

Människans närmaste släkt

Lärohandledning

I denna bioinformatikövning som inspirerats av Nobelpriset 2022 till Svante Pääbo tar man fram ett fylogenetiskt träd baserat på mitokondrie-DNA och tar reda på vilken art som hör hemma på vilken gren i släktrödet med hjälp av några ledtrådar.

Syftet med övningen är att öva på problemlösning utifrån genetiska data och att träna på att tolka fylogenetiska träd.

För vem? Övningen har prövats ut av lärare och elever på både högstadiet och gymnasiet. Lärarutvärderingen pekar på stor spridning i hur lätt eller svårt elever har att klara övningen. Detta återspeglas även i lärarnas skattning i hur lång tid de uppfattat att övningen tar för elever. Beroende på elevgruppens förkunskaper kan de ledtrådar som ges sist i elevinstruktionen delas ut redan från början eller bara vid behov om du vill utmana eleverna mer.

Hur lång tid tar det? Elevutvärderingarna visade på ganska stor spridning i hur lång tid det tog att göra själva datorövningen. De flesta har angett mellan 30-60 minuter för att göra alla uppgifter. Beroende på belastning i datortrafik ingår viss väntetid för att analysen av datat ska bli klart.

Vad behövs? Datorer med internetuppkoppling. Möjlighet att ladda ner en fil med mtDNA-sekvenser. Elevinstruktionen digitalt och/eller utskrivet.

På följande sidor ger vi dels förslag på hur övningen kan ramas in i biologiundervisningen, förslag på svar till uppgifterna samt några didaktiska råd baserade på forskning om användning av bioinformatik i skolan.

Mer information

Övningen "Människans närmaste släkt" finns på Bioreurs webbplats under Resurser och Bioinformatik.

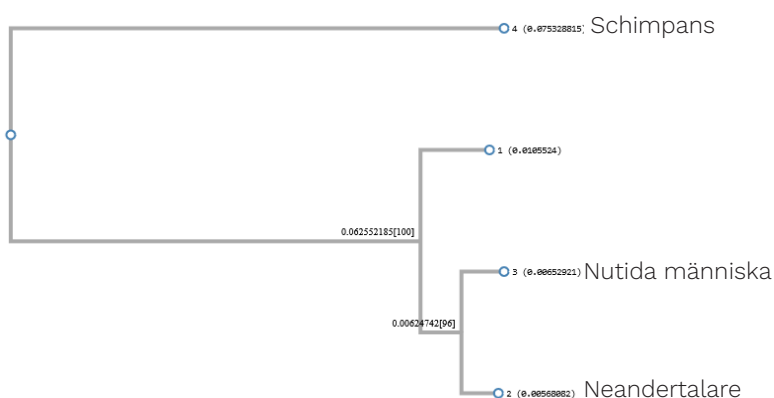
Svarsförslag och kommentarer till uppgifterna

Uppgift 1

- Den nutida människan passar in på sekvens 3 med motivering att den sekvensen är exakt lika lång som referensgenomet för människor som anges i den inledande faktatexten (16569 bp). OBS! Det finns en variation mellan människor i exakt hur långt mtDNA är, men vi har valt att använda referensgenomet som en ledtråd. Bland de övriga sekvenserna finns längder mycket nära detta värde medan en sekvens (för schimpans) avviker mer (16554 bp). Det kan vara bra att diskutera styrkan i denna typ av "bevis". Enbart längden på mtDNA säger egentligen inte så mycket.
- Sekvensen som är mest lik den nutida människan är sekvens 2 vilket man ser genom att titta igenom Score-värdena i resultattabellen. Score 98.654 för den parvisa jämförelsen mellan sekvenserna är både det högsta värdet för alla jämförelser, och framförallt det högsta värdet för alla jämförelser för gjorts för just sekvens 3 med de övriga.
- Sekvens 4 borde tillhöra schimpansen med argumentet att den borde vara mest olika de andra som alla är olika typer av människor. Sekvens 4 har dels en avvikande längd på sitt mtDNA (se uppgift 1a) och dels ger alla jämförelser med sekvens 4 låga Score-värden.

Uppgift 2

- En skiss kan ritas för hand, men även laddas ned från resultatsidan. I elevinstruktionen finns en figur med förklarande text på sidan 4. Ett tillägg är att man kan summera värden som anges inom parentes efter numret på sekvenserna i spetsarna på grenarna (som motsvarar avståndet till närmaste förgreningspunkt) med de värden som ges på förgreningarna för att få fram olika distansmått.
- Nutida människan skrivs in på grenen med siffran 3 på.
- Schimpans skrivs in på grenen för sekvens 4.



Uppgift 3

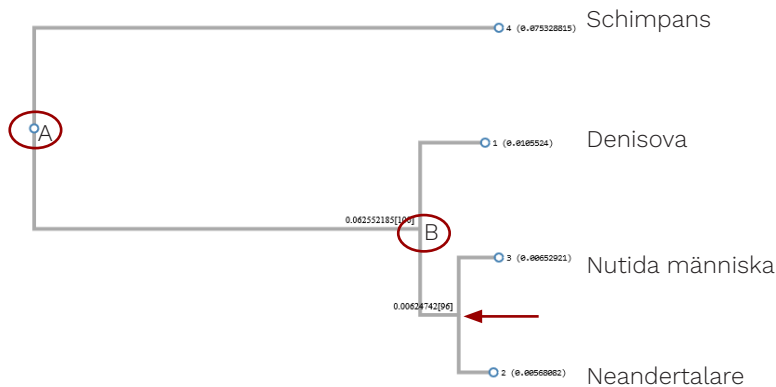
- Neandertalare motsvarar sekvens 2.
- Neandertalare skrivs in på grenen med siffran 2, det vill säga närmast släkt till nutida människan i detta släktträd.

Uppgift 4

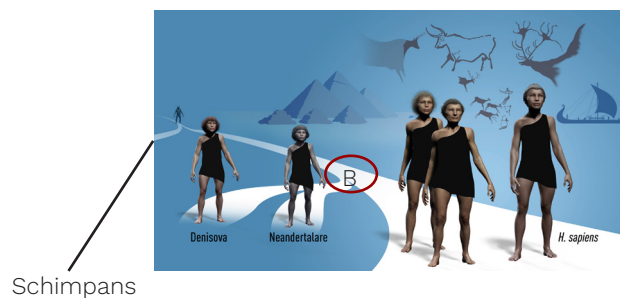
Denisovamänniskan kan placeras ut på den sista grenen, sekvens nummer 1.

Uppgift 5

Den närmaste släktingen till den moderna människan enligt släkträdets man får fram utifrån mtDNA är neandertalare. Den röda pilen pekar på den närmaste gemensamma förfadern för dessa två människoarter.



- Markeringen "A" i figuren ovan visar den gemensamma föregångaren för schimpansen och alla människoarter.
- Markeringen "B" i figuren ovan visar den gemensamma föregångaren för alla människoarter i uppgiften.



© The Nobel Committee for Physiology or Medicine.
Ill. Mattias Karlén

Uppgift 6

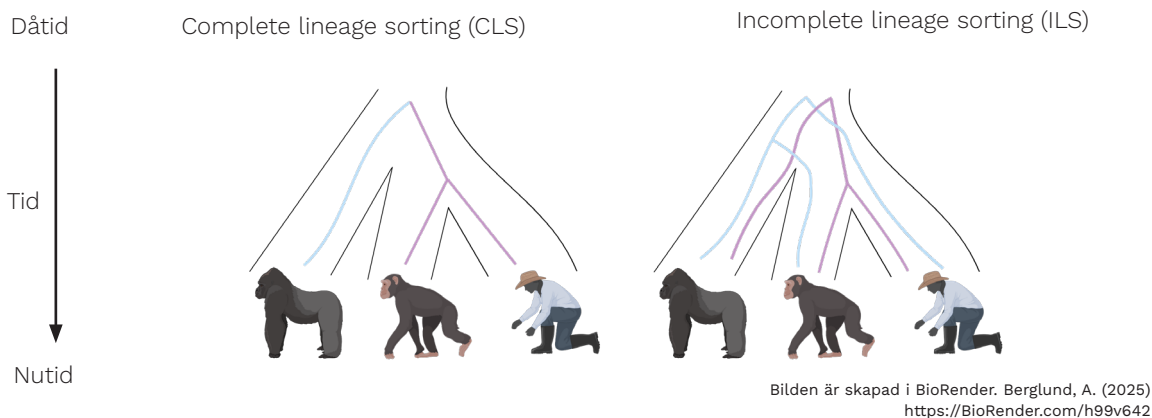
- Se "B" i figuren ovan.
- En släktgren för schimpans finns inritad i figuren ovan. Notera att den ska nå lika långt fram i kanten som nu levande människor eftersom denna illustration visar en skillnad i grenlängd mellan nu levande och utdöda arter.
- Nutida människor står längre fram i bilden då vi lever idag till skillnad ifrån de arter som dött ut (som har grenar som avslutas en bit bakåt i bild).
- I illustrationen visas att neandertalare är närmast släkt med denisova och de är båda lika nära släkt med moderna människor då de delar samma gemensamma förfader (se markering "B"). Det som visas i illustrationen sammanfattar undersökningar av mer data (nDNA) än bara mtDNA som använts i övningen. Mer om utmaningar med att tolka släktskap utifrån genetiska data beskrivs nedan i avsnittet "Något om genträd och släktskapsträd".

Något om genträd och släktskapsträd

Vi försöker spåra släktskap mellan olika arter genom att studera likheter och skillnader i deras arvs massa. Genernas DNA-sekvenser av nukleotiderna A, T, C, G ger många olika "datapunkter". Som nämntes tidigare är det viktigt att sekvenser som ska jämföras är homologa, det vill säga att en position i DNA som jämförs mellan olika arter har ett gemensamt ursprung. Homologa positioner (nukleotider) som är identiska mellan de arter vi jämför kan sägas vara konserverade, de har bevarats utan att mutationer har förändrat dem. Det är dock mutationerna som kan ge ledtrådar till hur olika arter förhåller sig till varandra. Mutationer ger upphov till variabla positioner (polymorfier).

Beroende på om vi jämför mycket avlägset besläktade eller närbesläktade arter kan olika delar av arvs massan vara mer eller mindre lämplig. Väljer man ett område med hög grad av konservering så får man kanske inga polymorfa datapunkter mellan närbesläktade arter då det har gått för kort tid för att några mutationer ska ha hunnit ske. Väljer man istället att titta på sekvenser i arvs massan som muterar väldigt ofta så kan det bli svårt att se några enhetliga mönster, särskilt om det gäller mer avlägset besläktade arter.

Oavsett vilka delar av arvs massan vi använder som data är det vi får fram så kallade "genträd". Om de allra flesta generna (eller datapunkterna) i arvs massan man testat visar exakt samma resultat för förgreningar i "genträdet" kan man med säkerhet dra slutsatser om arternas släktskapsförhållande. Men om olika genträd visar olika förgreningsmönster blir det svårare att tolka släktskapet mellan arterna. Genträd är alltså inte samma sak som släktskapsträd. Det är inte ovanligt att mtDNA-genträd visar en sak och nDNA-genträd visar en annan "historia". En förklaring till detta är så kallad ofullständig linjesortering (*incomplete lineage sorting, ILS*). För att förklara detta fenomen kan man jämföra med *complete lineage sorting (CLS)*, se bild till vänster nedan).



I bilden visas släktskapsträd som breda banor (linjer i svart). Inuti visas genträd i ljusblått och lila. Högst upp visas ljusblått och lila för att symbolisera att vid den tidpunkten fanns en variation i populationen av föregångare till dagens arter av gorilla, schimpans och människa. I CLS (vänster bild) sker en uppdelning så att de ljusblå och lila varianterna följer två olika "fåror". Den vänstra (ljusblå) ger upphov till gorilla medan den högra fåran (lila) ger upphov till schimpans och människa. Resultatet av denna fullständiga "uppsortering" är att genträdet helt överensstämmer med släktskapsträdet för gorilla, schimpans och människa. Genträdet säger samma sak som släktskapsträdet. I ILS (höger bild) sker ingen fullständig sortering utan de olika genvarianterna delas av fler än en av arterna. Beroende på vilket genträd man tittar på (det ljusblå eller det lila) så får man olika bild av släktskapet.

Man kan se ILS som ett problem - vi får svårare att kartlägga släktskapsförhållanden på grund av det. Men fenomenet ger intressant information om vad som har hänt med populationerna under evolutionens gång. Genom att kartlägga många olika genetiska datapunkter och studera hur vanligt ILS är kan man få en uppfattning om populationsstorlek och exempelvis uppskatta för hur länge sedan en uppdelning till olika arter har skett.

Slutligen en kommentar till att använda mtDNA för att kartlägga släktskapsförhållanden. Mitokondrier nedärvs från mamman via äggcellen. Med mtDNA sker ingen rekombination (vilket sker för nDNA vid meios, då könsceller bildas) vilket gör att det blir lättare att fastställa släkträd utifrån genträden. Men tiden som passerat sedan olika människoarter bildades är relativt kort. I mtDNA som är relativt litet (ca 16500 bp) finns därför förhållandevis få skillnader (mutationer) att dra slutsatser ifrån. En fördel med att använda nDNA är att det innehåller många många fler positioner som varierar mellan olika individer. Då kan man få fram många olika genträd och utifrån dem få en övergripande bild av vilket genträd som är vanligast och därmed mest sannolikt motsvarar det fylogenetiska släkträdet.

Några didaktiska råd

Forskning kring hur elever tar sig an och förstår bioinformatikövningar visar bland annat att de ofta klarar av själva knapptryckandet men att typen av instruktion belastar arbetsminnet olika mycket. Eleverna behöver ofta mest stöd i att knyta an görandet i datormiljön med själva biologin i den kontext man arbetar inom. Det finns flera didaktiska val som kan ge stöd för reflektion och koppling till det biologiska innehållet under genomförandet. Här ger vi några tips:

Arbeta själv eller i grupp? Ett förslag är att låta eleverna arbeta två och två vid en dator. De kan turas om att ha olika roller i själva genomförandet. En roll kan vara "Ordning och reda": läser instruktionen högt och leder på så sätt arbetet steg för steg. Den andra rollen kan vara "Nyfiken - varför?": klickar och genomför de olika stegen men har i uppgift att efter varje steg ställa frågan - varför gjorde vi det här? Ett annat alternativ är att gruppera eleverna om 3-4 per grupp där alla arbetar vid varsin dator men där en får rollen "Hänger du med?" för att se till att alla i gruppen tar sig framåt i övningen. Det är inte helt lätt att själv som lärare hantera frågor från många elever samtidigt i klassrummet. Det kan då underlätta att vissa problem löses inom elevgrupperna där de kan stötta varandra.

Utskriven eller digital instruktion? En fördel med att ha en utskreven instruktion är att man kan bocka av och skriva till anteckningar vartefter man tar sig framåt i övningen. En fördel med en digital instruktion är att länkarna i dokumentet är klickbara. I elevinstruktionen finns plats för att skriva korta svar på uppgifterna för hand. Nackdelen med enbart en utskreven instruktion är att man då behöver se till att man kommer till rätt webbplats för ClustalW (man behöver skriva in webbadressen eller söka fram den). Uppgift 3 bygger på att man ska kopiera en kortare sekvens för att matcha resultatet i ClustalW. Om man enbart har elevinstruktionen i pappersform behöver eleverna manuellt skriva in sekvensen i sökfönstret vilket kan ge skrivfel och tar viss onödig extra tid. Elevinstruktionen innehåller ett antal skärmdumpar med inringade områden för att uppmärksamma var man ska klistra in data och var man läser av olika resultat. Balansen mellan text och bild kan behöva anpassas beroende på elevers behov. Ett alternativ eller komplement till elevinstruktionerna kan vara att man som lärare visar på storbild först hur den digitala miljön ser ut som eleverna ska arbeta i. ClustalW är ett enkelt och användbart verktyg som går att använda för många andra uppgifter. I ett kortare videoklipp via [Bioresurs Youtubekanal](#) visar vi hur man arbetar i programmet ClustalW.

Lärostopp? Att stanna upp och gå igenom någon uppgift kan vara bra både för att få koll på hur det går för elevgrupperna men också ge tillfälle att gemensamt diskutera syftet med de olika steg som ingår i uppgiften. Visa delar av övningen via storbild. Här kan man även använda extra frågor som inte ingår i elevinstruktionen. Det kan hjälpa till att reda ut begrepp och fakta som kanske inte är självklara för alla, knyta an till begrepp som ni kanske har arbetat med tidigare för att underlätta det fortsatta arbetet. Några förslag på "checkfrågor":

Hur ser mtDNA-datat ut? Visa att sekvenserna i filen består av långa rader av kvävebaserna ATCG. Över 16500 bokstäver för varje sekvens. bp betyder baspar. Här kan man också ta upp en jämförelse: mtDNA-filen består av ca 1200 rader med vardera 70 bokstäver där filens storlek är 84 kb (kilobyte). Om vi skulle använt helgenomdata från kärnans DNA skulle vi behövt en fil med över 200 miljoner rader vilket motsvarar en fil på över 3 GB (gigabyte). Frågor som kan dyka upp här är varför det bara är en rad med bokstäver och inte två om man tänker på att vi lärt oss om DNAs byggnad med basparning A-T och G-C. Svaret blir då att det räcker att analysera den ena strängen just för att den komplementära basparningen. Det spelar ingen roll vilken sida vi jämför, men det förutsätter att vi jämför just samma sträng. Här kan man koppla till att DNA-strängarna har en riktning (3' och 5'-ändar skiljer sig åt kemiskt) vilket gör att vi kan veta vilken sträng vi jämför.

Hur ser en mutation ut? När man kört multiple sequence alignment och får fram resultatsidan så ser man hur de fem sekvenserna matchats, om man scrollar nedåt. Visa att alla positioner som har en stjärna (*) under är identiska mellan alla fem sekvenser. Leta efter positioner som saknar en stjärna för att visa exempel på mutationer. Här finns många exempel på där det skett ett utbyte av en kvävebas mot en annan. Det finns också exempel på insertioner och deletioner som visas med hjälp av streck (-). Om man tittar igenom datat ser man att särskilt sekvens 5 har avsnitt där det står flera "N" (istället för A, T, G eller C). Det är inte mutationer utan motsvarar positioner i sekvensen som visat sig vara svåra att få fram ett säkert resultat för när man sekvenserat materialet.

Vad finns det för koppling mellan mutationer och Score-värden? Ju fler skillnader som finns mellan två sekvenser desto lägre score-värde. Om vi tänker oss en ursprunglig sekvens så innebär fler mutationer över tid större skillnader mellan sekvenserna om vi följer dem framåt i tiden.

Hur ser kopplingen ut mellan mutationer och släktskap? Ett antagande som görs när man bygger släkträd på det sätt som görs i uppgiften är att ju fler skillnader som finns mellan två sekvenser desto längre tid har passerat sedan de tillhörde samma art. Om man vet hur många mutationer som sker per tidsenhet kan man räkna om antalet genetiska skillnader till tid - alltså få fram ett mått på hur länge sedan det är sedan två sekvenser var lika. Man brukar prata om en "molekylär klocka". Metoden är inte okontroversiell, för det varierar hur ofta mutationer uppkommer i olika delar av DNA och hos olika arter. Dessutom behöver man ta hänsyn till att organismer har olika generationstider om man vill räkna om till tid i termer av år.

Hur vet man vilken sekvens som är den "ursprungliga"? Det är svårt att veta. Det man ofta gör i den här typen av analyser är att man tar med en art som man vet är lite mer avlägset släkt, en så kallad "utgrupp". I den här uppgiften kan vi betrakta schimpansen som en utgrupp. I elevinstruktionen har vi kort nämnt att det finns olika sätt att bygga släkträd. I övningen används så kallade distansmått som grund för släktskapet som byggs upp. Det finns många olika metoder för att bygga släkträd som bygger på olika algoritmer och antaganden. Att enbart använda värden för hur stora skillnaderna är mellan olika arter ger oss inte tillräcklig information för att kunna uttala oss om hur evolutionen gått till. Men de är relativt snabba metoder som ändå ger intressanta mönster att diskutera.