

Database, 2018, 1–8 doi: 10.1093/database/bay006 Original article



# Original article

# The NCBI BioCollections Database

# Shobha Sharma\*, Stacy Ciufo, Elena Starchenko, Dakshesh Darji, Larry Chlumsky, Ilene Karsch-Mizrachi and Conrad L. Schoch

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, 45 Center Drive, Bethesda, MD 20892, USA

\*Corresponding author: Tel: 301-594-2606; Fax: 301-480-2918; Email: sharmash@ncbi.nlm.nih.gov

Citation details: Sharma,S., Ciufo,S., Starchenko,E. *et al.* The NCBI BioCollections Database. *Database* (2018) Vol. 2018: article ID bay006; doi:10.1093/database/bay006

Received 30 October 2017; Revised 5 January 2018; Accepted 8 January 2018

# Abstract

The rapidly growing set of GenBank submissions includes sequences that are derived from vouchered specimens. These are associated with culture collections, museums, herbaria and other natural history collections, both living and preserved. Correct identification of the specimens studied, along with a method to associate the sample with its institution, is critical to the outcome of related studies and analyses. The National Center for Biotechnology Information BioCollections Database was established to allow the association of specimen vouchers and related sequence records to their home institutions. This process also allows cross-linking from the home institution for quick identification of all records originating from each collection.

Database URL: https://www.ncbi.nlm.nih.gov/biocollections

# Introduction

The BioCollections Database is a curated dataset of metadata for culture collections, museums, herbaria and other natural history collections connected to sequence records in GenBank. It is maintained and curated by the Taxonomy group at the National Center for Biotechnology Information (NCBI). Biocollection institution codes are unique across multiple types of collections and the database is used to support the 'structured voucher' annotation in the sequence entries submitted to International Nucleotide Sequence Database Collaboration (INSDC) (1). This broadly follows the Darwin Core (DwC) standard for biodiversity data (2) and is used to standardize usage across interconnected databases including GenBank the NCBI (3), as well as the European Nucleotide Archive (ENA) (4) and DNA Databank of Japan (DDBJ) (5).

Initially, the data were imported from Index Herbariorum (6), World Federation for Culture Collections (http://www. wfcc.info/), Insect and Spider Collections of the World (http:// hbs.bishopmuseum.org/codens/codens-r-us.html), Amphibian Species of the World (AMNH) (7) and the Catalog of Fishes (8). Only the institution codes that are listed in the BioCollections Database appear as 'structured voucher' in GenBank records. New repository records are added to the database as they are submitted to INSDC along with sequence data. Since the BioCollections Database is maintained at NCBI, the validation process is fast. Prior to inclusion in BioCollections, the new collections are validated to ensure that they are curated, are readily available to the public and

there is a contact person responsible for the collections. If a home institution has a catalogue page and provides us with URL formula, the vouchers in the sequence entries are hotlinked to specimen pages at the relevant collection (Figure 1). Personal collections are not normally included. Other directories of repositories are periodically reviewed to ensure that the NCBI BioCollections is up-to-date.

As the importance of specimen vouchers in biodiversity studies continues to grow, it is increasingly important to organize and annotate the data to allow users to easily access this information and confirm which collection houses the original sample. This newly released public resource is the source for building links between NCBI databases and external collections.

# **BioCollections Database overview**

In 2005, the Consortium for the Barcode of Life (CBOL; http://www.barcodeoflife.org) proposed linking sequence records to voucher specimens as part of the DNA Barcode data standard. This method was developed in collaboration with the Global Biodiversity Information Facility (http://www.gbif.org/) and other major biodiversity database initiatives. The NCBI BioCollections Database was created as a part of this global project to gather, update, manage and search biological collections information. In mid-2008, members of INSDC started annotating sequence entries that contained culture collection or specimen voucher information with structured voucher qualifiers.

The initial method proposed for linkage by CBOL used a structured data format based on the DwC data standards developed by the Biodiversity Information Standards (TDWG, formerly the Taxonomic Database Working Group). The DwC standard Triplet format for specimen data consists of three parts: the universally-recognized code for the institution that holds the voucher specimen; the institution's code for the collection in which the voucher specimen is kept and the unique specimen identifier, all separated by colons.

For example:

/organism='Spizella atrogularis' /specimen\_voucher='MVZ: Bird: 170231'

In many cases, a secondary collection code (such as a collection devoted to mammals or plants at a specific institution) is not utilized and in such cases the specimen data is indicated as a doublet only.

/organism='Enterococcus flavescens' /culture\_collection='ATCC: 49996'

This structured data field for voucher specimens was approved by the members of the INSDC in May 2005.

Structured Voucher Annotation:

There are three different types of qualifiers for annotating sequences from different source materials:

- 1. /culture\_collection for live microbial and viral cultures and cell lines deposited in curated culture collections.
- 2. /specimen\_voucher for a physical specimen in a curated museum, herbarium, frozen tissue collection or in laboratory (accessible to public). If the specimen was destroyed in the process of sequencing, electronic images (e-vouchers) are an adequate substitute for a specimen voucher.
- /bio\_material for source material in biological collections that do not fit into either the/specimen\_voucher or the/culture\_collection modifier categories, like physical specimens from zoos, aquaria, stock centers, germplasm repositories and DNA banks.

Another set of qualifiers may contain information from BioCollections. Submitters commonly use these fields to add voucher information but they are not 'structured,' hence, they don't get linked to Biocollections Database.

- 1. /isolate is recommended to identify specific individuals or samples from which the sequence data was originally obtained—this can include field numbers and a broad set of unique identifiers that will not be classified under strain or culture collection.
- 2. /strain is recommended for cultures in personal collection or laboratory.
- 3. /note for any comment or additional information about the organism.

Until recently, the BioCollections Database was only used internally by the members of INSDC, mainly to facilitate sequence annotation, although a public text-based data file was (and remains) available (ftp://ftp.ncbi.nih.gov/pub/ taxonomy/Cowner\_dump.txt). Over the years, the database has grown significantly. Each record now provides information about the institution that houses the collection, standard institution code, mailing address and associated webpage if available. If there are collections within an institution, they are listed within the institution record as collection codes. As of October 2017, there are over 7400 institution codes and ~300 collection codes listed in the BioCollections Database. Recognizing that this information can be useful to a broader scientific community, NCBI released this resource to the public in April of 2017.

# Search and retrieve data

Various search queries can be used to search the BioCollections Database using the search box on the BioCollections homepage. For example, searching with

(a) Spizella atrogularis voucher MVZ:Bird:170231 cytochrome oxidase subunit 1 (COI) gene, partial cds; mitochondrial

Sales Stabil								
<u>Go to:</u> [\]								
LOCUS DEFINITION	DQ433192 Spizella atrogularis vouc	bp DNA linea her MVZ:Bird:170231 cyt	ar VRT 14-JUL-201 tochrome oxidase	6				
ACCESSION	subunit 1 (COI) gene, par po433192	tial cds; mitochondrial						
VERSION	DQ433192.1							
SOURCE	BARCODE. mitochondrion Spizella at	rogularis (Black-chinne	ed sparrow)					
ORGANISM	Spizella atrogularis							
	Archelosauria; Archosauri	a; Dinosauria; Saurisch	hia; Theropoda;					
	Coelurosauria; Aves; Neog	nathae; Passeriformes;	Passerellidae;					
REFERENCE	1 (bases 1 to 599)							
AUTHORS	Kerr, K.C., Stoeckle, M.Y., Hebert, P.D.	Dove, C.J., Weigt, L.A.,	Francis, C.M. and					
TITLE	Comprehensive DNA barcode	coverage of North Amer	rican birds					
JOURNAL	Mol. Ecol. Notes 7 (4), 5	35-543 (2007)						
REFERENCE	2 (bases 1 to 599)							
AUTHORS	Hebert, P.D.N.	., Dove,C.J., Weigt,L.A	A., Francis, C.M. an	a				
TITLE	Direct Submission	indimentity Tentityte -	E Catanta					
JOURNAL	Department of Integrative	Biology, University of	f Guelph, Guelph,					
PRATINES	Ontario N1G2W1, Canada							
source	1599							
	/organism="Spize /organelle="mito	lla atrogularis" chondrion"						
	/mol_type="genom	to DNA						
	/db_xref="BOLD:C	DMV2038-05"						
	/db_xref="taxon: /country="USA: C	40208" alifornia"	X					
	/lat_lon="35.357	8 N 120.305 W"	URL rule to	link to MVZ page				
	/collected by="K	="20-May-1984" http://	/mvzarctos.berkeley	edu/SpecimenDetail.c	fm?guid=			
	/PCR_primers="fw	d_seq: ttctccaaccacaaag	gacattggcac,					
gene	<1>599	agataattccaaatcctg"						
CDS	/gene="COI"							
	/gene="COI"							
	/codon_start=1 /transl_table=2							
	/product="cytoch	rome oxidase subunit 1"	•					
	/product="cytoch /protein_id="ABK /translation="MV	rome oxidase subunit 1" <u>29354.1</u> " GTALSLLIRAELGQPGALLGDDQ	QVYNVIVTAHAFVMIFFMV	м				
	<pre>/product="cytoch /protein_id="ABK /translation="MV PIMIGGRONL/PIMI PPLAONLAHAGASVDL VLITAVLLLSLPVLA</pre>	rome oxidase subunit 1" 29354.1" GTALSLLIRAELGQPGALLGDDQ GAPOMAFRYGNNMSFWLLPFSFL AIFSLHLAGISSILGAINFITTA AGITMLLTDRNLNTTFFDPAGGG	27 ynv i vtahafvmi ffwy Lllasstveagvgtgify Ainmkppalsgygtplfvw 3″	M Y S				
M	/product*oytooh /protein_id="ABK /translation="WV PIMIOFONNUVPIMI PPLAGNLARAGASVDL VLITAVLLLLSLPVLA	rome oxidase subunit 1* 29354.1* GRINASLIRAELGOPGALLGDDG GAPDAAFFRANNMSFWLLPFST IFSLHLASTSIGAINFITT AGITMLLTDRNLNTTFFDPAGGO	QYINVIVTAHAFVMIFFMV LLLIASSTVERGYGTGMTV LINKEPPALSQYQTPLFVW 7"	M Y S				
	/product="cytoch /protein_id="ABK /translation="Wy PIMOGROMUVEMI PELAGNLAHAGASVOL VLITAVLLLSLPVIA	rome oxidase subunit 1* 29354.1* GTALSLIRAELGOPGALLGOD GARDMAFPROMONSFULJEPST AIFSLMLAGISSILGAINFITTA AGITMLLTORNLNTIFFDPAGGG	, 201NV IVTAHAFVMI FFMV LLLASTVERGVGTGMTV LINKEPPALGQYQTPLFVN 3"	M Y S NEWS: Arctos Workir	ng Group M	7		
Search	/product="cytoch /protein_id="ABK /translation="WV PIMOGROMUVPIMI PPLAONLAWAGASVOL VLITAVLLLLSLPVIA Collections Database MUSEUM OF VERTEBRATE Portals My Stuff /	rome oxidase subunit 1* 29354_1* GTALSLLTRAELGQPGALLGDDQ GAENAAFRANNGSFWLLPFSTI AFTSHLAGSTSUGAINFITTA AGITMLLTDRNLNTTFFDPAGGG EZOOLOGY boddHeip	, 201nu ivtahafumi ffmu Lllasstverguvernitu Innkfpalsquqtplfum 2 <sup>4</sup>	M Y S NEWS: Arctos Workir	ng Group M	]		
Search MVZ:Bird Spizella a	/product="cytoch /protein_id="ABK /translation="Wy PIMORPOWINPEMI PPLAONLAWAGASYDL VLITAVLLLSLPVLA Collections Database MUSEUM OF VERTEBRATE Ponals My Stuff / :170231 atrogularis cana	rome oxidase subunit 1" 2935.1.1" OTALSULTRAENGOPGALIADOD GENMAFFRANDMESFULFPST- AIFSLHLAGISSILGAINFITTA AGITMLLTDRNLNTTFFDPAGGG EZOOLOGY bootHeip La Panza Range, North America, Uni 20 May 1984	2025WIVTAHAFWIIFBW LLLLASTVEACVGTGWTV LLLLASTVEACVGTGWTV J <sup>9</sup> <sup>9</sup> 1 mi NW Pozo sumr ited States, California	M S NEWS: Arctos Workin nit a, San Luis Obispo Cor	ng Group M	]	tissue (fr	ozen); skeleto
Search MVZ:Bird Spizella a Identification	/product="cytoch /protein_id="ABK /translation="Wy PIMORPOWINVEM PPLAONLAWAGASYDL VLITAVLLLSLPVLA Collections Database MUSEUM OF VERTEBRATE Ponals My Stuff / :170231 atrogularis cana	rome oxidase subunit 1* 2935.1.1* OTALSULTRAENGOPAALLAODO GADNAEFRAMMORSHULPSTI AIFSLHLAGISSILGAINFITTA AGITMLLTDRNLNTTFFDPAGGG EZOOLOGY UboutHeip La Panza Range, North America, Uni 20 May 1984	2025WIVTAHAFVMIERW LILLASTVEACVGTONTV LILLASTVEACVGTONTV J <sup>9</sup> <sup>9</sup> 1 mi NW Pozo sumr ited States, California	M Y S NEWS: Arctos Workin nit a, San Luis Obispo Cor Collector(s)	ng Group M	]	tissue (fr	ozen); skeletor
Search WVZ:Bird Spizella a Identification Spizella at	/product="cytoch /protein_id="ABK /translation="WV PTMIGPENNUVPINI PPLASNLAHAGASVDL VLITAVLLLLSLPVIA MUSEUM OF VERTEBRATE Portals My Stuff / :170231 atrogularis cana	rome oxidase subunit 1* 2935.1.1* GRADSULTRAENGOPEALLADDO GAPDMAFFRIGNESFMLLPPSFL AFFOMLASTSUGAINTITA AGITMLLTDRNLATTFFDPAGGG EZOOLOGY bboutHelp La Panza Range, North America, Uni 20 May 1984	20YINYIVTAHAPVMIPPAV LLLLASTVPACVOTONTV LLLLASTVPALSQYQTPLEVN 3" (1 1 mi NW Pozo summ ited States, California	M S NEWS: Arctos Workin nit a, San Luis Obispo Cor Collector(s) Ned K. Johnson	ng Group M unty	)	tissue (fr	ozen); skeletor
search Search MVZ:Bird Spizella a Identification Spizella a Animalia; Black-chin	/product="cytoch /protein_id="ABK /translation="WV PIMOGROMUVPLAI PPLAONLAHAGASVDL VLITAVLLLLSLPVIA Collections Database MUSEUM OF VERTEBRATE Portals My Stuff / :170231 atrogularis cana IS Chordats, Aves, Passenformes, Pat Gorido Latabase	rome oxidase subunit 1* 29354.1* GRAISLIRAELGOPGALLGOO GRADMAFPRIMOSTILLEPST AFFEMILAGISSILGAINFITTA AGITHLLTORNLNTTFFDPAGGG EZOOLOGY booutHeip La Panza Range, North America, Un 20 May 1984	20YIN IVTAHAPVMI PPM LLLLASSTVEACVOTONTV LLLLASSTVEACVOTONTV INNERPALSQYQTPLEVN 3" ( 1 mi NW Pozo summ ited States, California cona	MY S NEWS: Arctos Workin nit a, San Luis Obispo Cor Collector(s) Ned K. Johnson Identifiere	ng Group M unty n 19	)	tissue (fr	ozen); skeleto
Search Spizella a Identification Spizella a Spizella a Black-chimalar, Black-chimalar, Black-chimalar,	/product="cytoch /protein_id="ABK /translation="WV PTMOGROMUVPLAI PPLAONLAHAGASYOL VLITAVLLLUSLPVIA Collections Database MUSEUM OF VERTEBRATE Portais My Stuff / :170231 atrogularis cana Its Chordata: Aves. Passenformes: Para Chordata: Aves. Passenformes: Para Mod Sparrow; Gorión barba negra do Sparrow; Gorión barba negra	rome oxidase subunit 1* 29354.1* GTALSILIRAELGOPGALLGOO GADMAFPENGHNSFULJEPST AFFELMLAGISSILGAINFITTA AGITMLLTORNLNTTFFDPAGGG E ZOOLOGY bboutHeip La Panza Range, North America, Uni 20 May 1984 userellidae, Spizella atogularis o boology, University of Califo	oversi uzasta Post Person Liblas Stveza over control Liblas Stveza over control Liblas Person over control Liblas Presson over control over the states, California cana cana conta, Berkeley on	M Y S NEWS: Arctos Workin nit a, San Luis Obispo Cor Collector(s) Ned K. Johnson Identifiers GenBank: DO4	ng Group M unty n IB 133192 IB	)	tissue (fr	ozen); skeleto
Search Search MVZ:Bird Spizella a Spizella at Animala, Black-chin Identifice 1999-01	/product="cytoch /protein_id="ABK /translation="Wy PEIAGROMILVEMI PPLAGNIAHAGASYOL VLITAVLLLSLPVIA Collections Database MUSEUM OF VERTEBRATE Portais My Stuff 2170231 atrogularis cana ts rogularis cana Chordata; Aves, Passenformes; Pas ned Sparrow, Gorido barba negra d by Museum of Vertebrate Zu -27	rome oxidase subunit 1* 2935.1.1* GTALSILIRAELGOPGALLGODG GADNAEPBOINDASFULJEPST AFFELALAGISSILGAINETITA AGITMLLTORNIANTIFEDPAGGG E ZOOLOGY UboutHelp La Panza Range, North America, Uni 20 May 1984 userellidae, Spizella atrogularis o coology, University of Califo	1 mi NW Pozo summited States, California cana	M Y S NEWS: Arctos Workir nit a, San Luis Obispo Cor Collector(s) Ned K. Johnson Identifiers GenBank: DQ4 MVZ frozen tiss	ng Group M unty n IE 133192 IB sue collection:	) FC-21528	tissue (fr	ozen); skeleto
Search Search MVZ:Bird Spizella a Spizella at Animala, Black-chin Identifice 1999-01 Nature c	/product="cytoch /protein_id="ABK /translation="Wy PIMOGROMUVPIAI PPLAGNLARAGASVOL VLITAVLLLSLPVIA Collections Database MUSEUM OF VERTEBRATE Portais My Sturt Portais My Sturt 170231 atrogularis cana Chordata; Avec, Passenformer; Pat need Sparrow; Gorido harba negra d by Museum of Vertebrate Zot -27 filD: legacy	rome oxidase subunit 1* 2935.1.1* GTALSILIRAELGOPGALIGDO GADNAEPSONAMSFILLEPST AIFSLELAGISSILGAINETITA AGITMLLIDRNLATTFFDPAGGG EZOOLOGY UboutHelp La Panza Range, North America, Uni 20 May 1984 serellidae. Spizella atrogularis coology, University of Califo	operativezanta Ponterenov Lizita Strueza over dever Lizita Strueza over dever ponterenover dever dever ponterenover dever dever ponterenover dever dever dever dever ponterenover dever dever dever dever dever ponterenover dever dever dever dever dever dever ponterenover dever dever dever dever dever dever dever ponterenover dever de ponterenover dever de ponterenover dever deve	M Y S NEWS: Arctos Workir nit a, San Luis Obispo Cor Collector(s) Ned K Johnson Identifiers GenBank: DQ4 MVZ frozen tiss collector number	ng Group M unty n I2 33192 I2 sue collection: er: 5412	) FC-21528	tissue (fr	ozen); skeleto
Search WVZ:Bird Spizella a Identification Spizella a Animaia Black-chir Identification Spizella at Animaia 1999-01 Nature o Location	/product="cytoch /protein_id="ABK /translation="Wy PIMOGROMUVPIAI PPLAGNIAHAGASYDL VLITAVLLLSLPVIA Collections Database MUSEUM OF VERTEBRATE Portais My Stut / :170231 atrogularis cana ts rogularis cana Chordata, Aves, Passenformes, Pas need Sparrow, Comin barba negra d by Museum of Vertebrate Zu -27 yf ID; legacy	rome oxidase subunit 1* 2935.1.1* GTALSILIRAELGOPGALIAGDO GADMAPEPANIMASFILIEPST AFFELELAGISSILGAINEITTA AGITMLITORNIANTTEPOPAGGO E ZOOLOGY La Panza Range, North America, Uni 20 May 1984 userellidae. Spizella atrogularis o boology, University of Califo	yvesvi vzakazyni resolvezowerower Liziastrvezowerowerowero Liziastrvezowerowerowero Liziastrvezowerowerowerowero P 1 mi NW Pozo summ ited States, California cana cana bornia, Berkeley on	NEWS: Arctos Workir nit a, San Luis Obispo Col Collector(s) Ned K. Johnso Identifiers GenBank: DQ4 MVZ frozen tiss collector numbr	ng Group M unty n 2 I33192 2 Sue collection: er: 5412	) FC-21528	tissue (fr	ozen); skeleto
search MVZ:Bird Spizella a Animala; Black-chi Identification Spizella at Animala; Black-chi Identification Spizella at Animala; Black-chi Identification Spizella Animala; Black-chi Identification Spizella Animala; Black-chi Identification Spizella Animala; Black-chi Identification Spizella Animala; Black-chi Identification Spizella Animala; Black-chi Identification Spizella Animala; Black-chi Identification Spizella Animala; Black-chi Identification Spizella Animala; Black-chi Identification Spizella Animala; Black-chi Identification Spizella Animala; Black-chi Identification Spizella Animala; Black-chi Identification Spizella Animala; Black-chi Animala; Black-chi Animala; Black-chi Animala; Black-chi Animala; Black-chi Animala; Black-chi Animala; Black-chi Animala; Black-chi Animala; Black-chi Animala; Black-chi Animala; Black-chi Animala; Black-chi Animala; Black-chi Animala; Black-chi Animala; Black-chi Animala; Black-chi Animala; Animala; Animala; Animala; Animala; Animala; Animala; Animala; Animala; Animala; Anima	/product="cytoch /protein_id="ABK /translation="WV PTMIGPENNUVPINI PPLASILARAGASVDL VLITAVLILLGLEVIA MUSEUM OF VERTEBRATE Portals My Stuff / :170231 atrogularis cana Is rogularis cana Chordata, Avec, Passenformer, Par ned Sparrow, Corrido harba partor Corridat, Avec, Passenformer, Par ned Sparrow, Corrido harba partor Just Sparrow, Corrido harba partor Just Sparrow, Corrido harba partor of Museum of Vertebrate Zr -27	rome oxidase subunit 1* 2935.1.1* GTALSILIRAELGOPGALIAGDO GADNAEPSONMARSHILEPST AIFSLHLAGISSILGAINFITTA AGITMLLTDENLATIFFOPAGGO E ZOOLOGY UboutHeip La Panza Range, North America, Uni 20 May 1984 userellidae, Spizella atrogularis o coology, University of Califo	yven i vzaka Pvali EPavi Lili ASTVEA overen veren ve veren veren v	NEWS: Arctos Workir nit a, San Luis Obispo Cor Collector(s) Ned K. Johnsor Identifiers GenBank: DQA MVZ frozen tiss collector number Parts IPart Name	ng Group M unty n E 133192 E sue collection: er: 5412	FC-21528	tissue (h	ozen); skeletor
search Spizella a Identification Spizella a Spizella a Spizella at Animalia; Black-chi Black-chi Black-chi Starte d Spizella at Animalia; Black-chi Starte d Spizella at Animalia; Black-chi Starte d Spizella at Animalia; Black-chi Starte d Spizella at Animalia; Black-chi Starte d Spizella at Animalia; Black-chi Starte d Spizella at Animalia; Black-chi Starte d Spizella at Animalia; Spizella at	/product="cytoch /protein_id="ABK /translation="WV PTMIOGROMUVPLAI PPLAONLAHAGASVDL VLITAVLLLUSLEVIA Portals My Stuff / :170231 atrogularis cana IS Chordata; Avec, Passenformes, Par ned Sparwo, Comris barba participant d by Museum of Vertebrate Zu -27 d ID: legacy	rome oxidase subunit 1* 29354.1* GRAISLIRAELGOPGALLGOO GRADMAFPRIANDSFILLEPST AFFEMELAGISSILGAINFITTA AGITHLLTORNLNTTFFDPAGGG EZOOLOGY booldegy La Panza Range, North America, Uni 20 May 1984 userellidae. Spizella atrogularis o poology, University of Califo	ovyny u vzaka zvali peny kulta strve a ovorom v a take peak sovorom v a take peak sovorom v a take peak sovorom v a take peak sovorom v ted States, California cana cana ornia, Berkeley on	MY S NEWS: Arctos Workin nit a, San Luis Obispo Cor Collector(s) Ned K. Johnson Identifiers GenBank: DQ4 MVZ frozen fiss collector number Parts Parts Name Skeleton	ng Group M unty n I I I I I I I I I I I I I I I I I I I	FC-21528 Disposition in collection	tissue (fr	ozen); skeletor
Search Spizella a Identification Spizella a Identification Spizella a Identification Spizella a Identification Spizella a Identification Identification Nature of Location	/product="cytoch /protein_id="ABK /translation="Wy PTINGGROMUVPLAI PPLAONLAHAGASYDL VLITAVLLLUSLPVIA Collections Database MUSEUM OF VERTERATE Portals MyStuff / :170231 atrogularis cana Chordata: Aves. Passenformes: Par med Sparrow; Gorián barba negra dby Museum of Vertebrate Zu -27 df ID: legacy	rome oxidase subunit 1* 29354.1* GTALSILIRAELGOPGALLGOO GADMAFPENGHNSFULJEPST AFFELALAGISSILGAINETITA AGITMLLTORNLATTFFDPAGGO E ZOOLOGY bboutHeip La Panza Range, North America, Uni 20 May 1984 Userellidae; Spizella atogularis o boology, University of Califo	oversi uzasta Post Person Liblasstvezavoronery atilskeppal.sovortery "  1 mi NW Pozo summ ited States, California cana comia, Berkeley on	NEWS: Arctos Workir nit a, San Luis Obispo Cor Collector(s) Ned K. Johnson Identifiers GenBank: DQ4 MVZ frozen tiss collector numbr Parts Parts Part Name Skeleton fissue (frozen)	ng Group M unty n B i33192 B sue collection: er: 5412 Condition unchecked unchecked	FC-21528 Disposition in collection in collection	tissue (fr	ozen); skeletor
Search MVZ:Bird Spizella at Animala: Black-chir Identifie 1999-01 Nature of Location Serke Berke	/product="cytoch /protein_id="ABK /translation="Wy PTINGGROMUVELM PPLAGNLARAGASYOL VLITAVLLLUSLPVLA Collections Database MUSEUM OF VERTEBRATE Portals My Stuff Portals My Stuff Portals My Stuff Portals My Stuff Cogularis cana Is rogularis cana Condata, Aves, Passenformes; Pat ned Sparow, Condo harba negra d by Museum of Vertebrate Zo -27 of ID: legacy	rome oxidase subunit 1* 29354.1* GTALSILIPAELGOPGALLGODG GTALSILIPAELGOPGALLGODG GTALSILIPAELGOPGALLGODG E ZOOLOGY boothelp La Panza Range, North America, Uni 20 May 1984 cserellidae; Spizella atrogularis of boology, University of Califordia coology, University of Califordia of collection na Yamamoto on 2006-01-03	oversi uzakazyal zeno Lilasstvezovoromi inimitea.sorozeri e 1 mi NW Pozo summ ited States, California cana ornia, Berkeley on	NEWS: Arctos Workir nit a, San Luis Obispo Cor Collector(s) Ned K. Johnson Identifiers GenBank: DQ4 MVZ frozen tiss collector number Parts Part Name Skeleton tissue (frozen) tissue (frozen)	ng Group M unty n E 33192 E sue collection: er: 5412 Condition unchecked unchecked	FC-21528 Disposition in collection in collection	tissue (fr	ozen); skeletor
Search MVZ:Bird Spizella a Identification Spizella a Animala, Black-chir Identifica 1999-01 Nature c Location Determ Higher	/product="cytoch /protein_id="ABK /translation="Wy PEIAGROMILVEMI PPLAGNIAHAGASYOL VLITAVLLLSLEVIA Collections Database MUSEUM OF VERTEBRATE Portais My Stuff Portais My Stuff i:170231 atrogularis cana Chordata; Aves, Passenformes; Pas need Sparrow, Gorido harba negra d by Museum of Vertebrate Zot -27 of ID: legacy HeyMapper ID ination Type: accepted place- assigned by Kristi r Geography: North America, 1	rome oxidase subunit 1* 2935.1.1* GTAISLIFRENGOPGALIGDO GADNAFPBONANSFILLEPST AFFELELAGISSILGAINETITA AGITMLLTDRNLATTFFDPAGGG E ZOOLOGY UboutHelp La Panza Range, North America, Uni 20 May 1984 serellide. Spizella atrogularis boology, University of California, Uni 20 Collection na Yamamoto on 2006-01-03 United States, California, United States, California, University of California, Universit	anna annna anna anna anna anna anna anna anna anna anna anna	M Y S NEWS: Arctos Workir nit a, San Luis Obispo Cor Collector(s) Ned K Johnson Identifiers GenBank: DQ4 MVZ frozen tiss collector numbr Parts Part Name skeleton tissue (frozen) tissue (frozen)	ng Group M unty n IP 33192 IP sue collection: er: 5412 Condition unchecked unchecked n) unchecked	FC-21528 Disposition in collection in collection transfer of custor	tissue (fr	ozen); skeleto
Search Search MVZ:Bird Spizella a Animalia; Black-chi Identifieti Identifieti Identifieti Identifieti Identifieti Detrem Higher	/product="cytoch /protein_id="ABK /translation="W/ PFINGPRINUVPLAI PPLAGNLAHAGASYDL VLITAVLLLUSLEVIA MUSEUM OF VERTEBRATE Porals My Stuff / :170231 attrogularis cana Is rogularis cana Chordata; Aves, Passenformes; Pat ned Sparw, Corrish patha patha (Di Di legacy HeyMapper Is) ination Type: accepted place assigned by Kristi ; Geography: North America, County Is mes	rome oxidase subunit 1* 2935.1.1* CRAISLITRAEUGOPAALLOOD GAPMAFPRIAMSTRLIPPST AFFSLALAGISSILGAINFITTA AGITALLIDRNLATTFFDPAGGG EZOOLOGY bootHeip La Panza Range, North America, Uni 20 May 1984 userellidae. Spizella atrogularis coology, University of Califo of collection na Yamamoto on 2006-01-03 United States, California, 1	y venu i vzaka Pou Lebov LLLASSTVEA ovorowy LLLASSTVEA ovorowy i ulter a state ovorowy p <sup>o</sup> 1 mi NW Pozo sum/ ited States, California cana cana cana cana San Luis Obispo	NEWS: Arctos Workin  nit  a, San Luis Obispo Cor  Collector(s)  Ned K. Johnson  Identifiers  GenBank: DQ4  MV2 frozen tiss collector numbro  Parts  Parts  Releton  tissue (frozen)  tissue (frozen)  Attributes	ng Group M unty 133192 E sue collection: er: 5412 Condition unchecked inchecked	FC-21528 Disposition in collection fransfer of custo	Issue (fr	ozen); skeletor
Search Spizella a Identification Spizella a Spizella a Spizella a Animalia; Black-chi	/product="cytoch /protein_id="ABK /translation="W/ PFINOGROMUVPLAI PPLAONLAHAGASYDL VLITAVLLLLSLEVIA VLITAVLLLLSLEVIA Portals My Stuff / :170231 atrogularis cana Chordata; Avec, Passenformes; Par med Sparrow, Comris barba med Chordata; Avec, Passenformes; Par med Sparrow, Comris barba med 27 fr ID: legacy HeyMapper Im ination Type: accepted place- assigned by Kristi Geography: North America, Just	rome oxidase subunit 1* 29354.1* GRAISLIFREIGOPRALLODO GRADNAFPRIAMSFILLEPST AFFEMELACISSILGAINFITTA AGITHLLTORNLATTFFDPAGGG EZOOLOGY biboutHelp La Panza Range, North America, Uni 20 May 1984 userellidae. Spizella atrogularis o boology, University of California, serellidae. Spizella atrogularis o boology, University of California, distribution on 2006-01-03 United States, California, e, 1 mi NW Pozo summit	yvyn u vzaka Pvil Penv Lulaast ve aovoron y Lubast ve aovoron y Lubast ve aovoron y Lubast ve aovoron y Lubast ve aovoron y 1 ml NW Pozo summ ited States, California cana cana cana cana sornia, Berkeley on	MY S NEWS: Arctos Workin mit a, San Luis Obispo Cor Collector(s) Ned K. Johnson Identifiers GenBank: DQ4 MVZ fozen fiss collector numbro Parts Part Name Skeleton tissue (frozen) tissue (frozen) Sex. male	ng Group M unty n I I I I I I I I I I I I I I I I I I I	FC-21528 Disposition in collection in collection transfer of custo	tissue (fr	ozen); skeletor
Search  Search  MVZ:Bird Spizella a  Arimala; Black-chin Identification  Location  Berke  Determ Higher  Spe College	/product="cytoch /protein_id="ABK /translation="Wy PTMOGROMUVPLAI PPLAONLAHAGASYDL VLITAVLLLUSLPVIA Collections Database MUSEUM OF VERTERATH Portals MyStuff / itTO231 atrogularis cana Chordata: Aves. Passenformes: Par med Sparrow; Gorión barba negra d by Museum of Vertebrate Zu -27 d ID: legacy leyMapper I ination Type: accepted place- assigned by Kristi r Geography: North America, County I mars cific Locality: La Panza Rang:	rome oxidase subunit 1* 29354.1* GTALSILIRAELGOPGALLGOO GADMAFPENGINSFULJEPST AFFELRIAGISSILGAINETTT AGITMLLTORNLATTFFDPAGGO E ZOOLOGY bootHeip La Panza Range, North America, Uni 20 May 1984 Isserellidae; Spizella atrogularis of boology, University of California, serellidae; Spizella atrogularis of boology, University of California, a Yamamoto on 2006-01-03 United States, California, a, 1 mi NW Pozo summit	yoven u vzaka kyal renov Lizla strvza kovoromy a tibila strvza ovoromy a tibila strvza ovoromy a tibila strvza ovoromy a tibila strvza ovoromy i ted States, California cana cana comia, Berkeley on san Luis Obispo	NEWS: Arctos Workin nit a, San Luis Obispo Cor Collector(s) Ned K. Johnson Identifiers GenBank: D04 MVZ frozen tiss collector numbr Parts Parts Part Name Skeleton tissue (frozen Attributes sex: male Museum of Verte cone of exite	ng Group M unty n B i33192 B sue collection: er: 5412 Condition unchecked ) unchecked n) unchecked brate Zoology, I	FC-21528 Disposition in collection in collection transfer of custor	Tissue (fr Qty 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	eley, 2005-1
Search WVZ:Bird Spizella a Animala Black-chir 1999-01 Nature of Location Serve Determ Higher Spe Colled	/product="cytoch /protein_id="ABK /translation="Wy PTMOGROMUVELAI PPLAONLAHAGASYDL VLITAVLLLUSLPVIA POTAIS MUSEUM OF VERTEBRATE Potais My Stuff Potais Cana Chordata, Aves, Passenformes; Par and Sparrow, Condo barba negra d by Museum of Vertebrate Zo -27 of ID: legacy HeyMapper I ination Type: accepted place assigned by Kristi r Geography: North America, County I merec cific Locality. La Panza Range ting Source: wild caught Event Date: 1984-05-20	rome oxidase subunit 1* 2935.1.1* GTALSILIPARIOGPANLIADO GADMAPPROMINISTILIADO E ZOOLOGY bootherp La Panza Range, North America, Uni 20 May 1984 cserellidae; Spizella atrogularis of boology, University of Califo of collection na Yamamoto on 2006-01-03 United States, California, 3 e, 1 mi NW Pozo summit	yorun i vzaka Pvali Prevo Lizla Strvea ovoromy inimerea Logrope Prevo p  1 mi NW Pozo summ ited States, California cana comia, Berkeley on San Luis Obispo	M Y S S NEWS: Arctos Workir nit a, San Luis Obispo Cor Collector(s) Ned K. Johnson Identifiers GenBank: DO4 MVZ frozen tiss collector number Parts Part Name skeleton tissue (frozen) tissue (frozen) tissue (frozen) Attributes sex: male Museum of Verte age: ad sk.	ng Group M unty n E 33192 E sue collection: er: 5412 Condition unchecked unchecked brate Zoology, I ebrate Zoology, I	FC-21528 Disposition in collection it collection transfer of custo Jniversity of Califor University of Califor	tissue (fr Qty 1 1 1 1 dy 1 nia, Berk	Remarks
Search Search MVZ:Bird Spizella at Animala, Black-chin Identification Spizella at Animala, Black-chin Identification Spizella at Animala, Black-chin Identification Spizella at Animala, Black-chin Identification Spizella at Animala, Black-chin Identification Spizella at Animala, Black-chin Identification Spizella at Animala, Black-chin Identification Nature of Berkee Determ Higher Spie Colleon	/product="cytoch /protein_id="ABK /translation="Wy PELAGNLARAGASYOL VLITAVLLLSLEVIA Collections Database MUSEUM OF VERTEBRATE Portais My Stuff Portais My Stuff i:170231 atrogularis cana Chordata; Aves, Passenformes; Pat ned Sparrow, Gorido barba negra d by Museum of Vertebrate Zu -27 of ID: legacy leyMapper I ination Type: accepted place assigned by Kristi r Geography: North America, Countly II mare cific Locality: La Panza Range ting Source: wild caught Event Date: 1984-05-20 Verbatim Date: 20	rome oxidase subunit 1* 2935.1.1* GTALSILIPARIGOPARALIGDO GTALSILIPARIGOPARALIGDO GENDARPSONAMSTRILEPST AFFELELAGISSILGAINETITA AGITMLLTDRNLATTFFDPAGGG E ZOOLOGY UboutHelp La Panza Range, North America, Uni 20 May 1984 serellide. Spizella atrogularis coology, University of California, UNI 20 May 1984 serellide. California, UNI 20 May 1984	anna annna anna anna anna anna anna anna anna anna anna anna	M Y S S NEWS: Arctos Workir nit a, San Luis Obispo Cor Collector(s) Ned K. Johnson Identifiers GenBank: DQ4 MVZ frozen tiss collector numbr Parts Part Name Skeleton tissue (frozen) tissue (frozen) tissue (frozen) tissue (frozen) Museum of Vert age: ad sk. Museum of Vert age: ad sk.	ng Group M unty n E 33192 B sue collection: er: 5412 Condition unchecked unchecked unchecked brate Zoology, t ebrate Zoology, t	FC-21528 Disposition in collection in collection transfer of custo Jniversity of Califor	tissue (fr Qty) 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	Remarks
Search Spizella a Spizella a Identification Spizella a Animalia; Black-chi B	/product="cytoch /protein_id="ABK /translation="WW PTMOGROMUVPLAI PPLAONLAHAGASYDL VLITAVLLLELEVIA WISEUM OF VERTEBRATE Portals My Stuff / introgularis cana is rogularis cana Chordata, Aves, Passenformes; Pat ned Sparwc, Gornió barba and Chordata, Aves, Passenformes; Pat ned Sparwc, Gornió barba and d by Museum of Vertebrate Zr -27 fi ID: legacy HeyMapper I ination Type: accepted place assigned by Kristi r Geography: North America, County I mee cific Locality: La Panza Range ting Source: wild caught Event Date: 1984-05-20 Verbatim Date: 20 ation Status: unverified	rome oxidase subunit 1* 2935.1.1* CRAISILIRAELGOPGALLGOD GAPMAFPRIANDSTILEPSTI AFFILITARILGOPGALLGOD GAPMAFPRIANDSTILEPSTI AFFILITARISSILGAINFITTA AGITHLITDRNLNTTFFDPAGGG EZOOLOGY thoutHeip La Panza Range, North America, Uni 20 May 1984 userellidae. Spizella atrogularis o boology, University of Califo of collection na Yamamoto on 2006-01-03 United States, California, 3 e, 1 mi NW Pozo summit May 1984	yven i vzala Poli Poly LLLASSTVEA overen LLLASSTVEA overen i indere Pala overen poly i mi NW Pozo sum i ted States, California cana coma coma ornia, Berkeley on	NEWS: Arctos Workin nit a, San Luis Obispo Cor Ned K. Johnso Identifiers GenBank: DQ4 MVZ frozen tiss collector number Parts Parts Parts Partsue (frozen) tissue (fr	ng Group M unty n 2 33192 2 sue collection: er: 5412 Condition [unchecked [unchecked [unchecked] [unchecked] [unchecked] brate Zoology, I brate Zoology, I brate Zoology, I brate Zoology, I brate Zoology, I	FC-21528 Disposition in collection fransfer of custo Jniversity of Califor University of Califor	tissue (fr Qty 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	Remarks
Search MVZ:Bird Spizella a Identification Spizella a Spizella a Spizella a Participation Black-char	/product="cytoch //product="cytoch //translation="W/ PELAONLAHAGASYDL VLITAVLLLSLEVIA Portals MUSEUM OF VERTEBRATE Portals My Stuft Portals My	rome oxidase subunit 1* 29354.1* GRASSLIFRAEUGOPARALLODO GRADMAFPRIANDSTILLEPST AFFEIALAGISSILGAINFITTA AGITHLITORNIANTIFFDPAGGG EZOOLOGY bboutHelp La Panza Range, North America, Uni 20 May 1984 userellidae, Spizella atrogularis o boology, University of Califo of collection na Yamamoto on 2006-01-03 United States, California, 1 e, 1 mi NW Pozo summit May 1984	1 ml NW Pozo summited States, California cana ornia, Berkeley on	MY S NEWS: Arctos Workin nit a, San Luis Obispo Cor Ned K. Johnso Identifiers GenBank: DQ4 MVZ frozen fiss collector numbr Parts Part Name Skeleton tissue (frozen) tissue (frozen) tissue of Verte age: ad sk. Museum of Vert reproductive da Museum of Vert	ng Group M unty n II IS3192 II Sue collection: er: 5412 Condition unchecked n) unchecked iunchecked brate Zoology, t trace fat ebrate Zoology, t trace fat ebrate Zoology, t testis 6x4	FC-21528 Disposition in collection in collection transfer of custo Jniversity of Califor University of Califor University of Califor University of Califor University of Califor	tissue (fr Qty 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	Remarks
Search  Search  MVZ:Bird Spizella a Animala, Black-formala, Control Berke  Location  Spizella Berke  Location  Determ  Higher  Spe Collec  Verific Coordii	/product="cytoch //product="cytoch //translation="Wy PTINOGROMUVELAI PPLAONLAHAGASYDL VLITAVLLLUSLEVIA Collections Database MUSEUM OF VERTERATE Portais My Stuff / :170231 atrogularis cana Chordata: Aves. Passenformes: Par mad Sparrow; Gorián barba negra dby Museum of Vertebrate Zi -27 df ID: legacy leyMapper I ination Type: accepted place assigned by Kristi r Geography. North America, County I mass cific Locality: La Panza Rang ting Source: wild caught Event Date: 1984-05-20 Verbatim Coordinates: 35 Datum: North America Data Verbatim Coordinates: 35 Datum: North America Data	rome oxidase subunit 1* 2935.1.1* OrnASSLIFRAEUGOPGALLGODG GAPMAFPBOINDSFILLEPST: AFFSLAFLAGTSSILGAINETTTA AGITMLLTDRNLMTTFFDPAGGG EZOOLOGY bootHeip EZOOLOGY bootHeip Ca Panza Range, North America, Uni 20 May 1984 eserellidae; Spizella atrogularis of coology, University of Califo of collection na Yamamoto on 2006-01-03 United States, California, 1 e, 1 mi NW Pozo summit May 1984 35726-120.30679 un 1927	ana bornia, Berkeley on	NEWS: Arctos Workin nit a, San Luis Obispo Cor Collector(s) Ned K. Johnson Identifiers GenBank: D04 MVZ frozen tiss collector numbr Parts Parts Part Name Skeleton tissue (frozen) tissue (frozen) tissue (frozen) tissue of Verte age: ad.sk. Museum of Verte age: ad.sk. Museum of Verte age: ad.sk.	ng Group M unty n E i33192 E sue collection: er: 5412 Condition unchecked n) unchecked n) unchecked brate Zoology, t ebrate Zoology, t ebrate Zoology, t ebrate Zoology, t	FC-21528 Disposition in collection in collection transfer of custo University of Califor University of Califor University of Califor University of Califor University of Califor	tissue (fr Qty) 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	Remarks
Search WVZ:Bird Spizella a Animala Biack-chin 1999-01 Nature c Location Berker Higher Spe Collect Verific Coordin	/product="cytoch //product="cytoch //protein_id="ABK //translation="W/ PELAGNLAHAGASYDL VLITAVLLLUSLEVIA PELAGNLAHAGASYDL VLITAVLLLUSLEVIA Portals My Stuff Portals Portal Portals Portal Portal Portal Portals Portal Portals Portal Portals Portal Portals Portal Portal P	rome oxidase subunit 1* 2935.1.1* GRAISLIPARINGPEANLIGDEG GRAINAFANDESTLIPARING FORMATTERPEANDESTLATERING FORMATTERPEANDESTLATERITTA AGITMLLTDRNLATTERPEANDE E ZOOLOGY boothHelp La Panza Range, North America, Uni 20 May 1984 cserellidae; Spizella atrogularis o boology, University of Califo of collection na Yamamoto on 2006-01-03 United States, California, 3 e, 1 mi NW Pozo summit May 1984 3572/-120.30679 mi 1927	1 mi NW Pozo summited States, California cana ornia, Berkeley on	NEWS: Arctos Workin nit a, San Luis Obispo Cou Collector(s) Ned K. Johnson Identifiers GenBank: DO4 MVZ frozen tiss collector number Parts Parts Parts Parts Parts Parts Parts Parts Parts Attributes sex: male Museum of Verte age: ad sk. Museum of Verte Museum of Verte age: ad sk. Museum of Verte age: ad sk. M	ng Group M unty n E 33192 E sue collection: er: 5412 Condition unchecked unchecked brate Zoology, I ebrate Zoology, I ebrate Zoology, trace fat ebrate Zoology, ta: testis 6x4. ebrate Zoology, ebrate Zoology,	FC-21528 Disposition in collection in collection fransfer of custo University of Califor University of Califor University of Califor University of Califor University of Califor	tissue (fr Qty 1 1 1 1 dy 1 nia, Berr mia, Berr oacal pr oacal pr mia, Berr oacal pr mia, Berr	Remarks
Search Search MVZ:Bird Spizella at Animala; Black-chi Identification Spizella at Animala; Animala	/product="cytoch /protein_id="ABK /translation="WV PTMOGROMUVPLAI PPLAONLARAGASYDL VLITAVLLLGLEVIA MUSEUM OF VERTEBRATE Portals My Stud / :170231 atrogularis cana is rogularis cana Chordata; Avec, Passenformer; Pat atrogularis cana Chordata; Avec, Passenformer; Pat atrogularis cana Chordata; Avec, Passenformer; Pat atrogularis cana is rogularis cana Chordata; Avec, Passenformer; Pat atrogularis cana Chordata; Avec, Passenformer; Pat atrogularis cana Chordata; Avec, Passenformer; Pat atrogularis cana Chordata; Avec, Passenformer; Pat atrogularis cana Chordata; Avec, Passenformer; Pat rogularis cana Chordata; Avec, Passenformer; Pat atrogularis cana Chordata; Avec, Passenformer; Pat atrogularis cana Chordata; Avec, Passenformer; Pat atrogularis cana Chordata; Avec, Passenformer; Pat d by Museum of Vertebrate Zi -27 d ID: legacy	rome oxidase subunit 1* 2935.1.1* OrnASSLIPRELOGPERALLOPDG CARLINGPOG CARLINGPOG CARLINGPOG CARLINGPOG E ZOOLOGY USOUTHEIP E ZOOLOGY E Z	ana ornia, Berkeley on	M Y S S NEWS: Arctos Workir nit a, San Luis Obispo Cor Collector(s) Ned K. Johnson Identifiers GenBank: DQ4 MVZ frozen tiss collector numbr Parts Parts Parts Parts Parts Parts Parts Parts Parts Attributes sex: male Museum of Vert age: ad sk. Museum of Vert weight: 10.90 g Museum of Vert	ng Group M unty n 2 33192 2 sue collection: er: 5412 Condition unchecked unchecked unchecked prate Zoology, l brate Zoology, trace fat ebrate Zoology, tat: testis 6x4.	FC-21528 Disposition in collection fransfer of custo University of Califor University of Califor University of Califor University of Califor University of Califor	Itssue (fr Qty) 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	Remarks
search Spizella a Spizella a Authilication Spizella at Authilication Spizella at Authilication Spizella at Authilication Black-chil Black-chil Black-chil Black-chil Spizella at Authilication Spizella at Authilication Spizella at Authilication Spizella at Authilication Black-chil Black-chil Black-chil Spizella at Authilication Spizella at Authilication Authilication Spizella at Authilication Authilication Spizella at Authilication Authilic	/product="cytoch //protein_id="ABK /translation="W/ PTMOGROMUVPLAI PPLAONLAHAGASYDL VLITAVLLLUSLEVIA WISEUM OF VERTEBRATE Portals My Stuff / :170231 attrogularis cana IS Chordata; Aves, Passenformes; Par med Sparwo; Comró barba par d by Museum of Vertebrate Zu -27 fi ID: legacy leyMapper I ination Type: accepted place assigned by Kristi Geography: North America, Countly IE more cific Locality: La Panza Range ting Source: wild caught Event Date: 1984-05-20 Verbatim Date: 20 ation Status: unverified nates: 35.3572 / -120.30679 Verbatim Coordinates: 35 Datum: North American Dat Error: 1486 mi Geography: North American Dat Error: 1486 mi Geography: Source: Terrai USGS 1:24,000 Geographere France Count Math	rome oxidase subunit 1* 2935-1.1* GRASSLIFRAELGOPGRALLGOO GRASSLIFRAELGOPGRALLGOO GRADNAFPSIGNASTILGOPGRALLGOO GRADNAFPSIGNASTILGOINETTTA AGITHLLTORNLATTFFDPAGGO E ZOOLOGY biboutHelp La Panza Range, North America, Uni 20 May 1984 userellidae. Spizella atrogularis o boology, University of California, see a spizella atrogularis o see a spizella atrogulari	a govern u vzaka krvat i previ kulza strveza ovorom v a taka krvat po koncerne v a taka krvat po konce	NEWS: Arctos Workin nit n, San Luis Obispo Con Collector(s) Ned K. Johnson Identifiers GenBank: DQ4 MVZ frozen tiss collector number Parts Part Name Skeleton tissue (frozen) tissue (frozen) tissue (frozen) tissue of Vert age: ad.sk. Museum of Vert fat deposition: Museum of Vert fat deposition: Museum of Vert fat deposition: Museum of Vert Accession	ng Group M unty n 2 33192 2 sue collection: er: 5412 Condition unchecked unchecked unchecked innorhecked brate Zoology, t brate Zoology, t ebrate Zoology, ta: testis 6x4 ebrate Zoology, ebrate Zoology,	FC-21528 Disposition in collection fransfer of custo Jniversity of Califor University of Califor University of Califor University of Califor University of Califor University of Califor	Qty 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	Remarks Reev, 2005-1 keley, 2006- keley, 2006- keley, 2006- keley, 2006-

Figure 1. Example of structured voucher annotation. (a) GenBank flat file record for Spizella atrogularis, accession DQ433192 and URL formula to map the record to MVZ specimen page. (b) Specimen page at Museum of Vertebrae Zoology, University of California, Berkeley for Spizella atrogularis linked to GenBank record DQ431992.

MVZ will bring up Museum of Vertebrate Zoology, University of California at Berkeley and its collections (Figure 2). Some useful search fields are listed in Table 1.

The BioCollections Database is reciprocally linked to other databases like Nucleotide, Protein, Popset, EST and GSS. This allows users to find all related records that are from an institution of interest.

Users can download the BioCollections dataset by using 'Send To' -> 'File' option, located at the upper right corner of the search results page. **Summary** will download text file with data based on number of records selected using checkbox from page. **CSV** will download comma separated values. **XML** will download XML file with data based on number of records selected using checkbox from page. In each case the user can select specific entries to include in the download by using checkboxes. The data can also be downloaded as a pipe-delimited text file from the NCBI ftp site (ftp://ftp.ncbi.nih.gov/pub/taxonomy/biocollections/).

# Duplicated or ambiguous collections and codes

For various reasons, some institutions use more than one institution code. For example, University of Maryland uses MARY for its herbarium collection and UMDC for its museum collection. These are listed as separate records. If an institution changes the code for its collection or institution and adopts a new one, the old code is retained in the database as a synonym. Similarly, when there are several institution codes for the same collection, they are listed as synonyms.

When more than one institution uses the same code for their specimen, the International Organization for Standardization three letter country code is used to unique the collections. If the institutions are from the same country, a state code is added in addition to country code. The institution code that is already in the database is retained (without the country code) and the subsequent ones are registered with country codes (state codes where applicable).

For example, all the following institutions use UAM as their institution code. To distinguish between the collections, the institution codes are listed as:

University of Alaska, Museum of the North UAM University of Arkansas at Monticello UAM<USA-AR> University of Alabama, Malacology Collection UAM<USA-AL> Universidad Autonoma De Madrid culture collection of cyanobacteria UAM<ESP> Universidad de los Andes, Facultad de Ciencias UAM<VEN>

Since University of Alaska, Museum of the North (UAM) was the first one to be registered in the BioCollections Database, UAM is retained for University

of Alaska and the subsequent UAM codes are added with country and state codes. When a record is submitted to Genbank with an ambiguous code (ex: UAM), it prompts a consult so a curator can confirm the correct institution is listed.

#### **Challenges of DwC Triplet**

DwC Triplet creates an identifier for voucher specimens in the form <institution\_code>:<OPTIONAL collection\_ code>:<specimen\_id>. The problems with DwC Triplets as identifiers have been discussed before (9). There are many institutions that share the same institution code. We resolve this ambiguity by adding three letter country codes to the duplicated institution codes. This works well for our internal system i.e. to link BioCollections with GenBank records but may not find exact matches across other repositories. Adding to the problem, DwC Triplets are not formatted consistently and different collections codes could be used for a single institution. For example, we use UWBM: ORN: for University of Washington, Burke Museum Ornithology Collection, whereas VertNet Database (http://vertnet.org/) uses UWBM: BIRD: for the same collection. Furthermore, submitters are asked to fill in the voucher information when submitting sequences to GenBank but many don't provide that information, thus, many voucher specimens are fielded as/strain or/isolate in GenBank records and cannot be linked to BioCollections. We have over 600 000 ATCC records that are formatted correctly as/culture\_collection and are linked to BioCollections but there are about 76 000 ATCC records that are not formatted correctly and appear as/strain or/ isolate in GenBank records. We are working on improving processes to correct the legacy records for which the culture collections acronyms are not 'structured.' Additionally, GenBank has recently started to automatically structure selected culture collections codes in new entries submitted as/strain or/isolate if they are from DSM, CBS, JCM, ATCC, LMG, NBRC, CCUG and KCTC. We selected these culture collection codes based on number of type strains we have in the taxonomy database (Table 2). Going forward, we will expand this list to other institution codes as well. Also, we would like to encourage submitters to provide specimen vouchers in a structured format so that they can be correctly linked to BioCollections by emailing updated information to gb-dmin@ncbi.nlm.nih.gov.

Often, institutions change their codes or are merged with other institutions. Linking mechanisms that depend on metadata like institution codes are prone to break as the metadata changes. The biodiversity community has long recognized the need for globally unique identifiers

Page

Hild

Notebook Page Collection

Milton Hildebrand collection



Figure 2. Screenshot of BioCollections search page. (a) Showing result for institution code MVZ. (b) Biocollections Database entry for MVZ.

#### Table 1. Selected search field and queries

Query	Find by
Search by code	Institution codes or collection codes or combination
Institution code= [icode]	Retrieves all the entries that have UAM as institution code including the ones that are unique
UAM[icode]	
Unique institution code=[uicode]	Retrieves only the exact match
UAM[uicode]	UAM = University of Alaska, Museum of the North
Collection code= [ccode]	Retrieves all the institution entries that list mamm as collections
Mamm[ccode]	
All of the above - [all]	Retrieves all the entries that list UAM as institution code/collection code and synonyms
UAM[all or just UAM]	
Search by name	Partial or full institution or collection name
Institution name = [iname]	Retrieves all the entries that have Alaska in the institution name
Alaska[iname]	
Collection name = [cname]	Retrieves all the entries that have mammal in collection name
Mammal[cname]	
All names=[all]	Retrieves all entries that have Alaska in institution/collection name
Alaska[all]	
Search by properties	Modifier type
Collection type museum[prop]	Retrieves all museum entries
Collection type herbarium[prop]	Retrieves all herbarium entries
Collection type culture collection[prop]	Retrieves all culture collection entries

Complex queries can be built by specifying the search terms, their fields and the Boolean operations AND, OR and NOT.

Culture collection	No. of type strains in Taxonomy database
Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSM)	9117
Centraal bureau voor Schimmelcultures, Fungal and Yeast Collection (CBS)	7667
Japan Collection of Microorganisms (JCM)	6980
American Type Culture Collection (ATCC)	6063
Belgian Coordinated Collections of Microorganisms/LMG Bacteria Collection (LMG)	3383
NITE Biological Resource Center (NBRC)	3352
Culture Collection, University of Goteborg, Department of Clinical Bacteriology (CCUG)	2855
Korean Collection for Type Cultures (KCTC)	2762

(GUID) to share, link and track biocollections data (specimen records, images, taxonomic names and DNA sequences) that are scattered all around the world. Several different technologies like Life Science Identifiers, Digital Object Identifiers, HyperText Transfer Protocol (HTTP) Uniform Resource Identifier-based identifiers etc. have been discussed for this purpose. More recently the use of GUID to provide stable identifiers for biocollections has gained traction (10–12). We will consider using these options as they become universally used in future.

### **External resources**

Resources outside of NCBI are constantly reviewed to keep the NCBI BioCollections Database up to date. In the past, we have exchanged data with the Global Registry of Biodiversity, an online metadata resource that provides information on biodiversity collections (13). Recently, we imported about 300 institution codes from Index Herbariorum (6) and about 50 culture collections codes from World Federation of Culture Collections. Integrated Digitized Biocollections (iDigBio) is another resource that provides data and images for millions of biological specimens in electronic format (14) and ways to link iDigBio specimen records to GenBank sequences associated with those specimens should be further explored. In 2011, the Global Genome Biodiversity Network (GGBN) was created as a part of Global Genome Initiative to bridge the gap between biodiversity repositories, sequence databases and research (15). Through its Data Portal, GGBN aims to make biodiversity samples readily discoverable and accessible to the research community. We will continue to explore the possibilities of crosslinking and updating data in accordance with all these external resources.

We are in the process of cleaning and updating the information in the BioCollections Database and have already updated >200 records by contacting resource managers and asking them to verify and correct their relevant information. We are also requesting institutions to provide us with an URL rule to their catalogue page so we can cross link the data. At present, NCBI offers the ability for credible third party resources to link out directly from either sequence records or via taxonomic names in the Taxonomy Browser. LinkOut aims to facilitate another way to access relevant online resources and supplement information found in NCBI databases (16). Links could be expanded at these individual pages by collaborating with more biorepositories.

# Addditional uses of BioCollections Database and future steps

Ideally, taxonomic vouchers should be expertly identified samples deposited and stored in a facility that is accessible to researchers for further study and thus serve an important role in biological research (17). For prokaryote names to be validly published, its type strains must be deposited in two recognized culture collections, a rule set by International Committee on Systematics of Prokaryotes under the International Code of Nomenclature of Prokaryotes (18). Type strains in culture collections are the points of reference that other strains must be compared with when determining their taxonomic identity. Similarly, the International code for Nomenclature for algae, fungi and plants (19) and The International Code of Zoological Nomenclature (http://iczn.org/code) also requires the designation of a type specimen, albeit with slightly different rules. The designation of vouchers is an important part of establishing provenance in systematic research and allows for critical assessment. With the increasing use of molecular sequences analysis in the systematics, it is important to establish a mechanism to connect these two sets of data. Besides taxonomic identification, associated metadata can provide important information on geographic dispersal and DNA can potentially be obtained for further research.

There are >1 600 000 species-level taxonomy ids in the NCBI Taxonomy database and they are identified with varying degree of certainty, with almost 400 000 identified with a binomial name. Type specimens have an important role in this regard, by providing a clear reference for comparison. We currently have just over 36 000 names with type material annotations. The complete list of type

material annotations will be released as part of the taxonomy ftp files. Since 2013, GenBank curates type material in the Taxonomy Database and uses it to flag sequences from types in the sequence records. This has led to an improvement in the annotation of sequence records. Recently, GenBank has developed a protocol to identify and correct misidentified prokaryotic genomes, using Average Nucleotide Identity genome neighboring statistics in conjugation with reference genomes from type (20). In addition to this, GenBank, together with its collaborative partners in the INSDC, has accepted the addition of a new 'type material' qualifier for sequence records which will enable specific sequence records to be annotated automatically with information from the NCBI Taxonomy database (1). BioCollections Database can be used as a useful resource to facilitate the identification of the home institutions providing these important set of records and track these specimens. Furthermore, BioCollections can add value to other NCBI databases. In 2011, NCBI developed BioProject and BioSample databases to organize and integrate data across interdisciplinary resources and allow users to query across many NCBI databases to retrieve data relevant to their interest (21). BioSample can potentially include blood samples, cell cultures, individual organisms etc. that may come from culture collections, museums, herbaria or other repositories. Expanded links between BioCollections and BioSample database will help make these databases more comprehensive.

The individual biorepository pages in BioCollections can serve as a start site for users specifically interested in the breakdown of sequenced vouchers at a specific institution. For example, Smithsonian Institution, National Museum of Natural History shares specimens and DNA samples with collaborators worldwide. As a result, DNA sequence data is submitted to Genbank, ENA and DDBJ by a large number of submitters and are often not formatted correctly and therefore are not linked to BioCollections Database. USNM (National Museum, >29 000 total records) and US (National Herbarium, >16 000 total records) notations represent a large number of sequence records and are part of an important collaborative effort. The 'USNM' and 'US' strings were used to search the entire GenBank database, then manually checked to assure they referred to specimens as expected. This information was reported to Smithsonian where they were added to the databases of appropriate departments. Depending on the choice of the individual institution this can facilitate the linking of specimens to their sequence records. One option will be to provide LinkOuts to specific samples pages directly from sequence records.

We hope to expand the utility of the BioCollections Database in a similar fashion for other biocollections in future. In the meanwhile, this focused resource will continue to provide important institutional context to the large number of sequence records in the public sequence databases.

### Acknowledgements

The authors would like to acknowledge Scott Federhen and dedicate this paper in his memory. Scott initiated this resource at NCBI to promote linkage between sequence data and specimens and championed its release.

# Funding

Funding for Open Access charge: Intramural Research Program of National Institutes of Health, National Library of Medicine.

Conflict of interest. None declared.

# References

- Mizrachi, I., Takagi, T. and Cochrane, G. (2018) The International Nucleotide Sequence Database Collaboration. *Nucleic Acids Res.*, 46, D48–D51.
- 2. Wieczorek, J., Bloom, D., Guralnick, R. *et al.* (2012) Darwin Core: an evolving community-developed biodiversity data standard. *PLoS One*, 7, e29715.
- Benson, D.A., Cavanaugh, M., Clark, K. et al. (2018) GenBank. Nucleic Acids Res., 46, D41–D47.
- 4. Toribio, A.L., Alako, B., Amid, C. *et al.* (2017) European Nucleotide Archive in 2016. *Nucleic Acids Res.*, 45, D32–D36.
- 5. Mashima, J., Kodama, Y., Fujisawa, T. *et al.* (2017) DNA data bank of Japan. *DNA Data Bank Jpn.*, 45, D25–D31.
- Thiers,B. Index Herbariorum: a global directory of public herbaria and associated staff. New York Botanical Garden, New York. http://sweetgum.nybg.org/ih/ (15 October 2017, date last accessed).
- Frost,D. (2017) Amphibian species of the World: an online reference, Version 6.0. American Museum of Natural History, New York. http://research.amnh.org/herpetology/amphibia/index.html (15 December 2017, date last accessed).
- Eschmeyer, W.N., Fricke, R. and van der Laan, R. Catalog of fishes. http://researcharchive.calacademy.org/research/ichthyology/cata log/fishcatmain.asp (15 October 2017, date last accessed).

- 9. Guralnick, R., Conlin, T., Deck, J. *et al.* (2014) The trouble with triplets in biodiversity informatics: a data-driven case against current identifier practices. *PLoS One*, 9, e114069.
- Güntsch,A., Hyam,R., Hagedorn,G. *et al.* (2017) Actionable, long-term stable and semantic web compatible identifiers for access to biological collection objects. *Database*, 2017, 1–9.
- 11. Guralnick, R.P., Cellinese, N., Deck, J. *et al.* (2015) Community next steps for making globally unique identifiers work for biocollections data. *Zookeys.*, **494**, 133–154.
- Krishtalka,L., Dalcin,E., Ellis,S. et al. (2016) Accelerating the Discovery of Biocollections Data. GBIF Secretariat, Copenhagen. http://www.gbif.org/resource/83022
- Schindel, D.E., Miller, S.E., Trizna, M.G. *et al.* (2016) The global registry of biodiversity repositories: a call for community curation. *Global Registry Biodiv. Repositories: Biodivers Data J.*, 4, e10293.
- 14. Godden,G.T. and Soltis,P.S. (2014) A new iDigBio web feature links DNA banks and genetic resources repositories in the United States. In: Applequist WA Campbell LM, (eds). DNA Banking for 21st Century. St. Louis, MO: Missouri Botanical Garden, 173–181.
- Droege,G., Barker,K., Seberg,O. *et al.* (2016) The global genome biodiversity network (GGBN) data standard specification. *Database*, 2016, 1–11.
- Kwan,Y.K. (2013) LinkOut: linking to External Resources from NCBI Databases. In: *The NCBI Handbook [Internet]*, 2nd edn. National Center for Biotechnology Information (US), Bethesda (MD).
- 17. Culley, T.M. (2013) Why vouchers matter in botanical research. *Appl. Plant Sci.*, **1**, apps.1300076.
- Parker, C.T., Tindall, B.J. and Garrity, G.M. (2015) International Code of Nomenclature of Prokaryotes. *Int. J. Syst. Evol. Microbiol.*, http://ijs.microbiologyresearch.org/content/journal/ ijsem/10.1099/ijsem.0.000778#tab1.
- McNeill, J. et al. (2012) International Code of Nomenclature for algae, fungi, and plants (Melbourne Code), Vol. 154. In: McNeill (ed). Regnum Vegetabile. Koeltz Scientific Books, Königstein, p. 240.
- Federhen,S. (2015) Type material in the NCBI taxonomy database. Nucleic Acids Res., 43, D1086–D1098.
- Barrett, T., Clark, K., Gevorgyan, R. et al. (2012) BioProject and BioSample databases at NCBI: facilitating capture and organization of metadata. Nucleic Acids Res., 40, D57–D63.