

ArctosR: AN R PACKAGE TO INTERFACE WITH THE ARCTOS MUSEUM DATABASE API

HARLAN R. WILLIAMS^{1*}, JOCELYN P. COLELLA², VIJAY BARVE³, MICHELLE S. KOO⁴, MARLON E. COBOS²

1 Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS 66045, USA

2 Biodiversity Institute & Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, KS 66045, USA

3 Natural History Museum of Los Angeles County, Los Angeles, CA 90007 USA

4 Museum of Vertebrate Zoology, University of California, Berkeley, CA 94720, USA

Abstract: Arctos is an online museum collection management database that hosts data for >5 million observational and physical records of plants, animals, and artifacts from >60 institutions worldwide. For biological specimens, these data include geographic and temporal information, detailed parts such as tissue samples, and links to related specimen records (e.g., parasites) as defined by the extended specimen concept. Although some of the data in Arctos are accessible via aggregators such as the Global Biodiversity Information Facility, information of high relevance for other research questions, such as host-parasite associations, can only be accessed using the Arctos system. These data, which are missing from aggregators, have the potential to provide rich context for large-scale biodiversity, ecological, and evolutionary research, but there is no automated way to download bulk data from Arctos. To expand access to specimen data, we developed ArctosR, an R package that enables users to query the Arctos database programmatically, download bulk data, and use those data in other R-scripted workflows. Here, we describe the package functionality and outline five use cases for ArctosR: (1) a search for georeferenced occurrences, (2) finding records of a host species that had nematodes, (3) finding tissue samples preserved in ways that are suitable for molecular exploration, (4) downloading data to explore patterns of Bergmann's Rule, and (5) downloading data and using it in ecological niche modeling. Combined with the expanded and standardized specimen, parasite, and pathogen fields available in Arctos, ArctosR serves as a data bridge connecting biodiversity information to related biomedical investigations, enabling a more comprehensive OneHealth approach.

Keywords: biorepository, parasites, pathogens, specimen records, tissue samples, FAIR data principles

INTRODUCTION

Large-scale biodiversity research requires assembling information from across multiple sources to provide more complete, longitudinal, empirical data about a phenomenon of interest. This task becomes more complex when cross-institutional data come in different standards and formats, such as from multiple databases or data aggregators. While massive efforts in the last decades have liberated a wealth of data contained in the global network of natural history museums (Bakker et al. 2020), limited in-

teroperability among collection databases is still a barrier to effectively utilizing those data. This problem is somewhat mitigated by secondary-source aggregators such as the Global Biodiversity Information Facility (GBIF; Telenius 2011), the Global Genome Biodiversity Network (GGBN; Droege et al. 2016), and the Integrated Digitized Biocollection (iDigBio; Nelson 2014), which collect and standardize a subset of museum data from multiple primary sources. For example, GBIF extracts occurrence data and related fields, GGBN extracts tissue and molecular

* Corresponding author:
Email: harlanrhwilliams@gmail.com

Table 1. Description of functions available in the package ArctosR.

Function	Group	Description
<code>get_query_parameters</code>	Query	Fetches a data frame of all possible search parameters that can be used in a query.
<code>get_result_parameters</code>	Query	Fetches a data frame of all possible columns (fields) that can be requested from a query.
<code>get_record_count</code>	Query	Returns the total number of records that match a given query.
<code>get_records</code>	Query	Primary function to perform a query and retrieve records from <i>Arctos</i> .
<code>check_for_status</code>	Error checking	Checks if a query object contains a failed response.
<code>get_request_url</code>	Error checking	Extracts the URL used for a request within a query object for debugging.
<code>expand_column</code>	Data handling	Processes a column containing nested JSON data into nested data frames, facilitating access and indexing.
<code>response_data</code>	Data handling	Extracts the main data frame of records from a query object returned by <code>get_records()</code>
<code>save_response_rds</code>	Data handling	Saves the entire response object as an R-native RDS file.
<code>read_response_rds</code>	Data handling	Reads in a response saved as an RDS file.
<code>save_response_csv</code>	Data handling	Saves the retrieved records to one or more CSV files, with options for handling data expanded with <code>expand_column()</code> .

information, and iDigBio aggregates ecology and biodiversity-related metadata. The main drawback of these services, however, is that they only provide subsets of the total data available.

Arctos is a primary-source collection management software (CMS) that enables museum biorepositories to digitize and share specimen data using a standardized vocabulary that ensures consistency across institutions (Cicero et al. 2024). Unlike secondary-source aggregators, Arctos maintains the complete set of institution-supplied data, and it links related records within a Digital Extended Specimen framework (Galbreath et al. 2019; Hardisty et al. 2022). With data from 285 collections across over 60 institutions (by December 2025), Arctos functions as a unified platform for accessing rich, relational specimen information that supports interdisciplinary research (e.g., Cook et al. 2017; Cobos et al. 2025). For instance, it allows users to explore relationships such as specimen–part, host–parasite, or host–pathogen connections not common in general aggregators or other CMSs. Accessing Arctos directly, therefore, provides a richer picture of specimen data and lets users select specific fields that are most relevant to their work.

Arctos provides a web interface for users to access data via interactive queries and an API for programmatic data access (Cicero et al. 2024). The web interface is a user-friendly option for small queries, but pulling the amount of data required for large-scale research is cumbersome. Conversely, use of the API for programmatic data access presents significant technical challenges for many researchers. This is in part due to the complexities of handling paginated results and parsing hierarchical data structures, which highlights the need for user-friendly programmatic tools that streamline access to this primary-source information. An alternative way to gain access

to bulk data is via popular programming languages such as R (R Core Team 2025) or Python (Python Software Foundation 2025). For instance, R packages like *rgbif* (Chamberlain et al. 2018) and *spocc* (Chamberlain 2021) have substantially benefited researchers accessing GBIF data. A similar tool has been recently made available for iDigBio—*ridigbio* (Michonneau et al. 2024)—and libraries for Python that facilitate access to GBIF (e.g., Chamberlain and Boettiger 2017) are popular among researchers.

Following examples adopted for other databases and to reduce the challenges to accessing and utilizing information from Arctos, we introduce ArctosR, an R package that simplifies the bulk retrieval of data from Arctos. Our package enables users to request and download more data than is currently allowed by the web interface, making it practical for big-data projects. ArctosR also makes data requests more reproducible and sharable, providing metadata to replicate previous queries. As our package links directly to Arctos, it also facilitates appropriate acknowledgement of institutions hosting downloaded records. Altogether, the ArctosR package aims to facilitate access to specimen-backed information that can advance research in ecology, evolution, and public health, among other fields.

PACKAGE DESCRIPTION

ArctosR was developed to simplify the process of building repeatable queries and downloading data from Arctos via its API. The package offers tools to request data from Arctos, transform data into a format usable in R, and save data and associated metadata (e.g., timestamp of request, request URLs used). ArctosR was designed with minimal dependencies and relies on three packages for its core functionality: (1) *curl* (Ooms et al. 2025) for making web requests to the Arctos API over HTTPS (Hypertext Transfer Protocol Secure); (2) *jsonlite* (Ooms

2014) for parsing JavaScript Object Notation (JSON)-formatted responses from Arctos into R data structures; and (3) R6 (Chang et al. 2025) to enable development of the package with an object-oriented programming style. The streamlined set of functions in the package represents the most common types of requests that users send to Arctos and operations on responses that users may wish to perform (Table 1). These functions can be grouped as (1) query functions, (2) error checking functions, and (3) data handling functions. ArctosR is available from CRAN (<https://CRAN.R-project.org/package=ArctosR>), and its source code can be found on GitHub at <https://github.com/hrhwilliams/ArctosR>.

To use ArctosR securely, users must provide an API key issued by Arctos in each of their requests, which can be obtained by following the instructions at <https://arctosdb.org/arctos-api-policy/>. To abstract individual HTTP requests and responses and to provide a simple programmatic interface to Arctos, the objects in ArctosR are organized into Queries, Requests, and Responses. A Query represents a generic search in Arctos that should return data, such as a search to find specimens with a desired property (e.g., has tissue) or a relational search using one set of specimens to query another set (e.g., searching hosts to identify parasites or vice versa). Each Query is performed using one or more Requests, each of which represents a single HTTP request to the Arctos web API. A Query is executed, and its Request(s) are sent to Arctos. Arctos responds to each HTTP request with data relating to the search, and the ArctosR package combines these into a single Response object representing all of the data pertaining to the Query.

The data returned by Arctos are initially in JSON format, which contains both flat data (e.g., scientific name, collection date) and nested data (e.g., tables of tissues and parasites associated with a specimen). The package turns the JSON-encoded data into R data frames. Nested data can be difficult to work with manually, but ArctosR handles nested data by turning it into data frames within data frames (see *Handling nested data and saving results*). For scientific reproducibility, each query made with ArctosR saves metadata associated with that search, including query parameters, returned columns, the exact Uniform Resource Locator (URL) used for the request, and the date and time the request was made.

Queries and results

An Arctos query consists of a list of query parameters and user-provided values, and a list of desired return columns. A query parameter is a standardized name for an attribute of a record, such as `scientific_name`, and an

example user-provided value for this parameter would be the Latin binomial of a species (e.g., *Tamias alpinus*). Performing a query with that parameter and value would request all records from Arctos where the scientific name matches *Tamias alpinus*. Arctos returns its data in a table, and the return columns are the possible fields in that table, such as latitude (`dec_lat`) or longitude (`dec_long`). Arctos has over 100 different parameters to use in queries (Table S1), and over 100 possible columns that can be returned for each record (Table S2).

The web interface for Arctos guides users in building a query by providing an input form, with options to customize both the form and which columns are returned in the user's profile (Fig. 1). The equivalent interface in ArctosR is the `get_records()` function, which users can use to query any data available from Arctos's catalog search page. To use `get_records()`, users must find relevant query parameter names and result column names, which they can do by calling the `get_query_parameters()` and `get_result_parameters()` functions respectively. The `get_records()` function returns a Response object containing the data retrieved and metadata about the request. The data from the Response can be extracted by passing the Response object to the `response_data()` function, which returns the query data as a data frame. Every parameter from the `get_query_parameters()` function can be used as input to the `get_records()` function to create a query, and every column name from the `get_result_parameters()` can be passed in as a list to specify the columns returned by the query (Fig. 1).

To facilitate interoperability, many query parameters and return columns use a standardized vocabulary; that is, a restricted set of possible values that the parameter or column can have. Documentation for these can be found in the Arctos code tables, which can be accessed through the web portal (upon log-in to the Arctos web portal, go to the Search menu and select Code Tables) or via the API through the `get_query_parameters()` or the `get_result_parameters()` function. For instance, the part preservation code table (`ctpart_preservation`) defines controlled vocabulary for specimen parts (e.g., tissue, parasite), indicating whether a part is preserved in ethanol, formalin, RNAlater, or >30 other possible preservation modes. ArctosR also allows users to directly look up these parameters in R (see Example cases—Exploring query parameters).

Handling nested data and saving results

One challenge to the use of raw Arctos data has been the database's relational complexity. Many data portals

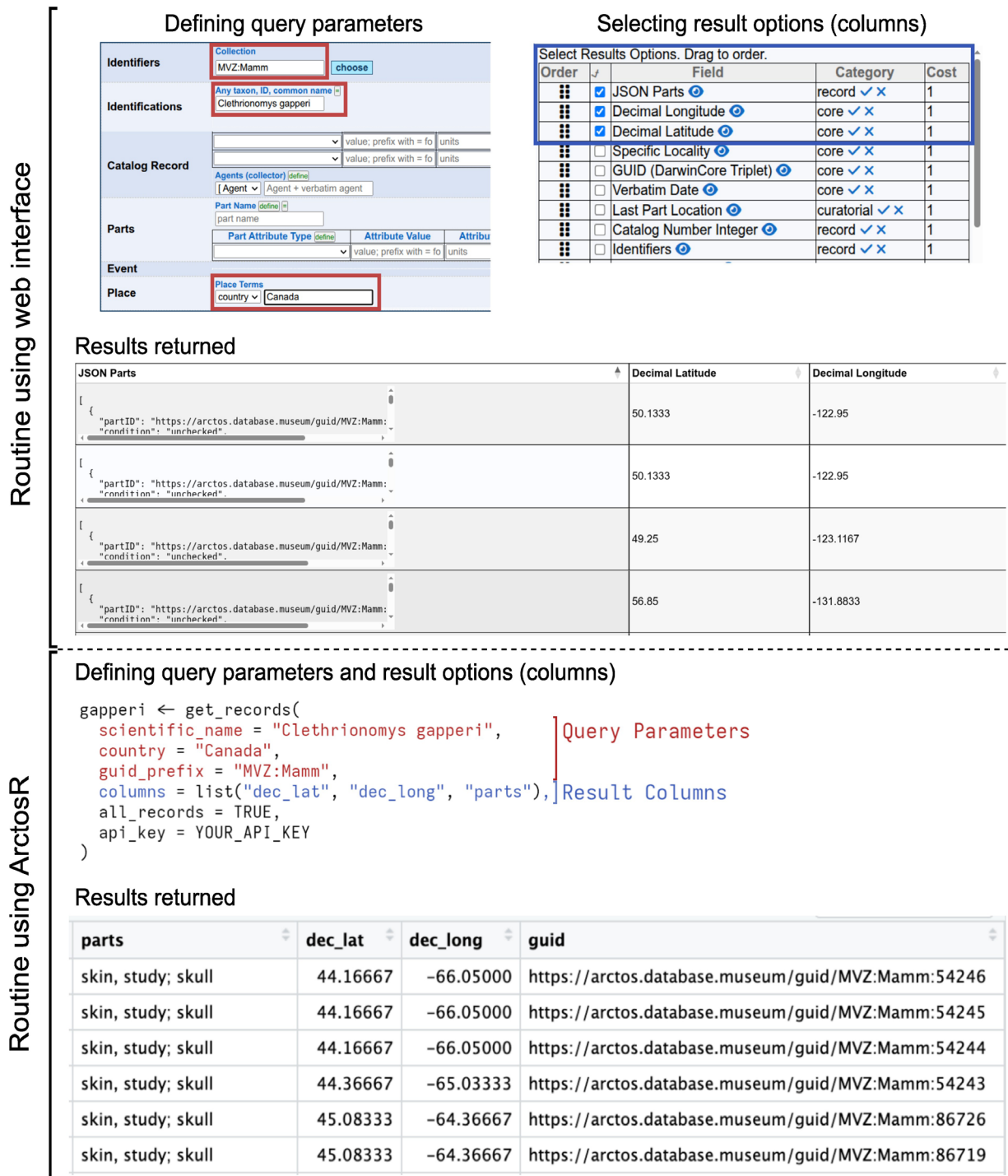


Figure 1. A query through the web interface and ArctosR for *Clethrionomys gapperi* collected in Canada and held in the Museum of Vertebrate Zoology mammal collection (MVZ:Mamm) with the result columns JSON parts, decimal latitude, and decimal longitude. Equivalent query parameters and result columns are highlighted using matching colors. The globally unique identifier (guid) field is automatically attached to records obtained from a query.

present data in the form of a table, but in Arctos, records can have multiple additional parts related to them, and some of those parts (e.g., parasites, embryos) may also represent separate specimen records. The information for parts is presented as nested tables in Arctos, which can be unintuitive to work with and not directly navigable in R. ArctosR handles these relational data by transforming nested tables into nested R data frames with the `expand_column()` function. Additional records or parts related to a given record can be queried using the `get_relationships()` function, making it easier to explore relational data.

Finally, ArctosR makes it easy for users to download and save data to be used later. The `save_response_csv()` function saves data from a response as a comma-separated values (CSV) plain text file, with the option of saving multiple CSVs for nested data. These CSV files can be read later with the base R function `read.csv()`. The `save_response_rds()` and `load_response_rds()` functions save and load a response as R-native data; when loaded, the response can be fully reconstructed.

EXAMPLE CASES

We provide five step-by-step examples of how researchers might use ArctosR to acquire data. All of the code to accomplish these examples can be found in the package's vignettes available on the package website: <https://hrhwilliams.github.io/ArctosR/> (See menu Articles). All the examples in the vignettes are performed as independent routines in R to illustrate how to use functions to explore parameter options and download data. The examples finish by displaying geospatial data using the `ggplot2` (Wickham 2011) package to create plots. Each vignette begins by querying Arctos for possible query parameters and result columns to build a data query.

Exploring query parameters and fields to be returned

Users are required to use their own API key to perform a query using ArctosR. Once the key is obtained, it can be stored in a variable and then used in later steps. For example, if the hypothetical key obtained is 5e5516a3-e45f-403e-bc00, the line of code to store the key in an object will look like this: `mykey <- "5e5516a3-e45f-403e-bc00"`. All future queries will need `mykey` to be set as the value provided to the argument `api_key`.

All queries via ArctosR are based on Arctos' standard vocabulary. The function `get_query_parameters()` helps to explore all arguments that can be defined later in the functions `get_record_count()` and `get_records()` to count and download records respectively.

The output of using `get_query_parameters()` is a data frame with five columns; the column `obj_name` contains the parameters that can be used to perform a query (Table S1). For instance, to query specimen records of the family Ixodidae and count the total number of records, the line of code will be `get_record_count(family = "Ixodidae", api_key = mykey)`, where `family` is one of the parameters returned by `get_query_parameters()`. Additional parameters can be provided to constrain the data to be retrieved. For example, to geographically limit the previous query to Canada, the code will be `get_record_count(family = "Ixodidae", country = "Canada", api_key = mykey)`. Multiple arguments can be used to make queries more specific; for example, locality information can be specified at various scales (e.g., county, province, state, country, etc.). Searches can also be restricted to a single collection (`guid_prefix`).

Records in Arctos can be associated with >100 fields (columns), which are divided into eight categories: attribute, classification, core, curatorial, identifiers, locality, record, trait. By default, if there are no specifically requested return columns, Arctos, and thus ArctosR, returns the columns belonging to the core category: GUID, decimal latitude, decimal longitude, coordinate error, verbatim date, specific locality, asserted state or province, asserted country, and identified as. If users require specific fields, the function `get_result_parameters()` can be used to consult all available column name options that can be defined. The names of the columns to be returned can be passed to the function `get_records()` as a `list()` in R, using the argument `columns`. A demonstration following the previous sequence of examples will be: `get_record_count(family = "Ixodidae", country = "Canada", columns = list("dec_long", "dec_lat"), api_key = mykey)`.

Example 1: Species occurrence query and plotting

In this example, we would like to query for *Tamias alpinus* at the Museum of Vertebrate Zoology that were collected in the state of California, and plot the coordinate data of those records on a map. First, we call `get_query_parameters()` to find the names of the exact query parameters to pass to ArctosR for our search. This function provides both the variable name that the API is expecting us to use, and documentation about that name. In this case, we will need to use: (1) `scientific_name`, (2) `guid_prefix`, and (3) `state_prov`. The `scientific_name` parameter requires that we provide a Latin binomial of our taxon of interest, the `guid_prefix` parameter should be the abbreviation of the collection, and `state_prov`

should just be the name of our state of interest. Next, we want to return the latitude and longitude of those records, and to find the names of those result parameters, we call `get_result_parameters()` and find the parameters `dec_lat` and `dec_long`.

Now that we know what variables to use in our search, we can call the `get_records()` function. We pass in all query parameters to the function directly, for example, as: `get_records(scientific_name = "Tamias alpinus")`, and any result parameters we want the table returned from Arctos to contain, we pass to the `columns` argument in a list, like: `get_records(columns = list("dec_lat", "dec_long"))`. (See code to run full example at <https://hrhwilliams.github.io/ArctosR/articles/species.html>). From this exercise, we obtained 424 georeferenced records of Alpine Chipmunks. All records were sampled mainly around the Sierra Nevada in the state of California (Fig. 2).

Example 2: Finding hosts infected with a parasite

ArctosR can query records by relation, for example, records of parasites recovered from hosts. In this example, we query for all specimen records of the genus *Microtus*, collected in Mongolia; filter by whether they were screened for the presence of parasites; and map where parasite-screened *Microtus* were sampled and whether or not nematodes (endoparasitic roundworms) were found in them. The subset of *Microtus* records from which nematodes were collected will either have links to the respective nematode records or will have nematodes listed as parts associated with each host record.

To perform this example, we use the `get_records()` function with the `genus` parameter set to the value "Microtus", the `country` parameter set to "Mongolia", and `filter_by` set to `list(examined_for = "parasite")`. Once the data were downloaded, we filtered records in base R to determine which of the Mongolian vole specimens examined for parasites had nematodes recovered and included as a "part" of the host. Finally, we visualized the results on a map of Mongolia with returned records of *Microtus*, colored by detection of nematodes (Fig. 3). See all code to run this example at <https://hrhwilliams.github.io/ArctosR/articles/parasite.html>.

A total of 886 *Microtus* records were downloaded in this example as of November 2025, with most found in northern Mongolia (Fig. 3). Records of *Microtus* for which nematodes were detected ($N = 111$) do not show a geographic pattern; however, nematodes were notably not found in areas closer to the center of the country. Par-

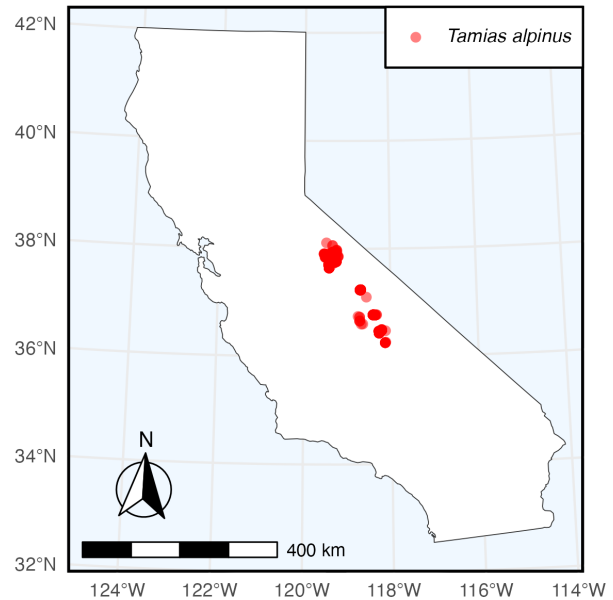


Figure 2. Geographic representation of records downloaded from Arctos using ArctosR for the species *Tamias alpinus* in the state of California.

asite preservation media details are returned together with specimen part information and can be used to determine which parasite specimens may be most appropriate for downstream morphological or molecular analyses.

Example 3: Finding samples suitable for molecular research across taxonomic synonyms

Identifying what parts are available for each specimen can inform sampling and experimental design for future research. For example, there is a hybridization zone between northern (*Clethrionomys rutilus*) and southern (*Clethrionomys gapperi*) red-backed voles in Canada (Runck et al. 2009; Wiens and Colella 2025). Investigations of hybrid dynamics in this system require molecular explorations of specimens collected within the hybridization zone. Complicating the query, the genus *Clethrionomys* used to be named *Myodes* (Kryštufek et al. 2020) and that name change may not yet be universally reflected by all collections hosting specimen data in Arctos. ArctosR can be used to find molecularly-suitable vole specimens by: querying for specimens of both species; filtering records by geographic proximity to the hybrid zone; and identifying the subset of specimens for which molecular quality tissue exists (e.g., tissue preserved in -80°C , LN_2 , DNA/RNA Shield, etc.). Tissues of interest can then be loaned from respective collections and used in downstream molecular applications (e.g., whole-genome sequencing, RNAseq, etc.). In ArctosR, we performed independent queries using

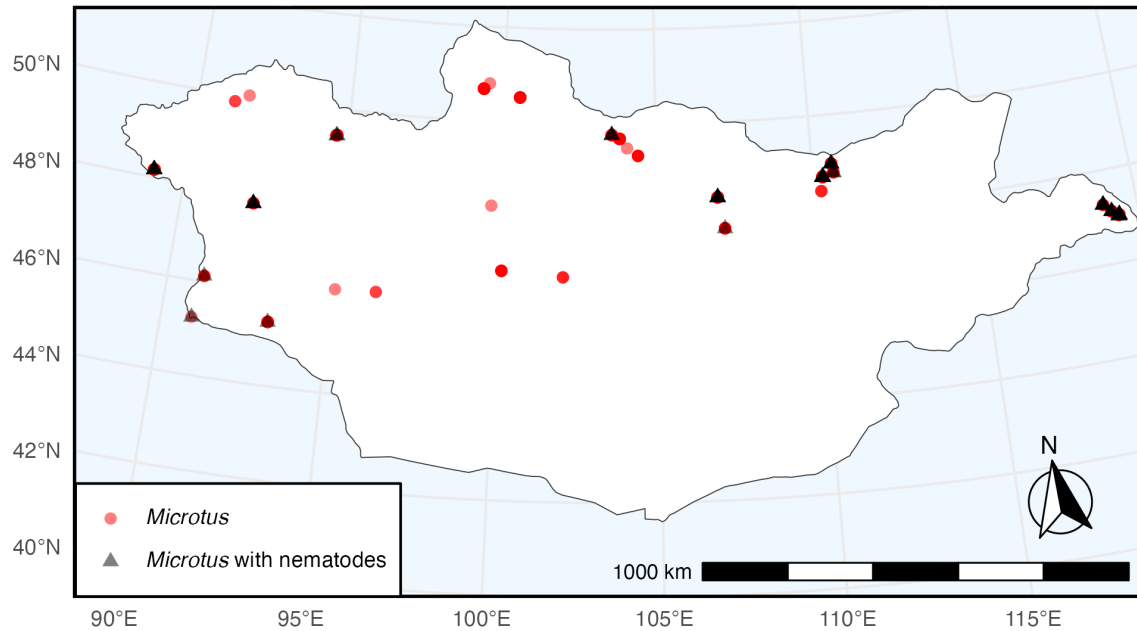


Figure 3. Geographic representation of records downloaded from Arctos using ArctosR for the genus *Microtus* in Mongolia, with filtering steps to identify specimens for which nematodes were detected.

the `get_records()` function for each species. Because of the recent taxonomic change, we performed four queries, one for each Latin binomial: *Clethrionomys gapperi*, *Clethrionomys rutilus*, *Myodes gapperi*, and *Myodes rutilus*. Currently, Arctos tracks taxonomic synonyms, but those data are not exposed to their API, thus necessitating queries under both *Clethrionomys* and *Myodes*. We set the `columns` argument to the value `list("dec_lat", "dec_long", "parts")`. Code to run this example is available at <https://hrhwilliams.github.io/ArctosR/articles/tissue.html>.

Our download contained 15,005 records for *C. gapperi* and 11,384 for *C. rutilus* (Fig. 4). Combined, a total of 10,899 of the georeferenced records had tissue suitable for molecular explorations of hybridization (i.e., frozen tissue). The information in the column `"rights"` (included by default in query results) can be used to identify which institution holds the tissue of interest and the catalog numbers for requesting loans.

Example 4: Exploring patterns of Bergmann's Rule

Arctos stores standard morphological measurements of specimens, among other metadata. These data can be used in quantitative analyses of species traits to explore patterns across geography, demographics, and time. In this example, we explore the idea that populations of the same species of vertebrate (and, generally, endotherms) in colder climates tend to have larger body sizes than those in warmer regions (Bergmann's Rule, Bergmann 1847). Considering latitude as a proxy for temperature (higher

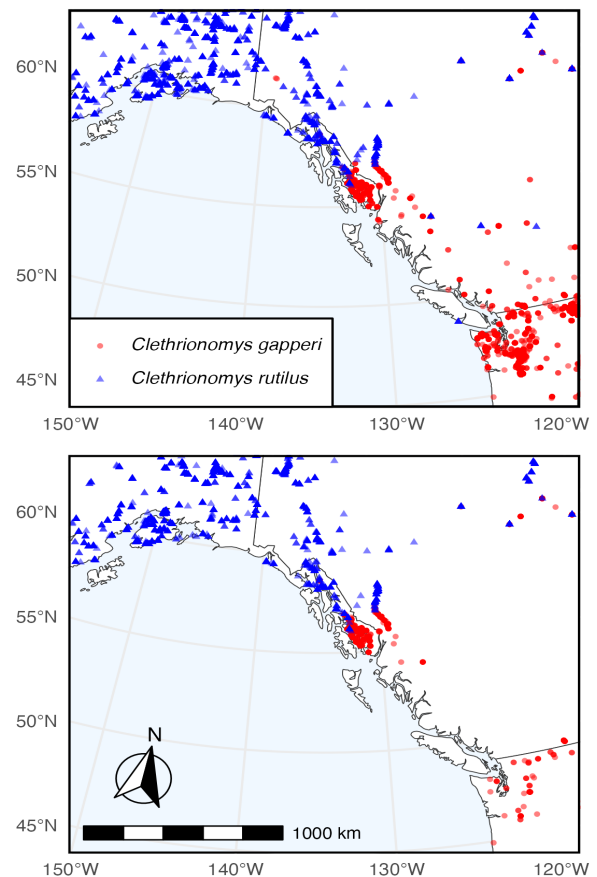


Figure 4. Geographic representation of records for the species *Clethrionomys gapperi* and *C. rutilus*, downloaded from Arctos using ArctosR. The records were filtered to show available tissue samples for the two species in a region of interest for the study of hybridization. The top panel shows all records in Arctos; the bottom panel shows only records with tissues suitable for the study.

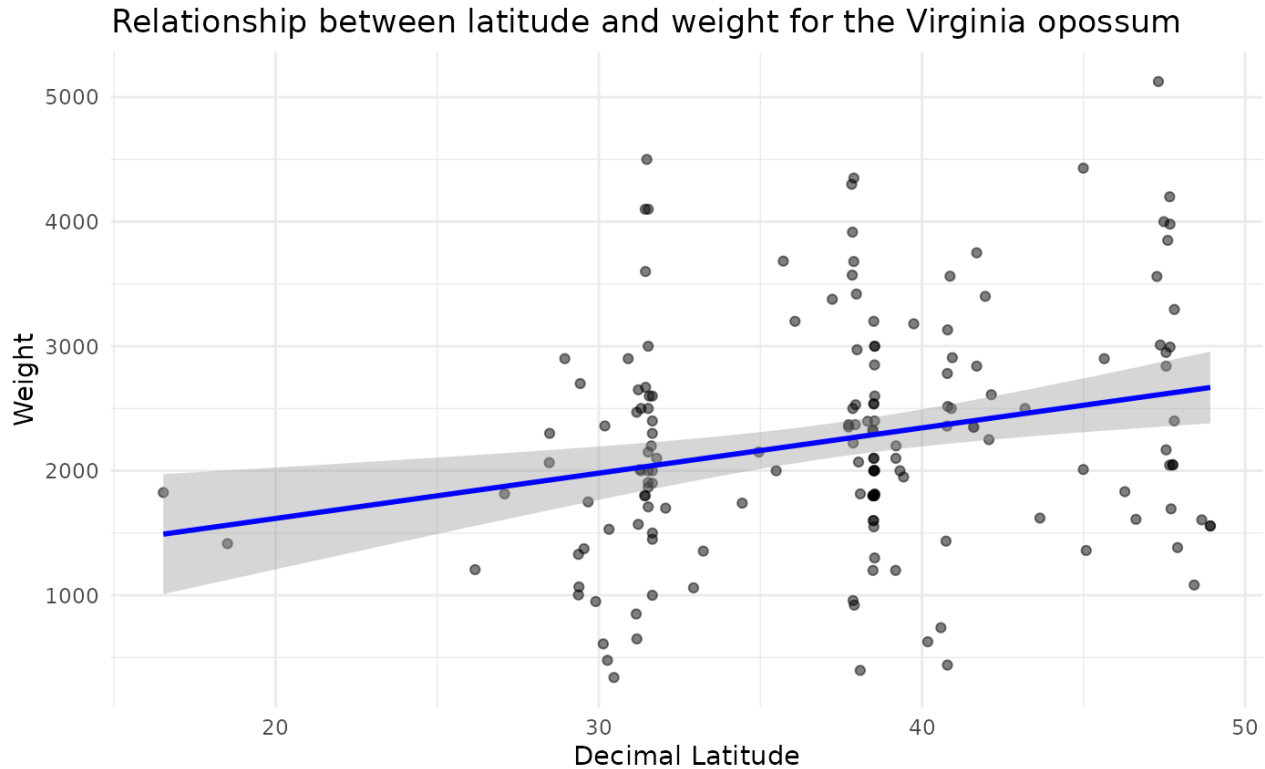


Figure 5. Exploration of Bergmann's Rule for *Didelphis virginiana* specimens.

latitudes equate to colder temperatures), endotherms are generally physically larger toward the poles. Here, we use specimen weight as a measure of size for Virginia Opossums (*Didelphis virginiana*), which have a broad distribution in North America. If this example follows Bergmann's Rule, we expect to see a positive relationship between latitude and weight.

We obtained data for *D. virginiana* using the `get_records()` function, setting the `columns` argument to the values `list("guid", "dec_lat", "dec_long", "weight", "age")`. We filtered data by age to keep records of adults and those with undetermined age. Then, we removed records with no weight information and worked with the values of weight to harmonize units. We explored and visualized the relationship of latitude and weight for adult individuals or those with weights > 300 g, per *Mammals of North America* (Kays and Wilson 2009). Code to run this example is available at https://hrhwilliams.github.io/ArctosR/articles/bergmann_rule.html.

We downloaded a total of 1868 *D. virginiana* records, and retained 164 records after data filtering steps. Our simple exploration of Bergmann's Rule pattern showed the expected positive relationship between latitude and weight (Fig. 5).

Example 5: Using Arctos occurrence data for ecological niche modeling

One common usage case of specimen occurrence data is for the study of species distributions via ecological niche modeling (ENM; Peterson et al. 2011). In these types of studies, occurrence data and layers of environmental variables are used to characterize conditions in which species can maintain populations. Models are then projected to geography to visualize areas with environments that are suitable for the species of interest. ArctosR can be used to download occurrence records and other packages to obtain environmental data, process information, and run ENM.

In this example, we downloaded occurrences for the Lone star tick *Amblyomma americanum* using the `get_records()` function, setting the `columns` argument to the value: `list("guid", "dec_lat", "dec_long")`. To download environmental data from the WorldClim database (Fick and Hijmans 2017), we used the package `geodata` (Hijmans et al. 2023). We used base R functions and the packages `terra` (Hijmans 2023) and `kuenm2` (Trindade et al. 2026) to filter, process, and clean our data. Data preparation for ENM, model selection routines, and predictions to geography were also produced using the `kuenm2` package. Code to run this example is

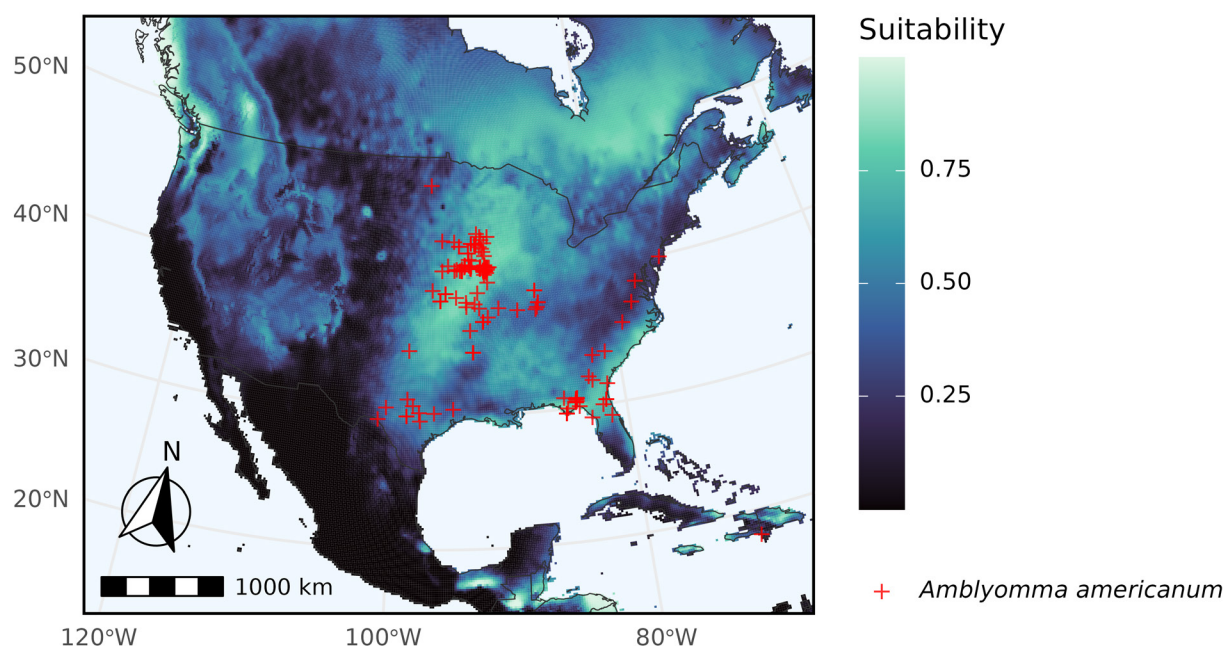


Figure 6. Geographic projection of downloaded records and suitable areas for the Lone Star tick *Amblyomma americanum*. Records shown are those resulting after data cleaning steps. Higher suitability is shown in lighter colors.

available at <https://hrhwilliams.github.io/ArctosR/articles/enm.html>.

We downloaded 220 records for *A. americanum*, and retained 95 after data cleaning (Fig. 6). Our simple ENM exercise predicted higher suitability close to our records with a clear low suitability band west of the Great Plains (Fig. 6). This example was intended only as a demonstration of one way to initiate data obtained with ArctosR in analyses with other R packages. Detailed ENM exercises include other steps not included here, such as spatial data thinning and analyses of extrapolation risks (Araújo et al. 2019).

DISCUSSION

The ArctosR package greatly improves accessibility and reproducibility to primary natural history data by providing a streamlined and simple interface for querying Arctos data. This provides access to >100 fields associated with each specimen record, many of which are not accessible through data aggregators like GBIF or GGBN. To our knowledge, Arctos is the only CMS that has integrated relational or nested queries; for example, users can search the database by host to find individuals that had parasites or search for a parasite species to find all hosts from which that parasite has been recovered. Arctos is also actively developing pathogen-related data standards, which are ab-

sent from most biodiversity data aggregators. Those fields include expanded microbial taxonomy for viruses, bacteria, and fungi; screening methods (e.g., serology, metagenomics, etc.); tissue quality (e.g., poor, excellent); and test-associated quantitative metrics (e.g., cycle threshold/ct values, etc.). The existing fields in Arctos and these new features significantly expand connections between museum biorepositories and the biomedical and public health communities (Colella et al. 2021; de Vries et al. 2025).

The FAIR (Findable, Accessible, Interoperable, and Reusable) guiding principles for data publishers aim to ensure that data will be reusable for future work and analyses made on those data will be reproducible (Wilkinson et al. 2016). In the age of “big data”, pooling data from disparate sources remains complicated by variation in data standards and vocabularies, cross-platform incompatibilities, and access limitations (Upham et al. 2021). Pathogen data derived from specimens and published openly on PHAROS (the Pathogen Harmonized Observatory, <https://pharos.viralemergence.org/>) or FDA-ARGOS (US Food and Drug Administration’s dAtabase for Reference Grade micRObial Sequences, Sichtig et al. 2019), for example, are not automatically linked back to the original specimen record in Arctos or other museum databases, limiting interdisciplinary synthesis and independent verification. Thus, future innovation and expanded interoperability

may spur innovation in the “circular economy” of museum data users/providers. Changing taxonomies similarly challenges data discovery and aggregation. In mammals alone, the number of recognized species has increased by nearly 25% since publication of the last major compendium in 2005 (Burgin et al. 2025). When names change, as illustrated by the generic shift from *Myodes* to *Clethrionomys* (Kryštufek et al. 2020) in Example 3, specimen databases, sequence repositories, and other data aggregators are not automatically or even simultaneously updated, leading to mismatches across resources (Garg et al. 2019; Uetz et al. 2021).

Arctos data are discoverable through the web portal and various aggregators (GBIF, GGBN) and accessible via the web, its API, and now ArctosR. Use of code tables and fixed vocabulary ensures that Arctos data are interoperable among collections. Arctos also represents data as a network of relationships or interactions, which better accommodates cases where multiple taxonomic identifications are associated with a single specimen record (e.g., new species, elevated subspecies, etc.). While these data are currently in Arctos, they are not currently available for download through the API; future versions of ArctosR plan to add support for querying outside taxonomy providers such as <https://globalnames.org> or GBIF for synonyms when a user performs a query with `scientific_name`. ArctosR further increases interoperability by facilitating the download of data, formatting specimen data as R data frames and local files (e.g., CSV text files) and making them immediately compatible with other analyses and software. In addition, by automating the retrieval of datasets, the package facilitates the data acquisition process itself and makes it more transparent, reproducible, and reusable. The ability to save a complete query object that contains both the data and the query used to obtain it allows other researchers to easily replicate an analysis, verify its results, or build upon the work.

The Global Museum concept (Bakker et al. 2020) posits that when individual natural history collections are joined together through an interoperable digital extended specimen network (Lendemer et al. 2020), they become more than the sum of their parts, enabling interdisciplinary insights at a global scale. This vision of a unified, queryable digital biodiversity network allows researchers to address broad biological, environmental, and societal questions that transcend the specimen holdings of any single institution. The Arctos Consortium, which is responsible for the Arctos CMS and is by its very nature a collection of collections, embodies this principle (Cicero et al. 2024). The full realization of the Global Museum, however, de-

pends not just on data sharing and interoperability but on how accessible and usable those data are for the scientific community. By facilitating data retrieval and usability, ArctosR plays a crucial role in contributing to this latter goal.

The expanded accessibility to Arctos data enabled by ArctosR significantly contributes to research efforts addressing biodiversity knowledge shortfalls (Hortal et al. 2015). Core specimen metadata support efforts to understand species’ distributions and evolutionary relationships. Arctos also provides unique information, such as relational data about the ecology of species (e.g., host-parasite and host-pathogen relationships), filling knowledge gaps about species’ interactions (i.e., the Eltonian shortfall). These data provide context essential for research advances in ecology and can be extended to public health applications (Colella et al. 2023). At the same time, ArctosR reduces barriers to accessing museum data, ensuring that diverse information types are findable, accessible, interoperable, and reusable in line with FAIR principles. By providing transparent, code-driven workflows that facilitate reproducibility, the package advances the Global Museum’s vision of a unified biodiversity knowledge base.

ACKNOWLEDGEMENTS

This R package was developed with support from the 2024 Google Summer of Code program. This material is based upon work supported by the National Science Foundation under Grant No. 2302677. We thank Dusty MacDonald for Arctos API and API key support, and Luke Wang, an undergraduate at the University of California Berkeley’s Museum of Vertebrate Zoology, for beta testing.

COMPETING INTERESTS

The authors have declared that no competing interests exist.

DATA AVAILABILITY

The package ArctosR can be found on CRAN at <https://cran.r-project.org/package=ArctosR>. The source code is available on GitHub at <https://github.com/hrhwilliams/arctosR>, and any issues can be reported at <https://github.com/hrhwilliams/ArctosR/issues>. All data used to produce the examples presented here can be obtained with the code provided in the example cases on GitHub. The code for the examples is available as part of the vignettes of the package at: <https://hrhwilliams.github.io/ArctosR/>.

REFERENCES

- Araújo, M. B., R. P. Anderson, A. M. Barbosa, C. M. Beale, C. F. Dormann, R. Early, R. A. Garcia, A. Guisan, L. Maiorano, B. Naimi, R. B. O'Hara, N. E. Zimmermann, and C. Rahbek. 2019. Standards for distribution models in biodiversity assessments. *Sci. Adv.* 5:eaat4858.
- Bakker, F. T., A. Antonelli, J. A. Clarke, J. A. Cook, S. V. Edwards, P. G. P. Ericson, S. Faurby, N. Ferrand, M. Gelang, R. G. Gillespie, M. Irestedt, K. Lundin, E. Larsson, P. Matos-Maraví, J. Müller, T. von Proschwitz, G. K. Roderick, A. Schliep, N. Wahlberg, J. Wiedenhoef, and M. Källersjö. 2020. The Global Museum: Natural history collections and the future of evolutionary science and public education. *PeerJ* 8:e8225.
- Bergmann C. 1847. Über die Verhältnisse der Wärmeökonomie der Thiere zu ihrer Grösse. *Gött. Stud.* 3:595–708.
- Burgin, C. J., J. S. Zijlstra, M. A. Becker, H. Handika, J. M. Alston, J. Widness, S. Liphardt, D. G. Huckaby, and N. S. Upham. 2025. How many mammal species are there now? Updates and trends in taxonomic, nomenclatural, and geographic knowledge. *J. Mammal.* 106:1082–1117.
- Chamberlain, S. 2021. spocc: Interface to species occurrence data sources. R package. DOI: 10.32614/CRAN.package.spocc
- Chamberlain, S. A., and C. Boettiger. 2017. pygbif: A Python interface to the GBIF API. *J. Open Source Softw.* 2:297.
- Chamberlain, S., V. Barve, D. McGlinn, D. Oldoni, L. Gefert, and K. Ram. 2018. rgbif: Interface to the Global Biodiversity Information Facility API. R package. DOI: 10.32614/CRAN.package.rgbif
- Chang, W., Posit Software, PBC. 2025. R6: Encapsulated classes with reference semantics. R package. DOI: 10.32614/CRAN.package.R6
- Cicero, C., M. S. Koo, E. Braker, J. Abbott, D. Bloom, M. Campbell, J. A. Cook, J. R. Demboski, A. C. Doll, L. M. Frederick, A. J. Linn, T. J. Mayfield-Meyer, D. L. McDonald, M. W. Nachman, L. E. Olson, D. Roberts, D. S. Sikes, C. C. Witt, and E. A. Wommack. 2024. Arctos: Community-driven innovations for managing natural and cultural history collections. *PLoS ONE* 19:e0296478.
- Cobos, M. E., J. L. Dunnum, B. Armien, P. González, E. Juárez, J. R. Salazar, J. A. Cook, and J. P. Colella. 2025. Selecting sites for strategic surveillance of zoonotic pathogens: A case study in Panamá. *EcoHealth*, DOI: 10.1007/s10393-025-01731-z.
- Colella, J. P., J. Bates, S. F. Burneo, M. A. Camacho, C. C. Bonilla, I. Constable, G. D'Elia, J. L. Dunnum, S. Greiman, E. P. Hoberg, E. Lessa, S. W. Liphardt, M. Londoño-Gaviria, E. Losos, H. L. Lutz, N. O. Garza, A. T. Peterson, M. L. Martin, C. C. Ribas, B. Struminger, F. Torres-Pérez, C. W. Thompson, M. Weksler, and J. A. Cook. 2021. Leveraging natural history biorepositories as a global, decentralized, pathogen surveillance network. *PLOS Pathog.* 17:e1009583.
- Colella, J. P., M. E. Cobos, I. Salinas, J. A. Cook, and T. P. Consortium. 2023. Advancing the central role of non-model biorepositories in predictive modeling of emerging pathogens. *PLOS Pathog.* 19:e1011410.
- Cook, J. A., K. E. Galbreath, K. C. Bell, M. L. Campbell, S. Carrière, J. P. Colella, N. G. Dawson, J. L. Dunnum, R. P. Eckerlin, V. Fedorov, S. E. Greiman, G. M. S. Haas, V. Haukialmi, H. Henttonen, A. G. Hope, D. Jackson, T. S. Jung, A. V. Koehler, J. M. Kinsella, D. Krejsa, S. J. Kutz, S. Liphardt, S. O. MacDonald, J. L. Malaney, A. Makarikov, J. Martin, B. S. McLean, R. Mulders, B. Nyamsuren, S. L. Talbot, V. V. Tkach, A. Tsvetkova, H. M. Toman, E. C. Waltari, J. S. Whitman, and E. P. Hoberg. 2017. The Beringian Coevolution Project: Holistic collections of mammals and associated parasites reveal novel perspectives on evolutionary and environmental change in the North. *Arct. Sci.* 3:585–617.
- de Vries, M., T. Kearney, A. Monadjem, P. J. Taylor, and W. Markotter. 2025. Beyond spillover: Leveraging zoonotic disease research to advance biodiversity conservation. *One Health* 21:101280.
- Droege, G., K. Barker, O. Seberg, J. Coddington, E. Benson, W. G. Berendsohn, B. Bunk, C. Butler, E. M. Cawsey, J. Deck, M. Döring, P. Flemons, B. Gemeinholzer, A. Güntsch, T. Hollowell, P. Kelbert, I. Kostadinov, R. Kottmann, R. T. Lawlor, C. Lyal, J. Mackenzie-Dodds, C. Meyer, D. Mulcahy, S. Y. Nussbeck, É. O'Tuama, T. Orrell, G. Petersen, T. Robertson, C. Söhngen, J. Whitacre, J. Wieczorek, P. Yilmaz, H. Zetzsche, Y. Zhang, and X. Zhou. 2016. The Global Genome Biodiversity Network (GGBN) data standard specification. *Database* 2016:baw125.
- Fick, S. E., and R. J. Hijmans. 2017. WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* 37:4302–4315.
- Galbreath, K. E., E. P. Hoberg, J. A. Cook, B. Armien, K. C. Bell, M. L. Campbell, J. L. Dunnum, A. T. Dursahinhan, R. P. Eckerlin, S. L. Gardner, S. E. Greiman, H. Henttonen, F. A. Jiménez, A. V. A. Koehler, B. Nyamsuren, V. V. Tkach, F. Torres-Pérez, A. Tsvetkova, and A. G. Hope. 2019. Building an integrated

- infrastructure for exploring biodiversity: Field collections and archives of mammals and parasites. *J. Mammal.* 100:382–393.
- Garg, A., D. Leipe, and P. Uetz. 2019. The disconnect between DNA and species names: lessons from reptile species in the NCBI taxonomy database. *Zootaxa* 4706:401–407.
- Hardisty, A. R., E. R. Ellwood, G. Nelson, B. Zimkus, J. Buschbom, W. Addink, R. K. Rabeler, J. Bates, A. Bentley, J. A. B. Fortes, S. Hansen, J. A. Macklin, A. R. Mast, J. T. Miller, A. K. Monfils, D. L. Paul, E. Wallis, and M. Webster. 2022. Digital extended specimens: Enabling an extensible network of biodiversity data records as integrated digital objects on the internet. *BioScience* 72:978–987.
- Hijmans, R. 2023. terra: Spatial data analysis. R package. DOI: 10.32614/CRAN.package.terra
- Hijmans, R., M. Barbosa, A. Ghosh, and A. Mandel. 2023. geodata: Download geographic data. R package. DOI: 10.32614/CRAN.package.geodata
- Hortal, J., F. de Bello, J. A. F. Diniz-Filho, T. M. Lewinsohn, J. M. Lobo, and R. J. Ladle. 2015. Seven shortfalls that beset large-scale knowledge of biodiversity. *Annu. Rev. Ecol. Evol. Syst.* 46:523–549.
- Kays, R. W., and D. E. Wilson. 2009. *Mammals of North America*. Princeton University Press.
- Kryštufek, B., A. S. Tesakov, V. S. Lebedev, A. A. Bannikova, N. I. Abramson, and G. I. Shenbrot. 2020. Back to the future: The proper name for red-backed voles is *Clethrionomys Tlesius* and not *Myodes Pal-las*. *Mammalia* 84(2):214–217. DOI: 10.1515/mammalia-2019-0067
- Lendemer, J., B. Thiers, A. K. Monfils, J. Zaspel, E. R. Ellwood, A. Bentley, K. LeVan, J. Bates, D. Jennings, D. Contreras, L. Lagomarsino, P. Mabee, L. S. Ford, R. Guralnick, R. E. Gropp, M. Revelez, N. Cobb, K. Seltsmann, and M. C. Aime. 2020. The extended specimen network: A strategy to enhance US biodiversity collections, promote research and education. *BioScience* 70:23–30.
- Michonneau, F., M. Collins, S. Chamberlain, K. Love, H. N. Morzaria-Luna, M. L. Gaynor, and J. Bennett. 2024. ridigbio: Interface to the iDigBio data API. R package. DOI: 10.32614/CRAN.package.ridigbio
- Nelson, G. 2014. iDigBio: The US National Science Foundation's National Resource for Digitization of Biological and Palobiological Collections. Vancouver, Canada.
- Ooms, J. 2014. The jsonlite package: A practical and consistent mapping between JSON data and R objects. *ArXiv StatCO* 1403.2805.
- Ooms, J., H. Wickham, Posit Software, PBC. 2025. curl: A modern and flexible web client for R. R package. DOI: 10.32614/CRAN.package.curl
- Peterson, A. T., J. Soberón, R. G. Pearson, R. P. Anderson, E. Martínez-Meyer, M. Nakamura, and M. B. Araújo. 2011. *Ecological Niches and Geographic Distributions*. Princeton University Press, Princeton.
- Python Software Foundation. 2025. Python: A programming language.
- R Core Team. 2025. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Runck, A. M., M. D. Matocq, and J. A. Cook. 2009. Historic hybridization and persistence of a novel mito-nuclear combination in red-backed voles (genus *Myodes*). *BMC Evol. Biol.* 9:114.
- Sichtig, H., T. Minogue, Y. Yan, C. Stefan, A. Hall, L. Tallon, L. Sadzewicz, S. Nadendla, W. Klimke, E. Hatcher, M. Shumway, D. L. Aldea, J. Allen, J. Koehler, T. Slezak, S. Lovell, R. Schoepp, and U. Scherf. 2019. FDA-ARGOS is a database with public quality-controlled reference genomes for diagnostic use and regulatory science. *Nat. Commun.* 10:3313.
- Telenius, A. 2011. Biodiversity information goes public: GBIF at your service. *Nord. J. Bot.* 29:378–381.
- Trindade, W. C. F., L. F. Arias-Giraldo, L. Osorio-Olvera, A. T. Peterson, and M. E. Cobos. 2026. kuenm2: Detailed development of ecological niche models. R package. DOI: 10.32614/CRAN.package.kuenm2
- Uetz, P., M. Koo, R. Aguilar, E. Brings, A. Catenazzi, A. Chang, R. Chaitanya, P. Freed, J. Gross, M. Hammermann, J. Hosek, M. Lambert, Z. Sergi, C. Spencer, K. Summers, R. Tarvin, V. Vredenburg, and D. Wake. 2021. A quarter century of reptile and amphibian databases. *Herpetol. Rev.* 52:246–255.
- Upham, N. S., J. H. Poelen, D. Paul, Q. J. Groom, N. B. Simmons, M. P. M. Vanhove, S. Bertolino, D. M. Reeder, C. Bastos-Silveira, A. Sen, B. Sterner, N. M. Franz, M. Guidoti, L. Penev, and D. Agosti. 2021. Liberating host–virus knowledge from biological dark data. *Lancet Planet. Health* 5:e746–e750.
- Wickham, H. 2011. ggplot2. *WIREs Comput. Stat.* 3:180–185.
- Wiens, B. J., and J. P. Colella. 2025. Unidirectional mitochondrial introgression despite limited nuclear admixture in North American Red-Backed Voles, *Clethrionomys rutilus* and *C. gapperi*. *Ecol. Evol.* 15:e72603.
- Wilkinson, M. D., M. Dumontier, Ij. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo,

O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. G. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018.

Table S1. List of query parameters

Parameter name	Description	Category
began_date	Earliest date/time at which an event occurred.	event
chronological_extent	Count of days between begin and end dates	event
coll_event_remarks	Event remark	event
collecting_event_name	Unique (within Arctos) event identifier	event
collecting_method	Event method	event
collecting_source	Event source	event
ended_date	Latest date/time at which an event occurred	event
evtAttributeSearchTable	All event attribute fields	event
habitat	Event habitat	event
specimen_event_remark	Catalog record event remark	event
specimen_event_type	Type of event	event
verbatim_date	Date as provided	event
verbatim_locality	Locality as provided	event
verificationstatus	Verification status of locality information	event
family	Family as provided in collection's preferred source(s)	identification
genus	Genus as provided in collection's preferred source(s)	identification
identification	Any name used in identification (taxon, common name, string values)	identification
identification_order	Identification level of acceptedness	identification
identification_publication	"Sensu" publication title	identification
identification_remarks	Identification remark	identification
identificationattributetable	Identification details and metadata	identification
identified_agent	Identification determiner (agent)	identification
kingdom	Kingdom as provided in collection's preferred source(s)	identification
made_date	Identification date	identification
phylclass	Class as provided in collection's preferred source(s)	identification
phylorder	Order as provided in collection's preferred source(s)	identification
phylum	Phylum as provided in collection's preferred source(s)	identification
species	Species (binomial) as provided in collection's preferred source(s)	identification
subfamily	Subfamily as provided in collection's preferred source(s)	identification
suborder	Suborder as provided in collection's preferred source(s)	identification
subspecies	Subspecies (as trinomial) or infraspecific term (as quadrinomial) as asserted by the collection's preferred taxonomy	identification
subtribe	Subtribe as provided in collection's preferred source(s)	identification
superfamily	Superfamily as provided in collection's preferred source(s)	identification
taxa_formula	Identification formula	identification
taxon_name	Any name used in identification or related classification(s) (taxon, common name)	identification
tribe	Tribe as provided in collection's preferred source(s)	identification
taxonomySearchTable	Taxonomic name in any taxonomy source. Taxonomy used in identifications: Search VALUE OF (and specify an operator), optionally AS RANK, optionally ACCORDING TO SOURCE	identification
anyid	Catalog record Identifiers including catalog number, GUID (DarwinCore triplet), accession number, and part barcodes	identifier
cat_num	Primary identifier assigned to catalog records	identifier
custom_identifier	Identifier type selected in user profile	identifier
genbank_fragment	LIMITED USAGE: comma-delimited lookup of GenBank fragments will be slow and not recommended. Use Identifiers search with full URLs for best results	identifier
guid	DarwinCore Triplet; primary catalog record identifier	identifier
guid_prefix	Collection responsible for the record. Turning this off will break most forms	identifier
otherIdSrchTbl	All catalog record identifier fields	identifier
relatedItemSrchTbl	Search for records by data from identifier-linked records	identifier
anybarcode	Barcode for any container, including those that do not directly hold parts	parts
barcode	Part barcode	parts
condition	Part condition	parts
is_tissue	Has tissue	parts
part_container_type	Container type of the container directly holding the part	parts

part_ctr_last_date	Date the part container was last updated or changed	parts
part_disposition	Part disposition	parts
part_remark	Part remark	parts
part_search	All part and part attribute fields	parts
partAttributeSearchTable	All part attribute fields	parts
partname	Part name	parts
attribute_meta_age	Age drawn from lithostratigraphy metadata	place
attribute_meta_term	Locality attributes for chronostratigraphy. Searches of a unit will find all units nested within it	place
continent	Continent as asserted	place
coordinateprecision	Maximum coordinate error	place
country	Country as asserted by the collection	place
county	County as asserted by the collection	place
elevationSearchTable	Elevation fields (max, min, units)	place
feature	Feature (locality attribute)	place
geog_srch_type	Control how spatial data are matched. Default is contains/within	place
locality_name	Unique (within Arctos) locality identifier	place
locality_remarks	Locality remark	place
locAttributeSearchTable	All locality attribute fields	place
map	Draw a polygon on a map	place
ocean	Ocean as asserted	place
place_term	Standardized place names drawn from spatial data	place
quad	Quad (locality attribute)	place
sea	Sea as asserted	place
spec_locality	Locality as asserted	place
state_prov	State or province as asserted	place
accession_agency	Agent acting in an accession role	record
acn_number	Accession number	record
archive_name	Archive name	record
cataloged_item_type	Type of record	record
collector	Agent acting in a collector role	record
culture_of_origin	Culture of origin attribute	record
culture_of_use	Culture of use attribute	record
description	Description attribute	record
entered_by	Operator who entered the record	record
entered_date	Date the record was entered	record
is_peer_reviewed	Publication type	record
loan_number	Loan number	record
loan_project_name	Name of project using catalog records	record
materials	Materials attribute	record
media_keywords	Media keywords	record
media_type	Media type	record
permit_issued_by	Permit Issuer (agent)	record
permit_issued_to	Permit issued to (agent)	record
permit_num	Permit number	record
permit_type	Permit type	record
portfolio_or_series	A portfolio is a number of prints presented as a group and often, though not necessarily, by the same artist and based on a related theme. A series can be defined as more than one related work grouped by the artist or in some logical way. Use if there is a title for the portfolio or series	record
project_name	Name of project contributing catalog records	record
project_sponsor	Project sponsor	record
publication_doi	Publication DOI	record
publication_title	Publication title	record
recordAttributeSearchTable	All catalog record attribute fields	record
remark	Catalog record remark	record

subject_matter	The subject matter of a work of art (sometimes referred to as its content) is the narrative, iconographic, or non-objective meaning conveyed by an abstract or a figurative composition. It is what is depicted in and by a work of art. It also covers the function of an object or architecture that otherwise has no narrative content. Getty CDWA GitHub Issue Formerly subject, a topic of the resource. Dublin Core	record
type status	Nomenclatural type	record

Table S2. List of return columns

Parameter name	Description	Category
culture of origin	Culture of origin attribute	attribute
culture of use	Culture of use attribute	attribute
detected	The attribute value was detected	attribute
examined for	A systematic examination for the attribute value was made	attribute
materials	Materials attribute	attribute
not detected	The attribute value was not detected	attribute
not examined for	The record was not examined for the given value	attribute
reproductive data	Record attribute value(s)	attribute
family	Family as provided in collection's preferred source(s)	classification
genus	Genus as provided in collection's preferred source(s)	classification
phylclass	Class as provided in collection's preferred source(s)	classification
phylorder	Order as provided in collection's preferred source(s)	classification
phylum	Phylum as provided in collection's preferred source(s)	classification
kingdom	Kingdom as provided in collection's preferred source(s)	classification
species	Species (binomial) as provided in collection's preferred source(s)	classification
subfamily	Subfamily as provided in collection's preferred source(s)	classification
suborder	Taxonomic Order as asserted by the collection's preferred taxonomy	classification
subspecies	Subspecies (as trinomial) or infraspecific term (as quadrinomial) as asserted by the collection's preferred taxonomy	classification
subtribe	Subtribe as provided in collection's preferred source(s)	classification
superfamily	Superfamily as provided in collection's preferred source(s)	classification
superorder	Taxonomic SuperOrder as asserted by the collection's preferred taxonomy	classification
tribe	Tribe as provided in collection's preferred source(s)	classification
coordinateuncertaintyin-meters		core
guid	DarwinCore Triplet; primary catalog record identifier	core
country	Country as asserted by the collection	core
dec_lat	Geographic latitude expressed as degrees with decimal fractions of degrees	core
dec_long	Geographic longitude expressed as degrees with decimal fractions of degrees	core
scientific_name	Identification ("scientific name") as determined by collection. Depending on the determination, most often genus plus species; often genus, species plus subspecies, but often a taxon less precise than genus	core
spec locality	Locality as asserted	core
state_prov	State or province as asserted	core
verbatim_date	Date as provided	core
attributedetail	All attribute data	curatorial
citation_data	citations	curatorial
citations	Citations are links between cataloged items and publications. Citations indicate that a cataloged item was referred to (cited) by the authors of a publication	curatorial
collecting_event_id	Event identifier issued by Arctos	curatorial
collector_agents	Collector-agents (not limited to role 'collector')	curatorial
entered by	Operator who entered the record	curatorial
id_history	All Identifications including metadata	curatorial
lastpartlocation	Special handler for part attribute "location."	curatorial
media_tags	TAGs	curatorial
remove_row	Keep/Remove rows; Turn this on to keep/remove rows from results. USAGE: Click the button on any number of rows, then requery or select Tools-->Keep/Remove Selected Rows	curatorial
rights	Terms and licenses	curatorial
AF	Alaska Frozen Tissue Collection, University of Alaska Museum of the North, Fairbanks, Alaska	identifiers
collection_preferred_identifiers	Identifiers selected by the collection for special prominence	identifiers
collector_number	A value that refers to a person's field catalog. For this number to be applied to a specimen, the person has to be the first collector of a specimen	identifiers
doi	DOI associated with the catalog record	identifiers

GenBank	GenBank DNA-sequence accession. The Nucleotide database is a collection of sequences from several sources, including GenBank, RefSeq, TPA and PDB. Genome, gene and transcript sequence data provide the foundation for biomedical research and discovery	identifiers
KT	Kansas State Biorepository, Kansas Tissue Number	identifiers
NK	"New Mexico Karyotype Number," a frozen tissue collection number for the Museum of Southwestern Biology	identifiers
Organism_ID	Identifier representing a DWC:Organism or individual, generally one comprised of multiple catalog records (which represent DWC:Occurrences)	identifiers
preparator_number	A value that refers to a record in the catalog of the person listed as the first preparator of a specimen	identifiers
public accn id	Part of the path to any public accession information	identifiers
associated_species	Other species noted in association with the collected specimens. Used in botanical collections	locality
began_date	Earliest date/time at which an event occurred	locality
coll_event_remarks	Collecting Event Remarks documents nonstandard information concerning a collecting event	locality
continent	Continent as asserted by the collection	locality
county	County as asserted by the collection	locality
datum	Asserted datum for the coordinates assigned to a locality	locality
day_collected	The day in which the specimen was collected	locality
ended_date	Latest date/time at which an event occurred	locality
event_assigned_by_agent	Agent asserting the relationship between the specimen and the place-time data	locality
event_count	Number of Events, of any type and including restricted access, attached to the record	locality
formation	Locality attribute formation	locality
georeference_protocol	Protocol used to obtain the values for the coordinates and the measure of precision	locality
georeference_source	The source of the coordinate information. Often the name of a map or gazetteer	locality
json_locality	Locality stack in JSON	locality
locality_count	Number of Localities, of any type and including restricted access, attached to the record	locality
locality_name	Assigned unique persistent locality identifier	locality
locality_remarks	Locality remark	locality
max_depth_in_m	Depth in meters	locality
max_elev_in_m	Maximum elevation for a locality converted to meters	locality
maximum_elevation	Maximum elevation of a locality	locality
member	Locality attribute member	locality
min_depth_in_m	Depth in meters	locality
min_elev_in_m	Minimum elevation for a locality converted to meters	locality
minimum_elevation	Minimum elevation of a locality	locality
month_collected	The month in which the specimen was collected	locality
ocean	Ocean as asserted by the collection	locality
orig_elev_units	The units of measurement used in the original reporting of an elevation	locality
orig_lat_long_units	The units used in the original report of geographic coordinates. For example, degrees/minutes/seconds, degrees/minutes/decimal-minutes, decimal degrees, or the UTM system	locality
sea	Sea as asserted	locality
specimen_event_remark	Catalog record event remark	locality
specimen_event_type	Type of event	locality
verbatim_coordinates	The geographic coordinates in the format in which they were reported by the collector. (Often absent)	locality
verbatim_locality	Locality as provided	locality
year_collected	The year in which the specimen was collected	locality
accn_number	Accession number	record
catalognumberint	Integer component of catalog number	record
collecting_method	Event method	record
collecting_source	Event source	record
collectors	collector role "collector"	record

copyright_holder	collector role "copyright holder"	record
creator	collector role "creator"	record
formatted_scientific_name	Scientific name including code-specific data and markup	record
full_taxon_name	Any term in a taxonomic hierarchy	record
habitat	Event habitat	record
id_sensu	A publication which defines the parameters used to make the identification. "This specimen is Taxon X in the sense of Publication Y."	record
id_date	The date on which the identification was made	record
identification_remarks	Any further description of a taxonomic determination, its basis, nature or circumstances	record
identified_by	The agent who made the identification	record
identifiers	All non relationship-forming identifiers and associated metadata	record
individualcount		record
media	Media include images, digital audio, and various kinds of documentation. More generally, media are digital files related to specimens, publications, and agents	record
othercatalognumbers	Identifier kind=value concatenation	record
partdetail	All part fields formatted as JSON	record
parts	Part names	record
preparators	collector role "preparator"	record
relatedcatalogeditems	Two cataloged items (usually biological individuals) may be associated with each other by kinship or some trophic relationship	record
remark		record
sci_name_with_auth	Taxonomic name followed by the name(s) of the person who originally published the name. Various specific conventions are dictated by nomenclatural commissions	record
subject	collector role "subject"	record
verificationstatus	Verification status of locality information	record
age	General description of the age. See also age class and numeric age	trait
ear_from_notch	Length of the ear from the notch to the tip	trait
hind_foot_with_claw	Length of the hind foot from the back of the ankle to the end of the longest claw. A numeric value plus units	trait
life_stage	Summary of values for record attribute life stage	trait
sex	Sex of the organism	trait
tail_lenth	length of tail	trait
total_length	Total length of the specimen. Fish: Total length (TL) is the length of a fish measured from the tip of the snout to tip of the longest lobe of caudal fin. Measurement done with caudal fin lobes compressed and measured at midline	trait
weight	Total body or object weight	trait