

Restaurering af østers i limfjorden

# bioinformatikøvelse - elevark

**Udarbejdet af**Homère Alves Monteiro, DTU AQUA og Søren Søgaard, Rysensteen Gymnasium

**Fag**Bioteknologi A og Biologi A

**Verdensmål**

[](https://www.verdensmaalene.dk/maal/12) [](https://www.verdensmaalene.dk/maal/14) [](https://www.verdensmaalene.dk/maal/15)

# bioinformatikøvelse - elevark

## Formål

Formålet med denne øvelse er at anvende bioinformatik til at analysere slægtskabet imellem de forskellige europæiske fladøsters. Denne information anvendes i et forskningsprojekt, som har til formål at undersøge, hvordan vi bedst muligt bevarer den population, vi har af fladøsters i Limfjorden.

Øvelsen har 3 dele, som står beskrevet, når øvelsen løbes igennem og følges kronologisk. De 3 dele illustrerer fint arbejdsgangen i et forskningsprojekt med ovennævnte formål.

## Introduktion

* Se [introduktionsdokumentet](https://www.verdensmaalene.dk/sites/default/files/introtekst_til_problemstillingen_med_at_opbygge_en_baeredygtig_oesterspopulation_5.docx), der følger med til projektet.
* Se filmene med ph.d.-studerende ved DTU Aqua, Homère Alves Monteiro, som kort beskriver forskningsprojektet:
  + [”Restaurering af østers i Limfjorden - bioinformatik og DNA-sekventering”](https://www.youtube.com/watch?v=Zl7LghOvDyQ&t=3s)
  + [​​​​​​​](https://www.verdensmaalene.dk/sites/default/files/bioinformatikoevelse_elevark.docx)["DNA-sekventering - metodefilm"](https://www.youtube.com/watch?v=EP5Q3KHF5IE&t=2s)

Oversigt over de forskellige østers og deres latinske navne

A picture containing text, mollusk, screenshot

Description automatically generated

*Fig.1: Forskelle i anatomi hos arter af østers*

- Reference mitokondrie genom fra *Pinctada maxima* (Perleøsters er ikke samme genus)

A group of coins

Description automatically generated with low confidence

*Fig.2: Perleøsters*

- Reference mitokondrie genom fra *Crassotrea Belcheri* (Tropisk Østers (Asien), er ikke samme genus)

A picture containing invertebrate, mollusk, indoor, seasnail

Description automatically generated

*Fig.3: Asiatisk fladøsters*

Oversigt over hvor fladøsters er høstet geografisk i Europa og Skandinavien

Et billede, der indeholder kort

Automatisk genereret beskrivelse

*Indsamlingssteder for fladøsters i Europa.*

Et billede, der indeholder kort

Automatisk genereret beskrivelse

*Indsamlingssteder for fladøsters i Skandinavien.*

## Datasæt

<https://www.verdensmaalene.dk/sites/default/files/datasaetoesters.zip>

Start med at downloade hele mappen med strukturer til fx dit skrivebord (et sted, der er nemt at finde).

## Online værktøj til bioinformatik

<https://mafft.cbrc.jp/alignment/server/add_fragments.html>

Gå til denne hjemmeside. Luk meget gerne mange af de andre vinduer, så du ikke skal lede efter denne side, hver gang du skal bruge den.

## Det eksperimentelle arbejde i 3 dele:

**Del 1 - Har vi fat i de rigtige sekvenser?**

Formålet med denne del af den bioinformatiske øvelse handler om, at vi skal tjekke, om de sekventerede mtDNA-sekvenser alle tilhører den europæiske fladøsters. Dette gøres ved at udforme et fylogenetisk træ over de fundne sekvenser, som vi formoder er fra den europæiske fladøsters, og sammenholde disse med en referencesekvens, som IKKE er fra den europæiske fladøsters.

Ses der en klynge af vores formodede europæiske fladøsterssekvenser, væk fra referencen, så vil de alle være af den ønskede art.

**Udførsel:**

Gå til online-værktøjet mafft på følgende link: <https://mafft.cbrc.jp/alignment/server/add_fragments.html>

Gå desuden til [datamappen](https://www.verdensmaalene.dk/sites/default/files/datasaetoesters.zip) for at gøre klar til at hente vores sekvenser fra sekventeringen samt referencesekvenserne.

Start med at downloade hele mappen med strukturer til fx dit skrivebord (et sted, der er nemt at finde).

Vi skal nu have uploadet vores sekvenser, som vi ønsker at undersøge for slægtskab imellem samt vores referencesekvens.

Scroll ned til dette skærmbillede:

Et billede, der indeholder tekst

Automatisk genereret beskrivelse

Gør derefter følgende:

1. Upload vores reference mitochondrial genome:

(Pinctada maxima : *file=PinctadamaximaMTgenome.fasta*) i feltet “Existing alignment”, hvilket gøres ved at ”choose file” under boksen, vælge mappen du lige har downloadet, vælge mappen RefMTfasta, vælge filen *PinctadamaximaMTgenome.fasta*

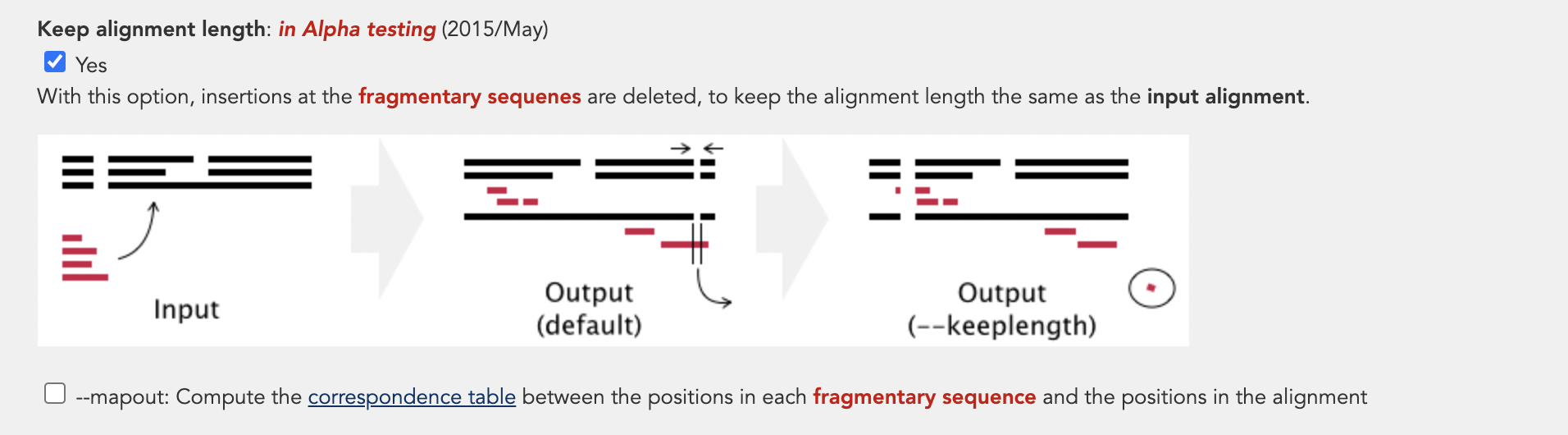
1. Upload vores samling af sekvenser af mitochondrial genome:

(file=Final\_fasta\_allMTgenome.fasta) i feltet ” Fragmentary sequence(s)”, hvilket gøres ved at ”choose file” under boksen, vælge mappen Merged\_dataset, vælge filen Final\_fasta\_allMTgenome.fasta

1. Et billede, der indeholder tekst

   Automatisk genereret beskrivelseJuster parametrene uniformative sequences til 50% (dette tales der mere om i del 2)
2. Et billede, der indeholder tekst

   Automatisk genereret beskrivelseSørg for, at følgende indstillinger er afkrydset i felterne, som det ses nedenfor:



Et billede, der indeholder tekst

Automatisk genereret beskrivelse

1. Tryk derefter submit og vent. Alt efter, hvor meget serveren anvendes, vil ventetiden være varierende.

Dette vil være det korrekte ”ventevindue”:

Et billede, der indeholder tekst

Automatisk genereret beskrivelse

1. Når beregningen er færdig, skal du vælge ***Phylogenetic tree*** og trykke ”go”

Et billede, der indeholder tekst

Automatisk genereret beskrivelse

1. Til sidst skal du vælge at vise det fylogenetiske træ ved fx ”view tree on Archaeopteryx.js

Et billede, der indeholder tekst

Automatisk genereret beskrivelse

**Del 2 - Problemløsning med ”dårlige sekventeringer”**

Formålet med denne del af den bioinformatiske øvelse handler om, at det umiddelbart lyder underligt, at Limfjordsøsters klynger så tæt med de spanske fladøsters, de er trods alt geografisk langt fra hinanden. Hvis vi i et alignment kun sammenholder de sekvenser, som vi i første tjek har set ligger i samme familie, og undersøger de ”underlige klynger” i rådata, kan vi måske finde evt. dårlige sekventeringer. Hvis vi går ind i rådata for dette alignment, kan vi se, at der i sekvenserne findes en del sekventeringer, der angiver en del baser som N. Dette er en sekventering, hvor sekventeringsmaskinen ikke kan give en sikker base i den pågældende placering.

Fjerner vi alle disse dårlige/usikre sekventeringer, skal vi tjekke, om dette ændrer på klyngerne.

1. Start med at gå 3 gange tilbage i dit browservindue, på den måde ender du tilbage ved vinduet, hvor der kan uploades sekvenser, men vi bevarer alle indstillingerne i programmet.
2. Upload vores reference mitochondrial genome:

file=*OstreaLuridaMTgenome.fasta)* i feltet “Existing alignment”, hvilket gøres ved at ”choose file” under boksen, vælge mappen du lige har downloadet, vælge mappen RefMTfasta, vælge filen *OstreaLuridaMTgenome.fasta*

1. Upload vores samling af sekvenser af mitochondrial genome:

(file=*DatasetMTgenomecombined.fasta* i feltet ” Fragmentary sequence(s)”, hvilket gøres ved at ”choose file” under boksen, vælge mappen Merged\_dataset, vælge filen *DatasetMTgenomecombined.fasta*

1. Tryk derefter submit og vent. Alt efter, hvor meget serveren anvendes, vil ventetiden stadig være kortere eller længere.

Dette vil være det korrekte ”ventevindue”

Et billede, der indeholder tekst

Automatisk genereret beskrivelse

1. Når beregningen er færdig, og I har orienteret jer i rådata, skal du vælge ***Phylogenetic tree*** og trykke ”go”

Et billede, der indeholder tekst

Automatisk genereret beskrivelse

1. Nu skal du vælge at vise det fylogenetiske træ ved fx ”view tree on Archaeopteryx.js

Et billede, der indeholder tekst

Automatisk genereret beskrivelse

1. Ud fra det fremkomne fylogenetiske træ er det nu muligt og se, om der er nogle sekvenser, som klynger lidt underligt. Hvis du går to skridt tilbage i browser, kan du igen se sekvenserne angivet med baser. Undersøg, om nogle af de underlige klynger har mange N baser.

Når resultatet vises, er det muligt at scrolle ned igennem alignmentsekvenserne og se, om der er mange N baser i sekvenserne. Se uddrag her:

Et billede, der indeholder tekst

Automatisk genereret beskrivelse

*Uddrag fra alignmentsekvenserne fra resultatoversigten*

**Del 3 - Det gode fylogenetiske træ, som kan give os rådgivning**

Formålet med denne del af den bioinformatiske øvelse handler om, at der nu er ryddet op i sekvenser med for mange usikre baseangivelser. Det vil altså være nødvendigt at foretage flere sekventeringer af indfangne fladøsters fra Limfjorden for at opnå samme antal individer i analysen.

Med de opryddede sekvenser kan der nu laves et nyt fylogenetisk træ, som kan anvendes til rådgivning om genetisk variation i europæiske fladøsters.

1. Start med at gå 3 gange tilbage i dit browservindue, på den måde ender du tilbage ved vinduet, hvor der kan uploades sekvenser, men vi bevarer alle indstillingerne i programmet.
2. Upload vores reference mitochondrial genome:

(*File= OstreaLuridaMTgenome.fasta* i feltet “Existing alignment”, hvilket gøres ved at ”choose file” under boksen, vælge mappen du lige har downloadet, vælge mappen RefMTfasta, vælge filen *OstreaLuridaMTgenome.fasta*

1. Upload vores rensede samling af sekvenser af mitochondrial genome: *(File2=Final\_goodGenome.fasta)* i feltet ” Fragmentary sequence(s)”, hvilket gøres ved at ”choose file” under boksen, vælge mappen Good\_genome, vælge filen *Final\_goodGenome.fasta*
2. Tryk derefter submit og vent. Alt efter hvor meget serveren anvendes, vil ventetiden stadig være kortere eller længere.

Et billede, der indeholder tekst

Automatisk genereret beskrivelseDette vil være det korrekte ”ventevindue”

1. Når beregningen er færdig, skal du vælge ***Phylogenetic tree*** og trykke ”go”
2. Til sidst skal du vælge at vise det fylogenetiske træ ved fx ”view tree on Archaeopteryx.js
3. Ud fra det fylogenetiske træ, hvor dårlige sekvenser nu er fjernet, vil det være muligt at analysere på slægtskabet imellem de europæiske fladøsters og hermed vurdere, hvordan vi bedst bevarer alle disse populationer.