

Sanger sekventering - fra mikroorganisme til DNA sekvens



Mikroorganismer, såsom bakterier og encellede eukaryote organismer er tit svære at identificere ud fra deres form, da mange er så små at det kan være svært at se forskel på dem i et mikroskop. En anden måde at identificere dem på, er ved at kortlægge dele af deres DNA, der er unikt for den enkelte organisme.

På alle sygehuse er der i dag molekylærbiologiske laboratorier, hvor man identificerer sygdomsfremkaldende bakterier ud fra deres DNA. Udstyr til DNA sekventering er desværre meget dyrt, derfor er det sjældent at finde det i et gymnasie laboratorium. Derfor bliver laboratorie delen af denne øvelse virtuel, og laves på HHMI's bacterial ID lab.

Aktivitet 1. Virtuel DNA Sanger sekventering af sygdomsfremkaldende bakterier.

I første del af denne øvelse skal I sekventere DNA fra bakterier fra 6 patientprøver ved at lave en kombination af PCR opformering af DNA og Sanger DNA sekventering.

Gå ind på følgende hjemmeside: <https://www.biointeractive.org/classroom-resources/bacterial-identification-virtual-lab>

Laboratorie øvelsen består af 6 forskellige steps. 1. Oprensning af DNA, 2. PCR opformering, 3. Oprensning af PCR produkt, 4. PCR Cycle sequencing, 5. DNA sekventering og 6. Blast søgning i NCBI sekvens databasen. I skal nu lave step 1-5. For hver del er der en række spørgsmål, som I skal svare på, og en ordliste, som hjælper jer med at forstå hvad der sker i laboratoriet, da det er på engelsk.

Oprensning af DNA.

1. Hvorfor er det nødvendigt at oprense DNA fra bakteriekolonierne?
2. Hvordan opnår man bakterie lysis i laboratoriet? (lysis=når bakterier sprænges)
3. Hvorfor skal prøven opvarmes til 100°C efter bakterie lysis?
4. Efter centrifugering af prøverne, hvor befinder DNA'et sig så i prøven, i pellet (bundfaldet) eller i supernatanten (væsken).

Ordforklaringer: Culture dish=petri skål, pathologist=læge med speciale i diagnose af sygdomme, proteolytic=ting der medfører proteolyse, dvs nedbrydning af proteiner.

PCR opformering.



5. Hvad består et master mix af?
6. Forklar hvad en universel primer er og hvorfor det er smart at bruge universelle primere, når man ikke ved, hvilken bakterie man har dyrket.
7. Hvorfor er det vigtigt at have både en positiv og en negativ kontrol når man laver PCR opformering?

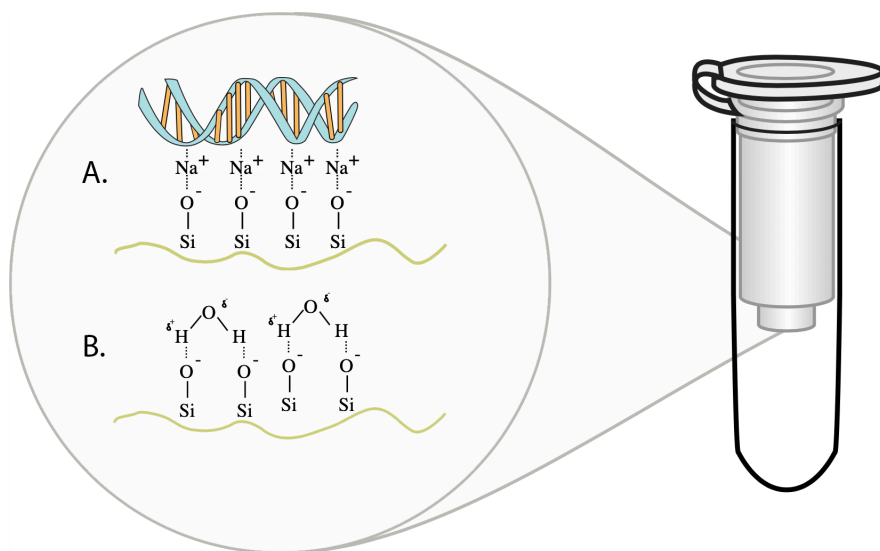
Det er 16s SSU rRNA genet der opformeres ved PCR reaktionen. Genet koder for det RNA, som den lille subunit af ribosomet i bakterier er opbygget af og er semikonservativt, dvs at der er områder, hvor der ikke er sket mutationer i DNA'et og hvor der næsten ikke er variation mellem arter, men også områder hvor der er sket mutationer og hvor der derfor er stor variation imellem arter.

8. Forklar hvorfor det er vigtigt at genet er semikonservativt, for at man kan bruge det til artsidentifikation vha PCR opformering med universelle primere.
9. PCR maskinen kører 30 cykler med 3 forskellige temperaturer. Forklar hvad der sker med DNA'et ved hver temperatur og hvad man kalder de 3 steps.

Ordforklaringer: Mastermix=samlet blanding af reagenser til PCR, der efterfølgende pipetteres ned i de enkelte PCR rør, deionized=vand der har fået fjernet ioner, svarer i grove træk til demineraliseret vand, vial=lille beholder,

Oprensning af PCR produkt.

10. Hvorfor er det smart først at lave en gelelektrofese af PCR produktet, den positive og den negative kontrol.
11. Hvorfor er det nødvendigt at oprense PCR produktet?



1

I det virtuelle laboratorie oprenses og koncentreres PCR produktet (det opformerede DNA) ved at bruges en micro-søjleoprensning. Søjlemateriale er lavet af silica materiale, der har negativt ladede oxygen atomer bundet til sig. PCR produktet tilsættes til kolonnen sammen med en bindings buffer, der består af en basisk saltopløsning, og trækkes igennem kolonnen ved at spinne den i en centrifuge (A). Herefter følger to vaskesteps med vaske buffer, der også består af en basisk saltopløsning, og som også trækkes igennem kolonnen ved centrifugering. Endelig frigøres PCR produktet ved at tilsætte en eluerings buffer, som primært består af rent vand (B). Ved at bruge relativt lidt elueringsbuffer opnås en høj koncentration af PCR produkt i elueringsbufferen.

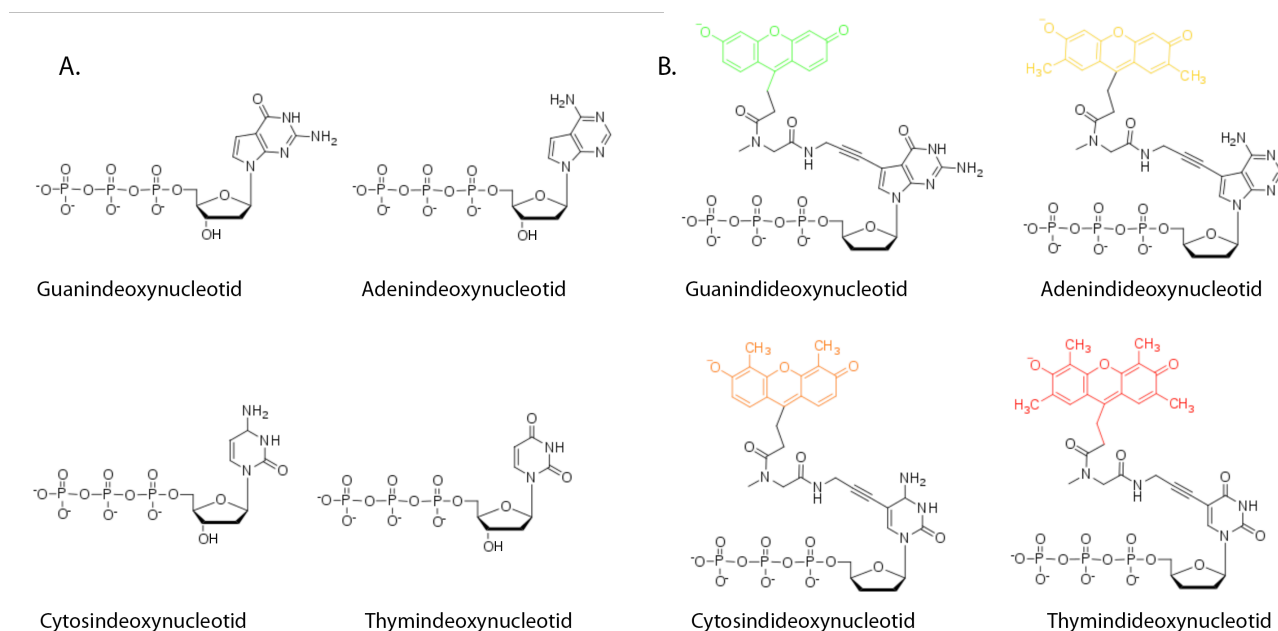
12. Forklar ud fra figuren ovenfor, hvad der sker når bindingsbufferen tilsættes (A), og hvorfor det er vigtigt at bufferen både indeholder meget salt (NaCl) og har en høj pH, dvs lavt indhold af H^+ ioner.
13. Forklar igen ud fra figuren ovenfor, hvad der sker når elueringsbufferen tilsættes (B), og hvorfor det er vigtigt at vand er et polært molekyle.
14. Overvej, hvorfor det er smart at vende filteret på hovedet inden at elueringsbufferen tilsættes?

Ordforklaringer: Prudent=klogt, lanes=baner, purifying=oprensning, microconcentrator=søjle, der kan opkoncentrere og oprense DNA, inverted=vendt på hovedet

¹ Modificeret efter https://commons.wikimedia.org/wiki/File:Qiagen_Mini_Spin_Column.svg

Cycle sequencing PCR

15. Hvad er formålet med at lave en cycle sequencing reaction?
16. Overvej hvorfor der kun tilsættes 1 primer per PCR cycle sequencing reaktion?



2

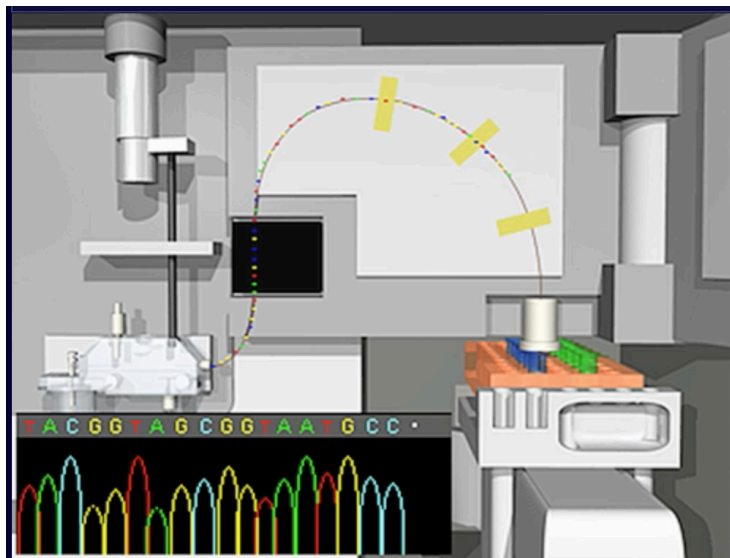
På figuren ovenfor ses de DNA nucleotider, der tilsættes mastermixet til en cycle sequencing PCR. A viser de almindelige DNA deoxynucleotider, der tilsættes flest af. B. viser de fire DNA dideoxynucleotider, der hver er mærket med et fluorescerende farvestof, og som kun tilsættes i små mængder.

17. Forklar, hvorfor at DNA polymerasen ikke kan sætte flere DNA nucleotider på efter at der er sat et DNA dideoxy nucleotid på den voksende DNA strengs kopi.
18. Hvorfor skal man bruge flere cycle sequencing reaktioner med hver sin primer, for at lave en DNA sekventering af et DNA stykke på ca. 1500 basepars længde?

Ordforklaringer: feasible=muligt, redundant=gentagelser (kopier af samme data), fluorescence tagged terminator=dideoxynucleotider mærket med fluorescerende farvestoffer, thermocycler=PCR maskine.

² Modificeret efter https://uk.wikipedia.org/wiki/Файл:Fluorescent_ddNTPs_for_Sanger_sequencing.svg

DNA sekventering



19. Hvorfor vandrer DNA, når der sættes strøm igennem en gel?
20. Hvorfor afskilles DNA fragmenterne af forskellige længde, når de vandrer igennem kapillær gelelektroforese gelen?
21. Hvordan bestemmer Sekventerings maskinen, hvilken DNA base der afslutter et specifikt DNA fragment.
22. Hvorfor er det svært at få bestemt DNA sekvensen ved meget korte og meget lange DNA fragmenter?

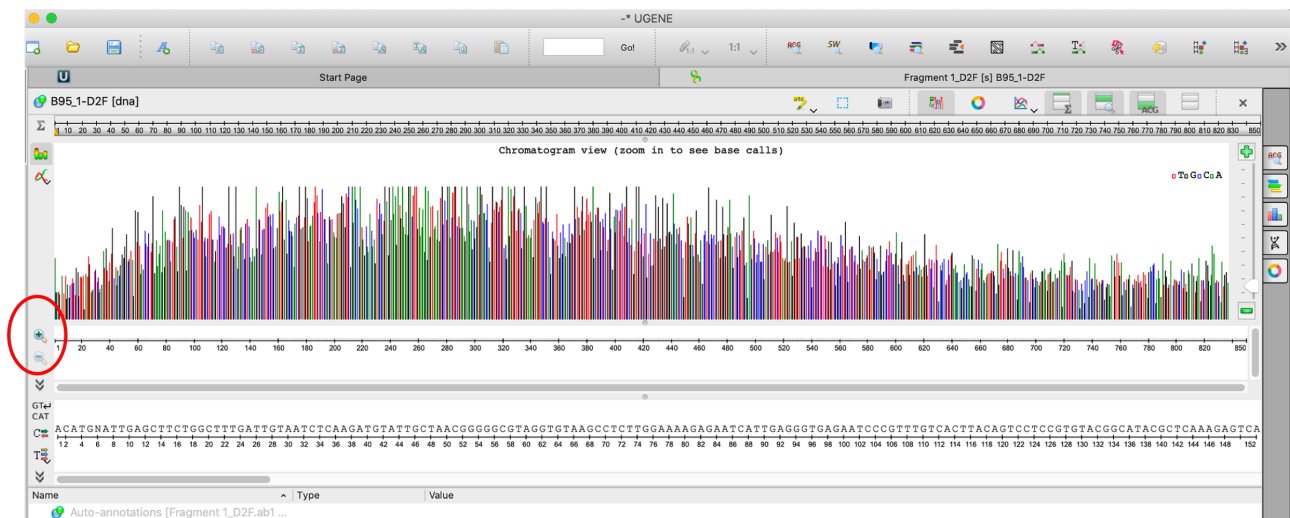
Ordforklaringer: automatic sequencer=DNA sekventeringsmaskine, current=strøm, capillary=kapillær dvs. meget tyndt rør, laser beam=laser stråle, flushes=skyller, tray=bakke, collated=sat sammen.

Næste trin af processen med at lave DNA sekvenser er at samle de enkelte korte DNA sekvenser fra DNA sekventeringen til lange DNA sekvenser. Dette trin er sprunget over i det virtuelle laboratorie, men I skal prøve at lave det her vha. bioinformatik programmet Ugene.

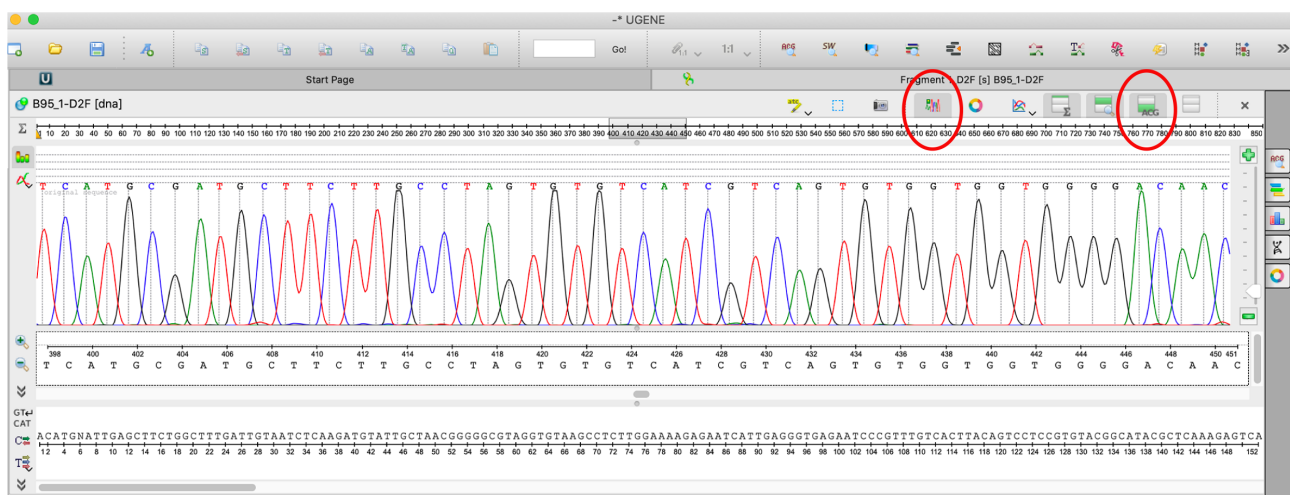
Aktivitet 2. Samling af lange DNA sekvenser.

Ved Sanger sekventering er der grænser for, hvor lange DNA fragmenter man kan sekventere ud fra en enkelt PCR cycle sequencing prøve. Meget korte og meget lange DNA fragmenter vil enten vandre alt for hurtigt eller langsomt forbi laserdetektoren, og derved give utydelige signaler. Derfor er det meget vigtigt at tjekke kvaliteten af de enkelte DNA sekvenser, for at kontrollere at den base som er angivet faktisk er placeret korrekt. Det skal I prøve i første del af denne aktivitet.

Åben filen Fragment 1_D2F.ab1 i Ugene. I får nu en skærm der ser ud som følger:



De enkelte kromatogram toppe er mast meget sammen, fordi programmet prøver at vise hele sekvensen på en gang. Klik derfor på + ikonet (markeret med rød cirkel) nogle gange, indtil at I har et billede med kromatogram toppe, der er tydelige at se.



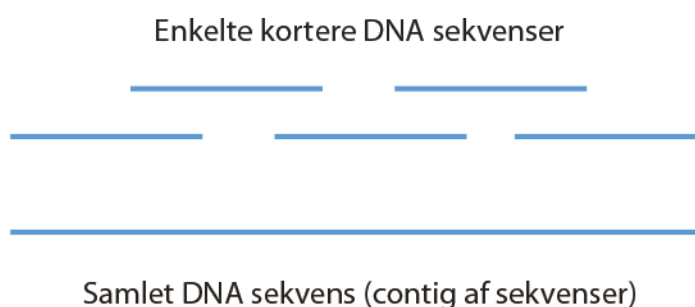
Hvis I ikke får et billede op som det viste skal I trykke på de to ikoner, der er markeret med rødt ovenfor. Så får I den ønskede billede.

1. Hvilke farver har A, G, C og T basernes kromatogramtoppe?

Scroll nu hen til begyndelsen af kromatogram toppene.

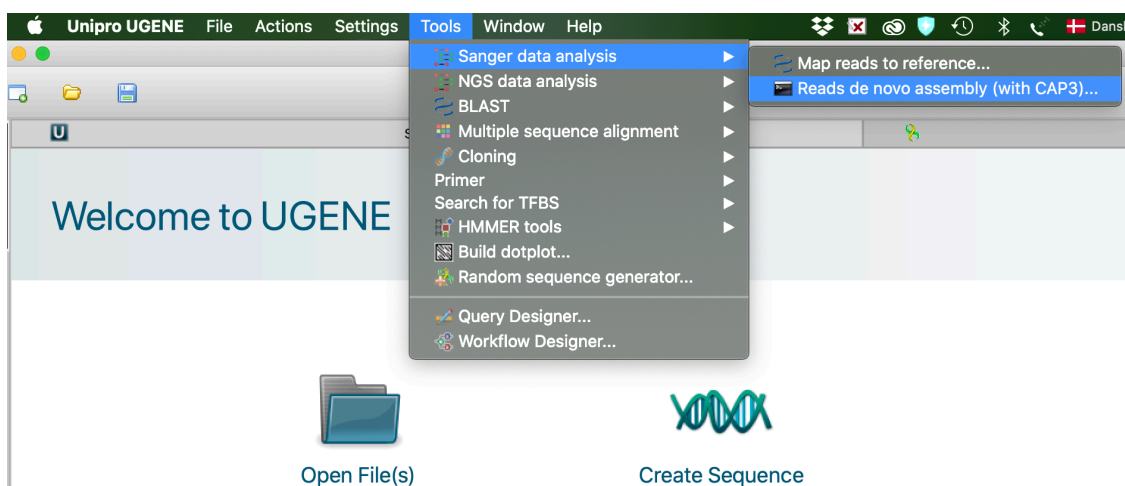
2. Se på position 766 i DNA sekvensen. Baseparret er angivet med et N, der angiver at det kan være en hvilken som helst af de 4 DNA baser. Forklar ud fra kromatogrammet, hvorfor basen er angivet med et N?

Når man har tjekket kvaliteten af de enkelte DNA sekvenser, er det tid til at sætte dem sammen til en lang sekvens. I valget af primere til PCR cycle sequencing sørger man for at vælge primere, der giver overlappende DNA sekvenser, se figuren nedenfor.



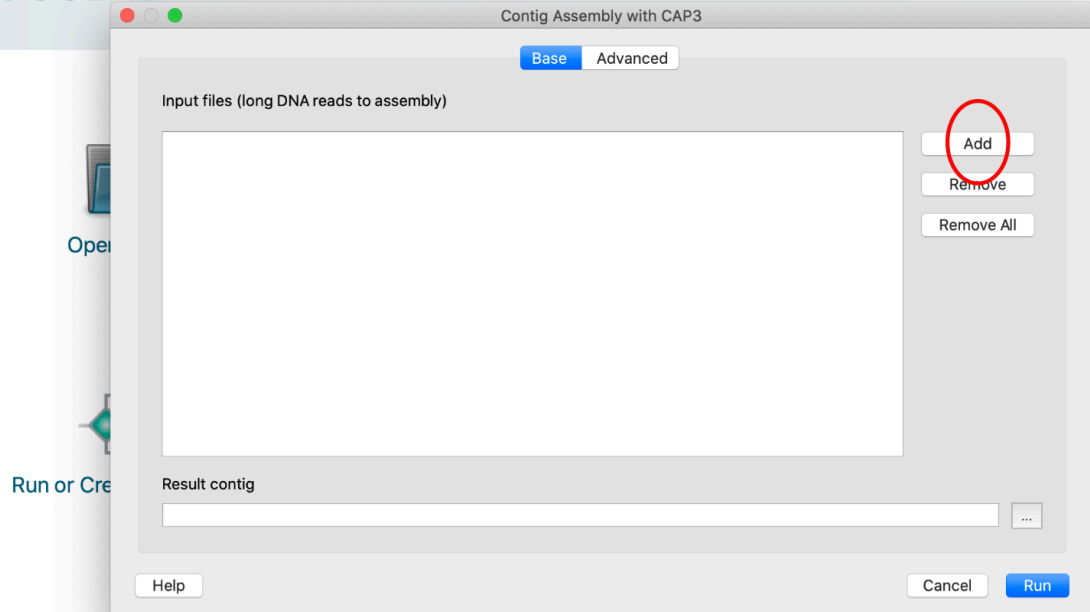
Man kan aligne (placere sekvenserne, så ens basepar overlapper), så man skaber en såkaldt Contig, dvs en konsensus sekvens ud fra alle de overlappende DNA fragmenter. Det skal I gøre nu.

Gå ind på fanebladet Tools i Ugene. Klik nu på Sanger Data analysis, og videre på reads de novo assembly (se figur nedenfor).



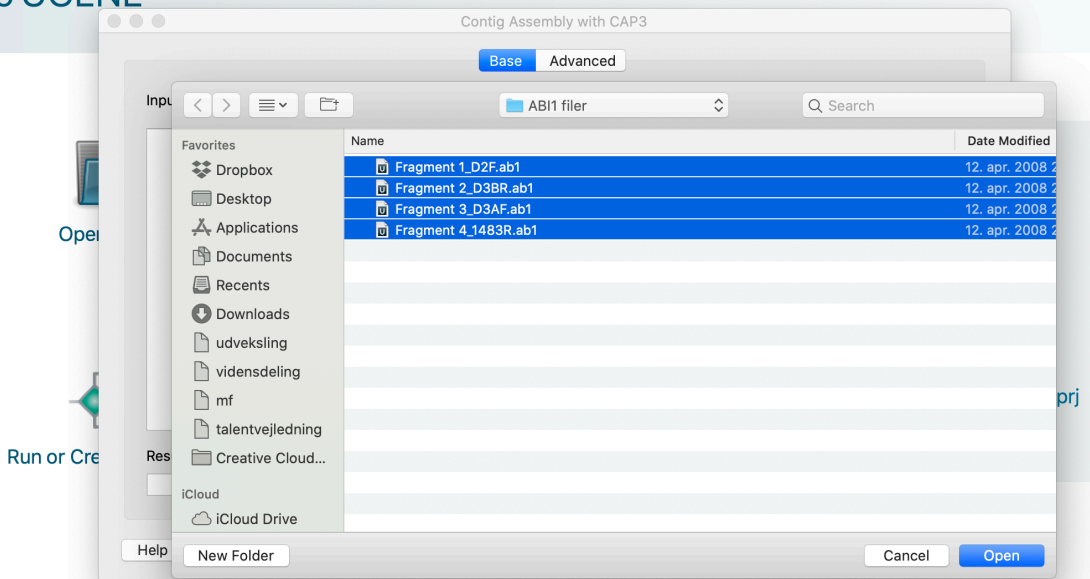
Følgende vindue åbner nu:

Welcome to UGENE



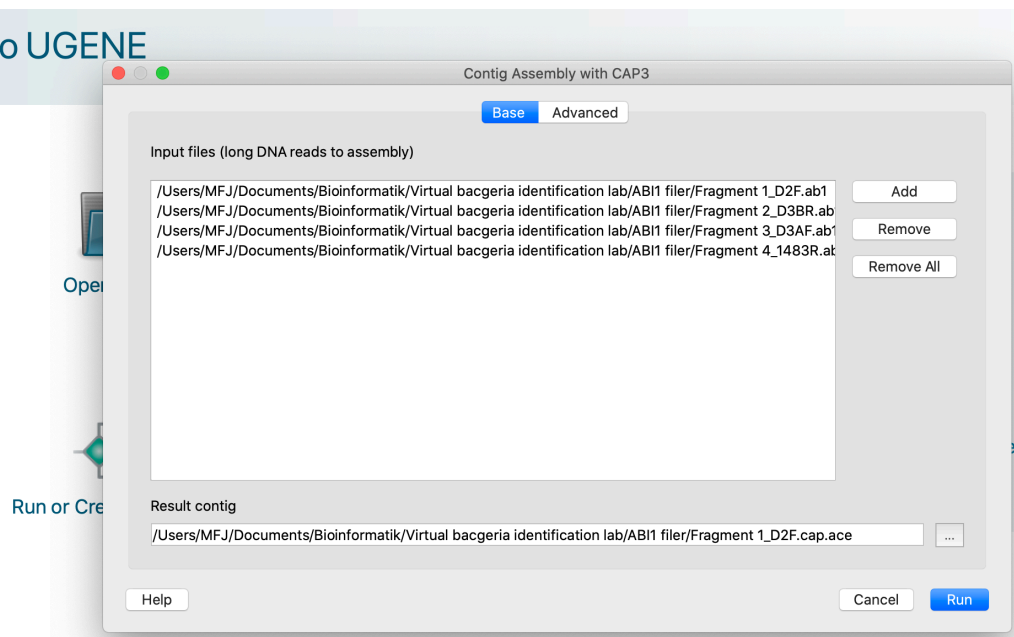
Klik på Add knappen. Du får nu følgende vindue åbnet:

Welcome to UGENE

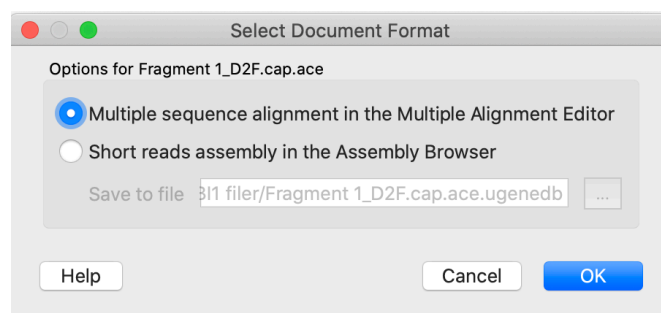


Vælg Fragment 1-4 ab1 filerne og tryk på open. Filerne indlæses nu i contig vinduet som vist på næste side.

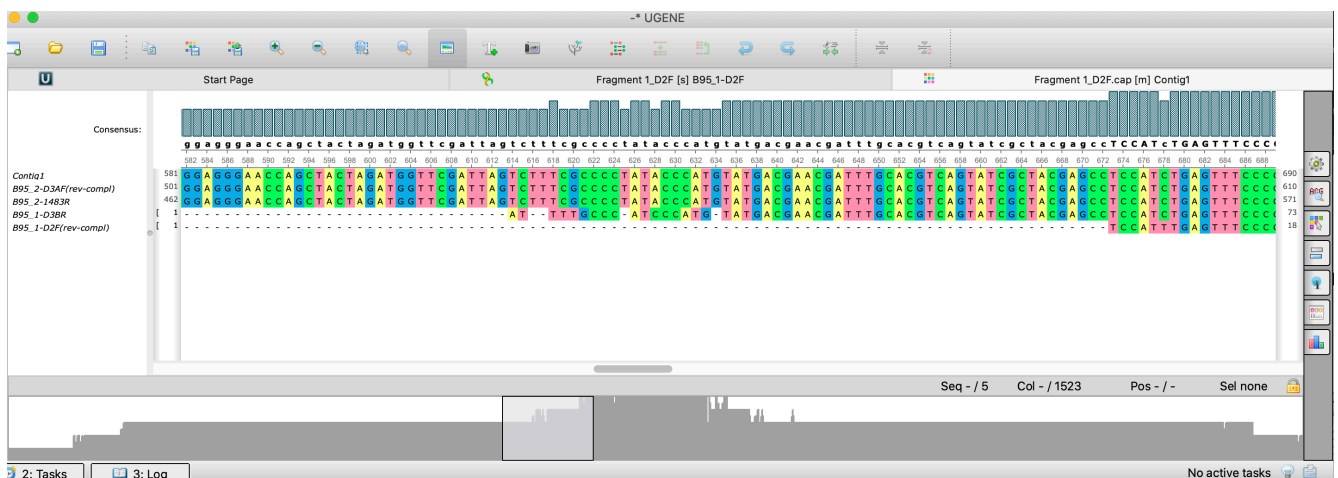
Welcome to UGENE



Tryk på Run. Du får nu følgende boks med to valgmuligheder.



Vælg Multiple Sequence Alignment og tryk ok. Programmet vil nu prøve at placere de fire sekvenser, så de overlapper og giver en lang Contig, dvs consensus sekvens. I får et skærmbillede som det vist nedenfor.



Se først på Contig sekvensen:

For at være sikker på at contig sekvensen er korrekt skal der helst være mindst 2 sekvenser, der overlapper det samme sekvensområde og som viser de samme baser. Det sker for første gang fra baseposition 117.

3. Hvor godt matcher de to sekvenser hinanden i den første del af overlappet?
4. Det område, man kan være mest sikker på er der hvor flest sekvenser overlapper. I hvilket sekvensområde er det?

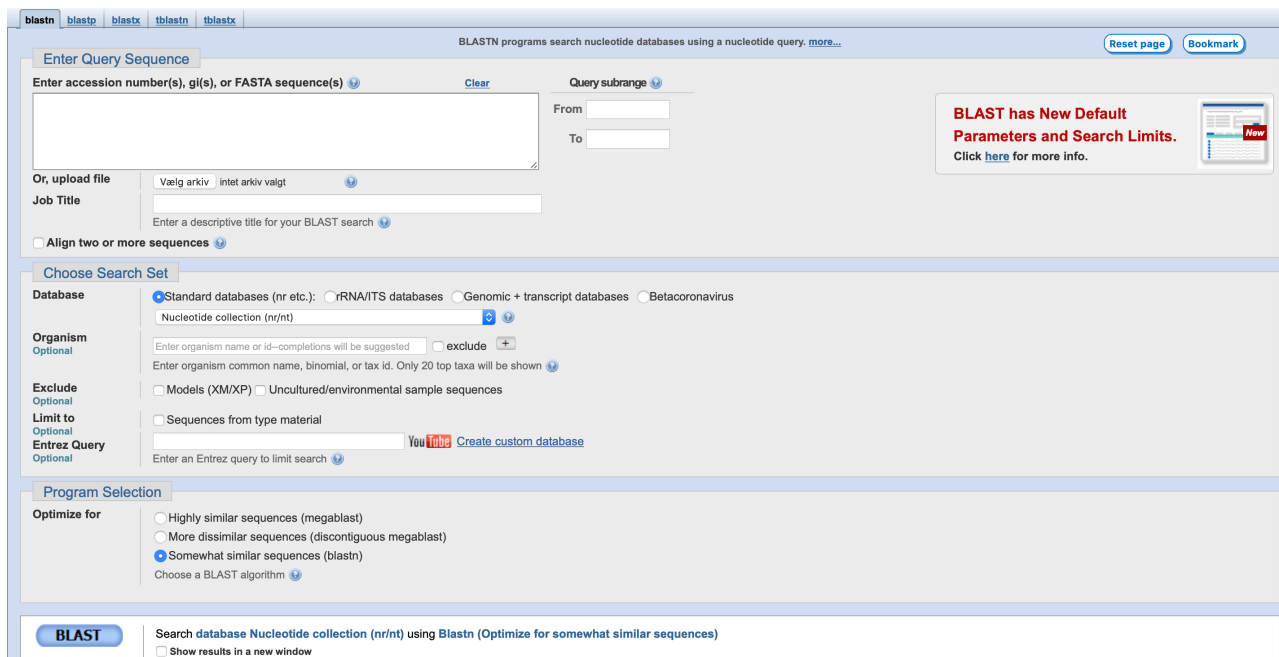
Den dannede sekvens kan nu kopieres og bruges til søgning i DNA databaser, hvilket I skal gøre i aktivitet 3. I bruger dog de sekvenser I får i det virtuelle laboratorium.

Aktivitet 3. DNA database søgning med BLAST (basic local alignment search tool)

Resultaterne fra de 6 prøver er nu klar til at blive identificeret. Dette gøres ved at søge i en offentlig database med alle publicerede DNA sekvenser som drives i et samarbejde mellem EU, Japan og USA. I den kan man sammenligne den fundne sekvens med alle sekvenserne i databasen, og se hvor identisk den er med sekvenserne i databasen.

DNA sekvensen, som I har bestemt i laboratoriet skrives med 4 bogstavskoden, nemlig ACTG. Koden angiver altid den kodende streng, og skrives i 5' til 3' retning. Den åbner op i et nyt vindue i det virtuelle laboratorium. Kopier sekvensen ved at markere hele sekvensen og tryk Command C.

Blast search laves på følgende hjemmeside: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>
Her vælger man nucleotide blast, hvorefter man kommer hen til denne side (det virtuelle laboratorie linker direkte til siden)





Kopier den bestemte bakteriesekvens ind i vinduet under Enter Query Sequence ved at trykke Command V. Tryk nu på Blast knappen nederst til venstre på siden. Hav tålmodighed, det kan godt tage et par minutter.

Hvad sker der nu? Den indsendte sekvens deles op i mindre stykker, der hver især matches op mod de mange millioner af sekvenser, der ligger i databasen vha. nogle matematiske algoritmer. Hvis der er match i et område, prøver søgemaskinen at lave små alignments, dvs at matche flere af fragmenterne ved at placere dem under hinanden i det fundne område. Jo bedre match jo højere score tildeles den fundne sekvens. Resultatet er en liste, hvor den øverste sekvens er den mest identiske og så med sekvenser listet, der er mindre og mindre identiske.

1. Skriv hvilke 6 bakterier I har identificeret og hvilken type prøve de kom fra.
2. Skriv et kort referat af hvilke sygdomme og hvilke symptomer man har, hvis man er inficeret med bakterierne.

I er nu færdige med at identificere de sygdomsfremkaldende bakterier. Men I virtuelle øvelser går alting jo også altid godt. I forsøget undersøgte I bakterier, der var blevet dyrket fra sput-, afførings- og blodprøver, hvorefter man opformerede deres 16S SSU rRNA gen med universelle primere. Denne metode giver ofte problemer i den virkelige verden.

3. Overvej, hvorfor det kan være et problem hvis man skal dyrke bakterierne før man kan kortlægge deres DNA.
4. Hvilke udfordringer tror I det kan give at bruge universelle primere til at opformere DNA fra bakterieprøver?