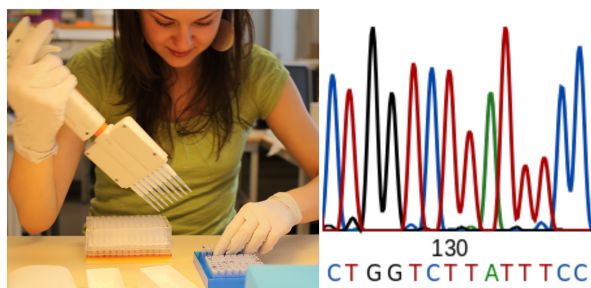


Streckkodning av den svenska floran och faunan – förutsättningar och utmaningar

Rasmus Hovmöller, Mattias Forshage och Fredrik Ronquist

PM från Naturhistoriska riksmuseet. 2017:1



DNA analys



Provtagning



By Everaldo Coelho and YellowIcon

Matchning mot referensbibliotek

Artlista

Chironomus aprilius
Chironomus pallidivittatus
Chironomus plumosus
Chironomus pseudothummi
Chironomus salinarius
Cladopelma virescens
Cladotanytarsus difficilis
Cladotanytarsus mancus
.....



Naturhistoriska
riksmuseet



Detta PM redovisar resultaten från ett utredningsuppdrag finansierat av Naturvårdsverket gällande förutsättningarna för DNA-streckkodning av svensk fauna och flora för miljöövervakning.

Eventuella frågor angående rapporten besvaras av:

Fredrik Ronquist
Niclas Gyllenstrand
Johannes Bergsten
Naturhistoriska riksmuseet
Box 50007
104 05 Stockholm
E-post: fornamn.efternamn@nrm.se

Rapporten bör citeras
Hovmöller, R., Forshage, M. & Ronquist, F. 2017. Streckkodning av svenska floran och faunan – förutsättningar och utmaningar. PM från Naturhistoriska riksmuseet. 2017:1. Naturhistoriska riksmuseets småskriftserie.

ISSN: 0585-3249

Innehållsförteckning

| | |
|---|-----------|
| Sammanfattning | 4 |
| Bakgrund | 6 |
| Streckkodning – ursprung och användningsområden | 6 |
| Integrerad taxonomi | 7 |
| Metastreckkodning | 7 |
| Streckkodsmarkörer | 9 |
| Nya möjligheter med NGS | 10 |
| Nuläget för DNA-referensbibliotek över den svenska floran och faunan | 12 |
| Befintliga databaser | 12 |
| Streckkodning i Sverige | 12 |
| Ett avgörande läge | 13 |
| Metoder för lägesstudien | 14 |
| Taxonomisk översikt | 16 |
| Djur | 16 |
| Växter och svampar | 19 |
| Orsaker till låg grad av streckkodning | 20 |
| Begränsningar hos BOLD | 21 |
| Ett svenskt streckkodningsinitiativ | 22 |
| Kunskapsläget | 22 |
| Tillgängligt material | 22 |
| Zoologiska museisamlingar | 22 |
| Nyare zoologiska insamlingsprojekt | 23 |
| Botaniskt material | 25 |
| Material från amatörbiologer | 25 |
| Framtagning av referensekvenser | 26 |
| Det fysiska arkivet | 27 |
| Det digitala arkivet | 28 |
| Slutord | 30 |
| Tack | 30 |
| Ordlista | 31 |
| Referenser | 34 |
| Bilaga | 36 |

Sammanfattning

Genetiska metoder för att analysera den biologiska mångfalden i miljöprover har utvecklats snabbt de senaste åren. I allmänhet bygger de på användning av nya, massivt parallella sekvenseringsmetoder ("Next Generation Sequencing" – NGS) och vissa genetiska markörer som är lämpade för dessa analyser, så kallade DNA-streckkoder. Med hjälp av referensbibliotek för de aktuella markörerna kan de genetiska analysresultaten översättas till kvalitativa och delvis också kvantitativa mått på förekomsten av namngivna arter. Översättningen till artnamn med hjälp av referensbibliotek är kritisk för att vi ska kunna koppla de genetiska resultaten till den befintliga kunskapen om faunan och floran.

Hittills har det inte funnits någon större systematisk ansats att bygga referensbibliotek för genbaserad miljöövervakning av den svenska floran och faunan. Trots det finns det många referenssekvenser som är relevanta för miljöövervakningen i Sverige. De är bland annat framtagna i olika forskningsprojekt och i streckkodningsprojekt i våra grannländer, och finns tillgängliga i internationella databaser som INSDC/GenBank (för alla sorters sekvenser), BOLD (för många olika streckkodsmarkörer), SILVA (för ribosomala markörer med särskilt fokus på mikroorganismer) och UNITE (för markören ITS på svampar). Vi genomförde en analys av tillgängliga data fokuserad på BOLD, den största streckkodsdatabasen. För den dominerande markören är referensbiblioteken i BOLD nästan kompletta för vissa grupper, som ryggradsdjur, och många andra grupper är också relativt väl täckta. Mycket arbete återstår dock för de flesta artrika grupper i vår flora och fauna, även för den dominerande markören. Särskilt för mindre välstuderade grupper finns det också många sekvenser i BOLD som inte är kopplade till något taxonomiskt namn utan bara till en molekylärt definierad grupp (så kallad BIN). Utan ytterligare taxonomiskt arbete är dessa sekvenser inte användbara i ett referensbibliotek.

NGS-metoderna har inneburit ett nytänkande när det gäller både analyserna av miljöprover och framtagandet av referensbiblioteken. Det blir nu allt vanligare att flera olika markörer används i analyserna, och inom en snar framtid kan det bli vanligt att sekvensera större regioner av genomet, som mitokondriegenomet eller kloroplastgenomet. Det här innebär att den taxonomiska upplösningen kommer att öka betydligt, så att inte bara arter utan också olika populationer kan åtskiljas. NGS-metoderna utvecklas snabbt och det mesta talar för att de inom kort också kan användas för att snabbt och billigt ta fram referensbibliotek för nya markörer, om man har tillgång till ett DNA-arkiv över de arter och populationer man vill studera.

Formas har tidigare finansierat streckkodning av svenska ryggradsdjur, vilket gör att vi i Sverige idag har ett embryo till ett nationellt DNA-arkiv. Det hanteras för närvarande av Centrum för genetisk identifiering (CGI) vid Naturhistoriska riksmuseet (NRM). Arkivet omfattar vävnadsprover och/eller DNA-extrakt av de allra flesta arter av svenska ryggradsdjur, samt en del andra grupper som varit föremål för forskning eller pilotprojekt inom miljöövervakningen, däribland fjädermyggor. Många grupper som streckkodats av svenska forskare har sekvenserats av Biodiversity Institute of Ontario (BIO) och DNA-extrakten finns idag kvar i frysar i Kanada. Att ha vävnadsprover eller DNA-extrakt i Sverige skulle dock vara en fördel för svenska forskare och miljöövervakare, till exempel vid framtagandet av referensbibliotek för nya markörer eller vid djupare genetiska analyser av materialet.

Tack vare Svenska artprojektet har vi i Sverige ett rikt och taxonomiskt välstuderat material som fortfarande endast till liten del streckkodats. Bara från Svenska artprojektets marina inventering och från Svenska Malaisefälleprojektets inventering av svenska insektsfaunan finns idag material av 5 000 arter som skulle kunna streckkodas och inkorporeras i ett nationellt DNA-arkiv. Närmare 4 000 av dessa arter har ännu inte någon referenssekvens i BOLD från nordiskt material.

Tills nyligen har referensbibliotek över streckkoder uteslutande tagits fram med traditionell Sanger-sekvensering. Den närmaste tiden finns det dessutom stora möjligheter att få svenskt material streckkodat av BIO utan kostnad genom samarbete med de norska (NorBOL) och tyska (GBOL) streckkodningsprojekten. På kort sikt är det därför lämpligt att utnyttja dessa samarbeten vid

streckkodning av svenskt material, och fokusera svenska insatser på informationshantering och logistik, bland annat för att försäkra att såväl referensexemplar som associerade vävnadsprover eller DNA-extrakt bevaras i svenska naturhistoriska samlingar. På längre sikt, när NGS-metoder tar över vid framtagningen av referensbiblioteken, kan det bli mer kostnadseffektivt att utnyttja de svenska sekvenseringsplattformar som idag erbjuder NGS-sekvensering till konkurrenskraftiga priser. Då skulle också informationshanteringen och logistiken bli enklare. Fördelar och nackdelar med de olika alternativen behöver kontinuerligt omvärderas allt eftersom tekniken, kostnaderna och övriga förutsättningar förändras.

Nätverket SweBOL har på senare år bidragit till att de svenska aktörerna är bättre samordnade. Flera svenska forskargrupper ligger också långt fram i utvecklingen av NGS-metoder för miljöövervakning, och genetiska metoder är numera ett temaområde inom den fortlöpande miljöanalysen på Sveriges Lantbruksuniversitet.

Sverige har alltså goda förutsättningar idag för att på kort tid se till att de svenska referensbiblioteken blir betydligt mer kompletta än idag. Det som saknas är i första hand personalresurser för att organisera och leda arbetet med att ta fram referensbibliotek för de mest använda markörerna och samtidigt bygga ett nationellt DNA-arkiv omfattande DNA-extrakt och vävnadsprover för framtida genetiska analyser av de streckkodade beläggexemplaren. Det finska streckkodningsprojektet FinBOL har visat att det med små personalresurser går att åstadkomma imponerande resultat, och förutsättningarna att få till stånd en kostnadseffektiv insats i Sverige som kan föra den genbaserade miljöövervakningen framåt är synnerligen goda.

Bakgrund

Den snabba utvecklingen av molekylärbiologiska metoder har gett miljöövervakare och taxonomer många nya verktyg för att inhämta och analysera genetisk information från arter, populationer och miljöprov (t.ex. Port m.fl., 2016; Thomassen m.fl., 2012; Valentini m.fl., 2016; Yoccoz m.fl., 2012). Just nu är vi mitt i den omvälvning som DNA-baserade tekniker har möjliggjort.

Molekylära analyser för att särskilja arter har utförts ända sedan molekylärbiologins pionjärtid under 1900-talets första årtionden. Ett genombrott skedde då Theodosius Dobzhansky och Alfred Sturtevant 1937 använde kromosomkartor för att upprätta ett släkträd över sjutton genetiska linjer av dagflugan *Drosophila pseudoobscura*. Den första riktiga sekvensanalysen publicerades 1963, då Margoliash m.fl. lyckades ta fram aminosyresekvensen från enzymet Cytokrom c från så skilda arter som människa, häst, gris, kanin, tamhöna, tonfisk och bagerijäst och beskrev skillnader och likheter mellan sekvenserna och hur det kunde återspegla släktskap.

I och med att PCR (Polymerase Chain Reaction, se ordlistan) och DNA-sekvensering blev allmänt tillgängliga metoder under 1990-talet har analyser av molekylära data blivit ett av de viktigaste verktygen för att rekonstruera släkträd och avgränsa arter. Samtidigt utvecklades metoder inom metagenomik, där en genetisk markör används för att analyseras artdiversiteten i ett ekologiskt sammanhang. Redan tidiga analyser (t.ex. Pace m.fl., 1985) visade att den dolda diversiteten bland bakterier var flera tiopotenser större jämfört med vad som gick att få fram genom odling. Även bland växter och djur har diversiteten ibland visat sig vara högre än väntat, då molekylära metoder kan påvisa kryptiska arter (se ordlistan). I en klassisk studie undersökte Smith m.fl (2006) parasitflugor (Tachinidae) som kläckts från fjärilslarver. Med streckkodning kunde de visa att det som antagits vara 16 arter av generalister antagligen borde splittras till 73 arter parasitflugor med snäva värdpreferenser.

Sedan 1982 har sekvenser deponerats på databaser inom INSDC (International Nucleotide Sequences Database Collaboration). I dagligt tal används ofta namnet "GenBank", som är den databas inom INSDC som drivs av amerikanska NCBI (National Center for Biotechnology Information), men INSDC omfattar också ENA (European Nucleotide Archive) och DDBJ (DNA Data Bank of Japan). Databaserna inom INSDC delar information med varandra och kan ses som en samlad enhet. Kombinerat med sökalgoritmen BLAST (Altschul m.fl, 1990) är det möjligt att identifiera nya eller okända sekvenser genom att de matchas och rankas i likhet mot de befintliga sekvenserna inom INSDC.

Streckkodning – ursprung och användningsområden

Benämningen streckkodning ("DNA barcoding") lanserades i en ofta citerad artikel av Hebert m.fl, 2003. Författarna jämförde DNA-sekvenser med streckkoder på butiksvoror: ett fåtal element (streck och mellanrum av olika bredd, eller olika nukleotider i DNA) kan bilda tillräckligt många olika kombinationer för att ge varje vara eller art en egen unik kod. Tanken är att DNA-streckkoderna ska vara jämförbara sinsemellan, lätta att sekvensera över stora delar av organismvärlden, och innehålla tillräcklig diversitet för att skilja även närstående arter. Författarna föreslog att forskarsamhället gemensamt skulle satsa på att sekvensera markören COI (Cytokromoxidas I, även förkortad COX eller CO1) över hela djurriket. De visade som illustration att COI kunde användas för att särskilja 200 arter av närbesläktade fjärilar. Senare forskning har visat att för organismgrupper som svampar och kärlväxter finns det andra markörer som är betydligt lämpligare för streckkodning än COI. Även för identifiering av djur finns det alternativa markörer som delvis har bättre egenskaper än COI. Att särskilja arter med DNA-streckkoder är inte heller alltid lätt, men det pågår intensiv utveckling av

analysmetoder för att förbättra avgränsningen av arter med hjälp av data från en eller flera DNA-streckkoder.

2004 etablerades Consortium for the Barcode of Life (CBOL) som ett internationellt samarbetsorgan, och projektet iBOL för initiativet att streckkoda så stor del av jordens liv som möjligt. Inom detta paraplyinitiativ har sedan ett stort antal delprojekt genomförts eller startats. Bland dessa finns flera nationella streckkodningsinitiativ (bland annat i Norge och Tyskland), men också systematiskt avgränsade projekt, där somliga som avser begränsade grupper (bin, koraller, fjärilar, stickmyggor, borrhugor, nattsländor, däggdjur, fiskar, och liv i polartrakterna) ofta har kommit långt, medan andra med vidare omfattning (svampar, marina organismer, etc) har betydligt längre kvar. För datahantering har Barcode of Life Data Systems (BOLD) lanserats. Det är en portal och databas för streckkoder med många funktionaliteter (se <http://boldsystems.org/>).

Integrerad taxonomi

Traditionell taxonomi bygger på att arter identifieras morfologiskt i första hand; de särskiljs från sina närmaste släktingar genom synliga skillnader i form, storlek och färg. Ofta är detta svårt, och ibland kan den morfologiska variationen rentav vara missvisande i det här avseendet. Ett inte ovanligt problem är kryptiska arter, som ser identiska ut morfologiskt men är åtskilda genetiskt.

Kombinationen av många olika typer av information i arbetet med att upptäcka och avgränsa arter kallas integrerad taxonomi. Genetiska data spelar här en central roll. Genetiska metoder är särskilt effektiva inom organismgrupper där det finns många svåridentifierade arter och gott om material, men där flaskhalsen i upptäckandet av nya arter är tidsåtgången att sortera och identifiera materialet för att utesluta redan beskrivna arter. Exempel på ett sådant material är små insekter som samlas i icke-selektiva fällor, t.ex. Malaisefällor, men också marina evertebrater som samlas med botten-skrapa. Det finns flera olika sätt att använda genetiska metoder för att analysera artsammansättningen och påskynda upptäckten av nya arter i den här typen av material, och metoderna är under snabb utveckling.

Att avgränsa arter med hjälp av genetiska data är inte trivialt. Hebert m.fl. (2003) betonade att skillnaden mellan arter i en markör som COI tenderar att vara betydligt större än skillnaderna inom arter. Sålunda kan man identifiera en kritisk tröskel, streckkodströskeln ("the barcoding gap"), där skillnaderna blir så stora att det måste handla om olika arter. Hebert m.fl. föreslog att denna tröskel låg på 3 % skillnad för fjärilsarter, men påpekade att nivån kunde variera mellan organismgrupper, och mycket forskning har sedan dess fokuserat på att hitta den relevanta streckkodströskeln för olika grupper. Numera finns det statistiska metoder för att avgränsa arter med hjälp av streckkodsdata oberoende av tröskelvärden, t.ex. General Mixed Yule Coalescent (Pons m.fl., 2006; Monaghan m.fl., 2009). Det står samtidigt klart att data från ett par olika markörer ger betydligt bättre möjligheter att avgränsa arter än data från bara en markör (Yang m.fl., 2010; Zhang m.fl., 2011, 2014).

Metastreckkodning

Metastreckkodning ("metabarcoding" på engelska) kombinerar DNA-baserad identifiering med de nya metoder för DNA-sekvensering som blivit tillgängliga de senaste åren, och som kan läsa av tusentals eller miljontals sekvenser samtidigt. De nya metoderna har radikalt sänkt tiden och kostnaden för sekvenseringsanalyserna. De nya metoderna kallas på engelska för "Next Generation Sequencing" och vi använder förkortningen NGS här. Andra vanliga beteckningar på engelska är "massively parallel sequencing" och "high-throughput sequencing". Den traditionella sekvenseringsmetoden, som analyserar en sekvens i taget, betecknas som Sanger-metoden.

Metastrekkodning har de senaste åren börjat användas för att undersöka artsammansättningen i miljöprover. Metoden kan vara känslig nog att detektera arter som inte alls representeras av några levande organismer utan endast genom DNA-spår i form av avskavda celler och andra rester från individer som finns eller har funnits i miljön som provet kommer från (jorden, vattnet). I en meta-strekkodningsstudie extraheras DNA direkt ifrån provet (jord- eller vattenprov), eller från organismerna i provet (fällprover från fallfällor, Malaisefällor etc), eventuellt efter att organismerna grovsorterats. Ur DNA-extraktet amplifieras sedan en eller flera strekkodsmarkörer (exempelvis COI och 16S för djur, ITS för svamp) med vanlig PCR-teknik (se ordlista). De DNA-strängar man amplifierar upp (som kallas amplikon) blir då en blandning av den amplifierade markören eller markörerna från alla organismer i provet.

Efter en NGS-körning av ett amplikon erhåller man en stor uppsättning sekvenserade DNA-strängar. Varje sekvenserad sträng kallas för en läsning ("read" på engelska) och mängden av dem kallas för ett bibliotek (ej att förväxla med ett referensbibliotek). Efter att fel har rättats grupperas läsningarna efter hypotetiska arthenheter eller OTUer (efter engelskans "Operational Taxonomic Unit"). Varje OTU behöver sedan matchas sedan mot en databas med referenssekvenser för att generera en artlista från provet. Om referensdatabasen är inkomplett, vilket ofta är fallet, kan det bli många okända taxa kvar i listan efter matchningen.

Metastrekkodning har redan använts en längre tid för att studera den mikroskopiska mångfalden, det vill säga bakterier och mikroskopiska djur, växter, svampar och andra eukaryoter. Tekniken har radikalt ändrat förutsättningarna för forskningen om dessa organismer. Flera undersökningar (exempelvis Handelsman m.fl., 1998) har visat att mångfalden av mikroorganismer är flera tiopotenser större än vad som tidigare varit känt. Mörkertalet hänger samman inte minst med att den stora majoriteten av bakterier och svampar inte går att odla på syntetiska medier, och på så sätt kunnat undgå upptäckt.

Metastrekkodning har visat sig vara användbart också för makroskopiska organismer. Metoden har exempelvis använts för att studera: leddjur i fällprover (översikt av Brandong-Mong m.fl., 2015), växter (Yoccoz m.fl., 2012), DNA-spår av groddjur och fisk från vattenprover (Ficetola m.fl., 2008; Goldberg m.fl., 2014; Thomsen m.fl., 2014), spår av däggdjur i jord (Andersen m.fl., 2012), och björnars matpreferenser (Stenset m.fl. 2016). Det tillkommer hela tiden nya exempel på tillämpningar inom miljöövervakning och ekologiska undersökningar.

Jämfört med traditionell inventering kan metastrekkodning vara en snabb och billig metod vid stora provvolymen eller när svåra, artrika grupper ska inventeras. Ofta är upplösningen också betydligt bättre. De genetiska metoderna kan till exempel identifiera kryptiska arter och livsstadier som är mycket svåra eller omöjliga att bestämma med traditionella metoder. Med lämpliga markörer kan man också få upplösning ner till populationsnivå, vilket är svårt eller omöjligt med traditionella metoder. För studier på populationsnivå handlar det dock oftast om andra protokoll än för "vanlig" metastrekkodning.

Att analysera all e-DNA-data från olika platser och olika tillfällen på art- eller populationsnivå kan ge mycket information om miljön men är också krävande tack vare den stora mängden data och den slumpmässiga variationen över tid och rum. Om det finns enskilda taxa som man har särskild anledning att övervaka (hotade arter, invasiva arter, avgörande funktionella grupper i ekosystemet etc) kan man sålla fram data för dessa kontinuerligt i den mån de ingår i referensbiblioteket. Det går också att mäta mångfalden på en mer övergripande nivå med olika diversitetsmått, och sedan följa utvecklingen på denna nivå över tid.

Genom möjligheten att associera alltfler sekvenser från genetiskt analyserade miljöprov med taxonomiska namn i ett växande referensbibliotek, långt efter att provet har sekvenserats, kan man också kontinuerligt öka kunskapen om dåligt kända arters uppträdande och habitatkrav, och därmed kan man identifiera deras roll som miljöindikatorer, och identifiera arter som kan vara hotade, o.s.v. Som alltid i miljöövervakningen är det kritiskt att ha tillgång till långa provtagningsserier och det finns all anledning att fundera de närmaste åren både över hur gamla övervakningsprogram kan växla över till att använda genetisk analysteknik, och vilka nya, genetiskt baserade övervakningsprogram, som kan bli aktuella att sätta igång.

Strekkodsmarkörer

Traditionell strekkodning av djur använder sig av en 658 baspar lång del av COI, en mitokondriell gen. COI lyftes fram som en lämplig markör redan i de första artiklarna om strekkodning (Hebert m.fl. 2003 osv.). Denna gen uppfyller många av de krav man kan ställa på en lämplig markör för strekkodning:

- Mitokondriellt DNA finns i många upplagor per cell, vilket underlättar DNA-extraktion och sekvensering under förhållanden som inte är optimala, till exempel vid analys av delvis nedbrutet material.
- Genen finns hos alla organismer som har mitokondrier.
- Det finns konserverade delar som ser nästan likadana ut hos alla arter, vilket gör det möjligt att amplifiera genen med ett och samma primerpar, så kallade universella primers. Senare studier har dock visat att det är betydligt svårare än man ursprungligen trodde att hitta riktigt bra generella primerpar för COI.
- Mitokondriellt DNA nedärvs endast på modernet (med vissa undantag, exempelvis musslor (Zourous m.fl, 1992)) och påverkas därför inte av rekombination (utbyte av genetisk information mellan kromosomer).
- Det mitokondriella genomet har en enklare struktur än kärngenomet och låter sig jämföras lättare mellan arter i och med att det saknar så kallade introner.

COI har också visat sig ha vissa nackdelar, bland annat att det hos djur ibland finns pseudogener av COI. Det är icke-fungerande kopior i det nukleära genomet som inte har samma fylogenetiska historia som den fungerande mitokondriella genen. Sekvenseras en pseudogen i stället för den äkta genen kan det ge missvisande resultat. Det har också visat sig svårt att i praktiken hitta bra primerpar med bred taxonomisk täckning.

Efter COI är 16S den mest använda markören för strekkodning av djur. Till skillnad från COI kodar 16S inte för ett protein utan för en RNA-sträng som ingår i ribosomer; 16S kallas därför ribosomalt DNA (rDNA). Även andra ribosomala markörer som 12S rDNA, 18S rDNA, D-regionen av 28S r-DNA och ITS (från engelskans "Internal Transcribed Spacer", en utfyllnadsregion mellan olika ribosomala gener) har föreslagits som strekkodningsmarkörer för djur. 12S och 16S är mitokondriella markörer som COI, vilket innebär liknande fördelar och nackdelar. 18S, 28S och ITS är däremot nukleära markörer som nedärvs genom andra mekanismer än mitokondrier, vilket gör att de kan berätta en annan fylogenetisk historia. När man ska skilja närstående arter åt har det visat sig vara synnerligen värdefullt att ha tillgång till flera markörer som nedärvs oberoende av varandra. Ribosomala gener har fördelen jämfört med proteinkodande gener att evolutionshastigheten varierar mer över sekvensen. Ribosomala gener har både starkt konserverade regioner, som gör det lätt att hitta universella primerpar, och hypervariabla regioner som kan användas för att skilja arter och populationer åt. Ribosomalt DNA medför emellertid också en särskild typ av svårigheter, eftersom längdskillnader i sekvenserna gör det problematiskt att jämföra dem i detalj och lista ut hur

evolutionen gått till. Detta ställer till problem vid rekonstruktion av släkträd men är av mindre betydelse vid artidentifiering.

COI förekommer även i växters mitokondrier, men det har visat sig att genen förändras för långsamt hos växter för att fungera för streckkodning (Kress m.fl., 2005). För gröna växter (landväxter och grönalger) är standardstreckkodningsregionen istället en del av kloroplastgenomet och utgörs av generna *rbcl*, *matK* och *trnL* (Hollingsworth m.fl., 2009; Taberlet m.fl., 2007). Hollingsworth m.fl. (2011) utvärderade ett större antal kloroplastmarkörer som kandidater till streckkodningsregion för gröna växter, men det visade sig vara svårt att hitta enstaka markörer som både är användbara för att identifiera växter till artnivå och möjliga att amplifiera med universella primerpar. *rbcl+matK* är en kompromisslösning, och utvecklingen går snarare mot att sekvensera fler markörer eller att använda hela kloroplastgenomet än att förlita sig på några enstaka markörer (Parks m.fl., 2009; Nock m.fl., 2011). BOLD tillåter numera uppladdning av hela kloroplastgenomet som streckkod för de som önskar.

Dentinger m.fl. (2011) jämförde COI och ITS som streckkodningsmarkörer för svamp. ITS förekommer i många kopior i kärnngenomet och visade sig fungera bättre än COI för svamp. ITS har en mer distinkt streckkodningströskel än COI och markören är också mer variabel, vilket gör den något bättre på att urskilja arter. Till skillnad från hos gröna växter och djur innehåller svampars COI-gen ofta långa introner (upp till 1500 baspar), något som komplicerar amplifiering från herbariematerial där DNA är fragmenterat.

Nya möjligheter med NGS

När DNA-streckkodning lanserades 2003 byggde all DNA-sekvensering på Sanger-metoden. NGS-metoderna slog igenom långt senare, och har ställt många tidigare idéer på huvudet. De NGS-metoder som används mest för metastreckkodning idag kan inte läsa så långa sekvenser som använts tidigare i streckkodsmarkörerna. Det har gjort det nödvändigt att arbeta med mindre delar av de tidigare markörerna, "ministreckkoder". Å andra sidan har det blivit lätt att samtidigt ta fram data för flera olika markörer, vilket har ökat intresset för att utvidga antalet streckkodsmarkörer.

Tills nyligen ansågs det nödvändigt att ta fram referensbibliotek med Sanger-metoden eftersom en NGS-läsning är förknippad med betydligt större risk för enstaka fel i sekvensen. Genom att läsa samma sekvens upprepade gånger kan man emellertid öka precisionen i NGS så att den överträffar Sanger-metoden. Antalet läsningar i en NGS-körning är nu så stort att man kan ta fram tusentals referenssekvenser i en enda körning. Med andra ord kan man bygga upp ett referensbibliotek för en ny markör mycket snabbt och till en låg kostnad – om man har ett DNA-arkiv över de aktuella arterna och populationerna, med extrakt förberedda för NGS-analyser.

Idag pågår en utveckling mot NGS-baserad metastreckkodning som inte baseras på PCR-amplifiering av utvalda markörer. I stället använder man NGS-metoder för att sekvensera hela eller delar av genomet, till exempel hela mitokondriegenomet eller hela kloroplastgenomet. Metoden kallas shotgun-sekvensering: många små, slumpmässigt valda bitar av genomet sekvenseras på måfå och pusslas sedan ihop till större regioner genom avancerad dataanalys. Än så länge kan shotgun-metoder inte konkurrera prismässigt med amplikon-baserade metoder för metastreckkodning, men det kan komma att ändras framöver. Intressant nog är shotgun-metoder precis som amplikon-baserade metoder starkt beroende av referensbibliotek. Det är svårt att pussla ihop en blandning av olika mitokondriegenom eller kloroplastgenom från många små sekvenssnuttar utan att ha tillgång till åtminstone några av de genomsekvenser man sannolikt kommer att påträffa i blandningen. Ju fler genomsekvenser man har tillgång till, desto lättare blir analysen av det blandade provet.

Kloroplastgenom kan man utan vidare rekonstruera direkt från shotgun-sekvensering av växt-DNA eftersom varje växtcell innehåller så många kloroplaster att kloroplast-DNA totalt dominerar NGS-läsningarna. Mitokondrier går också att direktsekvensera med shotgun-metoden utan PCR (Quispe-Tintaya m.fl., 2015), men det kräver anrikningsprocedurer och/eller större mängder material, vilket gör metoden mindre lämplig för små djur och museimaterial. Ett alternativ är att amplifiera hela mitokondriegenom med PCR, men metoden innebär liknande nackdelar som amplikonbaserad metastreckkodning, nämligen att fel kan uppstå i amplifieringen och att primerparen kanske inte amplifierar precis alla arter eller populationer man är intresserad av att detektera.

Den NGS-plattform som används mest idag för amplikon-baserad metastreckkodning är Illumina Mi-Seq. En Mi-Seq-körning kostar idag runt 14 000 kr och genererar 18 miljoner läsningar med det vanligaste protokollet. Det är tillräckligt mycket för att analysera mer än hundra miljöprover, beroende lite på hur stor diversitet det finns i proverna och hur viktigt det är att plocka upp de sällsynta sekvensvarianterna i varje prov. Till kostnaden för själva MiSeq-körningen kommer kostnaderna för att förbereda proverna och för databehandlingen. Kringkostnaderna kan vara betydande för enstaka prover eller pilotprojekt, men bör kunna bli låga per prov om man kör analyser i större skala med standardiserade metoder. Databehandlingen är i vilket fall krävande och det finns behov av att utveckla väldokumenterade och lättanvända standardmetoder som är lämpliga för miljöövervakning.

En NGS-analys av ett miljöprov skulle i bästa fall kunna ge en näst intill komplett lista på de OTUer som finns i provet. Begränsningarna sätts dels av primerparens täckningsgrad – förmågan att amplifiera alla arter och populationer av intresse – och dels av markörens taxonomiska upplösning – förmågan att skilja mellan närstående arter och populationer. Numera använder man ofta degenererade primerpar, det vill säga blandningar av snarlika primers. Det kan öka täckningsgraden betydligt. Beräkningar på NRM har till exempel visat att man teoretiskt skulle kunna detektera över 80 % av artdiversiteten av insekter i ett Malaisefälleprouv eller liknande miljöprov med en markör, och över 90 % med två markörer.

NGS-analyser kan alltså komma upp i imponerande prestanda i den kvalitativa analysen av provens taxonomiska sammansättning. Det är svårare att få en kvantitativ analys av antalet individer eller biomassan av olika taxa på grund av den stora variationen i antalet läsningar per individ eller per biomassa mellan olika arter. Därtill kommer vid amplikon-baserad analys kemirelaterade faktorer som påverkar hur effektivt olika sekvensvarianter amplifieras. Det finns dock hopp om att ytterligare forskning kan öka precisionen i den kvantitativa analysen.

Förutom teknisk utveckling krävs det insatser på två områden för att metastreckkodning ska bli riktigt användbar i svensk miljöövervakning. För det första behövs det referensbibliotek över de aktuella markörerna för att det ska gå att identifiera OTUerna med namngivna arter och på så vis koppla de nya analysresultaten till den taxonomiska kunskap vi redan har om den svenska floran och faunan. Med biblioteken som grund kan man också jämföra NGS-resultaten från olika markörer och underlätta framtida shotgun-analyser. För det andra behöver vi jämförande metodstudier som kan hjälpa oss att kalibrera resultaten från de nya genetiskt baserade metoderna mot resultaten vi redan har från existerande övervakningsprogram. Man kan behöva köra ett nytt genetiskt baserat program parallellt med ett existerande program under flera år innan man helt övergår till en ny metodik, så att man inte äventyrar möjligheten att analysera långsiktiga trender som innefattar bägge programmen. I resten av rapporten fokuserar vi på det första området, specifikt de utmaningar vi står inför när det gäller att bygga referensbibliotek för genetiskt baserad miljöövervakning.

Nuläget för DNA-referensbibliotek över den svenska floran och faunan

Befintliga databaser

Många svenska och utländska forskningsprojekt har under de senaste decennierna producerat genetiska sekvenser som är relevanta för ett DNA-referensbibliotek över den svenska floran och faunan. Det har emellertid inte funnits någon svensk infrastruktur för att samla information om det arbete som har gjorts. Eftersom de flesta vetenskapliga tidskrifter kräver att sekvenserna görs tillgängliga, har de i allmänhet deponerats i de internationella sekvensdatabaserna inom INSDC.

INSDC ställer tyvärr rätt låga krav på de metadata som associeras med sekvenserna, även om kraven har skärpts med åren. Det är sällan som en sekvens i INSDC är knuten till detaljerad information om den sekvenserade organismen, till exempel om när, var och hur den samlades in eller i vilken naturhistorisk samling (om någon) man kan hitta det som är kvar av organismen (beläggsexemplar eller med ett ofta använt engelskt ord "voucher"). Ännu mer sällan hittar man en bild av exemplaret eller information om sparade DNA-extrakt. Det mest problematiska med INSDC är dock att så många av de taxonomiska bestämningarna är dåligt dokumenterade eller direkt felaktiga.

BOLD är ett försök att skapa en sekvensdatabas bara för streckkodssekvenser. BOLD har betydligt högre krav på metadata än INSDC, vilket har gjort att BOLD – som drivs av Biodiversity Institute of Ontario (BIO) – på många sätt fungerar bättre som referensbibliotek. För att godkännas som streckkod i BOLD så måste till exempel sekvensfilerna ('trace files') samt information om de PCR-primers som använts skickas in. BOLD har dock tidvis haft väl stränga krav. Exempelvis dröjde det länge innan man accepterade ITS som officiell streckkodsmarkör för svampar, vilket bidrog till att svampforskarna byggde upp en egen databas för ITS, UNITE. UNITE drivs av en grupp vid universitetet i Tartu i Estland, ledd av professor Urmas Kõljalg. För många mikroorganismer är ribosomala markörer överlägsna, och där är de bästa referensdatabaserna SILVA, som drivs av ett tyskt konsortium, och det amerikanska RDP. En stor del av de data som finns i BOLD hör till olika forskningsprojekt som ännu inte gjort sina resultat offentliga. Det gör att endast en del av BOLD-data är offentligt tillgängliga och användbara i miljöövervakningen, även om data successivt släpps. De offentligt tillgängliga sekvenserna i samtliga databaser finns tillgängliga i INSDC, och numera kan sekvenserna där märkas med en särskild flagga om de uppfyller de aktuella metadatakraven för streckkodssekvenser.

Streckkodning i Sverige

För att få en överblick över de referenssekvenser som är relevanta för svensk miljöövervakning behöver man alltså extrahera de data som finns i system som BOLD, SILVA, RDP, UNITE och INSDC. Men även om Sverige hittills har saknat ett samlat initiativ inom det här området, till skillnad från länder som Norge och Tyskland, har vissa större insatser ändå gjorts. NRM var en av många institutioner runtom i världen som deltog i ett internationellt initiativ för att streckkoda världens fiskarter, ett av de första större streckkodningsprojekten i världen.

Formas stödde senare ett NRM-lett projekt för att streckkoda alla svenska ryggradsdjur. De flesta svenska arter av ryggradsdjur är nu streckkodade för COI, och många arter också för 16S. Som resultat av projektet skapades en webbsida för DNA-baserad artbestämning, Svenska DNA-nyckeln (<http://dna-nyckeln.se>), samt fröet till ett nationellt DNA-arkiv, omfattande DNA-extrakten från projektet. DNA-arkivet hanteras sedan 2013 av Centrum för genetisk identifiering (CGI), en facilitet

på NRM som tar emot uppdrag av myndigheter och organisationer gällande DNA-analys av miljöprover och annat biologiskt material. Förutom svenska ryggradsdjur omfattar DNA-arkivet idag också en del andra grupper, till exempel fjädermyggor från ett pilotprojekt fokuserat på bottenprover från Östersjön, finansierat av NV och HaV.

Försök att få fortsatt stöd för streckkodning av den svenska floran och faunan från Formas och Svenska artprojektet har inte varit framgångsrika; streckkodning faller lite grann mellan stolarna för dessa finansiärers ansvarsområden. För att trots detta få till bättre samverkan mellan de aktörer som är intresserade av att utveckla DNA-streckkodning i Sverige bildades för några år sedan nätverket SweBOL, Swedish Barcode of Life (<http://swebol.org/>). SweBOL ska bl.a. sprida information om behovet av att systematiskt bygga svenska referensbibliotek och om potentialen i genetiska metoder i miljöövervakningen och förvaltningen av våra svenska naturresurser. SweBOLs uppdrag är:

- A) att förankra arbetet med att bygga upp en referensdatabas för den svenska faunan och floran hos berörda politiker och myndigheter, potentiella finansiärer, i forskarsamhället, samt hos den intresserade allmänheten.
- B) Stödja projekt som syftar till att ta fram DNA-streckkoder för den svenska faunan och floran genom att tillhandahålla riktlinjer (standarder) bl.a. för:
 - Dokumentation och förvaring av nyinsamlat material som ska DNA-streckkodas
 - Sekvensering av streckkodsmarkörer för olika plattformar
 - Deponering av beläggmaterial och extraherat DNA
 - Deponering av DNA-sekvenser i relevanta samlings- och referensdatabaser (t.ex. BOLD)

Styrgruppen för SweBOL består idag av representanter från NRM, ArtDatabanken (SLU), Göteborgs Botaniska Trädgård och Göteborgs Universitet (GU). Ingen fast finansiering finns men ArtDatabanken och NRM lägger viss arbetstid i nätverket och NRM är tills vidare värd för SweBOLs sekretariat. En stor referensgrupp finns associerad till styrgruppen och nätverket når alla universitet och flertalet museer och relevanta myndigheter i Sverige.

SweBOL har en egen struktur för streckkoder i BOLD, d.v.s. projekt som följer vissa riktlinjer utöver BOLDs (bl.a. bevarande av DNA extrakt och beläggeexemplar med metadata i offentlig svensk naturhistorisk samling) kan läggas i en särskild SweBOL-mapp, vilken för närvarande förvaltas av Göteborgs Naturhistoriska Museum (GNM). SweBOLs ambition är att vidareutveckla och sprida dessa riktlinjer till fler svenska aktörer.

Förutom aktiviteterna inom SweBOL kan nämnas att Sverige för närvarande deltar i en EU Cost Action kallad DNAquanet. Projektet spänner över fyra år och 32 länder deltar. Projektet är framför allt inriktat på teknisk utveckling av provtagning i akvatiska miljöer för DNA-baserade metoder, men fem arbetsgrupper kommer också att hantera frågor om referensbibliotek, datalagging/data-vårdskap, pipelines för miljöövervakning, juridik och myndighetsutövande.

Ytterligare ett nätverk (EDNA – e-DNA) med särskilt fokus på DNA-baserade metoder inom miljöövervakning är under uppbyggnad. En grupp arbetar här med fokus på nationell infrastruktur. EDNA koordineras i nuläget av SLU, som även har DNA-baserade metoder som särskilt tema inom sin FOMA-verksamhet.

Ett avgörande läge

Det är olyckligt att det hittills har saknats resurser för samlad systematisk streckkodning av den svenska floran och faunan, eftersom vi i Sverige har ett unikt material, relativt nyinsamlat och ytterst välbestämt, av många organismgrupper som är dåligt representerade i de referensbibliotek som finns tillgängliga idag. Att vi har detta material kan vi delvis tacka Svenska Artprojektets olika

forskningsprojekt för, som ju haft ett för ändamålet lämpligt fokus på dåligt kända grupper, men framför allt de två stora inventeringar som har genomförts inom ramen för Artprojektet. Det handlar dels om Svenska Artprojektets marina inventering, som genomfördes 2006-2009, och dels Svenska Malaisefällepaketet (SMTP), vars fältedel genomfördes 2003–2006 men där sortering och bestämning av materialet pågår fortfarande.

I Norge har Norska Forskningsrådet satsat avsevärda resurser på systematisk streckkodning av floran och faunan i NorBOL (Norwegian Barcode of Life), och i Tyskland ligger en särskild regeringssatsning bakom ett motsvarande projekt, GBOL (German Barcode of Life). Både NorBOL och GBOL har varit mycket intresserade av det svenska materialet, och det material man har fått tillgång till har skickats till BIO i Kanada för Sanger-sekvensering av COI. När det gäller NorBOL har sekvenseringen gjorts utan avgift, medan GBOL i vissa fall till och med har betalat för det material man har fått tillgång till. I idealfallet skulle man i samband med den här streckkodningen vilja se att referensexemplaren accessionsfördes vid ett svenskt naturhistoriskt museum och registrerades i det nationella svenska systemet för naturhistoriska samlingar, DINA, med länkar till referenssekvenserna. Samtidigt skulle det vara värdefullt om inte bara referensexemplaren utan också vävnadsprover eller DNA-extrakten deponerades i Sverige för framtida genetiska analyser av det streckkodade materialet. I många fall kräver det att man repatrierar DNA-extrakten (frystorkade plattor) från Kanada efter streckkodningen, vilket innebär extra arbete och extra kostnad för den projektansvariga. För de forskare eller andra som driver streckkodningsprojekten finns det inga starka incitament att genomföra dessa steg, och svenska naturhistoriska museer har hittills inte haft möjlighet att avsätta de resurser som skulle krävas för att aktivt söka upp och arbeta med de aktuella projekten och erbjuda tjänsterna utan kostnad. Resultatet är att det inte finns några garantier idag för att beläggsexemplaren är korrekt accessionsförda vid svenska naturhistoriska museer med länkar till sekvenserna. I synnerhet gäller detta de enskilda forskarprojekten som anlitar NorBOL och GBOL, där beläggsexemplar och information lätt kan bli kvar hos forskargruppen. I de flesta fall går det inte heller att få tillgång till DNA-extrakt eller vävnadsprover för vidare genetiska analyser utan att först inhämta den projektansvarigas tillstånd och därefter antingen begära att få DNA-extrakten återsända till Sverige från Kanada, eller beställa ytterligare sekvensering genom labbet på BIO i Kanada.

I den översikt över nuläget vad gäller referensbibliotek för svenska floran och faunan vi presenterar nedan fokuserar vi på den makroskopiska mångfalden (flercelliga växter, svampar och djur) och på data som för närvarande är publikt tillgängliga på BOLD. De referenssekvenser som analyseras här är inte alltid kopplade till information som gör att det går att hitta beläggsexemplaret eller DNA-extrakten i svenska eller utländska samlingar, även om det i princip kan vara möjligt i många fall. En detaljerad analys av detta skulle dock kräva mer resurser än vad som funnits tillgängligt för det här projektet. Vår analys av befintliga referenssekvenser kompletteras med en grov översikt av de stora material som finns i Sverige och som skulle vara tillgängliga för systematisk streckkodning om SweBOL fick resurser för detta.

Metoder för lägesstudien

Vi tog fram en ögonblicksbild av nuläget för streckkodning av den svenska floran och faunan genom att samköra listor över streckkodade arter från BOLD med artlistor över Sveriges djur och växter från ArtDatabanken. En översiktlig presentation av resultaten finns i tabellform nedan (Tabell 1 och 2). De fullständiga resultaten finns i Bilaga 1, som kan laddas ner från <http://www.nrm.se/download/18.2a85cf3215e50a71056a553a/1509373421630/Lyrholm-smaskriftserien.zip>. Vi vill poängtera att det är en ögonblicksbild vi presenterar. Vi har inte kunnat inkludera resultat från pågående projekt eller sekvenser som inte är publika på BOLD än.

Listor över arter med svensk förekomststatus "Bofast och reproducerande" och de högre taxa de hör till har hämtats från Dyntaxa (<http://dyntaxa.se>), och kombinerats med Svenska Artprojektets "bedömning av forskningsbehovet 2015" (för högre taxa) och information om rödlistning 2015 från artfakta.se.

Listor över arter med tillgängliga streckkodssekvenser från Sverige och övriga Norden är hämtade via BOLDs API. De är begränsade inom ett taxon och geografi, exempelvis Acanthocephala[tax] och Sweden|Denmark|Norway|Iceland|Finland[geo]. Från BOLD (<http://boldsystems.org>) har också hämtats listor över arter som förekommer i hela databasen. Där ingår arter som går att få träffar mot genom sekvenssökningar, men där sekvens och exemplardata kan vara dolda för användare, samt arter där streckkodssekvensen är hämtad från exemplar insamlade utanför Norden. Dessa listor är hämtade via funktionen Species List – Progress under fliken Taxonomy på boldsystems.org.

Det är viktigt att notera att de data som hämtats från allmänna sökningar kan gälla OTUer som ännu inte kopplats till något taxonomiskt namn, eller OTUer som kopplats till samma namn i BOLD som vi använder för svenska artstocken, trots att det kan röra sig om genetiskt skilda arter, eller i varje fall distinkta populationer. Även inom vittspridda till synes enhetliga arter finns en geografiskt betingad genetisk variation (Bergsten m.fl., 2012). Referenssekvenser från de allmänna sökningarna är därför av begränsat värde för svensk miljöövervakning och vi fokuserar sålunda i första hand på de BOLD-sekvenser som har kopplats till taxonomiskt namn och som är framtagna från nordiskt material. Avgränsningen till Norden har vi gjort av praktiska skäl, och den kan diskuteras; i många fall kan till exempel nordtyska streckkodssekvenser vara mer relevanta än isländska. Allt eftersom precisionen i de genetiska analyserna ökar, och man vill urskilja mer av den geografiska variationen mellan underarter och populationer, kommer behovet av svenska referenssekvenser att stiga. Vi har därför också analyserat antalet referenssekvenser från svenska exemplar. Vi presenterar bara data för de markörer som har formell streckkodsstatus i CBOL. Om man tittar på andra markörer än dessa så har arbetet med referensbiblioteken generellt sett inte alls kommit lika långt.

För analysen användes program skrivna i skriptspråken Python och Shellscript. I första steget läses listan över arter från Dyntaxa in och annoteras med information från listan över arter med tillgängliga streckkodssekvenser från exemplar insamlade i Norden eller i Sverige. Matchning sker på fullständiga binomen (Släkte + artepitet). Arter matchas enligt kategorierna Sverige (sekvens från exemplar insamlade i Sverige), Norden (sekvens från exemplar insamlade i Norge, Finland, Danmark eller Island), BOLD (arten finns i databasen, men inga sekvenser från nordiska exemplar). Från BOLD räknas endast arter som förekommer i databasen under ett binomen. Sekvenser från icke fullständigt artbestämda individer (t.ex. *Lucanus* sp.), med namn delvis bestående av serienummer, eller arter med osäker bestämning (t.ex. *Lucanus* cf. *cervus*) räknas inte in i sammanställningen.

En möjlig felkälla är att BOLDs taxonomi inte helt överensstämmer med Dyntaxa. Dyntaxa håller generellt högre kvalitet rent taxonomiskt, men där det skiljer sig har BOLDs taxonomi använts för att få en bättre överblick av taxa med tillgängliga streckkoder.

Protozoa i vid bemärkelse är inte fullständigt undersökt, då BOLD och Dyntaxa har mycket olika system för deras taxonomiska indelning. Dessutom är SILVA och RDP mer relevant än BOLD för dessa organismgrupper. Vi har inte analyserat UNITE även om det skulle ha kunnat ge fler relevanta referenssekvenser för svampar än vad som finns i BOLD.

Flera faktorer bidrar till att siffrorna från BOLD kan vara något lägre än det faktiska antalet tillgängliga referenssekvenser. Streckkodningssekvenser för arter i Sveriges fauna som finns publicerade i INSDC ("GenBank"), men inte märkta som streckkod, kan saknas i BOLDs statistik. BOLD skördar inte heller data från INSDC som tillhör icke-officiella men ändå flitigt använda streckkoder som 16S. Detta gör t.ex. att alla de svenska däggdjur som ingick i det Formas-finansierade vertebratprojektet, och som

bara sekvenserades för 16S, saknas i BOLD och därmed i våra analyser. Data från INSDC är svårare att analysera än data från BOLD eftersom metadata (som insamlingslokal och datum) inte finns tillgängliga direkt från databasen, i varje fall inte från äldre poster, utan måste hämtas från den vetenskapliga litteraturen eller andra källor om den ens finns bevarad.

Taxonomisk översikt

Djur

Ryggradsdjuren har nästan fullständig täckning i BOLD. Endast 19 av 483 svenska arter (3,9 %) saknar helt streckkod, och från 342 av 483 arter (70,8 %) finns streckkoder från svenska exemplar (Fig. 1, Tabell 1, Bilaga 1). Streckkodssekvenser från i princip alla Sveriges däggdjur är framtagna, men de svenska exemplaren som ingick i Formasprojektet är hittills endast tillgängliga i INSDC (se ovan). Streckkodning av vertebrater skall dock inte ses som något avslutat, då det bland annat finns behov av ytterligare streckkoder för populationer av fiskar, amfibier och reptiler som är dåligt kända genetiskt, och som är under taxonomisk utvärdering.

De ryggradslösa djuren är streckkodade i mycket varierande grad. Här finns den största artrikedomen, och också de största kunskapsluckorna (Fig. 1). En grupp där streckkodning har prioriterats tack vare stöd från bland annat Svenska artprojektet är gördelmaskar (Clitellata). Redan idag finns streckkoder från nästan två tredjedelar av de svenska arterna tillgängliga i BOLD. Material finns från nästan hela den svenska faunan, och detta material är under bearbetning.

Bland eftersatta grupper av marina evertrebrater kan nämnas mossdjur (Bryozoa), där exemplar från Norden nästan helt saknas bland de arter som streckkodats, och 67 % av de svenska arterna saknas helt i BOLDs databas (se de utförliga resultaten i Bilaga 1). Här pågår det dock ett projekt under 2017 som omfattar streckkodning av 130 svenska arter. Andra eftersatta grupper är rundmaskar (Nematoda), plattmaskar (Platyhelminthes) och hjuldjur (Rotifera), där en stor majoritet av diversiteten inte finns representerad i BOLD (se Tabell 1 och Bilaga 1).

Även bland grupper som nässeldjur (Cnidaria) är exemplar från Norden dåligt representerade i BOLD (7 av 189 svenska arter: 7,8 %), även om det är relativt god täckning internationellt (132 av 189: 69,8 %). Andra grupper där mönstret återkommer med relativt god täckning internationellt av Sveriges fauna på BOLD, men få exemplar från Norden representerade bland de streckkodade individerna, kan nämnas tagghudingar (Echinodermata) (66 av 73 arter eller 90,4 % tillgängliga internationellt men bara 2,7 % från nordiska exemplar), och blötdjur (Mollusca) (446/673: 66,3 % internationellt, 7,58 % nordiska).

Insekter är relativt väl täckta internationellt. Hela 64,64 % av den svenska insektsfaunan finns streckkodad i BOLD internationellt, och 21,24 % om man endast räknar nordiska exemplar. Av de artikaste insektsordningarna är skalbaggar väl täckta internationellt (83,9 %), och överträffas vad gäller nordiska streckkoder procentuellt endast av nattsländor och fjärilar (43,4 % skalbaggar, 81,76 % fjärilar, 85,33 % nattsländor).

Tvåvingar (flugor och myggor) utgör nästan en tredjedel av den kända svenska insektsfaunan. 65,1 % är streckkodade, men endast 9,9 % från nordiska exemplar. Som exempel på att de publika data i

Lecanoromycetes (lavsvampar)



Agaricales (skivlingar)



Asterales (korgblommiga växter m fl)



Acari (kvalster)



Nematoda (rundmaskar)



Chordata (ryggsträngsdjur)



Fig. 1. Andelen streckkodade arter i BOLD för några organismgrupper

Antalet svenska arter inom respektive grupp visas med en ljusblå stapel (alla i samma skala). Den röda delen är andelen av dessa arter som finns representerade med minst en streckkodssekvens från nordiska exemplar i BOLD.

Tabell 1. Sammanställning av läget för streckkodning av valda delar av Sveriges fauna. Urvalet innehåller taxa på skilda nivåer. Se Bilaga 1 för en fullständig redovisning av resultaten.

| | Artantal | Streckkod i BOLD | Nordisk streckkod |
|------------------------------------|----------|------------------|-------------------|
| Ryggradsdjur | 483 | 96% | 83% |
| Benfiskar | 122 | 93% | 84% |
| Broskfiskar | 9 | 100% | 89% |
| Groddjur | 13 | 92% | 31% |
| Reptiler | 6 | 100% | 83% |
| Fåglar | 251 | 97% | 95% |
| Däggdjur | 78 | 96% | 53% |
| Rovdjur | 18 | 100% | 61% |
| Fladdermöss | 17 | 100% | 47% |
| Gnagare | 18 | 83% | 72% |
| Ringmaskar (Annelida) | 800 | 66% | 14% |
| Gördelmaskar (Clitellata) | 294 | 65% | 33% |
| Havsborstmaskar | 506 | 66% | 4% |
| Nässeldjur | 189 | 74% | 4% |
| Tagghudingar | 73 | 90% | 3% |
| Blötdjur | 673 | 74% | 8% |
| Snäckor | 457 | 74% | 5% |
| Musslor | 175 | 76% | 14% |
| Rundmaskar | 1024 | 12% | 0% |
| Leddjur | 30024 | 62,85% | 19,13% |
| Spindlar | 1818 | 55,72% | 4,40% |
| Kvalster | 1036 | 30,69% | 1,74% |
| Kräftdjur | 1567 | 41,74% | 2,68% |
| Storkräftor (Malacostraca) | 642 | 66,20% | 3,74% |
| Hoppkräftartade (Maxillopoda) | 623 | 17,01% | 1,77% |
| Musselkräftor | 192 | 25,00% | 0,00% |
| Insekter | 26530 | 64,55% | 21,18% |
| Hoppstjärtar | 308 | 56,82% | 1,95% |
| Skalbaggar | 4414 | 83,91% | 43,43% |
| Vivlar (Curculionidae) | 496 | 81,45% | 39,72% |
| Mycelbaggar (Leiodidae) | 125 | 75,20% | 27,20% |
| Kortvingar (Staphylinidae) | 1238 | 78,76% | 40,63% |
| Långhorningar (Cerambycidae) | 110 | 90,91% | 61,82% |
| Bladbaggar (Chrysomelidae) | 282 | 89,36% | 53,55% |
| Tvåvingar | 7849 | 65,09% | 9,92% |
| Minerarflugor (Agromyzidae) | 308 | 53,57% | 0,32% |
| Blomsterflugor (Anthomyiidae) | 335 | 54,63% | 0,00% |
| Gallmyggor (Cecidomyiidae) | 657 | 7,46% | 0,91% |
| Styltflugor (Dolichopodidae) | 350 | 59,43% | 0,00% |
| Husflugor (Muscidae) | 366 | 74,32% | 3,55% |
| Fjädermyggor (Chironomidae) | 572 | 79,90% | 20,28% |
| Svampmyggor (Mycetophilidae) | 420 | 80,71% | 0,24% |
| Halvvingar | 1803 | 67,05% | 2,27% |
| Bladlöss | 372 | 54,84% | 8,06% |
| Dvärgstritar (Cicadellidae) | 323 | 65,02% | 0,31% |
| Ängsskinnbaggar (Miridae) | 237 | 95,78% | 1,69% |
| Steklar | 8420 | 44,98% | 5,24% |
| Bracksteklar (Braconidae) | 1114 | 25,31% | 3,23% |
| Hyllhornssteklar (Diapriidae) | 320 | 21,25% | 0,00% |
| Brokparasitsteklar (Ichneumonidae) | 3002 | 51,93% | 3,70% |
| Fjärilar | 2681 | 92,80% | 81,76% |
| Djurlöss | 267 | 12,36% | 0,37% |
| Tripsar | 143 | 41,96% | 0,00% |
| Nattsländor | 225 | 94,67% | 85,33% |

BOLD underskattar läget kan här nämnas att av fjädermyggor i Sverige och Norge finns minst ca 75% streckkoder när ännu ej publika data inkluderas. Steklar har ännu lägre täckning (44,99 % internationellt; 5,24 % Norden). Man ska dock komma ihåg att ett stort antal arter av tvåvingar och steklar fortfarande återstår att upptäcka i Sverige. En uppskattning byggd på resultat från Svenska Artprojektet och Svenska Malaisefällexprojektet tyder på att det kan finnas så mycket som 4000 oupptäckta arter av insekter i landet, många av dem antagligen okända för vetenskapen. De flesta av dessa är tvåvingar och steklar. Det innebär att gott och väl 20 % av arterna av tvåvingar och steklar saknas i vår analys, och de flesta av dessa saknas troligen i BOLD.

Växter och svampar

Bland växterna är det övergripande mönstret att det är relativt god täckning internationellt, men relativt få arter streckkodade från nordiska exemplar (Fig. 1, Bilaga 1). För kärlväxter finns det streckkodssekvens från 45 % av de i Sverige förekommande arterna, men bara 5 % från nordiska exemplar. Bland de grupper där det har gjorts streckkodning i Norden kan nämnas ljungväxter (14 av 28 arter: 50 %). Däremot för korgblommiga växter, där en stor del av mångfalden utgörs av svårbestämda apomiktiska småarter, är siffran betydligt lägre (9 % av svenska arter från internationellt material, 1 % från nordiskt material).

För bladmossor är andelen streckkodade arter 85 % internationellt, och 12 % från nordiska exemplar, men högre för levermossor: 98 % internationellt, 24 % från nordiska exemplar.

Tabell 2. Sammanställning av läget för streckkodning av valda delar av växt- och svampflora. Märk väl att det är ett urval, och innehåller taxa på skilda nivåer.

| | | | |
|--|------|------|-----|
| Enhjärtbladiga växter | 521 | 91% | 10% |
| Orkidéer | 43 | 88% | 9% |
| Halvgräs | 136 | 98% | 26% |
| Gräs | 159 | 87% | 3% |
| Tvåhjärtbladiga (Magnoliopsida) | 3697 | 37% | 4% |
| Korgblommiga växter (Asterales) | 1983 | 9% | 1% |
| Ljungartade växter (Ericales) | 55 | 93% | 33% |
| Plisterartade växter (Lamiales) | 175 | 91% | 8% |
| Smörblommeväxter (Ranunculales) | 400 | 16% | 2% |
| Rosväxter (Rosales) | 249 | 67% | 5% |
| Barrträd | 9 | 100% | 22% |
| Ormbunkar (Polypodiopsida) | 39 | 95% | 36% |
| Lumnerartade (Lycopodiopsida) | 11 | 64% | 9% |
| Bladmossor | 702 | 85% | 12% |
| Levermossor (Marchantiopsida) | 248 | 98% | 24% |
| Rödalger (Rhodophytina) | 193 | 72% | 8% |
| Grönalger (Chlorophyceae) | 451 | 12% | 0% |
| Konjugater (Conjugatophyceae) | 303 | 1% | 0% |
| Sporsäcksvampar | 5848 | 36% | 7% |
| Lecanoromycetes | 1796 | 59% | 15% |
| Leotiomycetes | 944 | 20% | 2% |
| Sordariomycetes | 910 | 33% | 3% |
| Dothideomycetes | 1040 | 16% | 1% |
| Basidiesvampar | 4100 | 66% | 19% |
| Skivlingar (Agaricales) | 2208 | 74% | 23% |
| Soppar (Boletales) | 117 | 73% | 18% |
| Tickor (Polyporales) | 325 | 66% | 19% |
| Kremlor och riskor (Russulales) | 335 | 83% | 30% |

De skilda grupperna av alger är svårare att sammanställa, då BOLD och Dyntaxa använder olika taxonomiska system, vilket kan ge missvisande siffror. Gruppen egentliga grönalger (Chlorophyceae) har 451 reproducerande arter i Sverige, men bara 53 arter har streckkodssekvens i BOLD, varav inga från nordiska exemplar. Alger i stort verkar vara en eftersatt grupp, och med tanke på potentialen att använda alger i miljöövervakning är det angeläget med riktade insatser för streckkodning av dessa.

Bland svamparna är läget relativt gott för både basidiesvampar och för de typiska lavbildande svamparna (Lecanoromycetes), med 50-85 % av svenska arter streckkodade, varav 15-30 % på nordiskt material. Klart sämre är läget för de flesta sporsäcksvampar utöver Lecanoromycetes, med 15-35 % streckkodade varav 1-3 % på svenskt material (Fig. 1).

Orsaker till låg grad av streckkodning

Här berör vi några av de viktigaste orsakerna till att referensbiblioteken för vissa organismgrupper fortfarande är dåligt utvecklade.

Gruppen är dåligt känd taxonomiskt

Exempel: tvåvingar, steklar, rundmaskar, sporsäcksvampar

För flera grupper är många av de kända arterna insamlade och identifierade vid mycket få tillfällen och faunans sammansättning är fortfarande i stora delar okänd. De samlingar som finns är ibland äldre, och behöver ofta kontrollbestämmas av en expert med uppdaterade kunskaper. Även i de fall det finns stora mängder insamlat material är själva sorteringen och artbestämningen flaskhalsen i att göra det tillgängligt för streckkodning.

Specialutrustning krävs för insamling och bestämning

Exempel: djurlöss, kvalster, många grupper av marina evertebrater (plattmaskar, mossdjur).

Artrika grupper som kräver aktiv insamling och inte kan bestämmas i fält är underrepresenterade i referensbiblioteken. Många insekter och kvalster som lever som ektoparasiter på ryggradsdjur kan bara samlas in från levande värdar, t.ex. fåglar. Små marina evertebrater måste ofta bestämmas i det korta tidsspannet mellan insamling och konservering, både de som lever som plankton och de som lever i sediment (meiofauna). Bland exempelvis plattmaskar är det vanligt att själva exemplaren inte går att konservera: de går endast att identifiera levande under mikroskopförstoring vilket försvårar streckkodning avsevärt. Exemplar av sådana grupper som konserveras i sprit är förlorade för morfologisk identifikation, eftersom sprit gör mjuka vävnader hårda och ogenomskinliga. Exemplar som konserveras som mikroskoppreparat förlorar i stället allt DNA under fixeringsprocessen. Exemplar kan identifieras, fotograferas och sedan konserveras med sprit för DNA-extraktion, men det är en arbetskrävande process som kräver expertkunskaper i varje led.

Befintliga samlingar är inte användbara av tekniska skäl

Exempel: kvalster, tripsar, marina evertebrater

Här ingår grupper av organismer som vid preparation behandlas så att DNA förstörs eller tvättas bort. Kvalster identifieras under ljusmikroskop och skall då vara klarnade och placerade mellan glasskivor. I den processen används kemikalier för att lösa upp allt utom själva kutikulan. Även om det finns samlingar som är välbestämda och dokumenterade, så är allt DNA borttvättat från preparaten. Samma typ av prepareringsmetoder används också för många små evertebrater. Tidigare gjordes klarningen med kaliumhydroxid, men numera är det vanligt att använda proteinaser istället och ta till vara de upplösta vävnaderna för DNA-extraktion. Marina evertebrater som samlas från bottenprov har åtminstone tidigare ibland konserverats med formalin redan i provstadiet. Den preparationen behåller strukturer väl, och gör det möjligt att sortera ut och bestämma även små organismer ur provet. Tyvärr förstörs DNA i processen. Prov som behandlas med sprit blir inte lika välbevarade strukturellt, men DNA bevaras väl.

Begränsningar hos BOLD

Det är en del av BOLDs vision att de streckodssekvenser som görs tillgängliga också skall åtföljas av detaljerade metadata inklusive information om och fotografier av beläggexemplar. Tyvärr är en stor del av de data som finns i BOLD inte åtkomlig för besökare. Det beror i många fall på att sekvenserna hör till projekt som fortfarande arbetar aktivt på sina dataset och ännu inte är beredda att göra sina metadata offentligt tillgängliga. I andra fall saknas metadata fortfarande. Tillgången till data i BOLD styrs också av olika nationers/aktörers avtal med BOLD. I dag har Sverige inget avtal med BOLD.

Begränsningarna kan exemplifieras med statistik över publikt tillgängliga sekvensdata från nordiska exemplar av leddjur. Detta gäller den totala nordiska faunan, data är inte begränsade till i Sverige förekommande arter. Totalt finns 90 410 sekvenser tillgängliga i BOLD; 29 019 (32 %) av dem är endast namngivna på högre taxonomisk nivå, medan 28 109 (31 %) är identifierade till art (representerande totalt 6615 unika arter). 3 262 sekvenser är hämtade direkt från GenBank ("Mined from GenBank") och saknar metadata. Det är väl värt att notera att en stor del av referenssekvenserna för leddjur i BOLD (två tredjedelar) fortfarande saknar en mer precis taxonomisk identifiering. I en del fall finns dessa data men har inte släppts än, men för många sekvenser saknas denna information ännu helt. Mycket av det taxonomiska arbetet med BOLD-sekvenserna återstår alltså fortfarande.

Ett svenskt streckkodningsinitiativ

Vi kan konstatera att det finns behov av att komplettera de befintliga referensbiblioteken för svenska organismer. Här diskuterar vi förutsättningarna för ett svenskt streckkodningsinitiativ med mål att fylla de existerande luckorna, och hur ett sådant initiativ skulle kunna struktureras så att det blir av största möjliga värde för miljöövervakningen både på kort och lång sikt.

Kunskapsläget

Som redan nämnts är den taxonomiska kunskapen helt kritisk i uppbyggnaden av referensbiblioteken. Beläggexemplaren måste i något stadium av processen associeras med ett taxonomiskt namn på ett pålitligt sätt, och för väldigt många organismgrupper, ofta de artrikaste, krävs specialister för artbestämning. Att associera DNA-streckkoder med namn är som redan nämnts nödvändigt för att associera dem med befintlig kunskap och för att associera resultat från olika markörer med varandra.

Tillgången till taxonomisk expertis är ofta en av de svåraste flaskhalsarna i streckkodningsprojekt. Det är generellt ont om taxonomer och det nuvarande läget beskrivs av många som en "taxonomins kris" (se t.ex. Wheeler m.fl., 2004). Situationen i Sverige är dock bättre än i många andra länder. Dels har vi en gedigen taxonomisk tradition från Linné och framåt som gör att utgångsläget är ovanligt gott, dels har vi sedan 2002 haft Svenska Artprojektet som har inneburit en omfattande satsning på att kartlägga vår biologiska mångfald, inte minst genom forskning på dåligt kända organismgrupper. För många av de grupper som har studerats inom Artprojektets olika forskningsprojekt har vi unika förutsättningar tack vare tillgången till färskt och säkert bestämt material för streckkodning. Men ett streckkodningsprojekt är beroende av fortsatta sådana satsningar och av nätverk av taxonomer som kan kontrollera identiteten hos exemplar i museisamlingar och exemplar som kommer in. En viktig del för långsiktigheten i övervakning av den biologiska mångfalden i landet är att trygga återväxten och finansiera utbildning av taxonomer.

Behovet av taxonomisk kompetens blir inte heller överspelat även om man skulle få ett heltäckande referensbibliotek av streckkoder av kända organismer i landet. Tvärtom kan ett omfattande referensbibliotek bli ett viktigt hjälpmedel för taxonomen i arbetet med att identifiera och beskriva nya taxa för landet och för vetenskapen, reda ut svåra artkomplex, och bygga upp kunskaper om ekologiska preferenser, livscykel, levnadssätt osv även för de dåligt kända organismgrupperna, istället för att lägga ner en stor del av sin tid på triviala identifikationer för att sälla ut de intressanta exemplaren. Den här kunskapsuppbyggnaden om arterna i vår flora och fauna kommer också att bli alltmer värdefull för tolkningen av framtida miljöövervakningsresultat med NGS-metoder.

Tillgängligt material

Zoologiska museisamlingar

I svenska museisamlingar finns ett stort material av bestämda ryggradslösa djur. Museala våtsamlingar är ofta konserverade med formalin och sedan överförda till sprit för bevarande. Formalin bevarar form och färg bättre än sprit, men förstör möjligheten att utvinna DNA. Tills nyligen (1980-90 tal) bevarades sällan information om hur material konserverats. Detta gör det svårt att säga om enskilda våtsamlingar är användbara för DNA-extraktion. Har materialet konserverats i sprit, och aldrig behandlats med formalin eller torkat ut (vid uttorkning sjunker alkoholhalten, vilket påskyndar

nedbrytningen av DNA) är det ofta möjligt att utvinna DNA även ur äldre exemplar även om det innebär ett merarbete jämfört med nyinsamlat material.

Djur som är mycket små bevaras för det mesta som mikroskoppreparat, och är behandlade för att klarna och färga vävnader. För de minsta djuren, exempelvis många grupper av marina och limniska evertebrater som endast är uppbyggda av mjuk vävnad, är det inte möjligt att bevara exemplar som använts för DNA-extraktion. Belägget blir då ett foto och ett DNA-extrakt. För dessa grupper kan det finnas material i forskningsinstitutionernas frysar, och detta är värt att inventeras. För små evertebrater med hårt skal, exempelvis kräftdjur och insekter går det att extrahera DNA och samtidigt bevara ett beläggexemplar för morfologiska studier (Thomsen m.fl., 2009).

För torrsamlingar, främst av insekter, finns möjligheter att använda äldre material. Även här är materialets användbarhet helt beroende av hur det har preparerats och behandlats över tid. Espeland m.fl (2010) visade att DNA-kvalitén försämrades märkbart i torrpreparerade insekter som behandlats med skadedjurmedlet diklorvos. Tyvärr användes medlet i insektssamlingar på Naturhistoriska riksmuseet i Stockholm under större delen av 1900-talet, men insektssamlingarna i Göteborg och Uppsala har varit förskonade från diklorvosbehandling.

Stora delar av de befintliga samlingarna av svenska insekter härstammar från privatsamlingar som donerats till museer. Detta gäller särskilt skalbaggar och fjärilar, där större donationer av relativt nyinsamlat material kontinuerligt kommer in. Numera tas många sådana donationer inte emot av museerna eftersom grupperna redan är välrepresenterade i samlingarna. Materialet kan trots det på grund av att det är nyinsamlat vara värdefullt för uppbyggnaden av referensbibliotek. Denna typ av material är nästan helt oprövat för streckkodning, åtminstone i Sverige, men bör utvärderas systematiskt genom provsekvenseringar. Även av mer kritiska grupper som flugor och steklar kommer det in enstaka mycket värdefulla större samlingar av nyinsamlat material.

Streckkodning av vertebratfaunan är nästan heltäckande, med sekvenser från väldokumenterat material (med belägg och insamlingsdata) för de vanligaste streckkodsmarkörerna. Det har emellertid under senare år förts fram åsikten att vissa endemiska populationer av amfibier och reptiler uppfyller kraven på fullgoda arter, och bör höjas i taxonomisk nivå. Detta gäller exempelvis gotlandssnoken (*Natrix natrix gotlandica*) och den norduppländska populationen av gölgroda (*Pelophylax lessonae*). Reptiler och amfibier samlades tidigare i formalin, och det äldre materialet är inte användbart för DNA-extraktion, vilket innebär att nytt material måste samlas in.

Nyare zoologiska insamlingsprojekt

Under 00- och 10-talen har flera större inventeringsprojekt genomförts i landet. De två största är, som nämnts ovan, Svenska artprojektets marina inventering, som samlade in evertebrater i bottenfaunan på 378 lokaler längs den svenska västkusten 2006-2009 och Svenska Malaise-fällexperimentet som samlade insekter med 73 fällor under 2003-2006.

De flesta av den marina inventeringens lokaler är återbesök av lokaler som inventerats av Leonard Jägerskiöld 1921-1939. Totalt togs 527 prover, och materialet identifierades till största delen redan i fält. Metoder för insamling och konservering anpassades med tanke på DNA, och därför användes sprit istället för formalin där det var möjligt. För organismer som inte är möjliga att bevara i identifierbart skick i sprit togs vävnadsprov innan de konserverades i formalin. Enligt slutrapporten 2014 fanns registrerat ca 16 000 artbestämda fynd från 1 015 unika arter. Ett stort antal av dessa arter har ännu inte streckkodats (Tabell 3). Materialet är till allra största delen deponerat vid GNM. Streckkodning av materialet pågår (Sanger-sekvensering av COI på BIO i Kanada), i samarbete mellan SweBOL och NorBOL. Det skulle dock vara en fördel om takten i arbetet kunde ökas genom att

Tabell 3. Artbestämt material från Artprojektets marina inventering (SAMARIN) som är klart för streckkodning (data för COI för ett urval av grupper i materialet).

| Taxon | Svenska marina arter | Tidigare streckkodade (nordiskt mat.) | Icke tidigare streckkodade arter i SAMARIN |
|----------------------|----------------------|---------------------------------------|--|
| Polychaeta | 506 | 19 | 149 |
| Malacostraca | 640 | 24 | 84 |
| Cnidaria | 183 | 6 | 46 |
| Echinodermata | 73 | 2 | 54 |
| Bivalvia | 175 | 17 | 64 |
| Nudibranchia | 84 | 4 | 10 |

avsätta större resurser från svensk sida för att ta fram och skicka iväg material, inte minst med tanke på att NorBOL inte har en långsiktigt garanterad finansiering för sitt streckkodningsarbete. DNA-extrakt och tillhörande metadata återbördas idag inte till Sverige (se vidare nedan).

Malaisefällexprojektet är det första projekt som närmar sig en systematisk inventering av hela Sveriges insektfauna, med särskilt fokus på de dåligt kända grupperna bland tvåvingar och steklar. De 73 fällorna var placerade från Skånes sydkust till fjället Njuolla i Abisko och tömdes året runt under fältdelen av projektet. Det totala materialet består av 1 892 insektprover, som tillsammans innehåller uppskattningsvis 80 miljoner insekter. Hittills har ca 80 % av materialet sorterats till taxonomiska fraktioner (vanligen familjer men i flera fall lägre enheter än så) som är lämpade för vidare bearbetning av relevanta specialister. Mer än 600 000 exemplar har skickats ut för bestämning och 200 000 individer har bestämts så här långt. Det bestämda materialet omfattar ungefär 4 000 arter varav uppemot 1 000 är nya för Sverige och ungefär 500 av dem nya för vetenskapen. De nya arterna har främst hittats bland tvåvingar (både myggor ochflugor) och parasitsteklar. Flertalet av de 4 000 bestämda arterna har ännu inte streckkodats för COI (eller någon annan markör; Tabell 4).

Svenska Malaisefällexprojektet avsätter idag inga resurser för streckkodning av materialet. Det är dock ett rikt material som till stor del samlats och bevarats på ett sätt som gör det möjligt att använda det för streckkodning i större skala. På initiativ av enstaka forskare skickas små delar av materialet till GBOL, och andra mindre delar har streckkodats av NorBOL på norskt initiativ. I bägge fallen handlar det om Sanger-sekvensering av COI på BIO i Kanada, och det är bara bråkdelar av de tillgängliga arterna som har bearbetats så här långt. Metadata, DNA-extrakt och beläggsexemplar från streckkodningen via GBOL skickas till NRM för accessionsföring. Ett samlat svenskt streckkodningsinitiativ som systematiskt arbetade igenom materialet skulle kunna höja takten

Tabell 4. Artbestämt material från Malaisefällexprojektet (SMTP) som är klart för streckkodning.

| Taxon | Svenska arter | Tidigare streckkodade (nordiskt mat.) | Icke tidigare streckkodade arter i SMTP |
|--------------------------|---------------|---------------------------------------|---|
| Coleoptera | 4414 | 1917 | 109 |
| Diptera | 7849 | 779 | 1867 |
| Hymenoptera | 8420 | 441 | 988 |
| Lepidoptera | 2681 | 2192 | 21 |
| Övriga Arthropoda | 6660 | 415 | 267 |
| Totalt | 30024 | 5744 | 3252 |

dramatiskt och streckkoda tusentals arter som ännu inte finns representerade i de befintliga referensbiblioteken (Tabell 4).

Ett annat ambitiöst insamlingsprojekt inom ramen för Svenska Artprojektet, men av mer tillfällig karaktär, var en meiofaunaworkshop på Tjärnö 2007 där en grupp internationella experter studerade valda grupper av små marina djur, och fann åtminstone 430 arter varav 157 nya för Sveriges fauna och 27 sannolikt nya för vetenskapen (Willems m.fl., 2009). Med avseende på de särskilda svårigheter som redan nämnts med den här typen av material är det bara mindre delar av fångsten som finns tillgänglig för sekvensering i NRMs samlingar, och som kan läggas till ett nationellt DNA-arkiv.

Flera andra större insamlingsprojekt har ägt rum under de senaste åren, ofta inom ramen för enskilda forskningsprojekt i Svenska Artprojektet. Dels har dessa gett upphov till välbestämda referenssamlingar inom dåligt kända organismgrupper, som i några fall redan DNA-sekvenserats och i flera andra är direkt tillgängliga för streckkodning. Idag finns ingen komplett förteckning av detta material då det inte har varit en prioriterad arbetsuppgift inom Artdatabanken.

Det finns också en relativt stor mängd bulkmaterial (Malaisefällematerial m.m.) från museernas kontinuerliga insamlingsverksamhet och från projekt som finansierats av Artprojektet, och som är tillräckligt välbevarat för att kunna användas för streckkodning. För att materialet ska kunna användas för referensbiblioteken behöver resurser dock frigöras för att sortera upp materialet och identifiera det innan sekvensering av de intressanta arterna. NGS-metoder kan eventuellt användas för att accelerera processen, men någon form av uppsortering och taxonomisk bestämning av belägsexemplaren måste göras om materialet ska användas för att bygga referensbibliotek.

Botaniskt material

Kärlväxtmaterial i herbarier har visat sig vara användbart för DNA-extraktion. Detta förutsatt att materialet har torkats snabbt, och aldrig varit behandlat med sprit. Nyare kollektioner, från 1990-talet och framåt, är ofta samlade med tanke på DNA, och en del av det insamlade materialet har då frångilts och torkats med kiselsyrage (silicagel).

Herbariematerial av svampar (inklusive lavar), är även det användbart för DNA-extraktion. Mycket material är relativt gammalt (>50 år) och oprövat för molekylärgenetiska studier. Liksom för torra insektssamlingar är det önskvärt att herbariematerial av svamp och lavar undersöks mer systematiskt med avseende på hur de lämpar sig för DNA-extraktion och streckkodning.

Material från amatörbiologer

Det finska streckkodningsprojektet FinBOL har med mycket små resurser lyckats sekvensera en imponerande stor del av den finska floran och faunan. Det har återigen huvudsakligen handlat om Sanger-sekvensering av COI i samarbete med BIO; något nationellt DNA-arkiv har dessvärre inte byggts upp. FinBOLs framgångar bygger till mycket stor del på samarbete med amatörbiologer. Också i Sverige bör det finnas en betydande potential att få hjälp av våra många och kunniga amatörbiologer i att bygga ett nationellt DNA-arkiv.

Faktum är att SweBOL och CGI redan vid flera tillfällen kontaktats av amatörbiologer som är villiga att bidra med identifierat material, t.ex. från de korta och massmedialt uppmärksammade inventeringar som kallas Bioblitzar. I andra fall gäller det artavgränsningsproblem och nyheter för den svenska faunan och floran som uppmärksammas inom den nyfikenhetsdrivna utforskning av mångfalden som amatörer kontinuerligt bedriver. För många organismgrupper, särskilt bland de terrestra leddjuren men även inom andra grupper, är landets ledande expert i nuläget en amatörbiolog. De biologiska

föreningarna samlar både stor kompetens och ett storartat engagemang, som torde kunna bli avgörande för ett svenskt streckkodningsprojekt.

Det finns numera också ett enormt engagemang från allmänheten i Artportalen och sociala media som fokuserar på biologiska observationer, och det intresset når långt större grupper än bara de organiserade amatörbiologerna. Genom upprop och efterlysningar i föreningarna och i sociala media kommer utan tvivel ett svenskt streckkodningsprojekt kunna profitera på både viktig kompetens och specifikt material av särskilt utpekade måltaxa för DNA-referensbiblioteket, kanske som en särskild folkforskningsdel (av det engelska begreppet "citizen science") av streckkodningsprojektet.

Framtagning av referenssekvenser

Som framgått ovan har vi i Sverige idag unika förutsättningar för storskalig streckkodning i och med tillgången till ett stort och välbestämt material lämpat för streckkodning. Arbetet sker dock idag i långsam takt och utan nationell samordning. Prioriteringarna styrs till stor del av vilka forskningsprojekt som är aktiva inom området och inte av miljöövervakningens behov eller av de luckor som finns i referensbiblioteken. Under tiden försämras kvaliteten i det insamlade materialet. Jämfört med idag skulle ett samlat initiativ kunna innebära en avsevärd ökning i streckkodningstakten och en bättre styrning av arbetet. Samtidigt skulle vi kunna försäkra oss om att vi verkligen tar till vara den unika möjlighet vi har idag att streckkoda svensk flora och fauna till en låg kostnad och samtidigt lägga grunden till framtida genetiskt baserad miljöövervakning.

Ett samlat initiativ borde initialt undersöka möjligheterna att utöka samarbetet med GBOL och NorBOL för att få en större del av det svenska materialet streckkodat genom dessa projekt så länge de varar. Både GBOL och NorBOL streckkodar idag svenskt material utan kostnad, men villkoren för detta behöver klargöras i mer detalj om samarbetet ska utökas väsentligt. Även om kostnaden för streckkodningen täcks av andra behöver svenska resurser ändå sättas av för att identifiera vilket material som är prioriterat från svensk sida och för att täcka kostnaderna för att samla ihop materialet och skicka det till Kanada tillsammans med relevanta metadata i rätt format. Samtidigt behöver man försäkra sig om att beläggsexemplar accessionsförs vid något av de stora naturhistoriska museerna i Sverige och länkas till referenssekvenserna, och att DNA-extrakten repatrieras (se nedan). I Finland har man med stor framgång organiserat den här typen av arbete genom en nationell "streckkodningsgeneral" som har fokuserat helt på det logistiska arbetet förknippat med streckkodningen, och det är sannolikt att en liknande modell skulle kunna fungera väl i Sverige.

Både NorBOL och GBOL är tidsbegränsade satsningar och vi kan inte räkna med att få allt svenskt material streckkodat genom dem. På längre sikt behöver vi därför utvärdera alternativ. Den mesta streckkodning som sker idag fokuserar på Sanger-sekvensering av COI och andra standardmarkörer, och själva sekvenseringen och databearbetningen sker för det mesta hos BIO i Kanada. Sanger-sekvensering är relativt dyrt och BIO har genom stordrift kunnat erbjuda ett konkurrenskraftigt pris för sekvensering och databearbetning. Så länge Sanger-sekvensering används för streckkodningen så är det mycket som talar för att ett svenskt streckkodningsprojekt bör fortsätta att skicka material till BIO.

Teknikutvecklingen inom området är emellertid snabb och Sanger-sekvensering håller på att bli en föråldrad teknik. Vi räknar med att NGS-metoder inom några år kommer att erbjuda både billigare och mer tillförlitliga referenssekvenser än Sanger-sekvensering, särskilt om det finns intresse för att bredda antalet markörer som används i analyserna. Sverige kan idag erbjuda mycket konkurrenskraftiga priser för NGS-sekvensering tack vare de nationella sekvenseringsplattformarna som drivs av SciLifeLab i Stockholm och Uppsala. BIO i Kanada utforskar också de nya möjligheterna med NGS-plattformar, och det är idag oklart om de kommer att kunna erbjuda svenska användare

konkurrenskraftiga villkor vid streckkodning med NGS-metoder. Ett nationellt svenskt streckkodningsinitiativ behöver kontinuerligt utvärdera de olika alternativen.

Det fysiska arkivet

Som nämnts ovan är det inte bara viktigt att Sverige kommer igång med systematisk uppbyggnad av relevanta referensbibliotek för genetiskt baserad miljöövervakning. På längre sikt är det också viktigt att bygga upp ett nationellt DNA-arkiv över vår flora och fauna för framtida bruk, innehållande vävnadsprover och DNA-extrakt kopplade till tillförlitligt bestämda referensexemplar. Lämpligen konstrueras DNA-arkivet så att det lätt i framtiden kan användas för att snabbt och kostnadseffektivt ta fram nya referensbibliotek vid behov med automatiserad labb-teknik (robotar) och NGS-metoder.

Det traditionella sättet att bevara DNA-extrakt är djupfrysning. Erfarenheter har visat att frysning i ultrafrysar (-80°) ger endast marginellt bättre resultat än standardfrysar (-20°), och därmed är det svårt att motivera den betydligt högre kostnaden för ultrafrysar (se t.ex. Palkopoulou m.fl., 2015). Ytterligare en metod som används är frystorkning av DNA-extrakten. Ett alternativ som diskuterats mycket på senare år är att torka DNA-extrakten och förvara dem i rumstemperatur. Enligt jämförande studier (t.ex. Ivanova & Kuzmina, 2013) ger det lika bra bevarande som frysning och detta på ett sätt som är billigare, säkrare och mindre utrymmeskrävande. Man pipetterar ut DNA-extraktet på ett filterpapper och efter torkning förvarar man sedan filterpapperet torrt och mörkt. Frysning i buffertlösning är dock det som anses som säkraste bevaringsmetod och används i flera arkiv med höga krav för bevarande av humant DNA.

Det frö till nationellt DNA-arkiv som finns idag vid NRM omfattar huvudsakligen frysta DNA-extrakt av ryggradsdjur och några andra grupper, bland annat fjädermyggor som streckkodats i ett pilotprojekt fokuserat på bottenprover från Östersjön och finansierat av HaV. I nuläget förvaras DNA-extrakten i Miljöprovbanken på NRM, där man har erfarenhet av långvarig förvaring av biologiskt material sedan 60-talet. Infrastrukturen uppgraderas för närvarande, bland annat genom introduktionen av ett nytt lagringssystem anpassat för hantering av proverna i större skala med labb-robotar. Den fysiska lagringen av DNA-arkivet kommer dock successivt att behöva anpassas för de metoder som blir tillgängliga för rationell hantering av NGS-sekvensering av stora mängder prover.

DNA-extrakt som är framtagna vid streckkodning av svenskt material på BIO i Kanada finns frystorkade där och tillgången till sekvensinformationen och DNA-extrakten kontrolleras genom överenskommelser mellan respektive projektägare och BIO. Det är inte självklart att sekvenserna blir publikt tillgängliga omedelbart och att andra forskare kan få tillgång till DNA-extrakten. För svenska forskare som är intresserade av att gå djupare i de genetiska analyserna är det bekvämt om materialet finns på närmare håll, om inte BIO kontrakteras för vidare datagenerering. Även om kostnaderna för att återbördas DNA-extrakten täcks av en aktör som NRM är det inte självklart att man kan förmå projektägarna att se till att materialet blir återbördat utan särskilda incitament eller andra styrmedel. Det finns inte heller någon samsyn idag bland inblandade aktörer när det gäller avvägningen mellan fördelar och nackdelar med att ha kvar materialet i Kanada. En bra kompromissmodell som använts i GBOL-samarbeten är att dela DNA-extrakten så att de finns tillgängliga både i Sverige och Kanada efter streckkodningen. Det vore värdefullt om SweBOL kunde diskutera dessa frågor och enas om en rekommenderad modell som säkrar svenska forskares framtida tillgång till DNA-extrakt och vävnadsprover från de beläggexemplar som streckkodas.

Förutom DNA-extrakten och vävnadsproverna behöver det fysiska arkivet omfatta också de beläggexemplar från vilka DNA-extrakten kommer. Dessa kan med fördel spridas ut på Sveriges naturhistoriska museer baserat på var den bästa taxonomiska kompetensen och de mest relevanta samlingarna för den aktuella organismgruppen finns.

Det digitala arkivet

De referensbibliotek som tas fram över svensk flora och fauna måste åtföljas av relevanta informationssystem för hanteringen och tillgängliggörandet av såväl referenssekvenserna som metadata om dessa och om DNA-extrakten och beläggexemplaren. Sedan 2011 pågår ett arbete med att utveckla och implementera ett nationellt samlingshanteringssystem i Sverige, det så kallade DINA-projektet (DINA står för Digitalt Informationssystem för NATurhistoriska samlingar). Samarbetet omfattar i princip alla stora naturhistoriska samlingsinstitutioner i Sverige. Det nuvarande DINA-systemet hanterar redan ett stort antal svenska samlingar (se samlingsportalen <http://naturarv.se>), däribland beläggexemplaren i NRMs DNA-arkiv. Inom kort tillkommer det artbestämda materialet från SMTP och SAMARIN, de största svenska inventeringsmaterialen som är aktuella för DNA-streckkodning.

En ny version av DINA-systemet utvecklas sedan en tid av NRM i samarbete med bland annat Agriculture och Agri-Food Canada, som bidrar med en sofistikerad modul för hantering av sekvensdata, SeqDB. SeqDB har omfattande stöd för hantering av både Sanger-sekvensering och NGS-sekvensering, och erbjuder många funktioner både för streckkodning och NGS-baserade analyser av miljöprov. SeqDB tillåter bekväm uppladdning av referenssekvenser till relevanta databaser som INSDC och BOLD. Kanadensiska regeringen har nyligen lanserat ett sexårigt initiativ med en budget på totalt 30 miljoner CAD (200 miljoner kr) för att digitalisera alla samlingar i landet och stärka infrastrukturen. Där ingår bland annat en ytterligare förstärkning av SeqDB, som därmed står väl rustat för framtiden. I och med integreringen av SeqDB i DINA får Sverige en avancerad infrastruktur för den informationshantering som krävs av ett nationellt DNA-arkiv, oavsett om sekvenseringen görs i Sverige eller utomlands.

BIO i Kanada har överlägsen erfarenhet när det gäller att hantera informationsflödet kring DNA-streckkodning, och denna erfarenhet ligger till grund för BOLD-infrastrukturen som också är välfinansierad för lång tid framåt. När det gäller det svenska material som streckkodas av BIO och hamnar i BOLD gäller utmaningarna framför allt att se till att sekvenserna i BOLD länkas på korrekt sätt till de poster i DINA som berör de motsvarande beläggexemplaren, vävnadsproverna och DNA-extrakten som finns i svenska samlingar. Kopplingen mellan informationen i de olika systemen bygger idag på manuell hantering. Det finns viss risk för att fel uppstår i kopplingen, och hanteringen är inte så effektiv som den skulle kunna vara. I samband med ett tidigare streckkodningsprojekt finansierat av HaV diskuterades möjligheten att utveckla svenska inmatningslösningar som automatiserar delar av informationshanteringen, men i slutändan prioriterades inte detta. Denna fråga behöver undersökas på nytt i ljuset av de möjligheter som erbjuds av SeqDB.

Trots höga ambitioner finns det fortfarande många problem med data i BOLD, inte bara för att många av de sekvenserade exemplaren ännu inte har knutits till en säker bestämning, utan också för att en icke oansenlig del av bestämningarna är felaktiga. Ett svenskt streckkodningsinitiativ borde kunna se till att de svenska referensbiblioteken kan erbjuda ännu högre datakvalitet än BOLD och liknande system. Samtidigt ska referensbiblioteken på ett bättre sätt kunna koppla till de svenska samlingsdatabaser som innehåller information om beläggexemplaren och relevant information om rödlistestatus och annat av intresse gällande den svenska artstocken. Inte minst kan svenska referensbibliotek organiseras efter samma taxonomi som används för andra svenska system inom området, och på det viset skulle man kunna undvika problem kopplade till skillnader i den taxonomi som används i BOLD och den som används i Sverige.

En portal till de svenska referensbiblioteken skulle också kunna bli en naturlig inkörsport för svenska användare till de tjänster och den information som finns i de internationella referensbiblioteken. Ett

embryo till en portal finns i Svenska DNA-nyckeln (<http://dna-nyckeln.se>), som var ett resultat av det Formas-finansierade projektet för streckkodning av svenska ryggradsdjur. Långsiktigt bör utvecklingen av den här typen av system ske i nära samverkan med användare inom miljöövervakningen. Ett första steg i den riktningen tas inom ramen för ett Formas-finansierat utbildningsprojekt ("DNA för miljöövervakare") som pågår för närvarande och där svenska miljöövervakare får ökad kunskap om de nya möjligheter som öppnar sig för dem tack vara de nya genetiska teknikerna.

Slutord

Den genetiska miljöövervakningen utvecklas idag snabbt både i världen och i Sverige. För att knyta ihop resultaten från de nya metoderna med vår tidigare kunskap som bygger på taxonomiska namn är det viktigt att utveckla referensbibliotek för de användbara markörerna. Biblioteken kopplar en referenssekvens till ett välbestämt och noggrant dokumenterat beläggsexemplar i en publikt tillgänglig samling.

Till skillnad från många andra länder har vi fortfarande i Sverige inte fått igång ett systematiskt arbete med att bygga referensbibliotek och DNA-arkiv. Sverige har emellertid synnerligen goda förutsättningar för att sätta igång ett sådant arbete. Vi har gott om relevant material och expertis, och all den nödvändiga infrastrukturen. Det som saknas är i första hand personalresurser för att organisera och leda arbetet med att ta fram referensbibliotek för de mest använda markörerna och samtidigt bygga ett nationellt DNA-arkiv för framtida bruk. Det finska streckkodningsprojektet FinBOL har visat att det med små personalresurser går att åstadkomma imponerande resultat, och förutsättningarna att få till stånd en liknande kostnadseffektiv insats i Sverige som kan föra den genbaserade miljöövervakningen framåt borde vara synnerligen goda.

Tack

Vi vill till att börja med tacka Malin Strand som deltagit i processen och kommit med mycket värdefulla pusselbitar till rapporten. Även Niclas Gyllenstrand har kommit med konkreta bidrag, och Rasa Bukontaite gav elegant hjälp med figuren. Per Ericson har stött oss, liksom en lång rad forskare och museiarbetare som tog sig tid att diskutera streckkodning med oss eller tillhandahålla uppgifter: Johannes Bergsten, Christer Erséus, Oleksandr Holovachov, Ulf Jondelius, Charlotte Jonsson, Sven Kullander, Hans Mejlon, Michael Norén, Ellen Sandström och Mats Wedin. Tack till Rikard Sundin och Johan Liljeblad på ArtDatabanken för tillgång till data, assistans med Dyntaxa och goda råd. Markus Englund och Johan Nylander hjälpte till med programmering och datasammanställning. Naturvårdsverket finansierade rapporten, och vi vill särskilt tacka Ola Inghe för tålamod och stöd. Malin Strand och Ola Inghe lämnade många värdefulla kommentarer på det första utkastet till rapporten.

Ordlista

Amplikon – PCR-produkt. Används ofta om en blandad PCR-produkt från ett miljöprov. Där amplifieras en genetisk markör för alla arter som förekommer i provet.

Barcoding – se Streckkodning.

Beläggsexemplar – Vouchers, enskilda exemplar i museisamlingar som någon form av data hämtats ifrån, och som är tillgängliga för kontroll i efterhand. Är det t.ex. en DNA-sekvens från ett vävnadsprov från en fisk, så utgörs vouchern av resten av fisken i samlingen. I de fall då det fysiska exemplaret förstörs av t.ex. DNA-extraktion, så kan ett fotografi av det utgöra ett belägg.

Binomen – Ett vetenskapligt namn bestående av två led: släkte och artepitet. Utvecklat och populariserat av Carl von Linné i verket *Systema Naturae*. I släktet kan det ingå flera arter, men inom släktet måste vare artepitet vara unikt. Ekoxens binominala namn är *Lucanus cervus*. *Lucanus* är släktet, *cervus* artepitetet.

Bottenskrapa – Verktyg för att ta bottenprov i vatten. En tung metallram med ett nät som sänks ned och släpas längs botten efter ett fartyg.

DNA – Deoxiribonukleinsyra, självreplikerande makromolekyl som utgör en bas för livsprocesserna. Avläsning av sekvensen av nukleotidbaser längs med DNA utgör idag grunden för väldigt många forskningsområden inom och tillämpningar av biologin.

eDNA – Environmental DNA. Används om DNA från miljöprover, där det är vatten eller jord som utgör provet och inte vävnad från en organism. eDNA kan utgöras enbart av spår från djur, exempelvis vid detektion av fisk från vattenprover, där det räcker med de få molekyler DNA som förekommer fritt i vattnet.

Fauna – Den biologiska diversiteten av djur, t.ex. inom en grupp, eller på en plats, eller inom ett habitat eller ett land, eller globalt.

Flora – Den biologiska diversiteten av växter, t.ex. inom en grupp, eller på en plats, eller inom ett habitat eller ett land, eller globalt. I den här rapporten räknas även den biologiska mångfalden av svampar in i floran, fastän det faktum att svampar och växter inte är särskilt nära släkt gör att många hellre talar om svampdiversitet som funga.

Gen – Enhet för nedärvning av egenskaper. För det mesta använt om en region av DNA som innehåller koden för att tillverka ett specifikt protein.

Intron – Insprängda stycken av icke-kodande DNA i proteinkodande gener. Medför längdskillnader i gener för samma protein från olika arter.

Kloroplast – En organell i växtceller där fotosyntesen sker. Innehåller eget DNA. Streckkodningsmarkörer för växter, *rbcl*, *matK* och *trnL* är alla gener från kloroplastens genom.

Konservering – Den preparering som krävs för att ett djur eller en växt skall gå att bevara i en samling. Konservering kan skada DNA genom att förändra vävnaderna kemiskt för att de inte skall brytas ned.

Kryptiska arter – Arter som är mycket svåra att skilja från sina närmaste släktingar utan en genetisk analys.

Malaisefälla – Insektsfälla uppfunnen på 1930-talet av René Malaise, verksam vid Naturhistoriska riksmuseet i Stockholm. Påminner om ett tält med öppna sidor. Har revolutionerat insektssamlande.

Markör – En gen, del av gen eller definierat stycke DNA som kan sekvenseras och jämföras mellan arter.

Meiofauna – En sammanfattande term för diversiteten av små evertebrater som lever på botten och inuti sedimenten i marina och limniska habitat. Skiljs alltså genom storleken från både makrofauna och mikrofauna, och genom habitat/levnadssätt från plankton. Många taxonomiska grupper finns representerade, och studiet av denna fauna erbjuder särskilda praktiska svårigheter.

Metabarcoding – Processen att ta fram streckkodssekvenser från eDNA eller andra blandade prover. Används för att ta fram artsammansättningen i ett miljö- eller fällprov.

Metadata – Data som hör till. För en DNA-sekvens från INSDC ("GenBank") kan metadata vara exempelvis uppgifter om vilken markör det är, vilken organism den kommer ifrån och när och var organismen samlades eller provtogs.

Mitokondrie – Organell i djur, växt och svampceller. Har liksom kloroplasten ett bakteriursprung och innehåller sitt eget DNA. I mitokondrien sker cellandningen. De vanligaste markörerna för streckkodning av djur, COI och 16S, är båda mitokondriella markörer. Endast COI erkänns officiellt som streckkodsmarkör av CBOL.

NGS – Next Generation Sequencing. Sammanfattningsnamn på de metoder för DNA-sekvensering då många DNA-fragment sekvenseras parallellt. Fördelarna är stora mängder sekvensdata på kort tid, men nackdelen jämfört med standardsekvensering är att endast kortare fragment kan sekvenseras. Just nu ligger gränsen på ~400 baspar, och standardmarkören för streckkodning av djur (COI) är 658 baspar lång.

OTU – Operational Taxonomic Unit. En grupp organismer som kan grupperas ihop på genetisk likhet, men är distinkt skild ifrån närstående grupper. Används i genetiska analyser för kluster av sekvenser innan de kopplas ihop med ett formellt taxonomiskt namn.

PCR – Polymerase Chain Reaction. En metod där en genetisk markör kan kopieras upp och koncentreras. PCR-reaktionen innebär att man använder ett par korta DNA-sekvenser, eller "primers" från engelskan, för att hitta och amplifiera en specifik DNA-sekvens i provet. För att fånga upp så många arter som möjligt försöker man hitta primerpar som matchar DNA-sekvenser som är nästan identiska hos alla arter man är intresserad av. Den DNA-snutt som man amplifierar upp mellan sina primers kallas amplikon.

Primer – Korta bitar enkelsträngat DNA som behövs för PCR. Framställs syntetiskt. Primrar används i par, som matchar början och slutet på markören som skall amplifieras. Generella ('universella') primerpar passar samma markör hos en stor bredd av organismer. Specifika primerpar är optimerade för en snävare organismgrupp.

Referensbibliotek – en databas med sekvenser från en standardmarkör för streckkodning som används för identifiering genom att jämföra sekvenser från obestämda prov med biblioteket. Biblioteket byggs upp av sekvenser från välbestämda exemplar av organismer, vilka sparas som

beläggsexemplar vid en naturhistorisk samling eller liknande. Höga krav ställs på metadata kopplade till sekvenser i ett referensbibliotek.

Ribosomalt DNA eller rDNA – DNA som utgör koden för ribosomer, organeller i cellen som bygger proteiner. Till skillnad från de flesta strukturer i cellen utgörs ribosomer inte huvudsakligen av protein och fetter utan av nukleinsyran RNA. Ribosomalt DNA finns både i kärn- och mitokondriegenomet. De kodande delarna är 18S, 5.8S och 28S från kärnan, samt 12S och 16S från mitokondrien. 16S är den näst mest använda streckkodningsmarkören för djur. Till ribosomalt DNA räknas också de icke-kodande regionerna ITS1 och ITS2 som är standardmarkörer för streckkodning av svamp.

Sanger-metoden eller Sanger-sekvensering – Har blivit en samlingsbeteckning för traditionella sekvenseringsmetoder i kontrast till NGS, som då utmärks av att man läser av en markör från en organism i taget. Egentligen syftar termen på det arbetsflöde där en central del är att man bryter av DNA-kedjorna med vissa kemikalier för att kunna läsa dem.

Shotgun-sekvensering – Att ta fram mycket långa DNA-sekvenser genom en stor mängd avläsningar av korta fragment, som sedan pusslas ihop i datorn.

Streckkodning – Processen att ta fram en kort sekvens DNA av en markör som är befintlig hos en stor bredd av organismer och är variabel nog att vara unik för varje art, men generell nog att kunna amplifieras med universella primerpar.

Taxon (plural taxa) – En enhet i den biologiska taxonomin; i praktiken en art eller en grupp med arter. Enligt det linneanska systemet klassificeras organismer i en hierarki av taxa, där grundnivåerna är rike, fylum, klass, ordning, familj, släkte och art. Varje grupp som har ett namn i klassificeringen utgör ett taxon.

Torrssamling – De museisamlingar som förvaras torkade. Hit hör till exempel herbarier, insekter på nålar, skinnlagda däggdjur och ben.

Voucher – se Beläggsexemplar

Våtsamling – De museisamlingar som förvaras i vätska. Hit hör preparat i sprit och formalin, men också miljöprover och fällprover som samlats i sprit.

Referenser

- Agnarsson, I., Kuntner, M., Paterson, A., 2007. Taxonomy in a Changing World: Seeking Solutions for a Science in Crisis. *Syst Biol* 56, 531–539. doi:10.1080/10635150701424546
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215, 403–410. doi:10.1016/S0022-2836(05)80360-2
- Andersen, K., Bird, K.L., Rasmussen, M., Haile, J., Breuning-Madsen, H., Kjær, K.H., Orlando, L., Gilbert, M.T.P., Willerslev, E., 2012. Meta-barcoding of “dirt” DNA from soil reflects vertebrate biodiversity. *Molecular Ecology* 21, 1966–1979. doi:10.1111/j.1365-294X.2011.05261.x
- Bergsten, J., Bilton, D.T., Fujisawa, T., Elliott, M., Monaghan, M., Balke, M., Hendrich, L., Geijer, J., Herrmann, J., Foster, G., Ribera, I., Nilsson, A.N., Barraclough, T.G., Vogler, A.P., 2012. The Effect of Geographical Scale of Sampling on DNA Barcoding. *Systematic Biology* 61:851-869. doi:10.1093/sysbio/sys037
- Brandon-Mong, G.-J., Gan, H.-M., Sing, K.-W., Lee, P.-S., Lim, P.-E., Wilson, J.-J., 2015. DNA metabarcoding of insects and allies: an evaluation of primers and pipelines. *Bulletin of Entomological Research* 105, 717–727. doi:10.1017/S0007485315000681
- Dentinger, B.T.M., Didukh, M.Y., Moncalvo, J.-M., 2011. Comparing COI and ITS as DNA Barcode Markers for Mushrooms and Allies (Agaricomycotina). *PLOS ONE* 6, e25081. doi:10.1371/journal.pone.0025081
- Dobzhansky, T., Sturtevant, A.H., 1938. Inversions in the Chromosomes of *Drosophila pseudoobscura*. *Genetics* 23, 28–64.
- Espeland, M., Irestedt, M., Johanson, K.A., Åkerlund, M., Bergh, J.-E., Källersjö, M., 2010. Dichlorvos exposure impedes extraction and amplification of DNA from insects in museum collections. *Frontiers in Zoology* 7, 2. doi:10.1186/1742-9994-7-2
- Ficetola, G.F., Miaud, C., Pompanon, F., Taberlet, P., 2008. Species detection using environmental DNA from water samples. *Biology Letters* 4, 423–425. doi:10.1098/rsbl.2008.0118
- Goldberg, C.S., Strickler, K.M., Pilliod, D.S., 2015. Moving environmental DNA methods from concept to practice for monitoring aquatic macroorganisms. *Biological Conservation, Special Issue: Environmental DNA: A powerful new tool for biological conservation* 183, 1–3. doi:10.1016/j.biocon.2014.11.040
- Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J., Goodman, R.M., 1998. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & Biology* 5, R245–R249. doi:10.1016/S1074-5521(98)90108-9
- Hebert, P.D., Cywinska, A., Ball, S.L., others, 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B: Biological Sciences* 270, 313–321.
- Hollingsworth, P.M., Forrest, L.L., Spouge, J.L., Hajibabaei, M., Ratnasingham, S., van der Bank, M., Chase, M.W., Cowan, R.S., Erickson, D.L., Fazekas, A.J., Graham, S.W., James, K.E., Kim, K.-J., Kress, W.J., Schneider, H., van AlphenStahl, J., Barrett, S.C.H., van den Berg, C., Bogarin, D., Burgess, K.S., Cameron, K.M., Carine, M., Chacón, J., Clark, A., Clarkson, J.J., Conrad, F., Devey, D.S., Ford, C.S., Hedderson, T.A.J., Hollingsworth, M.L., Husband, B.C., Kelly, L.J., Kesanakurti, P.R., Kim, J.S., Kim, Y.-D., Lahaye, R., Lee, H.-L., Long, D.G., Madriñán, S., Maurin, O., Meusnier, I., Newmaster, S.G., Park, C.-W., Percy, D.M., Petersen, G., Richardson, J.E., Salazar, G.A., Savolainen, V., Seberg, O., Wilkinson, M.J., Yi, D.-K., Little, D.P., 2009. A DNA barcode for land plants. *Proc Natl Acad Sci U S A* 106, 12794–12797. doi:10.1073/pnas.0905845106
- Hollingsworth, P.M., Graham, S.W., Little, D.P., 2011. Choosing and Using a Plant DNA Barcode. *PLoS One* 6. doi:10.1371/journal.pone.0019254
- Ivanova, N.V., Kuzminova, M.L., 2013. Protocols for DNA dry storage and shipment at room temperature. *Molecular Ecology Resources*. 13 (5), 890-898 doi:10.1111/1755-0998.12134
- Margoliash, E., 1963. Primary structure and evolution of Cytochrome C. *Natl Acad Sci U S A* 50, 672–679.

- Meyer, C., Paulay, G. 2005. DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology* 3:e422
- Monaghan, M.T., Wild, R., Elliot, M., Fujisawa, T., Balke, M., Inward, D.J., Lees, D.C., Ranaivosolo, R., Eggleton, P., Barraclough, T.G., Vogler, A.P. 2009. Accelerated species inventory on Madagascar using coalescent-based models of species delineation. *Systematic Biology* 58, 298-311
- Nock, C.J., Waters, D.L.E., Edwards, M.A., Bowen, S.G., Rice, N., Cordeiro, G.M., Henry, R.J., 2011. Chloroplast genome sequences from total DNA for plant identification. *Plant Biotechnology Journal* 9, 328–333. doi:10.1111/j.1467-7652.2010.00558.x
- Pace, N.R., Stahl, D.A., Lane, D.J., Olsen, G.J., 1986. The analysis of natural microbial populations by ribosomal RNA sequences, in: *Advances in Microbial Ecology*. Springer, pp. 1–55.
- Palkopoulou, E., Mallick, S., Skoglund, P., Enk, J., Rohland, N., Li, H., Omrak, A., Vartanyan, S., Poinar, H., Götherström, A., Reich, D., Dalén, L., 2015. Complete genomes reveal signatures of demographic and genetic declines in the Woolly Mammoth. *Current Biology* 25, 1395-1400.
- Parks, M., Cronn, R., Liston, A., 2009. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biology* 7, 84. doi:10.1186/1741-7007-7-84
- Pons, J., Barraclough, T.G., Gomez-Zurita, J., Cardoso, A., Duran, D.P., Hazell, S., Kamoun, S., Sumlin, W.D., Vogler, A.P., 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology* 55: 595-609
- Port, J. A., O'Donnell, J. L., Romero-Maraccini, O. C., Leary, P. R., Litvin, S. Y., Nickols, K. J., Yamahara, K. M. and Kelly, R. P. 2016). Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA. *Mol Ecol*, 25: 527–541. doi:10.1111/mec.13481
- Quispe-Tintaya, W., White, R.R., Popov, V.N., Vijg, J., Maslov, A.Y., 2015. Rapid Mitochondrial DNA Isolation Method for Direct Sequencing, in: Weissig, V., Edeas, M. (Eds.), *Mitochondrial Medicine*. Springer New York, New York, NY, pp. 89–95.
- Smith, M.A., Wood, D.M., Janzen, D.H., Hallwachs, W., Hebert, P.D.N., 2007. DNA barcodes affirm that 16 species of apparently generalist tropical parasitoid flies (Diptera, Tachinidae) are not all generalists. *PNAS* 104, 4967–4972. doi:10.1073/pnas.0700050104
- Stenset, N.E., Lutnæs, P.N., Bjarnadóttir, V., Dahle, B., Fossum, K.H., Jigsved, P., Johansen, T., Neumann, W., Opseth, O., Rønning, O., Steyaert, S.M.J.G., Zedrosser, A., Brunberg, S., Swenson, J.E., 2016. Seasonal and annual variation in the diet of brown bears *Ursus arctos* in the boreal forest of southcentral Sweden. *Wildlife Biology* 22, 107–116. doi:10.2981/wlb.00194
- Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., Vermet, T., Corthier, G., Brochmann, C., Willerslev, E., 2007. Power and limitations of the chloroplast trn L (UAA) intron for plant DNA barcoding. *Nucl Acids Res* 35, e14–e14. doi:10.1093/nar/gkl938
- Thomsen, P.F., Elias, S., Gilbert, M.T.P., Haile, J., Munch, K., Kuzmina, S., Froese, D.G., Sher, A., Holdaway, R.N., Willerslev, E., 2009. Non-Destructive Sampling of Ancient Insect DNA. *PLOS ONE* 4, e5048. doi:10.1371/journal.pone.0005048
- Thomsen, P.F., Kielgast, J., Iversen, L.L., Møller, P.R., Rasmussen, M., Willerslev, E., 2012. Detection of a Diverse Marine Fish Fauna Using Environmental DNA from Seawater Samples. *PLOS ONE* 7, e41732. doi:10.1371/journal.pone.0041732
- Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P. F., Bellemain, E., Besnard, A., Coissac, E., Boyer, F., Gaboriaud, C., Jean, P., Poulet, N., Roset, N., Copp, G. H., Geniez, P., Pont, D., Argillier, C., Baudoin, J.-M., Peroux, T., Crivelli, A. J., Olivier, A., Acqueberge, M., Le Brun, M., Møller, P. R., Willerslev, E. and Dejean, T. 2016 Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Mol Ecol*, 25: 929–942. doi:10.1111/mec.13428
- Wheeler, Q.D., 2004. Taxonomic triage and the poverty of phylogeny. *Philos Trans R Soc Lond B Biol Sci* 359, 571–583.

- Willems, W.R., Curini-Galletti, M., Ferrero, T.J., Fontaneto, D., Heiner, I., Huys, R., Ivanenko, V.N., Kristensen, R.M., Kånneby, T., MacNaughton, M.O.M., Martínez Arbizu, P., Todaro, M.A., Sterrer, W., Jondelius, U., 2009. Meiofauna of the Koster-area, results from a workshop at the Sven Lovén Centre for Marine Science (Tjärnö, Sweden). *Meiofauna Marina* 17, 1-34.
- Yang, Z., Rannala, B. 2010. Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences* 107: 9264-9269.
- Yoccoz, N.G., Bråthen, K.A., Gielly, L., Haile, J., Edwards, M.E., Goslar, T., Von Stedingk, H., Brysting, A.K., Coissac, E., Pompanon, F., Sønstebo, J.H., Miquel, C., Valentini, A., De Bello, F., Chave, J., Thuiller, W., Wincker, P., Cruaud, C., Gavory, F., Rasmussen, M., Gilbert, M.T.P., Orlando, L., Brochmann, C., Willerslev, E., Taberlet, P., 2012. DNA from soil mirrors plant taxonomic and growth form diversity. *Molecular Ecology* 21, 3647–3655. doi:10.1111/j.1365-294X.2012.05545.x
- Zhang, C., Rannala, B., Yang, Z. 2014. Bayesian species delimitation can be robust to guide-tree inference errors. *Systematic Biology* 63: 993-1004
- Zhang, C., Zhang, D.X., Zhu, T., Yang, Z. 2011. Evaluation of a Bayesian coalescent method of species delimitation. *Systematic Biology* 60: 747-761
- Zouros, E., Freeman, K.R., Ball, A.O., Pogson, G.H., 1992. Direct evidence for extensive paternal mitochondrial DNA inheritance in the marine mussel *Mytilus*. *Nature* 359, 412–414. doi:10.1038/359412a0

Bilaga

Bilaga 1. Streckkodssekvenser av svenska arter i BOLD. Utförlig redovisning av analysresultaten. Arkiv (zip-fil) som kan laddas ned från <http://www.nrm.se/download/18.2a85cf3215e50a71056a553a/1509373421630/Lyrholm-smaskriftserien.zip>.

Tidigare utgivet i samma serie:

1. Förgiftar vi naturen? Tom Lötmarker 1966
2. Djuriskt/mänskligt beteende. Lennart Steen & Lars Fält 1967
3. Tandens i kultur, fantasi och verklighet. Tor Ørving 1968
4. Dinosaurier från Kina: dinosauriernas värld. Krister Brood 1989
5. Den svenska Sydpolsexpeditionen 1901 – 1903. Krister Brood 1989
6. Inventering av nissöga (*Cobitis taenia*) i Edsviken, Stockholms län, 2004. Basinventering inom Edsvikensamarbetet och Natura 2000. PM från Forskningsavdelningen, Naturhistoriska riksmuseet. 2004:1. Stefan Lundberg & Bo Delling 2004
7. Inventering av stormusslor i Albysjön, Tyresö kommun, 2004. Basinventering inom Tyresåsamarbetet. PM från Forskningsavdelningen,, Naturhistoriska riksmuseet. 2004:2. Stefan Lundberg 2004
8. Inventering av bottenfaunan i bäck mellan Flaten och Drevviken, Stockholms stad 2004. En naturvärdesbedömning utifrån bottenfaunans artrikedom. PM från Forskningsavdelningen, Naturhistoriska riksmuseet. 2004:3. Erland Dannelid & Stefan Lundberg 2004
9. Bottenfaunan i Sätträån, Stockholms stad 2004. Utvecklingen efter ett år med kontinuerligt vattenflöde. PM från Naturhistoriska riksmuseet. 2005:1. Christina Ekström & Stefan Lundberg 2005
10. Brunkullan (*Nigritella nigra*) i Jämtland och Härjedalen. Ekologi, populationsutveckling och skötsel aspekter. Slutrapport för "Aktion Brunkulla". PM från Naturhistoriska riksmuseet. 2005:2. Folke Björkbäck & Jim Lund vist 2005
11. Bottenfaunan i fem vattendrag runt Edsviken. Resultat från undersökningar 2004. PM från Naturhistoriska riksmuseet. 2006:1. Stefan Lundberg & Christina Ekström 2006
12. Inventering av stormusslor i Edsån, 2005. Basinventering inom Oxundaåns vattenvårdsprojekt. PM från Naturhistoriska riksmuseet. 2006:2. John Tapper & Stefan Lundberg 2006
13. Inventering av stormusslor i Fysingen, 2005. Basinventering inom Oxundaåns vattenvårdsprojekt. PM från Naturhistoriska riksmuseet. 2006:3. John Tapper & Stefan Lundberg 2006
14. Liv i vattnet vid Tisnaren. Bottenfaunaundersökningar i Tisnarens vattenområde, 2001. PM från Natur historiska riksmuseet. 2006:4. Stefan Lundberg & Urban Pettersson 2006
15. Miljöbokslut 2006. Naturhistoriska riksmuseets miljöledningssystem. PM från Naturhistoriska riksmuseet. 2007:1. Stefan Lundberg & Yvonne Arremo 2007
16. Mälarens stormusselfauna. Resultat från inventeringar längs Mälarens stränder. PM från Naturhistoriska riksmuseet. 2007:2. Stefan Lundberg & Ted von Proschwitz 2007

17. Mälarens stormussel fauna. Lokalbeskrivningar. PM från Naturhistoriska riksmuseet. 2007:3.
Stefan Lundberg & Ted von Proschwitz 2007
18. Miljöövervakningsstrategi för stormusslor. Utveckling av nationell miljöövervakning för sötvattenslevande stormusslor 2008. PM från Naturhistoriska riksmuseet. 2008:1.
Stefan Lundberg & Jakob Bergengren 2008
19. Inventering av stormusslor i Svennevadsån - Skogaån, Örebro län, 2007 - 2008: Miljöövervakning och utredning av åtgärdsbehov. PM från Naturhistoriska riksmuseet. 2008:2.
Stefan Lundberg, Urban Pettersson & John Tapper 2008
20. Miljöbokslut 2007, Naturhistoriska riksmuseets miljöledningssystem. PM från Naturhistoriska riksmuseet. 2008:3.
Stefan Lundberg & Yvonne Arremo 2008
21. Street Life under ytan. Resultat från dykinventering i Fyrisån inom Uppsala stad 2008. PM från Naturhistoriska riksmuseet. 2008:4.
Stefan Lundberg 2008
22. Miljöbokslut 2008, Naturhistoriska riksmuseets miljöledningssystem. PM från Naturhistoriska riksmuseet. 2009:1.
Stefan Lundberg & Yvonne Arremo 2009
23. DNA - baserade metoder för taxonomisk bestämning ("DNA barcoding"): Potentiella tillämpningar för effektivare miljöövervakning. PM från Naturhistoriska riksmuseet. 2009:2.
Thomas Lyrholm 2009
24. Genomförda naturvårdsåtgärder för bevarande av tjockskalig målarmussla *Unio crassus* i Svennevadsån - Skogaån, Örebro län, 2009. PM från Naturhistoriska riksmuseet. 2009:3. Naturhistoriska riksmuseets småskriftserie.
Stefan Lundberg, Urban Pettersson & John Tapper 2009
25. Uppföljning av naturvårdsåtgärder för bevarande av tjockskalig målarmussla *Unio crassus* i Svennevadsån - Skogaån, Örebro län, 2010. PM från Naturhistoriska riksmuseet. 2010:1. Naturhistoriska riksmuseets småskriftserie.
Stefan Lundberg & Urban Pettersson 2010
26. Nationell miljöövervakning av stormusslor i Norasjön, Södertälje kommun, 2010. PM från Naturhistoriska riksmuseet. 2010:2. Naturhistoriska riksmuseets småskriftserie.
Stefan Lundberg, Bo Ljungberg & Erik Wijnbladh 2010
27. Djurplankton i Tyresöfjärdarna: resultat från en undersökning i juni 2012. PM från Naturhistoriska riksmuseet. 2013:1. Naturhistoriska riksmuseets småskriftserie.
Jan-Erik Svensson & Stefan Lundberg 2013
28. Plankton i Tyresö-Flaten och Albysjön. Resultat från en undersökning i augusti 2013. PM från Naturhistoriska riksmuseet. 2014:1. Naturhistoriska riksmuseets småskriftserie.
Jan-Erik Svensson & Stefan Lundberg 2014