Identifying gastropod spawn from DNA barcodes: possible but not yet practicable

N. Puillandre, E. E Strong, P. Bouchet, M.-C. Boisselier, A. Couloux, S Samadi

Identifying gastropod spawn from DNA barcodes: possible but not yet practicable

N. Puillandre*[,1], E.E. Strong[2], P. Bouchet[3], M.-C. Boisselier[1], A. Couloux[4] & S. Samadi[1]

[1] Université Pierre et Marie Curie (UPMC), UMR 7138, Systématique, adaptation, évolution (UPMC/IRD/MNHN/CNRS), CP26, 57 rue Cuvier, 75231 Paris Cedex 05, France.

[2] Department of Invertebrate Zoology, Smithsonian Institution, National Museum of Natural History, MRC 163, P.O. Box 37012, Washington, DC 20013-7012, USA.

[3] Muséum National d'Histoire Naturelle, 57 rue Cuvier, 75231 Paris Cedex 05, France.

[4] GENOSCOPE, Centre National de Séquençage, Evry, France.

Keywords: DNA barcoding, egg capsules, Neogastropoda, Barcode of Life Database, GenBank

* Corresponding author: UMR 7138, Systématique, adaptation, évolution, UPMC; IRD; MNHN; CNRS, Service de systématique moléculaire (CNRS, IFR 101), Département systématique et évolution, Muséum National d'Histoire Naturelle, CP26, 57 rue Cuvier, 75231 Paris Cedex 05, France.; puillandre@mnhn.fr; +33 1 40 79 37 43

Running title: Barcoding gastropod spawn

Abstract

Identifying life stages of species with complex life histories is problematic as species are often only known and/or described from a single stage. DNA barcoding has been touted as an important tool for linking life history stages of the same species. To test the current efficacy of DNA barcodes for identifying unknown mollusk life stages, 24 marine gastropod egg capsules were collected off the Philippines in deep-water and sequenced for partial fragments of the COI, 16S and 12S mitochondrial genes. Two egg capsules of known shallow-water Mediterranean species were used to calibrate the method. These sequences were compared to those available in GenBank and the Barcode of Life Database (BOLD). Using COI sequences alone, only a single Mediterranean egg capsule was identified to species, and a single Philippine egg capsule was identified tentatively to genus; all other COI sequences recovered matches between 76% and 90% with sequences from BOLD and GenBank. Similarity-based identification using all three markers confirmed the Mediterranean specimens' identifications. A phylogenetic approach was also implemented to confirm similarity-based identifications and provide a higher-taxonomic identification when species-level identifications were not possible. Comparison of available GenBank sequences to the diversity curve of a well sampled coral reef habitat in New Caledonia highlights the poor taxonomic coverage achieved at present in existing genetic databases, emphasizing the need to develop DNA barcoding projects for megadiverse and often taxonomically challenging groups such as mollusks, to fully realize its potential as an identification and discovery tool.

Introduction

Identification of different life stages for species with complex life histories has long posed a challenge to biologists of almost every discipline as species are routinely known and/or described based only on a single stage. Identifying early life stages is particularly difficult for marine invertebrates including mollusks that often undergo dramatic metamorphoses from the egg to one or more larval stages before finally reaching juvenile and adult form. The issue is trivial in mollusk species with parental care, but these represent only a small minority, the vast majority dispersing their larvae or their spawn (e.g. Lebour 1937; Fretter & Graham 1962; Robertson 1974) sometimes vast distances through ocean currents.

Mollusk spawn are highly conservative within species and present features that are diagnostic at many taxonomic levels, providing another character set useful for understanding evolutionary relationships (e.g. Habe 1960; Robertson 1974; Bandel 1976a,b). But due to the difficulties inherent in identification, spawn have been described for only several hundred species at best, with the tropical mollusk fauna particularly poorly known. This limits the utility of spawn in systematic and ecological studies.

The process of identifying mollusk spawn typically relies on serendipitous discoveries of individuals actively depositing eggs, painstaking and time-consuming breeding and rearing of larvae *ex situ*, or circumstantial evidence (species presence, abundance) and a process of elimination using criteria including substrate, spawn morphology, developmental mode, ovum size, ovary/ovum color, etc. (e.g. D'Asaro 1970; Winner 1987; Gustafson et al. 1991). However, *in situ* observations or *ex situ* rearing of juveniles are often impractical or may be impossible, e.g. in vent (Lutz et al. 1986) and other deep-water species (present study).

Classic surveys of molluscan spawn have focused on northern temperate biomes with low numbers of species and are almost exclusively limited to species with intra-capsular development (crawl-away juveniles) allowing comparisons between near-hatching embryos and the protoconchs of known adults. Such conditions provide an adequately simple system facilitating the identification of larvae and spawn (Thorson 1940b). But even after careful scrutiny of the protoconchs, this approach may still be fallible and lead to incorrect identifications in a certain number of cases (e.g. Knudsen 1950) or will be inconclusive in poorly known systems and especially in species with planktonic larval development (e.g. Gustafson et al. 1991). In an extreme case, a parallel nomenclature was devised for egg capsules that could not be recognizably linked to benthic species (Tokioka 1950). Moreover, identification to species level can be complicated in some cases where spawn of several species are morphologically indistinguishable (e.g. simple gelatinous egg masses of some vetigastropods – Lebour 1937; Fretter & Graham 1962; some species of *Conus* – Kohn 1961).

DNA barcodes are a promising and expedient new tool for accurately identifying and linking the varied life history stages of single species (Schindel & Miller 2005). Indeed, DNA barcoding has demonstrated its capacity to do so successfully in several animal groups (Blaxter 2004; Steinke et al. 2005; Pegg et al. 2006; Thomas et al. 2005; Vences et al. 2005; Ahrens et al. 2007), and has already been used for marine fauna to link larvae and adults (Victor 2007). Such approaches have also been used to explore biodiversity, leading to the discovery of previously unrecognized species collected from their planktonic larvae while the difficult-to-collect adults are still unknown (Barber & Boyce 2006). Among mollusks, and especially bivalves, molecular approaches have been used to identify larvae using rRNA

probes for *in situ* hybridization (Le Goff-Vitry et al. 2007; Pradillon et al. 2007; Jones et al. 2008) or polymerase chain reaction-single strand conformation polymorphism (PCR-SSCP) combined with sequencing (Livi et al. 2006) of taxonomically informative fragments of 18S rDNA. Such methods have been used primarily to confirm the identity of target species using species-specific probes developed from known adults and as such, have mostly practical utility in, e.g., biomonitoring of target, commercial or invasive species. Although it has been claimed that these methods may have broader utility for species identification by overcoming some of the limitations of mitochondrial markers (introgression, pseudogenes) (Livi et al. 2006), 18S rDNA faces the limitation that it is a highly conserved gene and will have decreased ability to accurately distinguish between closely related species. Recently, DNA barcodes in combination with other gene fragments (H3, 16S, 18S) were used to link egg masses, larvae and adults in one family of gastropods (Naticidae) in the western Mediterranean – a geographically circumscribed area with a well documented fauna (Huelsken et al. 2008).

Here, we assess the capacity of the animal barcode (COI), two additional mitochondrial markers (16S, 12S) and the existing genetic databanks (GenBank, Barcode of Life Database - BOLD) to identify a set of unknown gastropod egg capsules collected in the megadiverse region of the Indo-Pacific as compared to identified gastropod egg cases from the well-documented Mediterranean fauna. We use a two-tiered process first involving similarity-based methods (the identification engine of BOLD and the BLAST algorithm on GenBank), previously used in DNA barcode identification (e.g. Wong & Hanner 2008), followed by phylogenetic reconstruction (using all available sequences in GenBank) to tentatively

117 determine the sister taxa of unidentified egg capsules, when similarity-based methods are not
118 effective.
119
120 **Material and methods**
121 *Sampling*
122 Twenty-four egg capsules were collected in the Philippines on scattered hard substrates (e.g.
123 stones, shells) from soft bottoms by trawling, between 150 and 1450 meters, during the
124 Aurora 2007 deep sea cruise off the east coast of Luzon. All were tentatively recognized as
125 neogastropod capsules as most share the lenticular shape typical of many neogastropod egg
126 cases (Bandel 1976b) (Fig. 1C-H); one egg capsule was identified as that of a species of
127 *Conus* (Conidae, Neogastropoda – Fig. 1B) with the flask-like shape characteristic of the
128 genus. All capsules were first photographed on the substrate, then placed in 95% ethanol. In
129 addition, two identified egg capsules (EC1, EC2) from the French Mediterranean coast were
130 used as a control, allowing us to assess the capacity of DNA barcodes to identify known
131 samples. Both were readily identifiable to species as *Coralliophila meyendorffii* broods egg
132 capsules in the mantle cavity, and adults of *Erosaria spurca* were found near the egg capsules
133 (Fig. 1A).
134
135 *Sequencing*
136 DNA was extracted from the whole egg capsule, using 6100 Nucleic Acid Prepstation system
137 (Applied Biosystem). Three gene fragments were amplified, corresponding to some of the
138 most represented molluscan genes in GenBank, and also to genes commonly used at the
139 species level (Hebert et al. 2003; Remigio & Hebert 2003): (i) a 658 bp fragment of the
140 Cytochrome Oxidase I (COI) mitochondrial gene using universal primers LCO1490 and

HCO2198 (Folmer et al. 1994), (ii) a 550 bp fragment of the 16S mitochondrial gene using primers 16SH (CGTGATCTGAGTTCAGACCGG) and 16SL (GTTTACCAAAAACATGGCTTC) and (iii) a 600 bp fragment of the 12S mitochondrial gene using primers 12SI (TGCCAGCAGYCGCGGTTA) and 12SIII (AGAGYGRCGGGCGATGTGT). All PCR reactions were performed in 25 µl, containing 3 ng of DNA, 1X reaction buffer, 2.5 mM $MgCl_2$, 0.26 mM dNTP, 0.3 µM of each primer, 5% DMSO and 1.5 units of Q-Bio Taq, QBiogene for all genes. Thermocycles used for COI gene are those described in Hebert et al. (2003); for 16S and 12S genes, they consisted in an initial denaturation step at 94°C for 4', followed by 30 cycles of denaturation at 94°C for 30'', annealing at 52°C for 16S and 54°C for 12S. The final extension was at 72°C for 10'. PCR products were purified and sequenced by the Genoscope. In all cases, both directions were sequenced to confirm accuracy of each haplotype (GenBank Accession Numbers: EU870520-EU870589).

Species identification

A two-tiered approach was employed to identify egg capsules, the first step involving similarity-based methods (identification engine of BOLD and BLAST search routine in GenBank) and a second step involving tree-based methods using Bayesian analysis of neogastropod sequences in GenBank. In the first step, each sequence was compared to available GenBank sequences using the BLASTn search routine as implemented in GenBank (default parameters). The best hit, as measured by percent maximum sequence identity, was retained. In addition, each sequence was compared to all available barcode records in BOLD, using the identification engine BOLD-IDS, with the option "searching all barcode records in BOLD". This provides a list of similar sequences with the associated taxon name and the

percent sequence similarity. Contrary to the BLAST algorithm, identification in BOLD is based on genetic distances, and is not influenced by sequence length (Ratnasingham & Hebert 2007). A cut off value of 1% sequence divergence was used for considering two sequences conspecific (Ratnasingham & Hebert 2007).

In the second step, all egg capsule sequences were included in phylogenetic analyses to assess which sequences form molecular operational taxonomic units (OTUs) and to evaluate the higher taxonomic placement of the OTUs. All neogastropod COI, 16S and 12S sequences were initially included, but to limit the total number of sequences, only one sequence per species was retained. An alignment with egg capsule sequences was generated using ClustalW multiple alignment implemented in BioEdit version 7.0.5.3 (Hall 1999) and only those sequences corresponding to the fragments sequenced for the egg capsules were retained. Ultimately, 159, 127 and 54 (for COI, 16S and 12S, respectively) GenBank sequences were used for the phylogenetic analyses. Phylogenetic trees were constructed using Bayesian inference with MrBayes (Huelsenbeck et al. 2001) (two Markov chains, 2000000 generations each with a sampling frequency of one tree each hundred generations, four parallel analyses).

Similarity-based methods (BLAST, BOLD-IDS) followed by phylogeny reconstruction were implemented for each gene, except for BOLD-IDS identifications as BOLD only contains COI sequences. Match scores are provided for BOLD searches with COI, and for BLAST searches with all three genes. For each sample, a final identification, corresponding to the best similarity score for the three genes (match score superior to 95%) and/or to the name of the sister taxa in the tree, is proposed based on the results of these analyses (see details in Table 1).

Results

Similarity-based identifications

Of the 26 samples (24 unknown, 2 known), only one of the known Mediterranean samples was identified to species using the BOLD identification engine (specimen EC2: 99.51% similarity with a sequence of *Erosaria spurca*), confirming the field-based identification (Table 1). Of the Philippine samples, only specimen EC8 was tentatively identified to genus, returning a match with a sequence of *Comitas* sp. (Turridae) at 97.84% similarity. All other COI sequences returned matches at 84% to 89% similarity with one or several sequences in BOLD, far exceeding the genetic distance considered to separate species (Hebert et al. 2003). Similarly, only the specimen of *Erosaria spurca* was identified to species with COI alone using the BLAST search routine in GenBank (98% identity); all other first hits returned matches at 76% to 90% sequence identity.

The two additional mitochondrial markers sequenced in this analysis (16S, 12S) did not fare much better. Again, identification of specimen EC2 (*Erosaria spurca*) was confirmed with the 16S gene (99% identity) and the 12S sequence of specimen EC1 returned a match with a sequence of *Coralliophila meyendorffii* at 99% identity, confirming the field-based identification of the second Mediterranean sample. These were the only matches at species level. However, twelve 16S egg capsule sequences produced matches with a 16S sequence of *Granulifusus niponicus* at between 95% and 98% identity, and the 16S sequence of specimen EC3 matched with a sequence of *Conus radiatus* at 95% identity; these results tentatively suggest a higher taxonomic identification at genus level (but see Discussion, below). All other

matches for the 16S gene ranged from 82% to 89% identity, while those for the 12S gene ranged from only 79% to 94% identity.

Tree-based identifications

Among the 26 samples, 11 molecular OTUs were recognized using phylogeny reconstruction (Table 1). As shown for the 16S gene (Fig. 2), the phylogenetic placement of these OTUs can be used to confirm similarity-based identifications, but also to suggest supra-specific identifications when identification to species level is not possible based on sequence similarity. For example, with the 16S gene, the 12 sequences comprising OTU 4 form a clade with the GenBank sequence of *G. niponicus* (posterior probability PP = 0.98) (Fig. 2). EC23 (OTU 10) and EC26 (OTU 11) form a clade with *Raphitoma linearis* (Conidae, Raphitominae) with a PP of 1. With the 12S gene (results not shown), OTU 8 (EC 12, 15) clusters in a group with EC8 composed exclusively of Turridae species (PP = 0.97). With the COI gene (results not shown), OTU 5 (EC5, 6 and 7) is closely related to a sequence of *Belomitra* sp. (Buccinidae) (PP = 1). This identification was already suggested with the BLAST search routine in GenBank, but only with 89% identity. The COI gene tree also identifies EC23 as a member of the Raphitominae (Conidae) and EC26 as a member of the Clathurellinae (Conidae), thereby refining the 16S familial placement of these two egg capsules.

Discussion

Eggs capsule identification

Identification of the egg capsules of *Erosaria spurca* collected in France illustrates the capacity of the animal DNA barcode to successfully link the different life stages of a single

gastropod species. However, only one of the two Mediterranean samples was identified by the COI barcode with a high level of precision (i.e. species level at >98% similarity). None of the Philippine egg capsules were identified with the same level of precision using COI alone. The threshold routinely used to consider two COI sequences as belonging to the same OTU ranges from 1% to 2% (Hebert et al. 2004; Bichain et al. 2007; Ratnasingham & Hebert 2007), and except for the specimen of *Erosaria spurca*, match scores for all other samples exceed this threshold. Similarly, only the Mediterranean samples retrieved matches at >98% sequence similarity for the 16S and 12S markers; species identification was confirmed for specimen EC2 (*Erosaria spurca*) with 16S (99% identity) and for specimen EC1 (*Coralliophila meyendorffii*) with 12S (also 99% identity).

For many of the remaining samples, a supra-specific identification was possible based on match scores or using a combination of similarity- and tree-based methods. For example, although a genus-level genetic threshold is difficult to specify and will vary greatly between taxa (Holland et al. 2004), a low genetic distance ($\leq$ 5%) tentatively supports a genus-level identification for OTU6 (97.84% – *Comitas* sp.) based on COI and for OTU3 (95% – *Conus radiatus*) and OTU4 (95-98% – *Granulifusus niponicus*) based on 16S (see Fig. 3). The positioning of OTUs in the phylogenetic tree confirms these similarity-based identifications, but can also refine them in some cases. For example, the COI sequence for EC23 returned matches with *Nannodiella* (Conidae, Clathurellinae) in BOLD (87.4%) and with *Gymnobela* (Conidae, Raphitominae) in GenBank (86%), suggesting a family-level identification (Conidae). In the COI gene tree, EC23 is more closely related to *Nannodiella*, thereby suggesting that EC23 is a Clathurellinae.

Three egg capsules (OTU7, 9) could not be assigned even a tentative supra-specific identification based on the combination of similarity- and tree-based methods (see Table 1). Hit scores for these samples for all three genes are inferior to 90%; BLAST and BOLD results indicate that when the first matching sequence displays low sequence similarity (i.e. <90%), the following matching sequences correspond (with lower or similar percentage) to completely different taxa. For example, BLAST results indicate that the COI sequence of EC3 matches first with *Lacuna pallidula* (Littorinidae) with a sequence identity of 85%, but then with *Nucella lamellosa* (Neogastropoda, Muricidae) with the same score, with *Ilyanassa obsoleta* (Neogastropoda, Nassaridae) at 84% identity and with *Urosalpinx perrugata* (Neogastropoda, Muricidae) also at 84%. Consequently, such scores are not useful for hypothesizing even superfamily-level identifications with a high degree of certainty. Moreover, the relationships of these samples were not robustly resolved in the phylogenetic analyses, preventing a family-level identification. However, this is not true in all cases. For example, OTU5 returned matches with *Belomitra* sp. (Buccinidae) in GenBank at 89% sequence identity, and this identification is supported by the phylogenetic analyses. In the case of specimens EC12 and EC15 (OTU8), BOLD and GenBank best hits for 12S correspond to turrid species with low sequence similarity (*Gemmula rosario* for EC12, 83%, and *Lophiotoma unedo* for EC15, 79%), but the 12S gene tree tentatively supports this identification at family rank (Turridae).

Morphology of eggs capsules

Specimen EC3 has a morphology typical of *Conus* species with its flask-like shape (Fig. 1 B), and DNA analysis confirms this hypothesis. Specimen EC5, identified as *Belomitra* sp. (Buccinidae) is globular (Fig. 1 E). Although *Belomitra* egg morphology has never been

described, this result is consistent with the literature, as similar egg capsule shapes have been already reported for several buccinid genera (Thorson 1935, 1940a). All other egg capsules examined in the present study share a similar lenticular shape, with an escape aperture at the apex of the capsule, and containing a variable number of eggs. This morphology is not common among fasciolarids (see e.g. Knudsen 1950) which instead usually have flask or vase-shaped capsules, sometimes with an undulating apical ridge or keel that surrounds the escape aperture. Here, lenticular capsules were tentatively identified as *Granulifusus* sp. (Fig. 1 D) which would be the first record of an egg case for this genus.

Database completeness

Accuracy of DNA-based identification is dependent, of course, on the maturity of existing genetic databases. The impediment that incomplete databases pose for accurate and precise species identification has already been acknowledged, but how incomplete are they, and how does existing taxonomic coverage accurately mirror known diversity? The present analysis clearly indicates that even well-known faunas are inadequately represented in existing databases as evidenced by the fact only one of two samples from the French Mediterranean fauna could be identified to species using COI alone. It is surprising that even a well-known species as *Coralliophila meyendorffii* is not represented in databases for the COI and 16S genes; indeed, there is not a single 16S coralliophiline sequence and only one for the COI gene in these databases – a clade with roughly 250 species worldwide that form ecologically important associations with Cnidaria. Phylogenetic analysis can provide a supra-specific identification in cases where similarity-based methods return matches with only low similarity, and is the best alternative when no sequences of the same species are present in the databases. However, tree-based identifications should be treated with an appropriate amount

of caution. The weak phylogenetic signal of each gene alone does not allow a robust resolution of the phylogenetic relationships among neogastropods and most family-level relationships are not well supported. Concatenation of all three genes to improve resolution of the tree is not possible due to high levels of terminal mismatch – yet another measure of the incompleteness of existing databases.

Tallying the number of sequences available provides a more concrete measure of database completeness, although we must assume that all specimens have been identified correctly. Despite this caveat, by this measure taxonomic diversity of mollusks is severely underrepresented, as less than 16,000 molluscan COI sequences are currently published in GenBank, ostensibly corresponding to 3,688 species, or less than 2% of the roughly 80,000 valid species (of which 53,000 are marine; Bouchet 2006) already described. But equally important is the taxonomic distribution of available sequences. In order to quantify taxonomic coverage, we compared gastropod species richness by family at a well-sampled tropical site in New Caledonia (Koumac - Bouchet et al. 2002) with the number of COI sequences available in GenBank over the last three years (Fig. 4). The results illustrate several important biases in family representation in GenBank. For example, five of the six most speciose families at Koumac (Triphoridae, Eulimidae, Pyramidellidae, Cerithiopsidae, Costellariidae) are each represented by less than three COI sequences. These families are all highly diverse and taxonomically complex with large numbers of minute species that classically have been overlooked in systematic studies. To complicate matters, they are almost exclusively predatory and typically occur in low abundance, with as many as 20% of the species known only from single specimens at any given site (Bouchet et al. 2002).

Conversely, some families are conspicuously over-represented, with the number of sequences available disproportionate to their known global diversity. Among the marine families, these over-represented groups tend to be large, charismatic “sea shells” of interest to collectors and hobby naturalists (Cypraeidae, Muricidae). For example, the Cypraeidae comprises ~230 species worldwide (Lorenz 2002) and are represented by 682 COI sequences in GenBank. Whereas for Turridae *s.l.* (Conoideans except Terebridae and *Conus*), the most speciose assemblage at Koumac and with about 4,000 named valid species (Tucker 2004) and perhaps as many as 10-20,000 in reality (Kantor et al. in press), there are only 92 COI sequences available. The other well-represented marine families (e.g. Littorinidae, Neritidae) are similarly disproportionately represented given their known diversity, and tend to be easily accessible and abundant in shallow, near shore habitats.

A comparison of the number of sequences available in GenBank between 2006 and 2008 shows a slight increase in the number of sequences for most families, with several showing more significant increases (e.g. *Conus*, Littorinidae, Neritidae, Turridae *s.l.*). The overall number of sequences and sudden increases in sequence availability can often be traced to the contributions of a single individual or research group. These may represent large numbers of sequences for a limited number of species (e.g. population genetic studies of Neritidae) or targeted systematic studies of particular taxonomic groups. For example, of the 92 COI turrid *s.l.* sequences available, all but one (109/110) were generated by our own research group. The potential impact of such targeted studies is evident when comparing the resolving power of existing databases including and excluding our own sequences (see Fig. 3). However, the majority of the most speciose families remain enormously underrepresented.

Conclusion

The COI gene, the universal barcode for animals, in association with two other mitochondrial markers, has demonstrated its ability to identify gastropod spawn, and thus constitutes an important tool for taxonomic identification of various animal life stages. However, our results indicate that sequence availability in existing genetic databases is disproportionately low given the known and estimated diversity of gastropods, and that the taxonomic coverage is highly biased towards shallow water species and/or highly collectible macro-mollusks. Consequently, barcodes are currently unable to provide species-level identifications for most unknowns, even for well-known, shallow water European species. In cases when species-level identification is impossible, tree-based methods are useful for refining the higher taxonomic placement of unidentified samples, but should be treated with caution. Barcoding efforts under development for targeted vertebrates (birds, fishes), "charismatic" invertebrates or economically important groups currently allow positive identification to species in ~99% of unknowns in some cases (e.g. Wong & Hanner 2008), but similar projects should be developed for megadiverse groups such as mollusks, not only to facilitate taxonomic expertise but also to enhance species discovery in such poorly known groups.

References

Ahrens D, Monaghan MT, Vogler AP (2007) DNA-based taxonomy for associating adults and larvae in multi-species assemblages of chafers (Coleoptera: Scarabaeidae). *Molecular Phylogenetics and Evolution* 44, 436–449.

Bandel K (1976a) Observations on spawn, embryonic development and ecology of some Caribbean lower Mesogastropoda. *The Veliger* 18, 249-270.

379 Bandel K (1976b) Spawning, Development and ecology of some higher neogastropoda from
380 the Carribean Sea of Colombia (South America). *The Veliger* 19, 176-193.
381 Barber P, Boyce SL (2006) Estimating diversity of Indo-Pacific coral reef stomatopods
382 through DNA barcoding of stomatopod larvae. *Proceedings of the Royal Society B*
383 273, 2053-2061.
384 Bichain JM, Boisselier MC, Bouchet P, Samadi S (2007) Delimiting species in the genus
385 *Bythinella* (Mollusca: Caenogastropoda: Rissooidea): molecular and morphometric
386 approachs. *Malacologia* 49, 291-311.
387 Blaxter ML (2004) The promise of a DNA taxonomy. *Philosophical Transactions of the*
388 *Royal Society of London B* 359, 669-679.
389 Bouchet P (2006) The magnitude of marine biodiversity. In: *The Exploration of Marine*
390 *Biodiversity - Scientific and Technological Challenges* (ed. BBVA F).
391 Bouchet P, Lozouet P, Maestrati P, Héros V (2002) Assessing the magnitude of species
392 richness in tropical marine environments : exceptionnally high numbers of molluscs at
393 a new Caledonian site. *Biological Journal of the Linnean Society* 75, 421-436.
394 D'Asaro CN (1970) Egg capsules of prosobranch mollusks from south Florida and the
395 Bahamas, and notes on spawning in the laboratory. *Bulletin of Marine Science* 20,
396 414-440.
397 Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of
398 mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates.
399 *Molecular Marine Biology and Biotechnology* 3, 294-299.
400 Fretter V, Graham A (1962) *British Prosobranch Molluscs; their functional anatomy and*
401 *ecology* Ray Society, London.

402 Gustafson RG, Littlewood DTJ, Lutz RA (1991) Gastropod egg capsules and their contents
403 from deep-sea hydrothermal vent environments. *Biological Bulletin* 180, 34-55.
404 Habe T (1960) Egg masses and egg capsules of some Japanese marine prosobranchiate
405 gastropod. *Bulletin of the Marine Biological Station of Asamushi* 10, 121-126.
406 Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis
407 program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41, 95-98.
408 Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through
409 DNA Barcodes. *Proceedings of the Royal Society B* 270, 313-321.
410 Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM (2004) Identification of birds through
411 DNA barcodes. *PLOS Biology* 2, 1657-1663.
412 Holland BS, Dawson MN, Crow GL, Hofmann DK (2004) Global phylogeography of
413 *Cassiopea* (Scyphozoa: Rhizostomeae): molecular evidence for cryptic species and
414 multiple invasions of the Hawaiian Islands. *Marine Biology* 145, 1119-1128.
415 Huelsenbeck JP, Ronquist F, Hall B (2001) MrBayes: bayesian inference of phylogeny.
416 *Bioinformatics* 17, 754-755.
417 Huelsken T, Marek T, Schreiber S, Schmidt I, Hollmann M (2008) The Naticidae (Mollusca:
418 Gastropoda) of Giglio Island (Tuscany, Italy): Shell characters, live animals, and a
419 molecular analysis of egg masses. *Zootaxa* 1770, 1-40.
420 Jones WJ, Preston CM, Marin R, Scholin CA, Vrijenhoek RC (2008) A robotic molecular
421 method for in situ detection of marine invertebrate larvae. *Molecular Ecology*
422 *Resources* 8, 540-550.
423 Kantor YI, Puillandre N, Olivera BM, Bouchet P (2008) Morphological proxies for
424 taxonomic decisions in turrids (Mollusca, Neogastropoda): a test of the value of shell
425 and radula characters using molecular data. *Zoological Science*, (in press).

426 Knudsen J (1950) Egg capsules and development of some marine prosobranchs from tropical
427 West Africa. *Atlantide Report* 1, 85-130.
428 Kohn AJ (1961) Studies on spawning behavior, egg masses, and larval development in the
429 gastropod genus Conus, Part I. Observation on nine species in Hawaii. *Pacific Science*
430 15, 163-179.
431 Le Goff-Vitry MC, Chipman AD, Comtet T (2007) *In situ* hybridization on whole larvae: a
432 novel method for monitoring bivalve larvae. *Marine Ecology progress Series* 343,
433 161-172.
434 Lebour MV (1937) The eggs and larvae of the British prosobranchs with special reference to
435 those living in the plankton. *Journal of the Marine Biological Association of the*
436 *United Kingdom* 22, 105-166.
437 Livi S, Cordisco C, Damiani C, Romanelli M, Crosetti D (2006) Identification of bivalve
438 species at an early developmental stage through PCR-SSCP and sequence analysis of
439 partial 18S rDNA. *Marine Biology* 149, 1149-1161.
440 Lorenz F (2002) *New Worldwide cowries. Descriptions of new taxa and revisions of selected*
441 *groups of living Cypraeidae (Mollusca: Gastropoda)* Conchbooks.
442 Lutz R, Bouchet P, Jablonski D, Turner R, Warén A (1986) Larval ecology of mollusks at
443 deep-sea hydrothermal vents. *American Malacological Bulletin* 4, 49-54.
444 Pegg GG, Sinclair B, Briskey L, Aspden WJ (2006) MtDNA barcode identification of fish
445 larvae in the southern Great Barrier Reef, Australia. *Scientia Marina* 70, 7-12.
446 Pradillon F, Schmidt A, Peplies J, Dubilier N (2007) Species identification of marine
447 invertebrate early stages by whole-larvae *in situ* hybridisation of 18S ribosomal RNA.
448 *Marine Ecology progress Series* 333, 103-116.

449 Ratnasingham S, Hebert PDN (2007) BOLD : The Barcode of Life Data System. *Molecular*
450 *Ecology Notes* 7, 355-364.
451 Remigio EA, Hebert PDN (2003) Testing the utility of partial COI sequences for phylogenetic
452 estimates of gastropod relationships. *Molecular Phylogenetics and Evolution* 29, 641-
453 647.
454 Robertson R (1974) Marine prosobranch gastropods: larval studies and systematics. *Thalassia*
455 *Jugoslavica* 10, 213-238.
456 Schindel DE, Miller SE (2005) DNA barcoding a useful tool for taxonomists. *Nature* 435, 17.
457 Steinke D, Vences M, Salzburger W, Meyer A (2005) TaxI: a software tool for DNA
458 barcoding using distance methods. *Proceedings of the Royal Society B* 360, 1975-
459 1980.
460 Thomas M, Rahariviololoniaina M, Glaw F, Vences M, Vireites DR (2005) Montane Tadpoles
461 in Madagascar: Molecular Identification and Description of the Larval Stages of
462 *Mantidactylus elegans, Mantidactylus madecassus,* and *Boophis laurenti* from the
463 Andringitra Massif. *Copeia* 1, 174-183.
464 Thorson G (1935) Studies on the egg-capsules and development of Arctic marine
465 prosobranchs. *Medd. Om Gronland* 100, 20-73.
466 Thorson G (1940a) Notes on the egg-capsules of some North-Atlantic prosobranchs of the
467 genus *Troschelia*, *Chysodomus*, *Volutopsis*, *Sipho* and *Trophon*. *Videnskabelige*
468 *Meddelelser fra Dansk naturhistorisk Forening i Kobenhavn* 104, 251-265.
469 Thorson G (1940b) *Studies on the egg masses and larval development of Gastropoda from the*
470 *Iranian Gulf. Danish Scientific Investigations in Iran, Part 2* Ejnar Munksgaard,
471 Copenhagen.

Tokioka R (1950) Droplets from the plankton net. V. New names for egg capsules of littorinid gastropods. *Publication of the Seto Marine Biological Laboratory* 1, 151-152.

Tucker JK (2004) Catalog of Recent and fossil turrids (Mollusca: Gastropoda). *Zootaxa* 682, 1-1295.

Vences M, Thomas M, Bonett RM, Vieites DR (2005) Deciphering amphibian diversity through DNA barcoding: chances and challenges. *Philosophical Transactions of the Royal Society B: Biological Sciences* 360, 1859-1868.

Victor BC (2007) *Coryphopterus kuna*, a new goby (Perciformes: Gobiidae: Gobiinae) from the western Caribbean, with the identification of the late larval stage and an estimate of the pelagic larval duration. *Zootaxa* 1526, 51-61.

Winner BE (1987) *A field guide to molluscan spawn. Vol. 1*. EBM, North Palm Beach, Florida.

Wong EH-K, Hanner RH (2008) DNA barcoding detects market substitution in North American seafood. *Food Research International* 41, 828-837.

Acknowledgments

The AURORA 2007 cruise (Principal Investigators: Marivene Manuel, NMP, and Philippe Bouchet, MNHN) on board M/V DA-BFAR was a joint project between the Philippine Bureau of Fisheries and Aquatic Resources (BFAR), National Museum of the Philippines (NMP), and Muséum National d'Histoire Naturelle (MNHN). It was made possible by a grant from the Lounsbery Foundation, and was carried under a *Census of Marine Life / Census of Margins* umbrella. This work was supported by the "Consortium National de Recheche en Génomique", and the "Service de Systématique Moléculaire" of the Muséum National

d'Histoire Naturelle (IFR 101). It is part of the agreement n°2005/67 between the Genoscope and the Muséum National d'Histoire Naturelle on the project "Macrophylogeny of life" directed by Guillaume Lecointre". We are grateful to André Hoareau and Jacques Pelorce for the two specimens collected in France. We are also pleased to thank Barbara Buge (MNHN) for the pictures of the egg capsules and Yuri Kantor (Russian Academy of Sciences) for discussions about egg capsule morphology.

Figure Legends

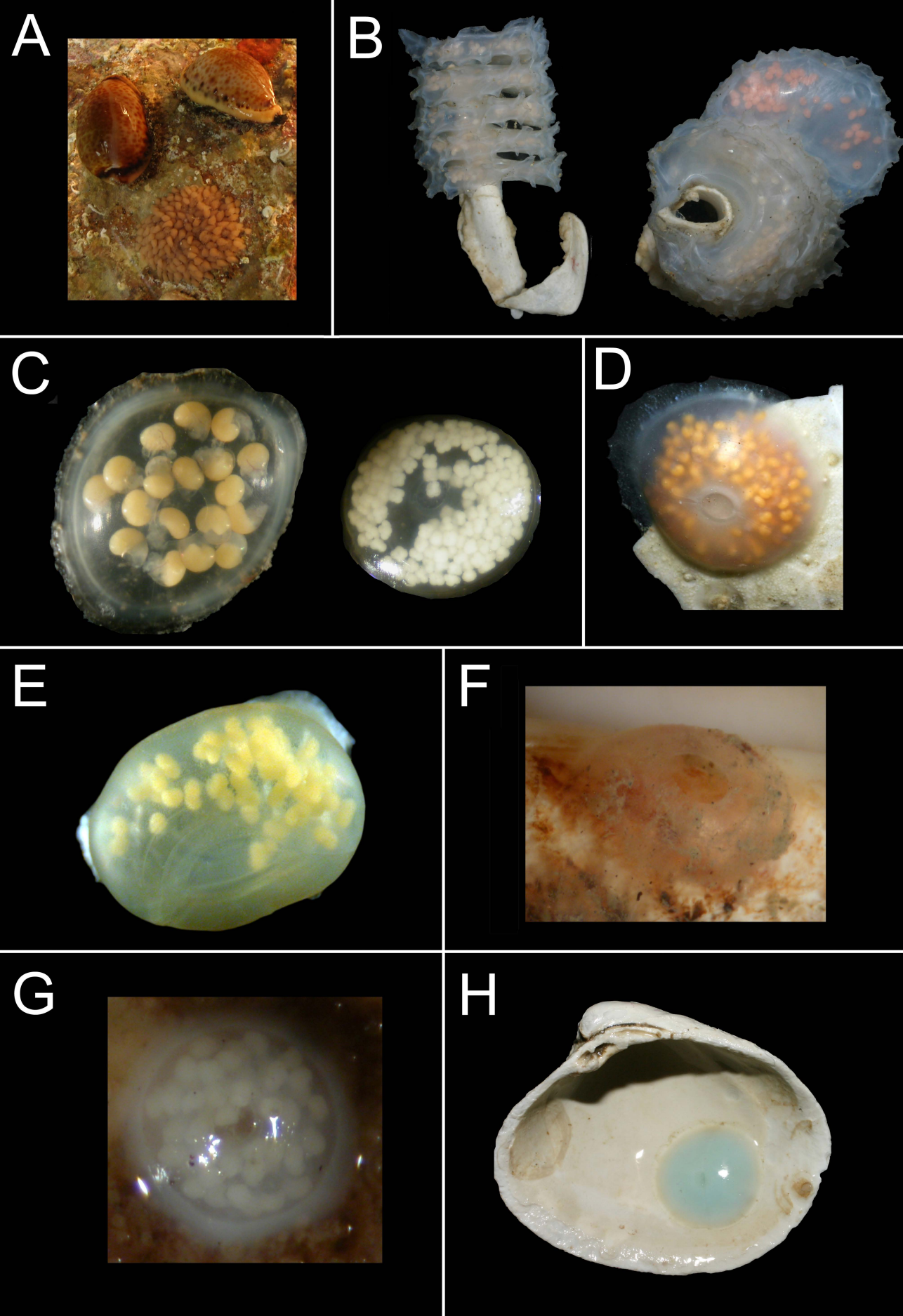
Fig. 1: Egg capsules used in molecular analyses and resulting identifications. A. EC2 (*Erosaria spurca* – egg capsule with adults) B. EC3 (*Conus* sp.) C. EC9 (left) and EC11 (*Granulifusus* sp., at two different stages) D. EC8 (*Comitas* sp.) E. EC5 (*Belomitra* sp.) F. EC14 (unidentified caenogastropod) G. EC23 (Conidae, Clathurellinae) H. EC26 (Conidae, Raphitominae) fixed on a bivalve shell. Photo credits: André Hoareau (A) and Barbara Buge (MNHN) (B-H). Color photos are also published in MorphoBank (http://morphobank.geongrid.org/permalink/?P232)

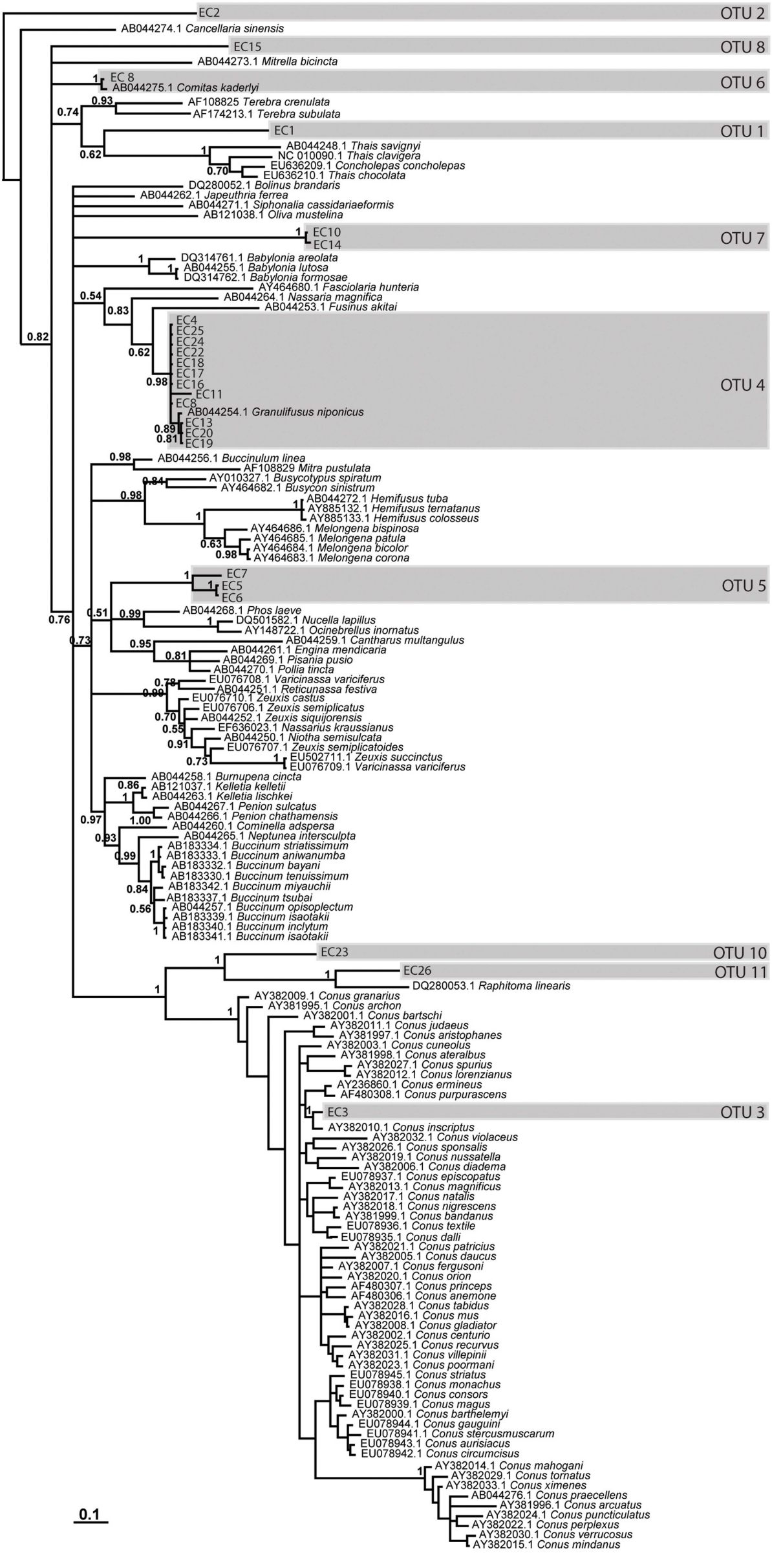
Fig. 2: Bayesian phylogram generated from all 16S neogastropod sequences in GenBank and 23 neogastropod egg capsule specimens from the Philippines and France. *Erosaria spurca* (EC2), a non-neogastropod, is the outgroup.

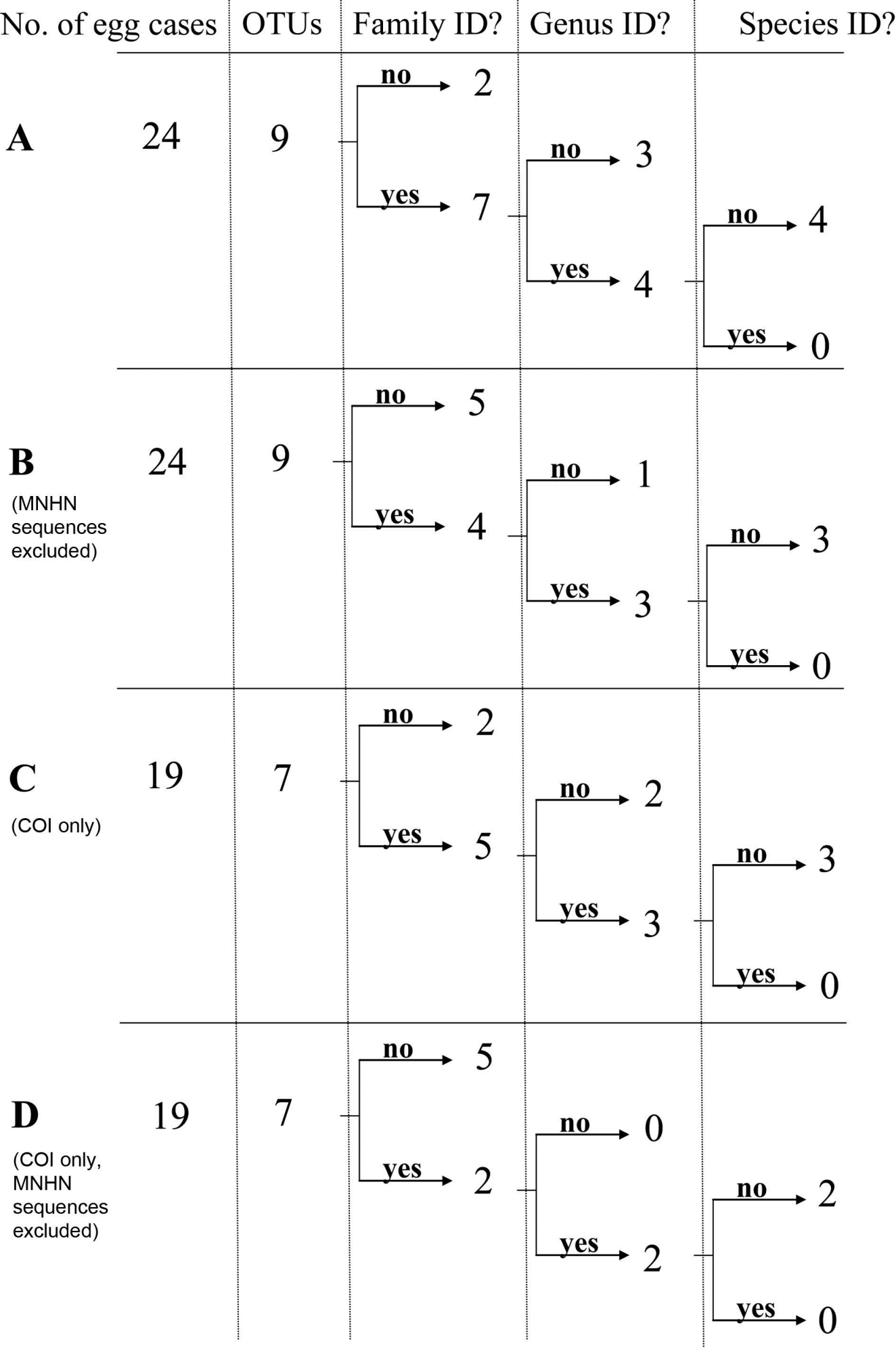
Fig. 3: Number of positive identifications of unknown Philippine egg capsules at different taxonomic levels based on all genes (A, B) or COI only (C, D). A and C include all GenBank sequences; B and D exclude sequences produced by our own research group, demonstrating the impact of a single research group on the global dataset.

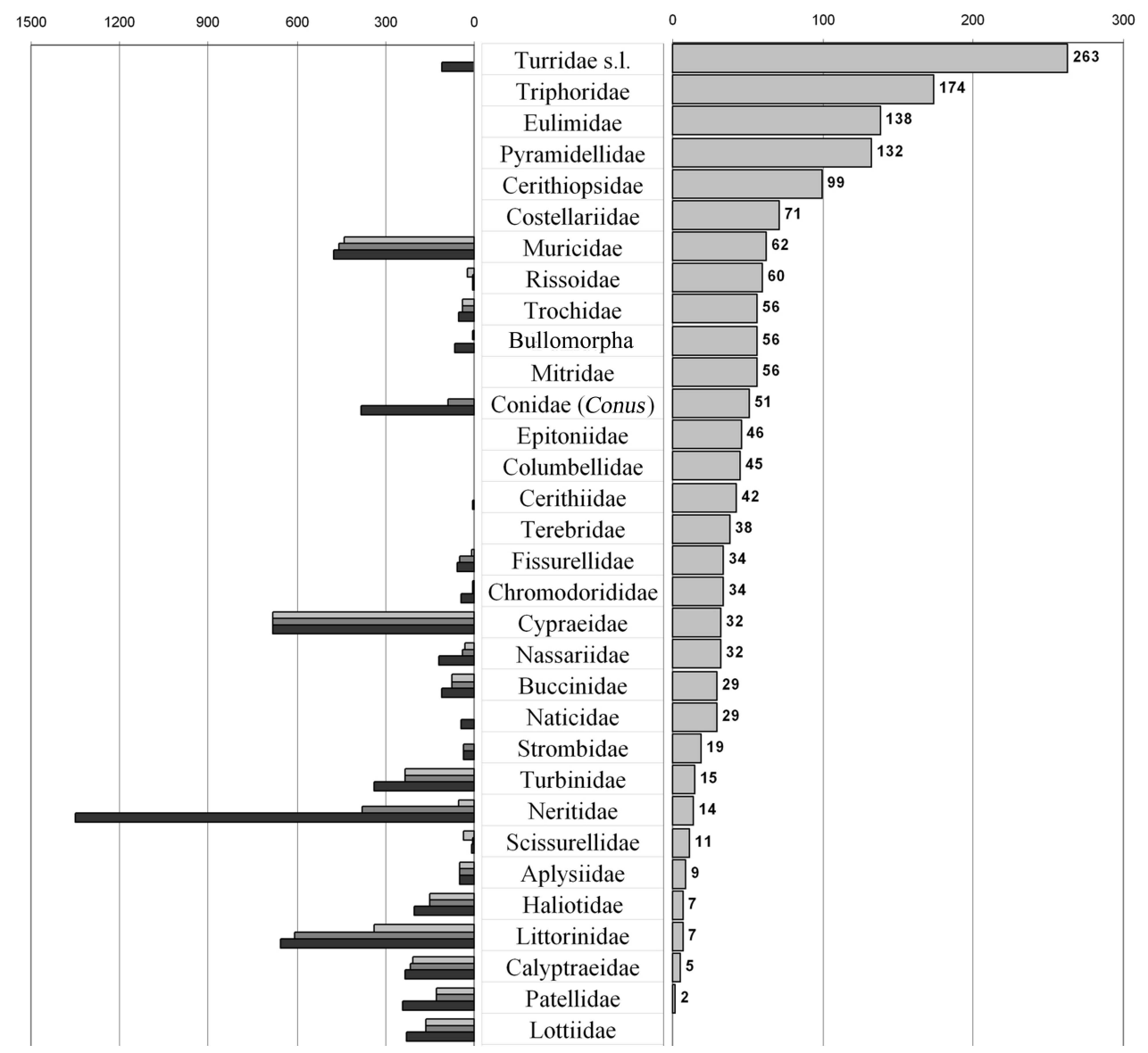
Fig. 4: GenBank coverage and actual marine gastropod diversity for gastropod families. Histogram on the left indicates number of COI sequences in GenBank: light gray – July 28, 2006; dark gray – April 6, 2007; black – July 4, 2008. Histogram on the right shows species richness of gastropod families at a well-sampled tropical site (Koumac, New Caledonia; Bouchet *et al*. 2002). Only groups with more than 10 species in Koumac or more than 100 sequences in GenBank (in 2006) are represented here. The decreasing of the number of

sequences for some groups through time is due to the fact that some sequences were removed from GenBank.


A
B
C
D
E
F
G
H


EC2
OTU 2
AB044274.1 Cancellaria sinensis
EC15
OTU 8
AB044273.1 Mitrella bicincta
EC 8
AB044275.1 Comitas kaderlyi
OTU 6
AF108825 Terebra crenulata
AF174213.1 Terebra subulata
EC1
OTU 1
AB044248.1 Thais savignyi
NC 010090.1 Thais clavigera
EU636209.1 Concholepas concholepas
EU636210.1 Thais chocolata
DQ280052.1 Bolinus brandaris
AB044262.1 Japeuthria ferrea
AB044271.1 Siphonalia cassidariaeformis
AB121038.1 Oliva mustelina
EC10
EC14
OTU 7
DQ314761.1 Babylonia areolata
AB044255.1 Babylonia lutosa
DQ314762.1 Babylonia formosae
AY464680.1 Fasciolaria hunteria
AB044264.1 Nassaria magnifica
AB044253.1 Fusinus akitai
EC4
EC25
EC24
EC22
EC18
EC17
EC16
EC11
EC8
AB044254.1 Granulifusus niponicus
EC13
EC20
EC19
OTU 4
AB044256.1 Buccinulum linea
AF108829 Mitra pustulata
AY010327.1 Busycotypus spiratum
AY464682.1 Busycon sinistrum
AB044272.1 Hemifusus tuba
AY885132.1 Hemifusus ternatanus
AY885133.1 Hemifusus colosseus
AY464686.1 Melongena bispinosa
AY464685.1 Melongena patula
AY464684.1 Melongena bicolor
AY464683.1 Melongena corona
EC7
EC5
EC6
OTU 5
AB044268.1 Phos laeve
DQ501582.1 Nucella lapillus
AY148722.1 Ocinebrellus inornatus
AB044259.1 Cantharus multangulus
AB044261.1 Engina mendicaria
AB044269.1 Pisania pusio
AB044270.1 Pollia tincta
EU076708.1 Varicinassa variciferus
AB044251.1 Reticunassa festiva
EU076710.1 Zeuxis castus
EU076706.1 Zeuxis semiplicatus
AB044252.1 Zeuxis siquijorensis
EF636023.1 Nassarius kraussianus
AB044250.1 Niotha semisulcata
EU076707.1 Zeuxis semiplicatoides
EU502711.1 Zeuxis succinctus
EU076709.1 Varicinassa variciferus
AB044258.1 Burnupena cincta
AB121037.1 Kelletia kelletii
AB044263.1 Kelletia lischkei
AB044267.1 Penion sulcatus
AB044266.1 Penion chathamensis
AB044260.1 Cominella adspersa
AB044265.1 Neptunea intersculpta
AB183334.1 Buccinum striatissimum
AB183333.1 Buccinum aniwanumba
AB183332.1 Buccinum bayani
AB183330.1 Buccinum tenuissimum
AB183342.1 Buccinum miyauchii
AB183337.1 Buccinum tsubai
AB044257.1 Buccinum opisoplectum
AB183339.1 Buccinum isaotakii
AB183340.1 Buccinum inclytum
AB183341.1 Buccinum isaotakii
EC23
OTU 10
EC26
OTU 11
DQ280053.1 Raphitoma linearis
AY382009.1 Conus granarius
AY381995.1 Conus archon
AY382001.1 Conus bartschi
AY382011.1 Conus judaeus
AY381997.1 Conus aristophanes
AY382003.1 Conus cuneolus
AY381998.1 Conus ateralbus
AY382027.1 Conus spurius
AY382012.1 Conus lorenzianus
AY236860.1 Conus ermineus
AF480308.1 Conus purpurascens
EC3
OTU 3
AY382010.1 Conus inscriptus
AY382032.1 Conus violaceus
AY382026.1 Conus sponsalis
AY382019.1 Conus nussatella
AY382006.1 Conus diadema
EU078937.1 Conus episcopatus
AY382013.1 Conus magnificus
AY382017.1 Conus natalis
AY382018.1 Conus nigrescens
AY381999.1 Conus bandanus
EU078936.1 Conus textile
EU078935.1 Conus dalli
AY382021.1 Conus patricius
AY382005.1 Conus daucus
AY382007.1 Conus fergusoni
AY382020.1 Conus orion
AF480307.1 Conus princeps
AF480306.1 Conus anemone
AY382028.1 Conus tabidus
AY382016.1 Conus mus
AY382008.1 Conus gladiator
AY382002.1 Conus centurio
AY382025.1 Conus recurvus
AY382031.1 Conus villepinii
AY382023.1 Conus poormani
EU078945.1 Conus striatus
EU078938.1 Conus monachus
EU078940.1 Conus consors
EU078939.1 Conus magus
AY382000.1 Conus barthelemyi
EU078944.1 Conus gauguini
EU078941.1 Conus stercusmuscarum
EU078943.1 Conus aurisiacus
EU078942.1 Conus circumcisus
AY382014.1 Conus mahogani
AY382029.1 Conus tornatus
AY382033.1 Conus ximenes
AB044276.1 Conus praecellens
AY381996.1 Conus arcuatus
AY382024.1 Conus puncticulatus
AY382022.1 Conus perplexus
AY382030.1 Conus verrucosus
AY382015.1 Conus mindanus
0.1


No. of egg cases
OTUs
Family ID?
Genus ID?
Species ID?
A
24
9
no
2
yes
7
no
3
yes
4
no
4
yes
0
B
(MNHN sequences excluded)
24
9
no
5
yes
4
no
1
yes
3
no
3
yes
0
C
(COI only)
19
7
no
2
yes
5
no
2
yes
3
no
3
yes
0
D
(COI only, MNHN sequences excluded)
19
7
no
5
yes
2
no
0
yes
2
no
2
yes
0


1500
1200
900
600
300
0
0
100
200
300
Turridae s.l.
Triphoridae
Eulimidae
Pyramidellidae
Cerithiopsidae
Costellariidae
Muricidae
Rissoidae
Trochidae
Bullomorpha
Mitridae
Conidae (Conus)
Epitoniidae
Columbellidae
Cerithiidae
Terebridae
Fissurellidae
Chromodorididae
Cypraeidae
Nassariidae
Buccinidae
Naticidae
Strombidae
Turbinidae
Neritidae
Scissurellidae
Aplysiidae
Haliotidae
Littorinidae
Calyptraeidae
Patellidae
Lottiidae
263
174
138
132
99
71
62
60
56
56
56
51
46
45
42
38
34
34
32
32
29
29
19
15
14
11
9
7
7
5
2

							COI		
							BOLD		Genbank
ID	**Station**	**COI**	**16S**	**12S**	**N°**	**OTU**	**ID**	**%**	**ID**
P3		EU870569	EU870545	EU870520	EC1	1	x	x	*Fluminicola virens*
P4		EU870570	EU870546	EU870521	EC2	2	*Erosaria spurca*	**99,51**	*Erosaria spurca*
496	CP2665	EU870571	EU870547	EU870522	EC3	3	*Conus sulcatus*	88,94	*Conus venulatus*
648	CP2678	EU870572	EU870548	EU870523	EC4	4	*Busycon carica*	87,19	*Busycon sinistrum*
694	CP2684	EU870573	EU870549	EU870524	EC5	5	*Batillaria multiformis*	87,63	*Belomitra sp.*
695	CP2684	EU870574	EU870550	EU870525	EC6	5	*Batillaria multiformis*	87,45	*Belomitra sp.*
696	CP2684	EU870575	EU870551	EU870526	EC7	5	*Busycon sinistrum*	85,74	*Belomitra sp.*
861	CC2700	EU870576	EU870552	EU870527	EC8	6	*Comitas*	**97,84**	*Pyrgulopsis glandulosa*
898	CP2707	EU870577	EU870553	EU870528	EC9	4	*Busycon carica*	87,26	*Nucella lapillus*
1057	CP2712	EU870578	EU870554	EU870529	EC10	7	*Lacuna pallidula*	84,64	*Ilyanessa obsoleta*
1285	CC2724		EU870555	EU870530	EC11	4			
1402	CP2727			EU870531	EC12	8			
1404	CP2727	EU870579	EU870556	EU870532	EC13	4	*Busycon carica*	87,61	*Ilyanessa obsoleta*
1405	CP2727		EU870557		EC14	7			
1409	CP2727		EU870558	EU870533	EC15	8			
1646	CP2735	EU870580	EU870559	EU870534	EC16	4	*Busycon carica*	87,09	*Busycon sinistrum*
1647	CP2735	EU870581	EU870560	EU870535	EC17	4	*Busycon carica*	87,3	*Busycon sinistrum*
1648	CP2735	EU870582	EU870561	EU870536	EC18	4	*Busycon carica*	86,99	*Busycon sinistrum*
1681	CC2743	EU870583	EU870562	EU870537	EC19	4	*Busycon carica*	87,26	*Acanthinucella punctulata*
1682	CC2743	EU870584	EU870563	EU870538	EC20	4	*Busycon carica*	87,26	*Acanthinucella punctulata*
1689	CC2745			EU870539	EC21	9			
1690	CC2745	EU870585	EU870564	EU870540	EC22	4	*Busycon carica*	87,37	*Busycon sinistrum*
1691	CC2745	EU870586	EU870565	EU870541	EC23	10	*Nannodiella*	87,4	*Gymnobela sp.*
1692	CC2745	EU870587	EU870566	EU870542	EC24	4	*Busycon carica*	87,19	*Busycon sinistrum*
1693	CC2745	EU870588	EU870567	EU870543	EC25	4	*Busycon carica*	87,19	*Busycon sinistrum*
1775	CP2751	EU870589	EU870568	EU870544	EC26	11	*Gymnobela*	87,52	*Nannodiella*

		16S			12S		
	Tree	Genbank		Tree	Genbank		Tree
%		ID	%		ID	%	
76		*Conus circumcisus*	82		*Coralliophila meyendorffii*	**99**	*C. meyendorffii*
98		*Erosaria spurca*	**99**		*Cypraea annulus*	94	
90	*Conus sp.*	*Conus radiatus*	**95**	*Conus sp.*	*Conus textile*	90	*Conus*
86		*Granulifusus niponicus*	**98**	*G. niponicus*	*Ilyanassa obsoleta*	86	
89	*Belomitra sp.*	*Buccinum opisoplectum*	89		*Ilyanassa obsoleta*	84	
89	*Belomitra sp.*	*Buccinum opisoplectum*	89		*Ilyanassa obsoleta*	83	
89	*Belomitra sp.*	*Burnupena cincta*	89		*Gemmula sogodensis*	83	
85	*Lophiotoma*	*Comitas kaderlyi*	**98**	*C. kaderlyi*	*Gemmula rosario*	84	
87		*Granulifusus niponicus*	**98**	*G. niponicus*	*Ilyanassa obsoleta*	87	
84		*Buccinulum linea*	86		*Ilyanassa obsoleta*	79	
		Granulifusus niponicus	**95**	*G. niponicus*	*Ilyanassa obsoleta*	87	
					Gemmula rosario	83	Turridae
86		*Granulifusus niponicus*	**98**	*G. niponicus*	*Ilyanassa obsoleta*	85	
		Buccinulum linea	85				
		Penion chathamensis	85		*Lophiotoma unedo*	79	Turridae
86		*Granulifusus niponicus*	**97**	*G. niponicus*	*Ilyanassa obsoleta*	86	
86		*Granulifusus niponicus*	**98**	*G. niponicus*	*Ilyanassa obsoleta*	86	
85		*Granulifusus niponicus*	**98**	*G. niponicus*	*Ilyanassa obsoleta*	86	
84		*Granulifusus niponicus*	**98**	*G. niponicus*	*Ilyanassa obsoleta*	86	
84		*Granulifusus niponicus*	**98**	*G. niponicus*	*Ilyanassa obsoleta*	86	
					Lophiotoma polytropa	84	Cypraeidae
86		*Granulifusus niponicus*	**97**	*G. niponicus*	*Ilyanassa obsoleta*	86	
86	Raphitominae	*Conus consors*	86	*Conidae*	*Ilyanassa obsoleta*	82	
86		*Granulifusus niponicus*	**97**	*G. niponicus*	*Ilyanassa obsoleta*	87	
86		*Granulifusus niponicus*	**97**	*G. niponicus*	*Ilyanassa obsoleta*	87	
87	Clathurellinae	*Raphitoma linearis*	87	*Conidae*	*Lophiotoma acuta*	81	

Final ID
C. meyendorffii
Erosaria spurca
Conus sp.
Granulifus sp.
Belomitra sp.
Belomitra sp.
Belomitra sp.
Comitas sp.
Granulifus sp.
?
Granulifus sp.
Turridae
Granulifus sp.
?
Turridae
Granulifus sp.
Granulifus sp.
Granulifus sp.
Granulifus sp.
Granulifus sp.
?
Granulifus sp.
Conidae (Raphitominae)
Granulifus sp.
Granulifus sp.
Conidae (Clathurellinae)