Contribution to a reference library of DNA barcodes for Colombian freshwater fishes

Manuela Mejía Estrada, ⓘD Luz Fernanda Jiménez-Segura, Iván Soto Calderón

Contribution to a reference library of DNA barcodes for Colombian freshwater fishes

Manuela Mejía Estrada[‡], Luz Fernanda Jiménez-Segura[‡], Iván D Soto Calderón[‡]

‡ Universidad de Antioquia, Medellín, Colombia

Corresponding author: Manuela Mejía Estrada (manuela.mejiae@udea.edu.co), Luz Fernanda Jiménez-Segura (luz.jimenez@udea.edu.co), Iván D Soto Calderón (ivan.soto@udea.edu.co)

Abstract

Background

The proposal for DNA Barcoding was and motivated by the mismatch between the low number of taxonomists contrasting with the large number of species. The method requires the construction of reference collections of DNA sequences that represent existing biodiversity. Freshwater fishes are key indicators for understanding biogeography around the world. Colombia, with 1610 species of freshwater fishes, is the second richest country in the world in terms of this group. However, as genetic information for the species continues to be limited, the contribution to a reference library of DNA barcodes for Colombian freshwater fishes highlights the importance of biological collections and seeks to strengthen inventories and taxonomy of such collections in future studies.

New information

This dataset contributes to the knowledge on the DNA barcodes and occurrence records of 96 species of freshwater fishes from Colombia. The species represented in this dataset correspond to an addition of 39 species to the BOLD public database. Forty-nine specimens were collected in Atrato Basin and 708 in Magdalena-Cauca Basin during the period from 2010 to 2020, two species (Loricariichthys brunneus and Poecilia sphenops) are considered exotic to the Atrato, Cauca and Magdalena Basins and four species (Oncorhynchu mykiss, Oreochromis niloticus, Parachromis friedrichsthalii and Xiphophorus helleri) are exotic to Colombian hydrogeographic regions. All specimens are deposited in the CIUA collection at the University of Antioquia and have their DNA barcodes made publicly available in the Barcode of Life Data System (BOLD) online database. This distribution dataset can be freely accessed through the Global Biodiversity Information Facility (GBIF).

Keywords

freshwater fishes, occurrence records, Colombian biodiversity, exotic species, DNA barcode, COX1

Introduction

Neotropical freshwater fishes constitute the most diverse continental vertebrate fauna on Earth, with more than 6,200 named species compressed into < 0.5% the total land-surface area and represent the greatest phenotypic disparity and functional diversity of any continental ichthyofauna (Albert et al. 2020). This fauna is still in a pioneering stage of discovery, with dozens of new species being described each year. The current pace of discovery indicates that the actual Neotropical freshwater fishes diversity could exceed 9,000 species, meaning as many as one-third of the species in the wild have yet to be described (Reis et al. 2016).

Colombia has 1610 species of freshwater fishes (FF) described, consolidating itself as the second richest country in terms of this group in the world. In particular, the trans-Andean Basin of the Magdalena and Cauca Rivers presents altitudinal gradients that testify to geological and climatic events where 232 species of fish have been registered, of which 57% are endemic. Atrato Basin, part of the Pacific hydrogeographic region, has 128 described species, 32 being endemic to this region (DoNascimiento et al. 2018). Despite the efforts of research bodies in collecting data, much remains to be determined about the diversity of FF in the country. The huge diversity of these organisms, the shortage of specialised taxonomists and the difficulties in identifying many species are the main obstacles for fully addressing this lack of knowledge.

The DNA barcode is an approach that aims to perform a rapid identification of species of a taxonomic group, based on the amplification of a fragment of the sequence of the mitochondrial gene cytochrome c oxidase subunit I (COX1) and its comparison to those previously sequenced from morphologically identified specimens (Kress and Erickson 2012, Hebert et al. 2003). As the efficacy of COX1 in the discrimination of fish species has previously been demonstrated (Thu et al. 2019, Ude et al. 2020, Ward et al. 2005), this can be very useful for the identification of Colombian ichthyofauna, considering the high diversity of fish in the country and the rate of ten species of freshwater fishes described per year (DoNascimiento et al. 2017).

This contribution to a reference library of DNA barcodes for Colombian freshwater fishes contains records of 757 specimens collected in three basins of Colombia, all of which were morphologically identified to species level, when possible, for a total of 94 species (673 records to species level) and 63 genera (84 records to genus level). All specimens have their DNA barcodes made publicly available in the Barcode of Life Data System (BOLD). Overall, this paper is a contribution to sharing and publicly disseminating the occurrence

records and DNA barcodes of specimens from the reference collection of the University of Antioquia in order to increase the available information on Colombian freshwater fishes.

General description

Purpose: This dataset aims to provide information from a COX1 gene sequence database for freshwater fish species present in Atrato, Cauca and Magdalena Basins in Colombia (Fig. 1), facilitating the identification of species by molecular methods and also being a tool for future metabarcoding studies and monitoring of basins, as well as highlighting the importance of biological collections.

Project description

Title: The title "Contribution to a reference library of DNA barcodes for Colombian freshwater fishes" refers to the publication of sequences of the cytochrome oxidase 1 gene generated in the Ichthyology Collection of the University of Antioquia of the fishes catalogued there which occur in the Atrato, Cauca and Magdalena Basins in Colombia.

Personnel: Manuela Mejía Estrada (Project Developer, Student)

Iván D. Soto Calderón (Project Mentor, Teacher)

Luz Fernanda Jiménez Segura (Project Coordinator, Teacher)

Study area description: The Atrato, Cauca and Magdalena Basins in Colombia (Fig. 2).

Design description: Freshwater fish specimens were collected in the field, using different fishing methods, morphologically identified and DNA barcoded.

Funding: This project was funded by "Empresas publicas de medelin, EPM" under the "BIO" agreement No. CT-2017-001714 with the University of Antioquia.

Sampling methods

Study extent: The Atrato, Cauca and Magdalena Basins, Colombia.

Sampling description: The analysed material was collected from 138 different localities in the Atrato, Cauca and Magdalena Basins in Colombia. Sampling was conducted between 2010 and 2020 on a wide range of habitats, using different fishing methods. Collected specimens were fixed and stored in alcohol and a portion of muscle or fin was stored in 96% ethanol for downstream molecular analysis. Morphological identification was performed, based on taxonomical keys and descriptions from literature.

DNA was extracted using the QIAgen Dneasy Blood & Tissue kit (Hilden, Germany), following the manufacturer's protocol. The COX1 fragment was amplified using the

primers FishF1 (5'-TCAACCAACCACAAAGACATTGGCAC-3') and FISHR1 (5'-TAGACTTCTGGGTGGCCAAAGAATCA-3') (Ward et al. 2009). The products were sent to a commercial laboratory to be purified and sequenced by the Sanger Method. The forward and reverse sequences were edited and assembled using Geneious Prime (2019) software and inspected manually.

Step description: Specimens were collected from 138 different localities of three Colombian Basins (Atrato, Cauca and Magdalena). Sampling was conducted from 2010 to 2020 and consisted of the use of different fishing methods to collect specimens. Collected specimens were stored in 70% ethanol. A tissue sample from muscle or fin was removed and stored in 96% ethanol, from which DNA was extracted and the COX1 DNA barcode fragment was sequenced. Data generated was submitted to BOLD and is part of the data from the Ichthyology Collection of the University of Antioquia in GBIF.

Geographic coverage

Description: Atrato, Cauca and Magdalena Basins, continental Colombia.

Coordinates: 4.53888 and 8.89651 Latitude; -76.81916667 and -73.55097222 Longitude.

Taxonomic coverage

Description: This dataset is composed of data relating to 757 specimens of freshwater fishes that occur in Colombia; 673 specimens were identified to the species level and 84 to genus. Overall, 94 species in 24 families are represented in the dataset Suppl. material 1. The families Characidae, Astroblepidae and Loricariidae account for 71% of the total collected specimens. One family (Sciaenidae) is represented by a single sequence, whereas seven families (Apteronotidae, Aspredinidae, Callichthyidae, Cynolebiidae, Erythrinidae, Salmonidae (exotic) and Scianidae) are represented by a single species. Fig. 3 illustrates examples of species that are part of the CIUA dataset.

Taxa included:

Rank	Scientific Name
species	*Ageneiosus pardalis*
species	*Andinoacara latifrons*
species	*Apteronotus eschmeyeri*
species	*Argopleura magdalenensis*
genus	*Astroblepus*
species	*Astroblepus grixalvii**
species	*Astroblepus chapmani**

species	*Astroblepus guentheri**
species	*Astroblepus homodon**
species	*Astroblepus latidens**
species	*Astroblepus micrescens**
species	*Astroblepus trifasciatus**
species	*Astroblepus unifasciatus**
genus	*Astyanax*
species	*Astyanax atratoensis**
species	*Astyanax gisleni*
species	*Astyanax magdalenae*
species	*Astyanax microlepis*
species	*Bryconamericus caldasi*
species	*Brycon henni*
species	*Brycon moorei*
species	*Brycon rubricauda**
species	*Bunocephalus colombianus**
species	*Caquetaia kraussii*
species	*Carlastyanax aurocaudatus*
species	*Cetopsorhamdia boquillae**
species	*Cetopsorhamdia nasus**
genus	*Chaetostoma*
species	*Chaetostoma leucomelas**
species	*Chaetostoma milesi**
species	*Chaetostoma thomsoni**
species	*Characidium phoxocephalum*
genus	*Characidium*
species	*Coptodon rendalli*
species	*Cordylancistrus pijao**
genus	*Creagrutus*
species	*Creagrutus affinis*
species	*Creagrutus brevipinnis*

species	*Creagrutus caucanus**
species	*Creagrutus magdalenae*
species	*Curimata mivartii*
species	*Cynodonichthys magdalenae**
species	*Cyphocharax magdalenae*
species	*Dasyloricaria filamentosa**
species	*Eigenmannia humboldtii*
species	*Eigenmannia virescens*
species	*Geophagus pellegrini*
species	*Geophagus steindachneri*
species	*Gephyrocharax melanocheir**
species	*Gilbertolus alatus**
species	*Gilbertolus atratoensis**
genus	*Hemibrycon*
species	*Hemibrycon antioquiae*
species	*Hemibrycon boquiae**
species	*Hemibrycon fasciatus**
species	*Hemibrycon caucanus**
species	*Hemibrycon palomae**
species	*Hemibrycon rafaelense**
species	*Hemibrycon raqueliae**
family	*Heptapteridae*
species	*Hoplias malabaricus*
species	*Hoplosternum magdalenae*
species	*Ichthyoelephas longirostris*
species	*Hypostomus hondae**
genus	*Imparfinis*
species	*Lasiancistrus caucanus*
species	*Leporellus vittatus*
species	*Leporinus striatus*
species	*Loricariichthys brunneus**

species	*Megaleporinus muyscorum*
species	*Mesoheros atromaculatus*
species	*Oncorhynchus mykiss*
species	*Oreochromis mossambicus*
species	*Oreochromis niloticus*
species	*Parachromis friedrichsthalii*
species	*Parodon magdalenensis*
species	*Parodon suborbitalis**
species	*Pimelodella chagresi*
species	*Pimelodella macrocephala**
species	*Pimelodus grosskopfii*
species	*Pimelodus yuma*
species	*Plagioscion magdalenae*
species	*Poecilia caucana*
species	*Poecilia sphenops*
species	*Prochilodus magdalenae*
species	*Psalidodon fasciatus*
genus	*Pseudopimelodus*
species	*Pseudoplatystoma magdaleniatum*
species	*Pterygoplichthys undecimalis**
species	*Rhamdia guatemalensis*
species	*Roeboides dayi*
species	*Saccodon dariensis*
species	*Salminus affinis*
species	*Sorubim cuspicaudus*
species	*Spatuloricaria gymnogaster*
species	*Steindachnerina atratoensis**
species	*Sternopygus aequilabiatus*
species	*Sturisomatichthys leightoni**
species	*Sturisomatichthys panamensis**
species	*Trachelyopterus insignis*

genus	*Trichomycterus*
species	*Trichomycterus banneaui*
species	*Trichomycterus caliensis**
species	*Trichomycterus chapmani**
species	*Trichomycterus striatus*
species	*Triportheus magdalenae*

Temporal coverage

Data range: 2010-1-01 - 2020-3-01.

Collection data

Collection name: Colección de Ictiología de la Universidad de Antioquia CIUA

Collection identifier: Registro Nacional de Colecciones Biológicas: 168

Parent collection identifier: CIUA

Specimen preservation method: ethanol 70%

Usage licence

Usage licence: Creative Commons Public Domain Waiver (CC-Zero)

Data resources

Data package title: Barcoding CIUA

Resource link: http://www.boldsystems.org/index.php/Public_SearchTerms?searchMenu=records&query=ciua&taxon=

Alternative identifiers: dx.doi.org/10.5883/DS-CIUA01

Number of data sets: 1

  Data set name: CIUA01

  Download URL: http://www.boldsystems.org/index.php/Public_SearchTerms?query=DS-CIUA01

  Data format: xml, fasta

Description: The Barcoding CIUA Database: CIUA01 dataset can be downloaded from the Public Data Portal of BOLD (dx.doi.org/10.5883/DS-CIUA01) in xml format and sequences as fasta files, including 18 columns with the information described below Suppl. material 2. Alternatively, BOLD users can log-in and access the dataset via the Workbench platform of BOLD. All records are also searchable within BOLD, using the search function of the database.

The Barcoding CIUA will continue sequencing Colombian freshwater fishes for the BOLD database, with the main objective of achieving a representative coverage of the Colombian fishes species.

All records of the collection have been reported to the SiB database (Jiménez-Segura et al. 2021) in Darwin Core Standard Format (Suppl. material 3) that includes 58 Columns with the information shown and described in Table 1.

Column label	Column description
Project Code	Unique Code for the project
Process ID	Unique identifier for the sample
SampleID	ID for the specimen in BOLD Database
BIN	Barcode Index Number system identifier
Catalogue Num	Number of the record in the collection
COI-5P Seq. Length	Length of the sequence
Identification	Current identification of the record
Institution-Institution Storing	Name of the institution that has physical possession of the voucher specimen
Museum ID	Unique number of identification for the record at the museum where it is stored
Phylum	Phylum to which the record belongs
Class	Class to which the record belongs
Order	Order to which the record belongs
Family	Family to which the record belongs
Genus	Genus to which the record belongs
Species	Species to which the record belongs
Country	The full, unabbreviated name of the country where the organism was collected
Latitude	The geographical latitude (in decimal degrees) of the geographic centre of a location
Longitude	The geographical longitude (in decimal degrees) of the geographic centre of a location

Acknowledgements

References

- Albert JS, Dagosta F, Tagliacollo VA, et al. (2020) Diversification of Neotropical freshwater fishes. Annual Review of Ecology, Evolution and Systematics 51: 27-53. https://doi.org/10.1146/annurev-ecolsys-011620-031032
- DoNascimiento C, Herrera-Collazos EE, Herrera-R. G, Ortega-Lara A, Villa-Navarro F, Usma Oviedo JS, Maldonado-Ocampo J (2017) Checklist of the freshwater fishes of Colombia: a Darwin Core alternative to the updating problem. ZooKeys 708: 25-138. https://doi.org/10.3897/zookeys.708.13897
- DoNascimiento C, Collazos EEH, Maldonado-Ocampo JA, et al. (2018) Lista de especies de peces de agua dulce de Colombia/Checklist of the freshwater fishes of Colombia. v.2.10. Asociación Colombiana de Ictiólgos. https://doi.org/10.15472/numrso
- Hebert PN, Cywinska A, Ball S, deWaard J (2003) Biological identifications through DNA barcodes. Proceedings of the Royal Society of London. Series B: Biological Sciences 270 (1512): 313-321. https://doi.org/10.1098/rspb.2002.2218
- Jiménez-Segura LF, Herrera Pérez J, Ospina Pabón JG, Hernández Zapata M (2021) Colección de Peces de la Universidad de Antioquia. v.5.5. Dataset/Occurrence. . Universidad de Antioquia. URL: https://doi.org/10.15472/lkcff8
- Kress WJ, Erickson D (2012) DNA barcodes: Methods and protocols. DNA Barcodes3-8. https://doi.org/10.1007/978-1-61779-591-6_1
- Reis RE, Albert JS, Di Dario F, Mincarone MM, Petry P, Rocha LA (2016) Fish biodiversity and conservation in South America. Journal of Fish Biology 89 (1): 12-47. https://doi.org/10.1111/jfb.13016
- Thu PT, Huang W, Chou T, Van Quan N, Van Chien P, Li F, Shao K, Liao T (2019) DNA barcoding of coastal ray-finned fishes in Vietnam. PLoS One 14 (9): e0222631. https://doi.org/10.1371/journal.pone.0222631
- Ude G, Igwe D, Brown C, Jackson M, Bangura A, Ozokonkwo-Alor O, Ihearahu O, Chosen O, Okoro M, Ene C, Chieze V, Unachukwu M, Onyia C, Acquaah G, Ogbonna J, Das A (2020) DNA barcoding for identification of fish species from freshwater in Enugu and Anambra States of Nigeria. Conservation Genetics Resources 12 (4): 643-658. https://doi.org/10.1007/s12686-020-01155-7
- Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PN (2005) DNA barcoding Australia's fish species. Philosophical Transactions of the Royal Society B: Biological Sciences 360 (1462): 1847-1857. https://doi.org/10.1098/rstb.2005.1716

- Ward RD, Hanner R, Hebert PDN (2009) The campaign to DNA barcode all fishes, FISH-BOL. Journal of Fish Biology 74 (2): 329-356. https://doi.org/10.1111/j.1095-8649.2008.02080.x
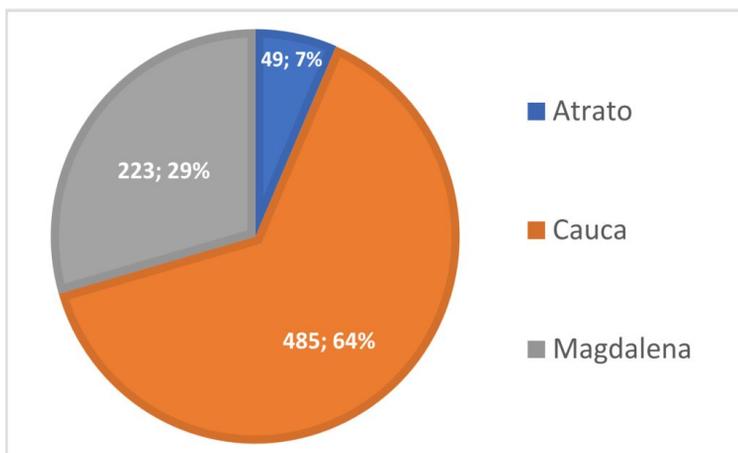
Figure 1.

Number of sequences generated in the Basins of the Atrato, Cauca and Magdalena Rivers in Colombia.
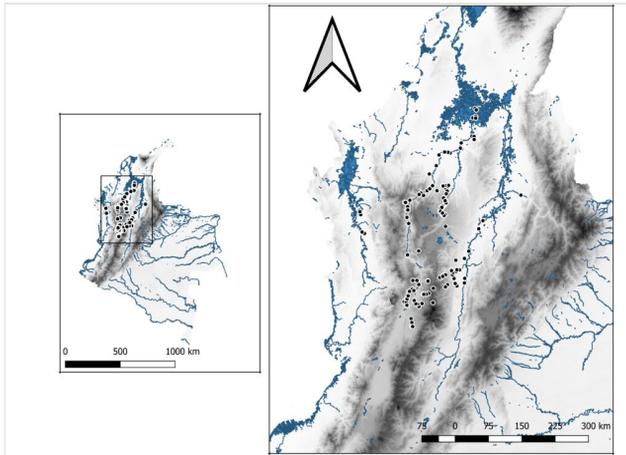
Figure 2.

Map of the localities where freshwater fish samples were collected in Atrato, Cauca and Magdalena Basins in Colombia.

Figure 3.

Examples of the diversity of species that are part of the dataset of "Contribution to a reference library of DNA barcodes for Colombian freshwater fishes". All photos are by Juan Guillermo Ospina.

a: *Astroblepus grixalvii*
b: *Chaetostoma thomsoni*
c: *Hoplosternum magdalenae*
d: *Trichomycterus* sp.

Table 1.

Columns labels included in the Darwin Core standard format of the Ichthyology Collection of the University of Antioquia and their description.

Column label	Description
ocurrenceID	Unique identifier code in the collection for the record
institutionCode	The name (or acronym) in use by the institution having custody of the object(s) or information referred to in the record.
basisOfRecord	The specific nature of the data record.
CollectionCode	The name, acronym, coden or initialism identifying the collection or dataset from which the record was derived.
catalogNumber	An unique identifier for the record within the collection.
type	The nature or genre of the resource.
language	A language of the resource.
institutionID	An identifier for the institution having custody of the object(s) or information referred to in the record.
collectionID	An identifier for the collection or dataset from which the record was derived.
recordedBy	A list of names of people, groups or organisations responsible for recording the original Occurrence.
individualCount	The number of individuals represented present at the time of the Occurrence.
LifeStage	The age class or life stage of the biological individual(s) at the time the Occurrence was recorded.
preparations	A list of preparations and preservation methods for a specimen.
eventDate	Date-time when the event was recorded.
year	The four-digit year in which the Event occurred, according to the Common Era Calendar.
habitat	A category or description of the habitat in which the Event occurred.
continent	The name of the continent in which the Location occurs.
waterBody	The name of the water body in which the Location occurs.
island	The name of the island on or near which the Location occurs.
country	The name of the country or major administrative unit in which the Location occurs.
countryCode	The standard code for the country in which the Location occurs.
stateProvince	The name of the next smaller administrative region than country (state, province, canton, department, region etc.) in which the Location occurs.

county	The full, unabbreviated name of the next smaller administrative region than stateProvince (county, shire, department etc.) in which the Location occurs.
municipality	The full, unabbreviated name of the next smaller administrative region than county (city, municipality etc.) in which the Location occurs.
locality	The specific description of the place.
verbatimLocality	The original textual description of the place.
minimumElevationInMetres	The lower limit of the range of elevation (altitude, usually above sea level), in metres.
maximumElevationInMetres	The upper limit of the range of elevation (altitude, usually above sea level), in metres.
locationRemarks	Comments or notes about the Location.
verbatimCoordinates	The verbatim original spatial coordinates of the Location.
verbatimLatitude	The verbatim original latitude of the Location.
verbatimLongitude	The verbatim original longitude of the Location.
verbatimCoordinateSystem	The coordinate format for the verbatimLatitude and verbatimLongitude or the verbatimCoordinates of the Location.
decimalLatitude	The geographic latitude (in decimal degrees, using the spatial reference system given in geodeticDatum) of the geographic centre of a Location.
decimalLongitude	The geographic longitude (in decimal degrees, using the spatial reference system given in geodeticDatum) of the geographic centre of a Location.
geodeticDatum	The ellipsoid, geodetic datum or spatial reference system (SRS) upon which the geographic coordinates given in decimalLatitude and decimalLongitude as based.
coordinateUncertaintyInMetres	The horizontal distance (in metres) from the given decimalLatitude and decimalLongitude describing the smallest circle containing the whole of the Location.
georeferencedBy	A list of names of people, groups or organisations who determined the georeference (spatial representation) for the Location.
georeferencedDate	The date on which the Location was georeferenced.
georeferenceProtocol	A description or reference to the methods used to determine the spatial footprint, coordinates and uncertainties.
georeferenceRemarks	Notes or comments about the spatial description determination, explaining assumptions made in addition or opposition to the those formalised in the method referred to in georeferenceProtocol.
identifiedBy	A list of names of people, groups or organisations who assigned the Taxon to the subject.
dateIdentified	The date on which the subject was determined as representing the Taxon.

identificationQualifier	A brief phrase or a standard term ("cf.", "aff.") to express the determiner's doubts about the Identification.
scientificName	The full scientific name, with authorship and date information, if known.
kingdom	The full scientific name of the kingdom in which the taxon is classified.
phylum	The full scientific name of the phylum or division in which the taxon is classified.
class	The full scientific name of the class in which the taxon is classified.
order	The full scientific name of the order in which the taxon is classified.
family	The full scientific name of the family in which the taxon is classified.
genus	The full scientific name of the genus in which the taxon is classified.
specificEpithet	The name of the first or species epithet of the scientificName.
infraspecificEpithet	The name of the lowest or terminal infraspecific epithet of the scientificName, excluding any rank designation.
taxonRank	The taxonomic rank of the most specific name in the scientificName.
scientificNameAutorship	The authorship information for the scientificName formatted according to the conventions of the applicable nomenclaturalCode.

Supplementary materials

Suppl. material 1: Barcoding CIUA01-Specimen details

Authors: Manuela Mejia Estrada, Juliana Herrera, Omer Campo, Marcela Hernandez.
Data type: Record information, Specimen data.
Brief description: The file includes information about all records in BOLD for the Barcoding CIUA01 library. It contains collection, location and identification data.
Download file (146.02 kb)

Suppl. material 2: Barcoding CIUA01 Library

Authors: Manuela Mejía Estrada, Omer Campo, Marcela Hernandez.
Data type: DNA Sequences, COX1 Sequences.
Brief description: COX1 sequences in fasta format. Each sequence is identified by the BOLD Sample ID, species name and sequence category, separated by pipe.
Download file (448.06 kb)

Suppl. material 3: Darwin Core Format of the Ichthyology Collection of the University of Antioquia.

Authors: Jiménez-Segura L F, Herrera Pérez J, Ospina Pabón J G, Hernández Zapata M
Data type: Occurrences and metadata asociated.
Brief description: Matrix in Darwin Core Standard Format of the Ichthyology Collection of the University of Antioquia.
Download file (6.83 MB)