



Engelsk bulldog  
Foto: Pleple2000



Tibetansk terrier  
Foto: Flickr user skaty222



Västgötaspets  
Foto: Sören T Eriksson



Dvärgpinscher  
Foto: Entheta



Chow chow  
Foto: Jurriaan Schulman

Alla bilder Wikimedia commons

Generell instruktion till övningar om

## Ärftliga sjukdomar och egenskaper hos hund

Text: artin Dahlö

Science for Life Laboratory (SciLifeLab) Uppsala universitet

*Övningen handlar om att titta på DNA-sekvenser från hundar och jämföra mot referenser för att se om vi kan upptäcka skillnader som kan ge förändringar eller sjukdomar.*

Alla individer inom en viss art har så gott som identiskt DNA, men enstaka nukleotider kan variera. Hos människa skiljer sig DNA-sekvensen åt på var tusende bas, d.v.s. två individer är till 99.9% identiska. Det är därför till exempel alla hundar ser ut som hundar, men alla är ändå unika och har sitt speciella utseende. Vissa har ljust hår och andra har mörkt, vissa har långa ben och andra korta.

Det är inte bara utseendet som varierar. Till exempel kan benägenheten att få sjukdomar vara olika stor eller också kan förmågan bryta ner olika ämnen i matsmältningen variera. Mycket av detta beror på små förändringar i DNA-sekvensen. När man sekvenserar allt DNA från en individ så kan man se alla dessa små skillnader och använda sig av informationen.

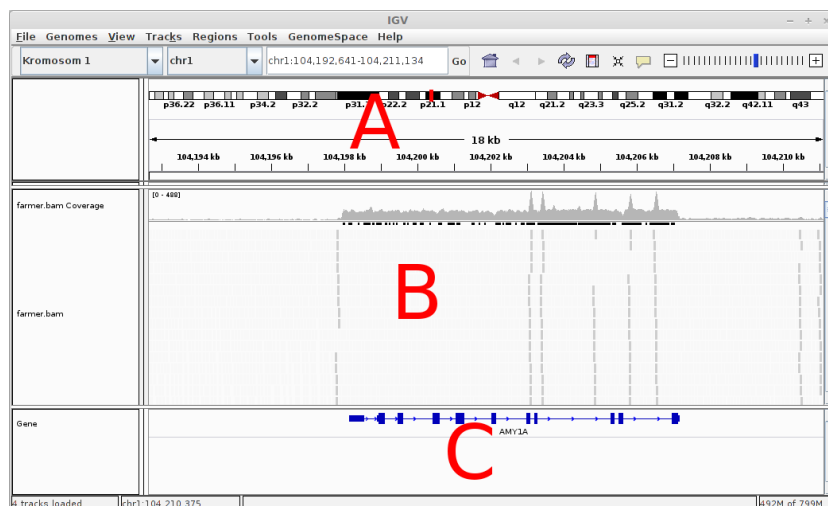
---

Övningen har utarbetats av forskare vid SciLifeLab, Uppsala, i samarbete med Nationellt resurscentrum för biologi och bioteknik och Biotopia

Övning under Bioresursdagarna 2014



För att titta på DNA-sekvenserna kommer vi att använda ett program som heter Integrated Genome Viewer (IGV).



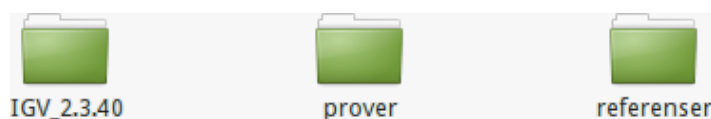
A. Referensgenom. Detta är som ett medelvärde av flera olika hundars genom. Man har utgått från en grupp av hundar och tagit den vanligast förekommande DNA-sekvensen. Det är till denna sammanslagna sekvens vi passar in vårt sekvenserade DNA från en viss hund.

B. Vårt prov. Allt DNA som sekvenserats har delats upp i bitar bestående av 150 nukleotider (A, T, C, och G), och därefter har ett datorprogram hittat platserna i genomet som dessa bitar kommer ifrån. Processandet tar cirka 4-5 timmar per prov för en modern dator om man har ett normalstort prov. Det vi tittar på nu är bara en liten del (cirka 0.01%) av ett normalt prov.

C. En referensannotering. Att endast känna till sekvensen av ett genom löser inga gåtor, man måste också veta vad sekvensen gör. Det är ungefär som att ha en karta, men inga städer eller namn utmärkta. Referensannoteringen motsvarar dessa namn och städer och märker ut var i genomet det finns något intressant. För att ta fram dessa annoteringar har det gjorts många experiment och tester för att visa att sekvensen i ett område har den funktion man påstår.

Materialet för denna labb hämtas från SciLifeLabs server, se [www.bioresurs.uu.se](http://www.bioresurs.uu.se), välj länken Tidningen Bi-lagan och därefter Bi-lagan nr 3 2014. Gå till den mapp där filerna för labben ligger och följ sedan nedanstående steg:

1. Öppna programmet IGV genom att först gå in i mappen "IGV\_2.3.40".



I mappen finns olika filer, vilken du ska öppna beror på datorns operativsystem.



#### **Windows: igv.bat**

kan stå bara igv om Windows är inställt på att gömma filändelser på kända filtyper. Detta skapar lite förvirring, eftersom filen igv.jar finns i samma mapp, som också kan få sin filändelse gömd. Filen med ikonen till vänster ska användas. Den är beskriven som en Windows Batch file.06\_bat\_windows\_icon.png Filen med ikonen till vänster ska användas. Den är beskriven som en Windows Batch file, eller Windows kommandofil.

#### **Mac: igv.command**

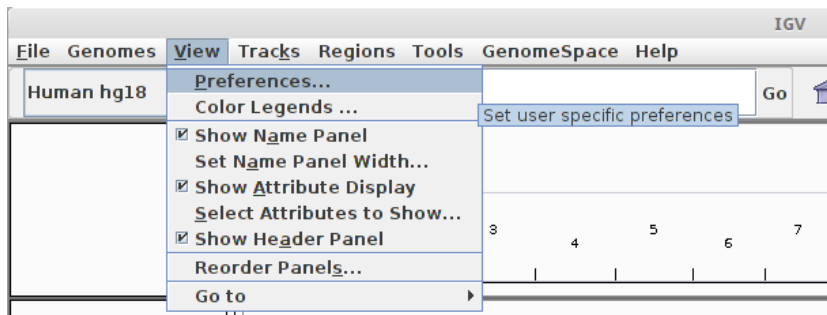
#### **Linux: igv.sh**

Övningen har utarbetats av forskare vid SciLifeLab, Uppsala, i samarbete med Nationellt resurscentrum för biologi och bioteknik och Biotopia

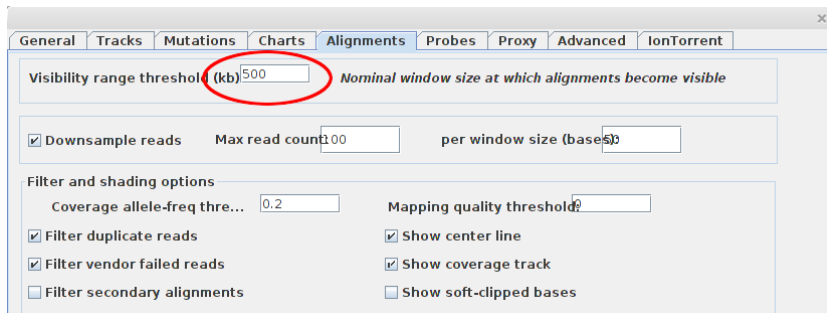
Övning under Bioresursdagarna 2014



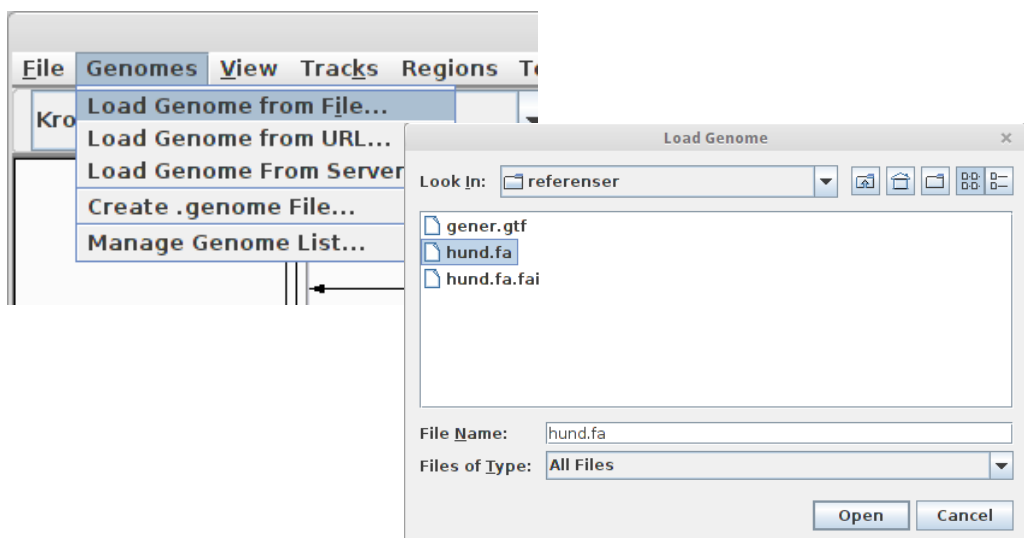
2. När IGV har startat så ska vi först ändra på en inställning så att vi kan titta på stora områden samtidigt. Klicka på **View - Preferences** i menyraden överst.



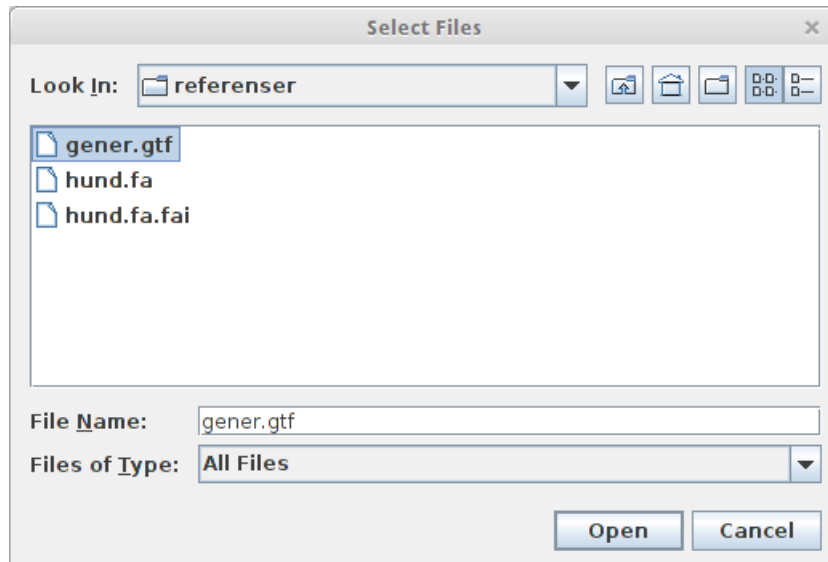
När det öppnats ett nytt fönster med en massa inställningar i så väljer du fliken **Alignments**, och ändrar värdet på det som kallas **Visibility range threshold**. Ändra det från standardvärdet 30 till **500**. Detta gör att programmet kommer att rita ut detaljer även när vi är väldigt utzoomade, vilket är bra om man vill kunna titta på stora områden. Tryck på **OK**-knappen nere i högra hörnet för att spara inställningen.



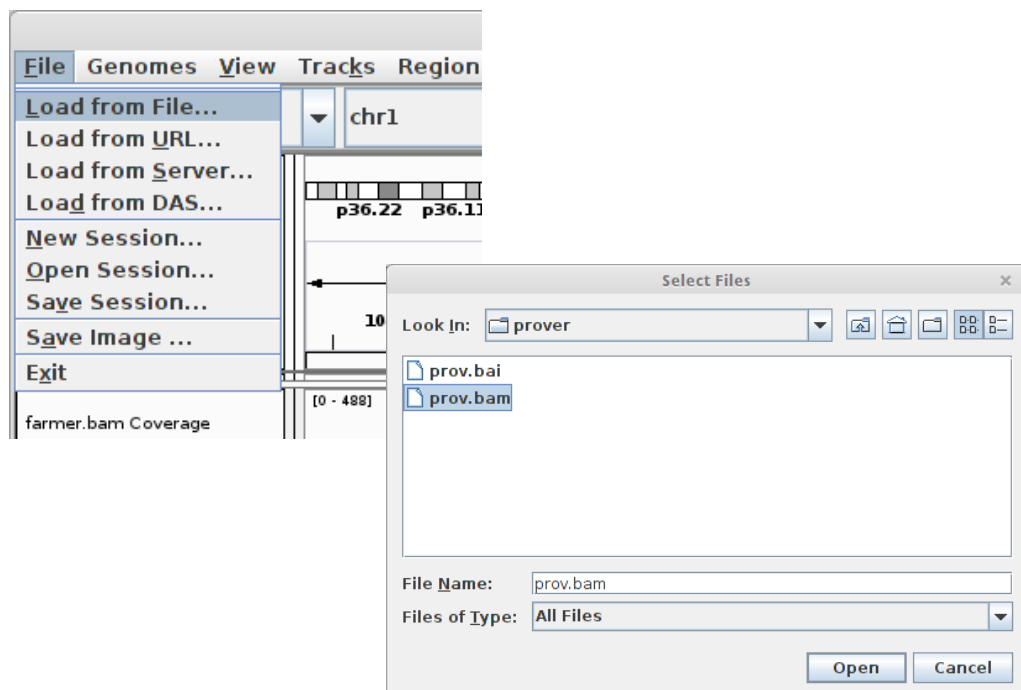
3. När det är klart så är vi redo att börja titta på data. Först måste vi välja vilket referensgenom vi vill visa tillsammans med våra prover. För att göra detta i IGV klicka på menyn **Genomes** högst upp till vänster i fönstret, och välj därefter **Load Genome from File...** Leta sedan rätt på den mapp som innehåller filerna som ingår i övningen. När du hittat den så öppnar du mappen **referenser** och väljer filen **hund.fa**



4. När vi har laddat hundens referensgenom så måste vi även ladda hundens referensannoteringar. Utan dessa kommer vi inte veta var generna sitter. För att ladda referensannoteringarna, klicka på menyn **File** högst upp till vänster i fönstret, och välj alternativet **Load from File...** och gå till mappen med filerna som tillhör labben. I mappen **Referenser** så väljer du filen **gener.gtf**



5. När referensgenomet och referensannoteringarna är laddade saknas endast proverna. Filerna för proverna kommer att heta olika beroende på vilken labb du gör. Här i exemplet används namnet prov.bam för att visa hur man öppnar en prov-fil. För att ladda dessa klicka på menyn **File** högst upp till vänster, och sedan på **Load from File...** igen, gå till mappen med filerna som tillhör labben. Denna gång går du in i mappen **prover** och väljer de filer som anges i respektive övning (filerna du väljer ska ha ändelsen **.bam**)



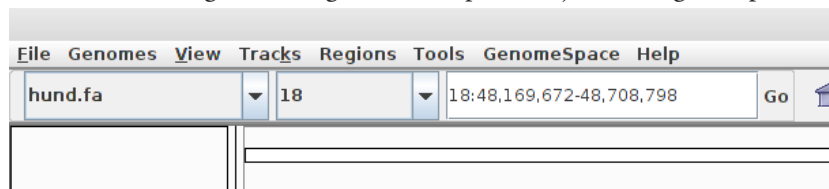
Övningen har utarbetats av forskare vid SciLifeLab, Uppsala, i samarbete med Nationellt resurscentrum för biologi och bioteknik och Biotopia

Övning under Bioresursdagarna 2014



Nu är alla prover och referenser laddade.

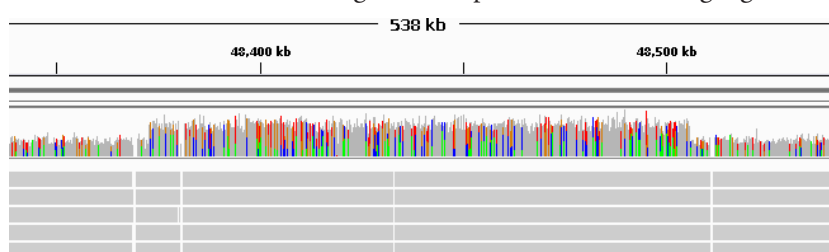
6. Genom genetiska associationsanalyser har man hittat en region i genomet som man misstänker har med sjukdomen/egenskapen att göra (se separat instruktion). För att zooma in till det området, skriv in adressen till det i sökfältet högst upp i programmet och tryck på **Go**. (Adressen anges i övningarna för respektive sjukdom/egenskap.)



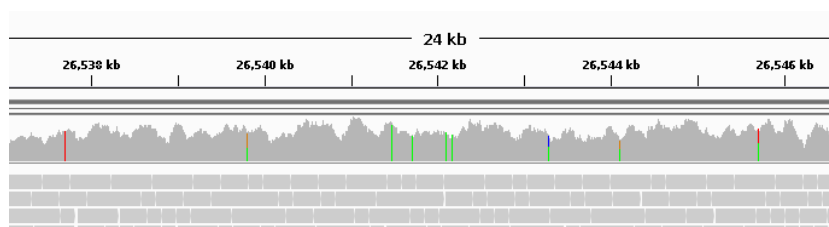
Nu har vi zoomat in på den specifika kromosomen.

För att se hur mycket DNA som sekvenserats från de olika platserna längs referensgenomet visar programmet en så kallad **coverage graph**. Ju högre nivå grafen har desto mer DNA har sekvenserats från just den platsen i genomet.

Ex 1) Större mutation: En hel gen har kopierats en eller flera gånger (se röd ruta)



Ex 2) Mindre mutation: En eller ett fåtal nukleotider har förändrats (se färgade streck).



Alla grå streck nedanför grafen är bitar med 150-nukleotider som har sekvenserats. De har staplats på varandra på de platser där bitarna överensstämmer bäst med referensgenomet. Ju fler DNA-bitar som täcker en position, ju högre coverage har positionen, och det är det som visas av en coverage graph.

Exempel 1

Här ser vi att den region vi är intresserad har högre coverage än sin omgivning. Genom att mäta hur stor skillnaden är mellan medel-coverage-nivån i hela provet och coverage på den intressanta regionen vet vi hur mycket extra genetiskt material från den regionen som finns i vårt prov. Här är det cirka 2x högre coverage (~30 vs ~60), vilket skulle indikera 2 kopior av regionen.

De färgade strecken man ser här och var i coverage graphen är positioner där sekvenserat DNA från den aktuella hunden skiljer sig från referensgenomet. Alla individer är något olika, så det är inte konstigt att vi har dessa skillnader.

Hade referensgenomet också haft 2 extra kopior hade alla dessa DNA bitar smetats ut över en längre sträcka och coverage hade varit på samma nivå som områdena i närheten.

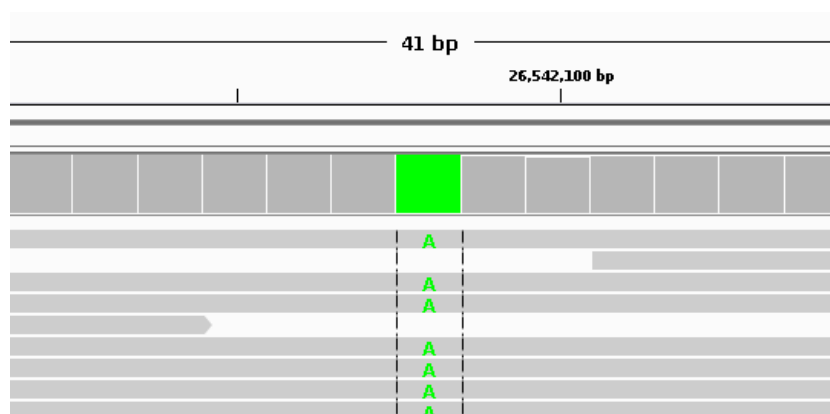
## Exempel 2

Här ser vi också färgade streck här och var i coverage graphen och det är positioner där hunden vi har sekvenserat skiljer sig från referensgenomet. Det är när dessa skillnader hamnar på fel ställen som de kan ge upphov till nya egenskaper eller sjukdomar.

Ett exempel på mutation finns på position **chr31:26,542,098**. Här ser man att alla DNA-bitar har ett A. DNA-bitar som är grå visar att de är identiska med referensgenomet, vilka nukleotider de består av anges ej.

I en övning som handlar om att hitta en position med en förändrad nukleotid, dubbelklickar man på det område som man vill zooma in på.

Man kan också zooma in genom att klicka på +/- ovan till höger.



## Fortsätt nu med övningarna på specifika sjukdomar/egenskaper

Övningarna handlar om antingen dubbleringar av gener eller förändringar av enstaka nukleotider. Fortsätt med en av övningarna som handlar om specifika egenskaper eller sjukdomar hos hundar. För frågeställningar och diskussionsfrågor se respektive övning.