

Discovering the uncharacterizable world under the sea using genetic tools

Keynote Edu-lab 2

Jana Dömel



UNIVERSITÄT
DUISBURG
ESSEN

Open-Minded

Acknowledgment

HOCHSCHULWETTBEWERB

Im Wissenschaftsjahr 2016*17 – Meere Ozeane

Zeigt eure Forschung!

Eine Initiative des Bundesministeriums
für Bildung und Forschung

Wissenschaftsjahr 2016 * 17

MEERE
UND OZEANE

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

wissenschaft : im dialog

Structure of scientific work

- Introduction
 - Overview
 - Hypothesis
- Material
- Methods
- Results
- Discussion

General Introduction

Subject: Biology

Useful background knowledge: Genetic basics (DNA structure)

- Southern Ocean
- Antarctic toothfish
 - Adaptation
 - Commercial fishing
- Threats
 - Illegal fishing
 - Sold on the market using wrong labels

Hypothesis 1:

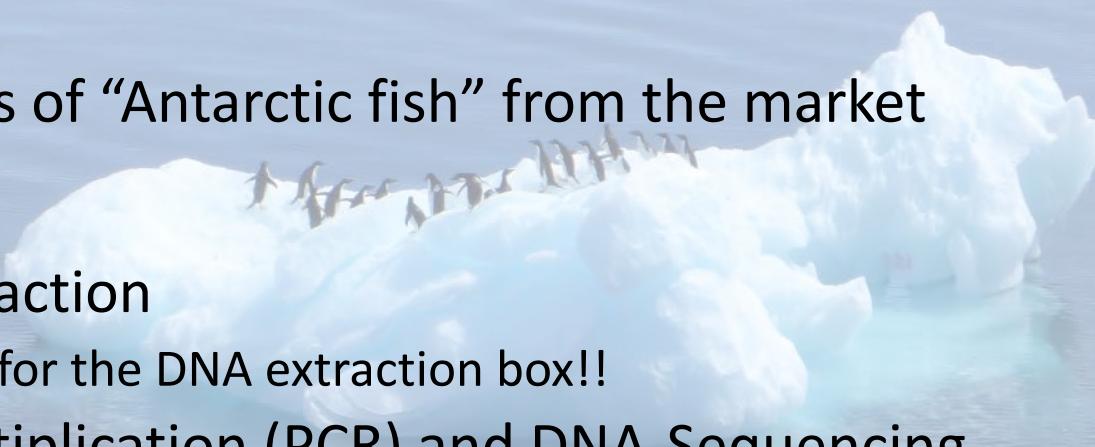
Genetic analyses help us to identify species. This can help us to find out which fish is sold on the market and if it was illegally caught.

Dissostichus mawsoni

Worksheet 1:

Genetic analyses of Antarctic toothfish

- Material
 - 7 samples of “Antarctic fish” from the market
- Methods
 - DNA extraction
 - apply for the DNA extraction box!!
 - DNA multiplication (PCR) and DNA-Sequencing
 - watch YouTube videos provided or others



Sequence Analyses

BLAST search

<https://blast.ncbi.nlm.nih.gov/Blast.cgi>

The screenshot shows the NCBI BLAST homepage. At the top, there are links for NIH, U.S. National Library of Medicine, NCBI National Center for Biotechnology Information, and user accounts (janasd, My NCBI, Sign Out). A banner at the top right announces "Magic-BLAST 1.2.0 released" with a message about a new version of the BLAST RNA-seq mapping tool. Below the banner, the "Basic Local Alignment Search Tool" is introduced with a brief description. The main section is titled "Web BLAST" and features three large buttons: "Nucleotide BLAST" (nucleotide > nucleotide), "blastx" (translated nucleotide > protein), and "tblastn" (protein > translated nucleotide). To the right, there is a "Protein BLAST" button (protein > protein). Below these buttons, there is a section for "BLAST Genomes" with a search bar and options for Human, Mouse, Rat, and Microbes.

The screenshot shows a DNA sequence from a BLAST search. The sequence is presented in FASTA format with a red border around it. The sequence starts with >DNA1 and continues with several other entries (DNA2, DNA3, DNA4, DNA5, DNA6, DNA7, DNA8) each preceded by a > symbol. The sequence itself consists of a long string of nucleotide bases (A, T, C, G).

```
>DNA1
CCTTATGATGGGGCCCCGACATAGCTTCCCCGTATGAATAATATGAGCTTCTGACTCCTGCCCTCATTCCTCTTAC
>DNA2
CCTCATGATGGGGCCCCGACATGGTTCCCTCGAATAAAATATGAGTTCTGGCTCTCCCTCCCTCTTCTACTCTT
>DNA3
CTCATCCCCCTATGATGGGGCCCCGACATAGCTTCCCCGTATGAATAATATGAGCTTCTGACTCCTGCCCTCATTCCTCTTAC
>DNA4
CTCATGATGGGGCCCCGACATGGTTCCCTCGAATAAAATATGAGTTCTGGCTCTCCCTCCCTCTTCTACTCT
>DNA5
TCATGATGGAGCCCCGACATGGTTCCCGAATAAAATATGAGTTCTGGCTCTCCCTCCCTCTTCTACTACT
>DNA6
CCTCTTAGCAGGAACACTCCACCTGGAGCCTCGTAGACCTAACCATCTCTCTTACACCTAGCAGGTGTCTCTTAT
>DNA7
TCCCCCTCATGATGGAGCCCCGACATGGTTCCCGAATAAAATATGAGTTCTGGCTCTCCCTCCCTCTTACTA
```

The screenshot shows the "Standard Nucleotide BLAST" search interface. At the top, there are tabs for "blast", "blastp", "blastn", "tblastn", and "tblastx". The main search area has a "Enter Query Sequence" input field with a placeholder "Enter accession number(s), g(i)s, or FASTA sequence(s)". Below this is a "Query subrange" section with "From" and "To" fields. There are also fields for "Or, upload file" (with a "Choose File" button) and "Job Title". Under "Choose Search Set", there are options for "Database" (Human genomic + transcript, Mouse genomic + transcript, Others (nr etc.)), "Organism" (Optional), "Exclude" (Optional), "Limit to" (Optional), and "Enter Query" (Optional). In the "Program Selection" section, there are radio buttons for "Highly similar sequences (megablast)", "More dissimilar sequences (discontiguous megablast)", "Somewhat similar sequences (blastn)", and "Choose a BLAST algorithm". At the bottom, there is a large blue "BLAST" button and a red call-to-action button labeled "Start analysis!" with the text "Search database using your query sequence above. Then, choose Optimize for highly similar sequences" below it.

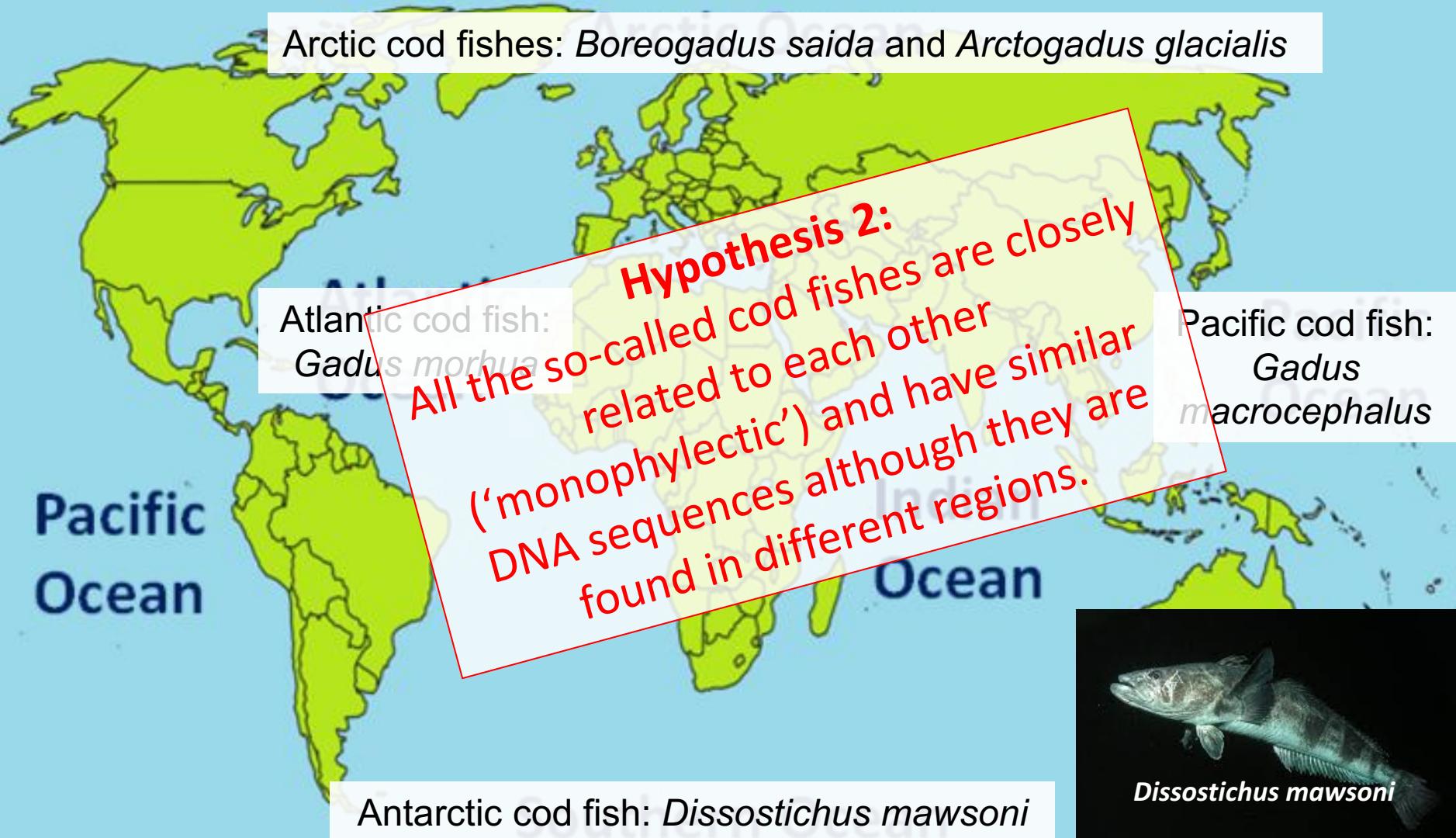
Results and Discussion

- DNA 1
- DNA 2
- DNA 3
- DNA 4
- DNA 5
- DNA 6
- DNA 7



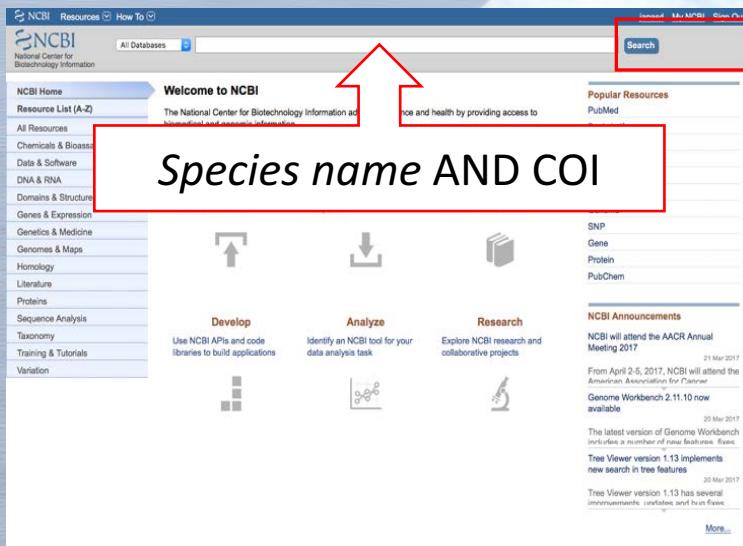
Worksheet 2:

Phylogenetic analyses of cod fishes



Material

- DNA-Sequences from NCBI:
<https://www.ncbi.nlm.nih.gov>



Welcome to NCBI

Species name AND COI

Search

Popular Resources: PubMed

NCBI Home Resource List (A-Z) All Resources Chemicals & Bioassays Data & Software DNA & RNA Domains & Structure Genes & Expression Genetics & Medicine Genomes & Maps Homology Literature Proteins Sequence Analysis Taxonomy Training & Tutorials Variation

Develop Analyze Research

SNP Gene Protein PubChem

NCBI Announcements

NCBI will attend the AACR Annual Meeting 2017

From April 2-5, 2017, NCBI will attend the American Association for Cancer Research Annual Meeting.

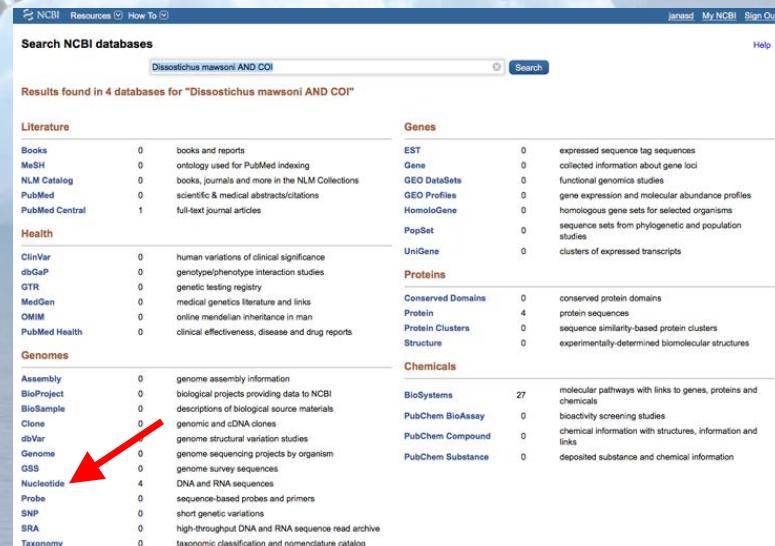
Genome Workbench 2.11.10 now available

The latest version of Genome Workbench includes a number of new features. View more.

Tree Viewer version 1.13 implements new search tree features

Tree Viewer version 1.13 has several improvements. View more.

More...



Search NCBI databases

Dissostichus mawsoni AND COI

Results found in 4 databases for "Dissostichus mawsoni AND COI"

Literature

Books	0	books and reports
MeSH	0	ontology used for PubMed indexing
NLM Catalog	0	books, journals and more in the NLM Collections
PubMed	0	scientific & medical abstracts/citations
PubMed Central	1	full-text journal articles

Health

ClinVar	0	human variations of clinical significance
dbGaP	0	genotype/phenotype interaction studies
GTR	0	genetic testing registry
MedGen	0	medical genetics literature and links
OMIM	0	online mendelian inheritance in man
PubMed Health	0	clinical effectiveness, disease and drug reports

Genes

EST	0	expressed sequence tag sequences
Gene	0	collected information about gene loci
GEO DataSets	0	functional genomics studies
GEO Profiles	0	gene expression and molecular abundance profiles
HomoloGene	0	homologous gene sets for selected organisms
PopSet	0	sequence sets from phylogenetic and population studies
UniGene	0	clusters of expressed transcripts

Proteins

Conserved Domains	0	conserved protein domains
Protein	4	protein sequences
Protein Clusters	0	sequence similarity-based protein clusters
Structure	0	experimentally-determined biomolecular structures

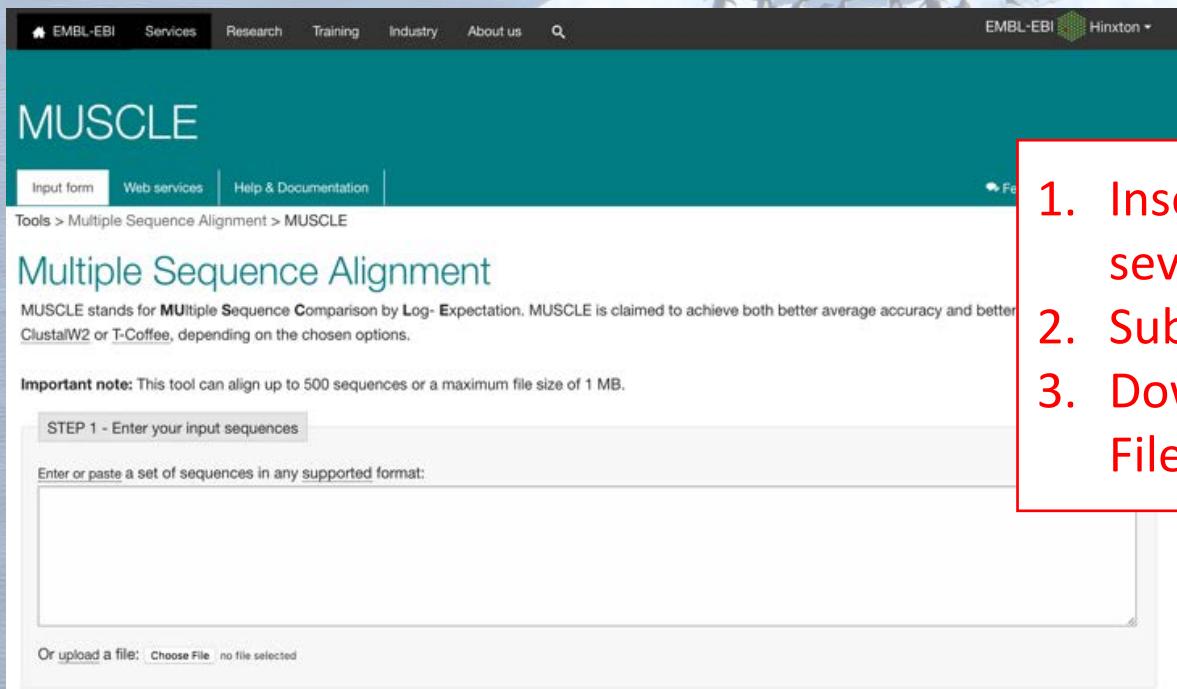
Genomes

Assembly	0	genomic assembly information
BioProject	0	biological projects providing data to NCBI
BioSample	0	descriptions of biological source materials
Clone	0	genomic and cDNA clones
dbVar	0	genomic structural variation studies
Genome	0	genome sequencing projects by organism
GSS	0	genome survey sequences
Nucleotide	4	DNA and RNA sequences
Probe	0	sequence-based probes and primers
SNP	0	short genetic variations
SRA	0	high-throughput DNA and RNA sequence read archive
Taxonomy	0	taxonomic classification and nomenclature catalog

Create a FASTA file with several DNA sequences.

Methods

- Sequence alignment with MUSCLE:
<http://www.ebi.ac.uk/Tools/msa/muscle/>



MUSCLE

Input form Web services Help & Documentation

Tools > Multiple Sequence Alignment > MUSCLE

Multiple Sequence Alignment

MUSCLE stands for MUltiple Sequence Comparison by Log-Expectation. MUSCLE is claimed to achieve both better average accuracy and better ClustalW2 or T-Coffee, depending on the chosen options.

Important note: This tool can align up to 500 sequences or a maximum file size of 1 MB.

STEP 1 - Enter your input sequences

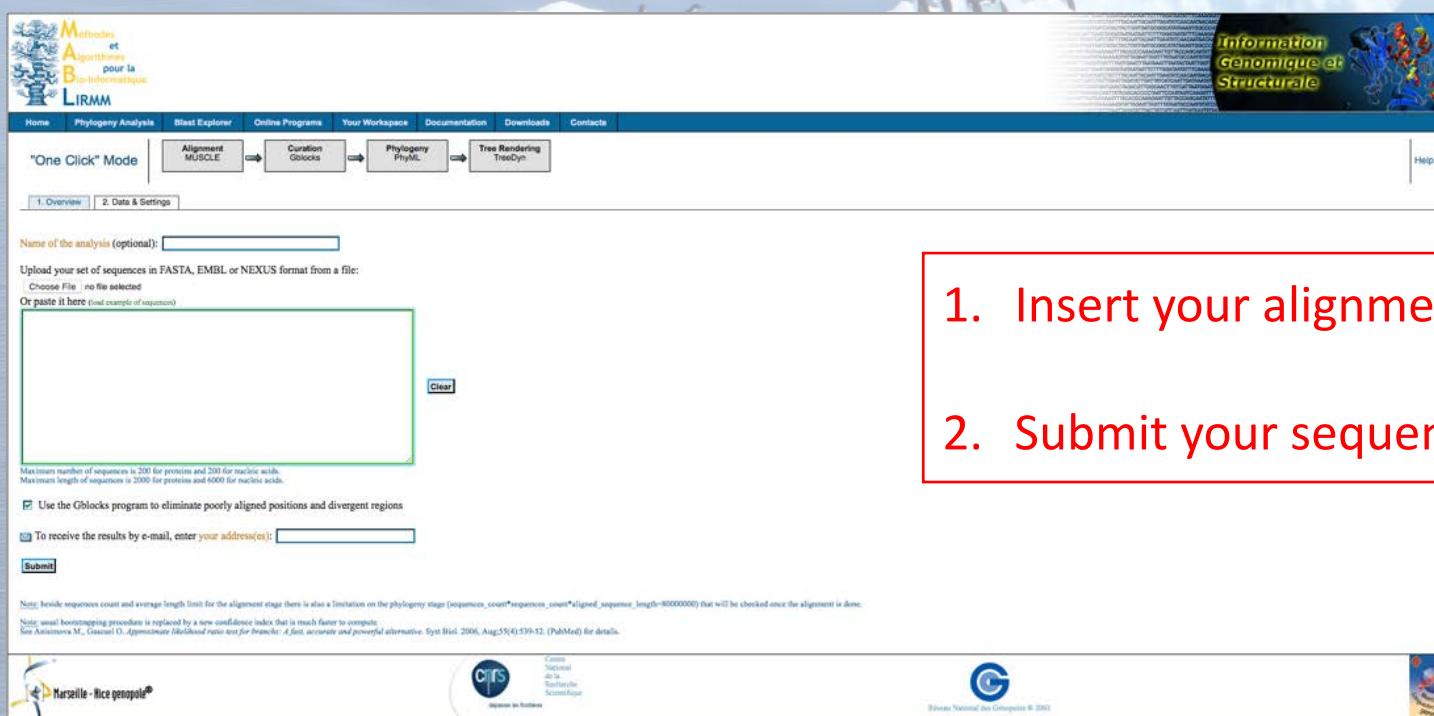
Enter or paste a set of sequences in any supported format:

Or upload a file: Choose File no file selected

1. Insert FASTA file with several DNA sequences.
2. Submit your sequences.
3. Download Alignment File.

Methods

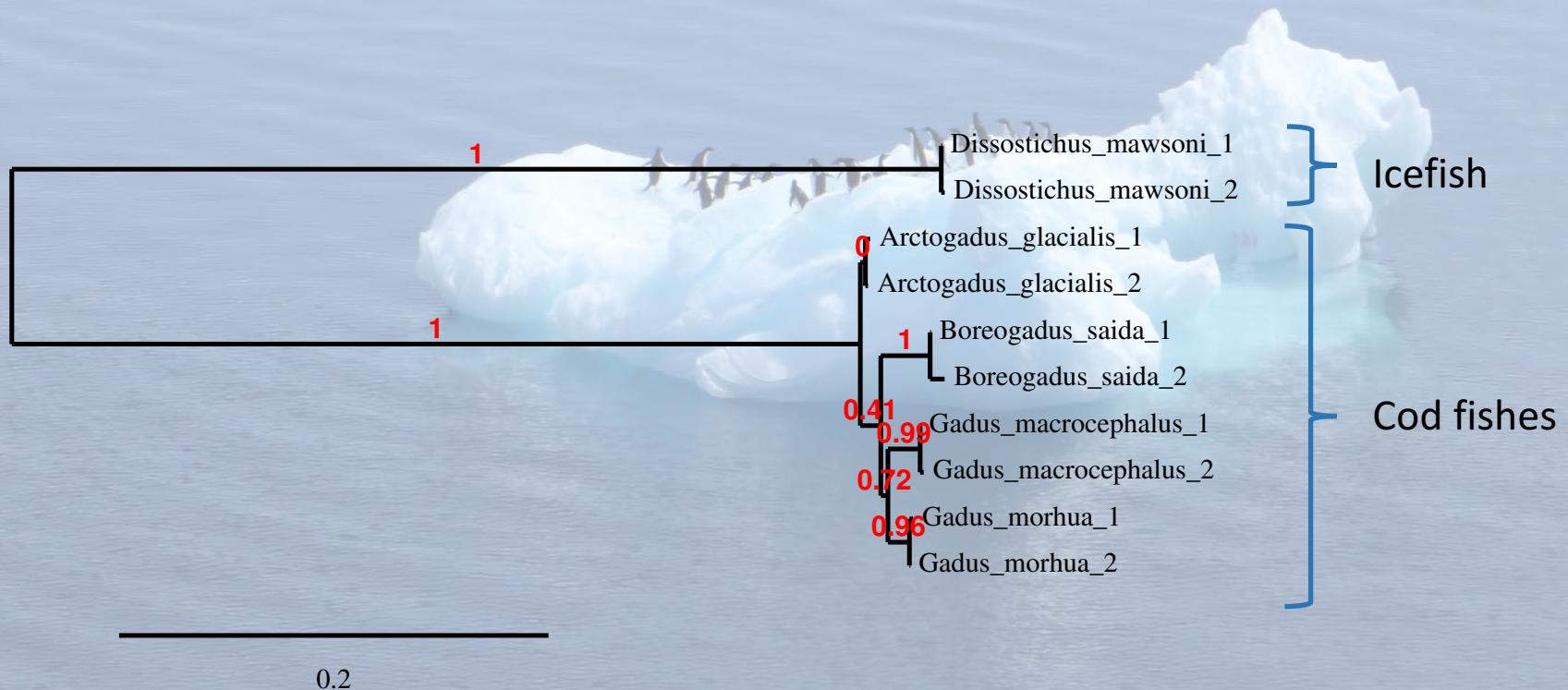
- Reconstruction of phylogenetic tree:
http://www.phylogeny.fr/simple_phylogeny.cgi



The screenshot shows the Phylogeny.fr website's 'One Click' mode interface. At the top, there is a navigation bar with links: Home, Phylogeny Analysis, Blast Explorer, Online Programs, Your Workspace, Documentation, Downloads, and Contacts. Below the navigation bar, a banner features the LIRMM logo and the text 'Information Génomique et Structurale'. The main area has a blue header 'One Click' Mode with arrows pointing from 'Alignment MUSCLE' to 'Curation Gblocks', then to 'Phylogeny PhyML', and finally to 'Tree Rendering TreeDyn'. Below this, there are two tabs: '1. Overview' (selected) and '2. Data & Settings'. A text input field is labeled 'Name of the analysis (optional):' followed by a file upload input field with the message 'Upload your set of sequences in FASTA, EMBL or NEXUS format from a file: Choose File : no file selected'. Below that is a text input field labeled 'Or paste it here (load example of sequence)' with a large green-bordered text area for pasting sequences. A 'Clear' button is located next to the text area. At the bottom of the form, there are several checkboxes: 'Use the Gblocks program to eliminate poorly aligned positions and divergent regions', 'To receive the results by e-mail, enter your address(es):' with a text input field, and a 'Submit' button. A note at the bottom states: 'Note: besides sequences count and average length limit for the alignment stage there is also a limitation on the phylogeny stage (sequences_count*sequences_count*aligned_sequence_length<=8000000) that will be checked once the alignment is done.' and 'Note: usual bootstrapping procedure is replaced by a new confidence index that is much faster to compute. See Anisimova M., Gascuel O. Approximate likelihood ratio test for branch: A fast, accurate and powerful alternative. Syst Biol. 2006, Aug;55(4):559-62. (PubMed) for details.'

1. Insert your alignment.
2. Submit your sequences.

Results and Discussion



A photograph of two Adelie penguins standing on a large, light-colored ice floe. They are facing each other, appearing to interact. The background is a vast, icy landscape with more ice floes and a distant, snow-covered mountain range under a clear sky.

Questions?
Suggestions?
Ideas?

Thank
you!!!

