

RESEARCH ARTICLE

Hidden biodiversity in Neotropical streams: DNA barcoding uncovers high endemism of freshwater macroinvertebrates at small spatial scales

Luis F. De León^{1,2,3*}, Aydeé Cornejo⁴, Ronnie G. Gavilán^{5,6}, Celestino Aguilar^{2,3}

1 Department of Biology, University of Massachusetts Boston, Boston, MA, United States of America, **2** Centro de Biodiversidad y Descubrimiento de Drogas, Instituto de Investigaciones Científicas y Servicios de Alta Tecnología (INDICASAT AIP), Panamá, República de Panamá, **3** Smithsonian Tropical Research Institute, Balboa, Panamá, **4** Instituto Conmemorativo Gorgas de Estudios de la Salud, Panamá, República de Panamá, **5** Centro Nacional de Salud Pública, Instituto Nacional de Salud, Lima, Perú, **6** Escuela Profesional de Medicina Humana, Universidad Privada San Juan Bautista, Lima, Perú

* luis.deleonreyna@umb.edu



OPEN ACCESS

Citation: De León LF, Cornejo A, Gavilán RG, Aguilar C (2020) Hidden biodiversity in Neotropical streams: DNA barcoding uncovers high endemism of freshwater macroinvertebrates at small spatial scales. *PLoS ONE* 15(8): e0231683. <https://doi.org/10.1371/journal.pone.0231683>

Editor: Michael A. Chadwick, King's College London, UNITED KINGDOM

Received: March 28, 2020

Accepted: July 8, 2020

Published: August 7, 2020

Copyright: © 2020 De León et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: COI sequences associated with this study are available in GenBank (<https://www.ncbi.nlm.nih.gov>) under accession numbers KX039451 - KX039650 and KU980966 - KU981004. Data are also available on the Barcode of Life Data System (BOLD) systems, under project code INVPA: <http://www.boldsystems.org/>. Taxonomic and molecular diversity data have been provided in Supporting information (S1 Table).

Funding: Financial support was provided by a grant from the Consortium for the Barcode of Life to L. F.

Abstract

Aquatic macroinvertebrates play a crucial role in freshwater ecosystems, but their diversity remains poorly known, particularly in the tropics. This “taxonomic void” limits our understanding of biodiversity patterns and processes in freshwater ecosystems, and the scale at which they operate. We used DNA barcoding to estimate lineage diversity (and the diversity of unique haplotypes) in 224 specimens of freshwater macroinvertebrates at a small spatial scale within the Panama Canal Watershed (PCW). In addition, we compiled available barcoding data to assess macroinvertebrate diversity at a broader spatial scale spanning the Isthmus of Panama. Consistently across two species delimitation algorithms (i.e., ABGD and GMYC), we found high lineage diversity within the PCW, with ~ 100–106 molecular operational taxonomic units (MOTUs) across 168 unique haplotypes. We also found a high lineage diversity along the Isthmus of Panama, but this diversity peaked within the PCW. However, our rarefaction/extrapolation approach showed that this diversity remains under-sampled. As expected, these results indicate that the diversity of Neotropical freshwater macroinvertebrates is higher than previously thought, with the possibility of high endemism even at narrow spatial scales. Consistent with previous work on aquatic insects and other freshwater taxa in this region, geographic isolation is likely a main factor shaping these patterns of diversity. However, other factors such as habitat variability and perhaps local adaptation might be reshaping these patterns of diversity at a local scale. Although further research is needed to better understand the processes driving diversification in freshwater macroinvertebrates, we suggest that Neotropical streams hold a high proportion of hidden biodiversity. Understanding this diversity is crucial in the face of increasing human disturbance.

De León. Additional funding was provided by Secretaría Nacional de Ciencia, Tecnología e Innovación (SENACYT; Grants No. ITE12-002 and FID16-116 to L. F. De León), Sistema Nacional de Investigación (SNI; to C. Aguilar), and the University of Massachusetts Boston (to Luis F. De León). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Aquatic macroinvertebrates are a fundamental component of freshwater environments. They mediate important processes such as food web dynamics, energy flow, and nutrient cycling, and therefore play a central role in sustaining the biodiversity and functioning of freshwater ecosystems [1–3]. However, the diversity of Neotropical freshwater macroinvertebrates remains poorly described, and even less is known about the processes that drive their diversity, and the scale at which they operate [4]. For instance, despite considerable efforts by local taxonomists (e.g., [5–8]), most of the published literature use genus and family as a standard taxonomic unit for Neotropical macroinvertebrates (e.g., [9–14]). This is partially due to the complexity of these communities, which are often composed of multiple life-stages existing at the interface between the terrestrial and aquatic environment [15,16]. Another limitation is the low efficiency of traditional morphological methods, which are generally time-consuming, and highly variable in the quality of identification across taxa and experts.

This “taxonomic void” has important consequences for our general understanding of biodiversity patterns and processes, both in Neotropical environments and globally. For example, species diversity is generally expected to increase at lower latitudes [17,18], but no consensus has been reached for macroinvertebrates, given the current lack of taxonomic knowledge [19–22]. Within the Neotropics, our current understanding of the drivers of species diversity in benthic macroinvertebrates is also limited [23–25].

Similar to other freshwater taxa (e.g., [26,27]), spatial isolation is likely a major factor driving diversification in macroinvertebrates, but few studies have tested this expectation [25,28,29]. In particular, Múrria et al. [25] found a high frequency of unique haplotypes associated with the geographical distance across watersheds in Panama. While confirmatory, these findings are not surprising, given the large geographic distance among the watersheds included in Múrria et al. [25]. However, patterns of haplotype (or lineage) diversity at smaller scales (e.g., among streams within watersheds), where dispersal and gene flow might be less restricted, have received less attention. We use DNA-barcoding to assess patterns of lineage diversity (and the diversity of unique haplotypes) in freshwater macroinvertebrates in four streams within the Panama Canal Watershed (PCW). In addition, we compiled available barcoding data [25] to contrast macroinvertebrate diversity at a broader spatial scale, among eight streams along the Isthmus of Panama.

Assessing the patterns and drivers of macroinvertebrate diversity at different scales is particularly relevant, given the increasing rate of environmental degradation in Neotropical regions [30–33]. This includes alterations such as introduction of alien species [34,35], the development of megaprojects [36], habitat degradation, water pollution, and climate change [30,33,37,38]. As a consequence, a large portion of this biodiversity risks being lost before discovery.

Material and methods

Study sites and sample processing

Samples were collected from four streams within the PCW (Frijolito, Frijoles, Trinidad, and Indio) between April and May of 2013 (Fig 1). Frijolito (09°08'57.9" N, 79°43'53.2" W) and Frijoles (09°09'08.2" N, 79°44'05.3" W) are typical Neotropical streams separated by approximately 300 m and located inside Soberanía National Park. These streams are surrounded by dense secondary forest and present low levels of disturbance. Río Trinidad (8°58'28.50" N, 79°57'23.9" W) is located approximately 30 km West Río Frijoles in an agricultural landscape dominated by pasture, but it has abundant riparian vegetation. Río Indio (09°12'04.1" N,

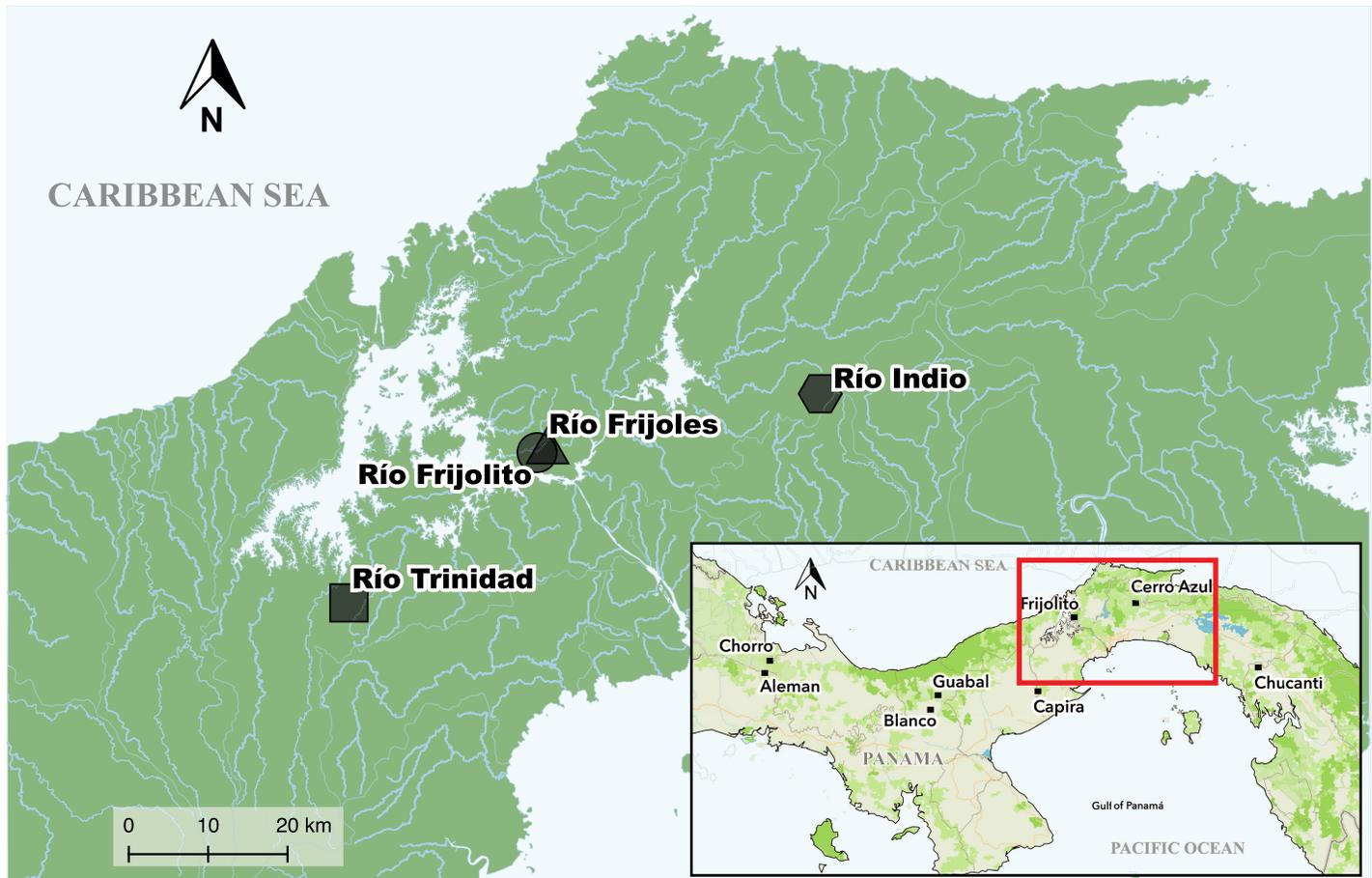


Fig 1. Sampling sites of macroinvertebrates in the Panama Canal Watershed. Names on the inset map indicate the sites previously sampled by Múrria et al. [25].

<https://doi.org/10.1371/journal.pone.0231683.g001>

079°24'20.4" W), located 35 Km east of Frijoles, is intermediately disturbed and is surrounded by secondary forest with dense riparian vegetation. At each site, we haphazardly collected aquatic macroinvertebrates using standard kick-netting from the two dominant habitats types (riffles and pools). The sampling effort was approximately two hours at each site. All samples were sorted in the field and immediately preserved in 95% ethanol. Sampling permit was obtained from the Autoridad Nacional del Ambiente de Panamá (Permit No. SC/A-44-12).

In the laboratory, specimens were morphologically identified to the lowest possible taxonomic level (i.e., family or genus) using taxonomic keys for Neotropical macroinvertebrates [6,8,9,39]. However, given the low accuracy of morphological identification, and the fact that less than 50% of the individuals were successfully identified to species level using our barcoding data (see [results](#)), we focused our analyses and discussion on lineage rather than morphological diversity. Representative specimens have been deposited in the invertebrate collection at Colección Zoológica Dr. Eustorgio Méndez (CoZEM) at Instituto Conmemorativo Gorgas de Estudio de la Salud in Panama City (Voucher numbers: B001 –TR020).

DNA sequencing

Tissue samples were obtained from the hind leg or part of the body of each specimen, and total DNA was extracted by using the DNeasy Blood & Tissue kit (Qiagen, CA, USA), according to

the manufacturer's instructions. A standard sequencing protocol [40] was used to amplify the full-length 658 base pair (bp) of the COI barcode region using the following primers sets: LCO1490/HCO2198 [41] and LepF1/LepR1 [42]. All PCR products were verified on a 1% agarose gel, and purified with EXO-SAP-IT (USB Corp., Cleveland, Ohio, U.S.A.). The protocol included adding 1 μ l of 1 U/ μ l Shrimp Alkaline Phosphatase (SAP), 0.5 μ l of 20 U/ μ l Exonuclease I (EXO), and 5 μ l of amplified product, and then incubating as indicated in the manufacturer's protocol [43]. This product was sequenced using an Applied Biosystems Genetic Analyzer (ABI 3130xl, Applied Biosystems, Carlsbad, California). Sequences were aligned in Geneious V7.03 [44], using the MAFFT 7.313 [45] tool and the L-INS-i algorithm. Sequence alignments were also inspected by eye in Geneious to confirm overall sequence quality. We did not find gaps or stop codons in any of the sequences. Project sequences, together with the information on collected specimens, are available on the Barcode of Life Data System (BOLD systems, under project code INVPA: <http://www.boldsystems.org/>). Project sequences are also available in GenBank (accession numbers: KX039451-KX039650, KU980966-KU981004).

Data analysis

To confirm morphological identification for our sequenced specimens, we performed BLAST searches for publicly available sequences in GenBank. We created a final dataset comprising unique haplotypes from our study and including three COI sequences retrieved from GenBank that were used as an outgroup. The retrieved sequences were *Thermobia domestica* (GenBank NC006080), *Atelura formicaria* (GenBank NC01119), and *Tricholepidion gertschi* (GenBank NC005437).

We estimated phylogenetic relationships among taxa using maximum likelihood (ML) searches in IQ-TREE v 1.6 [46] and Bayesian inference (BI) in BEAST v 2.4.6 [47] as implemented on the CIPRES Science Gateway [48]. The best-fit model of nucleotide substitution for the dataset, selected using jModelTest 2.0 [49] based on the Bayesian Information Criterion, was GTR+I+G. To determine node support for the IQ-TREE we used 10000 ultrafast bootstraps [50] and 1000 Shimodaira-Hasegawa-like approximate likelihood ratio test replicates [51]. BI analysis was executed with an uncorrelated lognormal relaxed clock and coalescent prior, with the default settings of BEAUti for the remaining parameters. We performed two runs of 2.0×10^7 generations, and sampled trees every 5000 generations. Trace logs and species trees for the two runs were combined using LogCombiner v 2.4.8 [52]. We used Tracer v. 1.6 [53] to ensure that effective sample size (ESS) values for all parameters were above 200 and to determine the burn-in. Finally, output trees were summarized as maximum clade credibility (MCC) trees using mean node heights after discarding 25% of generations as burn-in using TreeAnnotator v1.8.4 [54].

We then assessed the diversity of "molecular species" by estimating molecular operational units (MOTUs; [55] using the software MEGA 7.0 [56] and the BOLD analyses tools [57]. Sequence divergence was estimated using the Kimura-2-Parameter (K2P) model with 1000 bootstrap estimates in MEGA7. This a standard model that has been extensively used in barcoding studies [58,59]. The Barcode Index Number (BIN) system [57] was used as delimitation criterion for the assignment of MOTUs across the full dataset. This method uses a 2.2% in sequence divergence cut-off, but updates this value according to the distribution of divergence among sequences in the dataset. Recent studies suggest that using a single divergence cut-off may not be appropriate for every organism (reviewed in [60], such as in the case of diverse non-tropical chironomids [60,61]. But, given that our sample of this taxon was small, we assumed a single cut-off value for the entire dataset. However, future work should evaluate the optimal level of genetic divergence to delimit biological species in Neotropical macroinvertebrates.

In addition, given recent concerns with the use of the K2P model for species delimitation (e.g., [62]), we applied two additional single-locus analyses to confirm lineage diversity: the Bayesian General Mixed Yule Coalescent model (GMYC; [63]) and the Automatic Barcode Gap Discovery (ABGD; [64]). The GMYC approach uses branch lengths to determine the transition from intraspecific to interspecific relationships [63]. The ABGD algorithm allows to sort DNA sequences into “hypothetical species” based on the gaps in the distribution of intra- and inter-specific genetic divergence in a given sample [64]. Although the two approaches differ in their properties (i.e., tree branch length vs. distribution gaps), we used them to confirm the patterns of species delimitation.

To perform GMYC tree-based analyses, we used the ultrametric trees previously generated with BEAST. GMYC was performed using the single threshold parameter at the GMYC web-server (<https://species.h-its.org/gmyc/>). ABGD was carried out using the online version of ABGD software [64]; (<https://bioinfo.mnhn.fr/abi/public/abgd/abgdweb.html>). Default settings were used, however, distance matrices based on K2P distance calculated in MEGA7 were used as input. All analyses were run using a relative barcoding gap width (X value) set to 1.0. Only the recursive results were used because they allowed for different gap thresholds among taxa.

To compare patterns of spatial variation in genetic diversity (MOTUs), we quantified the number of shared species among sampling sites. We also estimated Fisher’s alpha index of diversity and Whittaker’s measure of β -beta diversity. Given that these analyses might be affected by variation in sampling size, we also used rarefaction and extrapolation methods [65,66] as implemented in the R package iNEXT [67]. This method allows for comparisons between sites while controlling for differences in abundance and sampling effort. For these analyses, we fit curves for the first three Hill numbers: species richness ($q = 0$), the exponential of Shannon entropy (“Shannon diversity”, $q = 1$), and the inverse Simpson concentration (“Simpson diversity”, $q = 2$), using individual-based abundance data. Given the limitation in sample size, we did attempt to make statistical inferences regarding differences in diversity among sites.

Finally, to assess lineage diversity and the diversity of unique haplotypes across watersheds spanning the Isthmus of Panama, we collected haplotype information (426 haplotypes: GenBank accession numbers KR134410-KR134835) from a previous study in this region [25]. After adding these sequences to our dataset, we generated multiple sequence alignments with MAFFT 7.313 [45] using the L-INS-i algorithm in Geneious V7.03. Then, we trimmed the sequences to the same fragment size and compared the previously reported haplotypes with the ones encountered in our dataset. To exclude redundancies prior to phylogenetic analyses, we applied DAMBE v. 6.4.11 [68] to identify and remove duplicate haplotypes from our dataset. In total, we found a total of 15 duplicate haplotypes between the two datasets. The combined dataset contained 12 sites (8 from Múrria et al. [25], and 4 from the present study). One site (Frijolito) was sampled during both studies, but we analyzed them separately to preserve independence between the two studies. We then applied the same rarefaction/extrapolation approach described above to generate rarefaction curves as a function of the number of individuals sampled. Although the collection method was similar (i.e., both studies used kick-nets during a given amount of time), the two studies may not be directly comparable due to differences in sampling effort and the overall study objective. Therefore, we did not attempt to make statistical inferences about the relative abundance of macroinvertebrates among sites or between studies. Instead, we only focused on compiling the total number of unique haplotypes or molecular species that are currently known within each site in this region. In addition, our ultimate goal was to explore the overall molecular endemism of aquatic macroinvertebrates, rather than providing precise estimates of species diversity in this region. Thus, substantially more research that includes more data, sites, and replication is needed to confirm if current patterns hold across the entire region.

Results

We collected approximately 300 specimens across the four sites; however, our analysis focused on the 224 individuals that were successfully barcoded (Table 1 and S1 Table). We were able to identify nearly 70% of individuals to genus level using the morphological approach, but species-level identification was only possible for 56 individuals (25%). Some of the most numerous taxa across sites included Leptophlebiidae (11.2% of individuals), Libellulidae (11.2%), Naucoridae (6.7%), Notonectidae (6.3%), Chironomidae (5.4%), Gerridae (4.9%), Hydropsychidae (4.0%), Perlidae (3.6%), and Baetidae (3.6%).

The 224 COI sequences revealed a total of 168 haplotypes (S1 Table). After comparing these haplotypes with molecular data from Múrria et al [25], we found 153 (~91%) unique haplotypes from the Panama Canal Watershed. In Frijolito, the site sampled by the two studies, we found a total of 46 haplotypes, 45 of which were unique to our study.

Our final COI dataset consisted of 171 terminals, including the 168 new barcode sequences, and 3 outgroup sequences retrieved from GenBank. The final aligned and pruned dataset contained 620 aligned positions, including gaps, with 371 variable sites, of which 359 were parsimony-informative (~96% of variable positions). As expected, we observed a hierarchical increase in the mean K2P genetic divergence with increasing taxonomic levels from within a species 0.38% (SE = 0.002), to within family 9.92% (SE = 0.01), to within order 19.75% (SE = 0.01). However, we were not able to identify our specimens to species level from our BLAST search, given that only around 50% of our sequences matched existing data in the public databases, and most of these matches corresponded to genus and family level only. Both ML and BI inference trees for all specimens showed well-defined clades at the level of order and family, with some differences in the topology, but overall support was higher for the BI tree, which we used to represent the number of molecular species (S1 Fig and S2 Fig).

Our species delimitation analyses yielded variable, but relatively high numbers of species. Specifically, GMYC detected 106 MOTUs (95% confidence intervals: 104–109), whereas ABGD found a total of 100 MOTUs (S1 Fig). These ABGD results were confirmed independently of the chosen model (Jukes-Cantor and Kimura) and were unaffected by changes of prior limits for intraspecific variation and threshold.

When looking at spatial patterns of diversity, we observed some overlap in the number of shared MOTUs as well as a considerable number of unique haplotypes in each river: Frijoles (65), Frijolito (43), Trinidad (31) and Indio (22) (Fig 2). This pattern was also supported by the Fisher's alpha diversity index, which showed variation in molecular species among sites: Frijoles 53.40, Frijolito 22.80, Trinidad 17.01, and Indio 67.63. Whittaker's index of β diversity also showed high species turn-over across sites (0.87). Similarly, our rarefaction/extrapolation analyses showed variation in species richness across sites: Frijoles (52), Frijolito (27), Trinidad (22) and Indio (22). However, the most striking pattern was a lack of saturation in the

Table 1. Macroinvertebrate lineage diversity in the Panama Canal Watershed.

Site	Ind.	Hap.	GMYC	ABGD
Trinidad (TRI)	45	33	22	21
Frijolito (FTO)	65	46	27	26
Frijoles (FES)	88	72	52	50
Indio (IND)	26	24	22	21
All	224	168	106	100

The data show the number of successfully barcoded individuals (Ind.), cox1 haplotypes (Hap.), and molecular species (based on GMYC and ABGD) collected at four sites.

<https://doi.org/10.1371/journal.pone.0231683.t001>

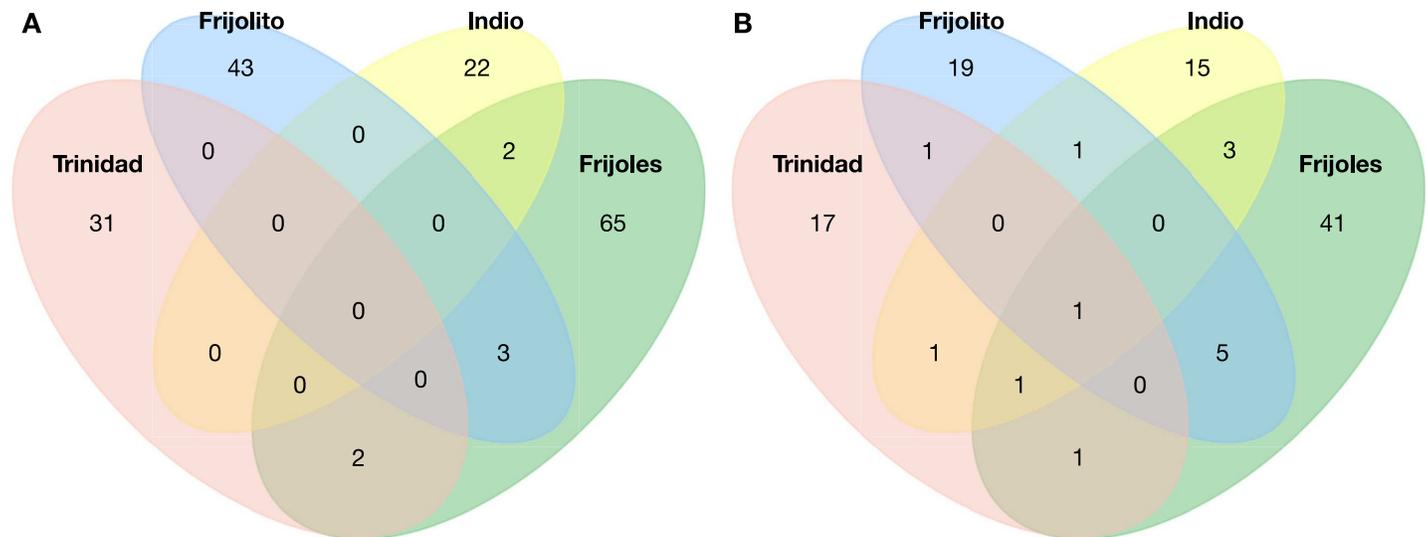


Fig 2. Distribution of molecular diversity in freshwater macroinvertebrates among streams within Panama Canal Watershed. Venn diagrams show the number of shared and unique haplotypes (A) and MOTUs (B) across four streams: Trinidad (pink), Frijolito (blue), Indio (yellow), and Frijoles (green).

<https://doi.org/10.1371/journal.pone.0231683.g002>

accumulation curves for most of the sites (Fig 3), and this pattern was consistent across the first three Hill numbers (S3 Fig). Similar results were found when looking at diversity Hill across the Isthmus of Panama using the compiled barcoding dataset (Fig 3). In particular, we observed substantial diversity of both MOTUs and unique haplotypes across sites, but the accumulation curves did not reach saturation for most of the sites (Fig 3). In addition, both molecular species and haplotype diversity tended to increase at sites within the PCW, in contrast to sites located in the eastern and western portion of the country (Fig 3). Data on assignment and diversity of MOTUs across study sites are available in the supplementary material (S1 Table).

Discussion

Using DNA-barcoding, we examined the diversity of freshwater macroinvertebrates at a small spatial scale, among four streams within the Panama Canal watershed (PCW). We also compiled existing barcoding data [25] to contrast macroinvertebrate diversity at a broader spatial scale, across eight streams along the Isthmus of Panama. Overall, we found high lineage diversity across sites within the PCW (Table 1; S1 Fig), and a large portion of these lineages appear to be unique to each site (Fig 2). In addition, our rarefaction/extrapolation approach showed that this diversity is still under-sampled across sites both within the PCW and along the Isthmus of Panama (Fig 3).

These findings confirm that the diversity of freshwater macroinvertebrates in Neotropical environments is largely under-studied [33,69,70], and could be much higher than previously thought. In particular, these findings highlight the fact that there is limited published research using genetic methods to study macroinvertebrate diversity in this region. This was reflected by one of our sites (Frijolito), which despite being sampled by a previous barcoding study [25], still showed a substantial number of novel haplotypes. In addition, the fact that only a small number of our specimens matched available sequences in public databases further highlights the potential for biodiversity discovery in Neotropical freshwater environments. This seems particularly relevant for taxa such as Hydropsychidae, Gerridae, Chironomidae, Leptophlebiidae, Libellulidae and Notonectidae, which showed high lineage/haplotype diversity across sites

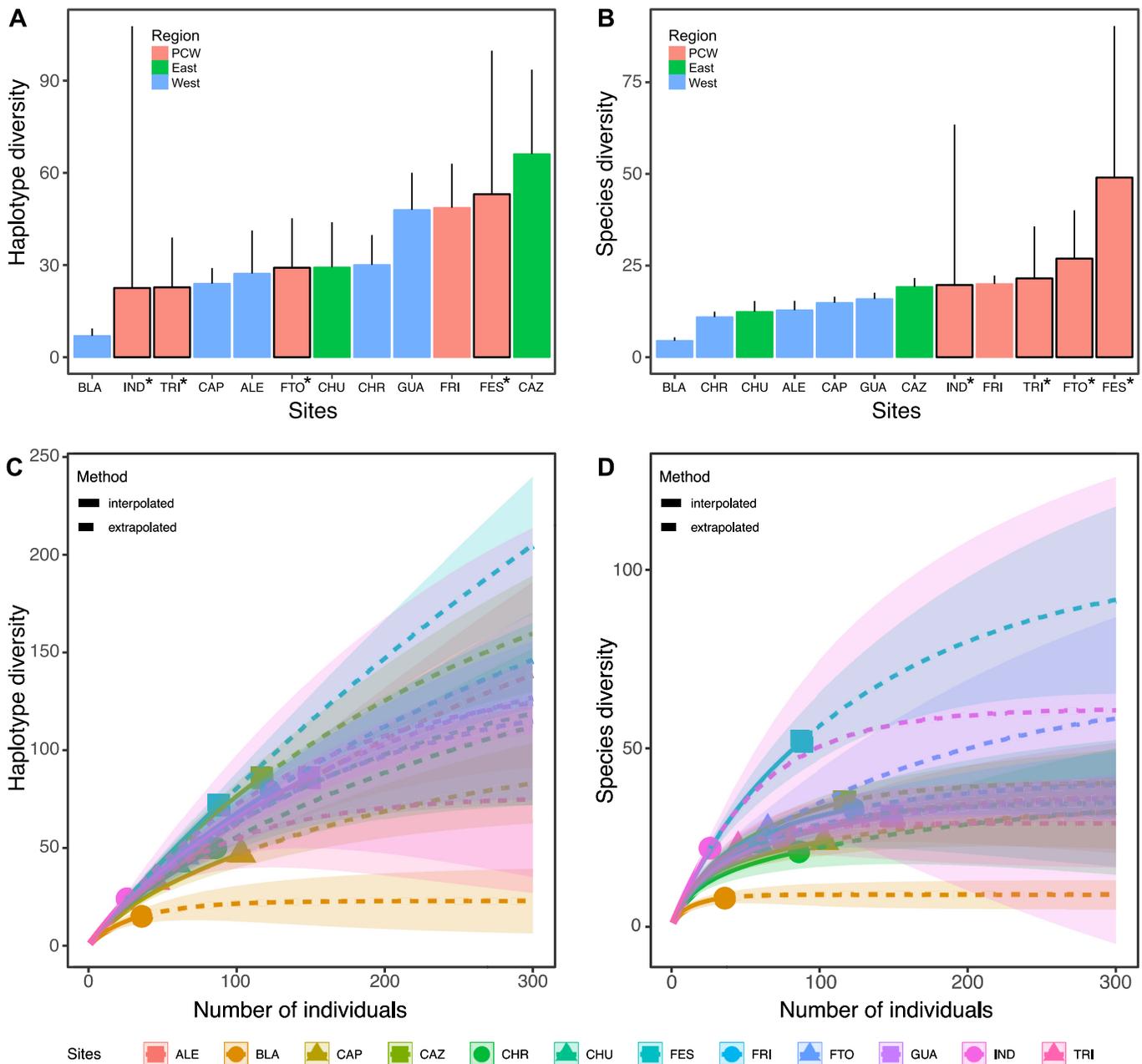


Fig 3. Molecular diversity in freshwater macroinvertebrates along the Isthmus of Panama. Panels show Simpson’s diversity index for both haplotype (A) and MOTUs (B), as well as rarefaction and extrapolation curves for both haplotypes (C) and MOTUs (D) diversity at each site. Sites are Alemán (ALE), Chorro (CHO), Blanco (BLA), Guabal (GUA), Capira (CAP), Frijolito (FRI), Cerro Azul (CAZ), and Chucantí (CHR) from Múrria et al. 2015; and Trinidad (TRI), Frijolito (FTO), Frijoles (FES) and Indio (IND) from present study (indicated with asterisks). Regions are indicated as East, West and the Panama Canal Watershed (PCW).

<https://doi.org/10.1371/journal.pone.0231683.g003>

(S1 Fig; S1 Table). Some of these taxa also showed high haplotype diversity in a previous molecular study across Panama [25], and are thought to hold a high number of undescribed species in the Central American Isthmus [6,69]. Unfortunately, our analysis is limited by relatively small sample size, particularly at one of our sites (Río Indio), where we were only able to

sequence 26 specimens. In addition, the fact that our rarefaction/extrapolation analysis showed a lack of saturation for most of the sites indicates that substantially more research is needed in this region. Overall, however, our results are in line with recent work showing high haplotype endemism among isolated watersheds across the Isthmus of Panama [25]. In fact, despite low sample size, we found at least 153 (~91%) novel haplotypes within the PCW. Thus, we expand on this previous work by highlighting the possibility that endemism of Neotropical macroinvertebrates can be substantial even within a single watershed.

Typically, diversification in freshwater organisms is marked by a strong geographic signature, where genetic divergence is facilitated by spatial isolation among populations [26,27]. However, the contribution of geographic isolation to the diversification of Neotropical freshwater macroinvertebrates has received little attention to date [25]. In addition, the fact that most macroinvertebrates are semiaquatic, and are likely to disperse during the adult stages [15,16] could limit genetic isolation among nearby stream communities. Yet, the possibility of high endemism even within a single watershed suggests that spatial isolation, habitat variability, and perhaps, local adaptation are important drivers of macroinvertebrate diversity. Another interesting finding was that macroinvertebrate diversity appeared to increase at sites located in Central Panama, specifically within the PCW (e.g., Frijoles, Frijolito, Trinidad). However, additional research is needed to confirm this pattern and to explore the underlying drivers. We encourage the application of more efficient tools such as DNA metabarcoding to facilitate this endeavor.

Our finding of high endemism at a small geographic scale is also relevant in the face of increasing anthropogenic disturbances [31,32,71–73]. Specifically, it suggests that small-scale local disturbances could have drastic consequences for the maintenance of a unique freshwater biodiversity—but this diversity is still unknown. We therefore predict that the current rate of species loss in freshwater ecosystems might be surpassing the rate of species discovery in Neotropical environments. Overall, however, further work is clearly needed to disentangle the contribution of other factors such as genetic drift, local adaptation, and environmental disturbance to persistence and diversification of Neotropical freshwater macroinvertebrates.

Taken together, our results confirm the expectation that the diversity of Neotropical macroinvertebrates remains under-studied. They also indicate that uncovering this hidden diversity is crucial to our understanding of the local and regional processes that shape biodiversity in Neotropical freshwater environments.

Supporting information

S1 Fig. Molecular diversity in freshwater macroinvertebrates from Central Panama. The Bayesian inference tree shows species delimitation analyses based on generalized mixed Yule coalescent (GMYC) and the automatic barcode gap discovery (ABGD). Black and grey blocks represent putative molecular species, with taxa sharing the same block corresponding to similar species. The numbers next to the nodes represent Bayesian posterior probability values. (PDF)

S2 Fig. Rarefaction and extrapolation curves for molecular diversity (MOTUs) at each site. Number at the top represent fit curves for the first three Hill numbers: species richness ($q = 0$), the exponential of Shannon entropy (“Shannon diversity”, $q = 1$), and the inverse Simpson concentration (“Simpson diversity”, $q = 2$), using individual-based abundance data. Sites are: Alemán (ALE), Chorro (CHO), Blanco (BLA), Guabal (GUA), Capira (CAP), Frijolito (FRI), Cerro Azul (CAZ), and Chucantí (CHR) from Múrria et al. 2015; and Trinidad (TRI), Frijolito (FTO), Frijoles (FES) and Indio (IND) (from the present study). (PDF)

S3 Fig. Phylogenetic tree determined by the Maximum Likelihood (ML). Data represent *cox1* sequences obtained from 224 freshwater macroinvertebrates collected within the Panama Canal Watershed. The numbers on the branches show nodal support. (PNG)

S1 Table. Molecular and taxonomic diversity of freshwater macroinvertebrates within the Panama Canal Watershed. For each specimen, we show a taxonomic group (i.e., order, family, and genus), molecular species identity (MOTUs: based on ABGD and GMYC), Haplotype identity, Genbank accession number, and sampling site. Sites are Trinidad (TRI), Frijolito (FTO), Frijoles (FES) and Indio (IND). (XLSX)

Acknowledgments

We dedicate this study to our friend Ruth G. Reina. Her passion for tropical freshwater biology served as an inspiration to this work. Logistical support was provided by the Smithsonian Tropical Research Institute. Field assistance was provided by Celestino Martínez, Nohelys Alvarado, Carlos Nieto, and Débora Delgado. Diana Sharpe provided valuable comments on an earlier version of the manuscript. We thank editor Dr. Michael A. Chadwick and four anonymous reviewers for providing feedback on our manuscript.

Author Contributions

Conceptualization: Luis F. De León.

Data curation: Luis F. De León, Aydeé Cornejo, Ronnie G. Gavilán, Celestino Aguilar.

Formal analysis: Luis F. De León, Ronnie G. Gavilán, Celestino Aguilar.

Funding acquisition: Luis F. De León.

Investigation: Luis F. De León, Aydeé Cornejo, Ronnie G. Gavilán, Celestino Aguilar.

Methodology: Luis F. De León, Aydeé Cornejo, Ronnie G. Gavilán.

Resources: Aydeé Cornejo, Ronnie G. Gavilán.

Supervision: Luis F. De León, Aydeé Cornejo.

Writing – original draft: Luis F. De León.

Writing – review & editing: Luis F. De León, Celestino Aguilar.

References

1. Wallace JB, Webster JR. The role of macroinvertebrates in stream ecosystem function. *Annu Rev Entomol.* 1996; 41: 115–139. <https://doi.org/10.1146/annurev.en.41.010196.000555> PMID: 15012327
2. Covich AP, Palmer MA, Crowl TA. The Role of Benthic Invertebrate Species in Freshwater Ecosystems: Zoobenthic species influence energy flows and nutrient cycling. *Bioscience.* 1999; 49: 119–127.
3. Covich AP, Austen MC, BÄRlocher F, Chauvet E, Cardinale BJ, Biles CL, et al. The Role of Biodiversity in the Functioning of Freshwater and Marine Benthic Ecosystems. *Bioscience.* 2004; 54: 767–775.
4. Al-Shami SA, Rawi CSM, Ahmad AH, Madrus MR, AL Mutairi K. Importance of regional diversity and environmental conditions on local species richness of aquatic macro-invertebrates in tropical forested streams. *J Trop Ecol.* 2014; 30: 335–346.
5. Pérez GR. Guía para el estudio de los macroinvertebrados acuáticos del Departamento de Antioquia. Fondo para la Protección del Medio Ambiente “José Celestino Mutis”; 1988.

6. Ramírez A, Paulson DR, Esquivel C. Odonata of Costa Rica: Diversity and checklist of species. *Revista de Biología Tropical*. 2000; 48: 247–254.
7. Posada-García JA, Roldán-Pérez G. Clave ilustrada y diversidad de las larvas de trichoptera en el noroccidente de Colombia / Illustrated key for the larvae of Trichoptera in the northwest of Colombia. *Caldasia*. 2003; 25: 169–192.
8. Springer M. Clave taxonómica para larvas de las familias del orden Trichoptera (Insecta) de Costa Rica. 1. 2006; 273–286.
9. Flowers RW. Diversity of stream-living insects in northwestern Panama. *J North Am Benthol Soc*. 1991; 10: 322–334.
10. Quintero D, Aiello A. *Insects of Panama and Mesoamerica: Selected Studies*. Oxford University Press; 1992.
11. Melo AS, Froehlich CG. Macroinvertebrates in neotropical streams: richness patterns along a catchment and assemblage structure between 2 seasons. *Journal of the North American Benthological Society*. 2001. pp. 1–16. <https://doi.org/10.2307/1468184>
12. Boyero L, Bailey RC. Organization of macroinvertebrate communities at a hierarchy of spatial scales in a tropical stream. *Hydrobiologia*. 2001; 464: 219–225.
13. Holzenthal RW, Blahnik RJ, Prather AL, Kjer KM. Order Trichoptera Kirby, 1813 (Insecta), Caddisflies. 2007. Available: <https://conservancy.umn.edu/handle/11299/196322>
14. Ligeiro R, Melo AS, Callisto M. Spatial scale and the diversity of macroinvertebrates in a Neotropical catchment. *Freshw Biol*. 2010; 55: 424–435.
15. Merritt RW, Cummins KW. *An introduction to the aquatic insects of North America*. Kendall Hunt; 1996.
16. Jacobsen D, Cressa C, Mathooko JM, Dudgeon D. 4—Macroinvertebrates: Composition, Life Histories and Production. In: Dudgeon D, editor. *Tropical Stream Ecology*. London: Academic Press; 2008. pp. 65–105.
17. Fischer AG. Latitudinal variations in organic diversity. *Evolution*. 1960; 14: 64–81.
18. Rosenzweig ML. *Species Diversity in Space and Time*. Cambridge University Press; 1995.
19. Jacobsen D, Schultz R, Encalada A. Structure and diversity of stream invertebrate assemblages: the influence of temperature with altitude and latitude. *Freshw Biol*. 1997; 38: 247–261.
20. Boyero L. Short Communication: Insect biodiversity in freshwater ecosystems: is there any latitudinal gradient? *Mar Freshwater Res*. 2002; 53: 753–755.
21. Wantzen KM, Ramírez A, Winemiller KO. New vistas in Neotropical stream ecology—Preface. *J North Am Benthol Soc*. 2006; 25: 61–65.
22. Boyero L, Ramírez A, Dudgeon D, Pearson RG. Are tropical streams really different? *J North Am Benthol Soc*. 2009; 28: 397–403.
23. Mayhew PJ. Why are there so many insect species? Perspectives from fossils and phylogenies. *Biol Rev Camb Philos Soc*. 2007; 82: 425–454. <https://doi.org/10.1111/j.1469-185X.2007.00018.x> PMID: 17624962
24. Dijkstra K-DB, Monaghan MT, Pauls SU. Freshwater biodiversity and aquatic insect diversification. *Annu Rev Entomol*. 2014; 59: 143–163. <https://doi.org/10.1146/annurev-ento-011613-161958> PMID: 24160433
25. Múrria C, Rugenski AT, Whiles MR, Vogler AP. Long-term isolation and endemism of Neotropical aquatic insects limit the community responses to recent amphibian decline. *Diversity and Distributions*. 2015. pp. 938–949. <https://doi.org/10.1111/ddi.12343>
26. Smith SA, Bermingham E. The biogeography of lower Mesoamerican freshwater fishes. *J Biogeogr*. 2005; 32: 1835–1854.
27. Aguilar C, Miller MJ, Loaiza JR, González R, Krahe R, De León LF. Tempo and mode of allopatric divergence in the weakly electric fish *Sternopygus dariensis* in the Isthmus of Panama. *Sci Rep*. 2019; 9: 18828. <https://doi.org/10.1038/s41598-019-55336-y> PMID: 31827183
28. Monaghan MT, Wild R, Elliot M, Fujisawa T, Balke M, Inward DJG, et al. Accelerated species inventory on Madagascar using coalescent-based models of species delineation. *Syst Biol*. 2009; 58: 298–311. <https://doi.org/10.1093/sysbio/syp027> PMID: 20525585
29. García-López A, Micó E, Múrria C, Galante E, Vogler AP. Beta diversity at multiple hierarchical levels: explaining the high diversity of scarab beetles in tropical montane forests. *J Biogeogr*. 2013; 40: 2134–2145.
30. Dudgeon D, Arthington AH, Gessner MO, Kawabata Z-I, Knowler DJ, Lévêque C, et al. Freshwater biodiversity: importance, threats, status and conservation challenges. *Biol Rev Camb Philos Soc*. 2006; 81: 163–182. <https://doi.org/10.1017/S1464793105006950> PMID: 16336747

31. Boyero L, Pearson RG, Dudgeon D, Ferreira V, Graça MAS, Gessner MO, et al. Global patterns of stream detritivore distribution: implications for biodiversity loss in changing climates. *Glob Ecol Biogeogr.* 2012; 21: 134–141.
32. Collen B, Whitton F, Dyer EE, Baillie JEM, Cumberlidge N, Darwall WRT, et al. Global patterns of freshwater species diversity, threat and endemism. *Glob Ecol Biogeogr.* 2014; 23: 40–51. <https://doi.org/10.1111/geb.12096> PMID: 26430385
33. Cornejo A, Tonin AM, Checa B, Tuñon AR, Pérez D, Coronado E, et al. Effects of multiple stressors associated with agriculture on stream macroinvertebrate communities in a tropical catchment. *PLoS One.* 2019; 14: e0220528. <https://doi.org/10.1371/journal.pone.0220528> PMID: 31393898
34. Martínez C, Chavarria C, Sharpe DMT, De León LF. Low Predictability of Colour Polymorphism in Introduced Guppy (*Poecilia reticulata*) Populations in Panama. *PLOS ONE.* 2016. p. e0148040. <https://doi.org/10.1371/journal.pone.0148040> PMID: 26863538
35. Sharpe DMT, De León LF, González R, Torchin ME. Tropical fish community does not recover 45 years after predator introduction. *Ecology.* 2017; 98: 412–424. <https://doi.org/10.1002/ecy.1648> PMID: 27861787
36. De León LF, Lopez OR. Biodiversity beyond trees: Panama's Canal provides limited conservation lessons for Nicaragua. *Biodivers Conserv.* 2016; 25: 2821–2825.
37. Ricciardi A, Rasmussen JB. Extinction Rates of North American Freshwater Fauna. *Conserv Biol.* 1999; 13: 1220–1222.
38. Castillo AM, Sharpe DMT, Ghalambor CK, De León LF. Exploring the effects of salinization on trophic diversity in freshwater ecosystems: a quantitative review. *Hydrobiologia.* 2018; 807: 1–17.
39. Roldán-Pérez G. Guía para el estudio de los macroinvertebrados acuáticos del Departamento de Antioquia. Bogotá, Universidad de Antioquia, 217p. 1988.
40. Ivanova Natalia V, Dewaard Jeremy R, Hebert Paul DN. An inexpensive, automation-friendly protocol for recovering high-quality DNA. 2006. <https://doi.org/10.1111/j.1471-8286.2006.01428.x>
41. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol.* 1994; 3: 294–299. PMID: 7881515
42. Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astrartes fulgerator*. *Proc Natl Acad Sci U S A.* 2004; 101: 14812–14817. <https://doi.org/10.1073/pnas.0406166101> PMID: 15465915
43. Werle E, Schneider C, Renner M, Völker M, Fiehn W. Convenient single-step, one tube purification of PCR products for direct sequencing. *Nucleic Acids Res.* 1994; 22: 4354–4355. <https://doi.org/10.1093/nar/22.20.4354> PMID: 7937169
44. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* 2012. pp. 1647–1649. <https://doi.org/10.1093/bioinformatics/bts199> PMID: 22543367
45. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013; 30: 772–780. <https://doi.org/10.1093/molbev/mst010> PMID: 23329690
46. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015; 32: 268–274. <https://doi.org/10.1093/molbev/msu300> PMID: 25371430
47. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, et al. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol.* 2014; 10: e1003537. <https://doi.org/10.1371/journal.pcbi.1003537> PMID: 24722319
48. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. 2010 Gateway Computing Environments Workshop (GCE). IEEE; pp. 1–8.
49. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 2012; 9:772–772.
50. Minh BQ, Nguyen MAT, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol.* 2013; 30: 1188–1195. <https://doi.org/10.1093/molbev/mst024> PMID: 23418397
51. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 2010; 59: 307–321. <https://doi.org/10.1093/sysbio/syq010> PMID: 20525638
52. Rambaut A, Drummond AJ. LogCombiner v1. 8.2. 2013.
53. Rambaut A, Suchard MA, Xie D, Drummond AJ. Tracer v1. 6. 2014. 2015.

54. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol.* 2012; 29: 1969–1973. <https://doi.org/10.1093/molbev/mss075> PMID: 22367748
55. Jones M, Ghoorah A, Blaxter M. jMOTU and Taxonator: Turning DNA Barcode Sequences into Annotated Operational Taxonomic Units. *PLoS ONE.* 2011. p. e19259. <https://doi.org/10.1371/journal.pone.0019259> PMID: 21541350
56. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol.* 2016; 33: 1870–1874. <https://doi.org/10.1093/molbev/msw054> PMID: 27004904
57. Ratnasingham S, Hebert PDN. A DNA-based registry for all animal species: the barcode index number (BIN) system. *PLoS One.* 2013; 8: e66213. <https://doi.org/10.1371/journal.pone.0066213> PMID: 23861743
58. Amouroux P, Crochard D, Germain J-F, Correa M, Ampuero J, Groussier G, et al. Genetic diversity of armored scales (Hemiptera: Diaspididae) and soft scales (Hemiptera: Coccidae) in Chile. *Sci Rep.* 2017; 7: 2014. <https://doi.org/10.1038/s41598-017-01997-6> PMID: 28515435
59. Havemann N, Gossner MM, Hendrich L, Morinière J, Niedringhaus R, Schäfer P, et al. From water striders to water bugs: the molecular diversity of aquatic Heteroptera (Gerromorpha, Nepomorpha) of Germany based on DNA barcodes. *PeerJ.* 2018; 6: e4577. <https://doi.org/10.7717/peerj.4577> PMID: 29736329
60. Lin X-L, Stur E, Ekrem T. DNA barcodes and morphology reveal unrecognized species in Chironomidae (Diptera). *Insect Syst Evol.* 2018; 49: 329–398.
61. Beermann AJ, Zizka VMA, Elbrecht V, Baranov V, Leese F. DNA metabarcoding reveals the complex and hidden responses of chironomids to multiple stressors. *Environmental Sciences Europe.* 2018; 30: 26.
62. Barley AJ, Thomson RC. Assessing the performance of DNA barcoding using posterior predictive simulations. *Mol Ecol.* 2016; 25: 1944–1957. <https://doi.org/10.1111/mec.13590> PMID: 26915049
63. Fujisawa T, Barraclough TG. Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent approach: a revised method and evaluation on simulated data sets. *Syst Biol.* 2013; 62: 707–724. <https://doi.org/10.1093/sysbio/syt033> PMID: 23681854
64. Puillandre N, Lambert A, Brouillet S, Achaz G. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Mol Ecol.* 2012; 21: 1864–1877. <https://doi.org/10.1111/j.1365-294X.2011.05239.x> PMID: 21883587
65. Colwell RK, Chao A, Gotelli NJ, Lin S-Y, Mao CX, Chazdon RL, et al. Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. *J Plant Ecol.* 2012; 5: 3–21.
66. Chao A, Gotelli NJ, Hsieh TC, Sander EL, Ma KH, Colwell RK, et al. Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecological Monographs.* 2014. pp. 45–67. <https://doi.org/10.1890/13-0133.1>
67. Hsieh TC, Ma KH, Chao A. iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods in Ecology and Evolution.* 2016. pp. 1451–1456. <https://doi.org/10.1111/2041-210x.12613>
68. Xia X. DAMBE6: New Tools for Microbial Genomics, Phylogenetics, and Molecular Evolution. *J Hered.* 2017; 108: 431–437. <https://doi.org/10.1093/jhered/esx033> PMID: 28379490
69. Springer M. Aquatic insect diversity of Costa Rica: state of knowledge. *Revista de Biología Tropical.* 2008; 56: 273–295.
70. Righi-Cavallaro KO, Roche KF, Froehlich O, Cavallaro MR. Structure of macroinvertebrate communities in riffles of a Neotropical karst stream in the wet and dry seasons. *Acta Limnol Brasil.* 2010; 22: 306–316.
71. Malmqvist B, Rundle S. Threats to the running water ecosystems of the world. *Environ Conserv.* 2002; 29: 134–153.
72. Bojsen BH, Jacobsen D. Effects of deforestation on macroinvertebrate diversity and assemblage structure in Ecuadorian Amazon streams. *Archiv für Hydrobiologie.* 2003; 158: 317–342.
73. Streebman JT, Gmyrek SL, Kidd MR, Kidd C, Robinson RL, Hert E, et al. Hybridization and contemporary evolution in an introduced cichlid fish from Lake Malawi National Park. *Mol Ecol.* 2004; 13: 2471–2479. <https://doi.org/10.1111/j.1365-294X.2004.02240.x> PMID: 15245419