



**HAL**  
open science

## Large-scale species delimitation method for hyperdiverse groups

N. Puillandre, M. V Modica, Y. Zhang, L. Sirovich, M.-C Boisselier, C. Cruaud, M. Holford, S. Samadi

### ► To cite this version:

N. Puillandre, M. V Modica, Y. Zhang, L. Sirovich, M.-C Boisselier, et al.. Large-scale species delimitation method for hyperdiverse groups. *Molecular Ecology*, 2012, 21 (11), pp.2671-2691. hal-02002432

**HAL Id: hal-02002432**

**<https://hal.science/hal-02002432v1>**

Submitted on 31 Jan 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Large Scale Species Delimitation Method for Hyperdiverse Groups**

2

3 N. Puillandre<sup>1</sup>, M. V. Modica<sup>2</sup>, Y. Zhang<sup>3</sup>, L. Sirovich<sup>3</sup>, M.-C. Boisselier<sup>1,4</sup>, C. Cruaud<sup>5</sup>,

4 M. Holford<sup>6</sup>, S. Samadi<sup>1,4</sup>

5

6 <sup>1</sup> “Systématique, Adaptation et Evolution”, UMR 7138 UPMC-IRD-MNHN-CNRS (UR  
7 IRD 148), Muséum National d'Histoire Naturelle, Département Systématique et Evolution,  
8 CP 26, 57 Rue Cuvier, F-75231 Paris Cedex 05, France.

9 <sup>2</sup> Dipartimento di Biologia Animale e dell’Uomo, “La Sapienza”, University of Rome,  
10 Viale dell’Universita` 32, 00185 Rome, Italy

11 <sup>3</sup> Laboratory of Applied Mathematics, Mount Sinai School of Medicine, NY, NY USA,

12 <sup>4</sup> Service de systématique moléculaire, UMS2700 CNRS-MNHN, Muséum National  
13 d'Histoire Naturelle, Département Systématique et Evolution, CP 26, 57 Rue Cuvier, F-  
14 75231 Paris Cedex 05, France.

15 <sup>5</sup> GENOSCOPE, Centre National de Séquençage, 2 rue Gaston Crémieux, CP 5706, 91057  
16 Evry Cedex France.

17 <sup>6</sup>The City University of New York-York College & The Graduate Center, and The  
18 American Museum of Natural History. NY, NY USA

19

20 Keywords: Barcoding, Conoidea, GMYC method, Integrative taxonomy, ABGD method,  
21 Turridae.

22

23 Corresponding author: Puillandre, Nicolas, “Systématique, Adaptation et Evolution”, UMR  
24 7138 UPMC-IRD-MNHN-CNRS (UR IRD 148), Muséum National d'Histoire Naturelle,

25 Département Systématique et Evolution, CP 26, 57 Rue Cuvier, F-75231 Paris Cedex 05,

26 France. Fax: +33 1 40 79 38 44. [puillandre@mnhn.fr](mailto:puillandre@mnhn.fr)

27

28 Running title: Large Scale Species Delimitation

29

30 **Abstract**

31           Accelerating the description of biodiversity is a major challenge as extinction rates  
32 increase. Integrative taxonomy combining molecular, morphological, ecological and  
33 geographical data is seen as the best route to reliably identify species. Classic molluscan  
34 taxonomic methodology proposes primary species hypotheses (PSHs) based on shell  
35 morphology. However, in hyperdiverse groups, such as the molluscan family Turridae, where  
36 most of the species remain unknown and for which homoplasy and plasticity of  
37 morphological characters is common, shell based PSHs can be arduous. A four-pronged  
38 approach was employed to generate robust species hypotheses of a 1000 specimen South-  
39 West Pacific Turridae dataset in which: (i) analysis of COI DNA Barcode gene is coupled  
40 with (ii) species delimitation tools GMYC (General Mixed Yule Coalescence Method) and  
41 ABGD (Automatic Barcode Gap Discovery) to propose primary species hypotheses that are  
42 then (iii) visualized using Klee diagrams, and (iv) evaluated with additional evidence, such as  
43 nuclear gene rRNA 28S, morphological characters, geographical and bathymetrical  
44 distribution to determine conclusive secondary species hypotheses (SSHs). The integrative  
45 taxonomy approach applied identified 87 Turridae species, more than doubling the amount  
46 previously known in the *Gemmula* genus. In contrast to a predominantly shell-based  
47 morphological approach, which over the last 30 years proposed only 13 new species names  
48 for the Turridae genus *Gemmula*, the integrative approach described here identified 27 novel  
49 species hypotheses not linked to available species names in the literature. The formalized  
50 strategy applied here outlines an effective and reproducible protocol for large scale species  
51 delimitation of hyperdiverse groups.

52 **INTRODUCTION**

53         The rapidly increasing rate of biodiversity extinction coupled with the magnitude of  
54 unknown biodiversity requires accurate and effective methods of species delimitation (Wiens  
55 2007). The onset of the 21<sup>st</sup> century has seen the development of technological advances that  
56 can accelerate the description of biodiversity (Wheeler 2009). One example of which are the  
57 DNA barcoding initiatives, which are an attempt to identify specimens at the species-level  
58 using a single-gene library (Hebert *et al.* 2003; Vernooy *et al.* 2010). DNA barcoding has  
59 proved effective in identifying larvae (Ahrens *et al.* 2007), processed biological products  
60 (Smith *et al.* 2008) or gut contents (Garros *et al.* 2008), as well as a taxonomic tool to aid in  
61 defining species, particularly when morphological characters are shown to be poor proxy of  
62 species boundaries (Taylor *et al.* 2006). For the bulk of undescribed biodiversity, the single-  
63 gene approach of DNA-barcoding project may be used, not to identify specimens, but as a  
64 primary glance, *i.e.* primary species hypotheses for approximating species descriptions  
65 (Goldstein& DeSalle 2011).

66         Problems linked to a single gene approach, such as the presence of pseudogenes  
67 (Lorenz *et al.* 2005), incomplete lineage sorting (Funk& Omland 2003) or introgression  
68 (Chase *et al.* 2005), accentuate the need for an integrated analyses for species identification.  
69 One strategy used to avoid single-gene pitfalls is to increase the gene sampling to two or  
70 more, if possible, unlinked, genes (see e.g. Boissin *et al.* 2008; Knowles& Carstens 2007;  
71 O'Meara 2010; Ross *et al.* 2010; Weisrock *et al.* 2006). Another approach is to challenge the  
72 patterns of diversity drawn using molecular data with other sources of evidence, such as  
73 morphological characters, ecological factors, geographic distributions, and other criteria (*e.g.*  
74 monophyly, reproductive isolation). This process of modification and validation of the species  
75 hypotheses that compiles various data and criteria, is referred to as integrative taxonomy  
76 (Barberousse& Samadi 2010; Dayrat 2005; De Queiroz 2007; Reeves& Richards 2011;

77 Samadi& Barberousse 2009; Schlick-Steiner *et al.* 2009; Wiens 2007; Will *et al.* 2005; Yeates  
78 *et al.* 2010). An integrative approach, starting with molecular characters, is particularly  
79 applicable for hyperdiverse groups, where most species are unknown and for which the  
80 quality of morphological characters as proxies for determining species boundaries is  
81 circumspect.

82         A group for which integrative taxonomy is particularly promising is the family  
83 Turridae *s.s.* (Bouchet *et al.* 2011), which are predatory marine snails including a large  
84 number of species, many of them being rare (Bouchet *et al.* 2009). Homoplasy and  
85 phenotypic plasticity of shell characters (Puillandre *et al.* 2010) renders traditional shell  
86 morphology based taxonomic approaches problematic for molluscs in general, and  
87 particularly for hyperdiverse and poorly known groups such as the Turridae. The Turridae are  
88 also a promising group to investigate because they are part of the Conoidea superfamily  
89 (Bouchet *et al.* 2011), which includes the genus *Conus* and the family Terebridae. Species  
90 diversity in the Conoidea is believed to be linked to the diversity of their venom (Duda 2008).  
91 Peptide toxins found in the venom of *Conus* snails, conotoxins, have been used extensively  
92 since the 1970's to characterize the structure and function of ion channels and receptors in the  
93 nervous system (Terlau& Olivera 2004). In 2004, the first conotoxin drug, ziconotide (Prialt)  
94 from *Conus magus*, became commercially available as an analgesic for chronic pain in HIV  
95 and cancer patients (Miljanich 2004; Olivera *et al.* 1987). In comparison to conotoxins, toxins  
96 from the Turridae, turritoxins, are not as well characterized and are an active area of study in  
97 the search for novel ligands that modulate the neuronal circuit and are promising therapeutic  
98 compounds (Lopez-Vera *et al.* 2004). Getting a grasp on the diversity of the Turridae would  
99 enhance the investigation of their peptide toxins similar to what is being done for the  
100 Terebridae (Holford *et al.* 2009).

101 This paper outlines a four–step methodology of integrative taxonomy to propose  
102 species hypotheses within hyperdiverse taxa (Fig. 1).

103 Step 1: Optimize taxon-coverage (Fig. 1, step 1). The sampling strategy for the  
104 Turridae included a large number of sampling events, covering a wide range of habitats and  
105 localities in order to increase the probability of sampling closely-related species and not  
106 overestimate the inter-specific differences (Hebert *et al.* 2004). In addition, multiplying the  
107 sampling events increases the probability of sampling several specimens for each species,  
108 even rare ones, providing a more accurate estimation of intra-specific variability (Eckert *et al.*  
109 2008; Lim *et al.* 2011).

110 Step 2: Construct Primary Species Hypotheses (PSHs). Sampled specimens are  
111 divided into PSHs based on the pattern of diversity of a single-gene, in this case the COI gene  
112 (Fig 1, Step 2). Several methods have been proposed for determining PSHs (Marshall 2006;  
113 Sites& Marshall 2003), but they make assumptions on the structure of the diversity within the  
114 sampling group. For example, the Population Aggregation Analysis (PAA) postulates that  
115 each population, defined a priori, includes only one species, which is not accurate when  
116 several morphologically similar species co-occur in sympatry (Kantor *et al.* 2008). In such  
117 cases, a phylogenetic approach, where species are more or less defined as terminal clades, is  
118 the solution commonly chosen (Fu& Zeng 2008; Puillandre *et al.* 2009). However, when the  
119 dataset is relatively large, exceeding several hundreds of specimens, it is difficult to  
120 objectively determine when a clade should be considered as a terminal leaf of a phylogenetic  
121 tree. Alternatively, two recently described bioinformatics tools, General Mixed Yule  
122 Coalescent (GMYC) (Monaghan *et al.* 2009; Pons *et al.* 2006) and Automatic Barcode Gap  
123 Discovery (ABGD) (Puillandre *et al.* 2011), define partitions of specimens using a well-  
124 defined criterion. GMYC uses a pre-existing phylogenetic tree to determine the transition  
125 signal from coalescent to speciation branching patterns. GMYC is generally considered an

126 effective method to detect species boundaries (Leliaert *et al.* 2009) even if it was argued that  
127 in some cases it could lead to an overestimation of the number of species (Lohse 2009).  
128 ABGD detects the breaks in the distribution of genetic pairwise distances, referred to as the  
129 “barcode gap”(Hebert *et al.* 2003), relying exclusively on genetic distance between DNA  
130 sequences. To construct reliable PSHs for the Turridae, a dataset of 1,000 COI sequences of  
131 Turridae were collected in the South-West Pacific and analyzed using both the GMYC and  
132 ABGD models.

133         Step 3: Visualization of PSHs using Klee diagrams. A recently developed method for  
134 processing genomic datasets, referred to as an “indicator vector” (Sirovich *et al.* 2009; 2010),  
135 produces an optimal classifier of a taxonomic group for biodiversity studies. This approach  
136 enables accurate quantitative display of affinities amongst taxa at various scales and extends  
137 to large genomic datasets. Indicator vectors are determined from each predefined set of  
138 nucleotide sequences (here the PSHs). The indicator vector for each PSH is used to build a  
139 structure matrix that accurately depicts affinities as correlations within and among-groups, or  
140 alternately as directly derivable distances. The structure matrix is presented as a color map,  
141 termed Klee diagram based on its resemblance to the works by the artist Paul Klee. Klee  
142 diagrams visualize the correlation patterns recovered for the PSHs, which are identified  
143 respectively from GMYC and ABGD (Fig. 1, Step 3).

144         Step 4: Consolidation of PSHs into secondary species hypotheses (SSHs). As stated  
145 before, delimiting species based on one gene is risky, and each PSH should be individually  
146 challenged using additional evidence. Additional criteria are used either to consolidate the  
147 PSHs when GMYC and ABGD are in agreement, or to choose the most likely option among  
148 alternate PSHs proposed by GMYC and ABGD (Fig. 1, Step 4). For the Turridae, SSHs were  
149 determined by analysis of additional gene sequences (rRNA 28S gene), geographic and  
150 bathymetric data, morphological characters, and using monophyly and gene flow criteria. In



151 the proposed hierarchy, the agreement of several independent genes is generally valuable  
152 evidence to support the existence of two (or more) independent evolutionary lineages  
153 recognized as species (Knowlton 2000). The definitive split of two lineages may be also  
154 supported by other sources of evidence, which includes intrinsic factors, such as the dispersal  
155 ability of individuals or their bathymetric preferences, and extrinsic factors, such as the  
156 geographic distribution of the habitats or the presence of geographic barriers. Figure 1 lists the  
157 different lines of evidence that can be used to delimit species.. Based on the morphological  
158 characters, proposed SSHs are then tentatively linked to the taxonomic names available in the  
159 literature. Using a sampling set of 1000 specimens, 87 Turridae SSHs are proposed based on a  
160 comparative analysis of bioinformatics species prediction tools GMYC and ABGD integrated  
161 with other available data. The strategy outlined in Figure 1 is specific for marine gastropods  
162 with internal fecundation, but could easily be adapted to other organisms with different life-  
163 history traits.

164

## 165 **MATERIAL AND METHODS**

### 166 *Sampling*

167 Specimens of Turridae were collected in different geographic regions: Taiwan  
168 (Taiwan 2004 expedition), Philippines (Panglao 2004 and 2005, Aurora 2007), Solomon  
169 Islands (Salomon 2, SalomonBOA 3), Vanuatu (BOA 1, Santo 2006), Chesterfield Islands  
170 (EBISCO) and New Caledonia (Norfolk 2 – Norfolk ridge) (Table S2). A fragment of the foot  
171 was clipped from anaesthetised specimens and preserved in 95% ethanol, while shells were  
172 kept intact for morphological analyses. The sampling strategy was designed to maximize the  
173 specific diversity within the set of collected specimens: (i) the prospected area is not  
174 comprehensive of the Turridae (they are present in other regions e.g. Africa, Central America)  
175 but corresponds to the centre of diversity of the Turridae (South-West Pacific, from

176 Philippines to Vanuatu), (ii) deep to shallow waters were explored (depth range 0-1762  
177 meters). All the specimens belonging to the family Turridae were analysed, without taking  
178 into account any kind of *a priori* species or population delimitation. This strategy would lead  
179 to potentially include several specimens for each species, but also to include potential cryptic  
180 species. One thousand specimens were analysed, and for each of them data corresponding to  
181 their sampling site (geographic coordinates, depth of collection) were databased (Barcode Of  
182 Life Database project “Conoidea barcodes and taxonomy”). All specimens and DNA extracts  
183 are stored in the Museum National d’Histoire Naturelle collection.

184

#### 185 *Sequencing*

186 DNA was extracted from a piece of foot, using a 6100 Nucleic Acid Prepstation  
187 system (Applied Biosystem). Two gene fragments were amplified: (i) a fragment of 658 bp of  
188 Cytochrome Oxidase I (COI) mitochondrial gene using universal primers LCO1490 and  
189 HCO2198 (Folmer *et al.* 1994) and (ii) a fragment of 900 bp of the rRNA 28S gene, involving  
190 D1, D2 and D3 domains, using the primers C1 and D3 (Jovelin& Justine 2001). For the COI  
191 gene, the primer LCO1490 was also used in combination with newly designed primers  
192 (COIH615: CGAAATYTNAATACNGCYTTTTTTTGA and COIHNP:  
193 GGTGACCAAAAAATCAAAAYARATG) when PCR were negative with HCO2198. All  
194 PCR reactions were performed in 25 µl, containing 3 ng of DNA, 1X reaction buffer, 2.5 mM  
195 MgCl<sub>2</sub>, 0.26 mM dNTP, 0.3 µM of each primer, 5% DMSO and 1.5 units of Q-Bio Taq  
196 (MPBiomedicals) for all genes. COI gene amplifications are performed according to Hebert et  
197 al. (2003); for 28S gene, the protocol consists of an initial denaturation step at 94°C for 4',  
198 followed by 30 cycles of denaturation at 94°C for 30'', annealing at 52°C and extension at  
199 72°C for 1'. The final extension was at 72°C for 10'. PCR products were purified and  
200 sequenced at Genoscope facilities. In all cases, both directions were sequenced using the

201 Sanger method to confirm accuracy of each haplotype sequence. All sequences were  
202 submitted to GenBank.

203

#### 204 *Phylogenetic Analyses of DNA Sequences*

205 The DNA sequences were manually (for the COI gene) or automatically (for the 28S  
206 gene) aligned using Clustal W as implemented in BioEdit version 7.0.5.3 (Hall 1999). Genetic  
207 distances were calculated between each pair of COI sequences. In order to evaluate the effect  
208 of multiple nucleotide substitutions on the distance between DNA sequences, three genetic  
209 distances are compared: (i) the uncorrected p-distance, (ii) the K2P distance, a model that  
210 corrects for multiple substitutions with different Ts (transitions) and Tv (transversions) rates,  
211 frequently used in DNA barcode analyses and (iii) the Tamura-Nei model (TN+I+G, with I =  
212 0.541 and G = 1.014), identified as the best-fitting distance (*i.e.* that corrects optimally for  
213 multiple substitutions) by Modelgenerator (Keane *et al.* 2006), following the hLRT criterion.  
214 The GTR+I+G (I = 0.817, G = 0.651) model was identified as the best-fitting model for the  
215 28S gene dataset. The Maximum Likelihood approach was conducted by determining the best  
216 tree over 20 independent runs using RAxML 7.2.3 (Stamatakis 2006). The GTRGAMMAI  
217 model was used for both genes. Robustness of the nodes was assessed with 100 bootstrap  
218 replicates (with five searches for each of them). Bayesian analyses were performed with  
219 BEAST 1.4.8 (Drummond& Rambaut 2007), using the best-fitting models identified with  
220 Modelgenerator. A relaxed lognormal clock with a coalescent prior, determined as the best  
221 fitting parameters to be used with the GMYC model (Monaghan *et al.* 2009), was used to  
222 generate the COI Bayesian gene trees that were used in conjunction with the GMYC model to  
223 delimit species. MCMC chains were run for 100M generations after which all ESS values  
224 calculated with Tracer 1.4.1 (Rambaut& Drummond 2007) were >200 (default burnin). Tree  
225 annotator 1.4.7 (<http://beast.bio.ed.ac.uk>) was used to analyse the MCMC outputs, using the

226 default parameters. COI Bayesian analyses were performed on all the obtained sequences; the  
227 other analyses were performed on haplotypes only to reduce computation time.

228

#### 229 *Automatic Barcode Gap Discovery (ABGD)*

230 Following the similarity criterion, genetic distances between specimens from the same  
231 species are supposed to be lower than genetic distances between specimens from different  
232 species, revealing a non-continuous distribution (Hebert *et al.* 2003). This barcode gap, *i.e.*  
233 the range of genetic distances not represented in the matrix of pairwise comparisons, can be  
234 used as a threshold offering primary species delimitation under the assumption that  
235 individuals within species are more similar than between species (genotyping clustering  
236 criterion – Mallet 1995). However, in some cases, this barcode gap does not correspond to a  
237 real discontinuity in the distribution, but only to a decrease of the distance frequency between  
238 the two modes of the distribution, *i.e.* the intra and interspecific distances overlap (Meier *et al.*  
239 2008). This can be due to incomplete lineage sorting, where the COI sequence of a specimen  
240 is more similar to a sequence of another species than to a sequence of the same species  
241 (Rosenberg & Tao 2008) or to an underestimation of genetic distances because of homoplasy.  
242 The Automatic Barcode Gap Discovery (ABGD) method aims at identifying a limit between  
243 the two distributions, even when they are overlapping. Starting from several *a priori*  
244 thresholds of genetic distances chosen by the user, ABGD will first compute the theoretical  
245 maximal limit of the intraspecific diversity (using a coalescent model) and then identify in the  
246 whole distribution of pairwise distances which gap, by definition superior to the maximal  
247 limit of the intraspecific diversity, potentially corresponds to the so-called “Barcoding gap”,  
248 *i.e.* a potential limit between intra and interspecific diversity. Inference of the limit and gap  
249 detection are then recursively applied to previously obtained groups to get finer partitions  
250 until there is no further partitioning. This method is described in detail in Puillandre *et al.*

251 (2011); we used the online version to analyse the dataset  
252 (<http://wwwabi.snv.jussieu.fr/public/abgd/>). MEGA was used to build the distance matrix  
253 using a TN model (with  $\alpha = 1.014$ ). ABGD default parameters were used, except the  
254 relative gap width (X) was set to 10 to avoid the capture of smaller local gaps.

255

### 256 *General Mixed Yule Coalescent Model (GMYC)*

257 The GMYC method, described in Pons *et al.* (2006) and Monaghan *et al.* (2009), is  
258 based on the difference in branching rates between speciation branching events (interspecific  
259 relationships) and coalescence branching events (intraspecific relationships) in a phylogenetic  
260 tree. This difference can be visualized as a switch between slow and fast rates of branching  
261 events in a lineage-through-time plot. The first step of the method is to compare the likelihood  
262 of the phylogenetic tree obtained with BEAST assuming a single branching process *versus* the  
263 likelihood of the same tree assuming a switch of branching rates between the two types of  
264 events. If such a switch is detected, its position is determined and placed in the tree, allowing  
265 the delineation of PSHs. Two versions of the method are applied here: in the single-threshold  
266 method (Pons *et al.* 2006), the switch between speciation and coalescence events is supposed  
267 to be unique; in the multiple-threshold method (Monaghan *et al.* 2009), each PSH defined  
268 with the single-threshold method is re-analysed one by one and can be divided in two, or  
269 fused with its sister-PSH, the hypotheses with the best likelihood being chosen. GMYC  
270 (multiple-threshold) proposes alternate hypotheses of species delimitations, and in this way is  
271 also similar to the ABGD method. The GMYC method with both the single and multiple-  
272 threshold models (Monaghan *et al.* 2009), implemented in the SPLITS package for R, was  
273 applied to the COI tree obtained with BEAST.

274

### 275 *Klee Diagrams*

276           Sirovich *et al.* (2009; 2010) provides a framework for translating nucleotide symbol  
277 sequences into numerical vectors; in a manner that links Euclidean vector distances to the  
278 customary symbol substitution (Hamming) distance. This leads to calculation of the angle,  $\theta$ ,  
279 between pairs of vectorized sequences and from this yields their correlation  $\cos \theta$ . Under  
280 proper normalization the corresponding Hamming distance is given by  $1 - \cos \theta$ . For  
281 collections of genomically defined taxa this formalism leads to the determination of a  
282 classifier for each taxa, called its indicator vector. The indicator vector of a taxa is obtained  
283 under the condition that it is maximally correlated with the taxa, and simultaneously that it is  
284 minimally correlated with all other taxa. The matrix of inter-taxa correlations (the structure  
285 matrix in physics), in image form the Klee diagram, is intrinsic to the data and independent of  
286 evolutionary models. It distinguishes differences among species with high information density  
287 and faithfully displays quantitative taxa relations.

288           As mentioned above the usual taxonomic distance matrix is reciprocally related to the  
289 Klee diagram and so can generate a taxonomic tree. However unlike trees, which lose  
290 distance accuracy with size the Klee diagram faithfully retains its accuracy at all scales. A  
291 Klee diagram may show some variation in appearance if sequence variance plays a role.  
292 Experience dictates that this is not a factor, and on the contrary variance is usually slight  
293 enough so that taxa averages can reasonably replace taxa ensembles in the calculations.

294           In the present study, Klee diagrams are used to compare and evaluate the results  
295 obtained by the two different species delimitation approaches used (ABGD and GMYC). One  
296 COI sequence in each of the PSH defined in two alternate PSHs partitions (the most inclusive  
297 and the less inclusive ones) were analyzed using the indicator vector approach and used to  
298 build matrices. Prediction tests were then performed to assign the sequences not used to build  
299 the matrices to PSHs. Areas of congruence are shown as blue, areas in conflict are shown in  
300 gradations of red and yellow.

301

302 *Analyses of Other Characters and Criteria*

303 Phylogenetic analyses

304 As the efficacy of ABGD and GMYC may be limited by the variation of evolutionary rates in  
305 the different species, the statistical support (bootstraps and posterior probabilities) calculated  
306 using RAxML and BEAST for each PSH recognized with the COI gene were reported.

307 Conflicts between the COI and 28S genes were also analysed by identifying which PSHs were  
308 sharing common 28S haplotypes and which ones were not monophyletic.

309 Genetic structure

310 When a PSH was present in at least two geographic populations, each of them including at  
311 least six specimens, the genetic structure was assessed among the different populations using  
312 Arlequin 3.1 (Excoffier *et al.* 2005). If a single PSH was present in several different  
313 geographic regions (among Taiwan, Philippines, Solomon Islands, Vanuatu, Chesterfield  
314 Islands and New-Caledonia – see Material and Methods, sampling) and in different localities  
315 within the geographic region, an AMOVA (with a 3,000 permutations tests) was performed. If  
316 only one hierarchical level was involved (different localities within a single geographic  
317 region),  $F_{st}$  between each pairs of populations was calculated. Network 4.5 (Median-Joining  
318 option) was used to construct haplotype networks.

319 Bathymetric distribution

320 Stations are characterized by starting and ending points that may correspond to different  
321 depths. This variation is sometimes up to 500m, and such stations may actually cover highly  
322 different environments. To minimize the effect of this imprecision, the depth data for stations  
323 with a significant discrepancy between the starting and ending points (>20m for shallow  
324 waters stations and >50m for deep water stations) were not considered. To reduce the bias in  
325 depth ranges, all the PSHs with only one specimen were also not considered in the estimation

326 of the bathymetrical distribution. It was then possible to conclude from the observation of the  
327 bathymetrical ranges of two PSHs if they were overlapping or not. To test the hypothesis that  
328 bathymetrical ranges could be underestimated by subsampling, a statistical test was designed  
329 to evaluate if the bathymetrical range of a given PSH could be obtained by subsampling.  
330 Details of the test and interpretations of the results are provided in the Figure S3.

### 331 Morphological analyses

332 The features of the shells of all analysed specimens were examined by several specialists of  
333 the Turridae, Yuri Kantor, Baldomero Olivera and Alexander Sysoev. Examinations of the  
334 shell were not performed “blindly” but taking into account the molecular taxonomy analyses.  
335 The goal was thus to determine if it was possible or not to find morphological differences  
336 between the PSHs, using shell characters as traditionally used in malacology. Considering  
337 available description in the malacological literature, each PSH was tentatively attributed *a*  
338 *posteriori* to available species name. When no name was available PSHs were numbered with  
339 the genus to which they were attributed.

### 340 Dispersion abilities

341 For such benthic organisms, dispersal abilities occur mainly during the larval stage.  
342 Furthermore, the accretionary growth of the protoconch (*i.e.* the shell formed by the embryo  
343 and/or the veliger larvae before metamorphosis) can be used to infer the mode of  
344 development, which constitutes the best proxy for the dispersal ability of a gastropod species  
345 when no other data are available (Jablonski& Lutz 1980). A multispiral protoconch suggests  
346 that the larva fed in the water column (*i.e.* planctotrophic species), and is thus able to disperse  
347 over large distances. Conversely, the dispersion abilities are supposedly reduced for a non  
348 planctotrophic species (*i.e.* with a paucispiral protoconch), even if some non planctotrophic  
349 species have been shown to disperse over wide distances, e.g. through passive larval transport  
350 (Parker& Tunnicliffe 1994). Dispersion abilities inferred from the protoconch morphology



351 were used to discuss the validity of the PSHs. When not broken, the protoconch of the  
352 analysed specimens was in most cases multispiral (~3 whorls or more), indicating important  
353 dispersal capacities. However, the PSHs identified as *Lophiotoma indica* (Table 1) were  
354 possessing reduced protoconchs with only 2 whorls.

355

### 356 *Turning PSHs into SSHs*

357 PSHs were considered and eventually turned in SSH following the workflow  
358 described in the Figure 1 (step 4). Mainly three types of data were analysed: (i) the  
359 presence/absence of shared haplotypes between PSHs and their reciprocal monophyly, (ii)  
360 geographical and bathymetrical distribution, considered in association with the dispersal  
361 abilities, and (iii) morphological variability. In cases where the various lines of evidence used  
362 to turn PSHs in SSHs are not conclusive, a conservative approach was followed to avoid an  
363 over-estimation of the species diversity and the creation of new species names that would be  
364 later synonymised. Each PSH can be considered as a single SSH, but the possibility that each  
365 of these SSHs includes several species cannot be ruled out.

366

## 367 **RESULTS**

### 368 *Turridae COI Gene Variability*

369 A set of 1000 specimens of Turridae was sequenced for a 658 bp fragment of the  
370 barcoding COI gene; 648 haplotypes were found, with 477 polymorphic sites and a high  
371 haplotypic diversity (0.995). Genetic pairwise distances for COI gene were computed using  
372 three different substitution models: (a) the p-distances, (b) the K2P distances, and (c) the  
373 Tamura-Nei (TN) distances. The distribution of genetic distances, whatever the substitution  
374 model used, displayed two modes separated by a rough gap (“barcode gap”) between 0.02 and  
375 0.04 (Fig. 2a-b). However, as shown in the Figure 2b, the number of pairwise comparisons for

376 genetic distances that corresponded to the barcode gap was lower for the TN distance than for  
377 the K2P distance and the p-distance. Consequently, the TN distances were used thereafter for  
378 Automatic Barcode Gap Discovery (ABGD) analyses.

379

#### 380 *Turridae Primary Species Hypotheses (PSHs)*

381 Species delimitation tools, Automatic Barcode Gap Discovery (ABGD) and General  
382 Mixed Yule Coalescence model (GMYC) were used to construct PSHs for the Turridae  
383 specimens sequenced with the COI gene. ABGD uses several *a priori* thresholds to propose  
384 partitions of specimens into PSHs based on the distribution of pairwise genetic distances. The  
385 numbers of PSHs defined with the ABGD method vary with the different *a priori* thresholds  
386 (Fig. 2c). Extreme threshold values lead to partitions where almost each haplotype is  
387 considered as a different PSH, or conversely where all haplotypes are placed in a single PSH.  
388 The other intermediate *a priori* thresholds lead to similar partitions with 87, 89 or 91 PSHs  
389 (the 87 and 91 PSH partitions are detailed in the Table 1).

390 Two versions of the GMYC method are applied: the single-threshold method (Pons *et*  
391 *al.* 2006) and the multiple-threshold method (Monaghan *et al.* 2009). For both versions of the  
392 method, the likelihood of the GMYC model ( $L_{\text{GMYCsingle}} = 10855.84$  and  $L_{\text{GMYCmultiple}} =$   
393  $10860.46$ ) was significantly superior to the likelihood of the null model ( $L_0 = 10770.74$ , p-  
394 value = 0). However, the partitions obtained are not identical: 95 PSHs were obtained for the  
395 single-threshold method (confidence limits: 86-107), and 102 with the multiple-threshold  
396 (confidence limits: 101-115). The likelihood of the two methods are not significantly different  
397 (p-value = 0.95).

398 Overall, the partitions obtained with the ABGD and GMYC are congruent. Among the  
399 103 PSHs listed in Table 1, 73 were obtained both with ABGD and the two GMYC methods

400 (Table 1, columns 2-5). In the phylogenetic tree of the Figure 3a each of the PSHs listed in  
401 Table 1 is represented by a single branch.

402

### 403 *Visualization of the PSHs Using Klee Diagrams*

404 The indicator vector method (Sirovich *et al.* 2009; 2010) was used to generate Klee  
405 diagrams for the 87 PSHs of the more inclusive partition (i.e. the partition with the lowest  
406 number of PSHs defined using ABGD method), and for the 103 PSHs of the less inclusive  
407 partition (i.e. the partition with the highest number of PSHs defined using the multiple-  
408 threshold GMYC method) (Fig. 3b-c). In the latter Klee diagram (Fig. 3c), a higher  
409 correlation is evident between pairs of PSHs that were considered as a single PSH by ABGD:  
410 PSHs 21+22, 25+26, 28+29, 32+33+34+35, 51+52, 55+56, 60+61, 68+69, 72+73, 82+83,  
411 84+85, 89+90, 91+92, 93+94 (Fig. 3b-c black arrows). Predictions tests performed using the  
412 vectors obtained for the 103 PSHs indicate that two PSH pairs (28+29 and 84+85) were  
413 recognized as belonging to the same species. In these two cases, the indicator vector analysis  
414 results provide support for the ABGD result rather than the multiple-threshold GMYC  
415 hypothesis. All other indicator vector analyses of ABGD and GMYC PSHs appear to be  
416 equally likely.

417

### 418 *Phylogenetic Analyses and 28S Gene*

419 Most of the PSHs defined with the COI gene (86 out of the 103 listed in Table 1,  
420 representing 708 specimens) were successfully sequenced for the 28S gene (Table 1). A 28S  
421 fragment of 908 bp after alignment displayed 228 haplotypes with 359 polymorphic sites and  
422 a haplotypic diversity of 0.979. Bootstraps and posterior probabilities are given for each PSH  
423 and each gene (COI and 28S) in Table 1. All the PSHs that included more than one specimen  
424 corresponded to highly supported clades with the COI gene (Bootstraps > 75, PP > 0.95),

425 except in 10 cases (PSHs 28, 29, 32+33+34+35, 52, 61, 68, 84, 89+90, 91, 93+94). Each of  
426 these ten cases corresponded to a pair of PSHs that were alternatively recognized as a single  
427 PSH or two different PSHs with ABGD and GMYC. In the less inclusive hypothesis, when  
428 one of the two PSHs corresponded to a weakly supported clade (*e.g.* PSHs 28 and 29), the  
429 alternate most inclusive hypothesis systematically corresponded to a highly supported clade  
430 (PSHs 28+29) (Table 1).

431         Of the 86 PSHs sequenced for the 28S gene, 61 were characterized by unique (*i.e.*  
432 diagnostic) 28S haplotypes; among them, 26 corresponded to monophyletic groups (15 with  
433 high statistical support) and 11 were non-monophyletic. The 25 other PSHs sequenced for the  
434 28S gene shared one or several 28S haplotypes with at least one other PSH. Among them, 12  
435 corresponded to pairs of PSHs that were recognized as a single PSH by either ABGD or  
436 GMYC (Table 1 and Fig. 4b) and 12 others corresponded to closely related PSHs with the  
437 COI gene, even if they were never recognized as a single PSH. In one case, PSH 47 + PSHs  
438 60-61, 28S haplotypes were shared between distant PSHs in the COI tree, and may  
439 correspond to different evolutionary histories for the two genes.

440

#### 441 *Geographic Distribution and Genetic Structure*

442         Among the 103 PSHs, 80 PSHs were restricted to a single geographic region (Taiwan,  
443 Philippines, Solomon Islands, Vanuatu, Chesterfield Islands or New Caledonia), 17 in 2  
444 different regions, and 6 in 3 or more. Among the 14 pairs or quadruplets of PSHs either  
445 recognized as a single PSH or as 2 or 4 different PSHs depending on the method, 6 of them  
446 were collected in different geographic regions and were thus considered allopatric and 8 were  
447 collected in at least one common area (at the same station for 7 of them), and are reported as  
448 sympatric (Table 1, geographic distribution column). The genetic structure among different  
449 sampling sites in a single PSH was calculated for eight different PSHs with the COI gene, and

450 for five with the 28S gene (Table 1, genetic structure column). All the  $F_{st}$  values are very low  
451 and only one is significant (Table S1).

452

### 453 *Bathymetric Distribution*

454       Among the 14 pairs or quadruplets of alternative PSHs, 9 included at least one PSH  
455 with only one specimen and were not analysed further. Another pair included PSHs with  
456 strictly non-overlapping bathymetric ranges (PSH 60-61), two pairs corresponded to  
457 subsamples of the association of two PSHs (28-29 and 84-85) and pair one included one PSH  
458 with bathymetric preferences (25-26). Finally, the quadruplet included two PSHs with  
459 bathymetric preferences (33-35) and two considered as a subsample of the association of the  
460 four PSHs (32, 34).

461

### 462 *Shell Morphology and Attribution to Species Names*

463       The shells of the specimens included in each PSH were examined. Based on the shell  
464 morphology, the PSHs were then tentatively assigned to a species-name available in the  
465 literature (Table 1, morphological ID). For 28 PSHs, shells of specimens corresponded to a  
466 unique morph and it was possible to link each of them to a unique species name; conversely,  
467 11 species names corresponded to shell features shared by several PSHs (39 PSHs affected).  
468 Two PSHs represented by a single juvenile specimen might not be attributed to a  
469 morphospecies attached to a species name. For 11 PSHs, shells corresponded to distinct  
470 morphospecies for which no species names were available and they were thus associated to a  
471 genus name and to a morphospecies number within each genus (*Gemmula* 3, 4, 8, 9, 11-14, 16  
472 and *Ptychosyrinx* 1-2). Finally, the 23 remaining PSHs corresponded to three different  
473 morphospecies, not attributed to a species-name: *Gemmula* 1 (PSHs 1, 16, 48, 60, 61, 80, 84,

474 85), 2 (PSHs 2, 3), 5 (PSHs 49, 87, 94, 95), 6 (PSHs 50-52), 7 (PSHs 62, 63), 10 (PSHs 82,  
475 83), 15 (PSHs 90, 91).

476

#### 477 *Consolidating Secondary Species Hypotheses (SSHs)*

478 Primary species hypotheses (PSHs) drawn using ABGD and GMYC were converted to  
479 SSHs according to the workflow presented in step 4 of Figure 1 and the criteria listed in Table  
480 1. Among the 103 PSHs listed in Table 1, 21, found monophyletic with the 28S gene, and 38,  
481 with unique 28S haplotypes, were converted to 59 SSHs. Of the 38 PSHs with unique 28S  
482 haplotypes, 24 were represented by specimens with identical sequences (i.e. a single  
483 haplotype was included in the phylogenetic analysis), preventing any test of the 28S  
484 monophyly for the corresponding PSH. Twenty PSHs were not sequenced for the 28S gene.  
485 Of this group 10 PSHs were converted to SSHs after analysis of other evidence (see table 1  
486 for details). Following a conservative approach, the remaining 11 PSHs without 28S sequences  
487 were converted to 5 SSHs, as there was no comparative evidence to support additional SSH  
488 assignments. Finally, 24 PSHs sharing 28S haplotypes were converted to 13 different SSHs  
489 following guidelines in step 4 of Figure 1. An example for which all the PSHs characters and  
490 criteria are congruent is shown in Figure 5.

491 There are only four cases where the PSHs were in agreement with ABGD and GMYC  
492 analyses, but were not directly converted to SSHs. For example, the PSHs 65 and 66 were  
493 considered to correspond to a single SSH, as they shared 28S haplotypes and they were not  
494 distinguished morphologically. Similarly, 28S variability, geographical and bathymetrical  
495 ranges, dispersal abilities and morphological analysis were decisive in discussing the 14 pairs,  
496 or quadruplets, of PSHs alternatively recognized either as a single PSH, or 2 to 4 different  
497 PSHs by the ABGD and GMYC analyses. They were turned into 21 SSHs (see details in  
498 Table 1). Three examples (one species, two species, or inconclusive species), corresponding

499 to three different conclusions that can be obtained following step 4 of Figure 1, are detailed  
500 here. (1) One species: PSHs 25 and 26 shared 28S haplotypes, were both found in the same  
501 geographic area, and in the same station for some of them, one of them (PSH 25) displayed  
502 bathymetric preferences and their larvae were weakly dispersive. PSHs 25 and 26 were  
503 interpreted as a single SSH (moreover including also the PSH 24), and the differences found  
504 in the COI were thought to correspond to intraspecific structure linked to the depth. Four  
505 other PSHs pairs (28-29, 55-56, 72-73 and 84-85) were similarly turned each in a single SSH.  
506 (2) Two species: PSH 21 and 22 were interpreted as two different species as they did not  
507 share 28S haplotypes, were found in two different geographic regions (Vanuatu and  
508 Solomons) without obvious barrier between them, and their larvae are highly dispersive.  
509 Additionally, bathymetric ranges for PSHs 21 and 22 did not overlap, which can be seen as  
510 ecological differences between the two species. Two other PSHs pairs (32-35 and 68-69) were  
511 similarly converted to two or four SSHs. (3) Inconclusive species: Following a conservative  
512 approach, PSHs 51-52 was considered as a single SSH as the supporting evidence was  
513 inconclusive. The 28S gene was not sequenced for these specimens, and they were found in  
514 the same geographic region, without bathymetric differences (see Material and Methods).  
515 Five other PSHs pairs (60-61, 82-83, 89-90, 91-92 and 93-94) were similarly converted to a  
516 single SSH following a conservative inclusive approach.

517

## 518 **DISCUSSION**

519 Illustrated here is a semi-automated integrative taxonomy strategy that uses a single-  
520 gene approach derived from “DNA-barcoding” to determine species as hypotheses that are  
521 consolidated using several additional lines of evidences through a process of modification and  
522 validation. The single-gene dataset analysed with bioinformatics species delimitation tools,  
523 such as ABGD and GYMC, is combined with biological (life-history traits), morphological

524 (shell characters) and ecological (bathymetric distribution, geographic barriers) data. This  
525 approach constitutes an efficient way for proposing primary species hypotheses (PSHs) for  
526 hyperdiverse groups especially when morphological characters are known to be problematic.

527         Using a predominantly shell-based morphological approach, over the last 30 years,  
528 only 13 new species names were proposed for the Turridae genus *Gemmula*. The integrative  
529 taxonomy approach described here (Fig. 1) identified 27 secondary species hypotheses  
530 (SSHs) within *Gemmula* that are not linked to available names, suggesting that 27 novel  
531 species names are needed to encompass the species diversity within this genus. Overall, the  
532 non-monophyletic genera *Lophiotoma* and *Gemmula* (Heralde *et al.* 2010) include 137  
533 species worldwide, around 100 of which are considered valid (Tucker 2004). In comparison,  
534 our analysis recognized 70 SSHs within these genera in the South-West Pacific alone,  
535 suggesting that the diversity of species in *Lophiotoma* and *Gemmula* has been underestimated.  
536 Moreover, in several cases morphologically very similar SSHs were found in a single  
537 population. These results confirm that taxonomic approaches based primarily on shell  
538 characters or even on *a priori* definitions of populations (Sites & Marshall 2003) may  
539 underestimate species diversity. However, it should be noted that in several cases the  
540 proposed SSHs are morphologically non-cryptic, as clear diagnostic shell-characters were  
541 identified. Whether these non-cryptic SSHs would have been detected using a traditional  
542 morphology-based species-delimitation approach is difficult to test. In the integrated  
543 methodology applied, the morphological analyses were not performed *a priori* but were based  
544 on finding any morphological differences between the molecularly-defined PSHs. It is then  
545 reasonable to think that some of the non-cryptic SSHs would have been detected by  
546 morphologists had the samples been previously described.

547         Key components of the integrative strategy presented here are a sampling design that  
548 covers both taxon and intraspecific diversity and visualization of the primary species



549 hypotheses proposed with ABGD and GMYC using the recently developed method of Klee  
550 diagram. For large datasets such as the 1,000 specimens used for the Turridae, Klee diagrams  
551 facilitate the visualization of PSH correlation patterns and rapidly identify borderline cases.

552 Overall, GMYC and ABGD recovered similar partitions within the Turridae dataset,  
553 as many PSHs are identical between the two methods (Table 1). Both ABGD and GMYC  
554 achieved their primary goal of proposing PSHs based on a criterion that is biologically  
555 justified, either empirically or theoretically. Most PSHs were similar among ABGD and  
556 GMYC methods, and other lines of evidences corroborated these primary hypotheses.  
557 Therefore, although ABGD and GMYC methods are not sufficient on their own to propose  
558 robust species hypotheses, they provided a primary partition that was close to the partition  
559 that was finally retained.

560 Conflicting cases were detected when one PSH defined by ABGD or GMYC is split in  
561 two by the alternate method (Fig. 3). Indicator vector analysis suggests in cases where conflict  
562 is detected that neither ABGD nor GMYC can be consistently preferred (Fig. 3b-c, black  
563 arrows). In conflicting cases ABGD and GMYC do not propose a unique threshold, but rather  
564 a range of possible partitions among which some PSHs are different. ABGD, by testing  
565 several a priori thresholds and by applying a recursive approach, and GMYC, with both the  
566 single and multiple threshold methods, are able to consider the heterogeneity among lineages  
567 of the rates of speciation and of coalescence that result in an overlapping distribution of the  
568 pairwise genetic distances (Fig. 2a and b). For eight out of nine pairs of conflicting  
569 hypotheses obtained with the single and multiple GMYC methods, the corroboration process  
570 turned into SSHs the PSHs proposed by the single-threshold method. For the 14 conflicting  
571 cases between ABGD and GMYC methods, the final SSH were defined as a PSH only by  
572 ABGD in 2 cases and only by GMYC in 1 case (Table 1). These findings suggest ABGD and  
573 GMYC are complementary and should be used together to increase the overall robustness of

574 the final partition to determine the set of PSHs fixed as SSHs. Compared to GMYC, ABGD  
575 may be considered as less refined in regards to underlying evolutionary processes, however,  
576 GMYC requires prior construction of a tree that must be ultrametric, which does not  
577 necessarily reflect the real divergence between species. Alternatively, ABGD is based solely  
578 on genetic distances calculated between each pair of COI sequences, allowing for the  
579 exploration of a range of thresholds and management of the heterogeneity of evolution rates.  
580 Furthermore, the short calculation time of the ABGD method, a few seconds vs several weeks  
581 to obtain a BEAST tree for the GMYC method (following the method described in Monaghan  
582 *et al.* 2009), allows a rapid comparison of different models of evolution for each dataset.  
583 Thus, ABGD would be easier to apply to very large datasets. Finally, both ABGD and GMYC  
584 are problematic when species are represented with only a few specimens (Lohse 2009;  
585 Puillandre *et al.* 2011), and as underlined in the results, PSHs with less than three specimens  
586 are generally difficult to discuss with other characters and criteria. A large proportion of rare  
587 species, represented by a low number of specimens, in a dataset is a common pattern,  
588 especially for marine gastropods (Castelin *et al.* 2011). SSH proposed for these samples are  
589 more susceptible to modification if new specimens are collected in the future.

590 In addition to analyses via ABGD and GMYC and to corroborate the hypotheses  
591 drawn from single gene with criteria, we examined patterns of diversity in the same sample  
592 set to give an evolutionary meaning to the proposed hypotheses. For example, testing the  
593 reciprocal monophyly of the PSHs on several genes indicate that the proposed species  
594 represent a unique evolutionary lineage. The SSHs proposed here all correspond to PSHs for  
595 which several characters and/or lines of evidence were congruent. However, not all SSHs are  
596 equally supported. For example, SSHs based on the lack of shared haplotypes on the 28S gene  
597 should be considered more carefully than SSHs confirmed by reciprocal monophyly with the  
598 nuclear gene. Bathymetric and geographic distributions, in association with dispersal abilities,

599 are not species delimitation criteria in themselves, but act as additional evidence toward one  
600 or another hypothesis further providing clues about the speciation process (Hyde *et al.* 2008).  
601 In figure 1, several patterns are interpreted as evidence for one or two species. For example,  
602 highly dispersive larvae for two PSHs in absence of geographical barrier, or in presence of  
603 bathymetrical differences when geographical ranges are overlapping, can be interpreted as  
604 evidence for the presence of two species, as large dispersal abilities would result in shared  
605 haplotypes if only one species was involved. However, in several cases, results are  
606 inconclusive. For example, when the two PSHs have non-dispersive larvae and discontinuous  
607 geographic ranges without barrier, the differences observed with the COI could be only due to  
608 geographic structuring. Here, the decision depends on the taxonomist's choice and we  
609 followed a conservative approach by considering only one species, even if two deep  
610 conspecific lineages (Padiál *et al.* 2010), potentially corresponding to incipient species, were  
611 revealed by the COI gene analysis. Finally, morphology, although a highly valuable character  
612 in numerous cases (Holynski 2010), is used here in the final step, where morphological  
613 differences are seen as additional evidence for the existence of different species, but knowing  
614 also that different species may share a highly similar morphology.

615         The borderline cases detected in the primary stage of the integrative protocol  
616 corresponded mostly to cases of recent divergence. These cases are of particular interest for  
617 understanding speciation processes. Using characters and criteria that are directly issued from  
618 evolutionary-based species criteria (phylogeny, reproductive isolation, phenetic divergence  
619 (Samadi & Barberousse 2006)) is not only useful to propose more robust hypotheses, but also  
620 to understand what induced and drove the speciation process (Padiál *et al.* 2010). For  
621 example, the distributions of most of the species illustrated in Figure 5 are restricted to one or  
622 two geographic regions, suggesting allopatric speciation. However, *Ptychosyrinx* sp. 2 and  
623 *Gemmula* sp. 4 are also distinguished by their bathymetric ranges, and could have diverged

624 under a parapatric model. Identifying the different factors promoting the speciation event,  
625 either linked to geographical isolation and genetic drift or linked to ecological differentiation  
626 and selective forces, requires model-based studies (Crow *et al.* 2010). However, in all cases,  
627 delimitating species with clear, robust and reproducible methods remains the first step.

628 PSHs proposed by ABGD and GMYC of species represented by large numbers of  
629 specimens or, conversely, by only a few specimens, i.e. uneven sampling, is an area of  
630 concern with the integrated method described. To test the effects of uneven sampling, the  
631 number of specimens in the two largest PSHs, PSH 36 with 94 specimens and PSH 74 with  
632 101 specimens, were reduced to 10. The PSHs defined with ABGD were unchanged as a  
633 result of the artificial minimization. Additionally, two datasets with three species each were  
634 simulated, where in the first dataset, each species was represented by 37 specimens and in the  
635 second, they were represented by 1, 10 and 100 specimens respectively. The Yule model was  
636 used to simulate the species tree, and a Kingman model in which genes from different species  
637 cannot coalesce was used for the gene tree (until they reach the common ancestral species). A  
638 theta of 10 was used for the mutations, with a sequence length of 1000 to obtain an average of  
639 1% divergence between 2 sequences of the same species. Under these conditions, ABGD  
640 (with a prior of 0.01) detected a mean of 3.67 species when the sampling is even, and 3.46  
641 species when the sampling is uneven among the 1,000 runs. The difference was subtle, but  
642 significant, with a p-value  $< 10^{-4}$ . Similar simulations performed with GMYC suggested that  
643 GMYC may overestimate the number of species in even and uneven sampling, and further  
644 detailed exploration of the effect of uneven sampling on species delimitation with both ABGD  
645 and GMYC are clearly needed. The simulation results suggest that the Turridae PSH  
646 delimitation could be slightly influenced by the evenness of the sampling and could explain  
647 why in several cases ABGD and GMYC underestimate or overestimate respectively the  
648 number of PSHs compared to the number of SSHs retained at the end of the analytical process

649 (Table 1). Increasing the sampling effort to reduce differences in specimen 649 numbers  
650 between PSH would reduce the potential biases witnessed in ABGD and GMYC. However,  
651 this recommendation is often hardly applicable as rare species are usually present in empirical  
652 studies.

653 As demonstrated for the Turridae, the integrative taxonomy strategy described here is  
654 compulsory for primary and secondary species delimitation hypotheses in hyperdiverse  
655 groups, and could be easily adjusted to any biodiverse group of organisms. In addition, the  
656 relative congruence between PSHs defined with ABGD and the final SSHs retained indicates  
657 that ABGD can be used as a proxy for species delimitation when only molecular data are  
658 available. ABGD can be applied in biodiversity analysis to quickly assess the biodiversity of  
659 an environmental sample and to facilitate comparative analysis in DNA metabarcoding.

660

661 **ACKNOWLEDGMENTS**

662 Key material for molecular studies originated from several expeditions to the  
663 Philippines and Vanuatu, funded via a consortium of agencies, including the Total  
664 Foundation, the French Ministry of Foreign Affairs, the Richard Lounsbery Foundation, the  
665 Philippines Bureau of Fisheries and Aquatic Research (BFAR) and the Niarchos Foundation;  
666 the Coral Sea and Solomon Islands took place on board R/V *Alis* from the Institut de  
667 Recherche pour le Développement (IRD). P. Bouchet and B Richer de Forges were P.I. for  
668 these cruises/expeditions. E. Strong and Y. Kantor are thanked for their role in molecular  
669 sampling during these expeditions. The phylogenetic analyses were performed on the vital-it  
670 clusters ([www.vital-it.ch](http://www.vital-it.ch)). The authors thank G. Achaz, A. Lambert and S. Brouillet for the  
671 development and the implementation of the ABGD method, A. Sysoev, B. M. Olivera and Y.  
672 Kantor for the taxonomic identifications, T. Barraclough for his assistance with the GMYC  
673 method, and P. Bouchet and G. Paulay for constructive comments on the manuscript. This  
674 work was supported by the "Consortium National de Recherche en Génomique" and the  
675 "Service de Systématique Moléculaire" (UMS 2700 CNRS-MNHN). It is part of the  
676 agreement n°2005/67 between the Genoscope and the Muséum National d'Histoire Naturelle  
677 on the project "Macrophylogeny of life" directed by Guillaume Lecointre. This project is  
678 partially funded by an Alfred P. Sloan foundation (B2010-37), NSF (0940108), and NIH-  
679 NIGMS (GM088096) grants to M. Holford. Molecular data were obtained through the DNA-  
680 barcoding workflow established at the MNHN thanks to the MARBOL grant from the Alfred  
681 P. Sloan foundation, PI D. Steinke, Co-PI P. Bouchet and S. Samadi.

682 REFERENCES

- 683 Ahrens D, Monaghan MT, Vogler AP (2007) DNA-based taxonomy for associating adults  
684 and larvae in multi-species assemblages of chafers (Coleoptera: Scarabaeidae).  
685 *Molecular Phylogenetics and Evolution* **44**, 436–449.
- 686 Barberousse A, Samadi S (2010) Species from Darwin onward. *Integrative Zoology* **5**, 187-  
687 197.
- 688 Boissin E, FÉral JP, Chenuil A (2008) Defining reproductively isolated units in a cryptic and  
689 syntopic species complex using mitochondrial and nuclear markers: the brooding  
690 brittle star, *Amphipholis squamata* (Ophiuroidea). *Molecular Ecology* **17**, 1732-1744.
- 691 Bouchet P, Kantor Y, Sysoev A, Puillandre N (2011) A new operational classification of the  
692 Conoidea (Gastropoda). *Journal of Molluscan Studies* **77**, 273-308.
- 693 Bouchet P, Lozouet P, Sysoev AV (2009) An inordinate fondness for turrids. *Deep-Sea*  
694 *Research II* **56**, 1724-1731.
- 695 Castelin, M., Puillandre, N., Lozouet, P., Sysoev, A., Richer de Forges, B., Samadi, S., 2011.  
696 Molluskan species richness and endemism on New Caledonian seamounts: Are they  
697 enhanced compared to adjacent slopes? *Deep-Sea Research I* **58**, 637-646.
- 698 Chase MW, Salamin N, Wilkinson M, *et al.* (2005) Land plants and DNA barcodes: short-  
699 term and long-term goals. *Philosophical Transactions of the Royal Society B* **360**,  
700 1889-1895.
- 701 Crow KD, Munehara H, Bernardi G (2010) Sympatric speciation in a genus of marine reef  
702 fishes. *Molecular Ecology* **19**, 2089-2105.
- 703 Dayrat B (2005) Towards integrative taxonomy. *Biological Journal of the Linnean Society* **85**,  
704 407-415.
- 705 De Queiroz K (2007) Species concepts and species delimitation. *Systematic Biology* **56**, 879-  
706 886.

707 Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling  
708 trees. *BMC Evolutionary Biology* **7**, 214.

709 Duda TF (2008) Differentiation of venoms of predatory marine gastropods: divergence of  
710 orthologous toxin genes of closely related *Conus* species with different dietary  
711 specializations. *Journal of Molecular Evolution* **67**, 315-321.

712 Eckert CG, Samis KE, Loughheed SC (2008) Genetic variation across species' geographical  
713 ranges: the central–marginal hypothesis and beyond. *Molecular Ecology* **17**, 1170-  
714 1188.

715 Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package  
716 for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**, 47-50.

717 Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of  
718 mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates.  
719 *Molecular Marine Biology and Biotechnology* **3**, 294-299.

720 Fu J, Zeng X (2008) How many species are in the genus *Batrachuperus*? A phylogeographical  
721 analysis of the stream salamanders (family Hynobiidae) from southwestern China.  
722 *Molecular Ecology* **17**, 1469-1488.

723 Funk DJ, Omland KE (2003) Species-level paraphyly and polyphyly: frequency, causes, and  
724 consequences, with insights from animal mitochondrial DNA. *Annual Review of*  
725 *Ecology Evolution and Systematics* **34**, 397-423.

726 Garros C, Ngugi N, Githeko AE, Tuno N, Yan G (2008) Gut content identification of larvae  
727 of the *Anopheles gambiae* complex in western Kenya using a barcoding approach.  
728 *Molecular Ecology Resources* **8**, 512-518.

729 Goldstein PZ, DeSalle R (2011) Integrating DNA barcode data and taxonomic practice:  
730 Determination, discovery, and description. *Bioessays* **33**, 135-147.



731 Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis  
732 program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**, 95-98.

733 Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through  
734 DNA Barcodes. *Proceedings of the Royal Society B* **270**, 313-321.

735 Hebert PDN, Stoeckle MY, Zemplak TS, Francis CM (2004) Identification of birds through  
736 DNA barcodes. *PLoS Biology* **2**, 1657-1663.

737 Heralde FM, Kantor Y, Astilla MAQ, *et al.* (2010) The Indo-Pacific *Gemmula* species in the  
738 subfamily Turrinae: aspects of field distribution, molecular phylogeny, radular  
739 anatomy and feeding ecology. *Philippine Science Letters* **3**, 21-34.

740 Holford M, Puillandre N, Terryn Y, *et al.* (2009) Evolution of the Toxoglossa venom  
741 apparatus as inferred by molecular phylogeny of the Terebridae. *Molecular Biology*  
742 *and Evolution* **26**, 15-25.

743 Holynski RB (2010) Taxonomy and the mediocrity of DNA barcoding – some remarks on  
744 PACKER *et al.* 2009: DNA barcoding and the mediocrity of morphology. *Arthropod*  
745 *Systematics & Phylogeny* **143**, 143-150.

746 Hyde JR, Kimbrell CA, Budrick JE, Lynn EA, Vetter D (2008) Cryptic speciation in the  
747 vermilion rockfish (*Sebastes miniatus*) and the role of bathymetry in the speciation  
748 process. *Molecular Ecology* **17**, 1122-1136.

749 Jablonski D, Lutz RA (1980) Molluscan larval shell morphology - ecological and  
750 paleontological applications. In: *Skeletal growth of aquatic organisms* (eds. Rhoads  
751 DC, Lutz RA), pp. 323-377. Plenum Press, New York.

752 Jovelin R, Justine J-L (2001) Phylogenetic relationships within the Polyopisthocotylean  
753 monogeneans (Plathyhelminthes) inferred from partial 28S rDNA sequences.  
754 *International Journal for Parasitology* **31**, 393-401.

755 Kantor YI, Puillandre N, Olivera BM, Bouchet P (2008) Morphological proxies for taxonomic  
756 decision in turrids (Mollusca, Neogastropoda): a test of the value of shell and radula  
757 characters using molecular data. *Zoological Science* **25**, 1156-1170.

758 Keane TM, Creevey CJ, Pentony MM, Naughton TJ, McInerney JO (2006) Assessment of  
759 methods for amino acid matrix selection and their use on empirical data shows that ad  
760 hoc assumptions for choice of matrix are not justified. *BMC Evolutionary Biology* **6**,  
761 1-17.

762 Knowles LL, Carstens BC (2007) Delimiting species without monophyletic gene trees.  
763 *Systematic Biology* **56**, 887-895.

764 Knowlton N (2000) Molecular genetic analyses of species boundaries in the sea.  
765 *Hydrobiologia* **420**, 73-90.

766 Leliaert F, Verbruggen H, Wysor B, Clerck OD (2009) DNA taxonomy in morphologically  
767 plastic taxa: Algorithmic species delimitation in the *Boodlea* complex (Chlorophyta:  
768 Cladophorales). *Molecular Phylogenetics and Evolution* **53**, 122-133.

769 Lim GS, Balke M, Meier R (2011) Determining species boundaries in a world full of rarity:  
770 singletons, species delimitation methods. *Systematic Biology* **60**, AA.

771 Lohse K (2009) Can mtDNA barcodes be used to delimit species? A response to Pons et al.  
772 (2006). *Systematic Biology* **58**, 439-442.

773 Lopez-Vera E, Heimer de la Cotera EP, Maillo M, *et al.* (2004) A novel structure class of  
774 toxins: the methionine-rich peptides from the venoms of turrid marine snails  
775 (Mollusca, Conoidea). *Toxicon* **43**, 365-374.

776 Lorenz JG, Jackson WE, Beck JC, Hanner R (2005) The problems and promise of DNA  
777 barcodes for species diagnosis of primate biomaterials. *Philosophical Transactions of*  
778 *the Royal Society B* **360**, 1869-1877.

779 Mallet J (1995) A species definition for the modern synthesis. *Trends in Ecology and*  
780 *Evolution* **10**, 294-299.

781 Marshall JC (2006) Delimiting species: comparing methods for mendelian characters using  
782 lizards of the *Sceloporus grammicus* (Squamata: Phrynosomatidae) complex.  
783 *Evolution* **60**, 1050-1065.

784 Meier R, Zhang G, Ali F (2008) The use of mean instead of smallest interspecific distances  
785 exaggerates the size of the "Barcoding Gap" and leads to misidentification. *Systematic*  
786 *Biology* **57**, 809-813.

787 Miljanich GP (2004) Ziconotide: neuronal calcium channel blocker for treating severe chronic  
788 pain. *Current Medicinal Chemistry* **11**, 3029-3040.

789 Monaghan MT, Wild R, Elliot, *et al.* (2009) Accelerated species inventory on Madagascar  
790 using coalescent-based models of species delineation. *Systematic Biology* **58**, 298-311.

791 O'Meara BC (2010) New heuristic methods for joint species delimitation and species tree  
792 inference. *Systematic Biology* **59**, 59-73.

793 Olivera BM, Cruz LJ, De Santos V, *et al.* (1987) Neuronal calcium channel antagonists.  
794 Discrimination between calcium channel subtypes using .omega-conotoxin from  
795 *Conus magus* venom. *Biochemistry* **26**, 2086-2090.

796 Padial JM, Miralles A, De la Riva I, Vences M (2010) The integrative future of taxonomy.  
797 *Frontiers in Zoology* **7**, 16.

798 Parker T, Tunnicliffe V (1994) Dispersal strategies of the biota on an oceanic seamount:  
799 Implications for ecology and biogeography. *Biological Bulletin* **187**, 336-345.

800 Pons J, Barraclough TG, Gomez-Zurita J, *et al.* (2006) Sequence-based species delimitation  
801 for the DNA taxonomy of undescribed insects. *Systematic Biology* **55**, 595-609.

802 Puillandre N, Baylac M, Boisselier MC, Cruaud C, Samadi S (2009) An integrative approach  
803 of species delimitation in the genus *Benthomangelia* (Mollusca: Conoidea). *Biological*  
804 *Journal of the Linnean Society* **96**, 696-708.

805 Puillandre N, Lambert A, Brouillet S, Achaz G (2011) ABGD, Automatic Barcode Gap  
806 Discovery for primary species delimitation. *Molecular Ecology*.

807 Puillandre N, Sysoev A, Olivera BM, Couloux A, Bouchet P (2010) Loss of planktotrophy  
808 and speciation: geographical fragmentation in the deep-water gastropod genus  
809 *Bathytoma* (Gastropoda, Conoidea) in the western Pacific. *Systematics and*  
810 *biodiversity* **8**, 371–394.

811 Rambaut A, Drummond AJ (2007) Tracer v1.4. Available from  
812 <http://beast.bio.ed.ac.uk/Tracer>

813 Reeves PA, Richards CM (2011) Species delimitation under the general lineage concept: an  
814 empirical example using wild North American Hops (Cannabaceae: *Humulus lupulus*).  
815 *Systematic Biology* **60**, 45-59.

816 Rosenberg NA, Tao R (2008) Discordance of species trees with their most likely gene trees:  
817 the case of five taxa. *Systematic Biology* **57**, 131-140.

818 Ross KG, Gotzek D, Ascunce MS, Shoemaker DD (2010) Species delimitation: a case study  
819 in a problematic ant taxon. *Systematic Biology* **59**, 162-184.

820 Samadi S, Barberousse A (2006) The tree, the network, and the species. *Biological Journal of*  
821 *the Linnean Society* **89**, 509-521.

822 Samadi S, Barberousse A (2009) Species: towards new, well-grounded practices. A response  
823 to Velasco. *Biological Journal of the Linnean Society* **96**, 696-708.

824 Schlick-Steiner BC, Steiner FM, Seifert B, *et al.* (2009) Integrative taxonomy: a multisource  
825 approach to exploring biodiversity. *Annual Review of Entomology* **55**, 421-438.

826 Sirovich L, Stoeckle M, Zhang Y (2009) A scalable method for analysis and display of DNA  
827 sequences. *PLoS ONE* **4**, e7051.

828 Sirovich L, Stoeckle M, Zhang Y (2010) Structural Analysis of Biodiversity. *PLoS ONE* **5**,  
829 e9266.

830 Sites JW, Marshall JC (2003) Delimiting species: a renaissance issue in systematic biology.  
831 *Trends in Ecology and Evolution* **19**, 462-470.

832 Smith PJ, McVeagh SM, Steinke D (2008) DNA barcoding for the identification of smoked  
833 fish products. *Journal of Fish Biology* **72**, 464-471.

834 Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses  
835 with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688-2690.

836 Taylor JW, Turner E, Townsend JP, Dettman JR, Jacobson D (2006) Eukaryotic microbes,  
837 species recognition and the geographic limits of species: examples from the kingdom  
838 Fungi. *Philosophical Transactions of the Royal Society B* **361**, 1947-1963.

839 Terlau H, Olivera BM (2004) *Conus* venoms: a rich source of novel ion channel-targeted  
840 peptides. *Physiological Review* **84**, 41-68.

841 Tucker JK (2004) Catalogue of recent and fossil turrids (Mollusca: Gastropoda). *Zootaxa* **682**,  
842 1-1295.

843 Vernooy R, Haribabu E, Muller MR, *et al.* (2010) Barcoding life to conserve biological  
844 diversity: beyond the taxonomic imperative. *PLoS Biology* **8**, e1000417.

845 Weisrock DW, Shaffer HB, Storz BL, Storz SR, Voss SR (2006) Multiple nuclear gene  
846 sequences identify phylogenetic species boundaries in the rapidly radiating clade of  
847 Mexican ambystomatid salamanders. *Molecular Ecology* **15**, 2489-2503.

848 Wheeler QD (2009) *The new taxonomy* CRC Press, Boca Ratan.

849 Wiens JJ (2007) Species delimitation: new approaches for discovering diversity. *Systematic*  
850 *Biology* **56**, 875-878.

851 Will KP, Mishler BD, Wheeler QD (2005) The perils of DNA Barcoding and the need for  
852 integrative taxonomy. *Systematic Biology* **54**, 844-851.  
853 Yeates D, Seago A, Nelson L, *et al.* (2010) Integrative taxonomy, or iterative taxonomy?  
854 *Systematic Entomology*.  
855  
856

857 **"Data Accessibility:**

858 - DNA sequences: Genbank accessions EU015659, EU015661, EU015664, EU015677,  
859 EU015681, EU015682, EU015684, EU015724, EU127874-EU127882, EU820248-  
860 EU821230 for COI gene and EU015543, EU015545, EU015548, EU015562, EU015566,  
861 EU015567, EU015569, EU015609, EU127883-EU127891, EU819556-EU820247 for 28S  
862 gene

863 - All samples are vouchered in the MNHN collection. They are all registered in the Barcode  
864 of Life Datasystem (BOLD), in the project "CONO – Conoidea barcodes and taxonomy".

865

866 **Figure legends**

867

868 Figure 1: Integrative taxonomy flowchart used to delimit species in the Turridae. Starting  
869 from COI sequences from numerous specimens (step 1), PSHs are proposed using both  
870 ABGD and GMYC (step 2), then visualized using Klee diagrams (step 3). Several other  
871 criteria and characters are analysed sequentially to turn PSHs into SSHs: first, a second  
872 independent marker (the 28S gene), then the geographic and bathymetric ranges, in  
873 association with the larval dispersion capacities, and finally the morphological differences are  
874 compared. In some cases, the evidence will not favor any of the two hypotheses, and the  
875 taxonomist will have to subjectively make a decision (using either a lumpers or splitter  
876 approach) waiting for more conclusive data.

877

878 Figure 2: Pairwise distribution for the COI gene and ABGD results. a) Distributions of p-  
879 distance, K2P distances and TN distances between each pair of specimens for the COI gene.  
880 b) Same results, but focusing on the barcode gap zone. c) ABGD results, with the number of  
881 PSHs obtained for each prior intraspecific divergence.

882

883 Figure 3: COI gene results. a) Bayesian COI gene tree with Posterior Probabilities (>0.8) and  
884 Bootstraps (>50) indicated next to each node. The 103 PSHs listed in Table 1 (first column)  
885 are represented each by a single branch (the intra- PSH trees are not shown). Black brackets  
886 indicate the PSHs that were subsequently grouped into one SSH. \*: PSHs with a shell  
887 illustration. b) Klee diagrams for the COI gene showing the correlations among indicator  
888 vectors for the less inclusive dataset corresponding to the 103 PSHs provided by the multiple-  
889 threshold GMYC (the black arrows point to the groups of PSHs recognized as a single PSH  
890 with ABGD); gradations of red and yellow color in the Klee diagram indicate areas of



891 conflict. c) for the most inclusive dataset corresponding to the 87 PSHs provided by ABGD  
892 (the black arrows point to PSHs that are divided into several PSHs by the multiple-threshold  
893 GMYC method).

894

895 Figure 4: 28S gene results. a) Klee diagrams for the 28S gene showing the correlations among  
896 indicator vectors for the 86 PSHs sequenced for this gene. b) Phylogenetic tree obtained with  
897 the 208 28S haplotypes (Bayesian analysis). Posterior Probabilities (>.8) and Bootstraps (>50)  
898 are reported for each node. Numbers at the tip of the branches refer to the PSH numbers  
899 (Table 1). Red star: monophyletic PSHs. Black arrow: haplotype shared by several PSHs.

900

901 Figure 5: Example of congruent SSHs. a) COI tree for the SSH *Ptychosyrinx* sp. 1, *X.*  
902 *gemmaoides*, *G. unilineata*, *Gemmula* sp. 4 and *Ptychosyrinx* sp. 2., corresponding depth of  
903 collection for each specimen is given. b) 28S tree for the corresponding SSHs. c) COI  
904 haplotype network for the same SSHs. Some specimens are illustrated for each SSH by their  
905 shells.

906

907 **Table**

908 Table 1: List of PSHs, as defined with the ABGD (M = More inclusive partition and L = Less inclusive partition) and GMYC (S = Single-  
909 threshold and M = Multiple-threshold) analyses of the COI gene. Number of specimens (N) and phylogenetic support are provided for both COI  
910 and 28S genes. Geographical, bathymetrical and morphological data are also provided. NA: non applicable (one or no specimen or one or no  
911 haplotype). Pa: paraphyletic; Po: polyphyletic. Phil.: Philippines; Sol.: Solomon Islands; Ches.: Chesterfield Islands; Van.: Vanuatu; N.-C.: New-  
912 Caledonia; Tai.: Taiwan. \*: indicates that at least one specimen from each of the corresponding PSHs was collected at the same station. The  
913 “Genetic structure” column lists the PSHs for which COI and 28S structure was tested (Table S1). The depth range index refers to the statistical  
914 tests explain in the Material and Methods section (a PSH number indicates that the test is significant for this PSH; n.s.: not significant).  
915 Morphological identification: *G.*: *Gemmula*; *P.*: *Ptychosyrinx*; *L.*: *Lophiotoma*; *T.*: *Turris*; *Td.*: *Turridrupa*; *I.*: *Iotyrris*; *X.*: *Xenuroturris*.  
916

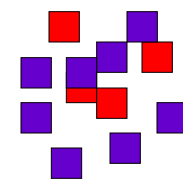
PSH	ABGD		GMYC		COI		28S			Geography			Depth		Morphological ID	SSH
	M	L	S	M	N	Support (ML/BA)	N	Support (ML/BA)	Haplotypes shared with	Region	Distribution	Genetic structure	Range (m)	Indices		
1	x	x	x	x	1	NA	0	NA		Sol			282-327		<i>G.</i> 1	<i>Gemmula</i> sp. 1
2	x	x	x	x	1	NA	1	NA		Phil			85-88		<i>Gemmula</i> sp. 2	<i>Gemmula</i> sp. 2
3	x	x	x	x	5	100/1	4	Pa/Pa		Phil			35-100		<i>G.</i> 2	<i>Gemmula</i> sp. 3
4	x	x	x	x	3	NA/1	2	99/1		Ches			345-413		<i>Td. armillata</i>	<i>Turridrupa armillata</i>
5	x	x	x	x	2	100/1	2	NA		Van			0-49		<i>Td. neojubata</i>	<i>Turridrupa neojubata</i>
6	x	x	x	x	1	NA	1	NA		Phil			6-8		<i>Td. Bijubata</i>	<i>Turridrupa cf. bijubata</i> 1
7	x	x	x	x	1	NA	1	NA		Van			0-49		<i>Td. albofasciata</i>	<i>Turridrupa albofasciata</i>
8	x	x	x	x	4	100/1	1	NA		Phil, Van			20-110		<i>Td. astricta</i>	<i>Turridrupa astricta</i>
9	x	x	x	x	1	NA	1	NA		Van			0-49		<i>Td. bijubata</i>	<i>Turridrupa cf. bijubata</i> 2
10	x	x	x	x	4	96/1	4	Po/Po		Phil, Van			0-49		<i>Td. bijubata</i>	<i>Turridrupa cf. bijubata</i> 3
11	x	x	x	x	1	NA	1	NA		Phil			593		<i>P.</i> 1	<i>Ptychosyrinx</i> sp. 1
12	x	x	x	x	2	100/1	1	NA		Ches			372-404		<i>X. gemmuloides</i>	<i>Xenuroturris gemmuloides</i>
13	x	x	x	x	15	89/1	11	75/1		Ches, N-C, Phil, Sol			410-741		<i>G. unilineata</i>	<i>Gemmula unilineata</i>
14	x	x	x	x	4	99/1	2	93/1		Sol			897-1057		<i>P.</i> 2	<i>Ptychosyrinx</i> sp. 2
15	x	x	x	x	24	100/1	5	76/1		Ches, N-C			440-1150		<i>G.</i> 3	<i>Gemmula</i> sp. 4
16	x	x	x	x	5	99/1	0	NA		Van			503-636		<i>G.</i> 1	<i>Gemmula</i> sp. 5

17	x	x	x	x	8	100/1	2	-	18	Van		131-308			<i>Lophiotoma cf. indica</i> 1
18	x	x	x	x	8	95/1	1	NA	17	Van, Sol		131-308		<i>L. indica</i>	
19	x	x	x	x	2	NA/1	1	NA		Van		83-339			<i>Lophiotoma cf. indica</i> 2
20	x	x	x	x	1	NA	1	NA		Phil		155-160		<i>L. tayabaensis</i>	<i>Lophiotoma tayabaensis</i>
21	x	x	x	x	1	99/1	1	NA		Sol	allopatric	150-160	NA	<i>L. friedrichbonhoefferi</i>	<i>Lophiotoma cf. friedrichbonhoefferi</i> 1
22	x	x	x	x	3	100/1	3	NA		Van		106-148			<i>Lophiotoma cf. friedrichbonhoefferi</i> 2
23	x	x	x	x	7	94/1	4	Pa/Pa		Phil		72-139		<i>L. bisaya</i>	<i>Lophiotoma bisaya</i>
24	x	x	x	x	7	95/1	6	-	26	Sol, Van		83-160			
25	x	x	x	x	2	76/1	1	NA	26	Phil	sympatric*	42-44	25	<i>L. indica</i>	<i>Lophiotoma cf. indica</i> 3
26	x	x	x	x	6	99/96	5	-	24, 25	Phil		42-79			
27	x	x	x	x	6	100/1	6	-	28, 29, 30, 31	Phil		219-318		<i>L. sikatunai</i>	<i>Lophiotoma sikatunai</i>
28	x	x	x	x	2	99/1	2	-	27, 29, 30	Ches	sympatric*	310-400	n.s.		
29	x	x	x	x	6	60/.82	4	-	27, 28, 30	Ches		267-400		<i>L. unedo</i>	<i>Lophiotoma unedo</i>
30	x	x	x	x	32	88/1	24	-	27, 28, 29, 31	Sol, N-C, Phil, Van		147-391			
31	x	x	x	x	68	93/1	14	-	27, 30	Sol, Van	COI	131-400			
32	x	x	x	x	4	96/1	3	-	36	Phil		229-400			<i>Lophiotoma cf. panglaoensis</i> 1
33	x	x	x	x	10	54/1	1	NA		Phil	sympatric*	182-346	33, 35	<i>L. panglaoensis</i>	<i>Lophiotoma cf. panglaoensis</i> 2
34	x	x	x	x	6	82/1	0	NA		Phil, Sol		173-400			<i>Lophiotoma cf. panglaoensis</i> 3
35	x	x	x	x	10	97/1	4	80/.97		Sol, Van		131-600			<i>Lophiotoma cf. panglaoensis</i> 4
36	x	x	x	x	94	99/1	75	Po/Po	32	Sol, Van	COI & 28S	350-659		<i>L. indica</i>	
37	x	x	x	x	1	NA	1	NA		Phil		8-22		<i>T. babylonia</i>	<i>Turris babylonia</i>
38	x	x	x	x	1	NA	1	NA		Van		112-148		<i>T. spectabilis</i>	<i>Turris spectabilis</i>
39	x	x	x	x	15	100/1	15	58/1		Van		0-55		<i>T. garnonsii</i>	<i>Turris garnonsii</i>
40	x	x	x	x	1	NA	1	NA		Van		0		<i>X. legitima</i>	<i>Xenuroturris legitima</i>
41	x	x	x	x	2	NA/1	2	NA		Van		20		<i>I. musivum</i>	<i>Iotyrris musivum</i>
42	x	x	x	x	3	86/1	2	55/.90		Van		0-49		<i>I. cingulifera</i>	<i>Iotyrris cingulifera</i>
43	x	x	x	x	4	98/1	4	Pa/Pa		Van		16-20		<i>I. devoizei</i>	<i>Iotyrris devoizei</i>
44	x	x	x	x	1	NA	1	NA		Phil		85-88		(juvenile)	<i>Gemmula</i> sp. 6
45	x	x	x	x	1	NA	0	NA		Phil		120		G. 4	<i>Gemmula</i> sp. 7
46	x	x	x	x	2	NA/	2	68/.99		Van		0-49		<i>G. lisajoni</i>	<i>Gemmula lisajoni</i>
47	x	x	x	x	4	100/1	3	-	60, 61	Van		0-49		<i>L. albina</i>	<i>Lophiotoma albina</i>
48	x	x	x	x	1	NA	1	NA		Van		266-281		G. 1	<i>Gemmula</i> sp. 8
49	x	x	x	x	1	NA	0	NA		Sol		173-379		G. 5	<i>Gemmula</i> sp. 9
50	x	x	x	x	1	NA	0	NA		Sol		286-423			<i>Gemmula</i> sp. 10
51	x	x	x	x	1	98/1	0	NA	NA	N-C	sympatric*	386-391	NA	G. 6	
52	x	x	x	x	2	60/.99	0	NA	NA	N-C		386-391			<i>Gemmula</i> sp. 11
53	x	x	x	x	8	93/1	6	90/1		Phil, Sol, Van		11-176			<i>Gemmula cf. monilifera</i> 1
54	x	x	x	x	9	100/1	9	80/1		Van		0-118			<i>Gemmula cf. monilifera</i> 2
55	x	x	x	x	8	100/1	7	92/1	56	Van	sympatric*	0-99	NA		<i>Gemmula cf. monilifera</i> 3
56	x	x	x	x	1	NA	1	NA	55	Van		0-49			
57	x	x	x	x	1	NA	1	NA		Phil		2-3			<i>Gemmula cf. hombroni</i> 1
58	x	x	x	x	1	NA	1	NA		Phil		85-88		<i>G. hombroni</i>	<i>Gemmula cf. hombroni</i> 2
59	x	x	x	x	18	100/1	16	93/1		Van		0-99			<i>Gemmula cf. hombroni</i> 3
60	x	x	x	x	5	82/1	2	-	47, 61	Phil, Sol		410-480			
61	x	x	x	x	19	100/1	17	-	47, 60	Sol, Van	sympatric	503-773	60, 61	G. 1	<i>Gemmula</i> sp. 12
62	x	x	x	x	1	NA	0	NA		Van		184-271		G. 7	<i>Gemmula</i> sp. 13
63	x	x	x	x	8	99/1	2	90/1		Sol		150-176			<i>Gemmula</i> sp. 14
64	x	x	x	x	20	100/1	18	Pa/Pa		Phil		98-356		G. 8	<i>Gemmula</i> sp. 15
65	x	x	x	x	2	100/1	1	NA	66	Phil		11-20		<i>L. jickelli</i>	<i>Lophiotoma jickelli</i>
66	x	x	x	x	8	93/1	8	-	65	Van		0-58			

67	x	x	x	x	6	100/1		6	Pa/.86				Phil			0-3		<i>L. polytropa</i>	<i>Lophiotoma polytropa</i>
68				x	11		53/1	8					Van			0-49	NA	<i>L. abbreviata</i>	<i>Lophiotoma abbreviata</i>
69	x	x	x	x	23	100/1	96/1	19	61/.99	54/.97			Van, Phil	sympatric*		0-49		<i>L. brevicaudata</i>	<i>Lophiotoma brevicaudata</i>
70				x	3	100/1		2	91/1				Van			0-49		<i>L. ruthveniana</i>	<i>Lophiotoma ruthveniana</i>
71	x	x	x	x	14	100/1		13	79/1				Van			0-49		<i>L. picturata</i>	<i>Lophiotoma picturata</i>
72				x	1		NA	1		NA			Phil			2-15	NA		
73	x	x	x	x	2	97/1	100/.99	2	73/1	NA			Van	allopatric		0-99		<i>L. acuta</i>	<i>Lophiotoma cf. acuta</i> 1
74				x	101	100/1		91	81/1				Van, Phil		COI & 28S	0-99			<i>Lophiotoma cf. acuta</i> 2
75	x	x	x	x	1	NA		0	NA				N-C			418-421		<i>G. rarimaculata</i>	<i>Gemmula cf. rarimaculata</i> 1
76				x	1	NA		0	NA				Phil			97-120		<i>G. monilifera</i>	<i>Gemmula cf. monilifera</i> 4
77	x	x	x	x	3	100/1		1	NA				N-C, Ches			175-370		<i>G. rarimaculata</i>	<i>Gemmula cf. rarimaculata</i> 2
78				x	4	100/1		4	91/1				Van, Phil			62-118		<i>G. hastula</i>	<i>Gemmula hastula</i>
79	x	x	x	x	77	100/1		64	53/1				Phil, Van		COI & 28S	35-196		<i>G. sogodensis</i>	<i>Gemmula cf. sogodensis</i> 1
80				x	1	NA		1	NA				Ches			330-331		<i>G. 1</i>	<i>Gemmula sp. 16</i>
81	x	x	x	x	1	NA		1	NA				Ches			627-741		<i>G. 9</i>	<i>Gemmula sp. 17</i>
82				x	23		-.93	16		74/1			Van			323-659			
83	x	x	x	x	1	97/1	NA	0	NA	NA			Sol	allopatric	COI	381-422	NA	<i>G. 10</i>	<i>Gemmula sp. 18</i>
84				x	12	100/1	-.87	11		-	85		Phil, Sol, Van			318-659		<i>G. 1</i>	<i>Gemmula sp. 19</i>
85	x	x	x	x	30		84/.97	24	56/1	-	84		Ches, Sol, Van	sympatric*	COI	345-636	n.s.		
86				x	2	NA/1		2	-		95		Van			350-400		<i>G. 11</i>	<i>Gemmula sp. 20</i>
87				x	1	NA		1	NA				Phil			342-358		<i>G. 5</i>	<i>Gemmula sp. 21</i>
88	x	x	x	x	1	NA		0	NA				Sol			630-836		<i>G. 12</i>	<i>Gemmula sp. 22</i>
89				x	1		NA	1	NA	NA			Ches			568-570		<i>G. 13</i>	
90	x	x	x	x	6	Po/Po	92/1	4	NA	70/1			Phil, Sol, Van	allopatric		416-786	NA	<i>G. 14</i>	<i>Gemmula sp. 23</i>
91				x	2		59/.98	0	NA	NA			Sol			484-836			
92	x	x	x	x	1	92/1	NA	0	NA	NA			Ches	allopatric		485-500	NA	<i>G. 15</i>	<i>Gemmula sp. 24</i>
93				x	1		NA	0	NA	NA			Ches			490-500		<i>G. 16</i>	
94	x	x	x	x	3	69/.96	100/1	1	NA	NA			Phil	allopatric		269-378	NA		<i>Gemmula sp. 25</i>
95				x	7	94/1		5	-		86		Van			350-600		<i>G. 5</i>	<i>Gemmula sp. 26</i>
96	x	x	x	x	1	NA		1	NA				Phil			422-431		(juvenile)	<i>Gemmula sp. 27</i>
97				x	30	98/1		17	Pa/Pa				Phil		COI & 28S	219-1762		<i>G. diomedea</i>	<i>Gemmula diomedea</i>
98	x	x	x	x	9	100/1		6	NA				Phil, Sol			65-160		<i>G. speciosa</i>	<i>Gemmula speciosa</i>
99				x	10	96/1		9	NA				Phil			85-137		<i>G. kieneri</i>	<i>Gemmula kieneri</i>
100	x	x	x	x	1	NA		0	NA				Taiwan			157-275		<i>G. cosmoi</i>	<i>Gemmula cf. cosmoi</i> 1
101				x	7	96/1		2	Po/Po				Sol			300-430		<i>G. martini</i>	<i>Gemmula martini</i>
102	x	x	x	x	24	98/1		16	91/1				Van			131-444		<i>G. cosmoi</i>	<i>Gemmula cf. cosmoi</i> 2
103				x	71	95/1		62	82/1				Phil			72-361		<i>G. sogodensis</i>	<i>Gemmula cf. sogodensis</i> 2



STEP 1



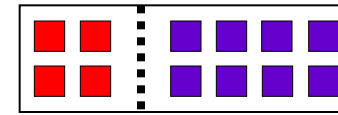
**SAMPLING**  
Large datasets, including numerous specimens from various localities and environments

STEP 2

GATGATCAATTGTAC  
GATGATCAGTTGTAT  
GACGATCAATTATAT  
GATGATCAGTTGTAT  
GATGATCAGTTGTAT

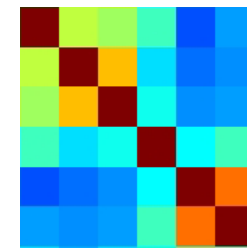
**DNA SEQUENCING**  
Monolocus approach (COI "Barcode fragment")

GMYC and ABGD (exploratory methods)

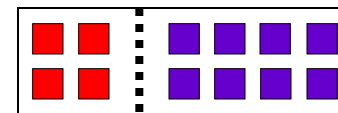


**PRIMARY SPECIES HYPOTHESES**

STEP 3



**KLEE DIAGRAMS**  
Graphical visualisation



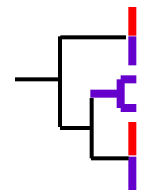
**PSH conversion to SSH : 1 or 2 SSHs?**

**COMPARATIVE EVIDENCE**  
Using additional characters and criteria

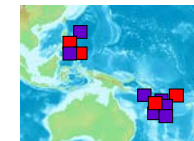
STEP 4

**In favor of 1 single SSH:**

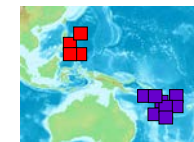
Shared haplotypes with 28S



Overlapping geographic range, with Bathymetric differences and weakly dispersive larvae



Non-overlapping geographic range without barrier and weakly dispersive larvae



*Morphology homogeneous*



**Inconclusive**

Overlapping geographic range, no bathymetric or morphological differences

Overlapping geographic range, bathymetric differences and non dispersive larvae

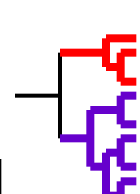
Non-overlapping geographic range without barrier and with non dispersive larvae

**In favor of 2 different SSHs:**

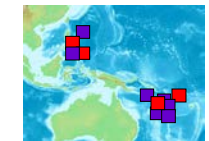
No shared haplotypes with 28S

and

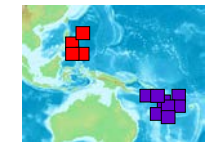
Reciprocal monophyly with both genes



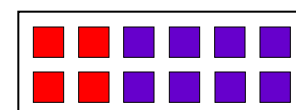
Overlapping geographic range, with bathymetric differences and dispersive larvae



Non-overlapping geographic range with barrier or without barrier but with dispersive larvae



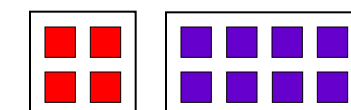
*Morphology differences*



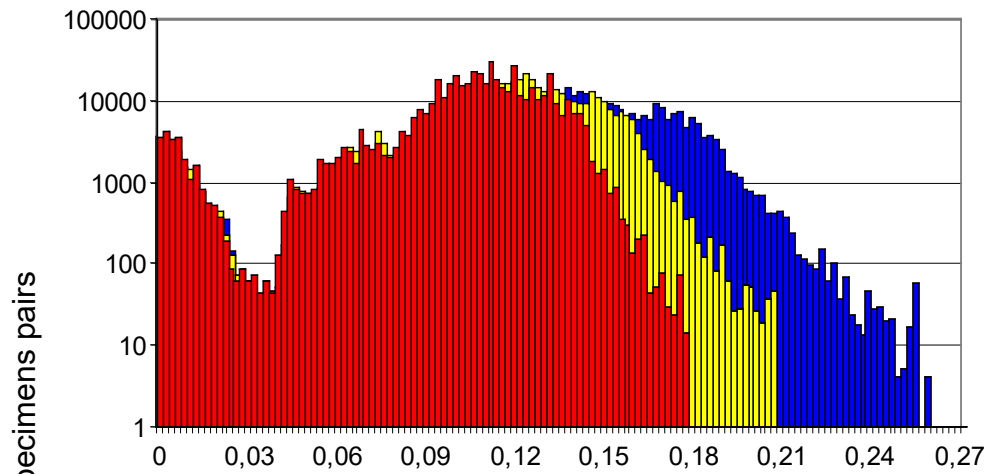
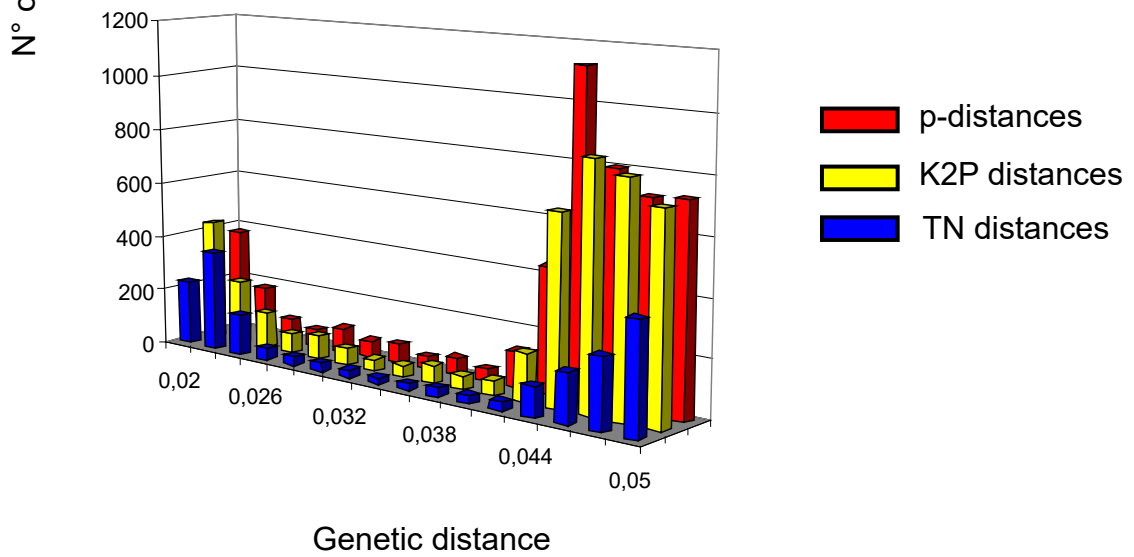
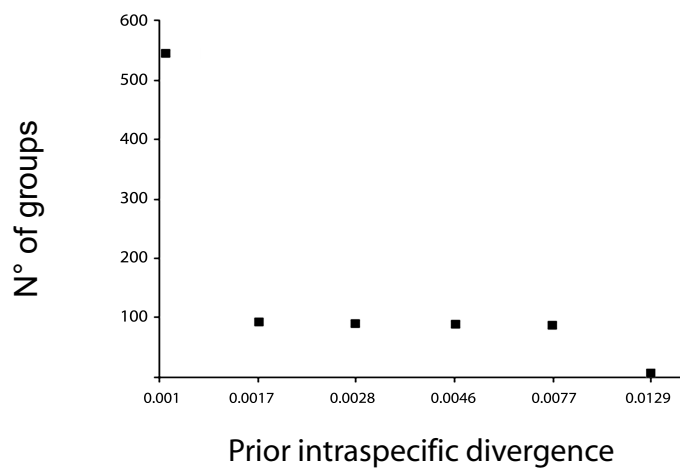
**LUMPER**

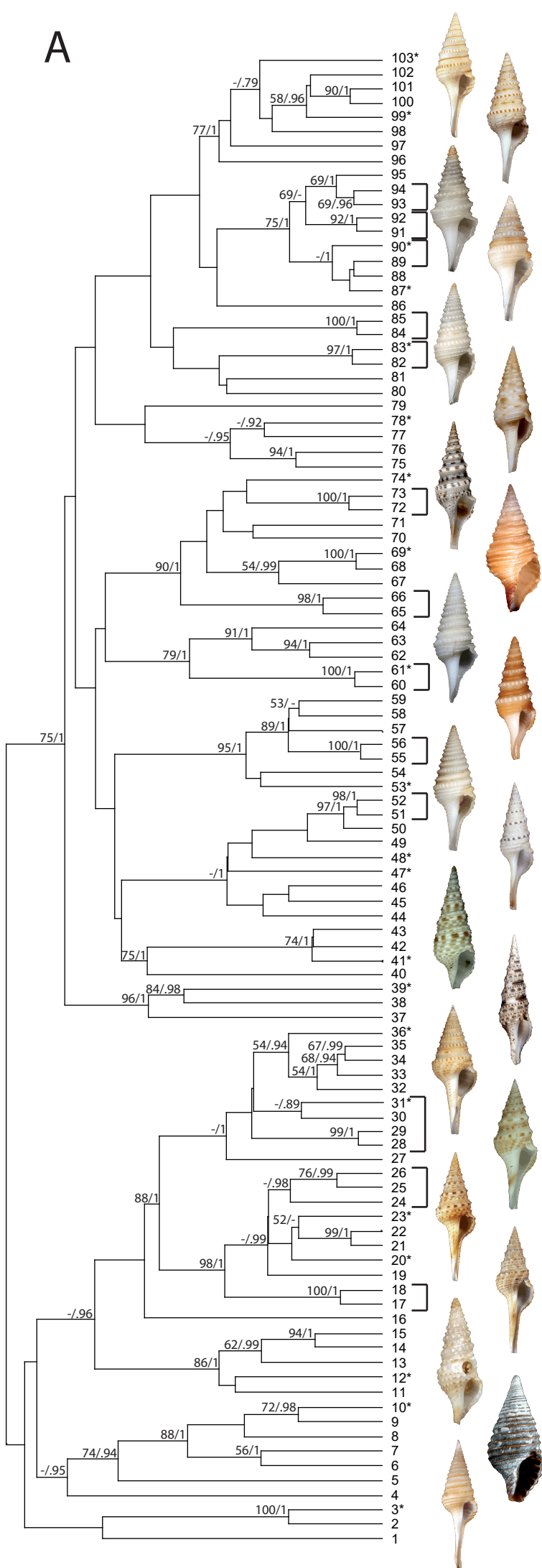
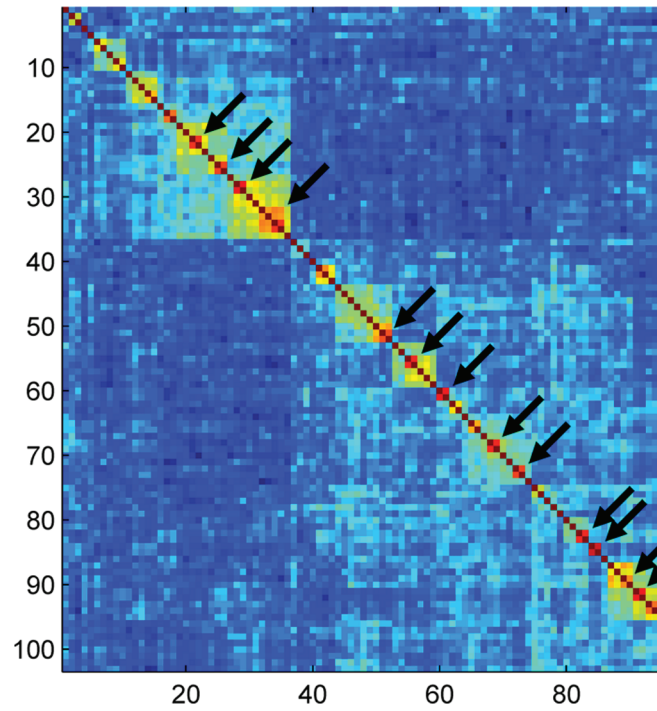
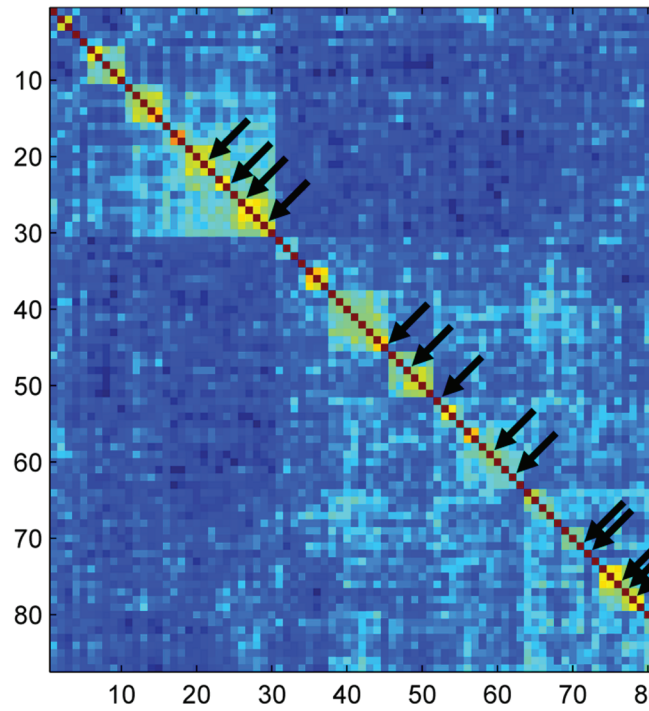
Taxonomist's choice

**SPLITTER**

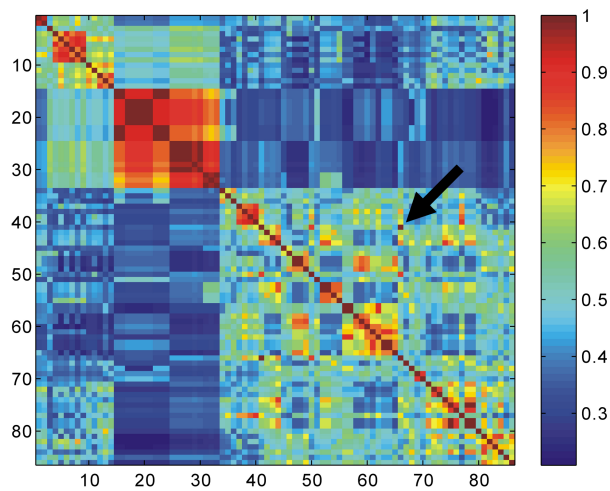
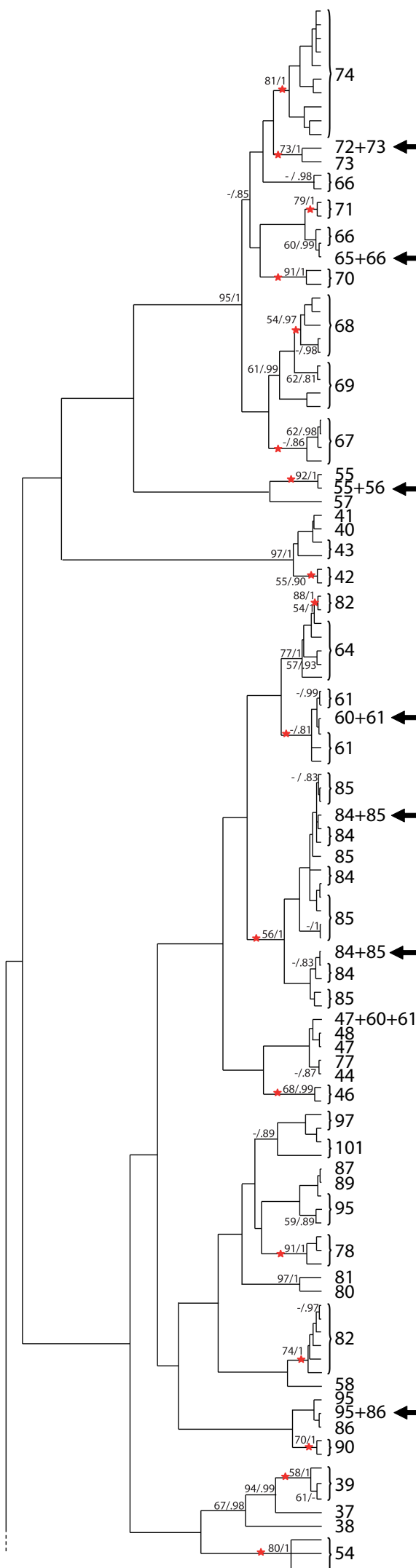
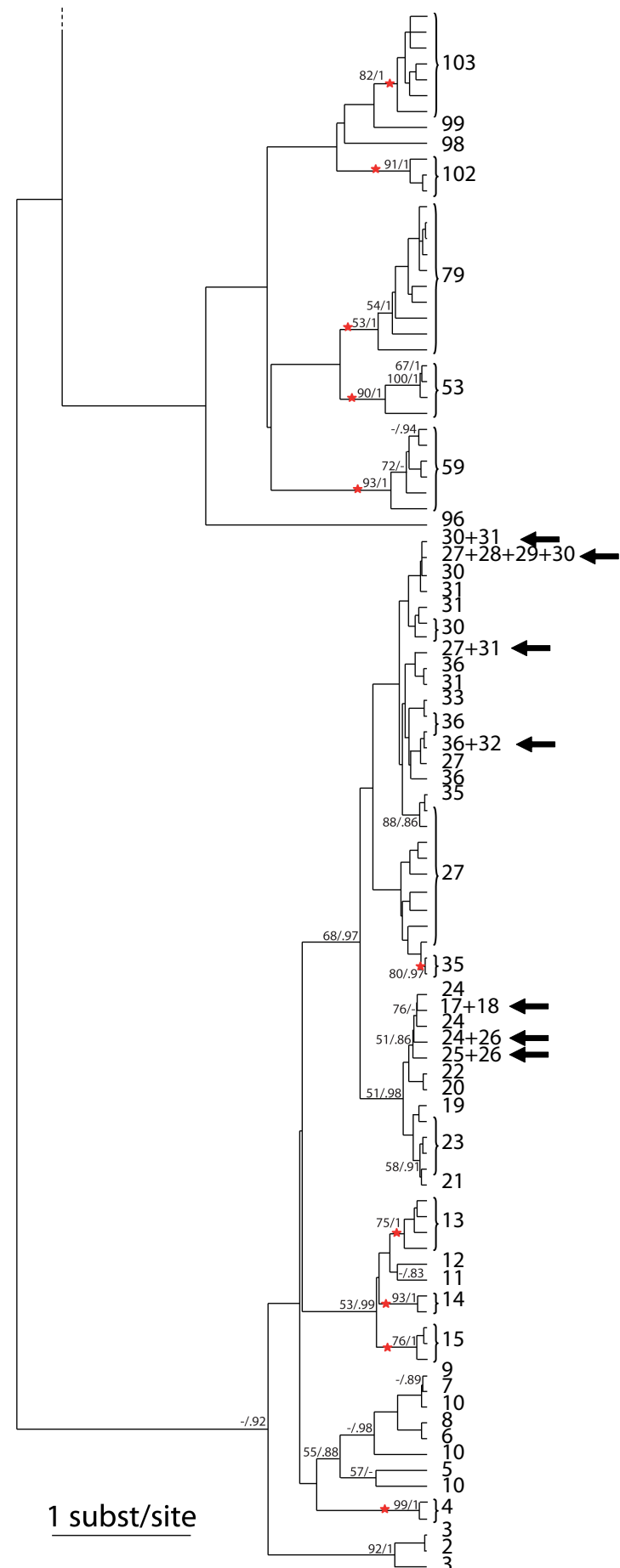


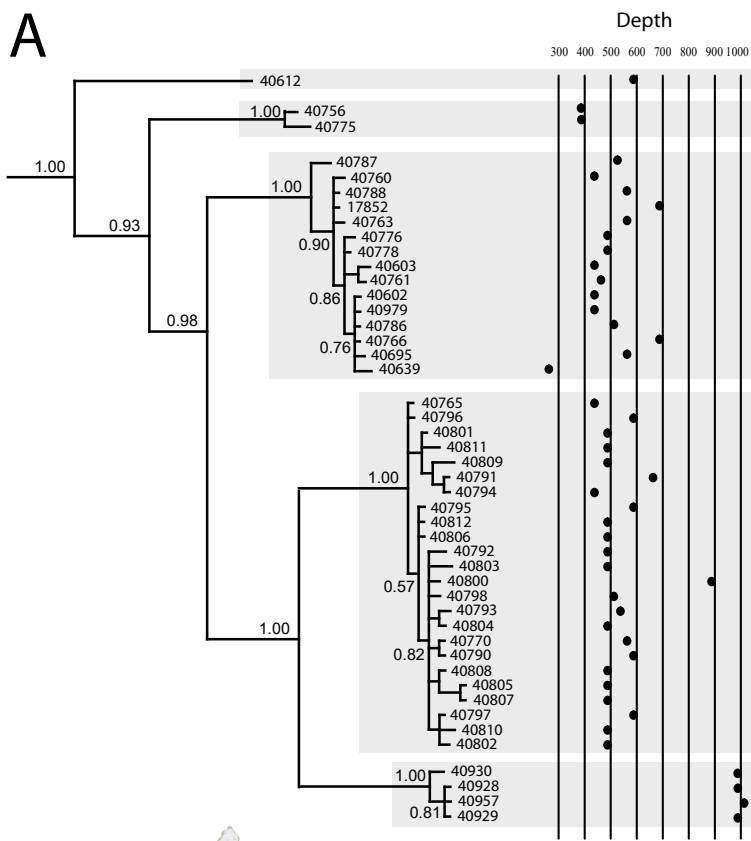
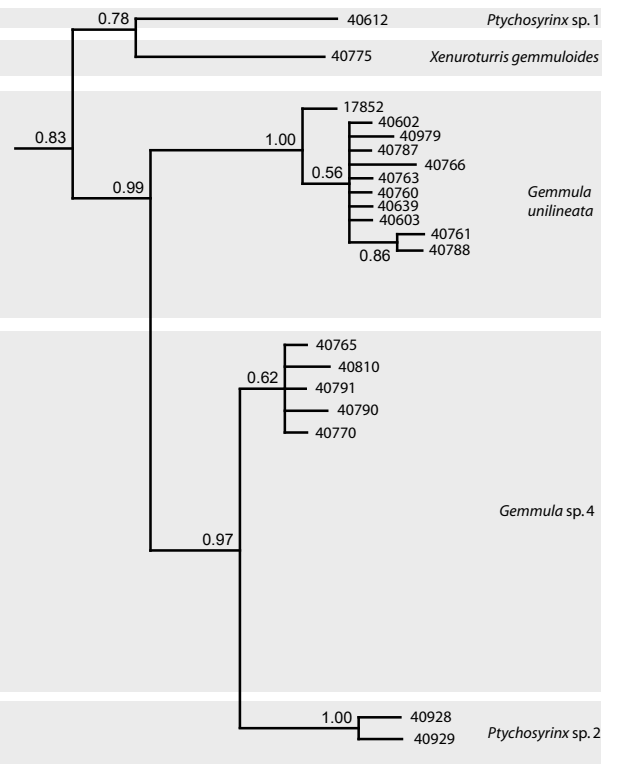
**SECONDARY SPECIES HYPOTHESES**

**A****B****C**

**A****B****C**



**A****B**

**A****B****C**