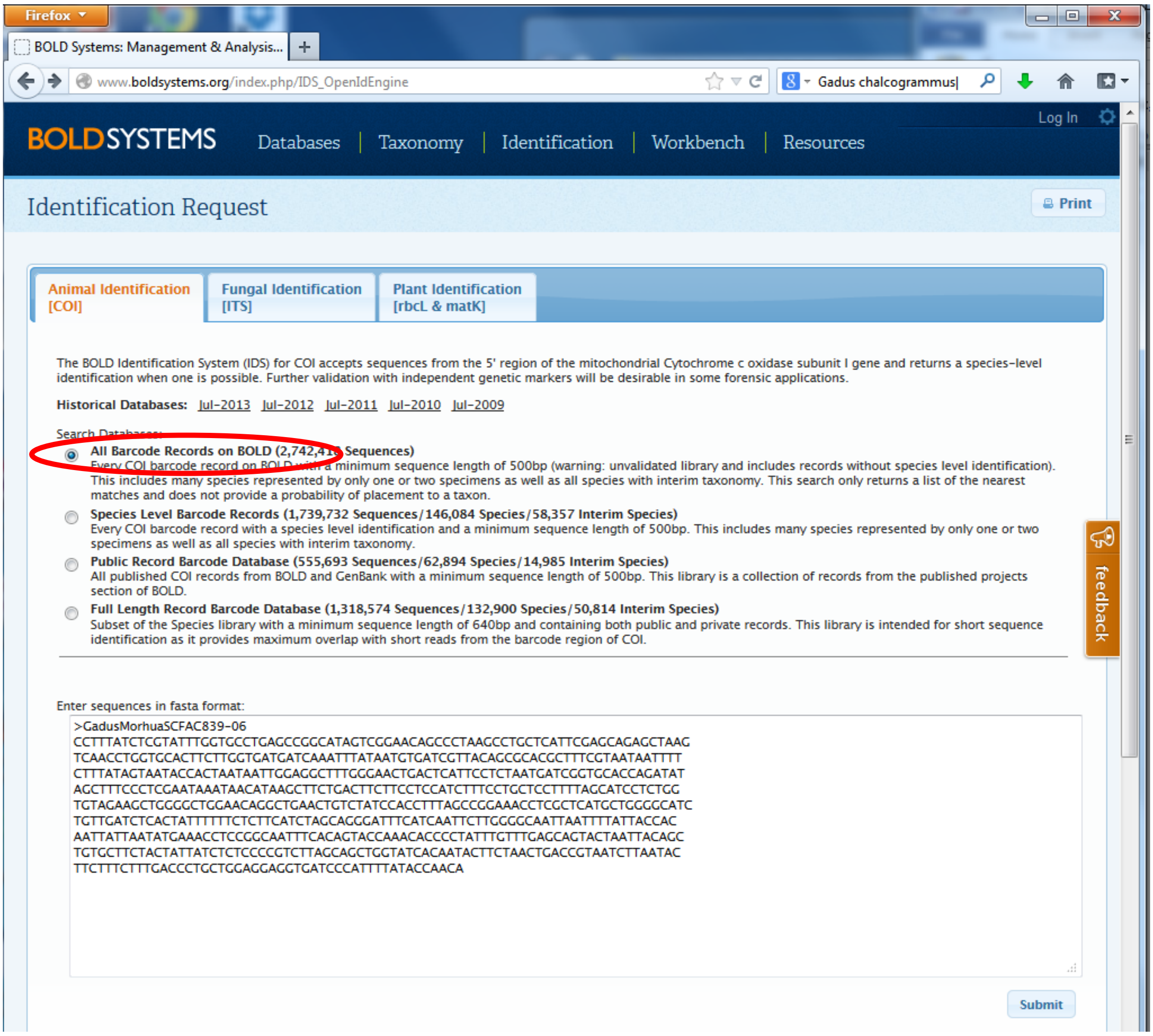
● **Public Record Barcode Database (555,6[illegible] Sequences/62,894 Species/14,985 Interim Species)**
All published COI records from BOLD and GenBank with a minimum sequence length of 500bp. This library is a collection of records from the published projects section of BOLD.

○ **Full Length Record Barcode Database (1,318,574 Sequences/132,900 Species/50,814 Interim Species)**
Subset of the Species library with a minimum sequence length of 640bp and containing both public and private records. This library is intended for short sequence identification as it provides maximum overlap with short reads from the barcode region of COI.

Enter sequences in fasta format:

```
>GadusMorhuaSCFAC839-06
CCTTTATCTCGTATTTGGTGCCTGAGCCGGCATAGTCGGAACAGCCCTAAGCCTGCTCATTCGAGCAGAGCTAAG
TCAACCTGGTGCACTTCTTGGTGATGATCAAATTTATAATGTGATCGTTACAGCGCACGCTTTCGTAATAATTTT
CTTTATAGTAATACCACTAATAATTGGAGGCTTTGGGAACTGACTCATTCCTCTAATGATCGGTGCACCAGATAT
AGCTTTCCCTCGAATAAATAACATAAGCTTCTGACTTCTTCCTCCATCTTTCCTGCTCCTTTTAGCATCCTCTGG
TGTAGAAGCTGGGGCTGGAACAGGCTGAACTGTCTATCCACCTTTAGCCGGAAACCTCGCTCATGCTGGGGCATC
TGTTGATCTCACTATTTTTTCTCTTCATCTAGCAGGGATTTCATCAATTCTTGGGGCAATTAATTTTATTACCAC
AATTATTAATATGAAACCTCCGGCAATTTCACAGTACCAAACACCCCTATTTGTTTGAGCAGTACTAATTACAGC
TGTGCTTCTACTATTATCTCTCCCCGTCTTAGCAGCTGGTATCACAATACTTCTAACTGACCGTAATCTTAATAC
TTCTTTCTTTGACCCTGCTGGAGGAGGTGATCCCATTTTATACCAACA
```

9. Alternatively, if the sample for example comes from a rare or exotic fish, there may be **no matching records** in the “Species Level Barcode Records” database that demonstrate high levels of sequence similarity. The screen grab and tree are still essential records, especially as BOLD may still be able to assign the sequence to a **genus** or **family**, which still provides potentially useful information (and is often enough to help check for mislabelling). An **additional search is also possible**, in this case, by selecting the “All Public Records on BOLD” (this is the broadest database). This may yield a stronger match and the results can be saved as above (with a screen-grab and tree, relating to this search). If *a-priori* information about the species that sample supposedly originates from is available (i.e. the label), it is also possible to check if a species is represented in the database within the taxonomy tab at the top of the window. For further information on troubleshooting see 7.10.



Identification Request

Animal Identification [COI] | Fungal Identification [ITS] | Plant Identification [rbcL & matK]

The BOLD Identification System (IDS) for COI accepts sequences from the 5' region of the mitochondrial Cytochrome c oxidase subunit I gene and returns a species-level identification when one is possible. Further validation with independent genetic markers will be desirable in some forensic applications.

Historical Databases: Jul-2013 Jul-2012 Jul-2011 Jul-2010 Jul-2009

Search Databases:

● **All Barcode Records on BOLD (2,742,418 Sequences)**
Every COI barcode record on BOLD with a minimum sequence length of 500bp (warning: unvalidated library and includes records without species level identification). This includes many species represented by only one or two specimens as well as all species with interim taxonomy. This search only returns a list of the nearest matches and does not provide a probability of placement to a taxon.

○ **Species Level Barcode Records (1,739,732 Sequences/146,084 Species/58,357 Interim Species)**
Every COI barcode record with a species level identification and a minimum sequence length of 500bp. This includes many species represented by only one or two specimens as well as all species with interim taxonomy.

○ **Public Record Barcode Database (555,693 Sequences/62,894 Species/14,985 Interim Species)**
All published COI records from BOLD and GenBank with a minimum sequence length of 500bp. This library is a collection of records from the published projects section of BOLD.

○ **Full Length Record Barcode Database (1,318,574 Sequences/132,900 Species/50,814 Interim Species)**
Subset of the Species library with a minimum sequence length of 640bp and containing both public and private records. This library is intended for short sequence identification as it provides maximum overlap with short reads from the barcode region of COI.

Enter sequences in fasta format:

```
>GadusMorhuaSCFAC839-06
CCTTTATCTCGTATTTGGTGCCTGAGCCGGCATAGTCGGAACAGCCCTAAGCCTGCTCATTCGAGCAGAGCTAAG
TCAACCTGGTGCACTTCTTGGTGATGATCAAATTTATAATGTGATCGTTACAGCGCACGCTTTCGTAATAATTTT
CTTTATAGTAATACCACTAATAATTGGAGGCTTTGGGAACTGACTCATTCCTCTAATGATCGGTGCACCAGATAT
AGCTTTCCCTCGAATAAATAACATAAGCTTCTGACTTCTTCCTCCATCTTTCCTGCTCCTTTTAGCATCCTCTGG
TGTAGAAGCTGGGGCTGGAACAGGCTGAACTGTCTATCCACCTTTAGCCGGAAACCTCGCTCATGCTGGGGCATC
TGTTGATCTCACTATTTTTTCTCTTCATCTAGCAGGGATTTCATCAATTCTTGGGGCAATTAATTTTATTACCAC
AATTATTAATATGAAACCTCCGGCAATTTCACAGTACCAAACACCCCTATTTGTTTGAGCAGTACTAATTACAGC
TGTGCTTCTACTATTATCTCTCCCCGTCTTAGCAGCTGGTATCACAATACTTCTAACTGACCGTAATCTTAATAC
TTCTTTCTTTGACCCTGCTGGAGGAGGTGATCCCATTTTATACCAACA
```

7.9 Quality Assurance

Extraction Control – negative control

This is included to check for extraction kit contamination. Only negligible DNA should be detected during quantification (<2ng/µl). If significant levels of DNA are detected, sterilize all equipment and repeat DNA extractions.

PCR Amplification – negative control

This is included to check for background laboratory contamination. No PCR product/band should be produced during procedure 7.4 PCR Product Check.

PCR Amplification – positive control

This should yield a strong PCR product/band (7.4 PCR Product Check) to ensure there are no issues with amplification.

7.10 Issues with Interpreting the Species Identification

1. **Every sample** should have **results** from searches in one (and occasionally two databases) with corresponding **screen-grabs and tree summarising** the results. BOLD may yield a completely unambiguous identification, but further interpretation of the results may be required to try and find the clearest species identification (as in 7.8). However, there will be **cases where** a species is lacking from the database, making a **species level identification impossible**. Despite this, the results may still yield other broader taxonomic information e.g. the genus or family the sample is likely to have originated from. It is important to note that the database is continually being updated and is becoming more comprehensive over time.

2. Another possible outcome is that despite examining the highest matching records and the tree, the **identification remains ambiguous** e.g. two species have 100% similarity to the uploaded sequence, so the end result is an ambiguous match to both. Some commercial groups of fishes, e.g. some *Thunnus* species of tunas and *Sebastes* species of "redfish", are very closely related/difficult to distinguish and further testing may be required to successfully identify them. In such circumstances **laboratories should** indicate on the official reporting that the **sample was identified to the genus level**, (e.g. *Thunnus* spp), and/or indicate the only two species creating ambiguity (e.g. *Thunnus albacares* or *Thunnus obesus*). However, this SOP is designed to be as universal as possible (i.e. applicable to the broadest range of fish products) and generates positive information to distinguish species (even if this may not always yield a match down to the species level).

3. In case the BOLD database is unable to identify your sequence, other publically available reference databases could be queried, e.g. GenBank (www.ncbi.nlm.nih.gov/). However, the correct use of these databases falls out of the scope of this SOP.

8. TROUBLE SHOOTING

Section 7.8 of the Standard Operating Procedure (SOP) explains step-by-step the methodology to correctly identify a sample using the Barcode of Life Database. The effectiveness of the SOP was tested by conducting a collaborative blind ring trial among 13 different laboratories. Yet, some inconsistencies were recorded in the correct identification of the samples. Here, guidelines are presented to address these specific issues in utilising BOLD for species identification.

8.1 How to interpret the results of the phylogenetic tree

BOLD uses neighbour-joining trees, which group sequences together hierarchically, based on the number of amino acid or nucleotide differences. The arrangement of the specimens in the tree is based on sequence similarities, with the sequences that are most similar placed closer together on the tree, and with the branch length proportional to the degree of similarity. The percentage of similarity between sequences can be measured against the legend (usually 2%). The longer the branch the more disparity between the sequences, as specimens of the same species have more similar sequences and cluster closer together than specimens from different species.

Robust species identification is straightforward when the unknown sequence is clustered within a monophyletic group containing reference sequences of just one species. In phylogenetics a clade or 'monophylum' is a group of species/records consisting of an ancestor and all its descendants. The ancestor may be an individual, a population or even a species. In Figure 1, a 'cladogram' or 'phylogenetic tree' of a biological group is depicted showing the last common ancestor at the bottom of the composite tree. The blue and red subgroups on the left and right hand side of the picture represent clades, or monophyletic taxonomic groups. Each shows the last common ancestor and all descendant branches. The green, central, subgroup is not a clade; it is not 'monophyletic'; rather it represents a 'paraphyletic' group, which is incomplete because the blue clade, although descended from it, is excluded.

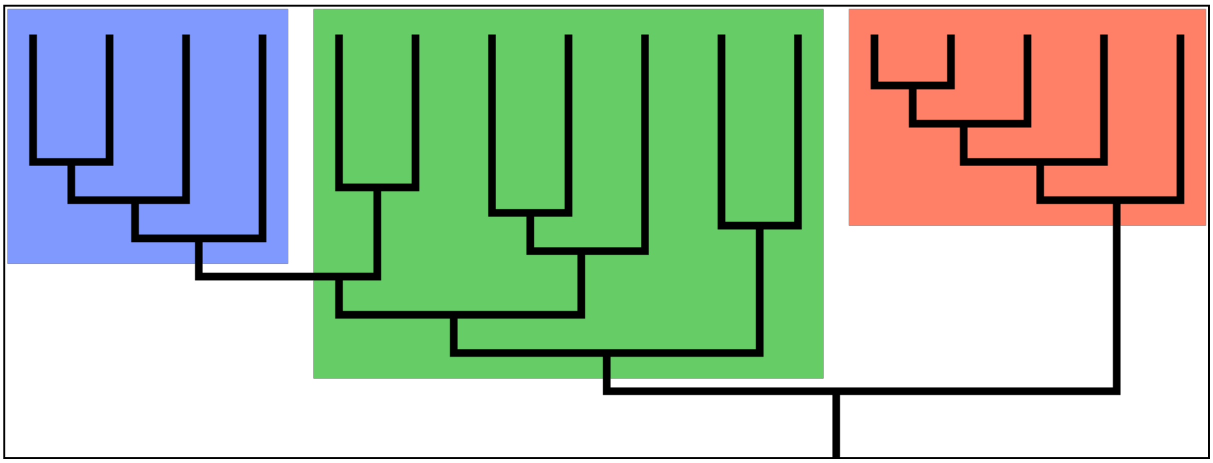


Figure 1 A tree or cladogram showing the last common ancestor at the bottom of the composite tree

Samples that fail to produce monophyletic trees, with species failing to form monophyletic groups can complicate the interpretation of results. Unexpected or ambiguous identification outcomes can reveal interesting findings, which could be associated with biologically relevant patterns, or they can reveal errors such as misidentification in the data base record or contamination of a sample (See below for a worked example; 8.2). For more information on how to build a Taxon ID Tree, and the parameters you can select to tailor your tree, please refer to the BOLD Handbook.

8.2 Does the sequences length influence the ability to identify the species?

The Taxon tree functionality allows generating dendograms from sequences using the Neighbour joining algorithm. This option is based on the distance matrix which is generated by aligning the sequences. BOLD provides an integrated alignment browser that allows users to analyse and edit sequences without needing to import them into a 3rd party software. The available BOLD alignment options are either based on amino-acid or nucleotide sequences. In turn, the distance matrix is calculated using the Kimura2 Parameter (default), Jukes Cantor or pairwise Distance models. More specifically, the tool identifies consensus bases from each group, compares them to those from the remaining sequences in other groups and then characterizes how unique each consensus base is.

Short sequences may influence the shape of the ID tree as they are less likely to align correctly with longer sequences and results should be interpreted with caution.

8.3 Identifying false identified species in the reference database

All submitted sequences on BOLD are assigned to a project, with a minimum of information requirements like the taxonomy, specimen detail and collection data. The person who creates a project is automatically assigned as the project manager. The process of creating a unique

sequence and project ID is useful to identify the sequences used throughout the barcoding process, but also as quality assurance.
Some records in the BOLD database may be flagged. This could indicate a contaminated sequence or a misidentified species. Flags serve two purposes: they act as an alert to inform project managers that an issue has been detected in their records, and they prevent a record from being included in the BOLD ID Engine and Taxonomy Browser.

Project managers can change the taxonomy of the sample or re-edit the sequence to resolve the flag. On the other hand, users that detect an issue with one of the records on BOLD Public Data Portal can log into BOLD to add a comment or a tag to that particular record. If you do not have a BOLD account, you can contact the BOLD support staff by emailing support@boldsystems.org. We strongly advice our SOP users to help the scientific community by alerting the BOLD support team in case of doubt. Through this collaborative action we all build on the quality assurance of the reference database.

Example:

Specimen sequences are identified by looking up the closest match of sequence similarity with the available records in the various BOLD search reference databases.

As part of the LABELFISH ring-trial, one of the test samples returned the following sequence:

```
TTNNAAAACGACGGCCAGTCGACTAATCATAAAGATATCGGCACCCTTTATCTCGTATTTGGTGCTTGAGCC
GGAATAGTAGGGACTGCCTTAAGTCTGCTCATTCGAGCGGAACTAAGCCAGCCTGGCGCCCTTTTAGGGGA
CGACCAAATCTATAATGTCATTGTTACAGCACACGCATTTGTAATAATTTTTTTCATAGTAATACCAATTATAAT
CGGAGGTTTCGGAAACTGACTTATTCCACTCATGATCGGTGCCCCCGACATAGCATTCCCCCGTATGAATAA
TATGAGCTTCTGACTCCTCCCCCCTTCATTCCTTCTACTCCTTGCCTCCTCTGGTGTTGAAGCCGGGGCCGG
TACTGGGTGAACAGTCTACCCACCACTAGCAGGGAACCTTGCCCACGCAGGTGCATCAGTTGACTTAACTAT
CTTTTCCCTCCACCTAGCCGGAATTTCATCCATTCTTGGGGCCATTAATTTCATTACTACCATTATTAATATGA
AACCCCCAGCCATTTCACAATACCAAACGCCACTATTTGTGTGAGCCGTCTTAATTACAGCTGTCCTTCTTCT
TCTGTCCCTCCCAGTACTTGCTGCCGGAATTACTATGCTCCTCACAGACCGAAACCTAAACACCACCTTCTTT
GACCCAGCCGGAGGAGGGGACCCAATTCTTTACCAACATCTTTTCTGATTCTTCGGACACCCTGAAGTGTCA
TAGCTGTTTCCNG
```

The BOLD search generated the outcome in Figure 2. At the top, the search result is returned, with no unambiguous match, but two potential species returned (*Boops boops* and *Oblada melanura*, top left), prompting BOLD to display the message

“A species match could not be made”, the queried specimen is likely to be one of the following”.

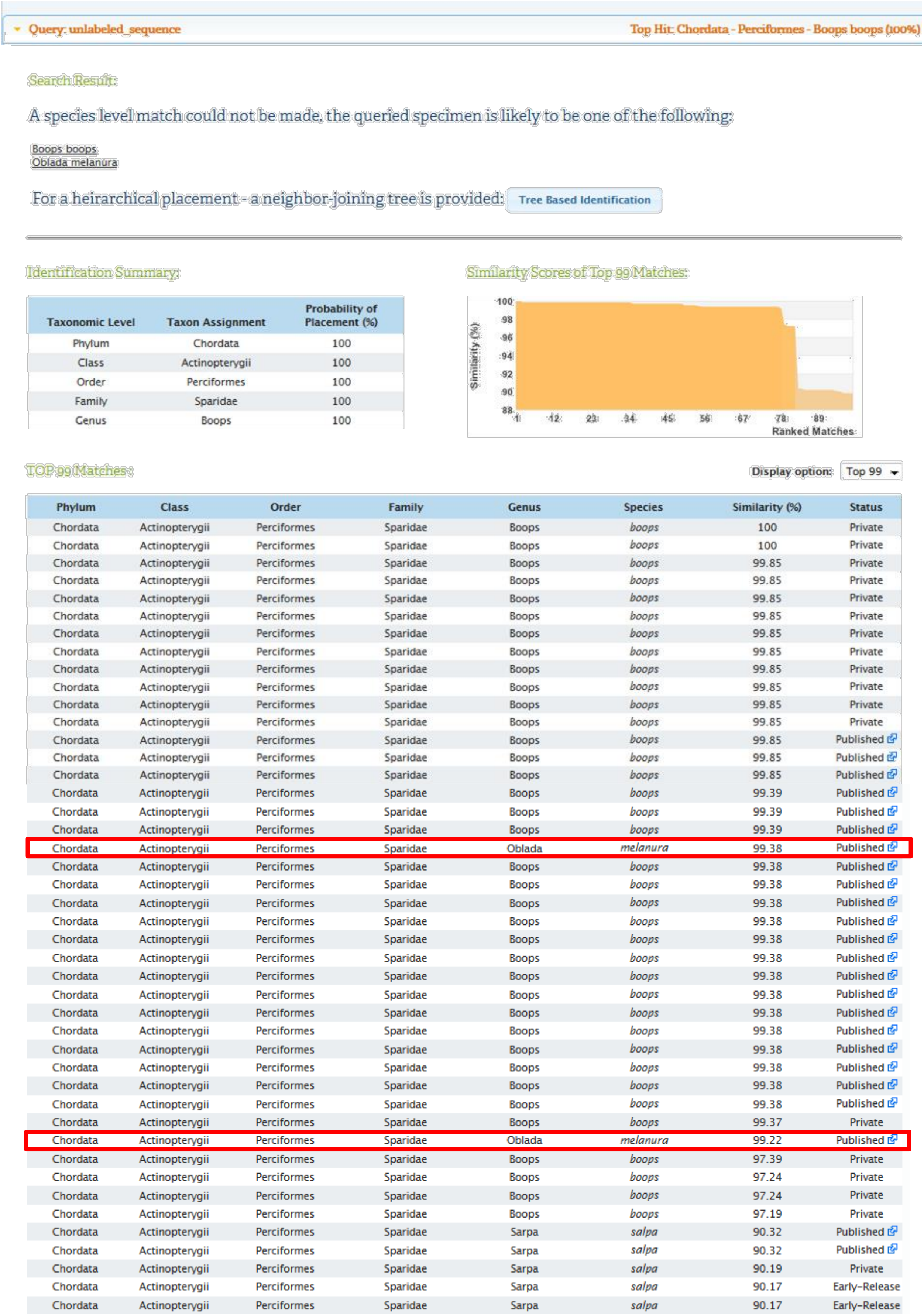
Only two of the top >99% matches are recorded as *Oblada melanura* (highlighted in red, Figure 2)

In order to make a robust identification several steps have to be taken:

- The graph entitled “Similarity Scores of Top 99 Matches” shows the percent similarity for each of 99 top matching records in the database against your consensus sequences. In this example, it is clearly illustrated only *Boops boops* are included in the top matches between 100% and 99.38% similarity.
- The “Top 99 matches” show that the lowest level of similarity (around 90%) appears with a species belonging to the same family, but different genus *Sarpa salpa.*

However, within the “Top 99 matches”, 2 records are reported, among *Boops boops*, as belonging to *Oblada melanura* (Figure 2), reducing the 100% certainty for correct species identification.

Apart from investigating the similarity levels of your sequences with the public database it is important to produce a simple Neighbour Joining tree (Figure 3).


▾ Query: unlabeled_sequence
Top Hit: Chordata - Perciformes - Boops boops (100%)
Tree Based Identification
Similarity (%)
100
98
96
94
92
90
88
1
12
23
34
45
56
67
78
89
Ranked Matches
Display option:
Top 99 ▾

Search Result:

A species level match could not be made, the queried specimen is likely to be one of the following:

Boops boops
Oblada melanura

For a heirarchical placement - a neighbor-joining tree is provided:

Identification Summary:

Taxonomic Level	Taxon Assignment	Probability of Placement (%)
Phylum	Chordata	100
Class	Actinopterygii	100
Order	Perciformes	100
Family	Sparidae	100
Genus	Boops	100

Similarity Scores of Top 99 Matches:

TOP 99 Matches :

Phylum	Class	Order	Family	Genus	Species	Similarity (%)	Status
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	100	Private
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	100	Private
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	99.85	Private
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	99.85	Private
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	99.85	Private
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	99.85	Private
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	99.85	Private
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	99.85	Private
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	99.85	Private
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	99.85	Private
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	99.85	Private
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	99.85	Private
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	99.85	Published
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	99.85	Published
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	99.85	Published
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	99.39	Published
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	99.39	Published
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	99.39	Published
Chordata	Actinopterygii	Perciformes	Sparidae	Oblada	*melanura*	99.38	Published
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	99.38	Published
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	99.38	Published
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	99.38	Published
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	99.38	Published
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	99.38	Published
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	99.38	Published
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	99.38	Published
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	99.38	Published
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	99.38	Published
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	99.38	Published
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	99.38	Published
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	99.38	Published
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	99.38	Published
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	99.38	Published
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	99.37	Private
Chordata	Actinopterygii	Perciformes	Sparidae	Oblada	*melanura*	99.22	Published
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	97.39	Private
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	97.24	Private
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	97.24	Private
Chordata	Actinopterygii	Perciformes	Sparidae	Boops	*boops*	97.19	Private
Chordata	Actinopterygii	Perciformes	Sparidae	Sarpa	*salpa*	90.32	Published
Chordata	Actinopterygii	Perciformes	Sparidae	Sarpa	*salpa*	90.32	Published
Chordata	Actinopterygii	Perciformes	Sparidae	Sarpa	*salpa*	90.19	Private
Chordata	Actinopterygii	Perciformes	Sparidae	Sarpa	*salpa*	90.17	Early-Release
Chordata	Actinopterygii	Perciformes	Sparidae	Sarpa	*salpa*	90.17	Early-Release

Figure 2 Partial print screen of the “Top 99 matches” table when conducting a BOLD search.

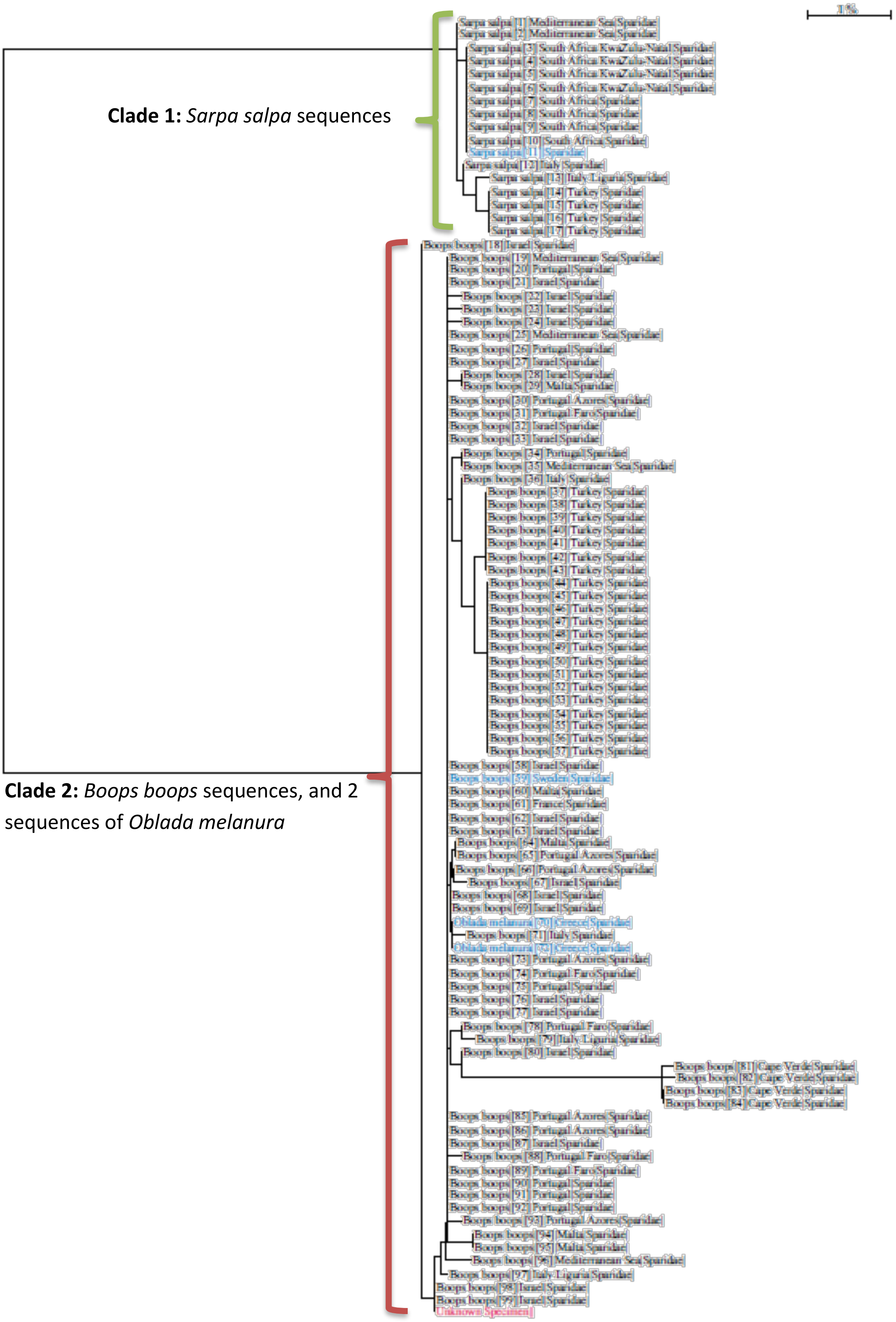
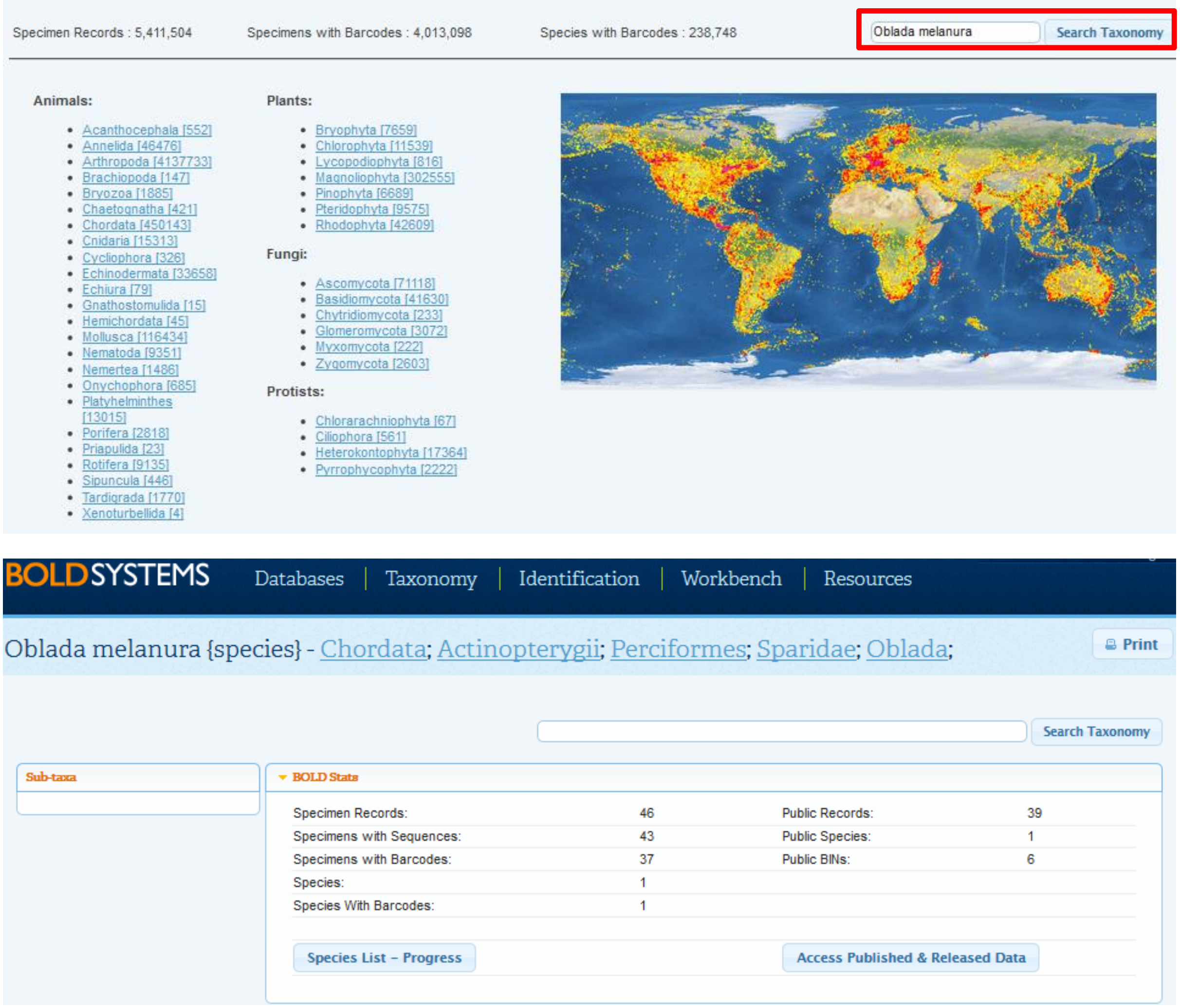

1%
Clade 1: Sarpa salpa sequences
Clade 2: Boops boops sequences, and 2 sequences of Oblada melanura
Sarpa salpa|[1]|Mediterranean Sea|Sparidae|
Sarpa salpa|[2]|Mediterranean Sea|Sparidae|
Sarpa salpa|[3]|South Africa KwaZulu-Natal|Sparidae|
Sarpa salpa|[4]|South Africa KwaZulu-Natal|Sparidae|
Sarpa salpa|[5]|South Africa KwaZulu-Natal|Sparidae|
Sarpa salpa|[6]|South Africa KwaZulu-Natal|Sparidae|
Sarpa salpa|[7]|South Africa|Sparidae|
Sarpa salpa|[8]|South Africa|Sparidae|
Sarpa salpa|[9]|South Africa|Sparidae|
Sarpa salpa|[10]|South Africa|Sparidae|
Sarpa salpa|[11]||Sparidae|
Sarpa salpa|[12]|Italy|Sparidae|
Sarpa salpa|[13]|Italy Liguria|Sparidae|
Sarpa salpa|[14]|Turkey|Sparidae|
Sarpa salpa|[15]|Turkey|Sparidae|
Sarpa salpa|[16]|Turkey|Sparidae|
Sarpa salpa|[17]|Turkey|Sparidae|
Boops boops|[18]|Israel|Sparidae|
Boops boops|[19]|Mediterranean Sea|Sparidae|
Boops boops|[20]|Portugal|Sparidae|
Boops boops|[21]|Israel|Sparidae|
Boops boops|[22]|Israel|Sparidae|
Boops boops|[23]|Israel|Sparidae|
Boops boops|[24]|Israel|Sparidae|
Boops boops|[25]|Mediterranean Sea|Sparidae|
Boops boops|[26]|Portugal|Sparidae|
Boops boops|[27]|Israel|Sparidae|
Boops boops|[28]|Israel|Sparidae|
Boops boops|[29]|Malta|Sparidae|
Boops boops|[30]|Portugal Azores|Sparidae|
Boops boops|[31]|Portugal Faro|Sparidae|
Boops boops|[32]|Israel|Sparidae|
Boops boops|[33]|Israel|Sparidae|
Boops boops|[34]|Portugal|Sparidae|
Boops boops|[35]|Mediterranean Sea|Sparidae|
Boops boops|[36]|Italy|Sparidae|
Boops boops|[37]|Turkey|Sparidae|
Boops boops|[38]|Turkey|Sparidae|
Boops boops|[39]|Turkey|Sparidae|
Boops boops|[40]|Turkey|Sparidae|
Boops boops|[41]|Turkey|Sparidae|
Boops boops|[42]|Turkey|Sparidae|
Boops boops|[43]|Turkey|Sparidae|
Boops boops|[44]|Turkey|Sparidae|
Boops boops|[45]|Turkey|Sparidae|
Boops boops|[46]|Turkey|Sparidae|
Boops boops|[47]|Turkey|Sparidae|
Boops boops|[48]|Turkey|Sparidae|
Boops boops|[49]|Turkey|Sparidae|
Boops boops|[50]|Turkey|Sparidae|
Boops boops|[51]|Turkey|Sparidae|
Boops boops|[52]|Turkey|Sparidae|
Boops boops|[53]|Turkey|Sparidae|
Boops boops|[54]|Turkey|Sparidae|
Boops boops|[55]|Turkey|Sparidae|
Boops boops|[56]|Turkey|Sparidae|
Boops boops|[57]|Turkey|Sparidae|
Boops boops|[58]|Israel|Sparidae|
Boops boops|[59]|Sweden|Sparidae|
Boops boops|[60]|Malta|Sparidae|
Boops boops|[61]|France|Sparidae|
Boops boops|[62]|Israel|Sparidae|
Boops boops|[63]|Israel|Sparidae|
Boops boops|[64]|Malta|Sparidae|
Boops boops|[65]|Portugal Azores|Sparidae|
Boops boops|[66]|Portugal Azores|Sparidae|
Boops boops|[67]|Israel|Sparidae|
Boops boops|[68]|Israel|Sparidae|
Boops boops|[69]|Israel|Sparidae|
Oblada melanura|[70]|Greece|Sparidae|
Boops boops|[71]|Italy|Sparidae|
Oblada melanura|[72]|Greece|Sparidae|
Boops boops|[73]|Portugal Azores|Sparidae|
Boops boops|[74]|Portugal Faro|Sparidae|
Boops boops|[75]|Portugal|Sparidae|
Boops boops|[76]|Israel|Sparidae|
Boops boops|[77]|Israel|Sparidae|
Boops boops|[78]|Portugal Faro|Sparidae|
Boops boops|[79]|Italy Liguria|Sparidae|
Boops boops|[80]|Israel|Sparidae|
Boops boops|[81]|Cape Verde|Sparidae|
Boops boops|[82]|Cape Verde|Sparidae|
Boops boops|[83]|Cape Verde|Sparidae|
Boops boops|[84]|Cape Verde|Sparidae|
Boops boops|[85]|Portugal Azores|Sparidae|
Boops boops|[86]|Portugal Azores|Sparidae|
Boops boops|[87]|Israel|Sparidae|
Boops boops|[88]|Portugal Faro|Sparidae|
Boops boops|[89]|Portugal Faro|Sparidae|
Boops boops|[90]|Portugal|Sparidae|
Boops boops|[91]|Portugal|Sparidae|
Boops boops|[92]|Portugal|Sparidae|
Boops boops|[93]|Portugal Azores|Sparidae|
Boops boops|[94]|Malta|Sparidae|
Boops boops|[95]|Malta|Sparidae|
Boops boops|[96]|Mediterranean Sea|Sparidae|
Boops boops|[97]|Italy Liguria|Sparidae|
Boops boops|[98]|Israel|Sparidae|
Boops boops|[99]|Israel|Sparidae|
Unknown Specimen|

Figure 3 A Neighbouring tree can be consulted when identifying a sequence on BOLD. The example shown here illustrates that the unknown sequence (in red) is clustered within a clade of *Boops boops* records, a second clade is shown composing only of *Sarpa salpa* sequences. This monophyletic clade is clearly well separated from our unknown sequence. Within the *Boops boops* clade however, two sequences are found originating from *Oblada melanura*.

In the above tree (Figure 3) the queried sequence is highlighted in red and is nested in the tree dominated by *Boops boops* sequences. Only two sequences in this clade are recorded as *Oblada melanura*, while the BOLD database contains many more sequences of this particular species, which in fact show a similarity score far lower than 1%. The threshold in nucleotide difference between species is commonly taken at 2% (Avise 2000). This raises suspicion whether or not these sequences are to be trusted. Further investigation would be recommended.

Consulting the BOLD Taxonomy browser allows users to examine the number of records available in BOLD of any species (Figure 4). For this specific case the Taxonomy Browser shows there are in total 43 sequence records of *Oblada melanura* and 103 *Boops boops.*



Specimen Records : 5,411,504 Specimens with Barcodes : 4,013,098 Species with Barcodes : 238,748

Oblada melanura Search Taxonomy

Animals:

- Acanthocephala [552]
- Annelida [46476]
- Arthropoda [4137733]
- Brachiopoda [147]
- Bryozoa [1885]
- Chaetognatha [421]
- Chordata [450143]
- Cnidaria [15313]
- Cycliophora [326]
- Echinodermata [33658]
- Echiura [79]
- Gnathostomulida [15]
- Hemichordata [45]
- Mollusca [116434]
- Nematoda [9351]
- Nemertea [1486]
- Onychophora [685]
- Platyhelminthes [13015]
- Porifera [2818]
- Priapulida [23]
- Rotifera [9135]
- Sipuncula [446]
- Tardigrada [1770]
- Xenoturbellida [4]

Plants:

- Bryophyta [7659]
- Chlorophyta [11539]
- Lycopodiophyta [816]
- Magnoliophyta [302555]
- Pinophyta [6689]
- Pteridophyta [9575]
- Rhodophyta [42609]

Fungi:

- Ascomycota [71118]
- Basidiomycota [41630]
- Chytridiomycota [233]
- Glomeromycota [3072]
- Myxomycota [222]
- Zygomycota [2603]

Protists:

- Chlorarachniophyta [67]
- Ciliophora [561]
- Heterokontophyta [17364]
- Pyrrophycophyta [2222]

Oblada melanura {species} - Chordata; Actinopterygii; Perciformes; Sparidae; Oblada;

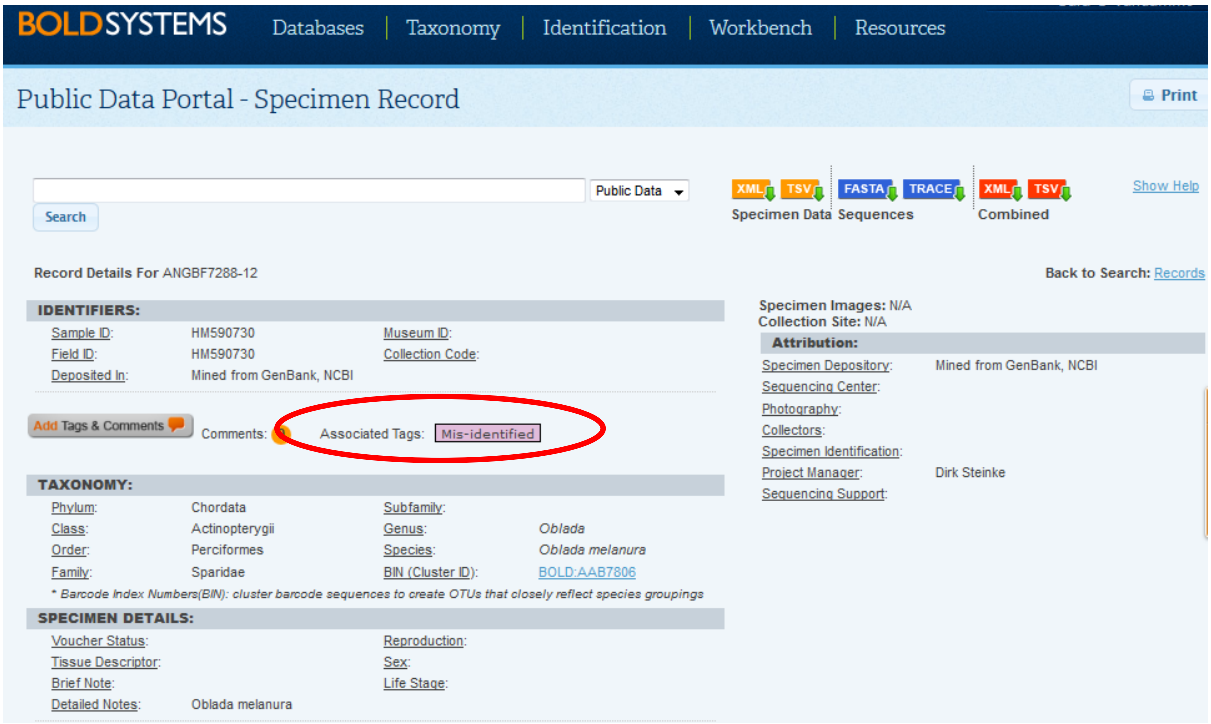
BOLD Stats

Specimen Records:	46	Public Records:	39
Specimens with Sequences:	43	Public Species:	1
Specimens with Barcodes:	37	Public BINs:	6
Species:	1		
Species With Barcodes:	1		

Figure 4 The top panel illustrates the BOLD Taxonomy search engine which is accessible through the home page by clicking on the Taxonomy tab. Whenever needed, taxa can be found by using the search option. The lower panel depicts the outcome for *Oblada melanura*, Illustrating that the BOLD reference database contains 43 sequences.

Second, in the “Top 99 matches” table (see Figure 2), one can check the details of all publically available records by clicking the blue arrow symbol: . The specimen record of both these *Oblada melanura* records appear to have been flagged and have received a tag illustrating the sequence could be misidentified (see red, Figure 5).

Consequently, the identification of the sequences as *Boops boops* is substantially strengthened as we have some grounds to discount the *Oblada melanura* records as being unreliable



Public Data Portal - Specimen Record

Record Details For ANGBF7288-12

Back to Search: Records

IDENTIFIERS:

Sample ID: HM590730
Field ID: HM590730
Deposited In: Mined from GenBank, NCBI
Museum ID:
Collection Code:

Comments:
Associated Tags: Mis-identified

TAXONOMY:

Phylum: Chordata
Class: Actinopterygii
Order: Perciformes
Family: Sparidae
Subfamily:
Genus: *Oblada*
Species: *Oblada melanura*
BIN (Cluster ID): BOLD:AAB7806

* *Barcode Index Numbers(BIN): cluster barcode sequences to create OTUs that closely reflect species groupings*

SPECIMEN DETAILS:

Voucher Status:
Tissue Descriptor:
Brief Note:
Detailed Notes: Oblada melanura
Reproduction:
Sex:
Life Stage:

Specimen Images: N/A
Collection Site: N/A

Attribution:

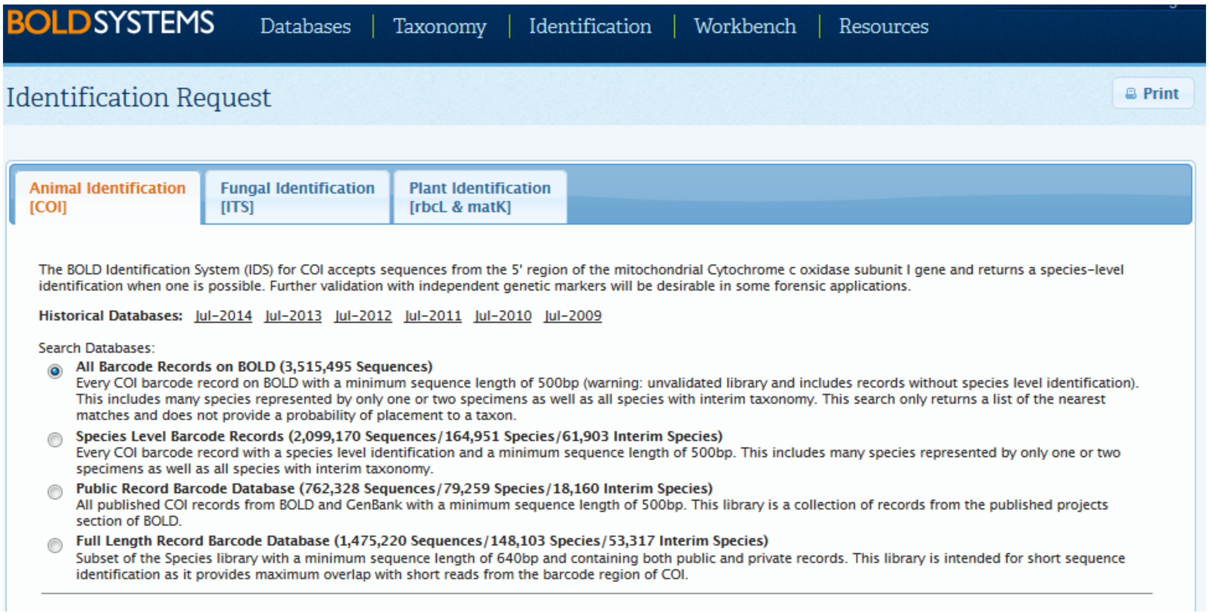
Specimen Depository: Mined from GenBank, NCBI
Sequencing Center:
Photography:
Collectors:
Specimen Identification:
Project Manager: Dirk Steinke
Sequencing Support:

Figure 5. Specimen record of one of the *Oblada melanura* sequences. Whenever a sequence has been flagged, or a tag has been added to indicate issues with the relevant sequences, this would be observed here.

8.4 Identifying unknown species

If a sample comes from a rare or exotic fish, there may be no matching records in the “Species Level Barcode Records” database that demonstrate high levels of sequence similarity. An additional search is also recommended by selecting the “All Public Records on BOLD”, this is the broadest database as it includes both sequences available on BOLD as well as on GenBank (Figure 6). This may allow the user to assign the sequence to a genus or family and still provide potentially useful information (this is one of the big advantages of utilising a big public database). The search is conducted and interpreted in the same fashion as the SOP but the selection of this expanded database must be recorded in the results.

.


BOLDSYSTEMS
Databases
Taxonomy
Identification
Workbench
Resources
Print
Animal Identification [COI]
Fungal Identification [ITS]
Plant Identification [rbcL & matK]

Identification Request

The BOLD Identification System (IDS) for COI accepts sequences from the 5' region of the mitochondrial Cytochrome c oxidase subunit I gene and returns a species-level identification when one is possible. Further validation with independent genetic markers will be desirable in some forensic applications.

Historical Databases: Jul-2014 Jul-2013 Jul-2012 Jul-2011 Jul-2010 Jul-2009

Search Databases:

● **All Barcode Records on BOLD (3,515,495 Sequences)**
Every COI barcode record on BOLD with a minimum sequence length of 500bp (warning: unvalidated library and includes records without species level identification). This includes many species represented by only one or two specimens as well as all species with interim taxonomy. This search only returns a list of the nearest matches and does not provide a probability of placement to a taxon.

○ **Species Level Barcode Records (2,099,170 Sequences/164,951 Species/61,903 Interim Species)**
Every COI barcode record with a species level identification and a minimum sequence length of 500bp. This includes many species represented by only one or two specimens as well as all species with interim taxonomy.

○ **Public Record Barcode Database (762,328 Sequences/79,259 Species/18,160 Interim Species)**
All published COI records from BOLD and GenBank with a minimum sequence length of 500bp. This library is a collection of records from the published projects section of BOLD.

○ **Full Length Record Barcode Database (1,475,220 Sequences/148,103 Species/53,317 Interim Species)**
Subset of the Species library with a minimum sequence length of 640bp and containing both public and private records. This library is intended for short sequence identification as it provides maximum overlap with short reads from the barcode region of COI.

Figure 6 screenshot of the type of reference database option one can pick from on the BOLD database. The selected one illustrated the “All Public Records” option.

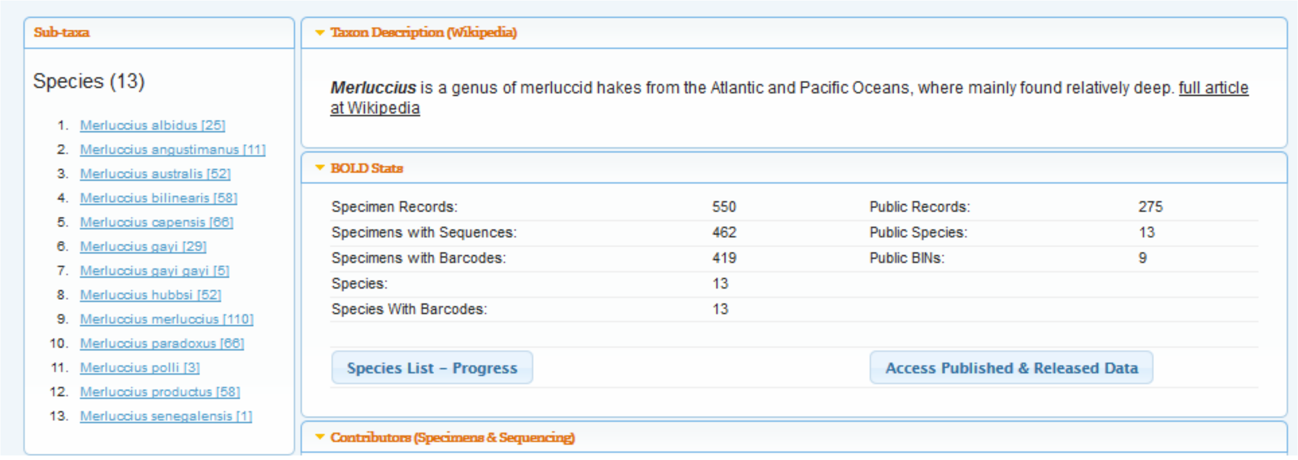
If *a-priori* information about the supposed species is available, i.e. from the label, it would be recommended to check if the species and the commercially significant counterparts belonging to the same genus are represented in the Taxonomy Browser. This Browser is a synthetic database that allows users to examine the progress of DNA barcoding by browsing through the different levels of the taxonomic hierarchy available on BOLD. Within this browser users are able to select between animal, plant, fungal and protist kingdoms and navigate from phylum to species level. To look up a specific taxon directly, use the search function by entering a taxonomic name into the search bar at the top of the Taxonomic Browser or on the BOLD main page.). Comparing this species list with database on marine organisms like Fishbase (*www.**fishbase**.org*) or World Register of Marine Species (WoRMS, *www.marinespecies.org*) allows users to determine whether all species belonging to a certain genus or family are available in the BOLD database:

Example:

Within the collaborative ring trial, two of the unknown samples belonged to the genus *Merluccius*. The Taxonomy Browser of BOLD provides an overview of all the species recorded belonging to this genus (Figure 7). Comparing this species list with database on marine organisms like Fishbase (*www.**fishbase**.org*) or World Register of Marine Species (WoRMS, *www.marinespecies.org*) shows that some species are not yet in the BOLD database:

- *Merluccius gayi peruanus*
- *Merluccius hernandezi*
- *Merluccius tasmanicus*
- *Merluccius patagonicus*

When a sequence from one of these species is queried in BOLD, high confidence identification will only be obtained for the genus level. Selecting the “All Public Records on BOLD”, will not increase the level of identification as neither of these species are included on GenBank. This should however not reduce the level of confidence in identifying the other *Merluccius* species.


Sub-taxa
Taxon Description (Wikipedia)
BOLD Stats
Species List – Progress
Access Published & Released Data
Contributors (Specimens & Sequencing)

Species (13)

1. Merluccius albidus [25]
2. Merluccius angustimanus [11]
3. Merluccius australis [52]
4. Merluccius bilinearis [58]
5. Merluccius capensis [66]
6. Merluccius gayi [29]
7. Merluccius gayi gayi [5]
8. Merluccius hubbsi [52]
9. Merluccius merluccius [110]
10. Merluccius paradoxus [66]
11. Merluccius polli [3]
12. Merluccius productus [58]
13. Merluccius senegalensis [1]

Merluccius is a genus of merluccid hakes from the Atlantic and Pacific Oceans, where mainly found relatively deep. full article at Wikipedia

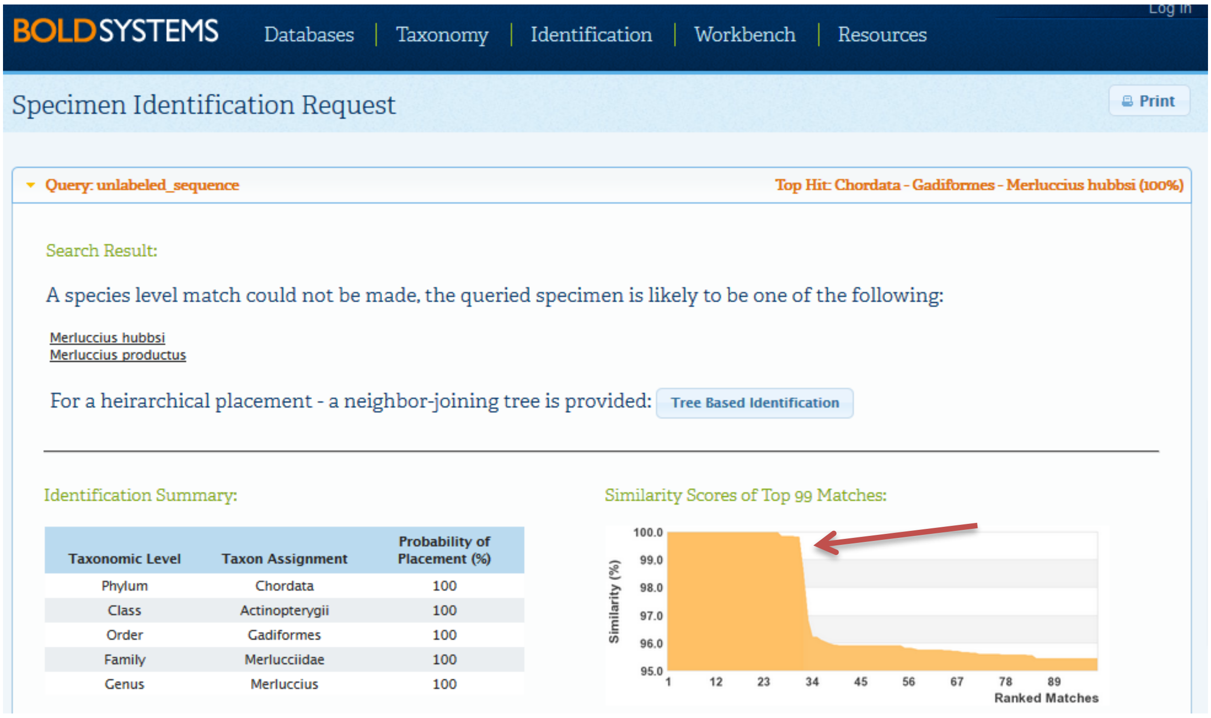
Specimen Records:	550	Public Records:	275
Specimens with Sequences:	462	Public Species:	13
Specimens with Barcodes:	419	Public BINs:	9
Species:	13		
Species With Barcodes:	13		

Figure 7 Taxonomy query in BOLD on the genus *Merluccius*.

An example is depicted in Figure 8 where the following sequence belonging to the *Merluccius* genus:

```
TTTGGTGCTTGAGCCGGCATAGTCGGAACAGCCCTAAGCCTGCTCATCCGGGCAGAACTAAGTCAACCCGG
CGCACTCCTGGGCGACGATCAAATTTATAACGTAATCGTCACGGCACACGCCTTCGTAATAATTTTCTTTATA
GTAATACCGTTAATAATTGGAGGCTTTGGAAACTGACTCGTTCCCCTAATGATCGGAGCCCCCGACATGGCC
TTCCCCCGAATAAATAACATAAGCTTCTGACTTCTTCCTCCGTCTTTCCTGCTCCTCCTAGCATCCTCCGGAG
TAGAAGCCGGAGCCGGGACAGGTTGAACAGTATACCCCCCTCTTGCAAGCAATCTTGCCCACGCTGGCGCC
AGCGTGGACCTCACTATTTTTTCACTTCACTTAGCAGGCGTTTCCTCAATTCTAGGAGCAATTAATTTCATTAC
TACTATTATTAATATGAAACCCCCTGCAATCTCACAGTACCAGACACCCCTCTTTGTTTGATCCGTCCTTATTA
CAGCTGTCCTCCTCCTACTCTCCCTGCCCGTCTTAGCCGCCGGCATCACAATACTACTAACTGACCGAAACC
TCAACACCTCCTTCTTTGACCCCGCCGGTGGAGGGGACCCCATCCTATACCAGCATTT
```

is identified using BOLD. The similarity score illustrate the sequences belongs to *Merluccius hubbsi* (% similarity 100-99.82, Figure 8). The next encountered species belongs to *Merluccius productus* (98.58%), after that the level of similarity drops steeply for *Merluccius albidus* (96.23%, see red arrow Figure 8).



Specimen Identification Request

Query: unlabeled_sequence

Top Hit: Chordata - Gadiformes - Merluccius hubbsi (100%)

Search Result:

A species level match could not be made, the queried specimen is likely to be one of the following:

Merluccius hubbsi
Merluccius productus

For a heirarchical placement - a neighbor-joining tree is provided:

Identification Summary:

Taxonomic Level	Taxon Assignment	Probability of Placement (%)
Phylum	Chordata	100
Class	Actinopterygii	100
Order	Gadiformes	100
Family	Merlucciidae	100
Genus	Merluccius	100

Similarity Scores of Top 99 Matches:

Figure 8 Specimen Identification output on BOLD. Top 99 Matches show the highest similarity scores are obtained for the species *Merluccius hubbsi* (100-98.58). Moving further down in the species list, the similarity score drops sharply to 96.82%.

Constructing a phylogenetic tree will help to further specify which species match has the highest confidence (Figure 9). The unknown sequence is clearly nested within sequences originating from *Merluccius hubbsi.* All other *Merluccius* species form a separate branch some distance away from our unknown specimen. Except for one sample recorded as *Merluccius productus* (Figure 9, blue arrow). Although this specific record does not have a tag added, the appearance of all other *Merluccius productus* records in another clade may indicate this particular sequence was indeed misidentified or should at least be treated with less confidence. As users you may want to raise your concern by contacting the BOLD support staff by emailing support@boldsystems.org or if you have a BOLD account by adding a comment or a tag to that particular record.

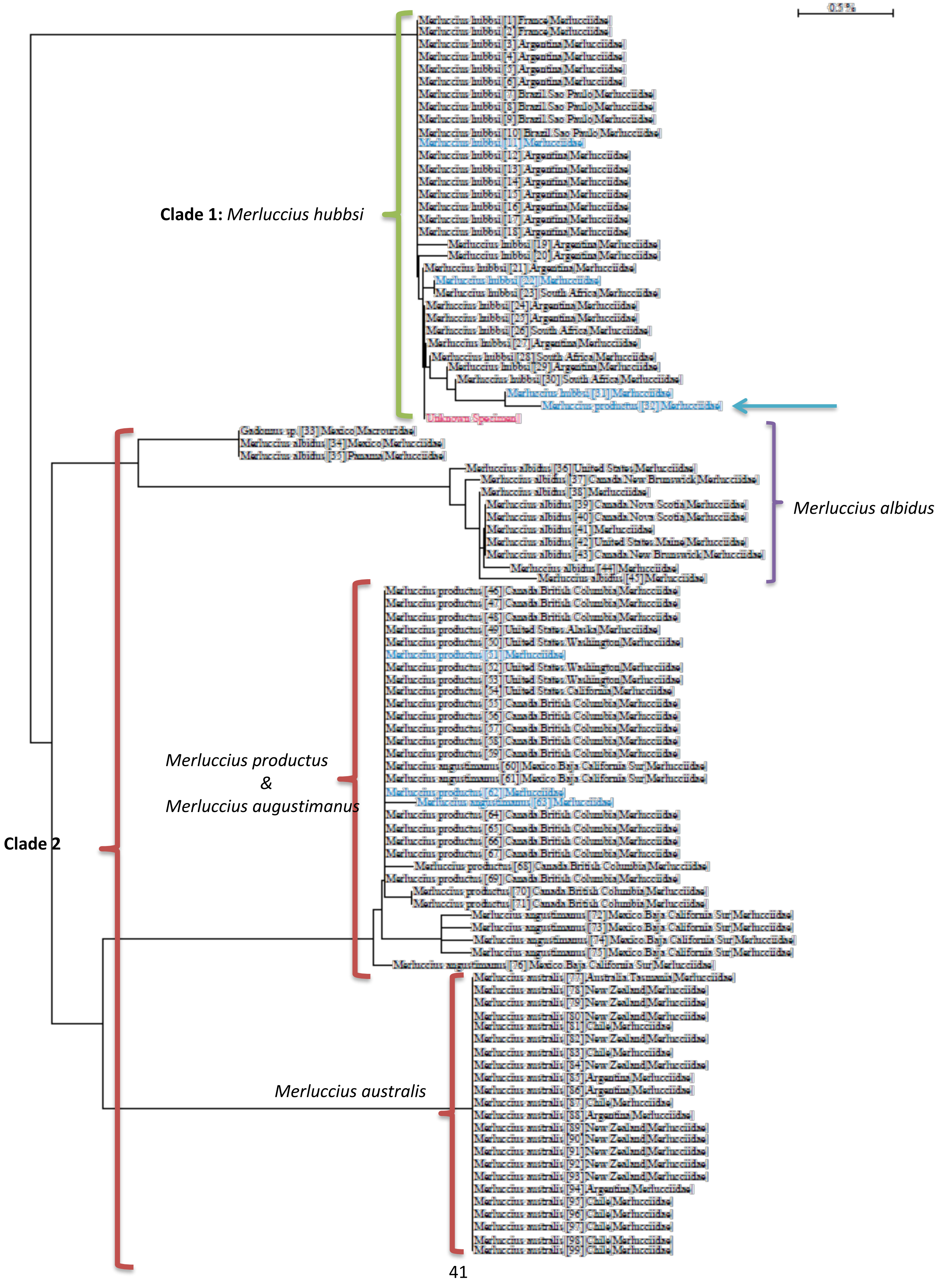
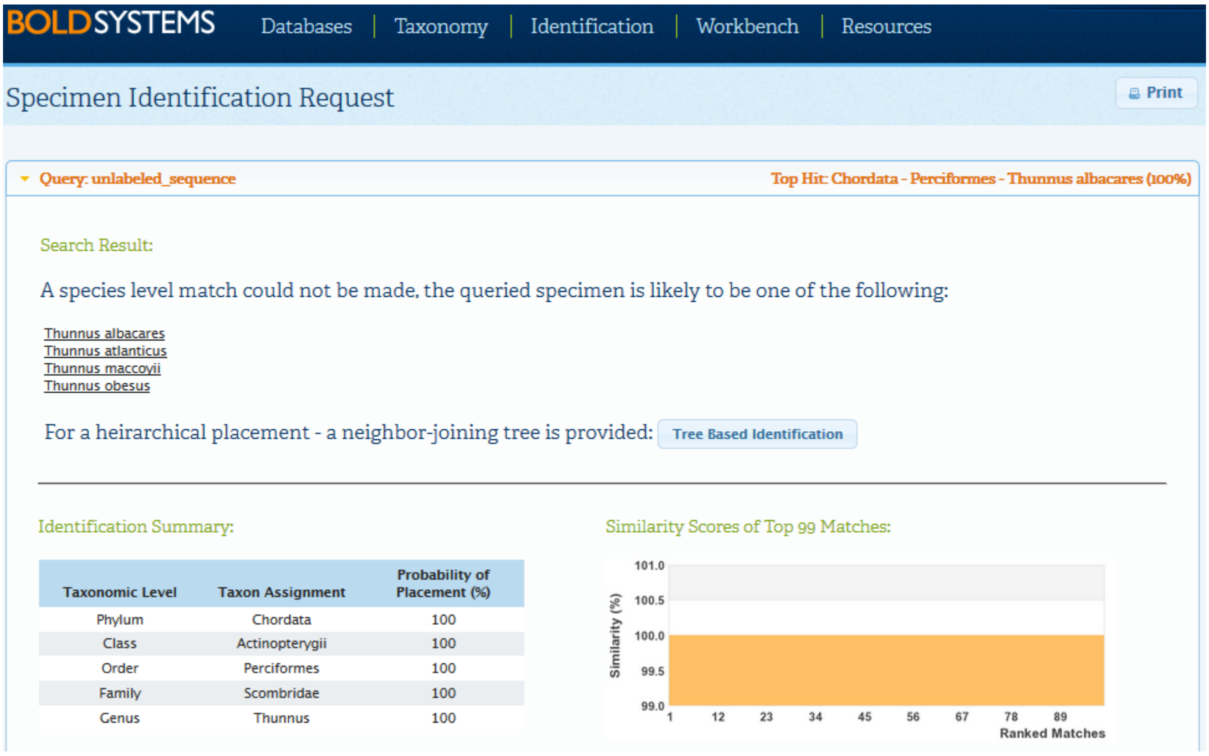

0.5 %
Clade 1: Merluccius hubbsi
Merluccius hubbsi|[1]|France|Merlucciidae|
Merluccius hubbsi|[2]|France|Merlucciidae|
Merluccius hubbsi|[3]|Argentina|Merlucciidae|
Merluccius hubbsi|[4]|Argentina|Merlucciidae|
Merluccius hubbsi|[5]|Argentina|Merlucciidae|
Merluccius hubbsi|[6]|Argentina|Merlucciidae|
Merluccius hubbsi|[7]|Brazil Sao Paulo|Merlucciidae|
Merluccius hubbsi|[8]|Brazil Sao Paulo|Merlucciidae|
Merluccius hubbsi|[9]|Brazil Sao Paulo|Merlucciidae|
Merluccius hubbsi|[10]|Brazil Sao Paulo|Merlucciidae|
Merluccius hubbsi|[11]|Merlucciidae|
Merluccius hubbsi|[12]|Argentina|Merlucciidae|
Merluccius hubbsi|[13]|Argentina|Merlucciidae|
Merluccius hubbsi|[14]|Argentina|Merlucciidae|
Merluccius hubbsi|[15]|Argentina|Merlucciidae|
Merluccius hubbsi|[16]|Argentina|Merlucciidae|
Merluccius hubbsi|[17]|Argentina|Merlucciidae|
Merluccius hubbsi|[18]|Argentina|Merlucciidae|
Merluccius hubbsi|[19]|Argentina|Merlucciidae|
Merluccius hubbsi|[20]|Argentina|Merlucciidae|
Merluccius hubbsi|[21]|Argentina|Merlucciidae|
Merluccius hubbsi|[22]|Merlucciidae|
Merluccius hubbsi|[23]|South Africa|Merlucciidae|
Merluccius hubbsi|[24]|Argentina|Merlucciidae|
Merluccius hubbsi|[25]|Argentina|Merlucciidae|
Merluccius hubbsi|[26]|South Africa|Merlucciidae|
Merluccius hubbsi|[27]|Argentina|Merlucciidae|
Merluccius hubbsi|[28]|South Africa|Merlucciidae|
Merluccius hubbsi|[29]|Argentina|Merlucciidae|
Merluccius hubbsi|[30]|South Africa|Merlucciidae|
Merluccius hubbsi|[31]|Merlucciidae|
Merluccius productus|[32]|Merlucciidae|
Unknown Specimen|
Gadomus sp.|[33]|Mexico|Macrouridae|
Merluccius albidus|[34]|Mexico|Merlucciidae|
Merluccius albidus|[35]|Panama|Merlucciidae|
Merluccius albidus|[36]|United States|Merlucciidae|
Merluccius albidus|[37]|Canada New Brunswick|Merlucciidae|
Merluccius albidus|[38]|Merlucciidae|
Merluccius albidus|[39]|Canada Nova Scotia|Merlucciidae|
Merluccius albidus|[40]|Canada Nova Scotia|Merlucciidae|
Merluccius albidus|[41]|Merlucciidae|
Merluccius albidus|[42]|United States Maine|Merlucciidae|
Merluccius albidus|[43]|Canada New Brunswick|Merlucciidae|
Merluccius albidus|[44]|Merlucciidae|
Merluccius albidus|[45]|Merlucciidae|
Merluccius albidus
Merluccius productus|[46]|Canada British Columbia|Merlucciidae|
Merluccius productus|[47]|Canada British Columbia|Merlucciidae|
Merluccius productus|[48]|Canada British Columbia|Merlucciidae|
Merluccius productus|[49]|United States Alaska|Merlucciidae|
Merluccius productus|[50]|United States Washington|Merlucciidae|
Merluccius productus|[51]|Merlucciidae|
Merluccius productus|[52]|United States Washington|Merlucciidae|
Merluccius productus|[53]|United States Washington|Merlucciidae|
Merluccius productus|[54]|United States California|Merlucciidae|
Merluccius productus|[55]|Canada British Columbia|Merlucciidae|
Merluccius productus|[56]|Canada British Columbia|Merlucciidae|
Merluccius productus|[57]|Canada British Columbia|Merlucciidae|
Merluccius productus|[58]|Canada British Columbia|Merlucciidae|
Merluccius productus|[59]|Canada British Columbia|Merlucciidae|
Merluccius angustimanus|[60]|Mexico Baja California Sur|Merlucciidae|
Merluccius angustimanus|[61]|Mexico Baja California Sur|Merlucciidae|
Merluccius productus|[62]|Merlucciidae|
Merluccius angustimanus|[63]|Merlucciidae|
Merluccius productus|[64]|Canada British Columbia|Merlucciidae|
Merluccius productus|[65]|Canada British Columbia|Merlucciidae|
Merluccius productus|[66]|Canada British Columbia|Merlucciidae|
Merluccius productus|[67]|Canada British Columbia|Merlucciidae|
Merluccius productus|[68]|Canada British Columbia|Merlucciidae|
Merluccius productus|[69]|Canada British Columbia|Merlucciidae|
Merluccius productus|[70]|Canada British Columbia|Merlucciidae|
Merluccius productus|[71]|Canada British Columbia|Merlucciidae|
Merluccius angustimanus|[72]|Mexico Baja California Sur|Merlucciidae|
Merluccius angustimanus|[73]|Mexico Baja California Sur|Merlucciidae|
Merluccius angustimanus|[74]|Mexico Baja California Sur|Merlucciidae|
Merluccius angustimanus|[75]|Mexico Baja California Sur|Merlucciidae|
Merluccius angustimanus|[76]|Mexico Baja California Sur|Merlucciidae|
Merluccius productus
&
Merluccius augustimanus
Clade 2
Merluccius australis|[77]|Australia Tasmania|Merlucciidae|
Merluccius australis|[78]|New Zealand|Merlucciidae|
Merluccius australis|[79]|New Zealand|Merlucciidae|
Merluccius australis|[80]|New Zealand|Merlucciidae|
Merluccius australis|[81]|Chile|Merlucciidae|
Merluccius australis|[82]|New Zealand|Merlucciidae|
Merluccius australis|[83]|Chile|Merlucciidae|
Merluccius australis|[84]|New Zealand|Merlucciidae|
Merluccius australis|[85]|Argentina|Merlucciidae|
Merluccius australis|[86]|Argentina|Merlucciidae|
Merluccius australis|[87]|Chile|Merlucciidae|
Merluccius australis|[88]|Argentina|Merlucciidae|
Merluccius australis|[89]|New Zealand|Merlucciidae|
Merluccius australis|[90]|New Zealand|Merlucciidae|
Merluccius australis|[91]|New Zealand|Merlucciidae|
Merluccius australis|[92]|New Zealand|Merlucciidae|
Merluccius australis|[93]|New Zealand|Merlucciidae|
Merluccius australis|[94]|Argentina|Merlucciidae|
Merluccius australis|[95]|Chile|Merlucciidae|
Merluccius australis|[96]|Chile|Merlucciidae|
Merluccius australis|[97]|Chile|Merlucciidae|
Merluccius australis|[98]|Chile|Merlucciidae|
Merluccius australis|[99]|Chile|Merlucciidae|
Merluccius australis

Figure 9 Phylogenetic tree of *Merluccius* species

8.5 Specific problems regarding the identification of Thunnus species

Due to recent divergence or introgression events, there are limitations associated with distinguishing between *Thunnus* species with the COI barcoding gene (these limitations are common amongst many methods). Especially problematic is the differentiation between i) *Thunnus albacares* and *Thunnus obesus (Vinas & Tudela 2009, Hanner et al. 2011,* Pedrosa-Gerasmio *et al.* 2012, Santini *et al.* 2013*)* and ii) *Thunnus thynnus, Thunnus orientalis* and *Thunnus alalunga* (Dawnay et al. 2007).

Identification of either *Thunnus albacares* or *Thunnus obesus* with high confidence without using additional methods (using additional markers like cyt*b* or D-loop) is therefore difficult. The sequence similarity of the COI gene between these two species is high as they share some haplotypes and consequently constructing the phylogenetic tree in BOLD will not provide more detailed insight (Figure 10). Nevertheless, a BOLD search will still provide reliable identification to the genus level e.g. *Thunnus*, *Sarda* and *Katsuwonus*, which may still provide valuable information. Below we illustrate the difficulty to identify a sample of *Thunnus alalunga.*



Specimen Identification Request

Query: unlabeled_sequence

Top Hit: Chordata - Perciformes - Thunnus albacares (100%)

Search Result:

A species level match could not be made, the queried specimen is likely to be one of the following:

Thunnus albacares
Thunnus atlanticus
Thunnus maccoyii
Thunnus obesus

For a heirarchical placement - a neighbor-joining tree is provided: Tree Based Identification

Identification Summary:

Taxonomic Level	Taxon Assignment	Probability of Placement (%)
Phylum	Chordata	100
Class	Actinopterygii	100
Order	Perciformes	100
Family	Scombridae	100
Genus	Thunnus	100

Similarity Scores of Top 99 Matches:

Figure 10 Specimen Identification output on BOLD for a *Thunnus albacares* sample. Clearly the BOLD database cannot distinguish between 4 different *Thunnus* species as some sequences exist under different names, which show a level of similarity of 100%.

First, the enquiry will reveal the sample belongs to the *Thunnus* genus, which should alert the user on possible difficulties for correct species level identification. Subsequently, one could focus only on the 100% similarity score matches as these are most robust. Depending on the chosen reference database, published and private or only published records, the 100% similarity scores will contain only *Thunnus alalunga* or *Thunnus alalunga* and *Thunnus obesus* respectively. The phylogenetic tree may further help determine to which species the unknown sample belongs.

The phylogenetic tree obtained from a *Thunnus alalunga* sample is depicted in Figure 11. The unknown sample is highlighted in red. Although the tree is unresolved (the different species in the tree are not represented in a single branch) which makes interpretation of the results complicated, we like to explain some insights into the species identification.

Three clusters can be identified in the tree:

- Cluster 1 contains only *Thunnus alalunga* sequences
- Cluster 2 consist of *Thunnus alalunga* sequences, and 2 *Thunnus obesus* records (see red arrow)
- Cluster 3 is nested within Cluster 2, but contains no *Thunnus alalunga* sequences

The longer branch of cluster 3 indicate lower similarity with the other two clusters (see 8.1), reducing the chance of our unknown sequences to be identified as either *Thunnus orientalis* or *Thunnus thynnus*.

Cluster 2 contains our unknown sample, 2 records of *Thunnus obesus* both not flagged, and sequences of *Thunnus alalunga.* One of the *Thunnus obesus* records showed up in the 100% similarity match. Elaborating on this, one could identify the sample as *Thunnus alalunga* based on the knowledge that out of the 87 sequences (composing clusters 1 & 2) only 2 records were from to *Thunnus obesus,* representing 2.3% chance of incorrect species identification. Additionally, the BOLD database contains 264 species records of *Thunnus obesus*, of which 190 are with barcode. In case the sample would be a *Thunnus obesus* more BOLD records of this species would be expected to appear. Additionally, when using the Public Record Barcode Database, instead of the Species Level Barcode Records (default reference database option in BOLD) 100% similarity matches will in most cases only return as either one of the two species, here 100% similarity returned *Thunnus alalunga*.

In such circumstances a laboratory could identify the sample as i) *Thunnus* species (very robust), ii) *Thunnus alalunga* (although can *Thunnus obesus* cannot be excluded 100%), or iii) the laboratory could suggest they need to perform an additional test if a 100% confident species identification is required.

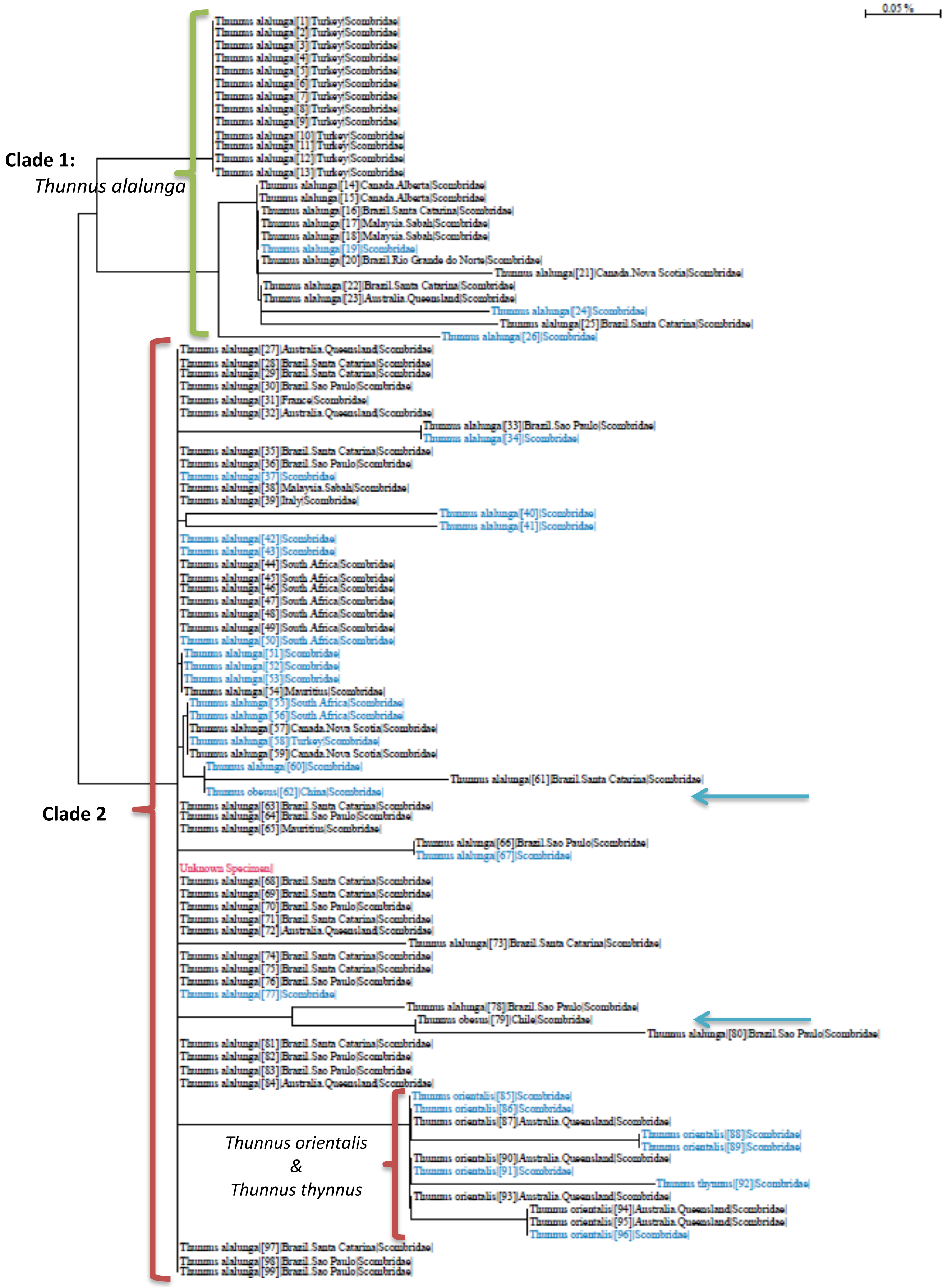

0.05 %
Clade 1:
Thunnus alalunga
Clade 2
Thunnus orientalis
&
Thunnus thynnus
Thunnus alalunga|[1]|Turkey|Scombridae|
Thunnus alalunga|[2]|Turkey|Scombridae|
Thunnus alalunga|[3]|Turkey|Scombridae|
Thunnus alalunga|[4]|Turkey|Scombridae|
Thunnus alalunga|[5]|Turkey|Scombridae|
Thunnus alalunga|[6]|Turkey|Scombridae|
Thunnus alalunga|[7]|Turkey|Scombridae|
Thunnus alalunga|[8]|Turkey|Scombridae|
Thunnus alalunga|[9]|Turkey|Scombridae|
Thunnus alalunga|[10]|Turkey|Scombridae|
Thunnus alalunga|[11]|Turkey|Scombridae|
Thunnus alalunga|[12]|Turkey|Scombridae|
Thunnus alalunga|[13]|Turkey|Scombridae|
Thunnus alalunga|[14]|Canada.Alberta|Scombridae|
Thunnus alalunga|[15]|Canada.Alberta|Scombridae|
Thunnus alalunga|[16]|Brazil.Santa Catarina|Scombridae|
Thunnus alalunga|[17]|Malaysia.Sabah|Scombridae|
Thunnus alalunga|[18]|Malaysia.Sabah|Scombridae|
Thunnus alalunga|[19]||Scombridae|
Thunnus alalunga|[20]|Brazil.Rio Grande do Norte|Scombridae|
Thunnus alalunga|[21]|Canada.Nova Scotia|Scombridae|
Thunnus alalunga|[22]|Brazil.Santa Catarina|Scombridae|
Thunnus alalunga|[23]|Australia.Queensland|Scombridae|
Thunnus alalunga|[24]||Scombridae|
Thunnus alalunga|[25]|Brazil.Santa Catarina|Scombridae|
Thunnus alalunga|[26]||Scombridae|
Thunnus alalunga|[27]|Australia.Queensland|Scombridae|
Thunnus alalunga|[28]|Brazil.Santa Catarina|Scombridae|
Thunnus alalunga|[29]|Brazil.Santa Catarina|Scombridae|
Thunnus alalunga|[30]|Brazil.Sao Paulo|Scombridae|
Thunnus alalunga|[31]|France|Scombridae|
Thunnus alalunga|[32]|Australia.Queensland|Scombridae|
Thunnus alalunga|[33]|Brazil.Sao Paulo|Scombridae|
Thunnus alalunga|[34]||Scombridae|
Thunnus alalunga|[35]|Brazil.Santa Catarina|Scombridae|
Thunnus alalunga|[36]|Brazil.Sao Paulo|Scombridae|
Thunnus alalunga|[37]||Scombridae|
Thunnus alalunga|[38]|Malaysia.Sabah|Scombridae|
Thunnus alalunga|[39]|Italy|Scombridae|
Thunnus alalunga|[40]||Scombridae|
Thunnus alalunga|[41]||Scombridae|
Thunnus alalunga|[42]||Scombridae|
Thunnus alalunga|[43]||Scombridae|
Thunnus alalunga|[44]|South Africa|Scombridae|
Thunnus alalunga|[45]|South Africa|Scombridae|
Thunnus alalunga|[46]|South Africa|Scombridae|
Thunnus alalunga|[47]|South Africa|Scombridae|
Thunnus alalunga|[48]|South Africa|Scombridae|
Thunnus alalunga|[49]|South Africa|Scombridae|
Thunnus alalunga|[50]|South Africa|Scombridae|
Thunnus alalunga|[51]||Scombridae|
Thunnus alalunga|[52]||Scombridae|
Thunnus alalunga|[53]||Scombridae|
Thunnus alalunga|[54]|Mauritius|Scombridae|
Thunnus alalunga|[55]|South Africa|Scombridae|
Thunnus alalunga|[56]|South Africa|Scombridae|
Thunnus alalunga|[57]|Canada.Nova Scotia|Scombridae|
Thunnus alalunga|[58]|Turkey|Scombridae|
Thunnus alalunga|[59]|Canada.Nova Scotia|Scombridae|
Thunnus alalunga|[60]||Scombridae|
Thunnus alalunga|[61]|Brazil.Santa Catarina|Scombridae|
Thunnus obesus|[62]|China|Scombridae|
Thunnus alalunga|[63]|Brazil.Santa Catarina|Scombridae|
Thunnus alalunga|[64]|Brazil.Sao Paulo|Scombridae|
Thunnus alalunga|[65]|Mauritius|Scombridae|
Thunnus alalunga|[66]|Brazil.Sao Paulo|Scombridae|
Thunnus alalunga|[67]||Scombridae|
Unknown Specimen||
Thunnus alalunga|[68]|Brazil.Santa Catarina|Scombridae|
Thunnus alalunga|[69]|Brazil.Santa Catarina|Scombridae|
Thunnus alalunga|[70]|Brazil.Sao Paulo|Scombridae|
Thunnus alalunga|[71]|Brazil.Santa Catarina|Scombridae|
Thunnus alalunga|[72]|Australia.Queensland|Scombridae|
Thunnus alalunga|[73]|Brazil.Santa Catarina|Scombridae|
Thunnus alalunga|[74]|Brazil.Santa Catarina|Scombridae|
Thunnus alalunga|[75]|Brazil.Santa Catarina|Scombridae|
Thunnus alalunga|[76]|Brazil.Sao Paulo|Scombridae|
Thunnus alalunga|[77]||Scombridae|
Thunnus alalunga|[78]|Brazil.Sao Paulo|Scombridae|
Thunnus obesus|[79]|Chile|Scombridae|
Thunnus alalunga|[80]|Brazil.Sao Paulo|Scombridae|
Thunnus alalunga|[81]|Brazil.Santa Catarina|Scombridae|
Thunnus alalunga|[82]|Brazil.Sao Paulo|Scombridae|
Thunnus alalunga|[83]|Brazil.Sao Paulo|Scombridae|
Thunnus alalunga|[84]|Australia.Queensland|Scombridae|
Thunnus orientalis|[85]||Scombridae|
Thunnus orientalis|[86]||Scombridae|
Thunnus orientalis|[87]|Australia.Queensland|Scombridae|
Thunnus orientalis|[88]||Scombridae|
Thunnus orientalis|[89]||Scombridae|
Thunnus orientalis|[90]|Australia.Queensland|Scombridae|
Thunnus orientalis|[91]||Scombridae|
Thunnus thynnus|[92]||Scombridae|
Thunnus orientalis|[93]|Australia.Queensland|Scombridae|
Thunnus orientalis|[94]|Australia.Queensland|Scombridae|
Thunnus orientalis|[95]|Australia.Queensland|Scombridae|
Thunnus orientalis|[96]||Scombridae|
Thunnus alalunga|[97]|Brazil.Santa Catarina|Scombridae|
Thunnus alalunga|[98]|Brazil.Sao Paulo|Scombridae|
Thunnus alalunga|[99]|Brazil.Sao Paulo|Scombridae|

Figure 11 Phylogenetic tree constructed in BOLD for a *Thunnus alalunga* sample.

8.6 References

Avise 2000 Phylogeography. The history and formation of species. Harvard University Press, Cambridge, MA.

Dawnay *et al*. 2007 Validation of the barcoding gene COI for use in forensic genetic species identification. *Forensic Science International*, **173**: 1-6.

Hanner *et al*. 2011 FISH-BOL and seafood identification: Geographically dispersed case studies reveal systemic market substitution across Canada. *Mitochondrial DNA*, **22**:106-122.

Ivanova *et al*. 2007 Universal primer cocktails for fish DNA barcoding. *Molecular Ecology Notes*, **7**: 544–548.

Pedrosa-Gerasmio et al. 2012 Discrimination of Juvenile Yellowfin (*Thunnus albacares*) and Bigeye (*T. obesus*) Tunas using Mitochondrial DNA Control Region and Liver Morphology. *PlosOne*, 7:e35604.

Santini *et al*. 2013 First molecular scombrid timetree (Percomorpha: Scombridae) shows recent radiation of tunas following invasion of pelagic habitat. *Italian Journal of Zoology*, **80**: 210-221.

Vinas & Tudela 2009 A validated methodology for genetic identification of tuna species (Genus *Thunnus*). PlosOne, 4: e7606.

Ward *et al*. 2005 DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B*, **360**: 1847–57.

Additional Resources:

General information and a solid background to DNA barcoding are available below, including access to the barcode of life online community (including a forum that can potentially provide troubleshooting advice);

http://www.barcodeoflife.org/

A comprehensive hand book for utilising the BOLD database is available;

http://www.boldsystems.org/index.php/Resources