

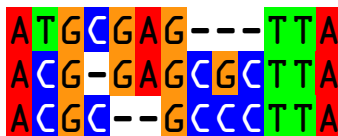
# Konsten att konstruera ett evolutionärt träd

Erik Lundin

Under stor del jorden existens har funnits har det funnits liv på den. Genom slumpmässiga förändringar så har livet förändras. Förändringar som ger en fördel kan förändra en population medan en förändring som ger en nackdel tenderar att göra individens liv kortvarigt. Det liv som finns just nu är bara en bråkdel av allt liv som någonsin funnits. Men framgångsrika arter utvecklas till nya arter som delar vissa drag. Alla fåglar har vingar, även om de kan vara oanvändbara. Det beror på att förfadern till alla fåglar hade vingar. Men alla arter med vingar är inte nära släkt med fåglar. Insekter har vingar men är inte nära släkt med fåglar. Även om det är enkelt att se att en fågel inte är nära besläktad med en fjäril så finns det andra livsformer som det inte är så lätt att se om de är släkt eller inte. För att se om två grupper av djur är nära släkt så brukade forskare leta efter drag som grupperna delade och på samma gång försöka undvika de drag som grupperna oberoende utvecklat. Numera använder forskare mer och mer molekylära sekvenser, såsom DNA och proteiner. Mängden molekylära sekvenser ökar för varje dag och de kan hanteras med datorer.

För att illustrera evolutionens gång brukar **evolutionära träd** användas där ett löv motsvarar en nu levande art medan en nod i trädet motsvarar en teoretisk anfader till alla grenar som växer ut från den. En nod och alla efterkommande arter kallas för en **monofyletisk** grupp och är av intresse för evolutionsbiologer. Många grupper är inte monofyletiska. Ett exempel är så kallade varmblodiga djur, då fåglar och däggdjurs gemensamma anfader inte var varmblodig.

För att tillverka ett evolutionärt träd från sekvenser så hämtas först sekvenserna från någon databas för sekvenser. Då en sekvens som fanns hos en art som bildar två arter har förändras så måste sekvenserna multi-inpassas. En multi-inpassning är när sekvenserna ställs vid varandra och luckor tillåts i vissa sekvenser så att de passar in så bra som möjligt, se figur 1.



```
ATGCGAG---TTA
ACG-GAGCGCTTA
ACGC--GCCCCCTTA
```

Figur 1: En multi-inpassning av tre korta DNA sekvenser. Här är första basen, de tre sista och två G lika hos alla tre. Det finns också ett ställe i varje sekvens där det är en lucka och två ställen där en bas är utbytt mot en annan bas.

En korrekt inpassning visar hur sekvenserna har ändrats av tidens gång och vilka delar som är med konserverade. Dessa ändringar används sedan för att göra ett träd. Då man kan ha för många sekvenser när man börjar så kanske man måste ta bort några sekvenser som inte passar in. Om man tar bort sekvenser så måste en ny multi-inpassning göras med de sekvenser som är kvar och ett nytt träd konstrueras.

För att göra inpassningen och trädkonstruktionen krävs ett verktyg för vardera. Om man tar bort sekvenser i trädet flera gånger så måste man byta mellan verktygen flera gånger. Att hämta sekvenserna kräver också användandet av speciella verktyg. För att förenkla detta så har jag konstruerat ett program som laddar hem sekvenser, multi-inpassar sekvenserna och konstruerar ett träd. Programmet, som heter AutoPhylo, låter också användaren ta bort sekvenser för att sedan göra om multi-inpassningen och trädkonstruktionen med de sekvenser som finns kvar. Programmet finns på <http://code.google.com/p/autophylo/>

Examensarbete i biologi, 15 hp, VT 2010

Institutionen för biologisk grundutbildning och Avdelningen för systematisk biologi, Uppsala  
Universitet

Handledare: Sandra Baldauf och Ding He