Published Manuscript

Cite this article:

Calderón-Gutiérrez F, Labonté JM, Gonzalez BC, Iliffe TM, Mejía-Ortíz LM, BordaE. 2024 Cryptic diversity patterns of subterranean estuaries. Proc. R. Soc. B 291: 20241483.

https://doi.org/10.1098/rspb.2024.1483

Cryptic diversity patterns of subterranean estuaries

Running title: Cryptic diversity of subterranean estuaries

Fernando Calderón-Gutiérrez[1*] (orcid.org/0000-0002-3181-4280)

Jessica M. Labonté[2] (orcid.org/0000-0003-0236-7634)

Brett C. Gonzalez[3,4] (orcid.org/0000-0001-6968-2677)

Thomas M. Iliffe[5] (orcid.org/0000-0002-4342-5960)

Luis M. Mejía-Ortíz[6] (orcid.org/0000-0003-0858-0751)

Elizabeth Borda[1] (orcid.org/0000-0003-1179-163X)

[1]Department of Natural Sciences, Texas A&M University San Antonio, San Antonio, Texas, USA

[2]Department of Marine Biology, Texas A&M University at Galveston, Galveston, Texas, USA

[3]Department of Invertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington DC, USA

[4]Minderoo-UWA Deep-Sea Research Centre, School of Biological Sciences and Oceans Institute, University of Western Australia, Perth, Western Australia, Australia

[5]Cadena Drive, Galveston, Texas, USA.

[6]Laboratorio de Biospeología y Carcinología, DDS, Universidad Autónoma del Estado de Quintana Roo, Campus Cozumel, Quintana Roo, Mexico.

[*]fercg12@gmail.com

Abstract

Subterranean estuaries are coastal ecosystems characterized by vertically stratified groundwater. The biota within these ecosystems is relatively understudied due to the inherent difficulty of accessing such extreme environments. The fauna inhabiting these ecosystems is considered vulnerable to extinction, and the presence of cryptic species has major implications for research and conservation efforts. Most species lack molecular data; however, the evaluation of genetic data for some taxa have revealed that undocumented species are common. This study employs molecular species delimitation methods and DNA barcoding through the analysis of publicly and newly generated sequences, including individuals from type localities and non-crustacean phyla; the latter typically overlooked in biodiversity assessments of subterranean estuaries. We analyzed 376 COI gene sequences and 154 16S rRNA gene sequences. The COI sequences represented 32% of previously described species and 50% of stygobiont species from the Yucatan Peninsula and Cozumel Island, while sequences of the 16S rRNA represented 14% of described species, and 22% of stygobionts. Our results revealed cryptic genetic lineages and taxonomic misidentification of species. As several species from these ecosystems are recognized as endangered, the use of molecular approaches will improve biodiversity estimates and highlight overlooked cryptic lineages in need of evaluation of conservation status.

Keywords

Cave fauna; stygobiont; troglobiont; anchialine; cave; species delimitation

Background

Subterranean estuaries, also known as anchialine caves, are coastal aquifer ecosystems characterized by vertically stratified groundwater, where a fresh to brackish meteoric lens is buoyed over one or more layers of brackish to marine groundwater, each separated by a halocline interface [1,2]. Typically, subterranean estuaries can be accessed via sinkholes that lead to complex networks of submerged cave

systems [3,4]. Anchialine caves are typically subdivided into three sections: (i) the entrance, including the photic zone that is nutrient rich and where photosynthesis dominates; (ii) cavern or twilight zone, the ecotone between the photic and aphotic regions; and (iii) the "true" cave zone, which is aphotic and often considered oligotrophic [5,6]. Subterranean estuaries are inhabited by *stygobionts* (i.e., species with distributions limited to groundwater), *stygophiles* (i.e., found in both groundwater and epigean environments), and *stygoxenes* (i.e., accidentals, or those found occasionally in groundwater habitats) [7,8]. Stygobionts exhibit high levels of endemism and have been shown to exhibit close affinities to deep-sea taxa [9], or, to epigean fauna from both freshwater and marine environments [10–12].

The Yucatan Peninsula (i.e., Mexican states of Campeche, Quintana Roo, Yucatan, eastern edge of Tabasco, northern Belize, and northern Guatemala) and Cozumel Island are considered prime examples of subterranean estuaries, geologically, hydrologically, and biologically based on over a century of research and exploration [3,13,14], and both landmasses are often treated as one biogeographic region for the study of stygobionts [15,16] (Figure 1). In this region, sinkholes are locally known as *cenotes* [3,17]. The Yucatan Peninsula is the region with the highest number of mapped underwater caves, including the longest underwater caves worldwide (i.e., Ox Bel Ha 436 km) [3,18]. Cozumel Island is separated by a 400 m deep channel from the Yucatan Peninsula [19], with an estimate of more than 200 cenotes [20], however, only six of them have been mapped and scientifically explored [21].

These ecosystems are vulnerable to anthropogenic pressure [22,23], meteorological events [24–26], and possible extinction of endemic species [27], yet the evolutionary affinities of most anchialine stygobionts remain poorly studied. The subterranean estuaries of the Yucatan Peninsula and Cozumel Island are considered hotspots for diverse endemic invertebrate species. Currently, most species descriptions and identifications, including recent publications, are based on morphological taxonomy [10,28–31], and primarily focused on crustaceans [15,32,33]. Five stygobiont species are co-recorded on both the

Yucatan Peninsula and Cozumel Island: *Bahadzia bozanici* Holsinger, 1992 [Amphipoda], *Barbouria cubensis* (von Martens, 1872) [Decapoda], *Calliasmata nohochi* Escobar-Briones, Camacho & Alcocer, 1997 [Decapoda], *Metacirolana mayana* (Bowman, 1987) [Isopoda], and *Yagerocaris cozumel* Kensley, 1988 [Decapoda] [17,30,34]. Type species representatives (i.e., used in the original species description) of *Bahadzia bozanici* and *Metacirolana mayana* include individuals from caves in both regions. The holotypes are from a cave in Cozumel Island (La Quebrada), while the paratypes are from caves on the Yucatan Peninsula (Carwash and Temple of Doom, respectively) and Cozumel Island [16,35].

Molecular methods have become widely used tools for the documentation of biodiversity and for the identification of cryptic and undocumented species lineages [36–38]. Particularly, DNA barcoding serves to establish genetic reference libraries for biodiversity assessments through metabarcoding and environmental DNA (eDNA), including species identification by non-specialists. While metabarcoding and eDNA analyses allows for community-level characterization of taxa without direct observation, the lack of an accurate and curated species reference library is an outstanding problem in need of resolution to address major research questions in subterranean biology [39–41].

Mitochondrial cytochrome c oxidase subunit I gene (COI) has been widely used for barcoding eukaryotic fauna, data of which has become widely available via public databases, such as GenBank and the Barcode of Life Data System v4 (BOLD) [39,42–47]. The ribosomal mitochondrial 16S rDNA (16S) gene has also been proposed and used as an alternative or complement to COI in DNA barcoding projects. The gene 16S allows higher amplification success across taxa than COI alone since it is a more conserved region, and in many taxa, 16S does not underestimate species diversity [38,48]. However, taxonomic and sequence contamination errors, including pseudogenes, in public databases can lead to species misidentification [10,49–51]. Therefore, to improve the knowledge of species distributions and clarify taxon identities, it is crucial that species descriptions also include deposited genetic vouchers, and that

efforts are made to resample species from type localities when molecular data are lacking. In turn, this will drastically improve future phylogenetic evaluations, allows molecular species delimitations, and resources for ecological studies.

This study aims to expand the current understanding of biodiversity within subterranean estuaries of the Yucatan Peninsula and Cozumel Island under the following aims: to re-evaluate the taxonomic status of stygobionts through molecular species delimitation analyses and create a DNA barcode library based on COI and 16S. These subterranean estuaries serve as an ideal case study to evaluate diversity patterns from this environment, and include sampling from type localities of select species, including non-crustacean representatives that are historically overlooked in subterranean estuary studies (e.g., Annelida, Echinodermata, Nemertea and Porifera). It was predicted that most stygobiont species with a broad distribution, especially those inhabiting different landmasses (i.e., Yucatan Peninsula and Cozumel Island), will be represented by species complexes.

Methods

Literature review of DNA barcoding of fauna from the subterranean estuaries

A species list from the caves of the Yucatan Peninsula and Cozumel Island was compiled and updated from Calderón-Gutierrez *et al*. [17] (Suppl. 1). An exhaustive literature review was conducted to obtain records of published and unpublished COI and 16S sequence data. There is ample evidence that the entrance and cave zones are distinct environments [5,7,52]. Consequently, only sequences from species previously reported from the cave zone were considered in this study. We considered as type locality only the sampling location of the holotype.

Sampling collection and study area

Organisms were collected by hand using cave diving techniques from the cave zone (Figure 1a), relaxed by cooling, until they stopped responding to physical stimuli, and preserved in 70-96% ethanol. Sampling efforts took place between 2011 and 2023 from: (a) seven anchialine caves (i.e., with presence of a halocline) on the Caribbean Coast of the Yucatan Peninsula: Aayin Aak (or Crustacea), Actun Ha (or Carwash), Chac Mool, Sac Actun (accessing through cenote Kalimba and Manatí), Murena (accessing through cenote Aak Kimin), Muk Ki'in (accessing through cenote Nohoch Pek), Ox Bel Ha (accessing through cenote Bang and Naharon); (b) six anchialine caves in Cozumel Island: Bambu, Chempita, Chun Ha, El Aerolito, La Quebrada (accessing through cenote S-1 and Km-1), and Tres Potrillos; and (c) two marine caves in Belize: Winter Wonderland (or Caye Chapel Cave) and Giant Cave (Figures 1-2). The remaining preserved organisms were stored at room temperature or at -20°C, and DNA extractions were stored at -20°C.

DNA extraction, amplification, sequencing, and sequence-based species delimitation

DNA extractions were performed using either the Qiagen DNeasy Tissue and Blood Kit, ethanol precipitation [53], or, by phenol-chloroform [54]. Final elution was in 50 µl Buffer AE (Qiagen) or Nuclease-free water (Promega). DNA quality was evaluated through UV-Vis spectrophotometer NanoDrop2000 (Thermo Scientific). The Folmer region of COI was amplified using universal [45,46,55], or specific primers [28,47,56]. The Palumbi region of 16S was amplified with the primers 16Sar/16Sbr [57]. PCR reaction mixtures totalled 12.5 µl and included GoTaq polymerase (Promega - 6.25 µl), RNAse-free water (4.25 µl), forward and reverse primers (0.5 µl each), and DNA template (1 µl). In some instances, we also included 0.25 µl of MgCl 50 µM, 0.25 µl BSA, and/or 2 µl of DNA template, and in each case, the water was adjusted to maintain a final volume of 12.5 µl. PCR products were visualized on 1% agarose gels stained with SYBR Safe (EDVOTEK) or GelRed (Biotium). Unsuccessful amplifications were reamplified using the same PCR settings with 2 µl of PCR product from the first reaction and 3.25

µl Nuclease-free water. Specific extraction method and successful PCR mixtures and conditions are available in Suppl. 2. Successful PCR products were purified using ExoSap-IT Express (Applied Biosystems) and sent to Azenta, Inc. (South Plainfield, NJ), or to the Genomics Core Lab at Texas A&M University - Corpus Christi for sequencing. Contigs were assembled, visually inspected, trimmed and cleaned using Geneious Prime ver. 2022.0.2 [58]. COI sequences were translated into amino acids and checked for stop codons in Geneious Prime [58].

Morphological taxonomic assignment was conducted with current literature and original species descriptions (e.g., [21,59–61]). Molecular taxonomic assignment was performed with the Identification System of BOLD (COI sequences); and with the sequences in the GenBank database using BLAST, applying a maximum 2% sequence divergence criterion (COI and 16S sequences) [42–44]. Eight molecular species delimitations methods were applied to taxa with either a >2% divergence, sequences from species with known cryptic species or taxonomic assignment problems. Species delimitation methods were performed by genus. Sequences were aligned with related sequences from GenBank using the MUSCLE Alignment algorithm in Geneious Prime [58,62]. Species were evaluated via: 1) Uncorrected pairwise distance (UPD); 2) Corrected pairwise distance (CPD); 3) Poisson Tree Process (PTP); 4) multi-rate Poisson Tree Process (mPTP); 5) Refined Single Linkage (RESL); 6) Automatic Barcode Gap Discovery (ABGD); 7) Assemble Species by Automatic Partitioning (ASAP); and 8) General Mixed Yule-Coalescent (GMYC). Uncorrected pairwise distance matrixes were constructed on Geneious prime [58]; corrected pairwise distance matrixes were constructed on Mega 11 [63,64] with a Kimura 2 parameter model; a 2% sequence divergence was applied [47,58,65]. PTP and mPTP were performed using an ultrametric and fully bifurcating tree (see below) with the default parameters (https://mptp.h-its.org/), unless otherwise stated in Suppl. 2 [66]. RESL was performed in BOLD [42]. A probability of 0.001 to split groups was used for ABGD (https://bioinfo.mnhn.fr/abi/public/abgd/abgdweb.html) and

ASAP (https://bioinfo.mnhn.fr/abi/public/asap/) [67,68]. In ABGD, the initial partition was considered [69]. For ASAP, the best partition was selected based on the ASAP-score, probability (p-rank) and ranked distance (w-rank). The ultrametric and fully bifurcating phylogenetic tree required for GMYC, and used on PTP and mPTP, was reconstructed using the Bayesian approach implemented in BEAUti and BEAST v2.6.6. Two Markov chains of 50,000,000 generations each were run independently with a strict clock (rate= 1). The optimal substitution model was identified though ModelTest based upon corrected AIC. (Suppl. 2). A pre-burn period of 10,000,000 was used, and trees were sampled every 10,000 states. Both runs were combined, and the maximum credibility tree was built with LogCombiner and TreeAnnotator v2.6.6 [70]. Finally, the GMYC was performed on the R platform v4.1.2 with the package splits v1.0-20 [71,72].

A final delimitation scheme was based on >50% consensus agreement among all methods (i.e., number of methods resulting in the same identification) to produce a robust delimitation. It has been recommended at least a 75% agreement for species represented by a singleton (i.e., represented by only one sequence) because each of these methods is affected by the lack of multiple sequences per species [39]. From a biodiversity conservation perspective, the presence of more than one species was assumed as the precautionary criterion and used a 50% consensus regardless of the number of sequences available.

Results

Biodiversity and barcoded anchialine invertebrates from the literature

A total of 228 species (45% crustaceans; n= 101 species; Figure 3, Suppl. 1-2) have been recorded in the literature pertaining subterranean estuaries of the Yucatan Peninsula and Cozumel Island, 42% (n=95; 80 nominal species) were classified as stygobionts. To date, 24 studies have included COI and 16S data (n=

11, COI & 16S; n= 7, COI only; n= 6, 16S only; Figure 3, Suppl. 1-2). These sequences represented individuals collected from 48 caves, of which most were concentrated near the “ring of cenotes” (Figure 2). Five species of crustaceans and one ophiuroid represented 59% of all available COI sequences in the literature, and four shrimp species represented 78% of all the available 16S sequences. In contrast, 25 species were singletons or doubletons with COI, and seven with 16S (Figure 3, Suppl. 2).

Prior to this study, non-crustacean barcoded species were limited to only one annelid, two stygobiont fish species, and twenty-one echinoderms. DNA was extracted from a total of 234 specimens representing eight metazoan phyla. A total of 100 COI and 48 16S sequences were successfully amplified and sequenced, representing 44 species (32 stygobionts) across six phyla: Annelida (3 species), Arthropoda (31 species), Echinodermata (6 species), Mollusca (1 species), Nemertea (1 species), and Porifera (2 species). Sequencing failed for all samples representing Chordata and Cnidaria. Including this study, 32% (73 species; 376 sequences) of all recorded species and 50% (48 species) of stygobionts have been barcoded with COI, while 14% (n= 31 species; 154 sequences) of all recorded species and 22% (21 species) of stygobionts have been barcoded with 16S (Figure 3, Suppl. 2). Crustaceans are the most widely barcoded (44 species, COI; 19 species, 16S) within this environment. Available sequences now represent 12 species from their type localities barcoded with COI and 16S (n= 5, COI & 16S; n= 7, COI only; Figure 3, Suppl. 2), including COI barcoding information for all remipede species of the genus *Xibalbanus* (Figure 5c), and COI and 16S of the only described representative of the family Anchialocarididae (*Anchialocaris paulini* Mejía-Ortíz, Yañez & López-Mejía, 2017). Newly generated sequences include the first COI sequences representing the infraorder Procaridídea, and the families Anchialocarididae, Agostocarididae, and Epacteriscidae; and the first 16S sequences for the order Stygiomysida, the family Tulumellidae, and the genera *Creaseriella,* and *Metacirolana*, and the second for the order Thermosbaenacea.

Species delimitation

Eight undescribed species lineages were identified: a) Crustacea: mysid shrimp *Antromysis* spp. (n= 2, Figure 5a), stygiomysid shrimp *Stygiomysis* spp. (n= 2, Figure 5b), isopod *Metacirolana* sp. (n= 1, Figure 4c) and thermosbaenacean *Tulumella* sp. (n= 1); and b) Annelida: *Macrochaeta* sp. (n= 1), and Flabelligeridae (n= 1). The only completely inconclusive results (i.e., a hypothesis was not supported by >50% of the analyses) from species delimitation methods corresponded to representatives of *Stygiomysis*, without species identification agreement on any sample (this study= 2 sequences, COI & 2, 16S; literature= 2, COI; sequence availability did not allow analyses with 16S; Table I, Figure 5b). Inconclusive results for at least one sample were also present for *Gorgonorhynchus* (this study= 4 sequences, COI; literature= 1, COI, not from a cave environment), and *Typhlatya* (this study= 29 sequences, COI & 20, 16S; literature= 56, COI & 74 from the Yucatan Peninsula and Cozumel Island). The analyses of *Antromysis,* Metacirolana, *Typhlatya* and *Xibalbanus* identified several cryptic lineages (Table I, Figures 4-5). See Suppl. 2 and Calderón-Gutiérrez et al., 2024 [73] for further details.

Table 1.- Summary of species delimitation analyses by genus. Number of described species are based on the literature and preliminary identification with Blast and Bold. Species delimitation analyses: Uncorrected pairwise distance (UPD), Corrected pairwise distance (CPD), Poisson Tree Process (PTP), multi-rate Poisson Tree Process (mPTP), Refined Single Linkage (RESL, only available for COI), Automatic Barcode Gap Discovery (ABGD), Assemble Species by Automatic Partitioning, and General Mixed Yule-Coalescent (GMYC). Consensus refers to number of species within each genus, as identified by the species delimitation analyses. IR indicate inconclusive results. [1]Species identification of *Typhlatya* follows Ballou et al (2022); *Typhlatya kakuki* was not included in the analysis with COI since there are no sequences available from the Yucatan Peninsula or Cozumel Island. (+) At least one published sequence identified as an undescribed species not previously reported; consensus identification with CO1

sequences OM456541, OM456542, OM456544, and OM456519; and 16S sequence OM458928 identifies them as from undescribed species. See Suppl. 2 and Calderón-Gutiérrez et al., 2024 [73] for further details.

		Species delimitation analyses																	
		CO1									16S								
	Described species	UPD	CPD	PTP	mPTP	RESL	ABGD	ASAP	GMYC	UPD	CPD	PTP	mPTP	ABGD	ASAP	GMYC	**CO1 Consensus**	**16S Consensus**
Antromysis	**1**	3	3	3	1	2	2	2	3								IR	
Creaseriella	**1**	1	1	3	1	1	1	2	2	1	1	2	1	1	1	1	**1**	**1**
Metacirolana	**1**	2	2	4	2	2	2	2	4	2	2	3	2	2	2	2	**2**	**2**
Stygiomysis	**2**	4	4	2	1	4	2	2	2								IR	
Typhlatya[1]	**6/7**[1]	10	10	10	9	6	6	7	8	9	9	9	7	7	7	9	**8**	**9**
Gorgonorhynchus	**1**	1	1	3	1	1	2	2	2	1	1	2	1	1	1	1	IR	**1**
Xibalbanus	**4**	4[+]	4[+]	4[+]	3	3	3	3	3								**3**	

Several stygobionts were confirmed to have an extensive distribution including: the isopod *Creaseriella anops* (Creaser, 1936) (n=8, COI; n=3, 16S; Figure 4a) from inland and coastal caves (205 km) within the Yucatan Peninsula. *Typhlatya* (Figure 4d) exhibited the most species with broad distributions, including the first records of *Typhlatya* in caves from Cozumel Island corresponded to *T. kakuki* Alvarez, Iliffe & Villalobos, 2005 (n= 1, 16S) and *T. iliffei* Hart & Manning, 1981 (n= 1, COI; n= 1, 16S). The type locality of *T. kakuki* is Shrimp Hole cave in Acklins Island, Bahamas, with sequences available (n= 7, COI; n= 13, 16S) from Bahamas, including the type locality, and Caicos. *Typhlatya iliffei* was described from Tucker's Town Cave in Bermuda, and has sequences available (n= 2, COI [Note: additional COI sequences are available, but do not include the Folmer region]; n= 23, 16S) from Bermuda, type locality representatives. Thus, species distribution range of *T. kakuki* and *T. iliffei* are 1,600 km and 2,575 km, respectively. The remipede *Xibalbanus cokei* (Yager, 2013) (n= 4, COI) from Winter Wonderland cave (type locality) was here identified as *X. tulumensis* (n= 9, COI) in all species delimitation methods, with pairwise COI similarities 96.6-98.7%. The latter supports an extensive a coastal distribution along the Yucatan Peninsula (360 km). Species identification and delimitation analyses also confirmed that the

241 species *Cirolana adriani* Ortiz & Cházaro-Olvera, 2015*, Creaseriella anops,* and *Typhlatya dzilamensis* are
242 identified as euryhaline, as they were distributed above and below the halocline (Figures 4-5, Suppl. 2).

243 **Discussion**

244 Our data indicate that species lineages from subterranean estuaries are more complex than expected.
245 The hypothesis was that most stygobionts species with a broad distribution, especially those inhabiting
246 different landmasses, will be represented by species complexes. However, we did not only identify
247 species complexes (e.g., *Antromysis, Metacirolana*), but also species with synonyms (e.g. *Xibalbanus*
248 *tulumensis*), species with broad distributions across spaces in the same (e.g., *Creaseriella anops,*
249 *Typhlatya dzilamensis* & *Xibalbanus tulumensis*) and different (e.g., *Typhlatya iliffei* & *T. kakuki*)
250 landmasses, and even across salinities (e.g., *Creaseriella anops* & *Typhlatya dzilamensis*). Inconclusive
251 results from the species delimitation methods in four genera further support the need of a higher
252 sequencing depth across the distribution range of the species within the subterranean (e.g. "*Typhlatya*
253 sp. A" has five sequences available collected across ~300km).

254 A detailed discussion by taxon, and biogeographic distributions is available on Suppl. 3.

255 *DNA barcoding and data availability*

256 The increase in sequence data availability from subterranean estuaries, especially among taxa from the
257 cave zones, provides: (a) a foundational molecular taxonomy reference to improve biodiversity
258 inventories [74], (b) genetic data for integrative ecological and evolutionary studies [10,75], (c) provide
259 information for further biodiversity studies to investigate not only subterranean estuary fauna, but also
260 overall biodiversity and phylogenetic relationships [76]. Nevertheless, to clarify taxonomic and
261 phylogenetic uncertainties, it is necessary to include in the analyses the type material (i.e., specimens
262 used in the original species description), or samples from the type locality (i.e., specimens collected from

263 the sampling site from the type series) [77]. This study revealed taxonomic uncertainties that need to be
264 addressed among the genera *Antromysis, Metacirolana, Typhlatya, Stygiomysis,* and *Xibalbanus* (Figures
265 4-5, Suppl. 2, see below). However, there is substantial difficulty in verifying new and existing species
266 without molecular data from type material. There are only four previously described species that have
267 COI sequences from the type material: the amphipod *Mayaweckelia troglomorpha* Angyal, 2018*,*
268 remipedes *Xibalbanus cozumelensis* and *Xibalbanus fuchscockburni,* and the ophiuroid *Ophionereis*
269 *commutabilis* (Figures 4-5, Suppl. 2) [28,47,60,78,79]. The sea star *Copidaster cavernicola* Solis-Marin &
270 Laguarda-Figueras, 2010 is DNA barcoded (COI) from the type locality. Of these, only *X. cozumelensis*
271 also has 16S sequences. In this study we obtained the first DNA barcodes from the type locality
272 representatives of the shrimps *Agostocaris zabaletai* (COI & 16S)*,* and *Anchialocaris paulini* (COI & 16S)*,*
273 the isopods *Metacirolana mayana* (COI & 16S), and *Cirolana adriani* (COI); and remipedes *Xibalbanus*
274 *cokei* (COI), and *Xibalbanus tulumensis* (COI)*.* It is recommended that any future taxonomic work should
275 include sequencing from type material or type localities for future reference of species records, and
276 even considering the reconstruction of full mitochondrial genomes and genomic data [80,81].

277 *Species records*

278 There are conflicting perspectives in regards to the documentation of cave biodiversity, with either
279 taxonomic records of cave micro-endemics inhabiting only one or two caves (e.g., *Xibalbanus* spp.,
280 Agostocarididae) [21,60] or regional cave cosmopolitans (e.g., *Antromysis cenotensis, Barbouria*
281 *cubensis*) [13,82], with most species identification and description, including >50% being described since
282 2000, based solely on morphology [10,30,83–85]. Furthermore, the paucity of ecological and
283 environmental data and/or number of sequenced species representatives and specimens from type
284 localities limits taxonomic evaluation and phylogenetic analyses.

In this study, molecular data compared against known species records identified seven obstacles to diversity assessments due to the biology of the species and the current state of biodiversity inventories of subterranean estuaries. Identified obstacles (labelled i-vii) are illustrated in the following four examples. Example 1: Species complexes (i) and lack of genetic data from type locality representatives (ii). *Antromysis cenotensis* is recorded as a single species across the Yucatan Peninsula [86], yet species delimitation indicated at least three species lineages (Figure 5a). The absence of sampling from or near the type locality (43 km to the nearest DNA barcoding sampling site) complicates the identification of this mysid. Example 2: Distinct described morphospecies within a single species (iii) and assumed micro-endemics (iv). The remipede *Xibalbanus cokei,* previously considered micro-endemic to a single cave in Belize [87], was here supported as a potential junior synonym of *X. tulumensis* (Figure 5c). Example 3: Poor taxonomic sampling for phylogenetic assessment (v) and low phenotypic diversity among syntopic species (vi). The remipede *Xibalbanus fuchscocburni* is represented by a single sequence [28], and lives in syntopy with the morphologically similar *X. tulumensis* [28]. Example 4: Limited ecological data (vii) further hinders biodiversity assessments and other studies (i.e., ecology, biogeography). The shrimp *Procaris mexicana* was described with only the cave name as the type locality (Cueva Quebrada, Chankanaab Park, Cozumel); without coordinates or other details [88].

Lastly, congenerics of *Typhlatya* (5 species) [10], *Stygiomysis* (2 species, this study), and *Xibalbanus* (2 species) [28] have been recorded as syntopic within the same cave system within a short distance, or even in the same cave passages. Low phenotypic diversity and/or limited understanding of ecological attributes led to misidentifications (e.g., *Typhlatya*), underrepresented (e.g., *Antromysis* and *Metacirolana*), and overrepresented (e.g., *Xibalbanus* and *Barbouria*) diversity [10,13,16,28,30,87]. Species misidentifications leading to incorrect evolutionary and ecological conclusions have already been identified on *Typhlatya* [10] and *Barbouria* [30]. For example, Ballou et al., [10] detected a >20%

misidentification rate on taxonomic, phylogenetic, and ecological studies on *Typhlatya*. Misidentification of *Typhlatya* includes the mitochondrial genome identified as *Typhlatya mitchelli* on the NCBI Reference Sequence (RefSeq, record NC035403) [10,89], however, it represents an undescribed species (*Typhlatya* sp. B). The above highlights that species records without molecular verification from subterranean estuaries may not be reliable beyond genus level, and likely does not reflect the potential diversity estimations due to the lack or limited integrative taxonomic evaluations [10,30]. Syntopic species with low phenotypic variation also restrict non-invasive visual ecological studies to genus-level on such taxa. It is recommended that any future biodiversity assessment and ecological study of fauna from the subterranean estuaries, especially from taxa identified with low phenotypic variation (e.g., *Antromysis, Stygomysis, Typhlatya, Xibalbanus*), include molecular identification [10,41,75,90–92]).

Challenges of DNA barcoding for understudied taxa

Fauna inhabiting caves, especially stygobionts, can have limited population size, and several species are threatened or vulnerable to extinction [27,93,94]. Diversity projects require sampling that are likely to be limited in number, due to access or low abundances, thus the evaluation for each specimen should be maximized for use in integrative taxonomy studies. This may include moving to downstream molecular approaches even when DNA extractions have “suboptimal” concentrations, which may lead to fewer successful PCR amplifications. Some species may have PCR inhibitors [74], particularly when collected from caves with high concentration of hydrogen sulphide. Another challenge with DNA barcoding is primer selection, truly universal primers do not exist, thus a multi-primer approach is needed for projects working with diverse taxonomic groups. In this case, COI primers by Geller *et al*. [46], had the best results with our samples, and thus an initial exploration with Geller’s primers is recommended. The amplification of 16S resulted as a feasible DNA barcoding alternative, with a higher success rate than COI when considering the use of the same PCR conditions across taxa/samples and a

single set of primers. Additionally, as observed in other studies [38,48], 16S resulted in a greater taxonomic/specimen coverage for some groups and did not underestimate species diversity [38].

Fundamental research questions requiring molecular information in subterranean biology as identified by Mammola et al., [40], such as: (a) Would the use of novel molecular methods provide new insights on subterranean biodiversity patterns and affect known patterns?; (b) What drives subterranean patterns of phylogenetic and functional diversity?; (c) What is the species richness pattern of subterranean organisms globally? These questions emphasize how next generation sequencing methods have the potential to improve the current understanding of biodiversity patterns and estimates. In order to respond to these questions, it is necessary to increase DNA barcoding data availability of the subterranean aquatic fauna. Among the major limitations of barcoding and/or metabarcoding projects for understudied ecosystems are taxonomic misidentifications and/or limited representation in sequence reference libraries [95]. Taxonomic misidentifications can be avoided with integrative taxonomy approaches that include type material or samples from type localities [96]; while PCR-free approaches like genome skimming circumvent PCR bias [80,81]. We also identify the need for mechanisms allowing amendment of public databases, such as GenBank and BOLD, allowing third parties updates, or adding alternative identification, thus leaving the original identity unchanged. Update capabilities of sequence's identification deposited in public databases will better represent the dynamic state of the taxonomy and science and increase the applicability of the vast molecular data publicly available.

Implications of DNA barcoding on the conservation of subterranean estuaries

Molecular analysis and species delimitation methods allowed for the identification and confirmation of 1) Syntopic species: five species of the shrimp *Typhlatya* in the Ox Bel Ha system [10]; two species of remipede genus *Xibalbanus* in Aayin Aak [28]; and two species of the stygiomysid *Stygiomysis* in the

Nohoch Pek system. 2) Identification of species complexes of the genera *Antromysis, Tulumella, Metacirolana,* and *Stygiomysis.* 3) Broad species – spatial - distribution ranges have been confirmed for ten species, including an isopod, shrimps and a remipede. 4) Broad species – salinity – distribution ranges have been confirmed for the isopods *Cirolana adriani* and *Creaseriella anops* and the shrimp *Typhlatya dzilamensis* [10]. Identifying the presence of cryptic species for conservation and management purposes implies that the species richness and vulnerability of the ecosystem are likely underestimated, especially for lesser-known microfauna such as *Antromysis cenotensis* belonging to a species complex, while other species may be micro-endemic to single caves (e.g., *Copisdaster cavernicola, Teinostoma brankovitsi* Rubio, Rolán, Worsaae, Martínez & Gonzalez, 2016, *Triacanthoneus akumalensis* Alvarez, Illife, Gonzalez & Villalobos 2012) [84,85,94]. For research projects, cryptic and syntopic species represent an opportunity to better understand evolutionary and ecological processes, especially because these ecosystems have simpler community structures and the semi-isolation characteristics of the subterranean estuary [10,36,97].

The presence of complex biodiversity patterns and inaccurate records of stygobionts taxa have also been reported in other regions (i.e., Europe, Middle East, Australia) after molecular re-evaluation [92,98,99], as such species complexes are likely a generality in aquatic subterranean ecosystems. Climate change and increasing anthropogenic pressures, such as rapid demographic growth, tourism activities, water pollution [22,23,27,100], are also concerning threats to subterranean ecosystems. Inaccurate species identification have major legal and logistic implications for research and conservation efforts, as some of the species are listed under extinction risk both in our study area (e.g., *Xibalbanus tulumensis, Antromysis cenotensis, Typhlatya pearsei*) [101], and in other regions worldwide such as Texas (e.g., *Stygobromus pecki, Lirceolus cocytus*) [102], the Canary Islands (e.g., *Speleonectes ondinae*) [103], and Australia (e.g., *Ophisternon candidum*) [104]; and evidence suggests that all stygobiont

species should be considered at risk of extinction [27,93]. Most species within subterranean estuaries are either recently described or remain undescribed, and recent studies suggest that these species are potentially the most vulnerable [27,84,85,93].

Conclusions

In this study, we identified the need of taxonomic status re-evaluation of stygobionts of the Yucatan Peninsula and Cozumel Island with an integrative approach, utilizing molecular methods to complement current morphological evaluations. Identified patterns of under- and over-descriptions have also been reported in other regions after molecular re-evaluation, therefore, these patterns should be generalized in aquatic subterranean ecosystems, and not limited to our study area. Reliable biodiversity records with correct species identifications and species distribution ranges are required to: (a) provide a foundation for continuing research in ecology, phylogenetics/genomics, evolution, biogeography, etc. at population to ecosystem levels; (b) evaluate the conservation status of the species; and (c); develop and implement proper conservation and management projects.

Acknowledgments

We thank Lauren Ballou and Heather Bracken-Grisom for her help in the laboratory work and for providing samples, German Yañez, Natalie Lynn Gibb, Roberto Rojo, Rory Okeefe, Arturo Mora, Marcelin Nebenhaus, Michel Vazquez, Skanda Cophield, James Petersen, Ilya Rosado, Anthony Tedeschi, Becky Kagan Schott, Chrissy Richards, Jason Richards, Ben Popik, KISS rebreathers USA exploration team, and Círculo Espeleológico del Mayab A. C. for their invaluable help during the fieldwork. We appreciate the logistic support by Fernando Álvarez, Nuno Simoes, and Efrain Chávez Solís. We also thank the landowners to allow us to access the caves. We deeply appreciate the help of Aurora Gaona in the identification of the copepod.

399 **References**

400 1. Bishop RE *et al.* 2015 'Anchialine' redefined as a subterranean estuary in a crevicular or
401 cavernous geological setting. *Journal of Crustacean Biology* **35**, 511–514. (doi:10.1163/1937240X-
402 00002335)

403 2. van Hengstum PJ, Cresswell JN, Milne GA, Iliffe TM. 2019 Development of anchialine cave
404 habitats and karst subterranean estuaries since the last ice age. *Sci Rep* **9**, 11907.
405 (doi:10.1038/s41598-019-48058-8)

406 3. Goldscheider N *et al.* 2020 Global distribution of carbonate rocks and karst water resources.
407 *Hydrogeol J* **28**, 1661–1677. (doi:10.1007/s10040-020-02139-5)

408 4. Culver DC, Deharveng L, Pipan T, Bedos A. 2021 An overview of subterranean biodiversity
409 hotspots. *Diversity (Basel)* **13**, 487. (doi:10.3390/d13100487)

410 5. Culver DC, Pipan T. 2009 *The biology of caves and other subterranean habitats*. United Kingdom:
411 Oxford University Press. (doi:10.1093/oso/9780198820765.001.0001)

412 6. Moseley M. 2009 Are all caves ecotones? *Cave and Karst Science* **36**, 53–58.

413 7. Moldovan OT, Kováč Ľ, Halse S. 2018 *Cave ecology*. Cham: Springer. (doi:10.1007/978-3-319-
414 98852-8)

415 8. Hart Jr. CW, Manning RB, Iliffe TM. 1985 The fauna of Atlantic marine caves: evidence of
416 dispersal by sea floor spreading while maintaining ties to deep waters. *Proceedings of the*
417 *Biological Society of Washington* **98**, 288–292.

9. Gonzalez BC, Martínez A, Worsaae K, Osborn KJ. 2021 Morphological convergence and adaptation in cave and pelagic scale worms (Polynoidae, Annelida). *Sci Rep* **11**, 10718. (doi:10.1038/s41598-021-89459-y)

10. Ballou L *et al.* 2022 An integrative re-evaluation of *Typhlatya* shrimp within the karst aquifer of the Yucatán Peninsula, Mexico. *Sci Rep* **12**, 5302. (doi:10.1038/s41598-022-08779-9)

11. Wilkens H. 1979 Regressive evolution and phylogenetic age: The history of colonization of freshwaters of Yucatan by fish and Crustacea. *Association for Mexican Cave Studies Bulletin* **8**, 237–243.

12. Von Rintelen K, Page TJ, Cai Y, Roe K, Stelbrink B, Kuhajda BR, Iliffe TM, Hughes J, Von Rintelen T. 2012 Drawn to the dark side: A molecular phylogeny of freshwater shrimps (Crustacea: Decapoda: Caridea: Atyidae) reveals frequent cave invasions and challenges current taxonomic hypotheses. *Mol Phylogenet Evol* **63**, 82–96. (doi:10.1016/j.ympev.2011.12.015)

13. Creaser EP. 1936 Crustaceans from Yucatan. *Carnegie Institution of Washington Publication* **457**, 117–132.

14. Ward WC, Weidie AE, Back W. 1985 *Geology and hydrology of the Yucatan and Quaternary Geology of Northeastern Yucatan Peninsula*. New Orleans Geological Society.

15. Alvarez F, Iliffe TM, Benitez S, Brankovits D, Villalobos JL. 2015 New records of anchialine fauna from the Yucatan Peninsula, Mexico. *Check List* **11**, 1–10. (doi:10.15560/11.1.1505)

16. Bowman T. 1987 *Bahalana mayana*, a new troglobitic cirolanid isopod from Cozumel Island and the Yucatan Peninsula, Mexico. *Proceedings of the Biological Society of Washington* **100**, 659–663.

17. Calderón-Gutiérrez F, Solís-Marín FA, Gómez P, Sánchez C, Hernández-Alcántara P, Álvarez-Noguera F, Yáñez-Mendoza G. 2017 Mexican anchialine fauna — With emphasis in the high biodiversity cave El Aerolito. *Reg Stud Mar Sci* **9**, 43–55. (doi:10.1016/j.rsma.2016.11.001)

18. Burger P. 2023 World long caves. See https://cave-exploring.com/index.php/long-and-deep-caves-of-the-world/world-long-caves/ (accessed on 23 October 2023).

19. Ward WC. 2004 Geology of coastal islands, Northeastern Yucatan Peninsula. In *Geology and Hydrogeology of Carbonate Islands* (eds HL Vacher, TM Quinn), pp. 275–298. Amsterdam: Elsevier. (doi:10.1016/S0070-4571(04)80029-3)

20. Rodríguez Castillo JF, Frausto-Martínez O, Colín Olivares O. 2021 Morphometry of karst depressions at detailed scale: El Cedral, Cozumel - Mexico. *Tropical and Subtropical Agroecosystems* **24**, 1–9.

21. Mejía-Ortíz LM, Yañez G, López-Mejía M. 2017 Anchialocarididae, a new family of anchialine decapods and a new species of the genus *Agostocaris* from Cozumel Island, Mexico. *Crustaceana* **90**, 381–398. (doi:10.1163/15685403-00003657)

22. Doehring DO, Butler JH. 1974 Hydrogeologic constraints on Yucatán's development: The peninsula's water resource is threatened by man-induced contamination. *Science (1979)* **186**, 591–595. (doi:10.1126/science.186.4164.591)

23. Moreno-Pérez PA, Hernández-Téllez M, Bautista-Gálvez A. 2021 In danger one of the largest aquifers in the world, the great Mayan aquifer, based on monitoring the cenotes of the Yucatan Peninsula. *Arch Environ Contam Toxicol* **81**, 189–198. (doi:10.1007/s00244-021-00869-5)

24. Brankovits D, Pohlman JW, Lapham LL. 2022 Oxygenation of a karst subterranean estuary during a tropical cyclone: mechanisms and implications for the carbon cycle. *Limnol Oceanogr* **9999**, 1–15. (doi:10.1002/lno.12231)

25. Kovacs SE, Reinhardt EG, Stastna M, Coutino A, Werner C, Collins S V., Devos F, Le Maillot C. 2017 Hurricane Ingrid and Tropical Storm Hanna's effects on the salinity of the coastal aquifer, Quintana Roo, Mexico. *J Hydrol (Amst)* **551**, 703–714. (doi:10.1016/j.jhydrol.2017.02.024)

26. Calderón Gutiérrez F, Iliffe TM, Borda E, Yáñez Mendoza G, Labonté J. 2023 Response and resilience of karst subterranean estuary communities to precipitation impacts. *Ecol Evol* **13**, e10415. (doi:10.1002/ece3.10415)

27. Mammola S *et al.* 2019 Scientists' warning on the conservation of subterranean ecosystems. *Bioscience* **69**, 641–650. (doi:10.1093/biosci/biz064)

28. Neiber MT, Hansen FC, Iliffe TM, Gonzalez BC, Koenemann S. 2012 Molecular taxonomy of *Speleonectes fuchscockburni*, a new pseudocryptic species of Remipedia (Crustacea) from an anchialine cave system on the Yucatán Peninsula, Quintana Roo, Mexico. *Zootaxa* **3190**, 31–46. (doi:10.11646/zootaxa.3190.1.2)

29. García-Machado E, Hernández D, García-Debrás A, Chevalier-Monteagudo P, Metcalfe C, Bernatchez L, Casane D. 2011 Molecular phylogeny and phylogeography of the Cuban cave-fishes of the genus *Lucifuga*: Evidence for cryptic allopatric diversity. *Mol Phylogenet Evol* **61**, 470–483. (doi:10.1016/j.ympev.2011.06.015)

30. Ditter RE, Mejía-Ortíz LM, Bracken-Grissom HD. 2020 Anchialine adjustments: an updated phylogeny and classification for the family Barbouriidae Christoffersen, 1987 (Decapoda: Caridea). *Journal of Crustacean Biology* **40**, 401–411. (doi:10.1093/jcbiol/ruaa037)

31. Marrón-Becerra A, Hermoso-Salazar M, Solís-Weiss V. 2018 *Hyalella maya*, a new hyalellidae species (Crustacea: Amphipoda) from a cenote in the Yucatan Peninsula, Mexico. *Journal of Cave and Karst Studies* **80**. (doi:10.4311/2017LSC0115)

32. Angyal D, Chávez-Solís EM, Liévano-Beltrán LA, Magaña B, Simões N, Mascaró M. 2020 New distribution records of subterranean crustaceans from cenotes in Yucatan (Mexico). *Zookeys* **911**, 21–49. (doi:10.3897/zookeys.911.47694)

33. Yañez-Mendoza G, Zarza-González E, Mejía-Ortíz LM. 2007 Sistemas anquihalinos. In *Biodiversidad acuática de la Isla de Cozumel* (ed LM Mejía-Ortíz), pp. 49–70. México: Universidad de Quintana Roo-CONABIO.

34. Álvarez F, Iliffe TM. 2008 Fauna anquihalina de Yucatán. In *Crustáceos de México: Estado actual de su conocimiento* (eds F Álvarez, GA Rodríguez-Almaraz), pp. 379–418. México, D.F.: Universidad Autónoma de Nuevo León.

35. Holsinger JR. 1992 Two new species of the subterranean Amphipod genus *Bahadzia* (Hadziidae) from the Yucatan Peninsula region of souther Mexico, with an analysis of phylogeny and biogeography of the genus. *Stygologia* **7**, 85–105.

36. Rastorgueff PA, Chevaldonné P, Arslan D, Verna C, Lejeusne C. 2014 Cryptic habitats and cryptic diversity: Unexpected patterns of connectivity and phylogeographical breaks in a Mediterranean endemic marine cave mysid. *Mol Ecol* **23**, 2825–2843. (doi:10.1111/mec.12776)

37. Yin Y, Yao LF, Hu Y, Shao ZK, Hong XY, Hebert PDN, Xue XF. 2022 DNA barcoding uncovers cryptic diversity in minute herbivorous mites (Acari, Eriophyoidea). *Mol Ecol Resour* **22**, 1986–1998. (doi:10.1111/1755-0998.13599)

38. Brasier MJ, Wiklund H, Neal L, Jeffreys R, Linse K, Ruhl H, Glover AG. 2016 DNA barcoding uncovers cryptic diversity in 50% of deep-sea Antarctic polychaetes. *R Soc Open Sci* **3**, 160432. (doi:10.1098/rsos.160432)

39. Shen Y, Hubert N, Huang Y, Wang X, Gan X, Peng Z, He S. 2019 DNA barcoding the ichthyofauna of the Yangtze River: Insights from the molecular inventory of a mega-diverse temperate fauna. *Mol Ecol Resour* **19**, 1278–1291. (doi:10.1111/1755-0998.12961)

40. Mammola S *et al.* 2020 Fundamental research questions in subterranean biology. *Biological Reviews* **95**, 1855–1872. (doi:10.1111/brv.12642)

41. Bush A *et al.* 2019 Studying ecosystems with DNA metabarcoding: lessons from biomonitoring of aquatic macroinvertebrates. *Front Ecol Evol* **7**, 434. (doi:10.3389/FEVO.2019.00434)

42. Ratnasingham S, Hebert PDN. 2007 Barcoding: BOLD The barcode of life database. *Mol Ecol Notes* **7**, 355–364. (doi:10.1111/j.1471-8286.2006.01678.x)

43. Hebert PDN, Ratnasingham S, de Waard JR. 2003 Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc R Soc Lond B Biol Sci* **270**, 96–99. (doi:10.1098/rsbl.2003.0025)

44. Agarwala R *et al.* 2018 Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **46**, 8–13. (doi:10.1093/nar/gkx1095)

45. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. 1994 DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol* **3**, 294–299.

46. Geller J, Meyer C, Parker M, Hawk H. 2013 Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. *Mol Ecol Resour* **13**, 851–861. (doi:10.1111/1755-0998.12138)

47. Bribiesca-Contreras G, Solís-Marín FA, Laguarda-Figueras A, Zaldívar-Riverón A. 2013 Identification of echinoderms (Echinodermata) from an anchialine cave in Cozumel Island, Mexico, using DNA barcodes. *Mol Ecol Resour* **13**, 1137–45. (doi:10.1111/1755-0998.12098)

48. Miglietta MP, Pruski S. 2023 Cryptic species in time and space: an assessment of cryptic diversity within eight nominal species of Hydrozoa (Cnidaria). *Proceedings of the Royal Society B: Biological Sciences* **290**, 20230851. (doi:10.1098/rspb.2023.0851)

49. Meiklejohn KA, Damaso N, Robertson JM. 2019 Assessment of BOLD and GenBank – Their accuracy and reliability for the identification of biological materials. *PLoS One* **14**, e0217084. (doi:10.1371/journal.pone.0217084)

50. Bensasson D, Zhang DX, Hartl DL, Hewitt GM. 2001 Mitochondrial pseudogenes: Evolution's misplaced witnesses. *Trends Ecol Evol* **16**, 314–321. (doi:10.1016/S0169-5347(01)02151-6)

51. Song H, Buhay JE, Whiting MF, Crandall KA. 2008 Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proc Natl Acad Sci U S A* **105**, 13486–13491. (doi:10.1073/pnas.0803076105)

52. Suárez-Moo P, Remes-Rodríguez CA, Márquez-Velázquez NA, Falcón LI, García-Maldonado JQ, Prieto-Davó A. 2022 Changes in the sediment microbial community structure of coastal and inland sinkholes of a karst ecosystem from the Yucatan peninsula. *Sci Rep* **12**, 1110. (doi:10.1038/s41598-022-05135-9)

53. Miglietta MP, Cunningham CW. 2012 Evolution of life cycle, colony morphology, and host specificity in the family Hydractiniidae (Hydrozoa, Cnidaria). *Evolution (N Y)* **66**, 3876–3901. (doi:10.1111/j.1558-5646.2012.01717.x)

54. Nishiguchi MK *et al.* 2002 DNA isolation procedures. In *Techniques in Molecular Systematics and Evolution* (eds R DeSalle, G Giribet, WC Wheeler), pp. 249–287. Basel: Birkhäuser Verlag. (doi:10.1007/978-3-0348-8125-8)

55. Meyer CP. 2003 Molecular systematics of cowries (Gastropoda: Cypraeidae) and diversification patterns in the tropics. *Biological Journal of the Linnean Society* **79**, 401–459. (doi:10.1046/j.1095-8312.2003.00197.x)

56. Hoareau TB, Boissin E. 2010 Design of phylum-specific hybrid primers for DNA barcoding: Addressing the need for efficient COI amplification in the Echinodermata. *Mol Ecol Resour* **10**, 960–967. (doi:10.1111/j.1755-0998.2010.02848.x)

57. Palumbi SR, Benzie J. 1991 Large mitochondrial DNA differences between morphologically similar penaeid shrimp. *Mol Mar Biol Biotechnol* **1**, 27–34.

58. Kearse M *et al.* 2012 Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649. (doi:10.1093/bioinformatics/bts199)

59. Gómez P, Calderón-Gutiérrez F. 2020 Anchialine cave-dwelling sponge fauna (Porifera) from La Quebrada, Mexico, with the description of the first Mexican stygobiont sponges. *Zootaxa* **4803**, 125–151. (doi:10.11646/zootaxa.4803.1.7)

60. Olesen J, Meland K, Glenner H, van Hengstum PJ, Iliffe TM. 2017 *Xibalbanus cozumelensis*, a new species of Remipedia (Crustacea) from Cozumel, Mexico, and a molecular phylogeny of *Xibalbanus* on the Yucatán Peninsula. *Eur J Taxon* **316**, 1–27. (doi:10.5852/ejt.2017.316)

61. Salazar-Vallejo SI, Sergio I, León-González J, Salaices-Polanco H. 1998 *Poliquetos (Annelida: Polychaeta) de México*. México: Universidad Autónoma de Baja California Sur.

62. Edgar RC. 2004 MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797. (doi:10.1093/nar/gkh340)

63. Tamura K, Stecher G, Kumar S. 2021 MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol Biol Evol* **38**, 3022–3027. (doi:10.1093/molbev/msab120)

64. Kimura M. 1980 A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* **16**, 111–120. (doi:10.1007/BF01731581)

65. Ward RD. 2009 DNA barcode divergence among species and genera of birds and fishes. *Mol Ecol Resour* **9**, 1077–1085. (doi:10.1111/j.1755-0998.2009.02541.x)

66. Kapli P, Lutteropp S, Zhang J, Kobert K, Pavlidis P, Stamatakis A, Flouri T. 2017 Multi-rate Poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo. *Bioinformatics* **33**, 1630–1638. (doi:10.1093/bioinformatics/btx025)

67. Puillandre N, Lambert A, Brouillet S, Achaz G. 2012 ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Mol Ecol* **21**, 1864–1877. (doi:10.1111/j.1365-294X.2011.05239.x)

68. Puillandre N, Brouillet S, Achaz G. 2021 ASAP: Assemble species by automatic partitioning. *Mol Ecol Resour* **21**, 609–620. (doi:10.1111/1755-0998.13281)

69. Maroni PJ *et al.* 2022 One Antarctic slug to confuse them all: the underestimated diversity of. *Invertebr Syst* **36**, 419–435. (doi:10.1071/is21073)

70. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014 BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* **10**, e1003537. (doi:10.1371/journal.pcbi.1003537)

71. R Core Team. 2024 *The R Project for Statistical Computing*. Ver. 4.3.3. The Comprehensive R Archive Network.

72. Fujisawa T, Barraclough TG. 2013 Delimiting species using single-locus data and the generalized mixed yule coalescent approach: A revised method and evaluation on simulated data sets. *Syst Biol* **62**, 707–724. (doi:10.1093/sysbio/syt033)

73. Calderón-Gutiérrez F, Labonté JM, Gonzalez B, Iliffe TM, Mejía-Ortíz LM, Borda E. 2024 Cryptic diversity patterns of subterranean estuaries [Data set]. *Zenodo*. https://zenodo.org/records/10530586 (doi:10.5281/zenodo.10530586)

74. Fontaneto D, Flot JF, Tang CQ. 2015 Guidelines for DNA taxonomy, with a focus on the meiofauna. *Marine Biodiversity* **45**, 433–451. (doi:10.1007/s12526-015-0319-7)

75. Couton M, Hürlemann S, Studer A, Alther R, Altermatt F. 2023 Groundwater environmental DNA metabarcoding reveals hidden diversity and reflects land-use and geology. *Mol Ecol* **32**, 3497–3512. (doi:10.1111/mec.16955)

76. Von Reumont BM *et al.* 2012 Pancrustacean phylogeny in the light of new phylogenomic data: Support for Remipedia as the possible sister group of Hexapoda. *Mol Biol Evol* **29**, 1031–1045. (doi:10.1093/molbev/msr270)

77. ICZN. 1999 *International code of zoological nomenclature*. London: International Trust for Zoological Nomenclature. (doi:10.5962/bhl.title.50608)

78. Angyal D, Chávez Solís E, Magaña B, Balázs G, Simoes N. 2018 *Mayaweckelia troglomorpha*, a new subterranean amphipod species from Yucatán state, México (Amphipoda, Hadziidae). *Zookeys* **735**, 1–25. (doi:10.3897/zookeys.735.21164)

79. Bribiesca-Contreras G, Pineda-Enríquez T, Márquez-Borrás F, Solís-Marín FA, Verbruggen H, Hugall AF, O'Hara T. 2019 Dark offshoot: phylogenomic data sheds light on the evolutionary history of a new species of cave brittle star. *Mol Phylogenet Evol* **136**, 151–163. (doi:10.1016/j.ympev.2019.04.014)

80. Trevisan B, Alcantara DMC, Machado DJ, Marques FPL, Lahr DJG. 2019 Genome skimming is a low-cost and robust strategy to assemble complete mitochondrial genomes from ethanol preserved specimens in biodiversity studies. *PeerJ* , 7543. (doi:10.7717/peerj.7543)

81. Tan MH, Gan HM, Bracken-Grissom H, Chan TY, Grandjean F, Austin CM. 2021 More from less: Genome skimming for nuclear markers for animal phylogenomics, a case study using decapod crustaceans. *Journal of Crustacean Biology* **41**, ruab009. (doi:10.1093/jcbiol/ruab009)

82. Martens E von. 1872 Über Cubanische Crustaceen. *Archiv für Naturgeschichte* **38**, 77–147.

83. Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B. 2011 How many species are there on earth and in the ocean? *PLoS Biol* **9**, e1001127. (doi:10.1371/journal.pbio.1001127)

84. Delić T, Trontelj P, Rendoš M, Fišer C. 2017 The importance of naming cryptic species and the conservation of endemic subterranean amphipods. *Sci Rep* **7**, 3391. (doi:10.1038/s41598-017-02938-z)

85. Liu J, Slik F, Zheng S, Lindenmayer DB. 2022 Undescribed species have higher extinction risk than known species. *Conserv Lett* , e12876. (doi:10.1111/conl.12876)

86. Liévano-Beltrán LA, Simoes N. 2021 Updated distribution of the mysid *Antromysis cenotensis* (Crustacea: Peracarida), a protected key species in the Yucatan Peninsula cenotes. *Diversity (Basel)* **13**, 154. (doi:https://doi.org/10.3390/d13040154)

87. Yager J. 2013 *Speleonectes cokei*, new species of Remipedia (Crustacea: Speleonectidae) from a submerged ocean cave near Caye Chapel, Belize. *Zootaxa* **3710**, 354–362. (doi:10.11646/zootaxa.3710.4.4)

88. Sternberg R V., Schotte M. 2004 A new anchialine shrimp of the genus *Procaris* (Crustacea: Decapoda: Procarididae) from the Yucatan Peninsula. *Proceedings of The Biological Society of Washington* **117**, 514–522.

89. O'Leary NA *et al.* 2016 Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* **44**, D733–D745. (doi:10.1093/nar/gkv1189)

90. Alter SE, Arroyave J. 2022 Environmental DNA metabarcoding is a promising method for assaying fish diversity in cenotes of the Yucatán Peninsula, Mexico. *Metabarcoding Metagenom* **6**, 337–347. (doi:10.3897/mbmg.6.89857)

91. West KM, Richards ZT, Harvey ES, Susac R, Grealy A, Bunce M. 2020 Under the karst: detecting hidden subterranean assemblages using eDNA metabarcoding in the caves of Christmas Island, Australia. *Sci Rep* **10**, 21479. (doi:10.1038/s41598-020-78525-6)

92. Stokkan M, Jurado-Rivera JA, Oromí P, Juan C, Jaume D, Pons J. 2018 Species delimitation and mitogenome phylogenetics in the subterranean genus Pseudoniphargus (Crustacea: Amphipoda). *Mol Phylogenet Evol* **127**, 988–999. (doi:10.1016/j.ympev.2018.07.002)

93. Calderón-Gutiérrez F, Sánchez-Ortiz CA, Huato-Soberanis L. 2018 Ecological patterns in anchialine caves. *PLoS One* **13**, e0202909. (doi:10.1371/journal.pone.0202909)

94. Mammola S *et al.* 2022 Towards evidence-based conservation of subterranean ecosystems. *Biological Reviews* **97**, 1476–1510. (doi:10.1111/brv.12851)

95. Fort A *et al.* 2022 Exhaustive reanalysis of barcode sequences from public repositories highlights ongoing misidentifications and impacts taxa diversity and distribution. *Mol Ecol Resour* **22**, 86–101. (doi:10.1111/1755-0998.13453)

96. Padial JM, Miralles A, De la Riva I, Vences M. 2010 The integrative future of taxonomy. *Front Zool* **7**, 1–14.

97. Mammola S. 2019 Finding answers in the dark: caves as models in ecology fifty years after Poulson and White. *Ecography* **42**, 1331–1351. (doi:10.1111/ecog.03905)

98. Matthews EF, Abrams KM, Cooper SJB, Huey JA, Hillyer MJ, Humphreys WF, Austin AD, Guzik MT. 2020 Scratching the surface of subterranean biodiversity: Molecular analysis reveals a diverse and previously unknown fauna of Parabathynellidae (Crustacea: Bathynellacea) from the Pilbara, Western Australia. *Mol Phylogenet Evol* **142**, 106643. (doi:10.1016/j.ympev.2019.106643)

99. Mamaghani-Shishvan M, Esmaeili-Rineh S, Fišer C. 2017 An integrated morphological and molecular approach to a new species description of amphipods in the Niphargidae from two caves in West of Iran. *Zool Stud* **56**, 33. (doi:10.6620/ZS.2017.56-33)

100. Iliffe TM. 2002 Conservation of anchialine cave biodiversity. In *Hydrogeology and biology of post-paleozoic carbonate aquifers* (eds JB Martin, CM Wicks, ID Sasowsky), pp. 99–102. Charles Town: Karst Waters Institute.

101. DOF. 2010 NOM-059-SEMARNAT-2010. Protección al ambiental-Especies nativas de México de flora y fauna silvestres-Categorías de riesgo y especificaciones para su inclusión, exclusión o cambio-Lista de especies en riesgo. *Diario Oficial de la Federación*, 30 December. , 78.

102. Hutchins BT. 2018 The conservation status of Texas groundwater invertebrates. *Biodivers Conserv* **27**, 475–501. (doi:10.1007/s10531-017-1447-0)

103. Koenemann S, Bloechl A, Martínez A, Iliffe TM, Hoenemann M, Oromí P. 2009 A new, disjunct species of *Speleonectes* (Remipedia, Crustacea) from the Canary Islands. *Marine Biodiversity* **39**, 215–225. (doi:10.1007/s12526-009-0021-8)

104. White NE, Guzik MT, Austin AD, Moore GI, Humphreys WF, Alexander J, Bunce M. 2020 Detection of the rare Australian endemic blind cave eel (Ophisternon candidum) with environmental DNA: implications for threatened species management in subterranean environments. *Hydrobiologia* **847**, 3201–3211. (doi:10.1007/s10750-020-04304-z)

Funding sources

This work was supported by the Consejo Nacional de Ciencia y Tecnología (CONACYT - grant No. 472273), CONACYT- Texas A&M University (Grant No. 2018-044-1), institutional support from Texas A&M University Galveston, The Mohamed bin Zayed Species Conservation Fund (Grant No. 200523736), Texas A&M University San Antonio (start-up funds), T3: Texas A&M Triads for transformation (Grant No. 688), and the National Science Foundation (Grant No. 2136322).

AI declaration

No, we have not used AI-assisted technologies in creating this article

Conflicts of interest

We declare we have no competing interests.

Data Accessibility

DNA sequences are available on the BOLD (COI) and Genbank (COI & 16S; accessions PQ230796 - PQ230895, COI; PQ219243:PQ219290, 16S) databases. Supplemental material, including alignments and results of each species delimitation are openly available in the Zenodo repository at https://zenodo.org/records/10530586 doi: 10.5281/zenodo.10530586 (Calderón-Gutiérrez et al., 2024 [73]). All accession numbers are available on the Supplemental material 2. Raw chromatograms generated in this study for COI and 16S are available in the Zenodo repository (Calderón-Gutiérrez et al., 2024 [73]).

Supplemental Information

Suppl. 1 - Extended literature complementing the species checklist, and molecular data.

Suppl. 2 - Excel document with the taxonomy and state of the art of COI and 16S barcoding the fauna of the subterranean estuary of the Yucatan Peninsula, sample information and laboratory procedures, available information of barcodes in the literature, results of the species delimitation methods of *Antromysis, Creaseriella, Gorgonorhynchus, Metacirolana, Stygiomysis, Typhlatya* and *Xibalbanus.*

Suppl. 3 – Extended discussion by taxa and biogeographic distributions.

Figure Legends

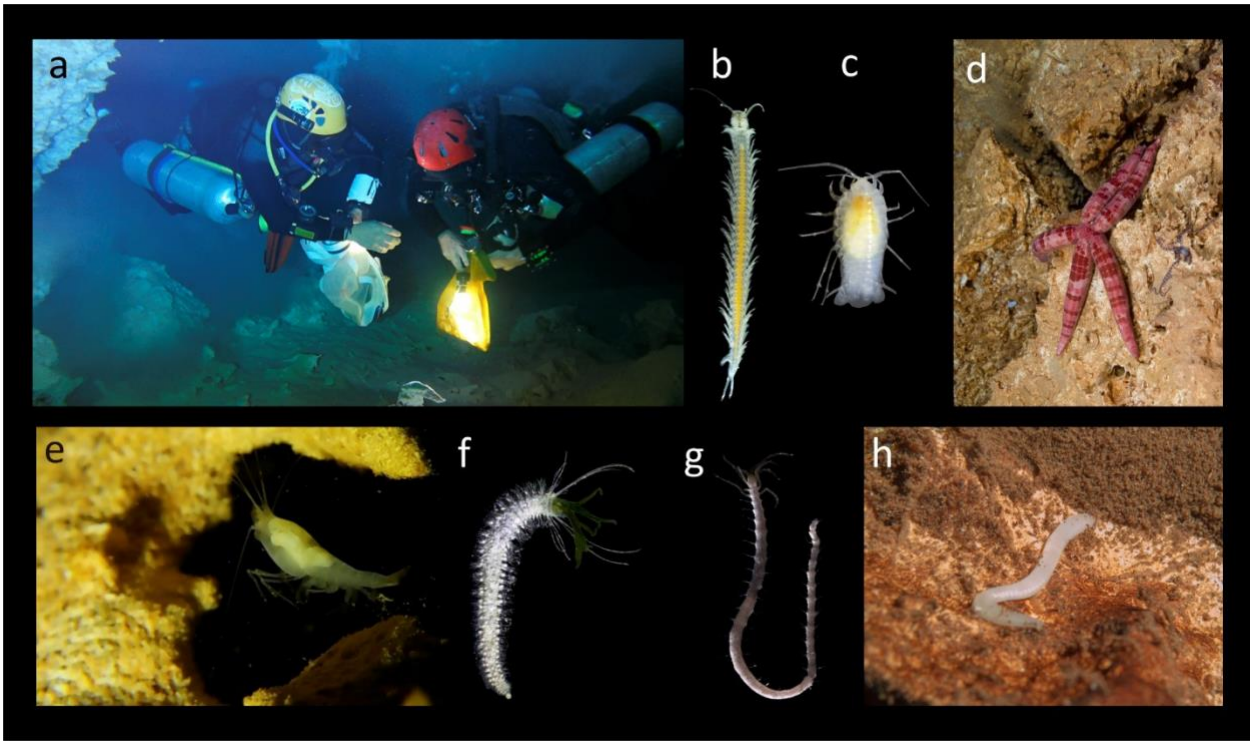
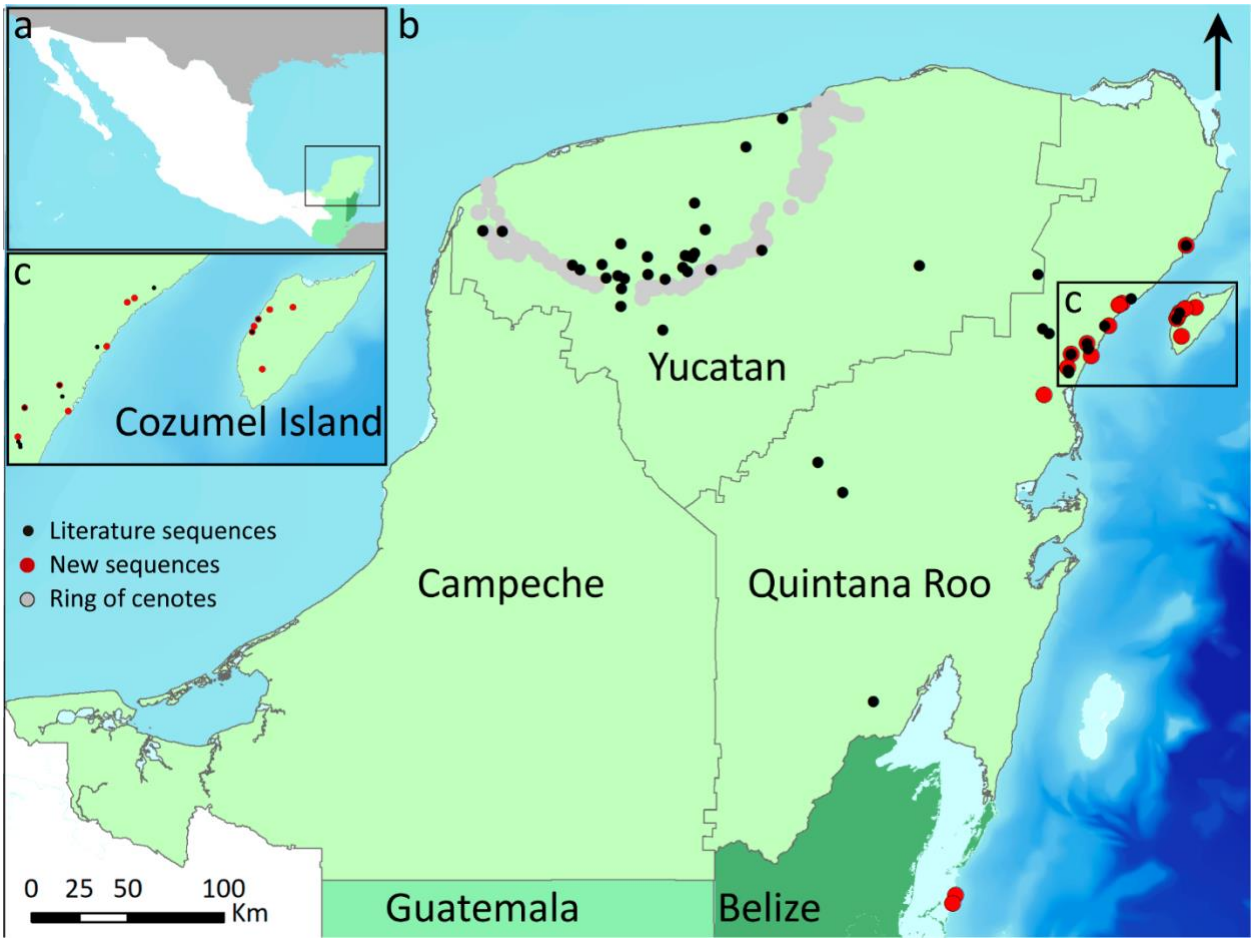
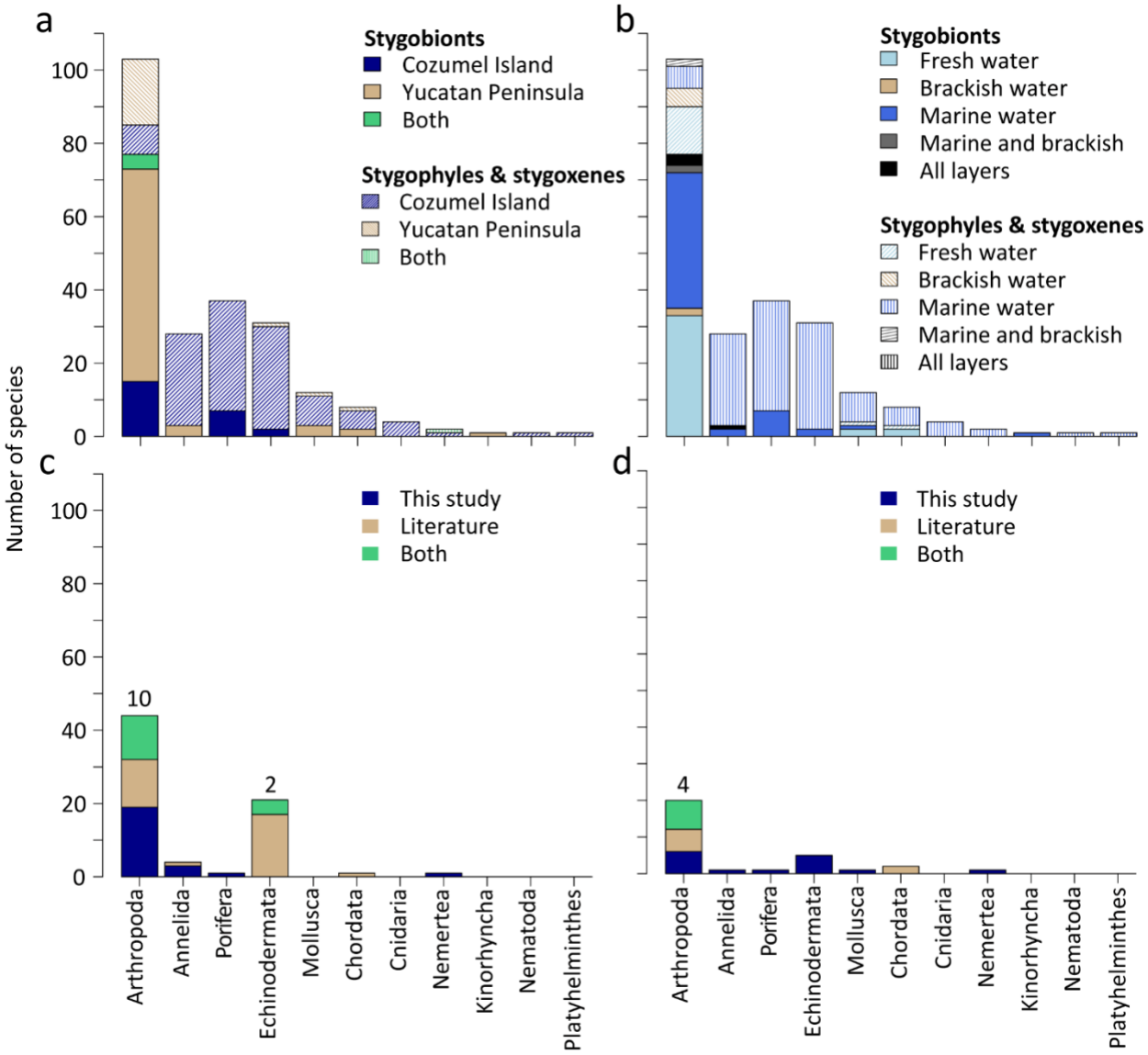


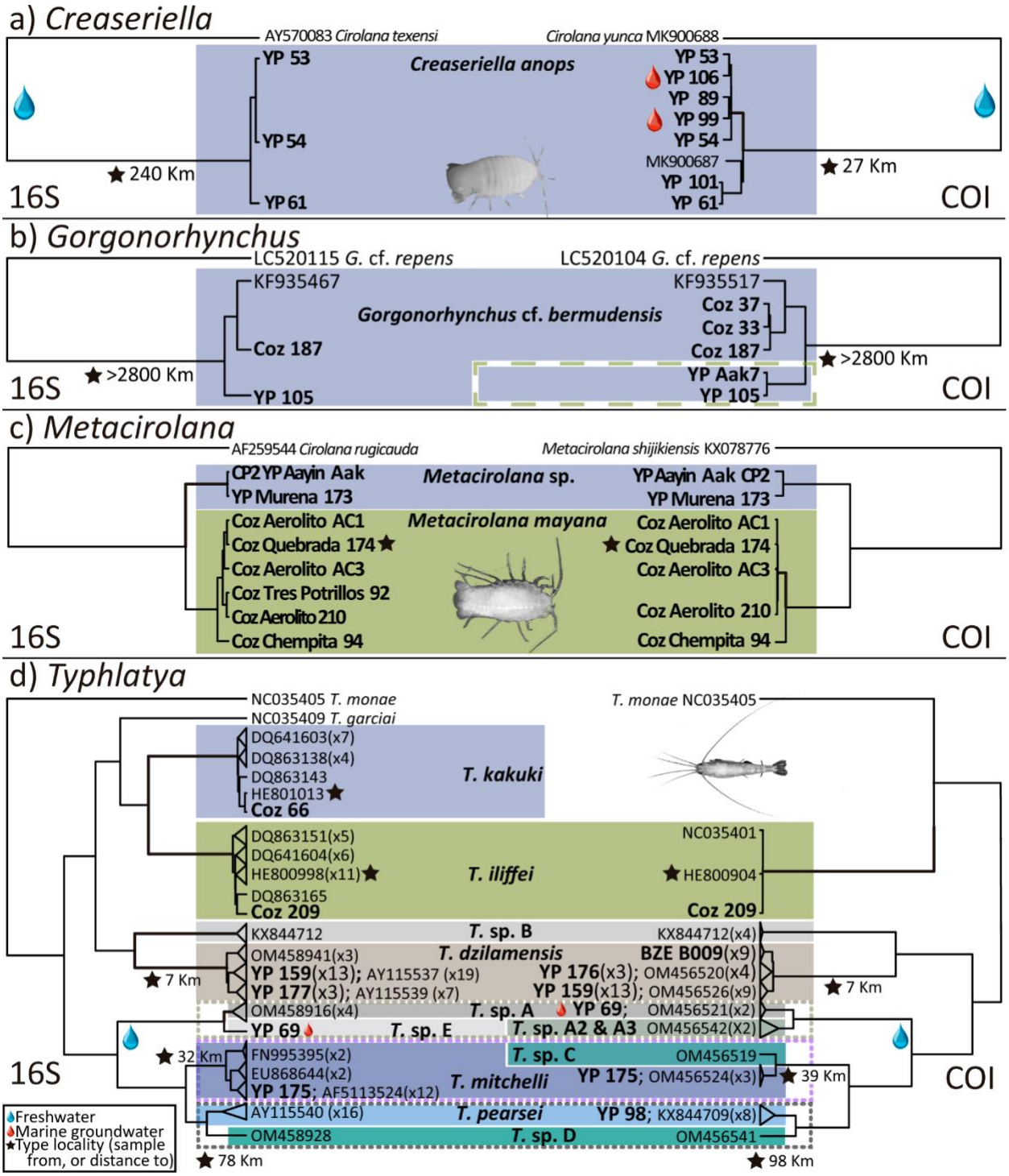
Figure 1.- Subterranean estuaries and their fauna. a) Diver sampling cave fauna. Representative stygobionts: Belize: b) *Xibalbanus tulumensis* (Yager, 1987). Cozumel Island: c) *Metacirolana mayana* (Bowman, 1987), d) *Copidaster cavernicola* Solis-Marin & Laguarda-Figueras, 2010. Yucatan Peninsula, Quintana Roo: e) *Typhlatya dzilamensis* Alvarez, Iliffe & Villalobos, 2005, f) Flabelligeridae, g) *Macrochaeta* sp., stygophile h) *Gorgonorhynchus* cf. *bermudensis*. Images a-e, h, F. Calderón-Gutiérrez, f-g, B. C. Gonzalez.



714 Figure 2.- Distribution of caves with barcoded subterranean species with COI and 16S from the literature
715 and newly sequenced in a) study area, b) Yucatan Peninsula and c) Cozumel Island. Species records
716 under a cumulative category (e.g. “Both”) are only counted once.



717 Figure 3.- Number of species in the subterranean estuary of Cozumel Island, and the Yucatan Peninsula
718 including Belize. Number of species recorded a) per region, and b) by water layer, c) species with COI
719 and d) 16S sequences available. Numbers above the bars (c & d), indicate species with sequences
720 available from the type locality within the study area.

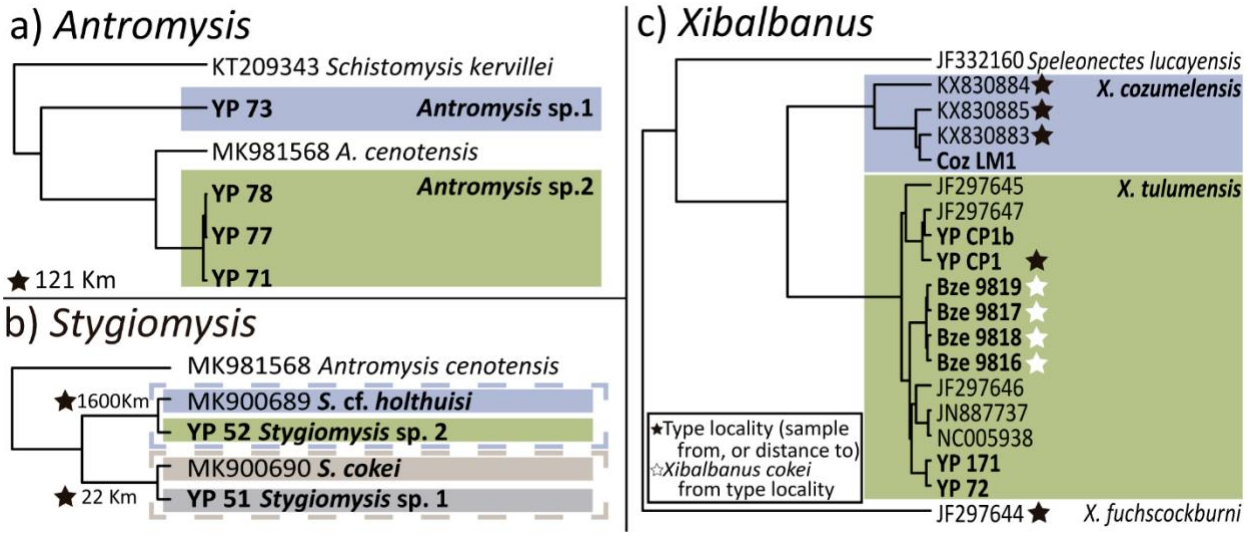


721 Figure 4.- Results of species delimitation for taxa with COI and 16S. a-d) Bayesian ultrametric and fully

722 bifurcating phylogenetic tree. ID-DNA name (bold and larger font) correspond to sequences new to this

723 study, with first 2-3 letters corresponding to the sampling region (BZE= Belize, Coz= Cozumel, YP=

724 Yucatan Peninsula), followed by sampling cave in (c). NCBI GenBank accession codes correspond to
725 sequences from the literature. Coloured boxes correspond to the proposed species designation
726 following a >50% consensus agreement among species delimitation methods; alternative results for
727 analyses with inconclusive results (i.e., a hypothesis was not supported by >50% of the analyses) are
728 represented with discontinuous lines; in d) identity considering published phylogenetic analyses with
729 mitochondrial and nuclear genes [10], and results from both genes are represented with dotted lines. d)
730 Grouped sequences have a representative ID, follow by the number of sequences in that group. Salinity
731 is displayed for lineages and specimens for species identified in this study as euryhaline. See Suppl. 2
732 and Calderón-Gutiérrez et al., 2024 [73] for further details.



733 Figure 5.- Results of species delimitation for taxa with COI. a-c) Bayesian ultrametric and fully bifurcating
734 phylogenetic tree. ID-DNA name (bold and larger font) correspond to sequences new to this study, with
735 first 2-3 letters corresponding to the sampling region (BZE= Belize, Coz= Cozumel, YP= Yucatan
736 Peninsula). Sequences from the literature are identified by their NCBI GenBank accession numbers.
737 Colored boxes corresponded to the proposed species designation following a >50% consensus
738 agreement among species delimitation methods, alternative results for analyses with inconclusive

739 results are represented with discontinuous lines. See Suppl. 2 and Calderón-Gutiérrez et al., 2024 [73]

740 for further details.