

Infrastruktur för DNA-streckkodning av svensk flora och fauna

Författare:

Erik Ersmark och Fredrik Ronquist, Enheten för bioinformatik och genetik,
Naturhistoriska riksmuseet

Sammanfattning

Den snabba utvecklingen inom DNA-teknik har ökat behovet av en effektiv och välintegrerad infrastruktur för DNA-streckkodning. I samarbete med SweBOL byggs nu en sådan upp på Naturhistoriska riksmuseet, inklusive både ett digitalt och fysiskt DNA-arkiv. Kopplingen mellan de två arkiven sker genom en rutinmässig skanning av DNA-extrakt, och en modul för registrering av digitala data, SeqDB. För att möta en ökande mängd analyser är systemet även utformat för att kunna automatiseras i framtiden. DNA-extrakt samlas nedfrysta i ett eget arkiv i Miljöprovbanken, medan DNA-data sparas och integreras i det nya digitala systemet för samlingshantering, DINA-Web. I samband med detta kopplas dessa DNA-arkiv för den svenska floran och faunan också till relevanta internationella nätverk.

Bakgrund.....	3
Dagsläget.....	3
DNA-laboratoriet vid NRM.....	3
CGI.....	4
Fysiskt DNA-arkiv.....	4
Digitalt DNA-arkiv.....	4
Naturarv och Svenska DNA-nyckeln.....	5
Internationella databaser.....	5
Ett DNA-streckkodat exempel från samlingarna - NRM20036012 ..	6
Ny infrastruktur för DNA-streckkodning.....	11
Uppgradering av det fysiska DNA-arkivet.....	11
Uppgradering av det digitala DNA-arkivet.....	13
Nya publika gränssnitt och internationella databaser.....	14
Slutord.....	16
Referenser.....	16

Bakgrund

Sverige har idag mycket goda förutsättningar för storskalig och högkvalitativ DNA-streckkodning av den inhemska floran och faunan (Hovmöller et al. 2017). Med DNA-streckkodning avser vi här framtagandet av referensbibliotek för standardiserad genbaserad bestämning. Vid Naturhistoriska riksmuseet (NRM) är en modern infrastruktur för insamling, sekvensering och lagring av DNA-streckkoder och andra DNA-markörer under uppbyggnad. Många delar används redan, medan andra håller på att etableras. Infrastrukturen består av både digitala och fysiska system för hantering, lagring och tillgängliggörande av streckkodsdata. Vi beskriver här dagsläget för infrastrukturen, samt den mer integrerade infrastruktur som är under etablering.

Utvecklingen av infrastrukturen för DNA-streckkodning vid Naturhistoriska riksmuseet sker i samarbete med SweBOL (Swedish Barcoding of Life), som har sitt sekretariat vid museet. SweBOL samordnar det svenska arbetet med streckkodning inom ramen för de internationella initiativ som finns inom området. SweBOL drivs idag med små resurser, vilket tyvärr innebär att det svenska streckkodningsarbetet fortfarande för närvärande till stor del består av små isolerade forskardrivna projekt. Några större streckkodningsprojekt har dock genomförts de senaste åren.

Dagsläget

Grunden för verksamheten vid Naturhistoriska riksmuseet är samlingarna, som idag uppgår till över elva miljoner enskilda föremål. Samlingarna utgör också grunden för den forskning som bedrivs vid museet. Materialet i samlingarna är inte minst värdefullt för de projekt som använder genetiska analysmetoder. Även om dessa metoder förvisso har varit etablerade sedan länge, har de utvecklats snabbt på senare tid och gjorts billigare och mer effektiva. Resultatet är en växande betydelse för DNA-analyser som metod, inte bara för forskningen, utan också för tillämpningar inom miljöövervakning och förvaltning. Här har DNA-streckkoder en stor betydelse, eftersom de möjliggör en effektiv identifiering av arter och, tillsammans med andra DNA-data, populationer. Analyserna kan exempelvis handla om sökandet efter en specifik rödlistad eller invasiv art, eller om att identifiera artsammansättningen i olika typer av miljöprover.

DNA-laboratoriet vid NRM

Idag DNA-analyseras en stor mängd material vid museet. Prover kan tas från vävnader av olika typer och av olika ålder, från moderna till historiska, och till och med från subfossila lämningar. Analyserna sker vid museets DNA-laboratorium, som drivs av Enheten för bioinformatik och genetik (BIO), och som fungerar som en resurs för hela museet. Även forskare från andra institutioner använder ofta laboratoriet. Här finns utrustning och kapacitet för de flesta typer av analyser, inklusive DNA-streckkodning. I dagsläget sker en stor del av sekvenseringen med traditionell Sanger-teknik. Allt fler projekt använder också sekvensering med den nya NGS-tekniken ("Next Generation Sequencing"), exempelvis för analyser av miljöprover – så kallad "metabarcoding". Vid NGS-

analyser utförs själva sekvenseringen i samarbete med SciLifeLab, medan de förberedande stegen och analyserna av resultaten utförs vid museet.

-
DNA-laboratoriet har deltagit i eller drivit flera större streckkodningsprojekt, bland annat fokuserade på den internationella fiskfaunan (FishTrace; <https://fishtrace.jrc.ec.europa.eu/project>), svenska faunan av ryggradsdjur (<http://dna-nyckeln.se>), och delar av svenska faunan av fjädermygggor (Brodin et al. 2013; www.boldsystems.org sök BSCHI).

Centrum för genetisk identifiering (CGI)

En resurs som är knuten till DNA-laboratoriet och BIO är Centrum för gGenetisk iIdentifiering (CGI) (<http://www.nrm.se/forskingochsamlingar/centrumforgenetiskidentifiering.18738.html>). CGI startades 2013 och är en serviceverksamhet som vänder sig till andra myndigheter och som erbjuder hjälp med olika typer av DNA-analyser. Verksamheten är uppdragsfinansierad, och kan också åta sig uppdrag inom museet, t.ex. gällande DNA-streckkodning eller analyser av miljöprover. I och med att många av uppdragen involverar just artidentifiering, finns en stor erfarenhet av arbete med olika streckkodsmärkörer och matchning av dessa mot referensdatabaser.-

Fysiskt DNA-arkiv

Redan idag finns en stor mängd provtaget material och DNA-extrakt lagrade på laboratoriet. För att lagringen ska kunna ske långsiktigt med maximalt bevarande av DNA, förvaras extrakten nedfrys. CGI har ett särskilt ansvar för de DNA-extrakt som används specifikt för DNA-streckkodning, och dessa förvaras i ett separat DNA-arkiv (Fig. 3). För att kunna expandera arkivet, har mer omfattande lagringsutrymme nyligen erhållits i Miljöprovbanken (MPB), som annars främst lagrar biologiskt material som analyseras för miljögifter. MPB drivs av Enheten för miljöforskning och -övervakning på museet.

Digitalt DNA-arkiv

De DNA-sekvenser och -data som genereras från samlingsmaterial har tidigare inte rutinmässigt inkluderats i några av museets arkiv eller samlingssystem. För att möjliggöra detta, pågår nu ett arbete i samband med att de digitala samlingssystemen uppdateras. Målet är att all information om samlingsobjekt ska vara digitalt tillgänglig och öppen. För DNA-streckkoder är det här av yttersta vikt, då det krävs en tydlig koppling till ett voucher-exemplar (belägg-exemplar), som kan lokaliseras och verifieras med avseende på art och ursprung. Att kunna koppla samlingsobjekt till DNA-extrakt har också ett stort värde, då detta möjliggör replikering av DNA-analyser eller att analysera extrakt med nya metoder.

I ett första steg i den här processen har de befintliga DNA-streckkoderna (sekvenser och data över voucher-exemplaren) framtagna vid museet importerats till "DINA-Specify", det centrala samlingshanteringssystem som idag används för ett antal av museets delsamlningar. DINA-Specify har tagits fram av DINA-projektet (DINA står för Digitalt Informationssystem för NAturhistoriska samlingar: https://www.dina-project.net/wiki/Welcome_to_the_DINA_project), som syftar till att bygga upp en nationell digital infrastruktur för naturhistoriska

samlingar. Utvecklingsarbetet inom DINA-projektet är idag inriktat på att ta fram nästa generation av systemet, kallat "DINA-Web", vilket bland annat omfattar betydligt bättre funktionalitet för hantering och tillgängligheten av DNA-data för forskare och andra nyttjare. Utvecklingsarbetet drivs tillsammans med internationella partners [som ingår i projektet](#), däribland Museum für Naturkunde i Berlin och Agriculture and Agri-Food Canada (AAFC) i Ottawa.

-DINA-Web beskrivs i större detalj nedan; här fokuserar vi på den nuvarande versionen, DINA-Specify. DINA-Specify (också kallad "DINA-Light" eller "DINA-Hybrid") består av ett antal komponenter utvecklade av DINA-projektet runt en kärna bestående av det amerikanska samlingshanteringssystemet Specify (<http://www.sustain.specifysoftware.org/>). Datalagringen sker i huvudsak i Specify-databasen, och dagligt samlingsarbete sker genom Specify-klienten, en Java-klient som körs på den lokala datorn. Till detta kommer fler web-gränssnitt utvecklade för att tillgängliggöra samlingsdata, däribland samplingsportalen Naturarv (<http://naturarv.se>) samt streckkodsportalen Svenska DNA-nyckeln (<http://dna-nyckeln.se>). DINA-projektet har också utvecklat en specialiserad klient för inventeringsprojekt, vilken idag används för Svenska Malaisefälleprojektets inventering av den svenska insektsfaunan (<http://www.stationlinne.se/sv/forskning/the-swedish-malaise-trap-project-smtp/>).

Vid registrering av samlingsföremål noteras, utöver ett individuellt katalognummer [\(ex. NRMXXXXXXX\)](#), vanligtvis även datum för insamling och registrering, geografiskt ursprung, samt vem som samlade in föremålet. Ofta finns ytterligare metadata, såsom koordinater, och även fotografier bifogat (Fig. 1). För DNA-streckkodade föremål, så kallade "voucher-exemplar", ställs extra höga krav på de data som registreras. Bland annat ska lokalen koordinat-sättas, föremålet ska fotograferas, och bestämningen av föremålet ska vara utförligt dokumenterad.

Naturarv och Svenska DNA-nyckeln

I andra ändan av systemet, där digitala data från samlingarna görs tillgängliga och sökbara, finns i dagens DINA-Specify ett system som innefattar en mängd naturhistoriska samlingar från hela Sverige, samlade i sökportalen Naturarv (<http://naturarv.se>). Naturarv omfattar alla föremål registrerade i DINA-Specify, inklusive de föremål som har DNA-streckkodats. En särskild portal, Svenska DNA-nyckeln (<http://dna-nyckeln.se>), har skapats för de streckkodade föremålen [och är under vidareutveckling](#). Genom portalen kan man bland annat söka med DNA-sekvenser mot det referensbibliotek för svenska ryggradsdjur som togs fram i ett FORMAS-finansierat projekt lett av museet. DNA-nyckeln tillåter också sökningar mot internationella databaser över streckkoder, som BOLD (www.boldsystems.org) och GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>).

Internationella databaser

Samtliga data från svenska samlingar som finns i Naturarv, exporteras regelbundet till det internationella nätverket GBIF (Global Biodiversity Information Facility) tillsammans med en mängd andra biodiversitetsdata från Sverige (se <http://www.gbif.se/portal/#/index> och <https://www.gbif.org/country/SE/summary>). GBIF samlar information från en mängd databaser och fynddata är representerade både i form av museiföremål, men även som inrapporterade observationer. [I dagsläget finns i allmänhet](#)

~~ingen~~[GBIF har nyligen introducerat möjligheten att dela](#) information om eller länk till eventuella DNA-data kopplade till föremålen i ~~GBIF, men sådana möjligheter är prioriterat att~~ [introduceras av GBIF internationellt inom kort](#) (GBIF Norge har redan börjat [implementera det](#) [utnyttja den nya funktionaliteten](#) (– se [www.gbif.no](#)) och andra länder beräknas följa under de kommande åren, däribland Sverige.)

De DNA-data som tas fram från museiföremål rapporteras vanligtvis i första hand till de etablerade internationella databaser som ingår i INSDC-samarbetet (International Nucleotide Sequence Database Collaboration), mest känt för amerikanska GenBank), där de görs tillgängliga och sökbara. Ofta sker detta i samband med publicering av de projekt där dessa data ingår. För DNA-streckkoder erbjuder BOLD (Barcode of Life Data Systems) och INSDC databaser som exklusivt ägnas dessa, och där man kan söka med t.ex. artnamn, katalognummer eller med en streckkodssekvens. Alla streckkodssekvenser ges här också ett specifikt sekvens-ID. DNA-streckkoder framtagna vid museet laddas rutinmässigt upp till både BOLD och INSDC.

Ett DNA-streckkodat exempel från samlingarna – NRM20036012

För att på ett tydligare sätt illustrera hur samlingsföremål som DNA-streckkodats idag representeras i de olika systemen, används nedan som exempel en varfågel (*Lanius excubitor*) från museets ornitologiska samling med katalognummer NRM20036012 (Fig. 1). Fågeln samlades in i Själevad, Ångermanland, i december 2002. Året därpå preparerades skinnet (skinn, vinge och partiellt skelett) på museet och katalogfördes av Ingrid Cederholm.



Figur 1: NRM20036012 i bild - fotografi av varfågelns preparerade skinn.

All denna data kan återfinnas i DINA-Specify (Fig. 2), där även en del ytterligare detaljer finns registrerade, som GPS-koordinater för insamlingsplatsen och fotografier.

I samband med [ett det Formas-finansierade](#) initiativet att streckkoda alla arter av svenska ryggradsdjur, extraherades DNA från varfågeln i november 2006. DNA-streckkoden (648bp från CO1) som togs fram, registrerades i BOLD (sekvens-ID: BISE353-08.CO1-5P) och Genbank (GU571937) följande år. Den gjordes också sökbar i Svenska DNA-nyckeln.

Förutom fåglar, så DNA-streckkodades även en mängd fiskar och däggdjur [i samma projekt](#). Alla DNA-extrakt samlades i prövrör som märktes med rörnummer och ursprungsföremålens katalognummer. Samlingen av extrakt-förvaras nedfrysta som en del av museets fysiska DNA-arkiv. Varje DNA-extrakt är kopplat till motsvarande voucher-föremål genom det ursprungliga katalognumret (Fig. 3).

I samband med DNA-streckkodningen, lades information om DNA-extrakten till i Specify (Fig. 2). På så sätt blev det möjligt att hitta föremål som i samlingarna inte bara bestod av fysiska vävnader ([skinn mm](#)), utan också utvunnet DNA.

▼ Collection Object

Catalog Number: 20036012	Accession:	Alt Cat Number:
Cataloger: Cederholm, Ingrid	<i>i</i>	Cataloged Date: 2003-01-14
		Project Number:

▼ Determinations

Taxon: Lanius excubitor	<i>i</i> <input checked="" type="checkbox"/> Current	
Preferred Taxon: Lanius excubitor		
Determined Date:	Determiner:	
		Type Status: None

14 ◀ [1 of 1] ▶ 11

Grid ▾

▼ Collecting Information

Field Number:	Method:	0
Start Date: 2002-12-17	End Date:	
Locality: Själevad; Ångermanland, Sweden; 63.2835000000, 18.6471000000		
Locality and Kartblad: 1914i		
Habitat Notes:		

▼ Collectors

Last Name	First Name	Remarks

▼ Preparations

Prep Type	Is On Loan	Preparation Attachments	Count	Prepared By	Storage
DNA	No				VE-DNA, NRM Vert...
partial skeleton	No				VE, NRM Vertebratz...
Skin	No				VE, NRM Vertebratz...
spread wing	No				VE, NRM Vertebratz...

Form ▾

▼ Col Obj Attribute

Sex: Male	Weight:	ReproCondition:	Age:
Body:	Stomach:	Gonads:	Molt:
Remarks:			

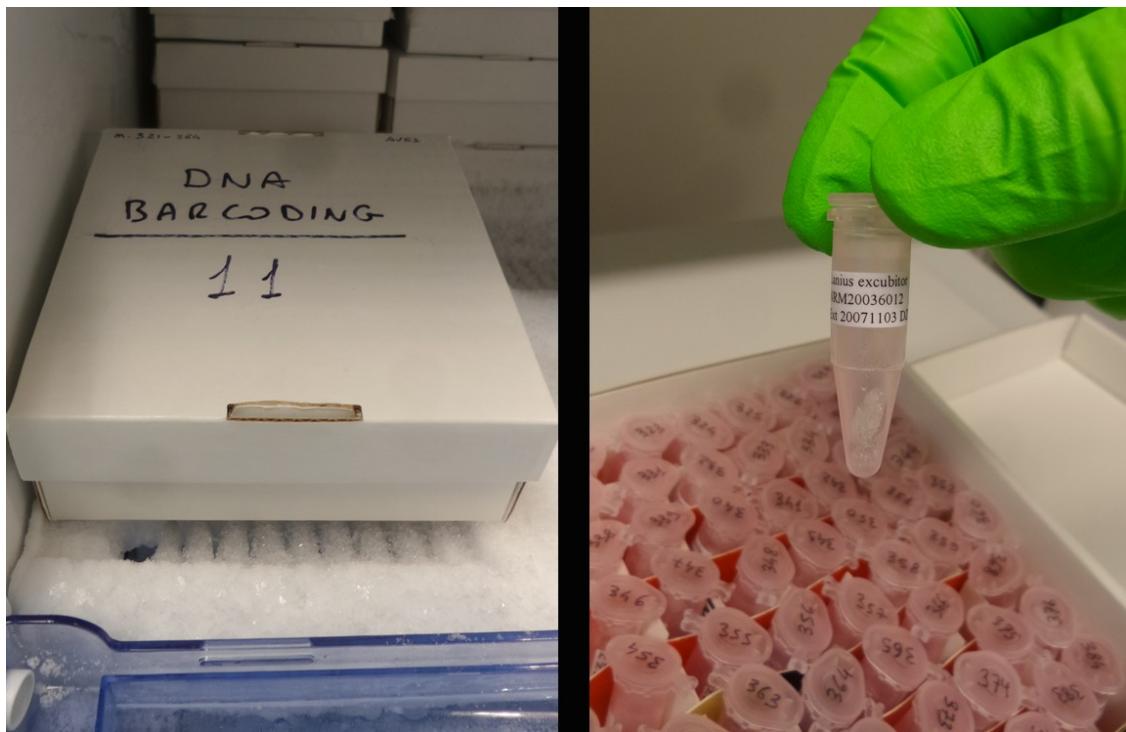
Attachments

▼ Collection Object Attachments *i* Grid ▾

NRM-20036012_dors.JPG	NRM-20036012_vent.JPG
-----------------------	-----------------------

Modified By Agent: Englund, Markus Modified Date: 2013-12-30

Figur 2: Data om varfågeln (NRM20036012) registrerade i Specify. DNA-extraktet finns noterat här (röd pil).



Figur 3: Dagens nedfrusna arkiv för DNA-streckkodat material från museet. Till höger ses DNA-extraktet tillhörande varfågeln, märkt med latinskt namn, katalognummer och datum för extraktionen (2007-11-03).

Den digitala informationen om varfågeln från Specify finns tillgänglig genom sökportalen Naturarv, där den bland annat är beskriven som ett "föremål med DNA-sekvens" (Fig. 4). För att hitta den sekvenserade DNA-streckkoden kan man till exempel söka med fågelns katalognummer i BOLDs databas. Här hittar man inte bara den övergripande informationen om varfågeln, som insamlingsdatum och -plats (Fig. 5a), utan även ingående information om DNA-streckkoden, som har ett specifikt sekvens-ID (BISE353-08.COI-5P) (Fig. 5b), och som enkelt kan laddas ned.

Varfågelns DNA-streckkod går också att hitta genom den Svenska DNA-nyckeln (Fig. 6). Här kan man söka direkt med en frågesekvens, som jämförs med en referensdatabas med data från svenska ryggradsdjur, men även med BOLDs och GenBanks databaser. På det här sättet kan man lätt via länkar hitta all data som finns i de olika databaserna för en viss art. De sekvenser som härrör från samlingsföremål listas med katalognummer länkade till Naturarv.

The screenshot shows the Naturarv website interface. At the top, there's a banner with images of museum specimens and the word 'Naturarv'. Below the banner, there are navigation links: 'Startsida', 'Samlingar', 'Samarbetspartners', 'Vanliga frågor och svar', 'Om Naturarv', 'Svenska', and 'English'. The main content area has a search bar with 'lanius' typed in, an 'Avancerad sökning' link, and a 'Valda: 1' section. Under 'Valda: 1', there's a checkbox 'Ta bort alla' and a list of detailed taxon information for Lanius excubitor. To the right, there are sections for 'Lägg till filter:' (with links to 'Föremål med koordinater', 'Typföremål', 'Föremål med bilder', 'Föremål insamlade i Sverige', and 'Föremål med DNA-sekvens'), 'Samling' (with links to 'NRM Entomology Coll...' and 'NRM Ornithology'), and 'Institution' (with a link to 'Naturhistoriska riksmuseet'). At the bottom, there are links for 'Visa på karta', 'Sök taxon i Naturforskaren', 'Rapportera fel', and a page number '(1 of 1)'.

Figur 4: Data om varfågeln i sökportalen Naturarv.

I GenBank återfinns varfågelns DNA-streckkod enklast via dess unika GenBank-ID ("GenBank accession number"), som både listas- och länkas i Svenska DNA-nyckeln och i BOLDs databas. I GenBanks databas finns också möjlighet att söka med frågesekvenser, artnamn och andra ID-nummer. Databasen är dock inte begränsad till streckkodsssekvenser, utan innehåller en stor variation av olika typer av DNA-sekvenser. Standardiserade streckkodsssekvenser identifieras dock genom nyckelordet 'BARCODE'.

I och med att nya DNA-sekvenser görs tillgängliga i GenBank vid publicering av vetenskapliga projekt, hittar man ofta ytterligare sekvensdata från en viss individ eller ett visst voucher-exemplar. När det gäller varfågelns streckkodssekvens så är den i GenBank kopplad till en publicerad artikel (Johnsen et al. 2010). Genom att söka med museets katalognummer och det latinska namnet kan man även hitta ytterligare två DNA-sekvenser från varfågeln, som tillsammans ingår i en annan artikel (Olsson et al. 2010).

BOLD SYSTEMS

DATABASES IDENTIFICATION TAXONOMY WORKBENCH RESOURCES LOGIN 

Record Details For BISE353-08 [Back to Search: Records](#)

IDENTIFIERS

Sample ID: BISE-Aves367 Museum ID: NRM20036012
Field ID: Collection Code:
Deposited In: Swedish Museum of Natural History

Tags

Comments

TAXONOMY

Phylum: Chordata Subfamily:
Class: Aves Genus: *Lanius*
Order: Passeriformes Species: *Lanius excubitor*
Family: Laniidae Subspecies:
BIN (Cluster ID): BOLD:AAC7309

* Barcode Index Numbers(BIN): cluster barcode sequences to create OTUs that closely reflect species groupings

SPECIMEN IMAGES: N/A

COLLECTION SITE:



A map of Northern Europe showing the location of the collection site in Sweden. The map includes labels for Norway (Norge), Sweden (Sverige/Sweden), Finland (Suomi/Finland), and Estonia (Eesti/Estonia). Major cities like Oslo, Stockholm, Helsinki, and Tallinn are marked. A red dot indicates the collection site near Stockholm. The map also shows the Gulf of Bothnia and various rivers and lakes.

Figur 5a: Övergripande data om varfågeln i BOLDs databas.

Figur 5b: DNA-streckkod från varfågeln i BOLDs databas.



Sökresultat

Nedan följer en tabell med resultatet av sökningen i en av DNA-nyckelns databaser. De träffsekvenser med högst likhet listas först. Klicka på de givna länkarna för mer information om den art som är mest lik söksekvensen.

Sökningen innehåller 1 frågesekvenser och är gjord med sökalgoritmen BLASTN 2.2.30+

[Spara resultatet \(csv\)](#)

[Ny sökning](#)

Sökningen är gjord i databasen Svenska ryggradsdjur (COI, 16S) [1882 sekvenser, 1122642 baspar]

Fråga 1 : 2017-12-21 10:20:21.933
Längd : 648

Sekvenser med stor likhet:

Katalognr	Vetenskapligt namn	Markör	GenBank acc.	BOLD sekv. ID	Likhet (%)	Poäng	E-värde
0 996541	Lanius excubitor	CO1	GU571938	B1SE146-08.CO1-SP.CO1	100	1169.87	0.0
0 2003012	Lanius excubitor	CO1	GU571937	B1SE353-08.CO1-SP.CO1	100	1169.87	0.0
Visa ytterligare träffar som är mindre lika (<99%)							
0 986403	Lanius collaris	CO1	B1SE265-08.CO1-SP.CO1	92	925.491	0.0	
0 20016267	Lanius collaris	CO1	B1SE086-07.CO1-SP.CO1	91	888.544	0.0	
0 20016248	Corvus frugilegus	CO1	B1SE413-08.CO1-SP.CO1	88	830.836	0.0	

Figur 6: Genom att söka med streckkodssekvensen i Svenska DNA-nyckeln kunde varfågeln hittas, med länkar till Naturarv, BOLD och GenBank.

Ny infrastruktur för DNA-streckkodning

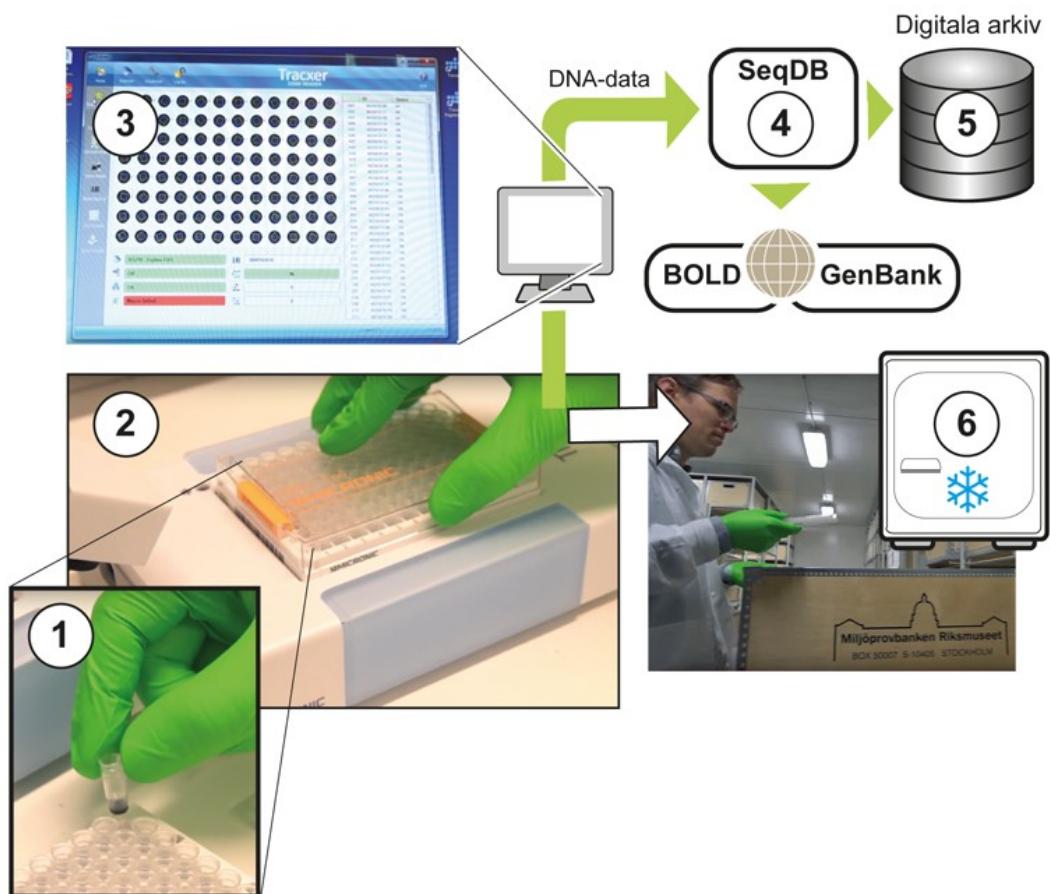
Både den fysiska och digitala infrastrukturen för DNA-streckkodning vid museet uppgraderas för närvarande. När det gäller den fysiska infrastrukturen så syftar uppgraderingen framför allt till att förbättra möjligheterna till rationell och i framtiden höggradigt automatiserad hantering av stora mängder DNA-prover. På den digitala sidan är syftet att skapa verktyg för smidig hantering av sekvensdata. Specify-systemet erbjuder idag inget grafiskt gränssnitt för hantering av sekvensdata, utan sekvensdata måste laddas in i databasen av databasexperter. Publicering av och länkning till sekvensdata i BOLD och INSDC/GenBank måste också skötas helt manuellt. Vi beskriver i korthet de viktigaste nyheterna nedan.

Uppgradering av det fysiska DNA-arkivet

Museet har nyligen förvärvat ett nytt system för rationell hantering av stora mängder prover producerat av Micronic (<https://www.micronic.com/>). Systemet bygger på att små prövrör med förtryckta identifierare (QR-koder) samlas i plattor som rymmer 96 rör vardera. Samtliga rör i en platta skannas och registreras digitalt, direkt i laboratoriet, med hjälp av en speciell scanner (Tracxer Code Reader RS210 CRYO) och tillhörande mjukvara (Tracxer Code Reader Software) (Fig. 7).

Vid skanningen tilldelas varje prövrör/extrakt på detta sätt automatiskt ett individuellt ID-nummer (NRM-DNA-XXXXXX). ID-numret knyts automatiskt till data som tidigare registrerats på laboratariet i samband med DNA-streckkodningen, omfattande katalognummer på samlingsföremålet från vilket extraktet kommer, samt grundläggande data som datum och namn på den som genomfört extraktionen. Dessa data importeras sedan automatiskt till DINA-Webs modul för DNA-data, SeqDB (se nedan), där föremålsdata och DNA-data knyts samman.

Micronic-rören är försedda med proppar som tillåter att man i efterhand kan föra in en injektionsnål och ta ut en del av DNA-extraktet för ytterligare analyser utan att behöva öppna själva rören. Därmed minimeras risken för hopblandning och kontaminering av extrakten. Plattorna kan även hanteras av robotsystem, vilket gör att framtida analyser av DNA-extrakten kan automatiseras.



Figur 7: 1) DNA-extrakt i rör med individuella QR-koder samlas på en platta, också den är märkt med en streckkod. 2) Skannern läser av både

streckkod och QR-koder. 3) Allt registreras digitalt med programmet Traxcer. 4) I SeqDB sker en koppling till katalognummer samt övrig data. DNA-data kan också registreras i olika format och direkt skickas till internationella databaser. 5) All digital data sparas och görs tillgänglig inom DINA-Web. 6) Plattorna med DNA-extrakt frysas ner och samlas i det nya DNA-arkivet i Miljöprovbanken.

Det nya systemet för registrering av DNA-extrakt innebär också införandet av nya rutiner. Praktiska beskrivningar finns redan nu samlade i en instruktionsmanual i DNA-laboratoriet, med utrymme för framtida korrigering och komplettering. En särskild manual gällande registrering av DNA-streckkoder utarbetas också av företrädare för SweBOL, för att säkerställa att all nödvändig information inkluderas i dessa fall.

Plattorna med DNA-extrakt placeras i det nya fysiska DNA-arkivet i Miljöprovbanken. Eftersom de är både små och kompakta, finns en kapacitet att lagra hundratals extrakt här. Även det befintliga arkiv som idag lagras i vanliga provrör (Fig. 3) kommer successivt att överföras till de nya plattorna och inkorporeras i DNA-arkivet. Den första plattan med befintliga DNA-extrakt har nyligen förts över till det nya systemet, men det kommer att ta tid innan hela DNA-arkivet har gått igenom.

Uppgradering av det digitala DNA-arkivet

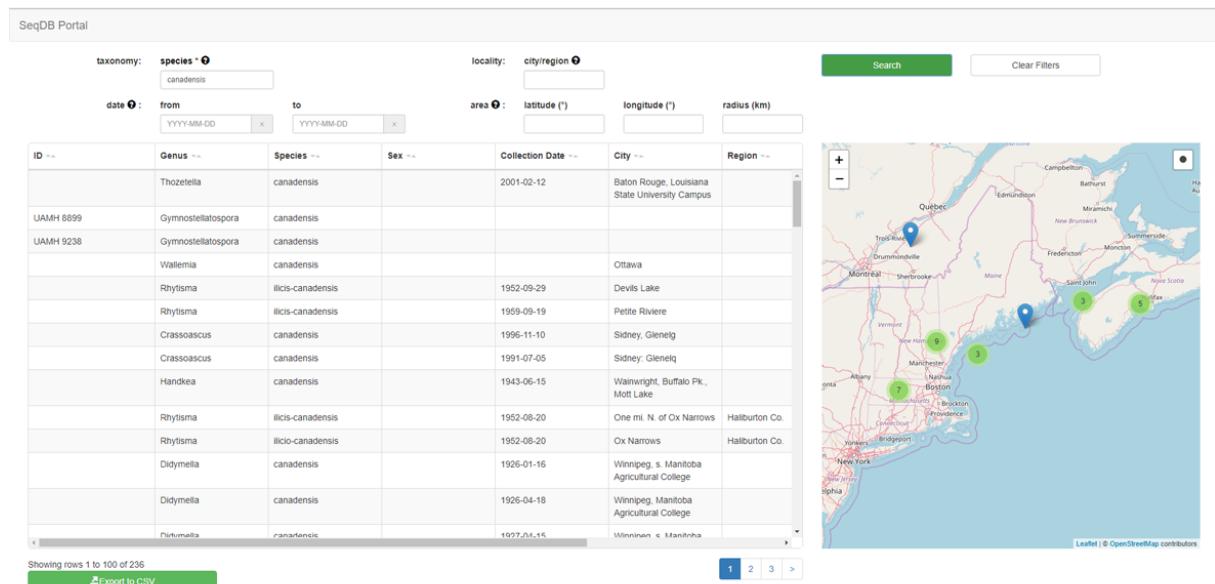
Inom ramen för DINA-Web utvecklas en särskild modul för hantering av DNA-sekvensdata. Den bygger på ett befintligt system, SeqDB, ursprungligen utvecklat av AAFC. SeqDB integreras nu av DINA-konsortiet med samlingshanteringen i DINA-Web, och anpassas för de behov av funktionalitet som finns på bland annat Naturhistoriska riksmuseet, där SeqDB redan är i drift sedan en tid. SeqDB kommer därmed att utgöra navet för den framtida digitala registreringen av DNA-sekvenser framtagna på museet.

SeqDB är redan från början fokuserat på datahanteringen i laboratorier för DNA-analyser av naturhistoriska samlingsföremål och miljöprover. Till skillnad från Specify har SeqDB grafiska, web-baserade gränssnitt, och gränssnitten stöder de flesta aktuella arbetsflödena (Fig. 8). SeqDB stöder till exempel både DNA-streckkodning och NGS-analyser av miljöprover, inklusive automatisk uppladdning av data till BOLD och INSDC/GenBank.

I samarbete mellan AAFC och Naturhistoriska riksmuseet har SeqDB nyligen fått stöd för automatisk import av data från Micronic-systemet. CGI använder också sedan en tid SeqDB i arbetet med det så kallade björnprojektet, där genotypning används för att identifiera individuella djur. Parallelt med arbetet på SeqDB håller DINA-konsortiet på att ta fram ett modernt, helt web-baserat system för samlingshantering, som kommer att ersätta Specify-delen i det nuvarande DINA-Specify-systemet. Arbetet med denna Collection Manager (CM) inom DINA-Web drivs av Naturhistoriska riksmuseet. Idag utvecklas en pilot-version av systemet för däggdjurssamlingen på museet, varefter andra samlingar runtom i Sverige står på tur. Målet är att utvecklingsarbetet på sikt kommer att omfatta samtliga svenska samlingar inom DINA-konsortiet.

All data som registreras och samlas i SeqDB, kommer även att kopplas till CM-delen av DINA-Web (Fig. 9). Förutom de funktioner för samlingshantering som redan finns i det nuvarande systemet, DINA-Specify, kommer det på så sätt bli

möjligt att få tillgång till och arbeta med all DNA-data som är knutna till samlingsföremålen.



Figur 8: Exempel från SeqDB:s publika sökportal med variabler som artnamn, geografiskt område och tidsperiod för insamling.

Nya publika gränssnitt och internationella databaser

På sikt kommer omdaningen av den digitala infrastrukturen att innebära nya publika gränssnitt. Inom ramen för DINA-projektet utvecklas nu en publik sökportal för genotypdata framtagna av CGI, vilken kopplas direkt till data i SeqDB. På sikt kommer Svenska DNA-nyckeln också att behöva omstöpas så att den kopplar till sekvensdata i SeqDB. I samband med detta är ambitionen att lägga till nya sökfunktioner, som att söka i DNA-streckkodsbiblioteken med artnamn och katalognummer.

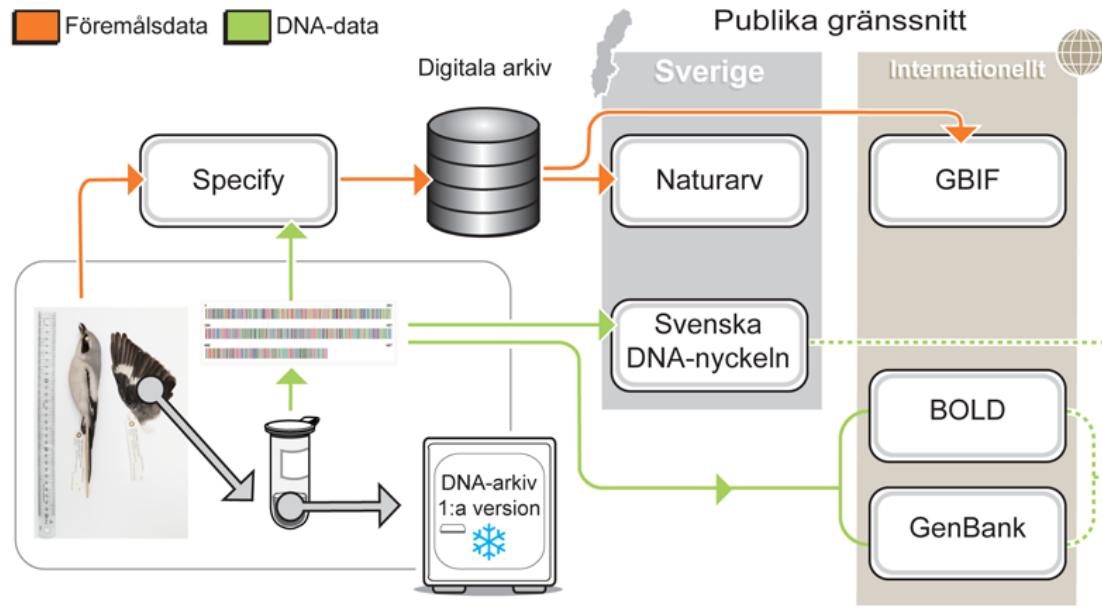
Vetenskapsrådet har nyligen givit Naturhistoriska riksmuseet i uppdrag att samordna införandet av en ny forskningsinfrastruktur för svenska biodiversitetsdata. Infrastrukturen går under namnet Svensk biodiversitetsatlas eller Biodiversity Atlas Sweden (BAS) (se <http://bioatlas.se>) och bygger på internationellt samarbete inom GBIF-nätverket och med gruppen av länder, institutioner och utvecklare runt Atlas of Living Australia (<http://ala.org.au>). BAS-systemet innehåller många av de funktioner som finns i Naturarv, och det är sannolikt att Naturarv på sikt ersätts av BAS, eller funktionalitet som utvecklas inom ramen för BAS-projektet. Genom BAS blir det möjligt att få tillgång inte bara till naturhistoriska samlningar, utan många andra typer av biodiversitetsdata, från

artobservationer och fotografier från allmänheten till molekylära biodiversitetsdata (Fig. 9).

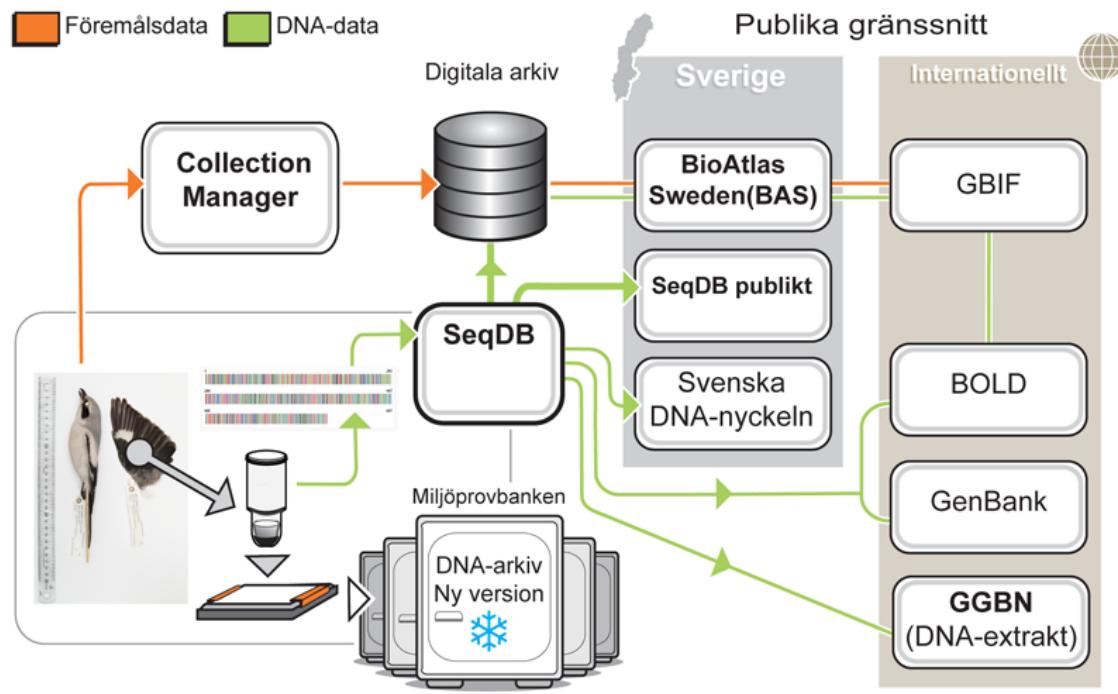
DINA-Web kommer precis som DINA-Specify att kontinuerligt göra data tillgängliga genom ett antal relevanta internationella nätverk. Publicering av sekvensdata till BOLD och INSDC/GenBank kommer att underlättas genom SeqDB-modulen, som beskrivits ovan. SeqDB innehåller också funktioner för att publicera data om vilka DNA-extrakt som finns tillgängliga i det lokala DNA-arkivet till nätverket GGBN (Global Genome Biodiversity Network). Detta internationella nätverk bildades 2011 just för att göra DNA-data mer tillgänglig, med tydliga kopplingar mellan publikationer, DNA-data och voucher-exemplar (Droege et al. 2013).

Data kommer också fortsättningsvis att göra tillgängliga via GBIF, och vi förväntar oss att GBIF framöver kommer att göra det möjligt att dela betydligt fler datatyper än man gör idag, inte minst olika typer av DNA-data.

DINA-Specify



DINA-Web



Figur 9: Jämförelse av hur föremåls- och DNA-data hanteras i dagens system (DINA-Specify), och i de nya system som etableras (DINA-Web)

Slutord

Redan idag kan Naturhistoriska riksmuseet erbjuda en omfattande fysisk och digital infrastruktur för DNA-streckkodning, och betydande erfarenhet av större DNA-streckkodningsprojekt. På senare år har mycket av den svenska DNA-streckkodningen skett utomlands, eftersom stora streckkodningsprojekt i bland annat Norge och Tyskland har kunnat erbjuda mycket fördelaktiga villkor och någon samlad finansiering för svensk streckkodning inte funnits. I en del av dessa projekt deponeras DNA-extrakt, föremål och tillhörande metadata på ett adekvat sätt i de fysiska och digitala arkiven på Naturhistoriska riksmuseet. Detta är dock upp till de enskilda forskare som driver projekten. Det finns således behov för ett nationellt program för streckkodning.

Genom arbete på fler fronter moderniseras nu den fysiska och digitala infrastrukturen för DNA-streckkodning på museet successivt. Delar av den nya infrastrukturen är redan implementerade och andra följer successivt efter under de kommande åren. Framförallt kommer den nya infrastrukturen att innehåra betydligt bättre möjligheter för museet att på ett rationellt sätt hantera stora mängder prover och därmed också kunna genomföra större projekt ~~som bygger~~ på omfattande DNA-analyser, inte minst DNA-streckkodningsprojekt.

Referenser

- Brodin, Y., Ejdung, G., Strandberg, J., & Lyrholm, T. (2013). Improving environmental and biodiversity monitoring in the Baltic Sea using DNA barcoding of Chironomidae (Diptera). *Molecular Ecology Resources*, 13(6), 996-1004.
- Droege, G., Barker, K., Astrin, J. J., Bartels, P., Butler, C., Cantrill, D., ... & Mackenzie-Dodds, J. (2013). The global genome biodiversity network (GGBN) data portal. *Nucleic acids research*, 42(D1), D607-D612.
- 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*.
- Johnsen, A., Rindal, E., Ericson, P. G. P., Zuccon, D., Kerr, K. C., Stoeckle, M. Y., & Lifjeld, J. T. (2010). DNA barcoding of Scandinavian birds reveals divergent lineages in trans-Atlantic species. *Journal of Ornithology*, 151(3), 565-578.
- Olsson, U., Alström, P., Svensson, L., Aliabadian, M., & Sundberg, P. (2010). The *Lanius excubitor* (Aves, Passeriformes) conundrum—Taxonomic dilemma when molecular and non-molecular data tell different stories. *Molecular Phylogenetics and Evolution*, 55(2), 347-357.