

# Geneious

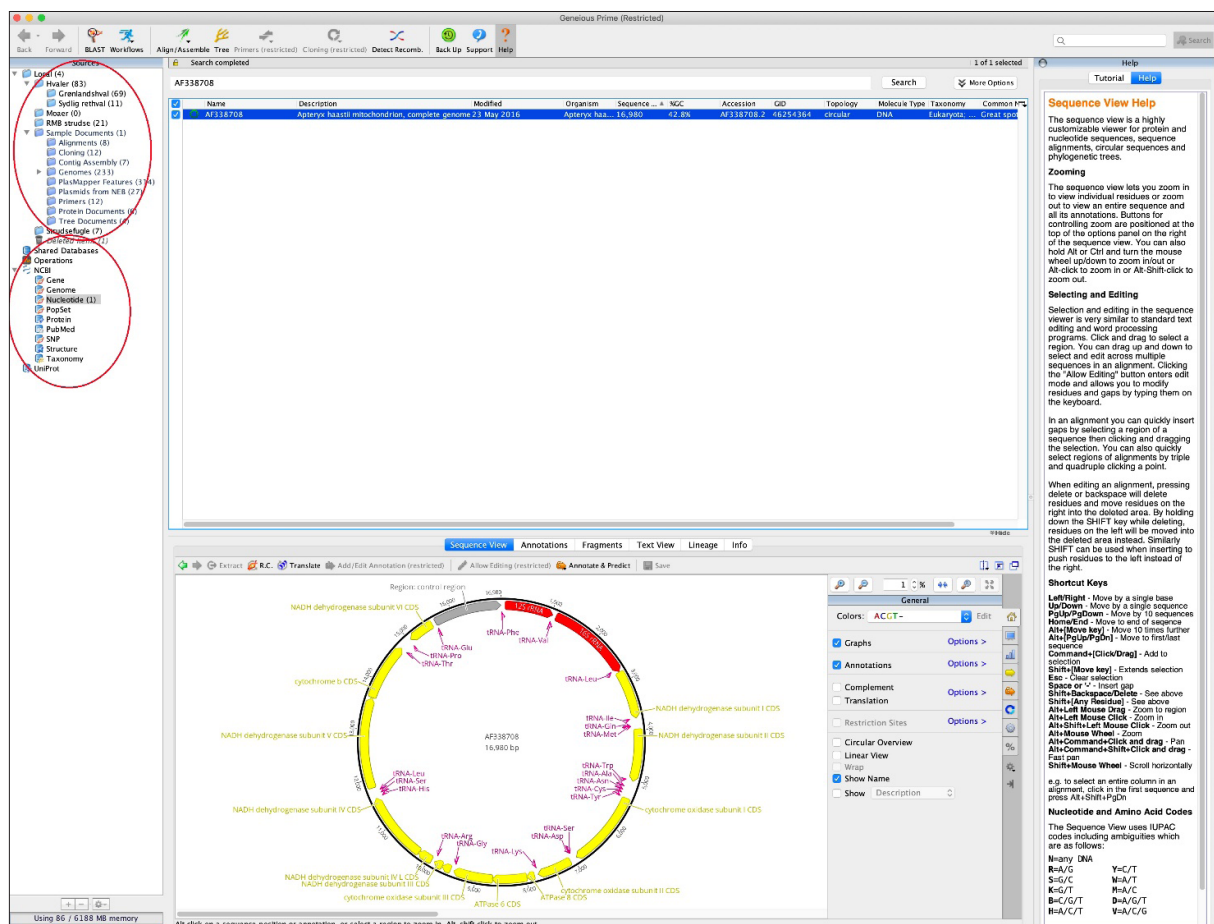
## – en manual til elevbrug

### Indhold

INTRODUKTION TIL GENEIOUS .....	2
DATABASERNE HOS NCBI .....	2
ORGANISÉR PROJEKTET .....	3
SØGNING .....	3
UDVIDET SØGNING .....	4
ARKIVÉR DE DOWNLOADEDDE SEKVENSER PÅ COMPUTEREN .....	6
RAFFINERING AF SEKVENSER .....	6
Ekstrahér gener i sekvens .....	7
Oversæt gener i sekvens til proteinsekvenser .....	7
EKSPORTÉR DATA .....	7
FREMSTILLING AF FYLOGENETISK TRÆ .....	8
Alignment .....	8
Alignmentscore .....	8
Tjek af alignment .....	9
Fremstilling af fylogenetisk træ .....	10
Ret stamtræets udseende .....	11
Danske navne i stamtræet .....	12
KORT OPSKRIFT PÅ AT LAVE FYLOGENETISKE STAMTRÆER I GENEIOUS .....	13

# Introduktion til Geneious

Når man starter Geneious, vises et vindue som vist i figur 1. Der er tre hovedpaneler. Til venstre et navigationspanel og i midten et panel med de fundne data. Til højre et hjælpepanel. Det kan være fornuftigt at lukke hjælpepanelet til højre; så bliver der mere plads til de øvrige dele af programmet. Alle paneler kan trækkes i så man kan se det, man vil, i et fornuftigt størrelsesforhold.



Figur 1. Arbejdsområdet i Geneious.

## Databaserne hos NCBI

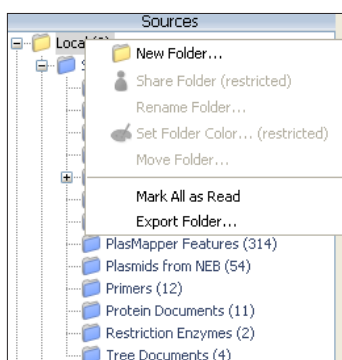
Når man skal fremstille et fylogenetisk stamtræ ud fra molekylære data, skal man have fat i de store databaser som indeholder DNA-sekvenser eller proteinsekvenser. Disse ligger på en database der hedder NCBI (det er faktisk flere databaser – og de er offentligt tilgængelige), og som det ses i venstre panel, markeret med den nederste røde ring på figur 1, kan man søge på mange forskellige former for molekylær information. Det der kan søges på, er vist i figur 2.

Søgning	Forklaring
GenE	Søgning på enkeltgener
GenomeE	Søgning på hele genomer – fx mtDNA
NucleotideE	Søgning på en sekvens af nucleotider (DNA eller RNA)
PopSet	Data der er brugt til at bestemme fylogeni for en population
Protein	Søgning på proteinsekvens (aminozyrer)
PubMed	Søgning på artikler og biblioteksinformation
SNP	Søgning efter SNP (variable mængde af gentagne nucleotider)
Structure	Struktur af fx proteiner
Taxonomy	Navnedatabase for de arter repræsenteret med molekylære data, der findes i NCBI's database

Figur 2. Tabel over de enkelte søgedatabaser hos NCBI.

## Organisér projektet

Når man begynder et projekt, skal man have oprettet en mappe til at gemme de data man kan finde i databaserne, så man senere kan behandle dem. Dette gøres ved at højreklikke på 'Locale' (se figur 3) og så klikke på 'New folder'. Der dannes så en ny mappe som kan navngives alt efter, hvad man undersøger. Det er også muligt at lave undermapper.

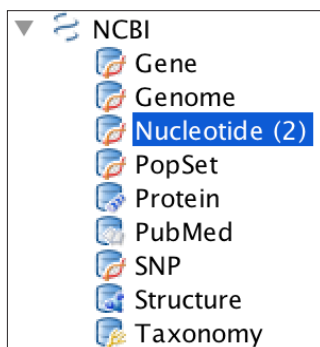


Figur 3. Oprettelse af lokal mappe på computeren. Det gøres i venstre panels øverste del.

## Søgning

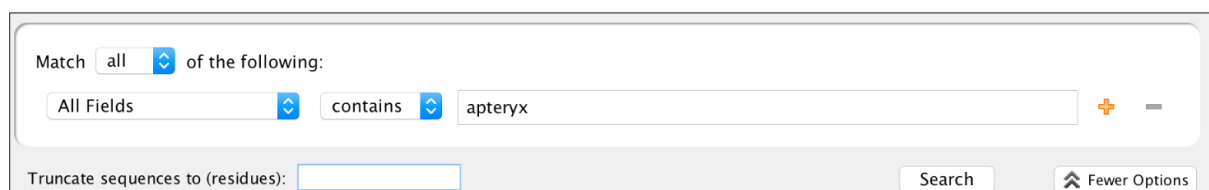
Når man søger i databaserne, er det vigtigt at finde ens data at sammenligne; det vil sige homologe sekvenser i betydningen 'sekvenser der koder for det samme eller er ens beliggende introns'. Man kan sige at hvis man skal finde genetiske forskelle, bør man når man skal undersøge nærtbeslægtede individer, fx i en population, bruge intergent DNA (introns). Når man skal undersøge fjernt beslægtede arter eller grupper, kan man fx anvende komplette mtDNA-genomer eller fx DNA der koder for 16S-underenheden af ribosomer, cytochrom B-genet osv.

Programmet er engelsk, og databaserne er organiseret på engelsk så man kan ikke søge med danske navne. Søger man på en bestemt art, er det bedst at søge på dens videnskabelige navn der består af to navnedele, som man kender det fra fx *Homo sapiens* (menneske). Man kan fx bruge Wikipedia eller 'The Tree of Life Web Project' <http://tolweb.org/tree> til at finde arters videnskabelige navne.



Figur 4. Søgning på sekvenser hos NCBI sker ved at klikke på det man søger efter – i dette tilfælde sekvenser af DNA eller RNA, og derefter søge i søgefeltet øverst i det midterste panel som vist i figur 5.

Klikker man i venstre panels nederste del på 'Nucleotide', som man se det i figur 4, bliver det øverste midterste felt til et søgefelt som ser ud som i figur 5.



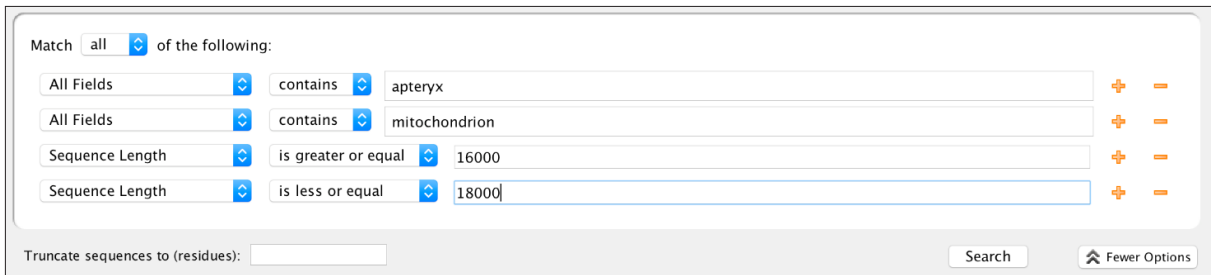
Figur 5. Sæmpel søgning på NCBI via Geneious.

## Udvidet søgning

Man kan udvide søgningen ved at klikke på det orange '+' til højre for søgefeltet, og man ændre hvad man vil søge efter, ved at ændre i boksen hvor der står 'All Fields'. Et eksempel på dette er vist i figur 6 hvor der søges efter slægten af kiwifugle (*Apteryx*). Der søges efter 'mitochondrion', svarende til mtDNA, og der søges efter sekvenslængder på mellem 16.000 og 18.000 basepar, svarende til hele mtDNA-genomer.

Søgningen dækker alle de indtastede karakteristika så længe der står 'Match all of the following:'. Rent computerteknisk betyder det at man søger med et såkaldt 'boolsk OG', hvilket betyder 'både og'. Man kan ændre 'all' til 'any', og så søger man med et såkaldt 'boolsk ELLER' hvilket betyder, at bare én af søgefelterne skal være sande for at få et hit.

Man skal stave korrekt når man søger. Der er ikke som hos fx Google, en funktion der spørger, om man mener noget andet, end det man har skrevet. Får man ikke noget hit i databasen, er der to mulige grunde. Enten findes sekvensen ikke; ellers har man stavet forkert. Dobbelttjek derfor altid stavningen.



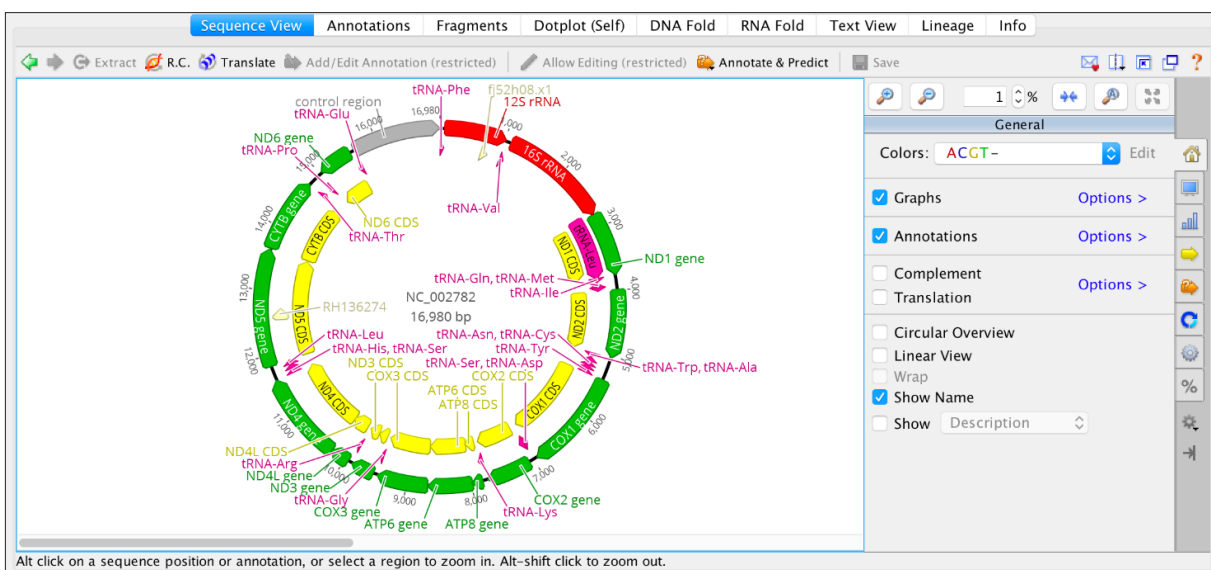
Figur 6. Udvidet søgning i Geneious.

Man kan derefter se på resultaterne af søgningen i det midterste panel i feltet nedenunder søgningen, som det ses i figur 7.

Name	Description	Modified	Organism	Sequence ...	%GC	Accession	GID
<input type="checkbox"/>	LK054805	Apteryx australis mantelli genome assembly, s... 15 Jun 2016	Apteryx aust...	16,674	42.9%	LK054805.1	840008365
<input type="checkbox"/>	KU695537	Apteryx australis mantelli mitochondrion, comp... 09 Mar 2016	Apteryx aust...	16,694	42.8%	KU695537.1	1002821075
<input type="checkbox"/>	AF338708	Apteryx haastii mitochondrion, complete genome 23 May 2016	Apteryx haa...	16,980	42.8%	AF338708.2	46254364
<input checked="" type="checkbox"/>	NC_002782	Apteryx haastii mitochondrion, complete genome 01 Feb 2010	Apteryx haa...	16,980	42.8%	NC_002782.2	46255316
<input type="checkbox"/>	GU071052	Apteryx owenii mitochondrion, complete genome 24 Jul 2016	Apteryx owenii	17,020	42.8%	GU071052.1	262179844
<input type="checkbox"/>	NC_013806	Apteryx owenii mitochondrion, complete genome 24 Feb 2010	Apteryx owenii	17,020	42.8%	NC_013806.1	288900678

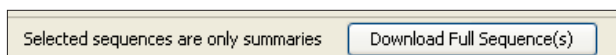
Figur 7. Søgeresultatet af søgningen i figur 6. Sekvensen NC\_002782 er markeret og vises derfor i midterste panels nederste del. Man bør tjekke at de sekvenser man finder, ligner hinanden.

Når man markerer en sekvens i resultatfeltet, som vist i figur 7, vises sekvensen i den nederste del af det midterste felt som illustreret i figur 8. Man kan klikke sig gennem alle de fundne sekvenser vist i figur 7 og man vil kunne se hvordan de ser ud nedenfor, svarende til at der vises en ny sekvens i vinduet vist i figur 8.

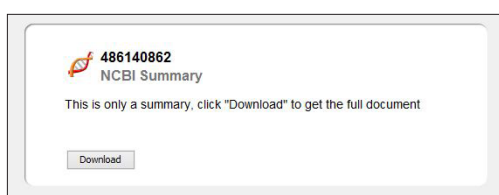


Figur 8. Den fundne sekvens vist med gener. Det er et stykke cirkulært mtDNA som også var det, søgningen i figur 6 ledte efter.

Er der ikke en tegning eller en sekvens i det nederste midterste panel, når man markerer en sekvens som vist i figur 7, vises en mulighed for at klikke på en knap for at downloade den pågældende fil. På figur henholdsvis 9 og 10 vises de to meddelelser Geneious kan give, hvis man skal aktivere et download af sekvensen ved klik på enten 'Download Full Sequence(s)' eller bare 'Download'.



Figur 9. Får man rigtigt mange hits på sin søgning, eller er sekvenserne meget store, downloades ikke hele sekvensen, men kun en beskrivelse. Passer beskrivelsen, kan man downloade sekvensen ved at klikke på 'Download Full Sequence(s)' eller 'Download' i figur 10.



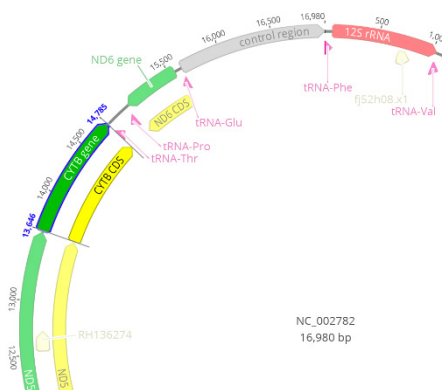
Figur 10. En anden måde hvorpå Geneious kræver et ekstra klik på 'Download' for at få lagret sekvensen lokalt så man kan arbejde med den.

## Arkivér de downloadede sekvenser på computeren

Når filen er hentet ned, skal den flyttes fra søgningen over i den mappe på den lokale computer man har lavet til sit projekt. Det gøres ved at trække den over i projektmappen i det venstre panel og slippe (drag and drop).

## Raffinering af sekvenser

På figur 8 ses et mitochondriegenom for en Kiwi. Proteindannende gener er markeret med grønne pile. RNA-dannende gener er markeret med røde pile, og grå områder er ikke-kodende DNA. Som vist på figur 11, kan man udvælge et gen ved at klikke på det hvorved de øvrige gener vises med svagere farve end det valgte gen.



Figur 11. Fra et mitochondriegenom er Cytochom-b-genet valgt ud.

### Ekstrahér gener i sekvens

Sekvensen for det valgte gen kan nu ekstraheres fra mitochondriegenomet ved klik på 'Extract' i panelet over sekvensen, og man kan herefter lave stamtræer ud fra kun cytochom-b-genet. Panelet er forstørret i figur 12, og man kan tydeligt se ikonet man skal klikke på for at ekstrahere sekvensdata fra en del af den fundne sekvens. Det kan være anvendeligt fordi man fx ikke kan finde fuldstændige mitochondriegenomer for alle arter. Hvis man gemmer direkte fra en søgning på NCBI, bliver man spurgt om hvor den ekstraherede sekvens skal gemmes. Ekstraheres en sekvens fra en gemt oprindelig sekvens, gemmes den ekstraherede sekvens automatisk i den mappe den oprindelige sekvens ligger i.

### Oversæt gener i sekvens til proteinsekvenser

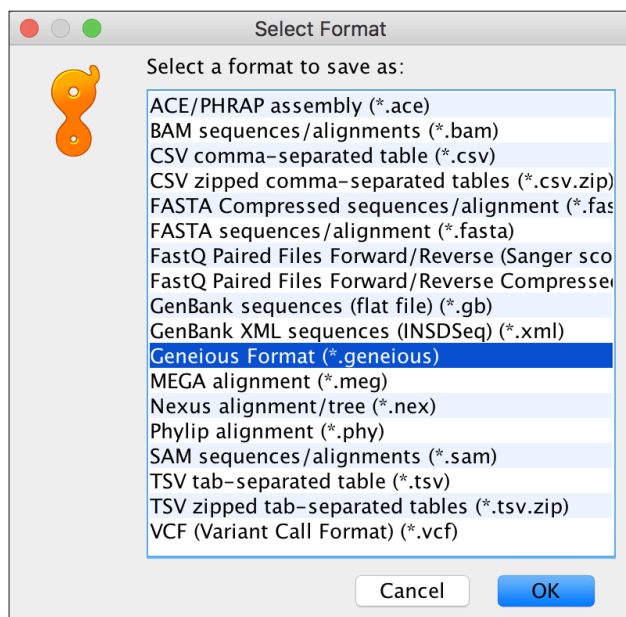
Et andet ikon der er vist på figur 12, er 'Translate'. På samme måde som man ekstraherer gener, kan man oversætte gener til proteinsekvenser. Man markerer genet som vist på figur 11 og klikker på 'Translate'. Dette er specielt anvendeligt hvis man ønsker at undersøge fylogeni mellem meget fjernt beslægtede arter.



Figur 12. Funktionerne 'Extract' og 'Translate' kan henholdsvis bruges til at arbejde med dele af sekvenser eller at arbejde med proteinsekvenser som oversættes fra DNA-sekvenser.

### Eksportér data

Ønsker man at anvende andre programmer, fx MEGA, til at analysere sine sekvensdata, skal man eksportere sine data fra Geneious og ud til et format som kan læses af andre programmer. Det gøres ved at klikke på det man ønsker at eksportere og derefter i menulinjen at klikke File → Export → Selected Documents. Herved fremkommer et vindue, som vist i figur 13, hvor man kan vælge mange forskellige eksportformater. \*.fasta-formatet er gennemgået i 'Regn med biologi', og det kan bruges af alle sekvensdataanalyseprogrammer. Det er derfor naturligt at eksportere sekvenser som '\*.fasta'-data. Har man allerede lavet et alignment, kan det med fordel eksporteres til et '\*.meg'-format (MEGA), hvis man vil analysere videre i MEGA. Enkeltsekvenser kan ikke eksporteres som '\*.meg'-format. Når man eksporterer en sekvens eller et alignment, giver programmet en advarsel om at man mister data ved eksport til et andet format end Geneious-formatet. De data man mister, er ikke sekvensdata så man kan roligt eksportere, uden at man skal frygte, at sekvenserne ikke længere er fuldstændige.

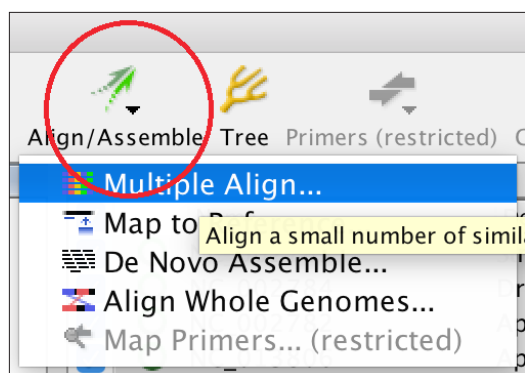


Figur 13. Eksportmuligheder fra Geneious. Figuren viser de eksportmuligheder man har for alignments i gratisudgaven af Geneious.

## Fremstilling af fylogenetisk træ

### Alignment

Når man har samlet de sekvenser man gerne vil fremstille et stamtræ over, i en mappe under 'Locale', markeres sekvenserne for de pågældende arter. Er det alle filerne i mappen, klikkes enten på <ctrl>+A (pc) eller <cmd>+A (Mac); er det ikke alle sekvenser i mappen, bruges museklik samtidig med at man holder <ctrl> nede (pc) eller <cmd> nede (Mac). Når de ønskede sekvenser er markerede, kan man klikke på alignmentknappen i menulinjen som vist i figur 14 (vælg 'multiple align').

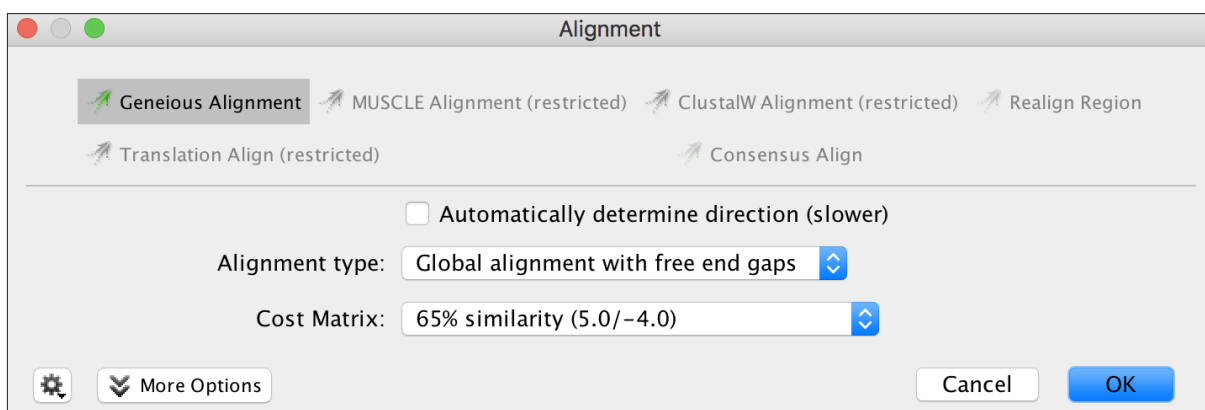


Figur 14. Fremstilling af alignment i Geneious.

### Alignmentscore

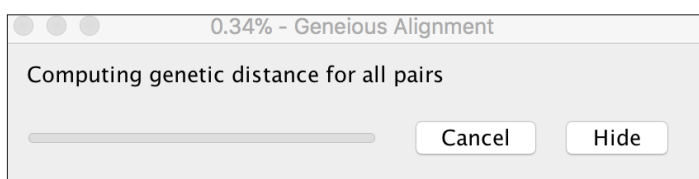
I gratisversionen er der kun én type alignment man kan vælge, nemlig 'Geneious Alignment', se figur 15. Man kan under menuboksen 'Alignment' ændre hvordan alignmentscores udregnes, men som oftest vil det ikke være nødvendigt. Geneious Alignment er en langsom proces.





Figur 15. Man kan ændre måden hvorpå man udregner alignmentscore under 'more options'.

Programmet fremstiller alignmentet – og det tager et stykke tid. Når det er omfattende data bestående af mange lange sekvenser, kan det virkeligt tage lang tid. Undervejs vises en markør som vist i figur 16, der viser hvor langt alignmentprocessen er forløbet. Det er dog sådan at de første 20 % tager lang tid, så går det lidt hurtigere frem til 30 % og derefter slutter processen lynhurtigt.



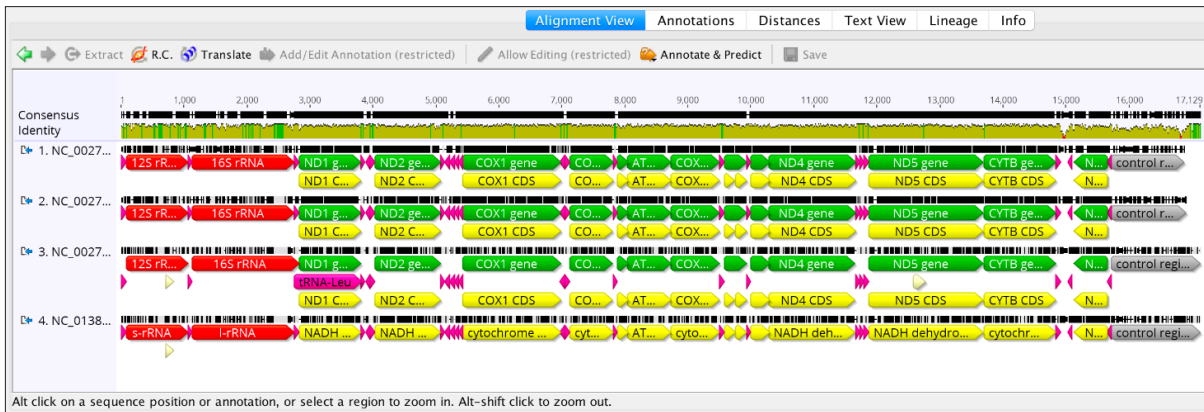
Figur 16. Informationsboks der viser fremgangen i dannelsen af alignmentet. Procentsatsen vises øverst i informationsboksen. De første 30 % tager lang tid, de sidste 70 % går lynhurtigt så regn med at alignmentet er færdigt ved 30 %.

### Tjek af alignment

Ved klik på alignment ('Nucleotide alignment') i det midterste panels øverste del (figur 17), kan man se alignmentet i det midterste panels nederste del (figur 18) hvor man kan konstatere, at de homologe gener står under hinanden (de er alignet).

	Name	Description	Modified	Organism	Sequence
<input checked="" type="checkbox"/>	NC_002785	Struthio camelus mitochondrion, complete geno...	01 Feb 2010	Struthio cam...	16,595
<input checked="" type="checkbox"/>	NC_002784	Dromaius novaehollandiae mitochondrion, com...	01 Feb 2010	Dromaius no...	16,711
<input checked="" type="checkbox"/>	NC_002782	Apteryx haastii mitochondrion, complete genome	01 Feb 2010	Apteryx haa...	16,980
<input checked="" type="checkbox"/>	NC_013806	Apteryx owenii mitochondrion, complete genome	24 Feb 2010	Apteryx owenii	17,020
<input checked="" type="checkbox"/>	Nucleotide alignment	Alignment of 4 sequences: NC_002785, NC_0...	03 Apr 2018 7:21 AM	-	17,129

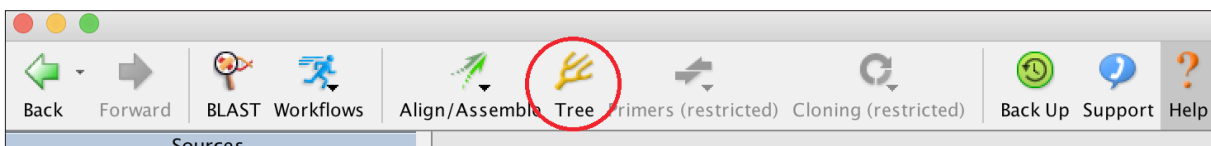
Figur 17. Klik på alignment for at undersøge om det ser rigtigt ud.



Figur 18. Alignment af mtDNA. Man kan se at alle sekvenser indeholder de samme gener. I den øvre del kan der ses en Identity-graf som viser, hvor stor overensstemmelse der er mellem de fire sekvenser. Det er tydeligt at der er mest overensstemmelse ved de grønne proteinkodende gener, mens de ikke kodende områder (grå) udviser mindst sammenfald.

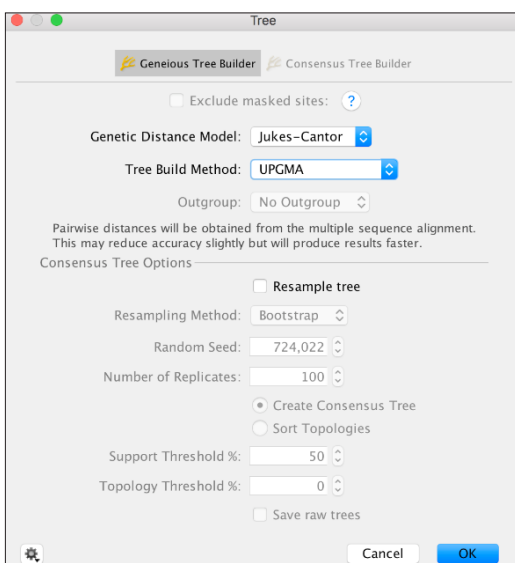
### Fremstilling af fylogenetisk træ

Herefter skal der laves et fylogenetisk stamtræ. Klik på 'Tree' i menuen som vist på figur 19.



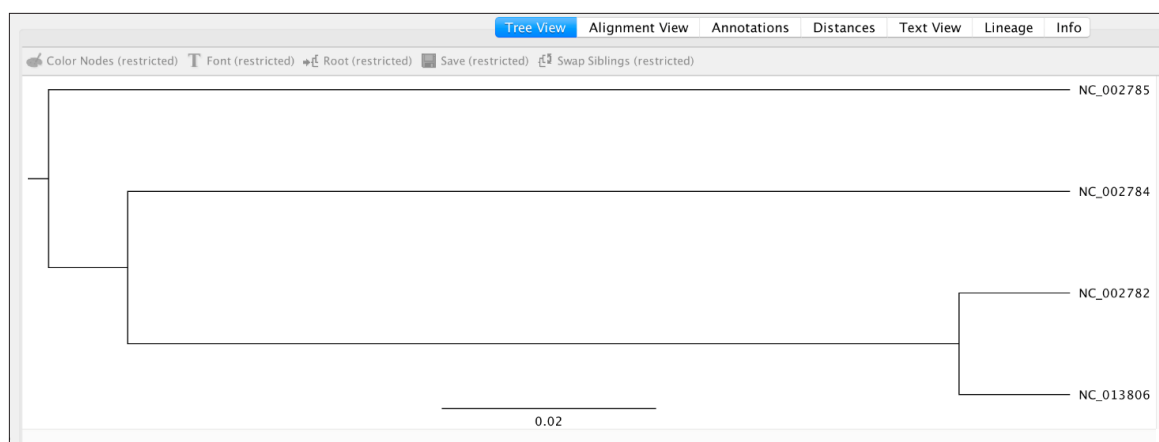
Figur 19. Menuen hvor man kommer ind i den del af programmet, der gør, at man kan fremstille et fylogenetisk træ.

Man skal nu vælge hvilken afstandsmodel der skal bruges, og hvilken trætype der skal bygges. Normalt er det fint at vælge 'Genetic Distance Model' som en 'Jukes-Cantor' og sætte 'Tree build method' til 'UPGMA', som vist i figur 20. Outgroup kan vælges, men normalt behøver man det ikke.



Figur 20. Parametre der kan sættes ved konstruktion af et fylogenetisk træ.

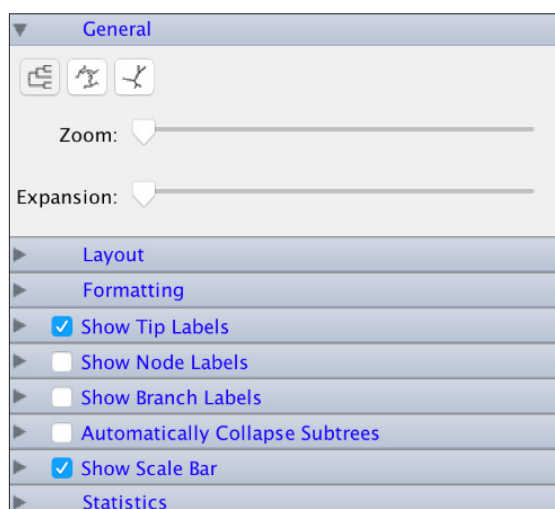
Ved klik på 'OK' får man så vist et stamtræ, der ser ud som vist i figur 21, og som man kan redigere udseendet af.



Figur 21. Stamtræet før den yderligere redigering.

### Ret stamtræets udseende

I arbejdsrummet panelet til højre for stamtræet kan man redigere udseendet af stamtræet. Værktøjerne fremgår af figur 22.

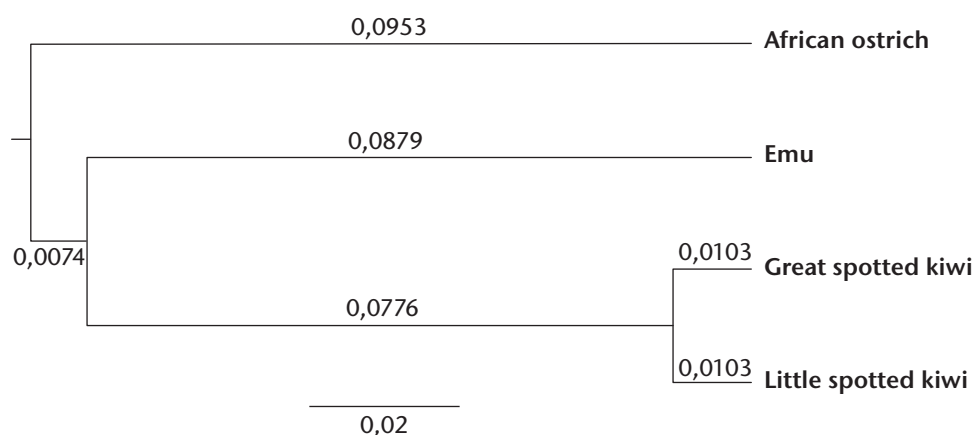


Figur 22. Værktøjer til redigering af stamtræets layout.

Man klikker på trekanten ud for den del man vil redigere; der er mange muligheder – prøv selv. Følgende bør gøres:

- ✓ Sæt almindelige navne på stamtræet: 'Show Tip Labels' → 'Display' → 'Common name'.
- ✓ Sæt flueben i 'Show Scale Bar', så man kan regne genetisk afstand ud.

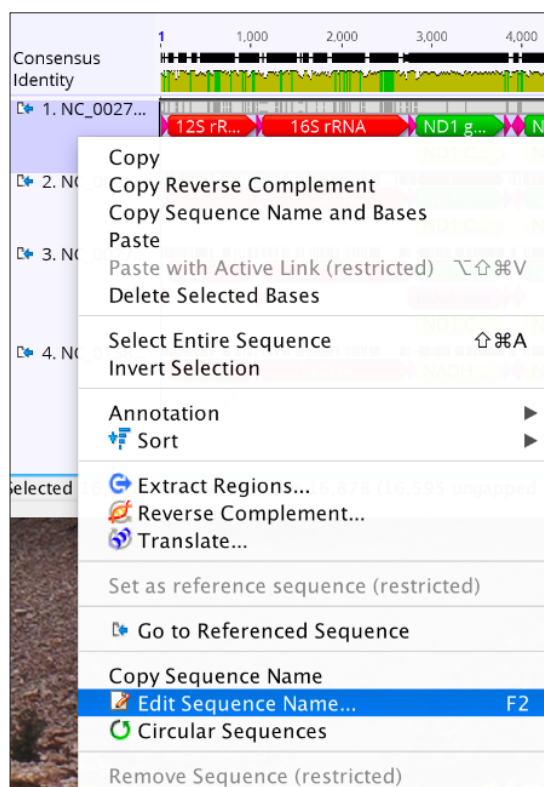
Når stamtræet er færdigt, kan man gemme det som et billede, man kan sætte ind i en rapport. Vælg 'File' → 'Save as image file'. Figur 23 viser hvorledes det endelige stamtræ kan se ud.



Figur 23. Stamtræ over fire strudsefugle.

### Danske navne i stamtræet

Vil man gerne have danske navne på stamtræet, er der to muligheder: Man kan i 'Paint' eller et andet billedbehandlingsprogram skrive de danske navne og overskrive/slette de engelske. En anden mulighed er vist i figur 24 og går ud på at man i alignmentet højreklikker på den enkelte sekvenser og omdøber angivelsen ved sekvenserne til danske navne i stedet. Herefter dannes et træ med de korrekte navne. Når man redigerer stamtræet efter at have omdøbt koderne for sekvenserne, skal man vælge 'Show Tip Labels' → 'Names'.



Figur 24. Omdøbning af sekvensnumre til danske navne.

## Kort opskrift på at lave fylogenetiske stamtræer i Geneious

- 1) Fremstil en lokal mappe til at gemme de molekylærbioologiske data i.
- 2) Søg efter fx mtDNA i nucleotide-databasen i NCBI eller efter et gen i gendatabasen. Det gælder om at finde nogle molekylærbioologiske data, der er homologe for de arter, man ønsker at undersøge, fx samme gen, samme slags protein eller et helt mtDNA.
- 3) Når brugbart materiale er fundet, trækkes det over i den lokale mappe der blev oprettet i punkt 1. Det kan være det først skal downloades, hvilket vil fremgå af det nederste panel i midten.
- 4) Når man har fundet de data, man skal bruge, markeres de, og der vælges 'alignment' (multipelt alignment). Der går et stykke tid.
- 5) Når alignment er færdigt, markers dette, og der trykkes i menuen øverst på 'Tree' (nemtest forståeligt er herunder at vælge 'Jukes-Cantor')
- 6) Gem træet som billedfil: File → Save as image.