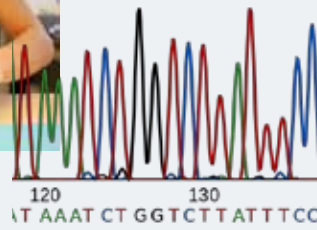


# DNA-streckkodning – så går det till



DNA-analys



Ill.: Everaldo Coelho and Yellow Icon

Matchning mot referensdatabas



Foto: S. Seyfert

Okänt prov



Artbestämning  
*Ardea cinerea*  
(Gråhäger)

Text: Johannes Bergsten  
Naturhistoriska Riksmuseet

Ända sedan metoder för att läsa DNA-sekvenser blev tillgängliga har DNA använts i frågeställningar kring arter, släktskap och identifiering. Eftersom DNA samlar på sig förändringar (mutationer) med tiden så blir DNA:t för olika arter mer och mer olika ju längre tid arterna är skilda från varandra.

Forskning på Naturhistoriska riksmuseet handlar mycket om biologisk mångfald och livets evolutionära historia. Eftersom DNA allt mer används i denna typ av forskning liksom inom andra samhällstillämpningar bildade museet förra året Centrum för Genetisk Identifiering (CGI). CGI kan ta olika typer av uppdrag som innefattar genetisk art- eller individidentifiering. Museet startade med DNA-baserad artidentifiering av fisk redan 2002 och 2006 sattes ett projekt igång för att bygga ett DNA-referensbibliotek över alla Sveriges ryggradsdjur. Projektet lade grunden till webbportalen Svenska DNA-nyckeln ([www.dnanyckeln.se](http://www.dnanyckeln.se)) som lanserades förra året.

Det var först 2003 som några forskare fick gehör i hela världen för att en specifik standardiserad genregion för djur skulle kunna bli

ett nytt artidentifieringsverktyg. Detta kom att kallas DNA Barcoding på engelska, vilket ibland försvenskas till DNA-barkodning eller ersätts med termen DNA-streckkodning.

Genregionen man bestämde sig för vad gäller djurriket var den mitokondriella genen cytochrom c oxidas subenhet I (COI). Då det visade sig att genen inte fungerade lika bra på alla organismgrupper, valde man senare andra gener för växter (kloroplastgenerna *rbcl* och *matK*) och svampar (den nukleära genen ITS).

För att kunna artidentifiera ett okänt prov behöver man först sekvensera den överenskomna genregionen – streckkoden. Hur det går till och viktiga saker att tänka på i anslutning till detta beskrivs av Markus Englund på s 17-18. DNA-sekvensen man får vid sekvenseringen är en textsträng med bokstäverna A, C, T, och G i olika kombinationer och upprepningar genom den cirka 650 bokstäver långa textsträngen. Bokstäverna representerar de fyra ämnena, så kallade kvävbaser (Adenin, Guanin, Cytosin och Tymin), som binder ihop DNA strukturen (den så kallade dubbelhelixen). Textsträngen med bokstäver kan man kopiera och klistra in i webbportalen Svenska DNA-nyckeln eller den internationella motsvarigheten BOLD för att jämföra med alla kända sekvenser i referensbiblioteket. Från detta kan en artbestämning oftast levereras, eller en

artlista om det var flera prover, förutsatt att organismgruppen för din okända sekvens finns väl representerad i referensdatabasen.

Man kan också få svaret att någon nära matching inte kunde hittas varför någon artidentifiering inte heller gick att göra. Det beror oftast på att arten som det okända provet kommer ifrån ännu inte har lagts in i referensbiblioteket. För att en webbportal som svenska DNA-nyckeln ska fungera väl krävs det en databas med ett välfyllt referensbibliotek av kända DNA-sekvenser som okända sekvenser kan jämföras med. Det krävs att referensbiblioteket innehåller sekvenser av hög kvalitet, helst av alla arter i en organismgrupp som en okänd sekvens kan tänkas komma ifrån, och att de är kopplade till bevarade referensexemplar (se nästa sida).

Vad gäller svenska DNA-nyckeln så har denna webbportal just lanserats och referensbiblioteket är i uppbyggnadsfas. Just nu finns ganska bra täckning av svenska ryggradsdjur, framförallt fåglar, däggdjur och fiskar. För många andra djurgrupper, till exempel insekter så kommer man i svenska DNA-nyckeln ofta få svaret att inga liknande sekvenser kunde hittas och därför kunde ingen artbestämning göras. Man kommer då att rekommenderas att klicka på en länk

till den internationella sidan BOLD som har en större representation men som inte automatiskt kan filtrera fram till exempel kända svenska arter. Ibland kan det också vara så att närstående arter inte går att skilja från varandra med streckkoden varvid flera artnamn kan komma upp som svar på den okända sekvensen. Detta händer inte lika ofta med en geografiskt begränsad databas som Svenska DNA Nyckeln, men kan hända om man istället använder den "långsamare" ribosomala genen 16s (som också finns i webbportalen). För att arter ska gå att skilja åt med en gen som förändras långsamt, krävs att de varit skilda åt en längre tid än för en snabbare evolverande gen.

Exempel på prover att identifiera i en biologilab kan vara en fjäder från skolgården, spillning från närmaste skogs/ängsmark (helst av växtätare eftersom rovdjursspillning kan ge en oläslig sekvens med en blandning av både rovdjuret och dess byte om man använder standardprotokoll) eller en fiskfilé från någon matbutik eller restaurang. Det kan också gärna vara insekter som håvas in eller samlas in med någon slags fälla, men det innebär att djuren avlivas. Djurprover är att föredra då referensbibliotek för växter än så länge är ganska dåliga. ►

# Varför heter det DNA-streckkodning?



Exempel på referensexemplar som finns bevarade på Naturhistoriska riksmuseet är blåkråka till vänster (Cat. id. NRM 20106015 ) och morkulla till höger (Cat. id. NRM 20046331), som de illustrerade streckkoderna nedan har sek-

venserats ifrån. I detta fall var det en morkulla från Fibysjön utanför Uppsala och en felflugen blåkråka från Ramsberg norr om Lindesberg. Bilderna går att hitta på den museigemensamma samlingsdatabasen [www.naturarv.se](http://www.naturarv.se).

En liten del av DNA-streckkodsekvensen från en blåkråka, som totalt omfattar cirka 650 bokstäver.

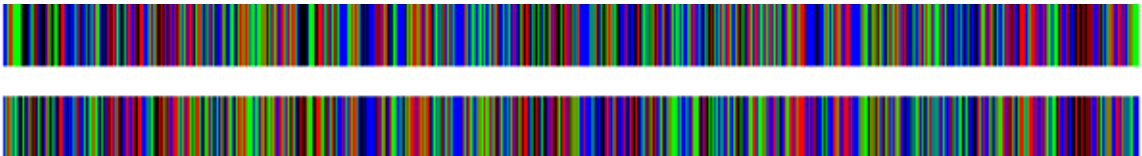
```
CTAATTTTTGGGGCCTGAGCGGGCATGGTTGGAACCGCCCTCAGCCTGCTCATTGCGCGAGAACTCGGTC  
AACCAGGAACCCTACTAGGAGACGACCAGATCTACAACGTAATCGTCACTGCCCATGCCTTCGTAATAATCT  
TCTTTATAGTCATACCAATCATAATCGGGGGCTTTGGAAACTGACTAGTCCCCCTTATAATCGGCGCCCCA...
```

En liten del av DNA-streckkodsekvensen från en morkulla, som totalt omfattar cirka 650 bokstäver.

```
CTAATCTTCGGTGCATGAGCTGGCATGGTCGGAACCGCCCTCAGCCTGCTTATTCGTGCAGAACTAGGCCAA  
CCAGGAACCCTCTTGGGAGATGACCAAATCTACAATGTAATCGTTACTGCTCATGCATTGTAATAATTTTCTT  
CATAGTTATACCAATCATGATCGGAGGATTTGGAAATTGACTAGTCCCACTCATAATCGGCGCCCCCGACAT..
```

En DNA-sekvens kan visualiseras på ett mer överskådligt sätt om man låter linjer i fyra olika färger representera ordningen av kvävebaserna A, C, T och G i genen. Då åskådliggörs också analogin med vanliga streckkoder för till exempel produkter och varför det kallas "DNA-streckkodning".

Bilderna nedan visar DNA-streckkoden för blåkråka överst och morkulla nederst med samma sekvenser som ovan fast för hela den 650bp långa sekvensen. Ser du skillnad? (Tips titta till exempel alldeles i början av streckkoden)



Färgkoden till bilden ovan är:

Tymin (T) = grön

Guanin (G) = svart

Adenin (A) = röd

Cytosin (C) = blå

Text och bilder är delvis hämtade från Svenska DNA nyckeln på [www.dnanyckeln.se](http://www.dnanyckeln.se) (texter skrivna av Johannes Bergsten). Där går det att läsa mer om bakgrund och tillämpningar.