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

Prepared by the Labelfish Consortium



December 2014

- 1. BACKGROUND
- 2. PURPOSE
- 3. SCOPE
- 4. DEFINITIONS & ABBREVIATIONS
- 5. PRINCIPLE OF THE METHOD
- 6. MATERIALS & EQUIPMENT
 - 6.1 Water
 - 6.2 Solutions, standards and reference materials
 - 6.3 Commercial kits
 - 6.4 Plastic-ware
 - 6.5 Equipment
 - 6.6 Other materials
 - 6.7 Electronic files / computer software

7. PROCEDURES

7.1 Sample preparation
7.2 DNA Extraction
7.3 PCR Amplification
7.4 PCR Product Check
7.5 DNA Sequencing
7.6 Raw Data Processing
7.7 Generating a consensus sequence
7.8 Identifying the species on the Barcode of Life Database
7.9 Issues with Interpreting the Species Identification
7.10 Quality Assurance

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 (<u>http://www.barcodeoflife.org/</u>);

"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 useItemDetailExample SupplierProduct codePipette tips (filtered)10, 20, 200 & 1000µlStarlabsS1120

PCR tubessingle, strip or 96-wellStarlabs1.5ml tubes1.5 mlStarlabs

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.

11402

S1615

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 μI 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 & MgCl₂)

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 <i>et al</i> . 2005
FishF2_t1	TGTAAAACGACGGCCAGTCGACTAATCATAAAGATATCGGCAC	Ward <i>et al</i> . 2005
FishR2_t1	CAGGAAACAGCTATGACACTTCAGGGTGACCGAAGAATCAGAA	Ward <i>et al</i> . 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 MgCl₂ (50 mM)
- 0.2 µL of each primer (0.01 mM)
- 0.4 μ L of the Bioline 10mM dNTP mix
- 0.1 µL of BIOTAQ *Taq* DNA Polymerase

6. Vortex master mix thoroughly.

7. Place 17 μ L 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
Таq	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 μ L 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 A nalysis.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: <u>http://www.mbio.ncsu.edu/bioedit/bioedit.html</u>

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;



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;



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.*



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).



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**.



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):



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&se ction=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.

. ()	BioEdi	t Sequence Alignment Editor - [C:\MyDocs\Source\2014_02_15_1CYTBHeavy\C	BH.fas]	
I. (🎾 File	Edit Sequence Alignment View Accessory Application RNA World	Wide Web O	ptions Window Help _ & ×
		New Alignment	Ctrl+N	
		Open	Ctrl+0	
		New from Clipboard		Charle
	MOC	New Text		ruler at: 1
	f	Open As Text		vII III Speed slow → a fast
		Save	Ctrl+S	
	СВ	Save As		CTTTATACAGGTATGAGCCGTAGTACAGGCCTCGGCCAAT
		Retrieve sequences from GenBank or GenPept		1 111
		Copy file name to clipboard		
		Impat		Sequence alignment file
		Import Import from Clinhoard		From tab delimited file (eg Eycel)
		Merge into Alignment based on a Reference Sequence		From tab delimited file (eg excel)
		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		
				-

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.

BIOEdit Ser La	ne Highment Editor - [C:\MyDocs\Source\2014_02_15_1CYTBHeavy\CBH.fas]	
File Edit S	sequence Alignment View Accessory Application RNA World Wide Web Options Window Help	
	Edit Sequence	
🔒 📇 Couri	Edit all selected	
Mode: Select / S	Select Positions	- 1
0 T D T	Open at cursor position	
E. I D T	Extract Positions	y 🖣 fast
	Go to Pubmed references	70 80 90 100 110
CB Forward CB Reverse	Overwrite/Retrieve sequences by gi number in title by HTTP to GenBank	SCCGTAGTACAGGCCTCGGCCAATGTGAAGATAAATACAAATAAGAAGAA CAGTTTGATGAAATTTCGGTTCGCTCTTAGGCCTATGCTTGATTTCCCAAA
_	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	Nucleotide Composition
		 Base composition and mass export (monoisotopic)
	Translate or Reverse-Translate (permanent)	Base composition and mass export with average masses
	Translate in selected frame (permanent)	Complement
	Toggle Translation Ctrl+(Reverse Complement Shift+Ctrl+R
	iii	RNA->DNA
	111	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
		Restriction Map Sorted Six-Frame Translation

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).



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).

*	Bi	lio	Edit	Seq	uenc	e Alig	gnm	ent E	ditor	- [Ur	ntitled	1]																		X	
>	ø	Fi	ile	Edit	Se	equen	ce	Alig	nmer	nt \	/iew	Acce	ssory A	Applic	cation	RNA	Wo	rld Wio	le Wel	b 0	ptions	Win	dow	Help)				-	8×	:
6	3	3	Ľ																												
E	3	ł	≞ [Cou	rier N	lew		•	11	•	в		21	otal se	equenci	es															
Мо	de	e:	Sele	- st 7 9	Slide	•	_		Se Po	electio	n: null						Seq Num	uence l berina l	Mask: Mask:	None None					Star rule	t rat 1					
	ſ	1	II)]	Г	D 🔒	h G	/D +	-			Ŧ				641 8	AT !!	15		6	ш	Ħ		Scroll [, fast]			
	-				-	i	, 		1			· · · I		1.1.1	· · · · · · · · · · · · · · · · · · ·			 ' ''	••••		1	<u>.</u>		,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	• •		70		••;		
CE	3	F	orv	ar	d										GG	CTGI	CAT	TATT	ACT	AGA	GGA	GAAG	GAT	TACC	CCG	ATG	TTTC	ATGI	TTC	TTTA	r
CE	3_	_R	leve	ers	e	TG	rcc	STCA	IGGG	5CAI	ACAC	ATA	SCCT	ACG	AAGG	CTGI	CAT	TATT	ACTA	AGA	AGGA	GAAG	JAT	TACC	CCG.	ATG	TTT	ATGI		TTTA.	ľ
																															1
																	C	`~·	\sim	-	<u>_</u>	<u></u>	nt	- ~ .	\sim	C	~~		n	~~	
																	C	.01	11	JI	211	Ie		.dl	y	3	eq	ue	:11	LE	

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.



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 TTCGAGCAGAGCTAAGTCAACCTGGTGCACTTCTTGGTGATGATCAAATTTATAATGTGA TCGTTACAGCGCACGCTTTCGTAATAATTTTCTTTATAGTAATACCACTAATAATTGGAGG CTTTGGGAACTGACTCATTCCTCTAATGATCGGTGCACCAGATATAGCTTTCCCTCGAAT AAATAACATAAGCTTCTGACTTCTTCCTCCATCTTTCCTGCTCCTTTTAGCATCCTCTGGT GTAGAAGCTGGGGCTGGAACAGGCTGAACTGTCTATCCACCTTTAGCCGGAAACCTCG CTCATGCTGGGGCATCTGTTGATCTCACTATTTTTTCTCTTCATCTAGCAGGGATTTCAT CAATTCTTGGGGCAATTAATTTTATTACCACAATTATTAATATGAAACCTCCGGCAATTTC ACAGTACCAAACACCCCTATTTGTTTGAGCAGTACTAATTACAGCTGTGCTTCTACTATT ATCTCTCCCCGTCTTAGCAGCTGGATGCAATACTTCTAACTGACCGTAATCTTAATAT TCTTTTGACCCTGCTGGAGGAGGTGATCCCATTTTATACCAACA

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 (<u>http://www.boldsystems.org/</u>) and select the "Identification" tab at the top of the webpage.



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.

	Firefox -	• ×
	BOLD Systems: Management & Analysis +	
	🔄 🔶 🕜 www.boldsystems.org/index.php/IDS_OpenIdEngine	• 🖬 •
	User Put BOLDSYSTEMS Databases Taxonomy Identification Workbench Resources	lic 🗘 📤
	Identification Request	rint
i.	Animal Identification Fungal Identification [ICOI] FIND FIND FIND FIND FIND FIND FIND FIND	
ii.	 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: <u>lul-2013</u> <u>lul-2012</u> <u>lul-2011</u> <u>lul-2010</u> <u>lul-2009</u> 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 markers will be non-movide a probability of placement to a taxon. Species Level Barcode Records (1,73) 732 Sequences/146,084 Species/58,357 Interim Species) Every COI barcode record with the minimum sequence length of 500bp. This includes many species represented by only one or two specimens as well as all species/14,985 Interim Species) Public Record Barcode Database (555,693 Sequences/62,894 Species/14,985 Interim Species) Public Record Barcode Database (1,318,574 Sequences/132,900 Species/50,814 Interim Species) Ful 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 500bp. This library is a collection of records from the published projects subset of the Species library with animum sequence length of 500bp. This library is a collection of records from the published projects section of BOLD. Ful 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 contatining bott public	 ∯ feedback
ii.	Entransport >> CadusMorhuaSCFAC839-06 CITTIATCCGTATTIGGTCGCCGCACCAGCCGCAACAGCCCCTAAGCCCTCCATTCCGAGCAGCGCAAG TCAACCTCGTCGCACTTCTTGCGCACTCAACATTATATATGTCGATCCGTCACCGCCCCCCTCCTCGTACTCGCACCAGCTTA TCAACCTCGCTCGCACTCTCTGCGACCAGCTTTGCGCACCAGCCTTTCCTGCATCAGCCCCTCCCCCCCC	н.

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_Identi	ificationRequest					~ C'	8 - G	adus ch	alcogramm	us 🔎	•	Â
OLD SYSTE	MS Database	es Taxonomy	Identifi	cation	Worł	denci	n 1	Resou	urces			User F	Public
pecimen Ider	ntification Req	luest											Print
Query: GadusMorh	uaSCFAC839-06						Top H	it: Cho	rdata -	Gadiforme	es - Gadu	s morhu	a (1009
Search Result:													
							6.1	C 11					
A species level	match could not i	be made, the quer	ied specim	ien is likel	y to b	e one	ofthe	e follo	owing	;:			
Gadus morhua													
Gadus chalcogramm	ius												
For a heirarch	ical placement - a	neighbor-joining	tree is pro	vided: Tr	ee Base	d Identi	fication						
	-		-										
					_								
Identification Su	mmary:			Similarity	Scores	of Top	99 Ma	tches:					
Identification Su	mmary:		_	Similarity	Scores	of Top	99 Ma	tches:					
Identification Su	mmary:	Probability of		Similarity : 100.0 99.8	Scores	of Top	99 Ma	tches:					
Identification Su Taxonomic Leve	mmary: I Taxon Assignmen	Probability of t Placement (%)		Similarity : 99.8 2 99.6	Scores	of Top	99 Ma	tches:					
Identification Su Taxonomic Leve Phylum	mmary: I Taxon Assignmen Chordata	Probability of t Placement (%) 100	0	Similarity 3	Scores	of Top	99 Ma	tches:					
Taxonomic Leve Phylum Class	mmary: I Taxon Assignmen Chordata Actinopterygii	Probability of Placement (%) 100 100		Similarity 3 100.0 99.8 99.6 99.6 Ajjue 99.2 99.2 99.0	Scores	of Top	99 Ma	tches:					
Identification Su Taxonomic Leve Phylum Class Order	mmary: I Taxon Assignmen Chordata Actinopterygii Gadiformes	Probability of Placement (%) 100 100		Similarity 3 100.0 99.8 99.6 99.4 99.4 99.2 99.2 99.0 99.0 98.8	Scores	of Top	99 Ma	tches:					
Taxonomic Leve Phylum Class Order Family	mmary: I Taxon Assignmen Chordata Actinopterygii Gadiformes Gadidae	t Probability of Placement (%) 100 100 100		Similarity : 100.0 99.8 (%) 99.6 (%) 99.4 99.2 99.0 99.0 98.8 98.6	Scores	of Top	99 Ma	tches:					
Taxonomic Leve Phylum Class Order Family Cenus	I Taxon Assignmen Chordata Actinopterygii Gadiformes Gadidae	t Probability of Placement (%) 100 100 100 100		Similarity (99.8 99.6 99.0 99.0 99.0 99.0 99.0 99.0 99.0	Scores	of Top	99 Ma	tches:	56	67 78	89		
Taxonomic Leve Phylum Class Order Family Genus	mmary: I Taxon Assignmen Chordata Actinopterygii Gadiformes Gadidae Gadus	robability of Placement (%) 100 100 100 100 100 100		Similarity : 100.0 99.8 (%) 99.6 11 99.6 99.4 11 99.6 98.6 98.6 98.4 1	Scores	of Top	99 Ma 34	tches: 45	56	67 78 Ra	s 89 nked Mate	ches	
Taxonomic Leve Phylum Class Order Family Genus	mmary: I Taxon Assignmen Chordata Actinopterygii Gadiformes Gadidae Gadus	t Probability of Placement (%) 100 100 100 100 100		Similarity : 90.0 99.6 99.6 99.6 99.0 99.0 99.0 99.0 99.0 99.0 99.0 99.0 99.4 98.6 98.4 1	Scores 12	of Top	99 Ma 34	tches: 45	56	67 78 Ra	8 89 nked Mate	ches	
Taxonomic Leve Phylum Class Order Family Genus	mmary: I Taxon Assignmen Chordata Actinopterygi Gadiformes Gadidae Gadus	t Probability of Placement (%) 100 100 100 100 100		Similarity - 99.8 (%) 99.4 (%) Apj. 99.0 99.0 99.0 99.0 99.0 99.0 99.4 98.8 98.6 98.4 1	Scores	of Top 23	99 Ma 34	tches: 45	56	67 78 Ra Displa	89 nked Mate	ches	
Taxonomic Leve Phylum Class Order Family Genus	mmary: I Taxon Assignmen Chordata Actinopterygii Gadiformes Gadidae Gadus	t Probability of Placement (%) 100 100 100 100 100		Similarity 3 100.0 99.8 99.4 99.4 99.4 99.4 99.4 99.4 99.4	Scores	of Top 23	99 Ma 34	45	56	67 78 Ra Displa	i 89 nked Mate	ches	0 🗸
Identification Su Taxonomic Leve Phylum Class Order Family Cenus TOP 20 Matches :	mmary: I Taxon Assignmen Chordata Actinopterygii Gadiformes Gadidae Gadus	rt Probability of Placement (%) 100 100 100 100 100 100	Family	Similarity 1 100.0 99.6 99.6 99.6 98.8 98.6 98.4 1	12	of Top 23	99 Ma 34	45	56	67 78 Ra Displa Similarity	r 89 nked Mate	ches Top 2	0 🗸
Identification Su Taxonomic Leve Phylum Class Order Family Cenus TOP 20 Matches : Phylum	mmary: I Taxon Assignmen Chordata Actinopterygii Gadiformes Gadidae Gadus Class	rt Probability of Placement (%) 100 100 100 100 100 100	Family	Similarity 1 (%) 99.6 99.6 99.4 99.0 99.8 98.6 98.4 1 Source Control	12 12	of Top 23	34 Speci	45 es	56	67 78 Ra Displa Similarity	t 89 nked Mate y option: (%)	ches Top 24 Statu	0 •
Identification Su Taxonomic Leve Phylum Class Order Family Genus TOP 20 Matches : Phylum Chordata	mmary: Taxon Assignmen Chordata Actinopterygii Gadiformes Gadidae Gadus Class	t Probability of Placement (%) 100 100 100 100 100 200 200 200 200 200	Family Gadidae	Similarity 1 100.0 99.8 99.6 99.4 99.	12 12 15	of Top	34 Speci morh	45 es	56	67 78 Ra Displa Similarity 100	t 89 nked Mate y option: (%)	ches Top 21 Statu Early-Rel	0 ▼ s lease
Identification Su Taxonomic Leve Phylum Class Order Family Genus TOP 20 Matches : Phylum Chordata Chordata	Taxon Assignmen Chordata Actinopterygii Gadiformes Gadidae Gadus Class Actinopterygii Class	t Probability of Placement (%) 100 100 100 100 100 200 500 500 500 500 500 500 500 500 5	Family Cadidae Cadidae	Similarity 1 100.0 99.8 99.6 99.4 99.4 99.4 99.4 99.4 99.4 99.4 99.4 99.4 99.4 99.4 99.4 99.4 99.4 99.4 1 90.6 99.4 99.6 99.4 1 90.6 99.8 99.6 99.4 1 90.6 99.8 99.6 90.4 90.6	12 12 15 15	of Top	99 Ma 34 Speci morh morh	45 es ua ua	56	67 78 Ra Displa Similarity 100 100	t 89 nked Mate y option: (%)	ches Top 21 Statu Early-Rel Privat	0 ▼ s lease se
Taxonomic Leve Phylum Class Order Family Cenus TOP 20 Matches : Phylum Chordata Chordata	Taxon Assignmen Chordata Actinopterygii Gadiformes Gadidae Gadus Class Actinopterygii Actinopterygii Catinopterygii	t Probability of Placement (%) 100 100 100 100 100 100 100 100 Cadiformes 0 Cadiformes 0	Family Gadidae Gadidae Gadidae	Similarity 1 99.8 99.6 99.4 99.4 99.4 99.4 1 99.4 1 99.4 1 90.0 98.8 98.6 98.4 1 0 0 0 0 0 0 0 0 0 0 0 0 0	12 12 15 15 15	of Top	34 Speci morh morh morh	45 es ua ua ua	56	67 78 Ra Displa Similarity 100 100	: 89 nked Mate y option: (%)	ches Top 24 Statu Early-Rel Privat Privat	0 ▼ s lease re
Identification Su Taxonomic Leve Phylum Class Order Family Genus TOP 20 Matches : Phylum Chordata Chordata Chordata Chordata	Taxon Assignmen Chordata Actinopterygi Gadiformes Gadidae Gadus Cadus Actinopterygi Actinopterygi Actinopterygi	t Probability of Placement (%) 100 100 100 100 100 100 100 Cadiformes Cadiforme Cadifo	Family Gadidae Gadidae Gadidae Gadidae	Similarity 1 (%) 99.6 99.6 99.4 99.0 99.8 98.6 98.4 1 98.8 98.6 98.4 1 98.8 98.6 98.4 1 98.8 98.6 98.4 1 98.8 98.6 98.4 1 98.8 98.6 98.8 98.8 98.6 98.8 98.6 98.8 98.6 98.8 98.6 98.8 98.6 98.8 98.6 98.8 98.6 98.8 98.6 98.8 98.8 98.8 98.8 98.6 98.8 98.	12 12 15 15 15 15	of Top	99 Ma 34 Specie morh morh morh morh	45 es ua ua ua ua	56	67 78 Ra Displa Similarity 100 100 100	i 89 nked Mate y option: (%)	Top 2 Statu Early-Rel Privat Early-Rel	0 ▼ s lease re re lease
Identification Su Taxonomic Leve Phylum Class Order Family Genus TOP 20 Matches : Phylum Chordata Chordata Chordata Chordata	Taxon Assignmen Chordata Actinopterygii Gadiformes Gadidae Gadus Cadus Actinopterygii Actinopterygii Actinopterygii Cachopterygi	t Probability of Placement (%) 100 100 100 100 100 Conternation of the second s	Family Gadidae Gadidae Gadidae Gadidae Gadidae	Similarity 1 100.0 99.8 99.6 99.4 99.4 99.2 99.4 99.	12 12 15 15 15 15 15 15	of Top	34 Speci morh morh morh	45 es ua ua ua ua ua ua	58	67 78 Ra Displa Similarity 100 100 100 100 100	a 89 nked Mate y option: (%)	ches Top 20 Statu Early-Rel Privat Early-Rel Early-Rel Early-Rel	0 ▼ s lease lease lease lease
Identification Su Taxonomic Leve Phylum Class Order Family Genus TOP 20 Matches : Phylum Chordata Chordata Chordata Chordata Chordata	Class Actinopterygii Ccass Actinopterygii Ccass Actinopterygii Ccass Actinopterygii Ccass Actinopterygii Ccacs Actinopterygii Ccacs Actinopterygii Ccacs Cca	t Probability of Placement (%) 100 100 100 100 100 100 100 100 100 10	Family Gadidae Gadidae Gadidae Gadidae Gadidae Gadidae	Similarity 1 100.0 99.8 99.6 99.4 99.2 99.2 99.4 99.	12 12 15 15 15 15 15 15 15 15	of Top 23	99 Ma 34 Speci morh morh morh morh morh	45 es ua ua ua ua ua ua ua	56	67 78 Displa Similarity 100 100 100 100 100 100	e 89 nked Mate y option: (%)	ches Top 24 Statu Privat Early-Rel Privat Early-Rel Early-Rel Privat	0 ▼ Is lease lease lease lease lease
Taxonomic Leve Phylum Class Order Family Cenus TOP 20 Matches : Phylum Chordata Chordata Chordata Chordata Chordata Chordata	Taxon Assignmen Chordata Actinopterygij Gadiformes Gadidae Gadus Class Actinopterygij Actinopterygij Actinopterygij Catinopter	t Probability of Placement (%) 100 100 100 100 100 100 100 100 100 10	Family Cadidae Cadidae Cadidae Cadidae Cadidae Cadidae Cadidae Cadidae	Similarity 1 99.0 99.6 99.6 99.4 99.0 99.0 99.0 99.0 99.0 99.0 99.0 99.4 99.0 99.4 99.0 99.4 1 99.0 99.4 99.0 99.4 99.0 99.4 99.0 99.4 99.0 99.4 99.0 99.0 99.4 99.0 90.0 9	12 12 15 15 15 15 15 15 15 15 15	of Top 23	99 Ma 34 Speci morh morh morh morh morh morh morh	45 es ua ua ua ua ua ua ua ua ua ua ua	56	67 78 Ra Displa Similarity 100 100 100 100 100 100 100 100	t 89 nked Mate y option: (%)	Top 21 Statu Privat Privat Early-Rel Early-Rel Early-Rel Privat Privat	0 ▼ Is lease lease lease lease lease lease lease lease lease lease lease
Identification Su Taxonomic Leve Phylum Class Order Family Genus TOP 20 Matches : Phylum Chordata Chordata Chordata Chordata Chordata Chordata Chordata	Taxon Assignmen Chordata Actinopterygii Gadiformes Gadidae Gadus Cadus Actinopterygii Actinopterygii Cactinopt	t Probability of Placement (%) 100 100 100 100 100 100 100 100 Cadiformes Cadiforme C	Family Gadidae Gadidae Gadidae Gadidae Gadidae Gadidae Gadidae Gadidae	Similarity 1 (%) 99.6 99.6 99.4 99.0 99.8 98.6 98.4 1 98.8 98.6 98.4 1 99.0 99.8 99.6 98.4 1 99.0 99.8 99.6 99.4 99.6 99.4 99.6 99.4 99.6 99.4 99.6 99.4 99.6 98.6 98.	12 12 15 15 15 15 15 15 15 15	of Top 23	99 Ma 34 Speci morh morh morh morh morh morh morh	45 es ua ua ua ua ua ua ua ua ua ua ua ua ua	56	67 78 Ra Displa Similarity 100 100 100 100 100 100 100 100 100	i 89 nked Mate y option: (%)	ches Top 2 Statu Privat Early-Rel Privat Privat Privat	0 ↓ s lease re lease lease lease iease iease iease iease iease iease iease
Taxonomic Leve Phylum Class Order Family Genus TOP 20 Matches : Phylum Chordata Chordata Chordata Chordata Chordata Chordata Chordata Chordata Chordata	Taxon Assignmen Chordata Actinopterygii Gadiformes Gadidae Gadidae Gadus Cadus Cadus Actinopterygii Cactinopte	t Probability of Placement (%) 100 100 100 100 100 Corder Cadiformes Cadiforme	Family Gadidae Gadidae Gadidae Gadidae Gadidae Gadidae Gadidae Gadidae Gadidae	Similarity 1 100.0 99.8 99.6 99.4 99.2 99.4 99.2 99.4 99.2 99.4 99.2 99.4 99.2 99.4 99.	12 12 15 15 15 15 15 15 15 15 15 15 15	of Top 23	99 Ma 34 Speci morh morh morh morh morh morh	45 45 ua ua ua ua ua ua ua ua ua ua ua ua ua	56	67 78 Ra Displa Similarity 100 100 100 100 100 100 100 100 100 10	a 89 nked Mate y option: (%)	ches Top 21 Statu Early-Rel Privat Early-Rel Privat Privat Privat Privat	0 ↓ Is lease re lease re re re re re re re re re
Identification Su Taxonomic Leve Phylum Class Order Family Genus TOP 20 Matches : Phylum Chordata Chordata Chordata Chordata Chordata Chordata Chordata Chordata Chordata Chordata Chordata Chordata Chordata		t Probability of Placement (%) 100 100 100 100 100 100 100 100 100 10	Family Gadidae Gadidae Gadidae Gadidae Gadidae Gadidae Gadidae Gadidae Gadidae	Similarity 1 100.0 99.8 99.6 99.4 99.4 99.2 99.4 99.2 99.4 99.	12 12 13 15 15 15 15 15 15 15 15 15 15 15	23	99 Ma 34 34 Speci morh morh morh morh morh morh morh morh	45 45 ua ua ua ua ua ua ua ua ua ua ua	56	67 78 Ra Displa Similarity 100 100 100 100 100 100 100 100 100 10	s 89 nked Mate y option: (%)	Top 21 Statu Privat Early-Rel Privat Early-Rel Privat Privat Privat	0

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.

	.org/index.php/105_identii	ficationRequest					1	⊽ C"	8 - 0	iadus cl	nalcogr	ammu	s >	' +	Â
LD SYSTE	MS Database	s Taxonomy	Ident	tificati	ion	Work	benc	h	Reso	urces				Use	r Publi
ecimen Idei	ntification Req	uest													🛎 Pri
Query: GadusMort	uuaSCFAC839-06							Top I	lit: Ch	ordata	Gadif	ormes	-Gadı	ıs morl	111a (10
Soarch Docult															
Search Result:															
A species level	match could not b	e made, the que	ried spec	imen	is likel	y to b	e one	ofth	e foll	owin	g:				
Cadus markus															
Gadus mornua Gadus chalcogramm	nus														
dadas charcogramm															
For a hoirarch	ical placement au	noighbor ioining	troo ic n	rouid	od: T	P	تحديد الدارية								
For a neirarch	icai piacement - a i	neignbor-joining	tree is p	rovia		ee Base	d Ident	ification	1						
dentification Su															
	mmary:			Sin	nilarity	Scores	of Top	99 M	atches	:					
	mmary:			Sin	nilarity	Scores	of Top	99 Ma	atches	:					
	mmary:	Probability of		Sin	100.0	Scores	of Top	99 M	atches	:					
Taxonomic Leve	mmary: I Taxon Assignment	Probability of Placement (%)		Sin	nilarity 100.0 99.8 99.6	Scores	of Top	99 M	atches	:					
Taxonomic Leve Phylum	I Taxon Assignment Chordata	Probability of t Placement (%) 100		Sin (%) A	100.0 99.8 99.6 99.4	Scores	of Top	99 M	atches	:			•		
Taxonomic Leve Phylum Class	I Taxon Assignment Chordata Actinoptervoji	Probability of Placement (%) 100		Sin (%)	100.0 99.8 99.6 99.4 99.2	Scores	of Top	99 M	atches	:			•		
Taxonomic Leve Phylum Class Order	I Taxon Assignment Chordata Actinopterygii Cadiformes	Probability of Placement (%) 100 100		Similarity (%)	100.0 99.8 99.6 99.4 99.2 99.0	Scores	of Top	99 M	atches	:					
Taxonomic Leve Phylum Class Order	Taxon Assignment Chordata Actinopterygii Gadiformes	Probability of Placement (%) 100 100 100		Similarity (%)	100.0 99.8 99.6 99.4 99.2 99.0 98.8 98.6	Scores	of Top	99 M	atches	:					
Taxonomic Leve Phylum Class Order Family	I Taxon Assignment Chordata Actinopterygii Gadiformes Gadidae	Probability of Placement (%) 100 100 100	1	Similarity (%)	100.0 99.8 99.6 99.4 99.2 99.0 98.8 98.6 98.4	Scores	of Top	99 M	atches	:					
Taxonomic Leve Phylum Class Order Family Genus	I Taxon Assignment Chordata Actinopterygii Gadiformes Gadidae Gadus	robability of Placement (%) 100 100 100 100 100	1	Similarity (%)	100.0 99.8 99.6 99.4 99.2 99.0 98.8 98.6 98.4 1	Scores 12	of Top 23	99 Ma 34	atches 45	56	67	78 Ran	89 ked Ma	tches	
Taxonomic Leve Phylum Class Order Family Genus	Taxon Assignment Chordata Actinopterygii Gadiformes Gadidae Gadus	Probability of Placement (%) 100 100 100 100 100		Similarity (%)	100.0 99.8 99.6 99.4 99.2 99.0 98.8 98.6 98.8 98.6 98.4 1	Scores	of Top	99 Ma 34	atches 45	56	67	78 Ran	89 ked Ma	tches	
Taxonomic Leve Phylum Class Order Family Genus	I Taxon Assignment Chordata Actinopterygii Gadiformes Gadidae Gadus	Probability of Placement (%) 100 100 100 100 100		Similarity (%)	nilarity 100.0 99.8 99.6 99.4 99.0 99.0 98.8 98.6 98.6 98.4 1	Scores	of Top	99 Ma 34	atches 45	56	67	78 Ran	89 ked Ma	tches	
Taxonomic Leve Phylum Class Order Family Genus	I Taxon Assignment Chordata Actinopterygii Gadiformes Gadidae Gadus	Probability of Placement (%) 100 100 100 100 100		Similarity (%)	nilarity 100.0 99.8 99.6 99.4 99.0 99.0 98.8 98.6 98.8 98.6 98.4 1	Scores	of Tor	99 Ma 34	45	56	67 D	78 Ran tisplay	89 ked Ma	tches : Top	20 🗸
Taxonomic Leve Phylum Class Order Family Genus TOP 20 Matches :	I Taxon Assignment Chordata Actinopterygii Gadiformes Gadidae Gadus	Probability of Placement (%) 100 100 100 100 100	Emily	Similarity (%)	nilarity 100.0 99.8 99.6 99.4 99.2 99.0 98.8 98.6 98.4 1	Scores 12	of Tor	9 99 Ma 34	45	56	67 D	78 Ran isplay	89 ked Ma optior	tches : Top	20 👻
Taxonomic Leve Phylum Class Order Family Genus COP 20 Matches : Phylum	I Taxon Assignment Chordata Actinopterygii Gadiformes Gadidae Gadus Class	Probability of Placement (%) 100 100 100 100 100 100	Family	Similarity (%)	100.0 99.8 99.6 99.4 99.2 99.0 98.8 98.6 98.4 1	Scores 12 15	of Top	999 Ma 34 Spec	45 ies	56	67 D Simil	78 Ran isplay arity (9	89 ked Ma optior	tches : Top Sta	20 🗸
Taxonomic Leve Phylum Class Order Family Cenus COP 20 Matches : Phylum Chordata	Taxon Assignment Chordata Actinopterygii Gadiformes Gadidae Gadus Class Actinopterygii G	Probability of Placement (%) 100 100 100 100 100 000 100	Family Gadidae	Similarity (%)	100.0 99.8 99.6 99.4 99.2 99.0 98.8 98.6 98.4 1 0 Gen Gad	I2 12 15	of Top	34 Spec mort	45 ies hua	56	67 D Simil	78 Rani isplay arity (9	89 ked Ma optior	tches : Top Sta Early-I	20 ▼ tus Release
Taxonomic Leve Phylum Class Order Family Genus TOP 20 Matches : Phylum Chordata Chordata	Taxon Assignment Chordata Actinopterygi Gadiformes Gadidae Gadus Class Actinopterygi G	Probability of Placement (%) 100 100 100 100 100 100 100	Family Gadidae Gadidae	Similarity (%)	nilarity 99.8 99.6 99.4 99.2 99.0 98.8 98.6 98.8 98.6 98.4 1 Gen Gad Gad	Scores 12 15 15	of Top 23	34 Spec mori mori	45 ies iua iua	56	67 D Simil	78 Rani iisplay 100 100	89 ked Ma optior	tches : Top Sta Early-I Priv	20 ▼ tus Release vate
Taxonomic Level Phylum Class Order Family Genus TOP 20 Matches : Phylum Chordata Chordata Chordata	Taxon Assignment Chordata Actinopterygii Gadiformes Gadidae Gadus Class Actinopterygii G Actinopterygii G Actinopterygii G	Probability of Placement (%) 100 100 100 100 100 100 100 100 100 10	Family Gadidae Gadidae Gadidae	Similarity (%)	nilarity 99.8 99.6 99.4 99.2 99.0 98.8 98.6 98.4 1 1 Gen Gad Gad	Scores 12 15 15 15	of Top 23	34 Spec mori mori mori	45 ies iua iua iua	56	67 D Simil	78 Ran iisplay arity (9 100 100	89 ked Ma optior %)	tches : Top Sta Early-I Prin Prin	20 ↓ tus Release rate rate
Taxonomic Leve Phylum Class Order Family Genus Colordata Chordata Chordata Chordata	Taxon Assignment Chordata Actinopterygi Gadiformes Gadus Class Actinopterygi Actinopterygi Gatinopterygi	e Probability of Placement (%) 100 100 100 100 100 100 adiformes adiformes adiformes adiformes	Family Gadidae Gadidae Gadidae Gadidae	Similarity (%)	nilarity 99.6 99.4 99.6 99.4 99.0 98.8 98.6 98.4 1 Gen Gad Gad Gad	Scores 12 15 15 15	of Top 23	34 Spec mort mort mort mort	45 ies hua hua hua hua	56	67 D Simil	78 Ran iisplay arity (\$ 100 100 100	89 kked Ma optior	tches : Top Sta Early-I Prin Prin Early-I	20 ▼ tus kelease rate kelease
Taxonomic Leve Phylum Class Order Family Cenus COP 20 Matches : Phylum Chordata Chordata Chordata Chordata	Taxon Assignment Chordata Actinopterygii Gadiformes Gadidae Gadus Class Actinopterygii G Actinopterygii G Actinopterygii G Actinopterygii G	Probability of Placement (%) 100 100 100 100 100 100 100 100 100 10	Family Gadidae Gadidae Gadidae Gadidae	Similarity (%)	nilarity 100.0 99.8 99.6 99.4 99.2 99.0 98.8 98.6 98.4 1 Gen Gad Gad Gad Gad	Scores 12 13 15 15 15 15	of Top 23	34 Spec mori mori mori mori	45 ies ihua ihua ihua ihua ihua	: 56	67 D Simil	78 Rani iisplay 100 100 100 100	89 Ma optior	tches : Top Sta Early-1 Priv Early-1 Early-1	20 🗸
Taxonomic Leve Phylum Class Order Family Genus COP 20 Matches : Phylum Chordata Chordata Chordata Chordata Chordata Chordata	Taxon Assignment Chordata Actinopterygii Gadiformes Gadidae Gadus Class Actinopterygii Gactino	Probability of Placement (%) 100 100 100 100 100 100 100 100 100 10	Family Gadidae Gadidae Gadidae Gadidae Gadidae	Similarity (%)	nilarity 100.0 99.8 99.6 99.4 99.2 99.0 98.8 98.6 98.4 1 0 Gen Gad Gad Gad Gad Gad Gad	12 12 13 15 15 15 15 15 15 15 15	of Top 23	34 Speco mori mori mori mori mori mori mori	45 ies ihua ihua ihua ihua ihua ihua	: 56	67 D Simil	78 Rani iisplay 100 100 100 100 100 100	89 ked Ma optior	tches : Top Sta Early-1 Prin Early-1 Early-1 Prin	20 🗸
Taxonomic Leve Phylum Class Order Family Cenus COP 20 Matches COP 20 Matches Chordata Chordata Chordata Chordata Chordata Chordata Chordata	Taxon Assignment Chordata Actinopterygii Gadiformes Gadidae Gadus Class Actinopterygii Gactino	Probability of Placement (%) 100 100 100 100 100 100 100 10	Family Gadidae Gadidae Gadidae Gadidae Gadidae	Sim (%) Aireininis	100.0 99.8 99.6 99.4 99.2 99.0 98.8 98.6 98.4 1 0 Gen Gad Gad Gad Gad Gad Gad	12 12 15 15 15 15 15 15 15 15	of Top 23	34 Speco mori mori mori mori mori mori mori	45 ies iuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuu	56	67 D Simil	78 Rani iisplay arity (? 100 100 100 100 100	89 ked Ma optior	tches : Top Sta Early-I Prin Early-I Prin Early-I Prin Prin	20 ▼ tus Release rate rate Release rate rate
Taxonomic Leve Phylum Class Order Family Genus COP 20 Matches : Phylum Chordata Chordata Chordata Chordata Chordata Chordata Chordata Chordata	Taxon Assignment Chordata Actinopterygi Gadiformes Gadiae Gadus Class Actinopterygi Actinopterygi Actinopterygi Gactinopterygi	Probability of Placement (%) 100 100 100 100 100 100 100 100 sidiformes iadiformes iadiformes iadiformes iadiformes iadiformes	Family Gadidae Gadidae Gadidae Gadidae Gadidae Gadidae Gadidae	Similarity (%)	100.0 99.8 99.6 99.4 99.2 99.0 98.8 98.6 98.4 1 Gen Gad Gad Gad Gad Gad Gad Gad Gad	12 12 15 15 15 15 15 15 15 15 15 15 15 15 15	of Top 23	34 34 Spece mori mori mori mori mori mori mori mori	45 ies iua iua iua iua iua iua iua iua	56	67 D Simil	78 Rani iisplay arity () 100 100 100 100 100	89 ked Ma optior	tches : Topp Stat Early-I Prin Early-I Prin Prin Prin Prin Prin	20 v tus Release rate Release Release Release rate
Taxonomic Leve Phylum Class Order Family Cenus COP 20 Matches : Phylum Chordata Chordata Chordata Chordata Chordata Chordata Chordata Chordata Chordata	Taxon Assignment Chordata Actinopterygii Gadiformes Gadidae Gadus Class Actinopterygii G Actinopterygii C Actinopterygii C Actinopterygii C Actinopterygii C Actinopterygii C Actinopterygii C Actinopterygii C Actinopterygii C Actinopterygii C Actinopterygii C Actinopterygii C Actinopterygii C Actinopterygii Actinopterygii C Acti	Probability of Placement (%) 100 100 100 100 100 100 100 100 100 10	Family Gadidae Gadidae Gadidae Gadidae Gadidae Gadidae Gadidae Gadidae	Simitarity (%)	nilarity 100.0 99.8 99.4 99.2 99.0 99.4 99.2 99.0 99.4 99.2 99.4 99.4 99.4 99.4 99.4 99.4	12 12 13 15 15 15 15 15 15 15 15 15 15	of To <u>r</u> 23	34 34 Speco mori mori mori mori mori mori mori mor	45 ies iua iua iua iua iua iua iua iua	56	67 D Simil	78 Rani isplay 100 100 100 100 100 100	89 Mado optior	tches : Top Sta Early-I Prin Early-I Prin Prin Prin Prin Prin	20 v tus Release rate Release
Taxonomic Leve Phylum Class Order Family Cenus COP 20 Matches : Phylum Chordata Chordata Chordata Chordata Chordata Chordata Chordata Chordata Chordata Chordata Chordata	I Taxon Assignment Chordata Actinopterygii Gadiformes Gadidae Gadus Class Actinopterygii G Actinopterygii C Actinopterygii C Actinopterygii C Actinopterygii C Actinopterygii C Actinopterygii C Actinopterygii C Actinopterygii C Actinopterygii C Actinopterygii C Actinopterygii C Actinopterygii C Actinopterygii C Actinopterygii Actinopterygii C Actinopterygii C Actinopterygii Actinopterygii Actinopterygii Actinopterygii Actinopterygii Actinopterygii Actinopterygii Actinopterygii Actinopterygii Actinopterygii Actinopterygii Actinopterygii Actinopterygii Actinopterygii Actinopterygii Actinopterygii Actinopterygii Actinopterygii Actinopterygii Actino	e Probability of Placement (%) 100 100 100 100 100 100 100 100 100 10	Family Gadidae Gadidae Gadidae Gadidae Gadidae Gadidae Gadidae Gadidae	Similarity (%)	nilarity 100.0 99.8 99.4 99.2 99.4 99.2 98.8 98.6 98.4 1 Cad Gad Gad Gad Gad Gad Gad Gad	12 12 13 15 15 15 15 15 15 15 15 15 15 15 15	of To <u>r</u> 23	34 34 Spec mori mori mori mori mori mori mori mori	45 ies iua iua iua iua iua iua iua iua iua iua	56	67 D Simil	78 Rani iisplay arity (* 100 100 100 100 100 100	89 optior	tches : Top Sta Early-1 Prin Early-1 Prin Prin Prin Prin Prin	20 ▼ tus kelease rate kelease kelease kelease rate rate rate

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*.

	.org/index.pnp/IDS_Identi	ificationRequest			ि े द 🛛	 Gadus chalcogrammus 	~ • n
OLDSYSTE	MS Database	es Taxonomy	Identifica	ation Workt	pench Re	sources	User Public
oecimen Ider	ntification Req	luest					🚇 Print
Query: GadusMorh	waSCFAC839-06				Top Hit:	Chordata - Gadiformes - Ga	adus morhua (100%
Search Result: A species level Gadus morhua Gadus chalcogramm	match could not l ws	be made, the que	ried specime:	n is likely to be	one of the f	ollowing:	
			•				
Identification Su	mmary:	Probability of	I. <	imilarity Scores o	of Top 99 Matcl	nes:	- I
Identification Su Taxonomic Level	mmary: 1 Taxon Assignmen	t Probability of t Placement (%)	I. <	100.0 99.8 99.6 99.4	of Top 99 Match	nes:	
Identification Sur Taxonomic Level Phylum	mmary: I Taxon Assignmen Chordata	Probability of t Placement (%) 100		100.0 99.8 99.6 99.4 99.2	of Top 99 Match	nes:	
Identification Su Taxonomic Level Phylum Class Order	mmary: I Taxon Assignmen Chordata Actinopterygii Cadiformer	Probability of Placement (%) 100 100		100.0 99.8 99.8 99.4 99.4 99.2 99.2 99.0	of Top 99 Match	nes:	
Identification Su Taxonomic Leve Phylum Class Order Esmity	mmary: I Taxon Assignmen Chordata Actinopterygii Gadiformes	Probability of Placement (%) 100 100 100		100.0 99.8 99.6 99.4 99.2 99.2 99.0 98.8 98.6	of Top 99 Match	nes:	
Identification Su Taxonomic Leve Phylum Class Order Family Genus	mmary: I Taxon Assignmen Chordata Actinopterygii Gadiformes Gadidae Gadus	t Probability of Placement (%) 100 100 100 100 100		imilarity Scores of 99.8 99.6 99.6 99.4 99.2 99.2 99.2 99.2 99.2 99.2 99.2	of Top 99 Match 23 34 4	5 56 67 78 Banked	
Identification Su Taxonomic Leve Phylum Class Order Family Genus TOP 20 Matches : Phylum	mmary: Taxon Assignmen Chordata Actinopterygii Gadiformes Gadidae Gadus Class	Probability of Placement (%) 100 100 100 100 100 100	Family	imilarity Scores of 99.8 99.6 99.6 99.4 99.2 99.0 98.8 98.6 98.4 1 12	of Top 99 Match 23 34 4 Species	5 56 67 78 Ranked I Display opti Similarity (%)	89 Matches ion: Top 20 ↓
Identification Su Taxonomic Leve Phylum Class Order Family Cenus TOP 20 Matches : Phylum Chordata	mmary: Taxon Assignmen Chordata Actinopterygii Gadiformes Gadidae Gadus Class Actinopterygii	t Probability of Placement (%) 100 100 100 100 100 100 200 0rder Cadiformes	Family Cadidae	imilarity Scores of 99.8 99.6 99.6 99.4 99.2 99.0 98.8 98.6 98.4 1 12 Genus Gadus	of Top 99 Match 23 34 4 Species morhua	5 56 67 78 Ranked I Display opti Similarity (%) 100	89 Matches ion: Top 20 ↓ Status Early-Release
Identification Su Taxonomic Leve Phylum Class Order Family Cenus TOP 20 Matches : Phylum Chordata Chordata	Taxon Assignmen Chordata Actinopterygi Gadidae Gadus Class Actinopterygi Catinopterygi	Probability of Placement (%) 100 100 100 100 100 100 200 200 200 200	Family Gadidae Gadidae	6 100 0 99.8 99.6 99.6 99.6 99.6 99.6 99.6 99.6 99.6 99.6 99.6 99.6 99.6 99.6 99.6 99.6 99.6 99.6 99.6 99.8 99.6 99.8 99.6 99.8 99.6 99.8 99.6 99.8 99.6 99.8 99.6 99.8 99.6 99.8 99.6 99.8 99.6 99.8 99.6 99.8 99.6 99.8 99.6 99.8 99.6 99.8 99.6 99.8 98.8 98.8 98.8 98.8 98.6 98.8 98.8 98.6 98.8 98.6 98.8 98.6 98.8 98.6 98.8 98.6 98.8 98.6 98.8 98.6 98.8 98.6 98.8 98.6 98.8 98.6 98.8 98.6 98.8 98.6 98.8 98.6 98.8 98.6 98.8 98.6 98.8 98.6 98.8 98.6 98.8 98.6 98.8 1 12	23 34 4 Species morhua morhua	5 56 67 78 Ranked I Display opti Similarity (%) 100 100	89 Matches ion: Top 20 V Status Early-Release Private
Identification Su Taxonomic Leve Phylum Class Order Family Genus TOP 20 Matches : Phylum Chordata Chordata	Taxon Assignmen Chordata Actinopterygi Gadidae Gadidae Cadus Actinopterygi Actinopterygi Actinopterygi	t Probability of Placement (%) 100 100 100 100 100 100 200 Eddiformes Eadiformes	Family Gadidae Gadidae Gadidae	Genus Gadus Gadus Gadus	23 34 4 Species morhua morhua morhua	5 56 67 78 Ranked I Display opti Similarity (%) 100 100 100	89 Matches ion: Top 20 V Status Early-Release Private Private
Identification Su Taxonomic Leve Phylum Class Order Family Cenus TOP 20 Matches : Phylum Chordata Chordata Chordata Chordata	Taxon Assignmen Chordata Chordata Gadiformes Gadidae Cadus Cadus Actinopterygii C Actinopterygii C Actinopterygii C	t Probability of Placement (%) 100 100 100 100 100 200 200 200 200 200	Family Cadidae Cadidae Cadidae Cadidae	Genus Gadus	23 34 4 Species morhua morhua morhua	5 56 67 78 Ranked Display opti Similarity (%) 100 100 100 100	89 Matches ion: Top 20 V Status Early-Release Private Private Early-Release
Identification Su Taxonomic Leve Phylum Class Order Family Genus TOP 20 Matches : Phylum Chordata Chordata Chordata Chordata Chordata	Taxon Assignmen Chordata Actinopterygii Gadiformes Gadiformes Gadidae Gadus Class Actinopterygii Actinopterygii Cactinopterygi	t Probability of Placement (%) Placement (%) 100 100 100 100 100 200 Corder Cadiformes Cadiformes Cadiformes Cadiformes Cadiformes Cadiformes	Family Gadidae Gadidae Gadidae Gadidae	Genus Gadus Gadus	23 34 4 Species morhua morhua morhua morhua morhua	5 56 67 78 Ranked Display opti Similarity (%) 100 100 100 100 100	89 Matches ion: Top 20 ↓ Status Early-Release Private Early-Release Early-Release Early-Release
Identification Su Taxonomic Leve Phylum Class Order Family Cenus TOP 20 Matches : Phylum Chordata Chordata Chordata Chordata	Taxon Assignmen Chordata Actinopterygii Gadiformes Gadiformes Gadus Class Actinopterygii Cacinopterygii Actinopterygii Cacinopterygii Cacinopterygii Cacinopterygii Cacinopterygii Cacinopterygii	t Probability of Placement (%) 100 100 100 100 100 100 200 Corder Cadiformes Cadiforme Ca	Family Gadidae Gadidae Gadidae Gadidae	100.0 99.8 99.6 99.6 99.6 99.6 99.7 99.2 99.2 99.8 98.6 98.4 98.8 1 12	23 34 4 Species morhua morhua morhua morhua morhua	5 56 67 78 Ranked ii. Display opti Similarity (%) 100 100 100 100 100 100	89 Matches ion: Top 20 ↓ Status Early-Release Private Early-Release Early-Release Early-Release Early-Release Private
Identification Su Taxonomic Leve Phylum Class Order Family Cenus TOP 20 Matches : Phylum Chordata Chordata Chordata Chordata Chordata	Taxon Assignmen Chordata Actinopterygii Gadiformes Gadiformes Gadidae Gadus Class Actinopterygii C Actinopte	t Probability of Placement (%) 100 100 100 100 100 100 100 100 100 10	Family Gadidae Gadidae Gadidae Gadidae Gadidae	imilarity Scores of 99.8 99.6 99.6 99.6 99.4 99.2 99.2 99.2 99.2 99.2 99.2 99.2 99.4 99.4 99.4 99.2 99.4 99.4 99.4 99.4 99.4 99.4 99.4 99.4 99.4 99.4 99.4 99.5 99.6 99.7 99.6 99.8 98.8 98.8 7 Cadus	23 34 4 Species morhua morhua morhua morhua morhua morhua	5 55 67 78 Ranked Display opti Similarity (%) 100 100 100 100 100 100 100	89 Matches ion: Top 20 ↓ Status Early-Release Private Early-Release Early-Release Early-Release Early-Release Private Private
Identification Su Taxonomic Leve Phylum Class Order Family Cenus TOP 20 Matches : Phylum Chordata Chordata Chordata Chordata Chordata Chordata Chordata Chordata Chordata Chordata Chordata	Taxon Assignmen Chordata Actinopterygii Gadiformes Gadiformes Gadus Cadus Cadus Actinopterygii Cacinopterygi Cacinopterygii Cacinopterygii Cacinopterygii Cacinopterygii Ca	t Probability of Placement (%) 100 100 100 100 100 100 100 200 Placement 100 200 200 200 200 200 200 200 200 200	Family Gadidae Gadidae Gadidae Gadidae Gadidae Gadidae Gadidae	1000 99.8 99.6 99.4 99.4 1 12 12	23 34 4 Species morhua morhua morhua morhua morhua morhua morhua morhua morhua morhua	5 55 67 78 Ranked I Display opti Similarity (%) 100 100 100 100 100 100 100 100 100	89 Matches ion: Top 20 ↓ Status Early-Release Private Private Private Private Private Private Private Private Private Private
Identification Su Taxonomic Leve Phylum Class Order Family Cenus TOP 20 Matches : Phylum Chordata Chordata Chordata Chordata Chordata Chordata Chordata Chordata Chordata Chordata	Taxon Assignmen Chordata Chordata Gadiformes Gadidae Cadidae Cadus Cadus Actinopterygii CACtinopterygii CACtin	t Probability of Placement (%) 100 100 100 100 100 100 100 100 200 100 1	Family Gadidae Gadidae Gadidae Gadidae Gadidae Gadidae Gadidae Gadidae Gadidae	imilarity Scores of 99.8 99.6 99.4 99.7 99.4 99.0 99.4 99.0 99.2 98.6 98.4 1 12 Cadus	23 34 4 Species morhua morhua morhua morhua morhua morhua morhua morhua morhua	5 56 67 78 Ranked 1 ii. Display opti Similarity (%) 100 100 100 100 100 100 100 10	89 Matches ion: Top 20 V Status Early-Release Private Private Private Private Private Private Private Private Private

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_Identifi	icationRequest	☆ マ C 🔠 - Ga	idus chalcogrammus 🔎 🖊 1	
BOLDSYSTEMS Databases	s Taxonomy Identification	Workbench Resou	Log	In 🔅 🔶
Specimen Identification Requ	uest		a F	Print
▼ Query: GadusMorhuaSCEAC839-06		Top Hit: Chor	rdata - Gadiformes - Gadus morhua (100%)
Search Result: A species level match could not be <u>Gadus morhua</u> <u>Gadus chalcogrammus</u>	e made, the queried specimen is li	ikely to be one of the follo	wing:	
BOLD Systems: Management & Analysis - Tree www.boldsystems.org/index.php/IDS_Identifi BOLD SYSTEMS	Result Mozilla Firefox			■ ×
Tree Result -			🖷 Pri	int
Tree Parameters : Data Moc Lab PDF tree : V Export Tree to Newick Format : D Taxonomy Report : V	atype iei iew Tres Download Tree iew Report			😵 feedback
Image List : Spreadsheet : NOTE : Query sequence will be marked up to date in terms of sequence accurate	Tew Image List Tew Spreadsheet If red on the tree with BOLD sequences in b cy and taxonomic identification and will be to	lack. GenBank sequences that ar marked in blue.	re not represented in BOLD may be	less

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*.



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).

Firefox 🔻		_	_		_			×
BOLD Systems: Managemen	t & Analysis +	_			_	- -	-	
www.boldsystems	.org/index.php/IDS_OpenIdE	ngine		☆ ⊽ C	8 - Gadus chal	:ogrammus 🔎	∔ 🏫	•
BOLD SYSTEM	S Databases equest	Taxonomy Identif	fication	Workbench	Resources		Log In @ Print	• • •
Animal Identification [COI]	Fungal Identification [ITS]	Plant Identification [rbcL & matK]						
The BOLD Identification 5 identification when one i Historical Databases: J Search Databases: All Barcode Recorr Every COI barcode This includes many matches and does i Species Level Barc Every COI barcode Every COI barcode Continued and the second continued of the Speci identification as it	system (IDS) for COI accepts sis possible. Further validation v ul-2013 Jul-2012 Jul-2011 Is on BOLD (2,742,418 Sequ record on BOLD with a minimu species represented by only on to provide a probability of pl ode Records (1,739,723 Seq record with a species level ide soft operation with interim taxe records from BOLD with Centar I Barcode Database (1,318,5 es library with a minimum sec provides maximum overlap wi	<pre>squences from the 5' region of vith independent genetic marka , jul=2010 jul=2009 ences) mm sequence length of 500bp (one or two specimens as well as acement to a taxon. uences/146,084 Species/58,3 ruffication and a minimum sequence le ynomy. uences/62,894 Species/14,98 k with a minimum sequence le 74 Sequences/132,900 Specie juence length of 640bp and cor th short reads from the barcodo the short reads from the barcodo protect of the sequence le sequence length of 640bp and cor the short reads from the barcodo protect of the sequence le sequence length of 640bp and cor the short reads from the barcodo protect of the sequence le sequence length of 540bp and cor the short reads from the barcodo protect of the sequence le sequence length of 540bp and cor the short reads from the barcodo protect of the sequence le sequence length of 540bp and cor the short reads from the barcodo protect of the sequence le sequence length of 540bp and cor the short reads from the barcodo protect of the sequence le sequence length of 540bp and cor the short reads from the barcodo protect of the sequence le sequence length of 540bp and cor the short reads from the barcodo protect of the sequence le sequence length of 540bp and cor the short reads from the sequence le sequence length of 540bp and cor the short reads from the sequence le sequence length of 540bp and cor the short reads from the sequence length of 540bp and cor the short reads from the sequence length of 540bp and cor the short reads from the sequence length of 540bp and cor the short reads from the sequence length of 540bp and cor the short reads from the sequence length of 540bp and cor the short reads from the sequence length of 540bp and cor the short reads from the sequence length of 540bp and cor the seq</pre>	the mitochondri ers will be desir warning: unvalit s all species with 357 Interim Spe under Solops. S5 Interim Spect ength of SOObp. es/50,814 Inter ntaining both pu ergion of COI.	al Cytochrome c oxi able in some forensi lated library and inc n interim taxonomy. cies) 500bp. This include es) This library is a coll- im Species) bilic and private reco	dase subunit I gene i c applications. udes records withou This search only retu s many species repre ection of records fror ords. This library is ir	and returns a specie t species level ident irns a list of the nea sented by only one n the published pro atended for short se	s-level ification). irest or two jects quence	။ ကြို့ feedback
Enter sequences in fasta CadusMorhuaSCFAC CCTITATCTCGTATTIT TCAACCTGCTGCACTT CTTTATAGTAATACCA ACCTTTCCCTCGCATA TGTAGAAGCTGCGCGC TGTTGATCTCACTATT AATTATTAATATCAAA TGTGCTTCTACTATTA TTCTTTCTTTGACCCT	format: 839-06 GGTGCCTGAGCCGGCATAGTC CTTGGTGATGATGATGAAGTTTAT, CTAATAATTGGAGGCTTTGGG AATAACATAACCTTCTGACTT GGAACAGGCTGAACTGTCTAT TTTTCTTCTTCATCTAGCAGCGC CCTCCGCGCAATTCACAGTAC TCTCTCCCCCGTCTTAGCAGCGT CCTCGGAGGAGGTGATCCCATT	СGААСАGCCCTAAGCCTGCTCA ЧАТСТGАTCGTTACAGCCCACG ЧАСТGАСТСАТТССТСТААТGG ТГССТССАТСТТТССТССТСТ ТСССТССАТСТТТССТССТ ССАССТТТАGCCGGAAACCTCC СТАТСАСАТТСТТТСАГС ЗСТАТСАСААТАСТТТСАГС ЗСТАТСАСААТАСТТСТААСТСА ТТАТАССААСА	ATTCGAGCAGAG CTTTCGTAATA TCGGTGCACCAA TTTTAGCATCCT GCTCATGCTGGC TTAATTTTATTAA ACCGTAATCTTA	CTAAG ITTT IATAT IGCATC IGCATC ICAC CACC ATAC			.t.	
							Submit	
•		"	11					P.

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.

BOLD Systems: Management	& Analysis +			X
+ > @ www.boldsystems.	org/index.php/IDS_OpenIdEr	igine	☆ マ C 🛛 🚼 - Gadus chalcogrammus 🔎 🖡 🏫	
BOLD SYSTEMS	Databases ' quest	Taxonomy Ider	Log In tification Workbench Resources	it
Animal Identification [COI]	Fungal Identification [ITS]	Plant Identification [rbcL & matK]		
identification when one is Historical Databases: Ju Search Databases: Ju All Barcode Record All Barcode Record Secret Coll barcode r This includes many matches and does n Species Level Barco Every Coll barcode r speciemens as well as Public Record Barco All published Col re section of BOLD. Full Length Record Subset of the Specie identification as it p Enter sequences in fasta ff ScadusMorhuaSCFAC8 CCITTATCICGTATTC TCAACCTGCTGCACTTTC CTITATACTAATACCAC ACCTTTCACCTCCACATTT CTICAACCTCCACGCCT	possible. Further validation w I-2013 Jul-2012 Jul-2011 s on BOLD (2,742,41) Seque cord on BOLD The mean imimu species represented by only o to provide a probability of pla de Records (1,739,732 Seque cord with a species level ider all species with interim taxo ode Database (555,693 Seque cords from BOLD and CenBan Barcode Database (1,318,57) s libaray with a minimum seque rovides maximum overlap with primat: 39-06 GTCCCTGACCGCCCCATACTCG CTGCTGATGATCAACTTTACAT TGCTGAACATCAACTTTCACTTCGCAACTGTCTTAT TGCTGATCAATCACCACCACTCTCAT CTCCTCCACTACTCTCACTTCTCATCTTCACTTCA	ith independent genetic m jul-2010 jul-2009 mees) m sequence length of 5001 ne or two specimens as we icement to a taxon. tences/146,084 Species/1 tiffication and a minimum : nomy. tences/62,894 Species/1 k with a minimum sequence /4 Sequences/132,900 Sp uence length of 640bp and h short reads from the ban GGAACAGCCCTAACCCTGC AGTGGATCGATTCGATCGCTAACCCCCCC AGTGGATCGATTCGATCGCTAACCCCCCC AGTGGATCGATTCGATCGATCGATCGATCGATCGATCGAT	arkers will be desirable in some forensic applications. pp (warning: unvalidated library and includes records without species level identification). Il as all species with interim taxonomy. This search only returns a list of the nearest s8,357 Interim Species) sequence length of 500bp. This includes many species represented by only one or two s985 Interim Species) e length of 500bp. This library is a collection of records from the published projects ecies/50,814 Interim Species) containing both public and private records. This library is intended for short sequence containing both public and private records. This library is intended for short sequence code region of COI. CCATTICCACCACACCTAAC CCCTTCG TCACTCACCACCACCTCA CCATTITIACCATCCTCG TCACTCACCACCACCTCAC CCATTITIATTACCACCACACCCCCCACCATE CCATTITIATTACCACCACACCCCCACCATE CCATTICACATITIATTACCACCACACCCCCCCACCATE CCATTICACATITIATTACCACCACACCCCCCCCACCACCACCACCACCACCAC	때 양 feedback
TICTTICTTICACCCTC	CTGGAGGAGGTGATCCCATT	TATACCAACA	.d. Submit	

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 ($<2ng/\mu I$). 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 (<u>www.ncbi.nlm.nih.gov/</u>). However, the correct use of these databases falls out of the scope of this SOP.

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

The two references cited in this SOP are:

Ivanova, NV, Zemlak, TS, Hanner, R, & Hebert, PDN (2007). Universal primer cocktails for fish DNA barcoding. *Molecular Ecology Notes*, **7**: 544–548.

Ward, RD, Zemlak TS, Innes BH, Last, PR, & Hebert, PDN (2005). DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B*, **360**: 1847–57.