

Case Oppgave: DNA-Strekkoding

Begreper:

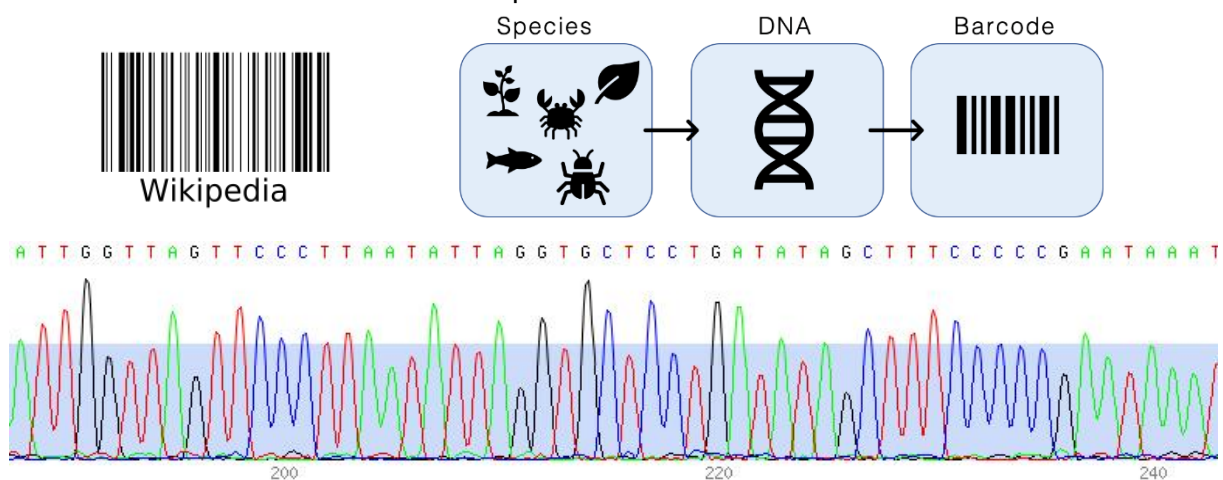
- DNA – Arvematerialet, består av de fire deoksyribonukleinsyrene adenine (A), cytosin (C), thymin (T) og guanin (G).
- Sekvensere – Lese av fysisk DNA til bokstaver som vi kan lese.
- Sekvens – DNA lest av som en sammenhengende streng slik vi finner DNA-et i cellen. En sekvens kan komme i ulike lengder når vi sekvenser DNA. Målet vår er å sette sammen mange sekvenser til en komplett sekvens.
- Database – Et område, ofte digitalt som inneholder informasjon om forskjellige ting, for eksempel DNA. Frøbanken på Svalbard kan være et fysisk eksempel på en database for forskjellige frø.

DNA-strekkoding er en metode for å identifisere forskjellige arter. En del av DNA-et vil være unikt mellom arter, hvis vi får tak i disse sekvensene kan vi bruke de til å identifisere arten. Dette fungerer på samme måte som strekkodene i matbutikken, vær eneste strekkode er unik for et produkt.

Etter at vi har tatt en DNA prøve, kan den sendes til et sekvenseringsfirma som vil lese og sekvensere de fysiske molekylene av DNA inn i en datamaskin som A, T, C eller G. De sekvensene vi vil få av firmaet kan vi prøve å sette sammen til en hel sammenhengende sekvens som vi kan bruke.

Ulike dyr og planter etterlater seg DNA hele tiden, vi kan utnytte dette til å identifisere hvilke organismer som finnes i ulike miljøer. Vi kan for eksempel ta en prøve av jord, hente ut DNA og finne ut hvilke organismer som har vært i vår jordprøve.

I denne oppgaven skal vi prøve å identifisere arter fra DNA som vi har levert til et sekvenseringsfirma. Vi har tatt prøver fra noen blomster i Norge, deretter har vi fulgt en oppskrift for å hente ut DNA-et fra prøvene våre som vi har så sendt til et sekvenseringsfirma. Vi fått tilbake 3 sett med DNA-sekvenser, se de tre siste sidene, vår oppgave vil være å sette sammen 3 komplette sekvenser for de 3 ulike settene med sekvenser. Disse skal vi deretter bruke til å identifisere hvilken plante vi har tatt prøve av og hvilke insekter som har besøkt denne planten.



Øverst til venstre: Eksempel på strekkode som vi finner i butikker, her er strekkoden til Wikipedia. Øverst til høyre: Grafisk framstilling på DNA-strekkoding. Nederst: Eksempel på en DNA-sekvens.

Instrukser:



1. For de tre ulike sekvens datasettene nedenfor skal dere sette sammen sekvensene til tre hele sammenhengende sekvenser. Disse sekvensene overlapper hverandre og deres oppgave er å finne overlappene, og deretter klippe og lime disse sammen. Som hjelp vil den første sekvensen være starten på den komplette sekvens.
 - a. Sekvensene som er i sekvenssettet vårt.
 - b. Finn hvor sekvensene overlapper.
 - c. Fjern overlappet i den ene sekvensen og behold overlappet i den andre, deretter lim sekvensene sammen til en komplett sekvens. Fortsett slik til alle sekvensene har blitt satt sammen til en komplett sekvens.

a

```
TTAGGGGCCATTAACTTTATCACAACCATTATTAACATGAAGC  
ATTAACATGAAGCCCCAGCTATTTCTCAATATCAAACCC
```

b

```
TTAGGGGCCATTAACTTTATCACAACCATTATTAACATGAAGC  
ATTAACATGAAGCCCCAGCTATTTCTCAATATCAAACCC
```

c

```
TTAGGGGCCATTAACTTTATCACAACCATTATTAACATGAAGCCCCAGCTATTTCTCAATATCAAACCC
```

2. Nå skal vi bruke et program mange biologer bruker for å identifisere hva slags organisme DNA sekvensen tilhører.

NCBI: National Center for Biotechnology Information er en database som inneholder DNA-sekvenser for alle organismer som har fått DNA-et sitt avlest, det er altså en database som inneholder mange sekvenser. NCBI har et verktøy som vi skal bruke, dette verktøyet heter BLAST (Basic Local Alignment Search Tool).

Dette verktøyet kan ta imot vår sekvens og søke gjennom hele databasen for en match.

Gå inn på denne linken og følg instruksene gitt på neste side:

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

a

b

c

a. Trykk på Nucleotide BLAST.

b. Lim inn sekvensen deres i området dere ser i bildet.

Trykk på blast. Denne knappen er nederst på siden, hvis dere ikke finner knappen må dere bla nedover.

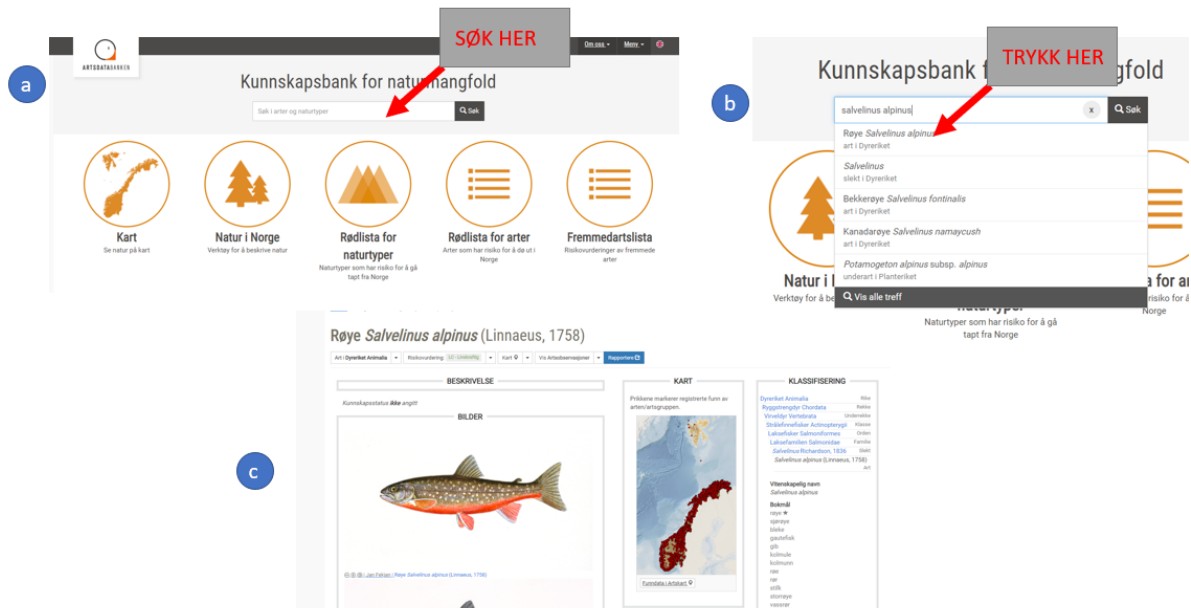
c. Først kommer det en side der programmet søker, den lar dere stå til programmet har funnet en match, det kan ta noen sekunder. Når siden har oppdatert seg skal dere komme til en side som har en liste med mange matches.

Kolonnen som heter Per.Ident. gir dere en match score i prosent, jo høyere denne er jo bedre. Artsnavnet (Her: *Salvelinus alpinus*) består av to deler, et slektsnavn (Her: *Salvelinus*) og et artsepetet (Her: *alpinus*).

Gratulerer dere har nå funnet en match til DNA-sekvensen deres!

- Nå vil vi finne ut litt mer om organismen vi har funnet, det latinske navnet sier oss ikke så mye hvis vi ikke vet hva det er. For å finne ut mer om de organismene som vi har i Norge kan vi bruke artsdatabanken. Artsdatabanken er en annen digital database som inneholder observasjoner over alle av organismer i Norge, alt fra planter, sopp,

fisk, insekter osv. Gå inn på <https://www.artsdatabanken.no/> og se under for instruksjer:



- Skriv artsnavnet du har funnet i søkeboksen, ikke trykk på søk men vent til det kommer opp resultater i en liste under.
- Trykk på det resultatet som inneholder artsnavnet.
- Her er det masse info om artet du har funnet, blant annet kan du se hvor den er funnet i Norge, om det er en truet art, og kanskje til og med noen bilder av organismen.

Oppsummering:

Før hadde vi bare en måte å finne ut hvilke organismer som finnes i et område. Da måtte vi gå ut og samle inn organismer, deretter måtte vi sitte i en stereolupe for å identifisere organismene. I dag gjøres dette fortsatt, men ved hjelp av DNA-strekkoding går arbeidet mye fortere.

Et hovedpoeng med å bruke DNA-strekkoding er at vi kan identifisere arter som har etterlatt seg DNA, dette betyr i prinsippet at vi faktisk ikke trenger å observere organismen, vi trenger bare DNA-et den har etterlatt seg. Dette kan være spesielt viktig når det er arter som er veldig vanskelig å finne, altså at de er sjeldne. Ved hjelp av DNA slipper vi å lete etter artene direkte.

Sekvensdata for sekvens 1:

TATTTACTCCAAATGTAGGAACAGGTTGAACTGTTT

TAAATTTTATTGTAACAATTATAATAATAAAAAAT

TAGGAACAGGTTGAACTGTTTATCCTCCTTTATCATCATATTTATTTCAATTCATC

TTTTTTCTCTTCATATAACAGGAATTTCTTCTATTATTGGATCATTAATTTTATTGT

CATCATATTTATTTCAATTCATCACCTTCTGTTGATATTGCAA

CTGTTGATATTGCAATTTTTCTCTTCATATAACAGGAATTTCT

Sekvensdata for sekvens 2:

GAACAGGATGAACAGTTTATCCCCCACTTTCATCTAATAT

ATCTAATATTGCACACAGAGGATCTTCTGTAGATTTAGCAATTTTCT

AGATTTAGCAATTTTCTCTCTTCATTAGCTGGGATTCTTCAATTTTAGGAG

TTTTAGGAGCAATTAATTTTATTACAACATCATTAAATATACGAGTAAAT

Sekvensdata for sekvens 3:

GGTATTGTGCAGGATGATGTTGGCCTCCCGTGAG

CCGTGAG CTCCATCGTCTCATGGTTGGTTGAAAATCGAGACCT

ATAGATGGTGGATGTGTTACGCACNAGACCAAATAATCATGTGCTGCTCTATT

TCGAGACCTTGGTAGCGTGTGCCATGATAGATGGTGGA
