**Molecular Phylogenetics and Evolution**
February 2024, Volume 191, Pages 107969 (10p.)
https://doi.org/10.1016/j.ympev.2023.107969 https://archimer.ifremer.fr/doc/00863/97476/



# Coupling DNA barcodes and exon-capture to resolve the phylogeny of Turridae (Gastropoda, Conoidea)

Paul Zaharias [1, *], Kantor Yuri I. [1, 2], Fedosov E. Alexander [1, 3], Puillandre Nicolas [1]

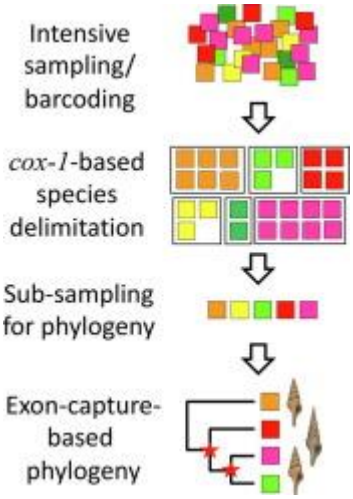
[1] Institut Systématique Evolution Biodiversité (ISYEB), Muséum National d'Histoire Naturelle, CNRS, Sorbonne Université, EPHE, Université des Antilles, 43 rue Cuvier, CP 51, 75005 Paris, France
[2] A.N. Severtsov Institute of Ecology and Evolution, Russian Academy of Sciences, Leninski prospect 33, 119071Moscow, Russian Federation
[3] Swedish Museum of Natural History, Box 50007, SE-104 05 Stockholm, Sweden 0851954209

* Corresponding author : Paul Zaharias, email address : paul.zaharias@mnhn.fr

**Abstract :**

Taxon sampling in most phylogenomic studies is often based on known taxa and/or morphospecies, thus ignoring undescribed diversity and/or cryptic lineages. The family Turridae is a group of venomous snails within the hyperdiverse superfamily Conoidea that includes many undescribed and cryptic species. Therefore 'traditional' taxon sampling could constitute a strong risk of undersampling or oversampling Turridae lineages. To minimize potential biases, we establish a robust sampling strategy, from species delimitation to phylogenomics. More than 3,000 cox-1 "barcode" sequences were used to propose 201 primary species hypotheses, nearly half of them corresponding to species potentially new to science, including several cryptic species. A 110-taxa exon-capture tree, including species representatives of the diversity uncovered with the cox-1 dataset, was build using up to 4,178 loci. Our results show the polyphyly of the genus Gemmula, that is split into up to 10 separate lineages, of which half would not have been detected if the sampling strategy was based only on described species. Our results strongly suggest that the use of blind, exploratory and intensive barcode sampling is necessary to avoid sampling biases in phylogenomic studies.

**Graphical abstract**




**Keywords** : phylogenomics, barcode, exon-capture, Conoidea, Turridae

## 1. Introduction

Recent advances in next-generation sequencing (NGS) considerably improved our understanding of phylogenetic relationships among living organisms, from the species level to the deepest branches of the tree of life. Indeed, our capacity to generate enough molecular data to resolve phylogenies, even difficult ones (e.g., in the case of rapid radiations – Van Damme et al., 2022) is no longer the main limiting factor. Robust phylogenies are now available for many taxa, which opened the door for phylogenetic comparative analyses that were until recently limited to a few model organisms such as vertebrates (e.g., Steppan and Schenk, 2017), angiosperms (e.g., Rose et al., 2018) or terrestrial arthropods (e.g., Legendre and Condamine, 2018).

What is now constituting the main difficulty in phylogenetic reconstructions is to sample adequately the diversity of the taxon of interest. This is particularly true in hyperdiverse, mostly poorly studied groups such as insects, annelids and molluscs, but even in some supposedly well-known groups (e.g., Cetaceans), for which recent phylogenies suggest new findings (Horreo, 2019). In a context where phylogenies are mandatory prerequisites to tackle evolutionary and biodiversity questions, it is important to remember that most phylogenetic comparative methods make the assumption that the input phylogenies are robust and complete (with a low proportion of missing taxa) – which is most often not the case. Consequently, phylogenetic comparative or diversification studies remain scarce in groups for which available phylogenies are known to be largely incomplete, such as marine gastropods (Williams and Duda 2008; Williams et al. 2013; Cunha et al. 2014; Postaire et al. 2014; Krug et al. 2015; Cunha et al. 2017; Abdelkrim et al. 2018; Phuong et al. 2019; Modica et al. 2020).

Taxon sampling is often incomplete in phylogenies for at least three reasons. First, it is sometimes impossible to gather all living species because the taxon is simply too diverse (e.g., Cabrero-Sañudo and Lobo, 2003). Second, most of the species remain unknown, and sample selection for phylogenetic reconstruction is often based on described species list, thus ignoring the undescribed diversity. Finally, described species often correspond to complexes of cryptic species, not necessarily closely related, and selecting samples to be included in a phylogeny based on morphospecies segregation can also lead to incomplete taxon sampling. To overcome these limitations, we propose a strategy in which the samples included in the phylogeny are selected based on a *de novo* species delimitation approach. In this approach, the barcode fragment of the mitochondrially encoded *cytochrome c oxidase I* (*cox-1*) gene is sequenced for all the available specimens, without any *a priori* sorting, either according to their identification or based on a morphospecies segregation. Consequently, specimens that would have been ignored either because they correspond to undescribed taxa or because they are morphologically undistinguished from others will be included in the final phylogeny. This strategy is similar to that of Puillandre et al., (2012) but extended from an alpha taxonomy purpose to both address both taxonomy and phylogenetics.

We apply this strategy to a family of marine gastropods, the Turridae (Conoidea), that includes 209 described species, but also many undescribed ones, and a large number of species complexes (e.g., Zaharias, Kantor, et al. 2020). The family Turridae initially enclosed the entire diversity of Conoidea excluding the cone snails

(Conidae) and auger shells (Terebridae), comprising 12 subfamilies (McLean, 1971). Morphological and anatomical considerations (Taylor, 1993) led to a new classification with most subfamilies elevated to the rank of family, with the Turridae now restricted to what earlier constituted the subfamily Turrinae. Our goal is to reconstruct a robust backbone phylogeny of Turridae without missing any deep lineages, and improve current estimates of the species diversity in the family. Thanks to recent efforts to clarify the alpha-taxonomy of the family (Abdelkrim et al., 2018a; Puillandre et al., 2017, 2012; Zaharias et al., 2020a), many *cox-1* sequences are available in public databases, to which we added original ones in order to increase species representation. We then used the species hypotheses obtained from the *cox-1* dataset to guide the selection of samples that will be included in a multigene phylogeny. Following an approach applied in related taxa (Abdelkrim *et al.*, 2018a; Phuong *et al.*, 2019) and proved to be efficient to resolve deep relationships (Zaharias et al., 2020b), we investigate the phylogeny of this group using an exon-capture approach.

## 2. Material and Methods

### *2.1 Species Hypotheses and Species Richness Estimations*

A total of 3,159 *cox-1* sequences were used to delimit Primary Species Hypotheses (PSHs) within Turridae (supp. Table ST1). Among them, 196 sequences come from online databases (NCBI and BOLD) while all other 2963 sequences (94%) were obtained from MNHN (Muséum national d'Histoire naturelle) specimens. Of the sequences obtained from MNHN specimens, 1505 were previously published in NCBI and/or BOLD (supp. Table ST1) and 1458 (46% of the total) *cox-1* sequences are newly published here. DNA from MNHN samples was extracted from small pieces of foot muscle by use of a Bioline Isolate II Genomic DNA extraction kit for animal tissue or using the Epmotion 5075 robot (Eppendorf), following the standard procedure of the manual. The barcode fragment (658 bp) of the mitochondrial *cox-1* gene was amplified using the universal primers LCO1490/HCO2198 (Folmer et al., 1994). Polymerase chain reactions (PCRs) were performed using a previously well-established protocol (Puillandre et al., 2017). Sequences were manually aligned with no ambiguity in the resulting alignment, as no insertion is known in Turridae in this mitochondrial region. We used ASAP (Puillandre et al., 2020) in default mode with a K80 model of substitution to delimit PSHs. ASAP (Assemble Species by Automatic Partitioning) is a distance-based species delimitation approach that uses pairwise genetic distances. Since it does not rely on a phylogeny, it allows the detection of species even if they are not monophyletic. While relying solely on genetic distances should be approached with caution, it has been demonstrated that ASAP are robust methods for species delimitation and are now widely used. Singleton (PSH with one individual) and doubleton (PSH with two individuals) delimited PSHs were checked to ensure that they were not artefacts generated by poorly cleaned sequences or missing data, by reconstructing a Maximum Likelihood (ML) tree using RAxML-NG (Kozlov et al., 2019) with a GTR+G model of substitution and 1,000 non-parametric bootstraps. Edges leading to singleton and doubleton PSHs were manually checked and if the edge was abnormally long and the associated sequence(s) contained lots of missing data or were sourced from

GenBank (making it impossible to check chromatograms), they were removed from the dataset to avoid an overestimation of the number of PSHs.

Additionally, we also ran Multi-rate Poisson tree processes (mPTP; Kapli et al., 2017), a single-locus species delimitation method under maximum likelihood that is phylogeny-based. The mPTP method counts the number of substitutions per branch and estimates the rates of branching events to detect which parts of the tree follow a speciation model (interspecific) and which follow a coalescent model (intraspecific). As for ASAP results, suspicious splits were corrected manually using the tree topology and branch lengths as well as the existing literature on species complexes (see Results section).

To estimate the total Turridae diversity, we used the Chao1 estimator (Gotelli and Chao, 2013), that is a function of the observed species richness (i.e., the number of PSHs), the number of singletons and the number of doubletons. The sampling fraction $\rho$ is defined as the ratio of sampled species (i.e., the total number of PSHs) over the total extant species diversity estimated using the Chao1 estimator. Because of sampling disequilibrium between genera (e.g., some geographical regions or some depth ranges are more sampled than others), we also estimated the species richness for each genus separately and then total species richness was estimated by summing species richness estimations across all genera.

### *2.2 Bait Design, Sampling, Library Preparation, Hybridization and Sequencing*

A target sequencing approach was used to recover exons for phylogenetic inference. First, 22 original (see supp. Table ST2) and 2 published Turridae transcriptomes (SRR1574907, SRR1574923) as well as 3 conoidean outgroups (1 Pseudomelatomidae – SRR1574922 and 2 Terebridae – SRR2059225, SRR2060990) were analyzed to recover a set of orthologous genes following the pipeline from (Zaharias et al., 2020b) and using *Lottia gigantea* (Simakov et al., 2013) as a reference-genome. RNA was preserved in RNA later and extracted using a Trizol protocol or the Qiagen RNeasy Micro kit (see supp. Table ST2), following the manufacturer's recommendations. Bioanalyzer traces were used to assess total RNA quality and determine suitability for sequencing. The cDNA libraries were prepared and sequenced either at the New York Genome Center or at the Evolutionary Genetics Lab at UC Berkeley. In New York, libraries were prepared using the automated polyA RNAseq library prep protocol and sequenced with Illumina HiSeq 4000 with 150-bp paired-end reads. In Berkeley, the KAPA Stranded mRNA-Seq kit was used to synthesize cDNA, ligate adapters using TruSeq HT adapters and barcode samples. Samples were then sequenced with Illumina HiSeq 2000 or 4000 with 100-bp paired-end reads. Transcriptomes were assembled following the exact pipeline of Zaharias et al., (2020b). Exon-boundaries were denoted by comparing the transcriptome sequences to the *L. gigantea* genome. Only exons > 120 bp were retained. Ancestral sequences were generated for each exon alignment using FastML v3.1 (Ashkenazy et al., 2012) from the following species pairs: (i) *Terebra subulata* (SRR2059225) OR *Cinguloterebra anilis* (SRR2060990) + any of the 24 Turridae (ii) *Crassispiria cerithina* (SRR1574922) + any of the 24 Turridae and (iii) *Gemmula congener* OR *Turridrupa jubata* + any of the remaining 22 Turridae. Then, loci with a GC content <30% or >70% were excluded because extreme GC contents can

reduce capture efficiency (Bi et al., 2012). Loci with repeats identified through the RepeatMasker v4.0.6 web server (Smit et al., 2015) were also removed. Then, we performed a self-blast with the target sequences via blastn v2.2.31 (evalue = 1e-10) and removed loci that did not blast to themselves with sequence identity>90%. A total of 19,427 sequences were sent to Mycroarray (Ann Arbor, MI) to synthesize a custom MYbaits-1 kit, comprising 19,966 bait sequences of 120 bp with ~54bp spacing (~2.2x tiling density). Overall, the bait design targeted about 500,000 bp, corresponding to 4,204 loci.

Specimens sequenced with the exon-capture approach were selected to represent all supported lineages in the *cox-1* ML tree (supp. Fig. SF1) and most identified PSHs. Particular attention was paid to sampling *Gemmula* lineages, because of its long-time suspected polyphyly (Puillandre et al., 2012). Most specimens in the final exon-capture dataset come from the MNHN collection, except *G. hindsiana* (Museum of Comparative Zoology, Harvard), *Polystira picta* and *G. deshayesii* (Natural History Museum of the UK, London) (supp. Table ST1 and ST3). In total, 172 samples were selected, and divided into three batches of 60, 60 and 52 specimens, respectively.

DNA was extracted using the E.Z.N.A.® Mollusc DNA Kit (Omega Bio-tek, Norcross, GA, USA), following the manufacturer's recommendations. Library preparation followed Meyer and Kircher (2010) with some minor modifications (see Abdelkrim et al. 2018). Resulting libraries were quantified and qualified through qPCR and fluorometry (Qubit). Depending on the library concentrations, 5–15 cycles of indexing PCR were conducted. Amplification started only after 20 cycles of indexing PCR in most of the batch 3 libraries. After cleanup and quantification, equal amounts (~120 ng) of indexed libraries were pooled by 10 samples (batches 1 and 2) or 13 samples (batch 3).

Capture was conducted following the MyBaits protocol v3.0 with a few modifications. As recommended, between 100 and 500 ng of each pool was used for the capture. Blocking Oligonucleotides from Integrated DNA Technologies were used instead of those provided by xGen. Capture was conducted for 24h at 60°C on a BIO-RAD C1000 touch thermal cycler. Post-capture libraries were cleaned-up following MyBaits protocol and quantified using Qubit. Each pool library was amplified through a PCR reaction of 10 cycles. At this stage, capture success was assessed using genes that could be amplified through PCR and that were supposed to be captured by our baits (positive controls) or not (negative controls). Controls were amplified using specific primers using qPCR prior to and after capture, expecting that the number of cycles needed for the PCR to reach the threshold cycle will be reduced in the case of positive controls, while delayed or no amplification will be observed in the negative controls. Finally, an Agilent 2100 Bioanalyzer was used for library characterization for each library of 10/13 samples. Each batch was sequenced on one lane of Illumina HiSeq 4000 paired-end (100 bp reads) (QB3, Berkeley, USA).

### *2.3 Data filtering and exon-assembly*

We trimmed reads for adapter contamination and quality using Trimmomatic v0.36 (Bolger et al., 2014), with ILLUMINACLIP option enabled, seed mismatch threshold = 2, palindrome clip threshold = 40, simple clip threshold of 15; SLIDING WINDOW option enabled, window size = 4, quality threshold = 20; MINLEN = 36; LEADING = 15; TRAILING = 15 and used flash v1.2.11 (Magoč and Salzberg, 2011) to merge mate reads. We generated assemblies using SPAdes v3.8.1 (Bankevich et al., 2012) and used cap3 (Huang and Madan, 1999) and cd-hit v4.6.5 (percent identity = 99%) to reduce redundancy in the assemblies. Ten Turridae transcriptomes (see supp. Table ST2) were added to the exon-capture bioinformatic pipeline, as well as the transcriptomes of three outgroups; *Mitra mitra* (Mitridae, SRR21742594), *Conus tessulatus* (Conidae, unpublished) and *Terebra subulata* (Terebridae, SRR2059224). We also added raw reads of 10 Turridae samples (MNHN-IM-2009-29350, MNHN-IM-2009-29322, MNHN-IM-2009-19093, MNHN-IM-2013-12760, MNHN-IM-2013-52064, MNHN-IM-2013-52063, MNHN-IM-2013-20432, MNHN-IM-2013-43303, MNHN-IM-2013-51290, MNHN-IM-2013-12759) from the exon-capture study of Abdelkrim et al., (2018b). We used blastn v2.2.31 (evalue = 1e-10, word size = 11) to associate contigs with the targeted loci. Then, we used EXONERATE v2.4.0 (Slater and Birney, 2005) under the est2genome model to redefine exon boundaries in our target sequences because many of the originally predicted exons were actually composed of several smaller exons. To estimate average heterozygosity, we mapped reads using bowtie2 v2.2.7 (Langmead and Salzberg, 2012) with the very sensitive local and no discordant options, marked duplicates using picard-tools v2.1.1 (http://broadinstitute.github.io/picard), and called single nucleotide polymorphisms (SNPs) using samtools v1.3 and bcftools v1.3 (Li et al., 2009). For each sample, we removed sequences if: 1) estimated heterozygosity was >2 SDs away from the mean, 2) they blasted to multiple reference targets, 3) less than 70% of the sequence full length had a coverage of 4X or more. Alignments for each locus were generated using mafft v7.222 (Katoh and Standley, 2013). Finally, custom python scripts were used on aligned loci for additional filtering. Alignments with ≥ 40% missing data were discarded. Sequences shorter than 50 nucleotides were removed. The 3' and 5' ends were trimmed using trimAl (Capella-Gutiérrez et al., 2009) if they were represented by less than 50% of the sequences.

### *2.4 Phylogenomic analysis*

Two sets of loci were used for phylogenetic inference, depending on taxon occupancy, i.e. the proportion of missing taxa in a given locus (Sayyari et al., 2017): DS1 with loci containing at least 4 taxa/locus and DS2 with loci containing at least 50 taxa/locus. Each dataset was then concatenated into a single matrix on which a maximum likelihood analysis was performed using IQ-TREE (Nguyen et al., 2015), with a GTR+F+R6 model selected using ModelFinder (Kalyaanamoorthy et al., 2017). We then applied 1,000 ultrafast bootstraps (UFBoot) (Hoang et al., 2018) on each dataset to obtain branch support. Individual trees for each locus were also generated for each dataset using IQ-TREE and the associated best substitution model found with ModelFinder following a BIC criterion. Then, the sets of trees for both datasets were used to infer a phylogeny using the coalescent-based approach implemented in ASTRAL-III (Zhang et al., 2018). These four analyses will be referred as DS1-AS, DS1-IQ, DS2-AS and DS2-

IQ. To account for extreme evolutionary rates, we also build a third dataset: following (Cunha and Giribet, 2019), the 10% slowest and the 20% fastest evolving loci of the DS2 dataset were identified using trimAl and removed. For this dataset (DS3), we inferred two trees using a concatenated (DS3-IQ) and a summary coalescent approach (DS3-AS) using the same parameters as for DS1 and DS2.

## 3. Results

### *3.1 Delimited PSHs and species richness*

The workflow and results of this study are summarized in a flowchart (Fig. 1). The best partition found by ASAP over the 3,159 *cox-1* sequences consisted of 207 PSHs (Fig. 2; see also ST1 for details). A careful inspection of the singleton and doubleton PSHs in the *cox-1* ML tree and checking the sequences led to the removal of 6 singletons (4 sequences from GenBank and 2 MNHN sequences with missing data), thus decreasing the final number of PSHs to 201. The mPTP output resulted in 206 delimited species, reduced to 196 after manually checking sequences or groups of sequences leading to suspicious splits. Partitioning conflict between the two methods is colored in yellow in ST1. Overall, ASAP and mPTP agree on 164 partitions, showing congruence between two very different methods. Contrarily to what the results may seem like, there are not significantly more cases of lumping with mPTP than there is with ASAP, but a few cases of over-lumping in mPTP (e.g., in *Lucerapex*, see ST1) explain the difference between the two methods in the final set of delimited species. We compared our ASAP PSHs with the corresponding species previously delimited and confirmed as Secondary Species Hypotheses (SSH) in four integrative taxonomy studies (Abdelkrim et al., 2018a; Puillandre et al., 2017, 2010; Zaharias et al., 2020a) (Table 1). We retrieved 23 of the 31 SSHs: one SSH (*X. legitima*) was split into two PSHs and 3 groups of SSHs were lumped, in all cases corresponding to closely related species. Our results thus generally agree with the SSHs found in previous studies and can thus be considered as robust proxies for species. For putative undescribed species we followed the PSH numbering from Puillandre et al. (2012). A total of 168 specimens, representative of the *cox-1* phylogenetic diversity (as determined using the ML tree from supp. Fig. SF1), were selected for the exon-capture dataset. The PSHs not represented in the exon-capture dataset are in all cases closely related to the PSHs included in the exon-capture dataset.

The Chao1 estimator calculated over the whole family is 220.8, suggesting a sampling fraction $\rho$ of 91.0%. The Chao1 estimator calculated by summing Chao1 estimators for each genus is of 231.8, suggesting a $\rho$ = 86.7%.

### *3.2 Targeted sequencing data*

On average, 5,477,758 reads per specimen were sequenced, with a high heterogeneity among specimens (from 2,764 to 40,868,381 reads, see supp. Table ST3). Moreover, some specimens did not pass the cleaning and filtering pipeline, despite having a non-

negligible number of reads, sometimes due to a high proportion of duplicate reads (see ST3). In particular, only five samples from the batch3 (out of 52) passed the exon-cleaning pipeline, probably because of the high number of indexing PCR cycles: many duplicates were found for most samples, possibly indicating that only a few targets were over-amplified during the indexing PCR step. We also removed multiple conspecific transcriptomes to keep one sample per PSH only. Our final sampling for the exon-capture phylogeny includes 110 Turridae PSHs, representing 49.8% (110/220.8) of the PSHs delimited with the *cox-1* dataset, and including most of the main lineages, with the exception of *Gemmula* sp. 27, a deep lineage, for which placement is not supported in the *cox-1* tree. We also note the poor results for the important specimen *Gemmula hindsiana* (type species of the polyphyletic *Gemmula*), for which only 14 loci were recovered. Overall, for these 110 PSHs, we recovered 4,180 targeted loci, with an average of 49 PSHs per locus.

### *3.3 Phylogeny*

Statistics for the five phylogenetic analyses are summarized in Table 2. Overall, the relationships remained very stable among all datasets. Most differences either concern a few short branches (e.g., the sister group of the "Nov. gen. A" clade), or the placement of a few samples with a lot of missing data (e.g., *Gemmula hindsiana*). The main difference between the DS3-IQ topology and the DS1-IQ / DS2-IQ topologies is the position of *Gemmula hindsiana* (see supp. Fig. SF3), either retrieved as sister to *G. closterion* (DS1-AS, DS2-AS, DS3-IQ) or as a singleton (DS1-IQ, DS2-IQ, DS3-AS). The overall support in the summary trees is lower than for the concatenated trees (Table 2), especially for the short branches (see also Fig. 3 and supp. Fig. SF3).

## 4. Discussion

### *4.1 Sampling strategy*

In most phylogenomic studies of marine gastropods (e.g., Cunha et al., 2022; Karmeinski et al., 2021), taxon sampling is based on known taxa and/or morphospecies, thus ignoring undescribed diversity and/or cryptic lineages. This is the major issue for phylogenies which aim at a full species-level sampling, but even in the case of backbone phylogenies, cryptic species can be an issue, since they are not always closely related to known ones (e.g., some of the *Gemmula* PSHs identified here) and might represent highly divergent lineages. Indeed, in particular in hyperdiverse and mostly unknown groups such as the superfamily Conoidea that includes the Turridae, taxonomic studies have revealed morphological convergences between species that diverged tens of millions of years ago (Abdelkrim et al., 2018b; Kantor et al., 2018). The use of an exploratory approach, consisting of sequencing all the available samples without any *a priori* sorting, revealed 201 PSHs (estimated 221-232 using Chao1) among the 3,159 *cox1* sequences available (including 1,458 original). This number nearly matches the number of described species (209) but needs to be put into perspective. Indeed, only for 110 of the 201 PSHs we were able to attribute a unique

species name (see supp. Table ST1), suggesting that only 55% of the PSHs are described. Therefore, a sampling based on valid taxa only would have missed 45% of the Turridae diversity sampled here, and a sampling based on morphospecies *a priori* delimited would have missed up to five (Nov. gen. A, B, G, H, I) of the independent "*Gemmula*-like" lineages (see below). A blind, exploratory, approach is thus mandatory to improve taxon sampling in such groups that contain many undescribed and/or cryptic species, and guarantee that the reconstructed phylogenies better represent the actual diversity. Indeed, phylogenetic trees are now mandatory prerequisites to tackle many evolutionary questions (e.g., comparative studies) and are also the main source of information for classifications revisions.

However, if this strategy overcomes biases when selecting among the available samples those that will be included in a phylogeny, it still lacks the taxa that have not been sampled in the field that could represent described or undescribed species. While our 'blind' approach is the result of 15 years of the 'Our Planet Reviewed' and the 'Tropical Deep-Sea Benthos' program, it still depends on financial means and sampling authorizations from diverse territories. While our sampling covers a wide diversity of tropical habitats (coral reefs, continental slope, thermal vents, rocky substrate, sandy bottoms, etc.), broad depths range (intertidal to bathyal zone) and multiple tropical geographic regions (extending sometimes beyond tropics), we are still missing some key regions (e.g., the Arabian Sea) or environments (abyssal plains). Our estimations of the total Turridae diversity suggest that several tens of species still remain to be collected, which is probably an underestimation given that most of the seafloor has never been sampled.

Another important bias is that *cox1* is a single (mitochondrial) marker, and although it has shown to be a good proxy for species delimitation, examination of an independently evolving nuclear marker is strongly recommended to corroborate the *cox1*-based species hypotheses. Design of new barcode markers is needed in the malacological community in order to make the blind approach presented in this study more robust to species delimitation biases.

### *4.2 Towards a new classification of the Turridae*

We were able to sequence one or several samples for 13 out of the 15 extant and valid genera listed in WoRMS (Table 3; Jul/2023). The two missing genera, *Epidirella* and *Decollidrillia*, should be further studied with molecular data. In most cases, the genera as defined in WoRMS (Costello et al., 2013) are found as clades in all phylogenies: the monophyly of *Gemmuloborsonia*, *Lucerapex*, *Polystira*, *Cryptogemma*, *Unedogemmula*, *Lophiotoma* and *Iotyrris* is confirmed. Furthermore, the genera represented by only one species in the trees are confirmed as independent lineages (*Kuroshioturris* and *Xenuroturris*). The two independent clades referable to *Turris* (*Turris* and '*Annulaturris'*/*Purpuraturris*) recovered by Fedosov et al. (2011) and confirmed by Chase et al. (2022) are also recovered as unrelated. The valid name and scope of these lineages will be discussed in an upcoming paper that will revise the generic classification of the Turridae. The species *Ptychosyrinx chilensis* Berry, 1968, formerly attributed to the genus *Ptychosyrinx* and recently synonymized with

*Cryptogemma* (Zaharias et al., 2020a), should be described as a new, monospecific genus. The major outcome of this phylogeny is the confirmation of the polyphyly of the genus *Gemmula* (Fig. 3; Table 3), which is split into 10 independent lineages (Nov. gen. A-J). We suggest that half of these lineages (Nov. gen. A, B, G, H, I) would not have been detected without our sampling strategy because we couldn't assign any existing name to any of the PSHs constituting them. These lineages are almost exclusively deep-water (more than 100 meters deep) and were retrieved mostly because of the sampling effort of the Tropical Deep-Sea Benthos program (Fig. 1). In addition to the difficulty of accessing these samples, the majority of the PSHs within these new lineages are morphologically very indistinct, making it difficult for taxonomists to recognize them as new species.

In the topology obtained with DS2-IQ (Fig. 3), *G. hindsiana*, the type species of the genus, does not cluster with any other species, but in the topologies obtained with DS3-IQ and the summary datasets, it is the sister species of *G. closterion* (although with low support). Pending on more data, especially for *G. hindsiana* (for which only 14 loci were recovered), we remain conservative and consider *G. closterion* as a true *Gemmula* as well. A new classification of the Turridae, based on the phylogenetic relationships inferred here and including the description of new genera, requires a considerable effort that extends beyond the scope of the present study and will be published in a separate article.

Another difficulty is the attribution of available species names to the PSHs, and eventually the description of new species: as illustrated for *Cryptogemma* (Zaharias et al., 2020a), *Xenuroturris*/*Iotyrris* (Abdelkrim et al., 2018a) and *Lophiotoma* (Puillandre et al., 2017), these steps are very time consuming and can be problematic in the case of unpreserved or lost name-bearing specimens. Therefore, species descriptions will be tackled in separate articles.

## 5. Conclusion

Careful sampling is the basis of any scientific study. In this study, our blind, exploratory and intensive barcode sampling showed great success at uncovering hidden diversity, with 11 potential new genera marine snails identified. In our case, a lot of this hidden diversity corresponds to deep-water and 'boring' shells for malacologists (small, grey and poorly ornamented specimens), which could explain why less descriptive taxonomic effort was put in those new lineages. We suggest that our blind sampling approach should be used more systematically, particularly in the hyperdiverse groups of marine invertebrates. Despite this effort, we suspect that many species could be unsampled because some major biogeographic regions remain undersampled, particularly in the deep sea, emphasizing the need for more expeditions to complete our knowledge of biodiversity.

Our barcode-sampling strategy led to the first robust backbone phylogeny of Turridae, which can now serve as a strong and reliable basis for taxonomic and evolutionary studies.

**Data availability**

The scripts used for the exon-capture post-sequencing processing are derived from https://github.com/markphuong/evolvability.phylogenomics.

All the *cox-1* sequences will be submitted to BOLD and GenBank upon acceptance of the manuscript for publication.

All the transcriptomes will be submitted to SRA (NCBI) upon acceptance of the manuscript for publication.

All the exon-capture data, alignments and trees will be made available through the French platform InDoRES (https://www.indores.fr/) upon acceptance of the manuscript for publication.

**Fundings**

This work was supported by the CONOTAX project funded by the French Agence Nationale pour la Recherche (ANR-13-JSV7-0013-01) and by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 865101) to N.P.

**Acknowledgments**

Specimens were obtained during research cruises and expeditions organized by the MNHN and ProNatura International as part of the Our Planet Reviewed program, and by the MNHN and the Institut de Recherche pour le Développement as part of the Tropical Deep-Sea Benthos program (TAIWAN 2004, TAIWAN 2013, DongSha 2014, NanHai 2014 and ZhongSha 2015 in the South China Sea; PANGLAO 2004, PANGLAO 2005 and AURORA 2007 in the Philippines; BIOPAPUA (dx.doi.org/10.17600/10100040), KAVIENG 2014 (dx.doi.org/10.17600/14004400), MADEEP (dx.doi.org/10.17600/14004000) and PAPUA NIUGINI (dx.doi.org/10.17600/18000841) in Papua New Guinea; CONCALIS (dx.doi.org/10.17600/8100010), EBISCO (dx.doi.org/10.17600/5100080), EXBODI (dx.doi.org/10.17600/11100080), KANACONO (dx.doi.org/10.17600/16003900), KANADEEP (dx.doi.org/10.17600/17003800), KOUMAC 2.1, KOUMAC 2.3, NORFOLK 2 (dx.doi.org/10.17600/3100030) and TERRASSES (dx.doi.org/10.17600/8100100) in New Caledonia; BOA1 (dx.doi.org/10.17600/5100060) and SANTO 2006 in Vanuatu, SALOMON 2 (dx.doi.org/10.17600/4100090) and SALOMON BOA3 (dx.doi.org/10.17600/7100070) in the Solomons, MUSORSTOM 9 (dx.doi.org/10.17600/97100020), PAKAIHI I TE MOANA and TARASOC (dx.doi.org/10.17600/9100040) in French Polynesia, ATIMO

VATAE (dx.doi.org/10.17600/10110040) and MIRIKY in Madagascar; BIOMAGLO (dx.doi.org/10.17600/17004000) in Mayotte and Glorieuses, INHACA 2011 and MAINBAZA in Mozambique, CORSICABENTHOS 1 in Corsica, KARUBENTHOS 2012 and KARUBENTHOS 2 (dx.doi.org/10.17600/15005400) in Guadeloupe, MADIBENTHOS in Martinique, GUYANE 2014 in French Guiana), and we are grateful to numerous cruise leaders and co-PIs: Wei-Jen Chen, Philippe Bouchet, Marivene Manuel Santos, Laure Corbari, Stephane Hourdez, Ludivina Labe, Peter Kee Lin Ng, Karine Olu-Le Roy, Bertrand Richer de Forges and Sarah Samadi (more information can be found at expeditions.mnhn.fr). Additional specimens were collected during 2011 field work organized in Western Australia by Hugh Morrison, with the support of the Western Australian Museum, during the 2017 field work in Nha-Trang Bay (Vietnam), supported by the joined Russian-Vietnamese Tropical Research and Technological center, and during the 2013 fieldwork in the Philippines associated with the Conus-Turrid project of the Marine Science Institute (MSI, UP Diliman). These expeditions operated under the regulations then in force in the countries in question and satisfy the conditions set by the Nagoya Protocol for access to genetic resources. Additional samples were provided by John Todd (Natural History Museum of the UK, London), Gonzalo Giribet (Museum of Comparative Zoology, Harvard), Alexander Sysoev (Zoological Museum of Moscow State University), Peter Stahlschmidt and Yasunori Kano (University of Tokyo). We thank Laetitia Aznar-Cormano for her help with the RNA extraction and Mark Phuong, Juliette Gorson and Mandë Holford for their help in the RNA sequencing. The analyses were performed on the Plateforme de Calcul Intensif et Algorithmique PCIA (UAR2700 2AD, MNHN). We also thank Barbara Buge for her help in curating the specimens and Mark Phuong and Jawad Abdelkrim for their help with bioinformatics scripts.

## References

Abdelkrim, J., Aznar-Cormano, L., Buge, B., Fedosov, A., Kantor, Y., Zaharias, P., Puillandre, N., 2018a. Delimiting species of marine gastropods (Turridae, Conoidea) using RAD sequencing in an integrative taxonomy framework. Molecular Ecology 27, 4591–4611. https://doi.org/10.1111/mec.14882

Abdelkrim, J., Aznar-Cormano, L., Fedosov, A.E., Kantor, Y.I., Lozouet, P., Phuong, M.A., Zaharias, P., Puillandre, N., 2018b. Exon-Capture-Based Phylogeny and Diversification of the Venomous Gastropods (Neogastropoda, Conoidea). Mol Biol Evol 35, 2355–2374. https://doi.org/10.1093/molbev/msy144

Ashkenazy, H., Penn, O., Doron-Faigenboim, A., Cohen, O., Cannarozzi, G., Zomer, O., Pupko, T., 2012. FastML: a web server for probabilistic reconstruction of ancestral sequences. Nucleic Acids Res 40, W580–W584. https://doi.org/10.1093/nar/gks498

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-

Cell Sequencing. Journal of Computational Biology 19, 455–477. https://doi.org/10.1089/cmb.2012.0021

Bi, K., Vanderpool, D., Singhal, S., Linderoth, T., Moritz, C., Good, J.M., 2012. Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. BMC Genomics 13, 403. https://doi.org/10.1186/1471-2164-13-403

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Cabrero-SaÑudo, F.J., Lobo, J.M., 2003. Estimating the number of species not yet described and their characteristics: the case of Western Palaearctic dung beetle species (Coleoptera, Scarabaeoidea). Biodiversity and Conservation 12, 147–166. https://doi.org/10.1023/A:1021247705018

Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Chang, Z.-Y., Liew, T.-S., 2021. A molecular phylogeny of Geotrochus and Trochomorpha species (Gastropoda: Trochomorphidae) in Sabah, Malaysia reveals convergent evolution of shell morphology driven by environmental influences. PeerJ 9, e10526. https://doi.org/10.7717/peerj.10526

Chase, K., Watkins, M., Safavi-Hemami, H., Olivera, B.M., 2022. Integrating Venom Peptide Libraries Into a Phylogenetic and Broader Biological Framework. Front Mol Biosci 9, 784419. https://doi.org/10.3389/fmolb.2022.784419

Costello, M.J., Bouchet, P., Boxshall, G., Fauchald, K., Gordon, D., Hoeksema, B.W., Poore, G.C.B., Soest, R.W.M. van, Stöhr, S., Walter, T.C., Vanhoorne, B., Decock, W., Appeltans, W., 2013. Global Coordination and Standardisation in Marine Biodiversity through the World Register of Marine Species (WoRMS) and Related Databases. PLOS ONE 8, e51629. https://doi.org/10.1371/journal.pone.0051629

Cunha, R.L., Assis, J.M., Madeira, C., Seabra, R., Lima, F.P., Lopes, E.P., Williams, S.T., Castilho, R., 2017. Drivers of Cape Verde archipelagic endemism in keyhole limpets. Sci Rep 7, 41817. https://doi.org/10.1038/srep41817

Cunha, R.L., Lima, F.P., Tenorio, M.J., Ramos, A.A., Castilho, R., Williams, S.T., 2014. Evolution at a Different Pace: Distinctive Phylogenetic Patterns of Cone Snails from Two Ancient Oceanic Archipelagos. Syst Biol 63, 971–987. https://doi.org/10.1093/sysbio/syu059

Cunha, T.J., Giribet, G., 2019. A congruent topology for deep gastropod relationships. Proceedings of the Royal Society B: Biological Sciences 286, 20182776. https://doi.org/10.1098/rspb.2018.2776

Cunha, T.J., Reimer, J.D., Giribet, G., 2022. Investigating Sources of Conflict in Deep Phylogenomics of Vetigastropod Snails. Systematic Biology 71, 1009–1022. https://doi.org/10.1093/sysbio/syab071

D, T.J., 1993. Foregut anatomy, feeding mechanisms, relationships and classification of the Conoidea (=Toxoglossa) (Gastropoda). Bull. nat. Hist. Mus., London (Zoology) 59, 125–170.

Fedosov, A., Watkins, M., Heralde, F.M., Corneli, P.S., Concepcion, G.P., Olivera, B.M., 2011. Phylogeny of the genus Turris: Correlating molecular data with radular anatomy and shell morphology. Molecular Phylogenetics and Evolution 59, 263–270. https://doi.org/10.1016/j.ympev.2011.01.019

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. Molecular Marine Biology and Biotechnology 6.

Gotelli, N.J., Chao, A., 2013. Measuring and Estimating Species Richness, Species Diversity, and Biotic Similarity from Sampling Data, in: Levin, S.A. (Ed.), Encyclopedia of Biodiversity (Second Edition). Academic Press, Waltham, pp. 195–211. https://doi.org/10.1016/B978-0-12-384719-5.00424-X

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., Vinh, L.S., 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol Biol Evol 35, 518–522. https://doi.org/10.1093/molbev/msx281

Horreo, J.L., 2019. New insights into the phylogenetic relationships among the oceanic dolphins (Cetacea: Delphinidae). Journal of Zoological Systematics and Evolutionary Research 57, 476–480. https://doi.org/10.1111/jzs.12255

Huang, X., Madan, A., 1999. CAP3: A DNA Sequence Assembly Program. Genome Res. 9, 868–877. https://doi.org/10.1101/gr.9.9.868

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nature Methods 14, 587–589. https://doi.org/10.1038/nmeth.4285

Kantor, Yu.I., Fedosov, A.E., Puillandre, N., 2018. New and unusual deep-water Conoidea revised with shell, radula and DNA characters. Ruthenica 28, 47–82. https://doi.org/10.35885/ruthenica.2018.28(2).1

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. https://doi.org/10.1093/bioinformatics/btx025

Karmeinski, D., Meusemann, K., Goodheart, J.A., Schroedl, M., Martynov, A., Korshunova, T., Wägele, H., Donath, A., 2021. Transcriptomics provides a robust framework for the relationships of the major clades of cladobranch sea slugs (Mollusca, Gastropoda, Heterobranchia), but fails to resolve the position of

the enigmatic genus Embletonia. BMC Ecol Evo 21, 226. https://doi.org/10.1186/s12862-021-01944-0

Katoh, K., Standley, D.M., 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol 30, 772–780. https://doi.org/10.1093/molbev/mst010

Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., Stamatakis, A., 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35, 4453–4455. https://doi.org/10.1093/bioinformatics/btz305

Krug, P.J., Vendetti, J.E., Ellingson, R.A., Trowbridge, C.D., Hirano, Y.M., Trathen, D.Y., Rodriguez, A.K., Swennen, C., Wilson, N.G., Valdés, Á.A., 2015. Species Selection Favors Dispersive Life Histories in Sea Slugs, but Higher Per-Offspring Investment Drives Shifts to Short-Lived Larvae. Syst Biol 64, 983–999. https://doi.org/10.1093/sysbio/syv046

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923

Legendre, F., Condamine, F.L., 2018. When Darwin's Special Difficulty Promotes Diversification in Insects. Syst Biol 67, 873–887. https://doi.org/10.1093/sysbio/syy014

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Magoč, T., Salzberg, S.L., 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27, 2957–2963. https://doi.org/10.1093/bioinformatics/btr507

McLean, J.H., 1971. A Revised Classification of the Family Turridae, with the Proposal of New Subfamilies, Genera, and Subgenera from the Eastern Pacific 14, 8.

Meyer, M., Kircher, M., 2010. Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. Cold Spring Harb Protoc 2010, pdb.prot5448. https://doi.org/10.1101/pdb.prot5448

Modica, M.V., Gorson, J., Fedosov, A.E., Malcolm, G., Terryn, Y., Puillandre, N., Holford, M., 2020. Macroevolutionary Analyses Suggest That Environmental Factors, Not Venom Apparatus, Play Key Role in Terebridae Marine Snail Diversification. Syst Biol 69, 413–430. https://doi.org/10.1093/sysbio/syz059

Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol 32, 268–274. https://doi.org/10.1093/molbev/msu300

Phuong, M.A., Alfaro, M.E., Mahardika, G.N., Marwoto, R.M., Prabowo, R.E., von Rintelen, T., Vogt, P.W.H., Hendricks, J.R., Puillandre, N., 2019. Lack of Signal for the Impact of Conotoxin Gene Diversity on Speciation Rates in Cone Snails. Systematic Biology. https://doi.org/10.1093/sysbio/syz016

Postaire, B., Bruggemann, J.H., Magalon, H., Faure, B., 2014. Evolutionary Dynamics in the Southwest Indian Ocean Marine Biodiversity Hotspot: A Perspective from the Rocky Shore Gastropod Genus Nerita. PLoS ONE 9, e95040. https://doi.org/10.1371/journal.pone.0095040

Puillandre, N., Brouillet, S., Achaz, G., 2020. ASAP: assemble species by automatic partitioning. Molecular Ecology Resources n/a. https://doi.org/10.1111/1755-0998.13281

Puillandre, N., Cruaud, C., Kantor, Y.I., 2010. Cryptic species in Gemmuloborsonia (Gastropoda: Conoidea). J Molluscan Stud 76, 11–23. https://doi.org/10.1093/mollus/eyp042

Puillandre, N., Fedosov, A.E., Zaharias, P., Aznar-Cormano, L., Kantor, Y.I., 2017. A quest for the lost types of Lophiotoma (Gastropoda: Conoidea: Turridae): integrative taxonomy in a nomenclatural mess. Zoological Journal of the Linnean Society. https://doi.org/10.1093/zoolinnean/zlx012

Puillandre, N., Modica, M.V., Zhang, Y., Sirovich, L., Boisselier, M.-C., Cruaud, C., Holford, M., Samadi, S., 2012. Large-scale species delimitation method for hyperdiverse groups. Molecular Ecology 21, 2671–2691. https://doi.org/10.1111/j.1365-294X.2012.05559.x

Rose, J.P., Kleist, T.J., Löfstrand, S.D., Drew, B.T., Schönenberger, J., Sytsma, K.J., 2018. Phylogeny, historical biogeography, and diversification of angiosperm order Ericales suggest ancient Neotropical and East Asian connections. Molecular Phylogenetics and Evolution 122, 59–79. https://doi.org/10.1016/j.ympev.2018.01.014

Sayyari, E., Whitfield, J.B., Mirarab, S., 2017. Fragmentary Gene Sequences Negatively Impact Gene Tree and Species Tree Reconstruction. Mol Biol Evol 34, 3279–3291. https://doi.org/10.1093/molbev/msx261

Simakov, O., Marletaz, F., Cho, S.-J., Edsinger-Gonzales, E., Havlak, P., Hellsten, U., Kuo, D.-H., Larsson, T., Lv, J., Arendt, D., Savage, R., Osoegawa, K., de Jong, P., Grimwood, J., Chapman, J.A., Shapiro, H., Aerts, A., Otillar, R.P., Terry, A.Y., Boore, J.L., Grigoriev, I.V., Lindberg, D.R., Seaver, E.C., Weisblat, D.A., Putnam, N.H., Rokhsar, D.S., 2013. Insights into bilaterian evolution from three spiralian genomes. Nature 493, 526–531. https://doi.org/10.1038/nature11696

Slater, G.S.C., Birney, E., 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6, 31. https://doi.org/10.1186/1471-2105-6-31

Smit, A.F.A., Hubley, R., Green, P., 2015. RepeatMasker Open-4.0.

Steppan, S.J., Schenk, J.J., 2017. Muroid rodent phylogenetics: 900-species tree reveals increasing diversification rates. PLoS ONE 12, e0183070. https://doi.org/10.1371/journal.pone.0183070

Van Damme, K., Cornetti, L., Fields, P.D., Ebert, D., 2022. Whole-Genome Phylogenetic Reconstruction as a Powerful Tool to Reveal Homoplasy and Ancient Rapid Radiation in Waterflea Evolution. Systematic Biology 71, 777–787. https://doi.org/10.1093/sysbio/syab094

Williams, S.T., Duda, T.F., 2008. Did Tectonic Activity Stimulate Oligo–Miocene Speciation in the Indo-West Pacific? Evolution 62, 1618–1634. https://doi.org/10.1111/j.1558-5646.2008.00399.x

Williams, S.T., Smith, L.M., Herbert, D.G., Marshall, B.A., Warén, A., Kiel, S., Dyal, P., Linse, K., Vilvens, C., Kano, Y., 2013. Cenozoic climate change and diversification on the continental shelf and slope: evolution of gastropod diversity in the family Solariellidae (Trochoidea). Ecology and Evolution 3, 887–917. https://doi.org/10.1002/ece3.513

Zaharias, P., Kantor, Y.I., Fedosov, A.E., Criscione, F., Hallan, A., Kano, Y., Bardin, J., Puillandre, N., 2020a. Just the once will not hurt: DNA suggests species lumping over two oceans in deep-sea snails (Cryptogemma). Zool J Linn Soc 190, 532–557. https://doi.org/10.1093/zoolinnean/zlaa010

Zaharias, P., Pante, E., Gey, D., Fedosov, A.E., Puillandre, N., 2020b. Data, time and money: evaluating the best compromise for inferring molecular phylogenies of non-model animal taxa. Molecular Phylogenetics and Evolution 142, 106660. https://doi.org/10.1016/j.ympev.2019.106660

Zhang, C., Rabiee, M., Sayyari, E., Mirarab, S., 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics 19, 153. https://doi.org/10.1186/s12859-018-2129-y

**Figure legends**

**Figure 1.** Flowchart summarizing the general strategy, from the blind, exploratory and intensive barcode sampling step to the taxonomic re-evaluation. Each step is described on the left while the results obtained in the present study are shown on the right.

**Figure 2.** *cox1* ML tree on which the ASAP species delimitations were mapped using iTOL (https://itol.embl.de/). Some of the 201 delimited species were removed when multiple names were overlapping for clarity. Because ASAP is non-tree based, a few delimited species (or species complex) appear paraphyletic (e.g., *Gemmuloborsonia karubar*/*incognita*). Boxes with multiple names correspond to cases where multiple previously named/delimited species were considered as a single partition by ASAP. The red circles indicate supra-specific nodes with Transfer Bootstrap Expectation (TBE) support > 80%.

**Figure 3.** IQ-TREE phylogenomic tree (DS2-IQ) of the Turridae. The ultrafast bootstrap support (UFboot) values are only shown for edges supported by less than 100. Despite their paraphyly in this dataset, *Gemmula hindsiana* and *G. closterion* are still considered in the same genus here because of the ambiguity of *G. hindisana*'s placement in other datasets. (T): Sequences obtained from transcriptomes. Occupancy matrix (taxa in rows) shown on the bottom left (detailed occupancy matrices available in supp. Fig. SF2). The other trees are available in the supplementary files (see supp. Fig. SF3).

**Tables**

**Table 1.** Comparison of Secondary Species Hypotheses (SSHs) from four previous studies with the matching Primary Species Hypotheses (PSHs) obtained in this study. The PSH numbering refers to the numbering in the ASAP partition (see Supp. Table S1). *The name *Gemmuloborsonia karubar* does not appear in the original article and was attributed a posteriori to the SSH formed by specimen MNHN-IM-2007-41918 (Puillandre et al., 2010).

**Table 2.** Tree statistics on all six analyses. Nb of loci: number of loci; Taxon occupancy: average number of taxa per loci; GT supported edges: proportion of gene tree edges with more than 80% of ultrafast bootstrap support; Aln. length: alignment length for each concatenated dataset with missing data proportion indicated in parenthesis; Mean ST support: mean species tree support indicated in ultrafast bootstrap (UFBoot) support for the concatenated dataset or Posterior Probabilities (PP) for the summary tree approach; Mean GT tree distance to ST: mean quartet distance of single-locus tree to the species tree.

**Table 3.** Taxonomic summary of the Turridae classification, with current listing of extant genera compared with our suggested generic revision (to be published in a separate article). * the total number of PSHs (200) doesn't match the number presented in Results section because we failed to attribute "*Gemmula* sp. 27" (= ASAP partition n°44) to any genus.

**Supplementary Material**

**Supplementary Table ST1.** List and description of all the barcoded Turridae specimens used in this study. ASAP partition results are shown in the two last columns. Yellow cells corresponds to cases were ASAP and mPTP disagree.

**Supplementary Table ST2.** Description of all the transcriptomes used in this study.

**Supplementary Table ST3.** Exon-capture results statistics.

**Supplementary Figure SF1.** A Maximum-Likelihood tree of the 3,159 *cox1* Turridae using RAxML-NG. Branch support inferred using 1,000 bootstrap trees.

**Supplementary Figure SF2.** Occupancy matrix (taxa in rows, exon loci in column) for each dataset.

**Supplementary Figure SF3.** All six species trees inferred from the three datasets (DS1-3) using either Astral (AS) or IQ-TREE (IQ).

Author Agreement Statement

We the undersigned declare that this manuscript is original, has not been published before and is not currently being considered for publication elsewhere.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

We understand that the Corresponding Author is the sole contact for the Editorial process.

He is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs

Signed by all authors as follows (October 30th 2023):

Paul Zaharias

Yuri I. Kantor

Alexander E. Fedosov

Nicolas Puillandre

| Study | SSHs (31) | matching PSHs (27) |
|---|---|---|
| Puillandre et al. (2010) | G. karubar* | 35 |
| | G. colorata | 22 |
| | G. clandestina | 99 |
| | G. neocaledonica | 97 |
| | G. moosai | 98 |
| Puillandre et al. (2017) | L. polytropa | 47 |
| | L. abbreviata | 67 |
| | L. brevicaudata | 67 |
| | L. jickelii | 20 |
| | L. picturata | 144 |
| | L. bratasusa | 91 |
| | L. vezzaroi | 84 |
| | L. kina | 114 |
| | L. semfala | 66 |
| | L. acuta | 19 |
| Abdelkrim et al. (2018) | X. legitima | 12, 13 |

| | | |
|---|---|---|
| | I. notata | 3 |
| | I. olangoensis | 3 |
| | I. musivum | 3 |
| | I. conotaxis | 3 |
| | I. cingulifera | 14 |
| | I. kingae | 11 |
| | I. devoizei | 11 |
| Zaharias et al. (2020) | C. periscelida | 193 |
| | C. phymatias | 108 |
| | C. praesignis | 1 |
| | C. tessellata | 54 |
| | C. aethiopica | 105 |
| | C. unilineata | 24 |
| | C. timorensis | 75 |
| | C. powelli | 56 |

| Dataset | Nb of loci | Taxon occupancy | GT supported edges | Aln. length | Mean ST support | Mean GT tree distance to ST |
|---|---|---|---|---|---|---|
| DS1-IQ | 4,178 | 49 | 28.8% | 634,645 (58%) | 97.3 UFBoot | 0.59 |
| DS1-AS | | | | - | 88.5 PP | 0.59 |
| DS2-IQ | 2,001 | 73 | 30.0% | 314,854 (37%) | 97.1 UFBoot | 0.61 |
| DS2-AS | | | | - | 91.4 PP | 0.61 |
| DS3-IQ | 1,380 | 73 | 29.8% | 214,812 (37%) | 95.0 UFBoot | 0.59 |
| DS3-AS | | | | - | 88.5 PP | 0.59 |

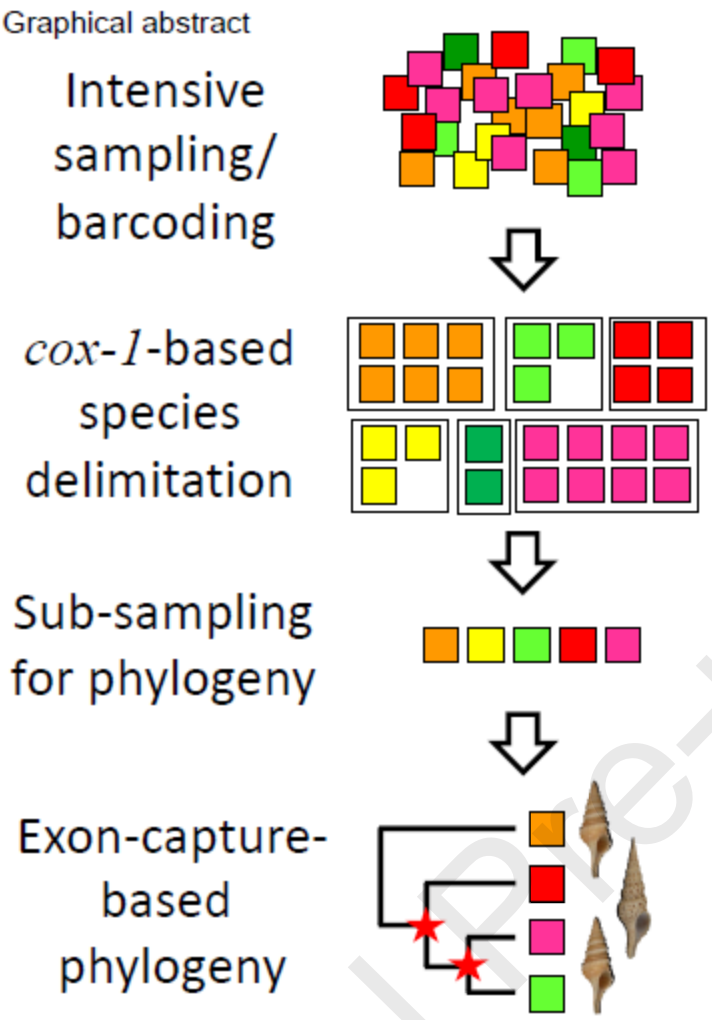
| Extant Genera (Worms Jul/2023) | Genus (this manuscript) | type species | Nb extant species (Worms Jul/2023) | Nb PSHs (this manuscript) | ASAP PSHs |
|---|---|---|---|---|---|
| *Cryptogemma* Dall 1918 | *Cryptogemma* | *Gemmula benthima* Dall, 1908 | 18 | 8 | 1, 24, 54, 56, 75, 105, 108, 193 |
| | "*chilensis*" clade | *Ptychosyrinx chilensis* S. S. Berry, 1968 | 1 | 1 | 148 |
| *Decollidrillia* Habe & Ki. Ito, 1965 | N.A. | *Decollidrillia nigra* Habe & Ito, 1965 | 1 | N.A. | N.A. |
| *Epidirella* Iredale, 1931 | N.A. | *Hemipleurotoma tasmanica* May, 1911 | 1 | N.A. | N.A. |
| *Gemmula* Weinkauff, 1875 | *Gemmula* | *Pleurotoma gemmata* Reeve, 1843 | 50 | 3 | 2, 60, 156 |
| | *Eugemmula* | *Eugemmula hawleyi* Iredale, 1931 | | 18 | 37, 53, 57, 65, 68, 71, 87, 88, 95, 112, 123, 131, 136, 137, 138, 149, 152, 153 |
| | *Gemmula* A | N.A. | | 3 | 49, 73, 115 |

| | | | | | |
|---|---|---|---|---|---|
| | *Gemmula* B | N.A. | | 3 | 89, 122, 175 |
| | *Gemmula* C | N.A. | | 1 | 155 |
| | *Gemmula* D | N.A. | | 2 | 39, 204 |
| | *Gemmula* E | N.A. | | 7 | 7, 16, 26, 40, 61, 64, 72 |
| | *Gemmula* F | N.A. | | 4 | 21, 43, 55, 63 |
| | *Gemmula* G | N.A. | | 6 | 42, 48, 52, 58, 76, 81 |
| | *Gemmula* H | N.A. | | 1 | 10 |
| | *Gemmula* I | N.A. | | 5 | 31, 51, 77, 41, 154 |
| | *Gemmula* J | N.A. | | 14 | 18, 32, 33, 36, 59, 70, 74, 78, 82, 83, 93, 121, 141, 181 |
| *Gemmulobors onia* Shuto, 1989 | *Gemmulobors onia* | *Gemmuloborsonia fierstinei* Shuto, 1989 † | 7 | 9 | 22, 28, 35, 97, 98, 99, 132, 164, 172 |
| *Iotyrris* Medinskaya & Sysoev, 2001 | *Iotyrris* | *Iotyrris marquesensis* Sysoev, 2002 | 9 | 5 | 3, 11, 14, 147, 179 |

| | | | | | |
|---|---|---|---|---|---|
| *Kuroshioturris* Shuto, 1961 | *Kuroshioturris* | *Gemmula* (*Kuroshioturris*) *hyugaensis* Shuto, 1961 † | 5 | 1 | 125 |
| *Lophiotoma* T. L. Casey, 1904 | *Lophiotoma* | *Pleurotoma tigrina* Lamarck, 1822 | 24 | 10 | 19, 20, 47, 66, 67, 84, 91, 114, 135, 144 |
| *Lucerapex* Wenz, 1943 | *Lucerapex* | *Pleurotoma casearia* Hedley & Petterd, 1906 | 8 | 14 | 100, 101, 102, 103, 104, 118, 120, 124, 126, 127, 128, 157, 163, 174 |
| *Polystira* Woodring, 1928 | *Polystira* | *Pleurotoma albida* G. Perry, 1811 | 25 | 22 | 129, 146, 161, 162, 165, 166, 167, 168, 169, 170, 176, 180, 190, 191, 195, 196, 197, 198, 199, 200, 201, 202 |
| *Purpuraturris* K. Chase, Watkins, Safavi-Hemami & B. M. Olivera, 2022 | *Annulaturris* / *Purpuraturris* | *Pleurotoma amicta* E. A. Smith, 1877 | 6 | 9 | 106, 107, 109, 143, 145, 177, 182, 185, 187 |
| *Turridrupa* Hedley, 1922 | *Turridrupa* | *Pleurotoma acutigemmata* E. A. Smith, 1877 | 17 | 16 | 23, 25, 29, 62, 69, 85, 94, 96, 110, 130, 140, 142, 158, 173, 178, 192 |

| | | | | | |
|---|---|---|---|---|---|
| *Turris* Batsch 1789 | *Turris* | *Murex babylonius* Linnaeus, 1758 | 25 | 11 | 17, 86, 92, 113, 139, 159, 160, 188, 189, 194, 207 |
| *Unedogemmula* MacNeil, 1961 | *Unedogemmula* | *Pleurotoma unedo* Kiener, 1839 | 7 | 24 | 4, 5, 6, 8, 9, 15, 27, 30, 34, 38, 45, 46, 50, 79, 80, 90, 111, 116, 117, 119, 133, 151, 203, 206 |
| *Xenuroturris* Iredale, 1929 | *Xenuroturris* | *Xenuroturris legitima* Iredale, 1929 | 5 | 3 | 12, 13, 134 |
| | | | **209** | **200*** | |

Graphical abstract



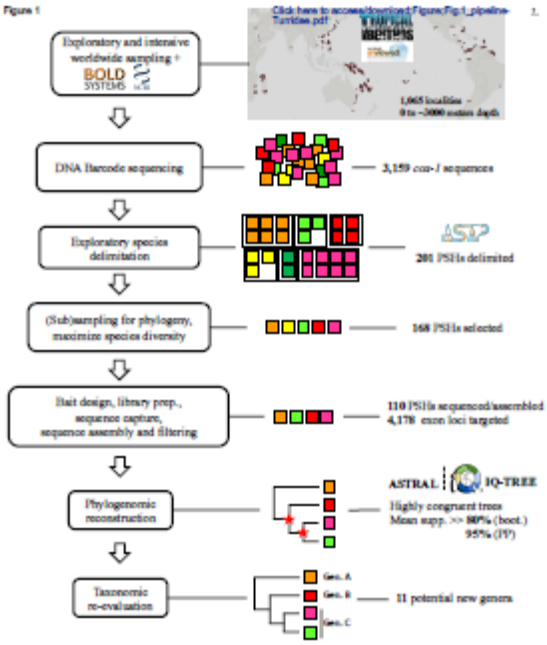


Figure 2