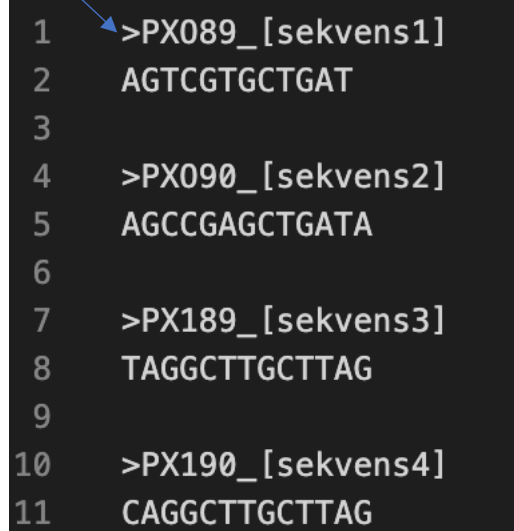


Brug af MEGA til alignment og fylogenetiske træer

En minimal vejledning til, hvordan man laver alignment og konstruerer træer ved brug af programmet MEGA <https://www.megasoftware.net/>, se hjemmesiden for detaljeret dokumentation.

Afsnit 1. Filer En meget almindeligt brugt fil – type der bruges til at opbevare sekvenser, er *fasta – filer*, dette er blot tekst – filer der overholder et bestemt format, *fasta - formatet*¹. Disse indeholder, udover sekvensen, en lille beskrivelse af den efterfølgende sekvens, beskrivelsen starter altid med >, f.eks.



```
1 >PX089_[sekvens1]
2 AGTCGTGCTGAT
3
4 >PX090_[sekvens2]
5 AGCCGAGCTGATA
6
7 >PX189_[sekvens3]
8 TAGGCTTGCTTAG
9
10 >PX190_[sekvens4]
11 CAGGCTTGCTTAG
```

For at MEGA kan genkende filen som en fasta – fil skal endelsen af filen være .fa, .fas eller .fasta. F.eks. hedder min fil

nukleotid.fas

ADVARSEL! MEGA brokker sig højlydt, hvis formatet ikke er overholdt eller hvis filen ikke har den korrekte endelse. Det SKAL være en ren tekstfil og endelsen SKAL være en som MEGA genkender.

¹ https://en.wikipedia.org/wiki/FASTA_format

Du kan f.eks. bruge følgende værktøjer til at redigere rene tekstfiler:

- PC: Notesblok
- MAC: TextEdit
- PC eller MAC: Visual Studio Code <https://code.visualstudio.com/> (mest til programmering, men kan altså også redigere en ren tekst - fil)

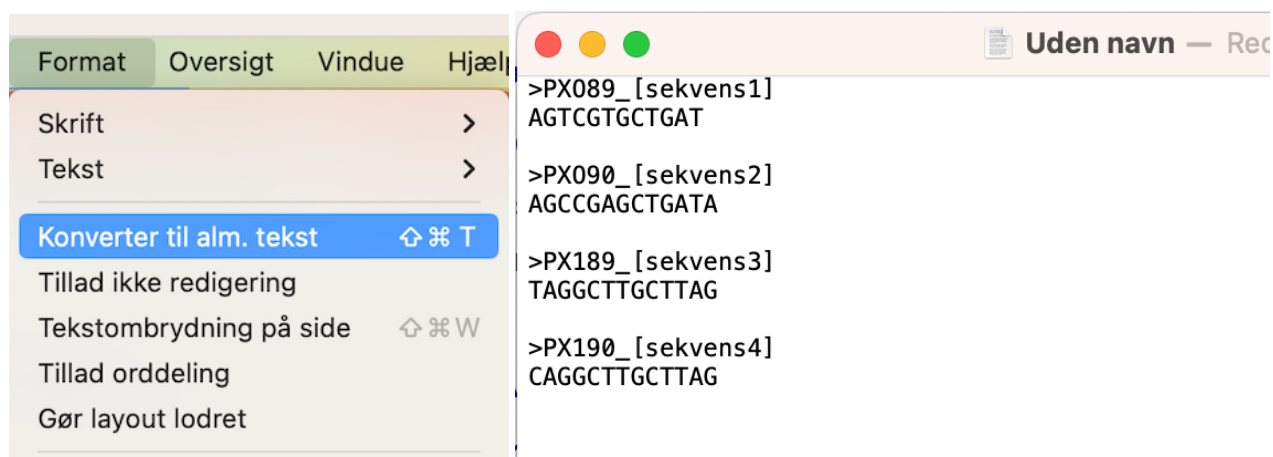
Du kan IKKE redigere rene tekstfiler med værktøjer som: Pages, Word og Excel.

Vi laver filen sammen (her i TextEdit, men samme i Notesblok):

Jeg laver et nyt dokument og taster data ind



For MAC – brugere: Jeg ændrer formatet til ren tekst (dette er ikke nødvendigt i Notesblok på PC)



Teksten skal se ud som til højre.

Gem filen

Gem som:

Mærker:

▼ Skrivebord

Q Søg

_Exerci.pdf

k >

ion >

7.s002.fas

ession.xlsx

rgsforløb 2 >

athwa.pdf

i. tekstkodning: Unicode (UTF-8)

Brug ".txt", hvis der ikke er nogen endelse.

Filnavnet skal ende på .fa, .fas eller .fasta

Når du navngiver, så må navnet ikke

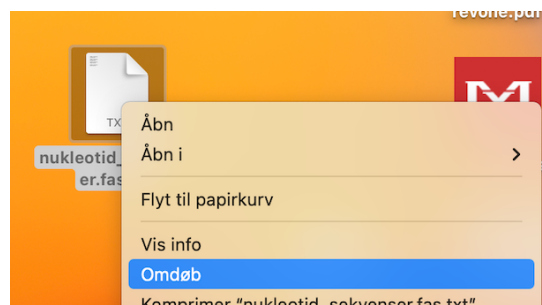
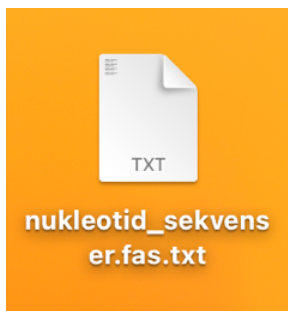
- starte med et tal eller specialtegn
- indeholde mellemrum (brug i stedet "snakecase" nukleotid_sekvenser, eller "camelcase" nukleotidSekvenser)

Bemærk, at filen ikke må ende på .txt

Annuller Gem

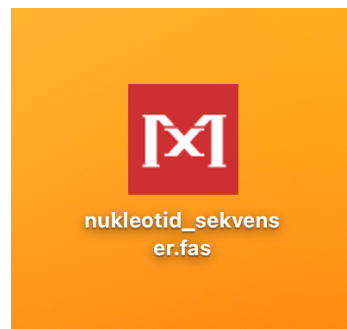
Hov jeg kom til at gemme som .txt alligevel

Ikke noget problem, du kan bare højreklikke på filen



Omdøb

MEGA kan nu genkende filen



Vi er nu klar til at bruge filen.

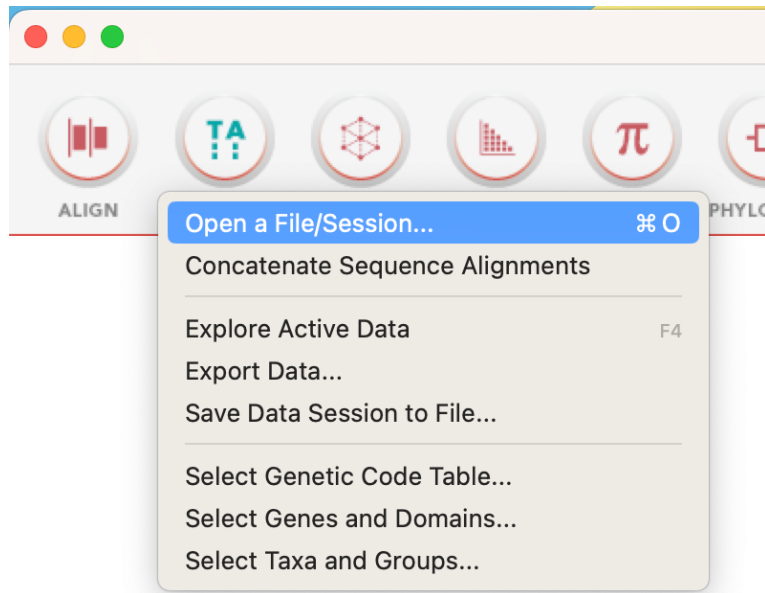
Læs mere om data og formater her

https://www.megasoftware.net/mega1_manual/DataFormat.html

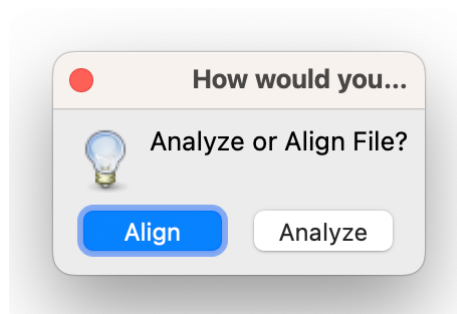
Afsnit 2. Alignment Vi vil lave multipel alignment af sekvenserne i den fasta – fil vi lige lavede.

Vi starter MEGA og vælger

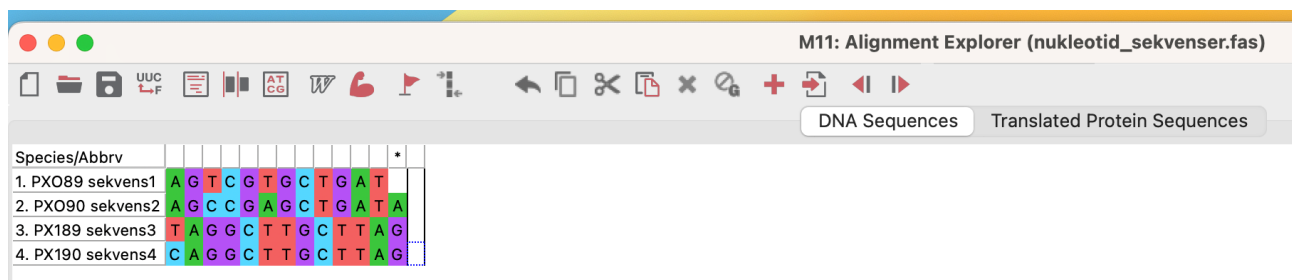
DATA > Open a File/Session



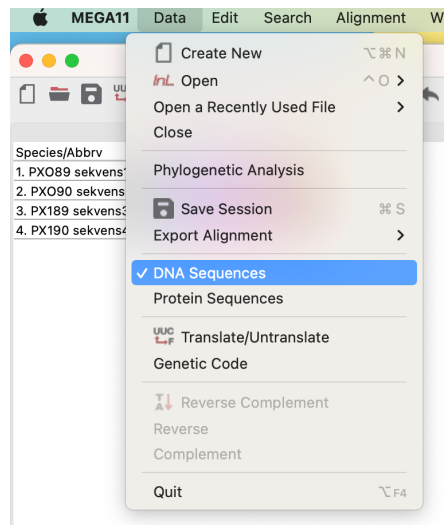
Når du har valgt filen *nukleotid_sekvenser.fas* skal du vælge om du vil lave alignment eller analysere. Vi vælger alignment



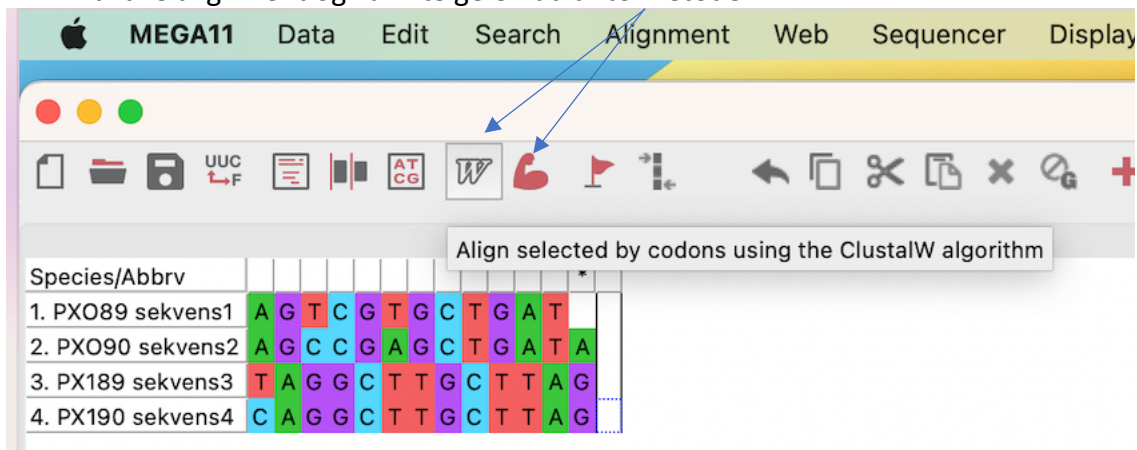
Dette giver os følgende skærbillede, hvor vi kan se sekvenserne



Vi sørger først for at MEGA bruger nukleotid – alfabetet (da vores fasta – fil bestod af DNA - sekvenser)



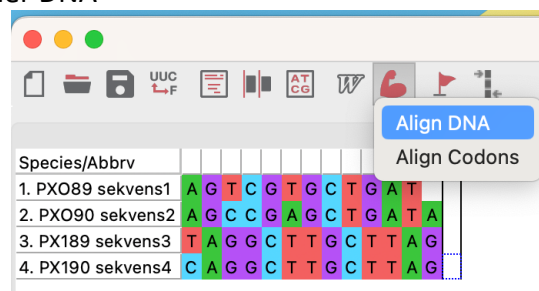
Vi vil nu lave alignment og kan vælge en ud af to metoder

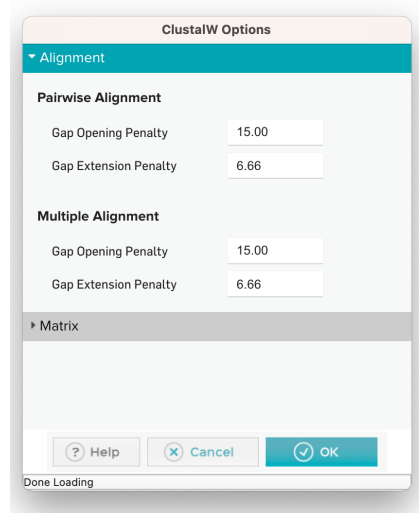


De to metoder er

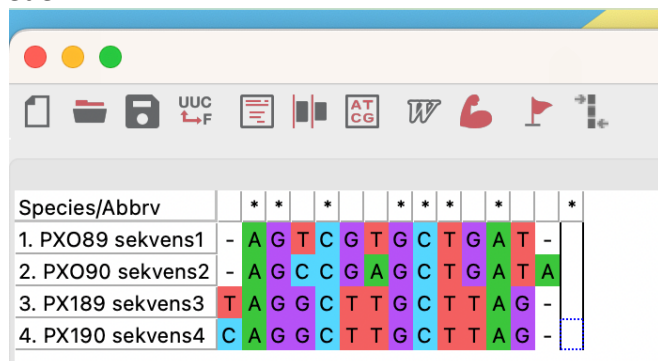
- ClustalW (<https://en.wikipedia.org/wiki/Clustal>) en ældre metode, men relativt forståelig
- MUSCLE ([https://en.wikipedia.org/wiki/MUSCLE_\(alignment_software\)](https://en.wikipedia.org/wiki/MUSCLE_(alignment_software))) en forholdsvis ny metode (2003), meget effektiv og kompliceret.

Jeg prøver ClustalW og aligner DNA



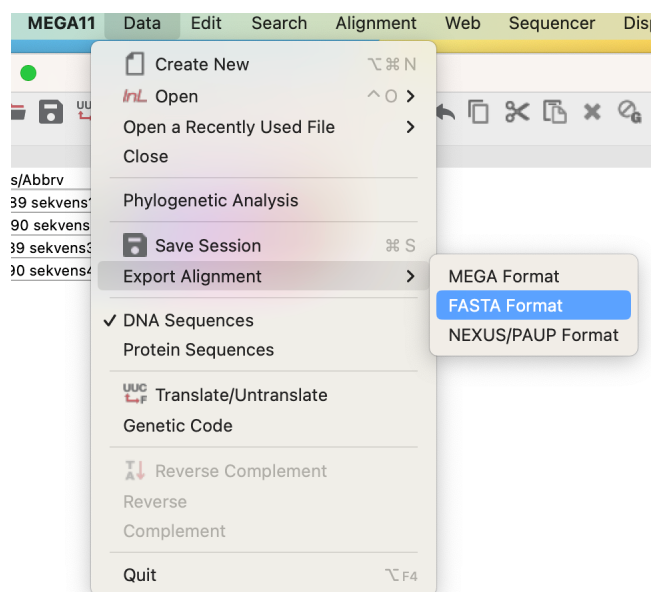


Sekvenserne er nu alignede

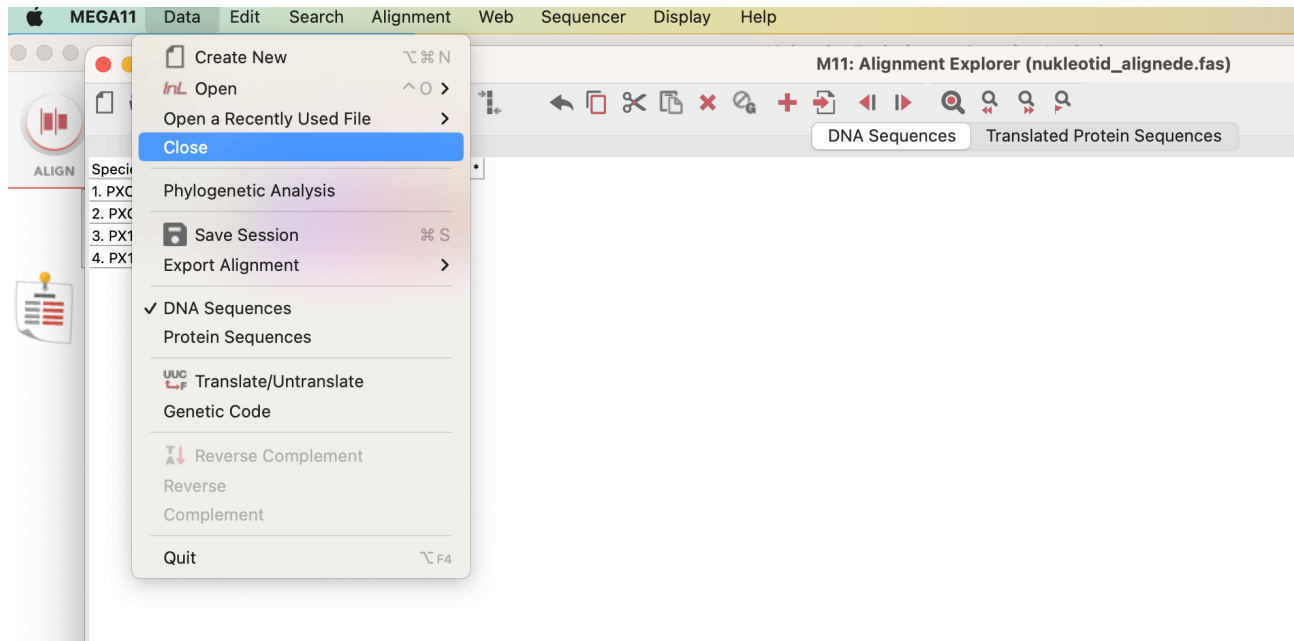


Vi gemmer filen med de alignede sekvenser i en ny fil *nukleotid_alignede.fas*

Data > ExportAlignment > Vælg format



Vi kan sagtens lukke ned for den grafiske fremstilling af sekvenserne



Dette afslutter alignment – delen. Du har nu lavet en fil med sekvenser, der er alignede med multipel alignment. Vi kan åbne filen i vores teksteditor



Bemærk, at der er indsat bindestreg – ved gaps.

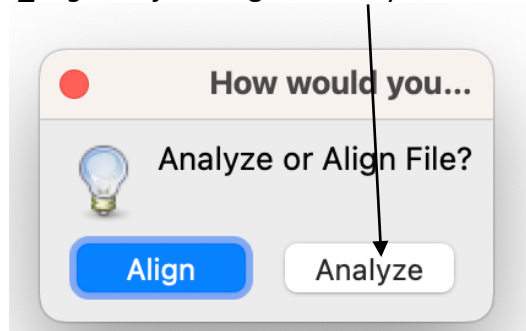
Afsnit 3. Beregn afstandsmatrix

Vi vil finde en afstandsmatrix, dette er ikke nødvendigt for fylogeni, men er jo rart at kunne

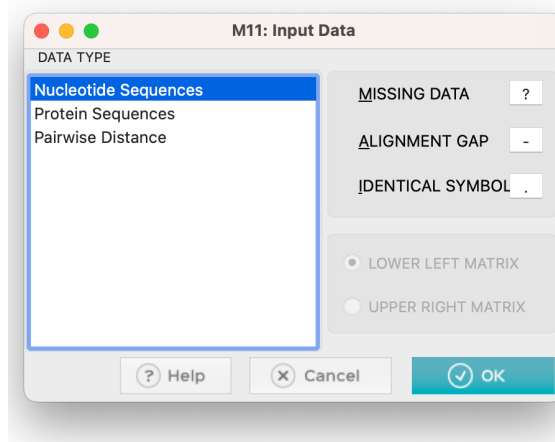
Vi starter MEGA og vælger (ligesom ovenfor)

DATA > Open a File/Session

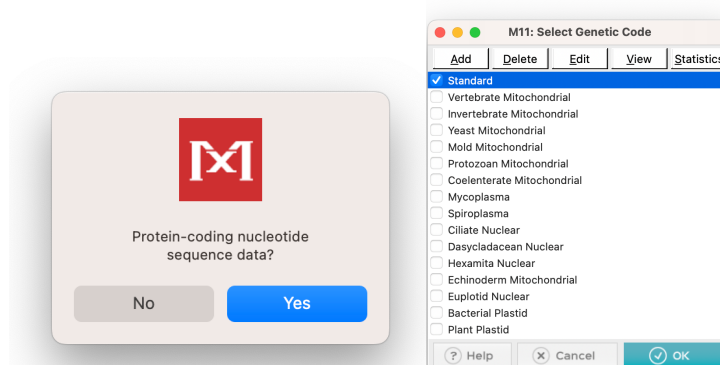
Når vi har valgt filen *nukleotid_aligned.fas* vælger vi Analyze



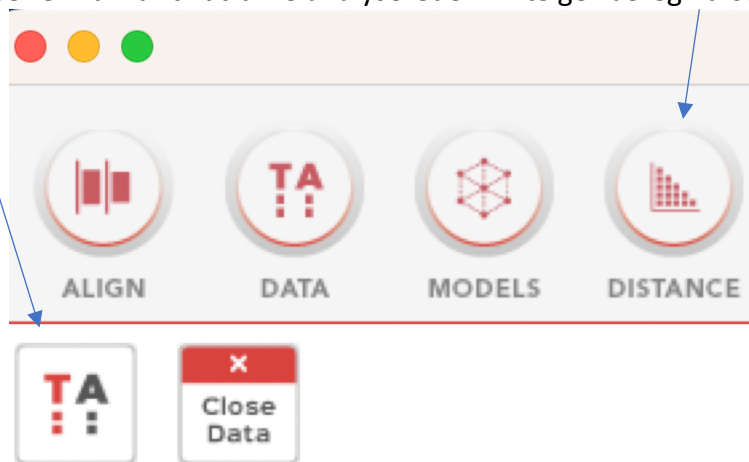
Vi holder øje med om det er en DNA – eller proteinsekvens



Vi vælger OK. Hvis vi ved at det er en proteinkodende DNA – sekvens, så kan vi endda få lov til at vælge, hvilken art sekvensen stammer fra. Jeg trykker bare OK



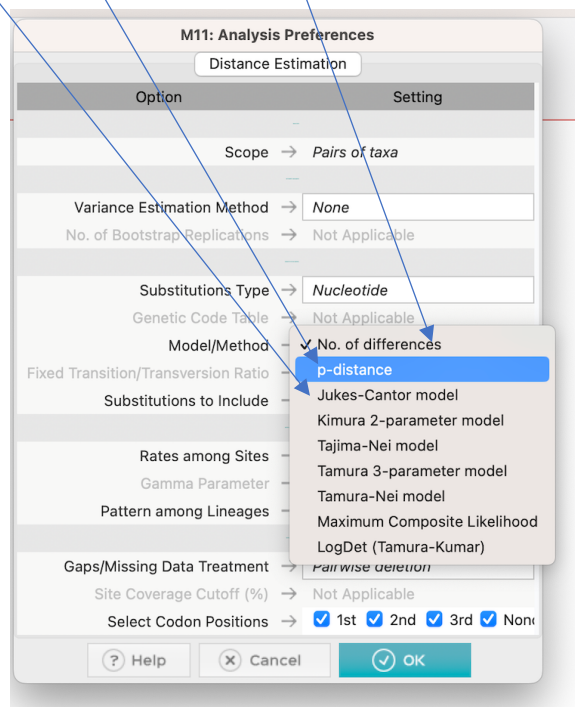
De alignede sekvenser er nu klar til at blive analyserede. Vi vælger beregn afstande



DISTANCE > Compute Pairwise Distance

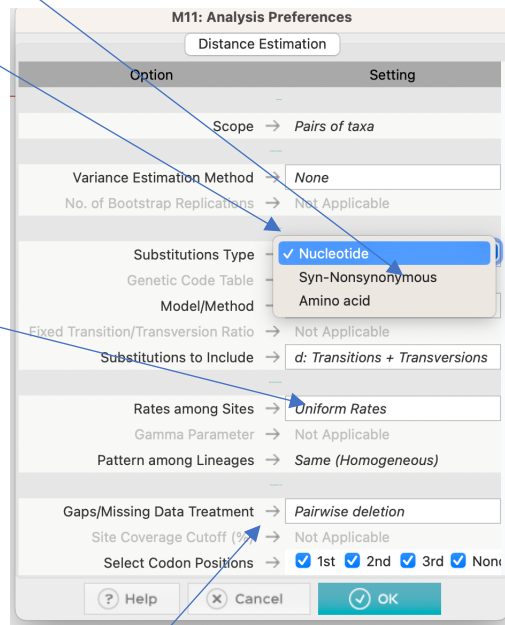
Vi har nu flere muligheder, f.eks. kan vi vælge mellem

- Antal forskelle i sekvenserne (en Hamming afstand)
- En p – afstand (procentvis afvigelse)
- En korrigeret p – afstand, hvor man bruger en bestemt model for molekylær evolution, f.eks. Jukes – Cantor modellen



Vi kan også fortælle, hvilke typer af substitutioner vi vil inkludere, jeg vælger blot alle nukleotidsubstitutioner, men det er også interessant at vi kan nøjes med at vælge synonyme substitutioner eller ikke – synonyme substitutioner (dette er nok for de mere øvede)

Her antages, at de enkelte positioner udvikler sig uafhængigt af hinanden. Dette er almindeligt.



Her er en forklaring på denne option

Pairwise-deletion option

In the pairwise-deletion option, sites containing missing data or alignment gaps are removed from the analysis as the need arises (e.g., pairwise distance computation). This is in contrast to the complete-deletion option in which all such sites are removed prior to the analysis.

Fra https://www.megasoftware.net/web_help/7/rh_pairwise_deletion_option.htm

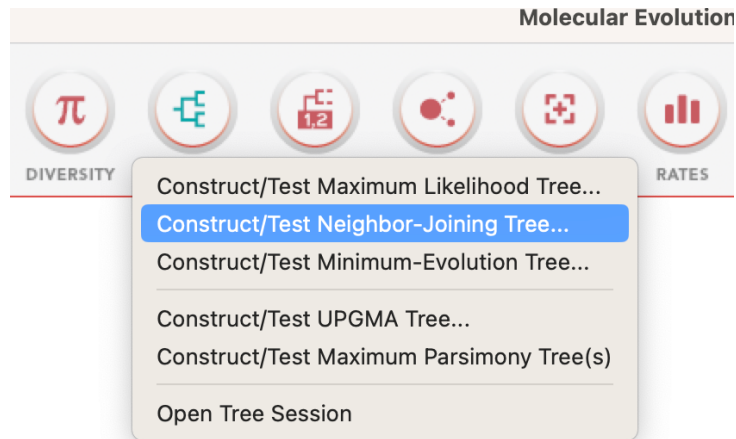
Jeg vælger blot Pairwise deletion. I alt giver dette følgende afstandsmatrix af p – afstande

	1	2	3	4
1. PX089 sekvens1 A	1.0000			
2. PX090 sekvens2 A	0.1667	1.0000		
3. PX189 sekvens3 B	0.3333	0.4167	1.0000	
4. PX190 sekvens4 B	0.3333	0.4167	0.0769	1.0000

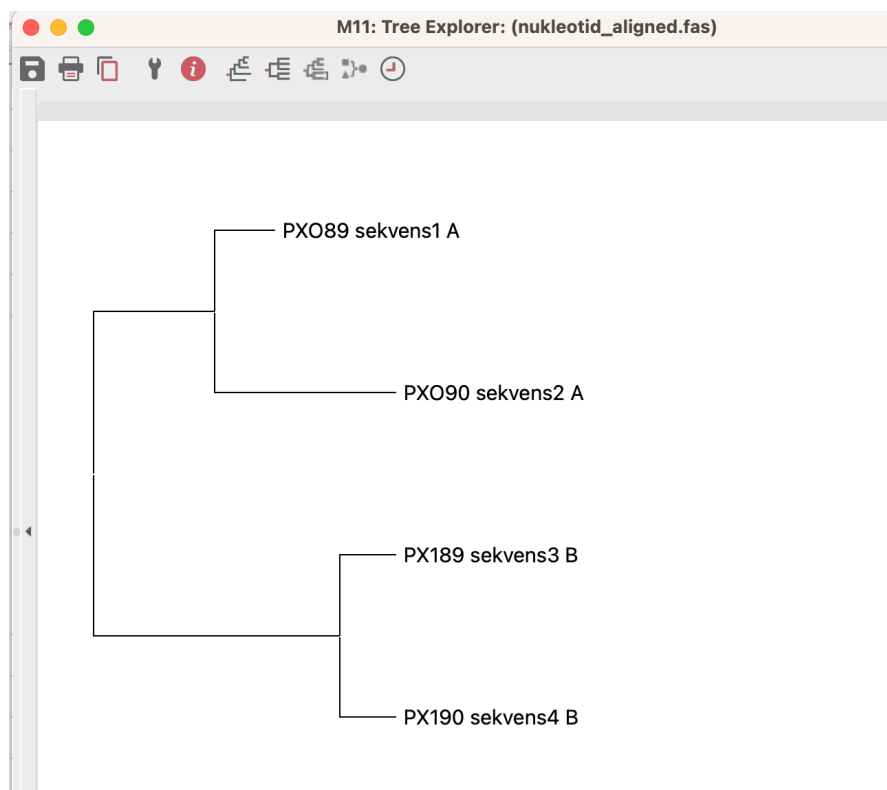
Afsnit 4. Fylogeni For at lave et fylogenetisk træ vælger vi

Phylogeny>Construct/Neighbor-Joining Tee

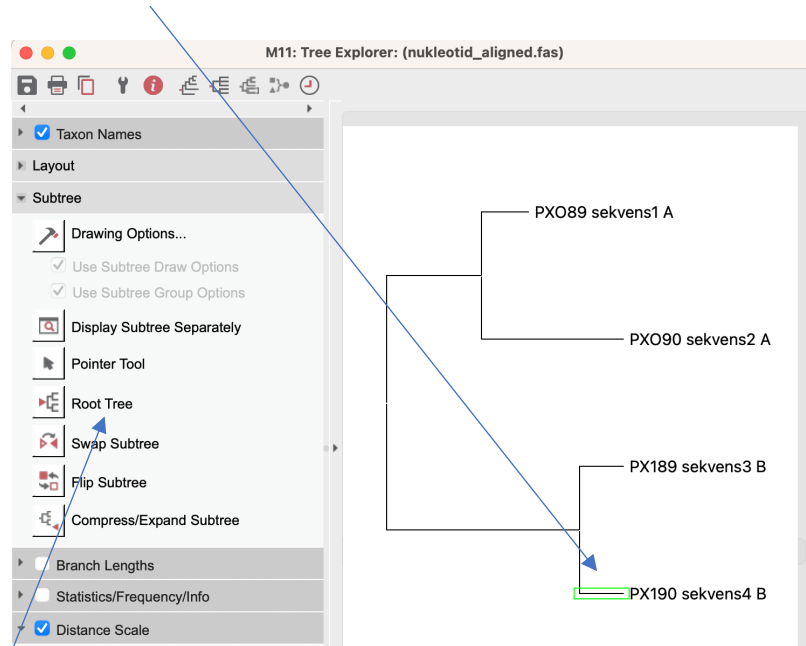
Vi kunne selvfølgelig også have valgt UPGMA Tree.



Jeg træffer de samme valg som på side 10 og får herved et træ uden rod.



Vi har til sidst mulighed for at vælge at taxa skal vælges som rod, lad os sige at vi ved at sekvens4 er outgroup. Marker grenen som vist her



Vælg derefter Root Tree, dette giver følgende træ med rod, hvor sekvens4 er outgroup.

