MOLECULAR ECOLOGY

Molecular Ecology (2012) doi: 10.1111/j.1365-294X.2012.05559.x

Large-scale species delimitation method for hyperdiverse groups

N. PUILLANDRE,* M. V. MODICA,† Y. ZHANG,‡ L. SIROVICH,‡ M.-C. BOISSELIER,*§ C. CRUAUD,¶ M. HOLFORD** and S. SAMADI*§

*'Systématique, Adaptation et Evolution', UMR 7138 UPMC-IRD-MNHN-CNRS (UR IRD 148), Muséum National d'Histoire Naturelle, Département Systématique et Evolution, CP 26, 57 Rue Cuvier, F-75231 Paris Cedex 05, France, †Dipartimento di Biologia e Biotecnologie 'Charles Darwin', 'La Sapienza', University of Rome, Viale dell'Universita' 32, 00185 Rome, Italy, ‡Laboratory of Applied Mathematics, Mount Sinai School of Medicine, One Gustave L. Levy Place, New York, NY 10029, USA, §Service de systématique moléculaire, UMS2700 CNRS-MNHN, Muséum National d'Histoire Naturelle, Département Systématique et Evolution, CP 26, 57 Rue Cuvier, F-75231 Paris Cedex 05, France, ¶GENOSCOPE, Centre National de Séquençage, 2 rue Gaston Crémieux, CP 5706, 91057 Evry Cedex, France, **Hunter College and Graduate Center, City University of New York, 95 Park Ave NY, NY 10065, USA, and The American Museum of Natural History, Central Park West at 79th Street, NY, NY 10024, USA

Abstract

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

Keywords: ABGD method, barcoding, Conoidea, GMYC method, Klee diagrams, Integrative taxonomy, Turridae

Received 21 October 2011; revision received 1 February 2012; accepted 10 February 2012

Introduction

The rapidly increasing rate of biodiversity extinction coupled with the magnitude of unknown biodiversity requires accurate and effective methods of species delimitation (Wiens 2007). The onset of the 21st century has seen the development of technological advances that can accelerate the description of biodiversity (Wheeler 2009), such as, DNA-barcoding initiatives, which are an attempt to identify specimens at the species-level using

Correspondence: Nicolas Puillandre, Fax: +33 1 40 79 38 44; E-mail: puillandre@mnhn.fr

a single-gene library (Hebert *et al.* 2003; Vernooy *et al.* 2010). DNA barcoding has proved effective in identifying larvae (Ahrens *et al.* 2007), processed biological products (Smith *et al.* 2008) or gut contents (Garros *et al.* 2008), as well as a taxonomic tool to aid in defining species, particularly when morphological characters are shown to be a poor proxy of species boundaries (Taylor *et al.* 2006). For the bulk of undescribed biodiversity, the single-gene approach of DNA-barcoding project may be used, not to identify specimens, but as a primary glance, that is, primary species hypotheses (PSHs) for approximating species descriptions (Goldstein & DeSalle 2011).

Problems linked to a single-gene approach, such as the presence of pseudogenes (Lorenz *et al.* 2005), incomplete lineage sorting (Funk & Omland 2003) or introgression (Chase *et al.* 2005), accentuate the need for an integrated analyses for species identification. One strategy used to avoid single-gene pitfalls is to increase the gene sampling to two or more, if possible, unlinked, genes (see e.g. Weisrock *et al.* 2006; Knowles & Carstens 2007; Boissin *et al.* 2008; O'Meara 2010; Ross *et al.* 2010). Another approach is to challenge the patterns of diversity drawn using molecular data with other sources of evidence, such as morphological characters, ecological factors, geographic distributions and other criteria (e.g. monophyly, reproductive isolation). This process of modification and validation of the species hypotheses that compiles various data and criteria is referred to as integrative taxonomy (Dayrat 2005; Will *et al.* 2005; De Queiroz 2007; Wiens 2007; Samadi & Barberousse 2009; Schlick-Steiner *et al.* 2009; Barberousse & Samadi 2010; Yeates *et al.* 2010; Reeves & Richards 2011). An integrative approach, starting with molecular characters, is particularly applicable for hyperdiverse groups, where most species are unknown and for which the quality of morphological characters as proxies for determining species boundaries is circumspect.

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

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

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

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

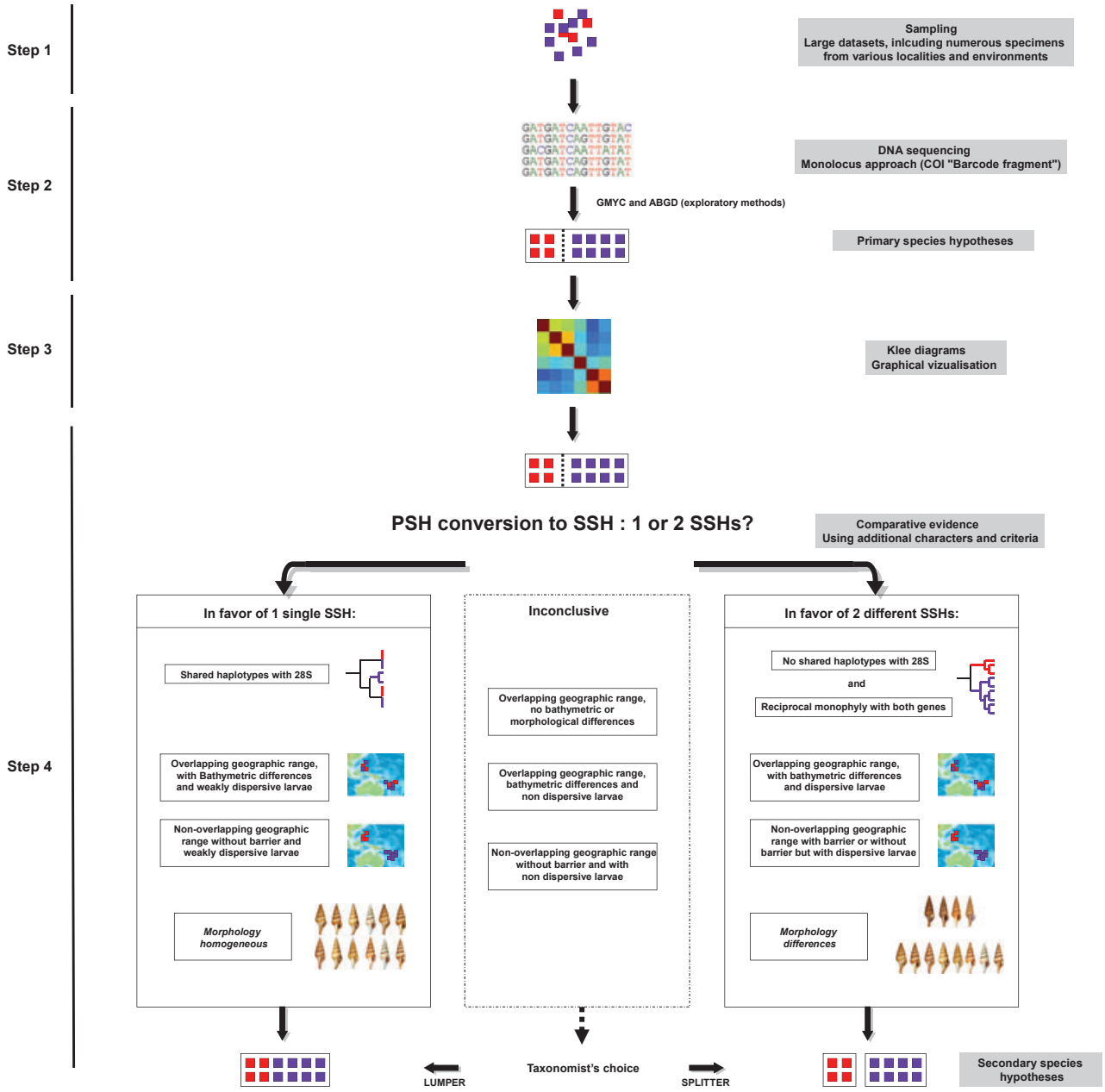

Fig. 1 Integrative taxonomy flowchart used to delimit species in the Turridae. Starting with COI sequences from numerous specimens (step 1), PSHs are proposed using both ABGD and GMYC (step 2) and then visualized using Klee diagrams (step 3). Several other criteria and characters are analysed sequentially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

species boundaries (Leliaert *et al.* 2009) even if it was argued that in some cases it could lead to an overestimation of the number of species (Lohse 2009). ABGD detects the breaks in the distribution of genetic pairwise distances, referred to as the 'barcode gap' (Hebert *et al.* 2003), relying exclusively on genetic distance between DNA sequences. To construct reliable PSHs for the Turridae, a data set of 1000 COI sequences of Turridae was collected in the South-West Pacific and analysed using both the GMYC and ABGD models.

Step 3: Visualization of PSHs using Klee diagrams. A recently developed method for processing genomic data sets, referred to as an 'indicator vector' (Sirovich *et al.* 2009, 2010), produces an optimal classifier of a

taxonomic group for biodiversity studies. This approach enables accurate quantitative display of affinities amongst taxa at various scales and extends to large genomic data sets. Indicator vectors are determined from each predefined set of nucleotide sequences (here the PSHs). The indicator vector for each PSH is used to build a structure matrix that accurately depicts affinities as correlations within and amongst groups, or alternately as directly derivable distances. The structure matrix is presented as a colour map, termed 'Klee diagram' based on its resemblance to the works by the artist Paul Klee. Klee diagrams visualize the correlation patterns recovered for the PSHs, which are identified, respectively, from GMYC and ABGD (Fig. 1, Step 3).

Step 4: Consolidation of PSHs into secondary species hypotheses (SSHs). As stated before, delimiting species based on one gene is risky, and each PSH should be individually challenged using additional evidence. Additional criteria are used either to consolidate the PSHs when GMYC and ABGD are in agreement or to choose the most likely option amongst alternate PSHs proposed by GMYC and ABGD (Fig. 1, Step 4). For the Turridae, SSHs were determined by the analysis of additional gene sequences (rRNA 28S gene), geographic and bathymetric data, morphological characters and using monophyly and gene flow criteria. In the proposed hierarchy, the agreement of several independent genes is generally valuable evidence to support the existence of two (or more) independent evolutionary lineages recognized as species (Knowlton 2000). The definitive split of two lineages may be also supported by other sources of evidence, which includes intrinsic factors, such as the dispersal ability of individuals or their bathymetric preferences, and extrinsic factors, such as the geographic distribution of the habitats or the presence of geographic barriers. Figure 1 lists the different lines of evidence that can be used in favour of either one or two species. Based on the morphological characters, proposed SSHs are then tentatively linked to the taxonomic names available in the literature. Using a sampling set of 1000 specimens, 87 Turridae SSHs are proposed based on a comparative analysis of bioinformatics species prediction tools GMYC and ABGD integrated with other available data. The strategy outlined in Fig. 1 is specific for marine gastropods with internal fecundation, but could easily be adapted to other organisms with different life-history traits.

Material and methods

Sampling

Specimens of Turridae were collected in different geographic regions: Taiwan (Taïwan 2004 expedition), Philippines (Panglao 2004 and 2005, Aurora 2007), Solomon Islands (Salomon 2, SalomonBOA 3), Vanuatu (BOA 1, Santo 2006), Chesterfield Islands (EBISCO) and New Caledonia (Norfolk 2—Norfolk ridge) (Table S2, Supporting information). A fragment of the foot was clipped from anaesthetized specimens and preserved in 95% ethanol, while shells were kept intact for morphological analyses. The sampling strategy was designed to maximize the specific diversity within the set of collected specimens: (i) the prospected area is not comprehensive of the Turridae (they are present in other regions, for example Africa, Central America) but corresponds to the centre of diversity of the Turridae (South-West Pacific, from Philippines to Vanuatu), (ii) deep to shallow waters were explored (depth range 0–1762 m). All the specimens belonging to the family Turridae were analysed, without taking into account any kind of a priori species or population delimitation. This strategy would lead to potentially include several specimens for each species, but also to include potential cryptic species. One thousand specimens were analysed, and for each of them, data corresponding to their sampling site (geographic coordinates, depth of collection) were database (Barcode of Life Database project 'Conoidea barcodes and taxonomy'). All specimens and DNA extracts are stored in the Museum National d'Histoire Naturelle collection.

Sequencing

DNA was extracted from a piece of foot, using a 6100 Nucleic Acid Prepstation system (Applied Biosystem). Two gene fragments were amplified: (i) a fragment of 658 bp of cytochrome oxidase I (COI) mitochondrial gene using universal primers LCO1490 and HCO2198 (Folmer *et al.* 1994) and (ii) a fragment of 900 bp of the rRNA 28S gene, involving D1, D2 and D3 domains, using the primers C1 and D3 (Jovelin & Justine 2001). For the COI gene, the primer LCO1490 was also used in combination with newly designed primers (COIH615: CGAAATYTNAATACNGCYTTTTTTGA and COIHNP: GGTGACCAAAAAATCAAAAYARATG) when PCR was negative with HCO2198. All PCRs were performed in 25 μl reaction mixture, containing 3 ng of DNA, 1× reaction buffer, 2.5 mM $MgCl_2$, 0.26 mM dNTP, 0.3 μM of each primer, 5% DMSO and 1.5 units of Q-Bio Taq (MPBiomedicals) for all genes. COI gene amplifications are performed according to Hebert *et al.* (2003); for 28S gene, the protocol consists of an initial denaturation step at 94 °C for 4′, followed by 30 cycles of denaturation at 94 °C for 30″, annealing at 52 °C and extension at 72 °C for 1′. The final extension was at 72 °C for 10′. PCR products were purified and sequenced at Genoscope facilities. In all cases, both directions were

sequenced using the Sanger method to confirm accuracy of each haplotype sequence. All sequences were submitted to GenBank.

Phylogenetic Analyses of DNA Sequences

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

Automatic Barcode Gap Discovery

Following the similarity criterion, genetic distances between specimens from the same species are supposed to be lower than genetic distances between specimens from different species, revealing a noncontinuous distribution (Hebert *et al.* 2003). This barcode gap, that is, the range of genetic distances not represented in the matrix of pairwise comparisons, can be used as a threshold offering primary species delimitation under the assumption that individuals within species are more similar than between species (genotyping clustering criterion—Mallet 1995). However, in some cases, this barcode gap does not correspond to a real discontinuity in the distribution, but only to a decrease in the distance frequency between the two modes of the distribution, that is, the intra- and interspecific distances overlap (Meier *et al.* 2008). This can be due to incomplete lineage sorting, where the COI sequence of a specimen is more similar to a sequence of another species than to a sequence of the same species (Rosenberg & Tao 2008) or to an underestimation of genetic distances because of homoplasy. The ABGD method aims at identifying a limit between the two distributions, even when they are overlapping. Starting from several a priori thresholds of genetic distances chosen by the user, ABGD will first compute the theoretical maximal limit of the intraspecific diversity (using a coalescent model) and then identify in the whole distribution of pairwise distances which gap, by definition superior to the maximal limit of the intraspecific diversity, potentially corresponds to the so-called Barcoding gap, that is, a potential limit between intra- and interspecific diversity. Inference of the limit and gap detection are then recursively applied to previously obtained groups to get finer partitions until there is no further partitioning. This method is described in detail in the study by Puillandre *et al.* (2011); we used the online version to analyse the data set (http://wwwabi.snv.jussieu.fr/public/abgd/). mega was used to build the distance matrix using a TN model (with $\alpha = 1.014$). ABGD default parameters were used, except the relative gap width (X) was set to 10 to avoid the capture of smaller local gaps.

General Mixed Yule Coalescent Model

The GMYC method, described by Pons *et al.* (2006) and Monaghan *et al.* (2009), is based on the difference in branching rates between speciation branching events (interspecific relationships) and coalescence branching events (intraspecific relationships) in a phylogenetic tree. This difference can be visualized as a switch between slow and fast rates of branching events in a lineage-through-time plot. The first step of the method is to compare the likelihood of the phylogenetic tree obtained with beast assuming a single branching process vs. the likelihood of the same tree assuming a switch of branching rates between the two types of events. If such a switch is detected, its position is determined and placed in the tree, allowing the delineation of PSHs. Two versions of the method are applied here: in the single-threshold method (Pons *et al.* 2006), the switch between speciation and coalescence events is supposed to be

unique; in the multiple-threshold method (Monaghan *et al.* 2009), each PSH defined with the single-threshold method is reanalysed one by one and can be divided into two, or fused with its sister-PSH, the hypotheses with the best likelihood being chosen. GMYC (multiple threshold) proposes alternate hypotheses of species delimitations, and in this way is also similar to the ABGD method. The GMYC method with both the single and multiple-threshold models (Monaghan *et al.* 2009), implemented in the SPLITS package for R, was applied to the COI tree obtained with BEAST.

Klee diagrams

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

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

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

Analyses of other characters and criteria

Phylogenetic analyses. As the efficacy of ABGD and GMYC may be limited by the variation of evolutionary rates in the different species, the statistical support (bootstraps and posterior probabilities) calculated using RAxML and BEAST for each PSH recognized with the COI gene was reported. Conflicts between the COI and 28S genes were also analysed by identifying which PSHs were sharing common 28S haplotypes and which ones were not monophyletic.

Genetic structure. When a PSH was present in at least two geographic populations, each of them including at least six specimens, the genetic structure was assessed amongst the different populations using Arlequin 3.1 (Excoffier *et al.* 2005). If a single PSH was present in several different geographic regions (amongst Taiwan, Philippines, Solomon Islands, Vanuatu, Chesterfield Islands and New Caledonia—see Material and methods, sampling) and in different localities within the geographic region, an AMOVA (with a 3000 permutations tests) was performed. If only one hierarchical level was involved (different localities within a single geographic region), F_{ST} between each pairs of populations was calculated. Network 4.5 (median-joining option) was used to construct haplotype networks.

Bathymetric distribution. Stations are characterized by starting and ending points that may correspond to different depths. This variation is sometimes up to 500 m, and such stations may actually cover highly different environments. To minimize the effect of this imprecision, the depth data for stations with a significant discrepancy between the starting and ending points (>20 m for shallow waters stations and >50 m for deep water stations) were not considered. To reduce the bias in depth ranges, all the PSHs with only one specimen were also not considered in the estimation of the bathymetrical distribution. It was then possible to conclude from the observation of the bathymetrical ranges of two PSHs if they were overlapping or not. To test the hypothesis that bathymetrical ranges could be underestimated by subsampling, a statistical test was designed to evaluate whether the bathymetrical range of a given PSH could be obtained by subsampling. Details of the test and interpretations of the results are provided in the Fig. S1 (Supporting information).

Morphological analyses. The features of the shells of all analysed specimens were examined by several specialists of the Turridae, Yuri Kantor, Baldomero Olivera and Alexander Sysoev. Examinations of the shell were

not performed 'blindly' but taking into account the molecular taxonomy analyses. The goal was thus to determine whether it was possible or not to find morphological differences between the PSHs, using shell characters as traditionally used in malacology. Considering available description in the malacological literature, each PSH was tentatively attributed a posteriori to available species name. When no name was available, PSHs were numbered with the genus to which they were attributed.

Dispersion abilities. For benthic organisms, such as marine snails, dispersal abilities occur mainly during the larval stage. Furthermore, the accretionary growth of the protoconch (i.e. the shell formed by the embryo and/or the veliger larvae before metamorphosis) can be used to infer the mode of development, which constitutes the best proxy for the dispersal ability of a gastropod species when no other data are available (Jablonski & Lutz 1980). A multispiral protoconch suggests that the larva fed in the water column (i.e. planctotrophic species) and is thus able to disperse over large distances. Conversely, the dispersion abilities are supposedly reduced for a nonplanctotrophic species (i.e. with a paucispiral protoconch), even if some nonplanctotrophic species have been shown to disperse over wide distances, for example, through passive larval transport (Parker & Tunnicliffe 1994). Dispersion abilities inferred from the protoconch morphology were used to discuss the validity of the PSHs. When not broken, the protoconch of the analysed specimens was in most cases multispiral (~3 whorls or more), indicating important dispersal capacities. However, the PSHs identified as *Lophiotoma indica* (Table 1) were possessing reduced protoconchs with only two whorls.

Turning PSHs into SSHs

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

Results

Turridae COI gene variability

A set of 1000 specimens of Turridae was sequenced for a 658-bp fragment of the barcoding COI gene; 648 haplotypes were found, with 477 polymorphic sites and a high haplotypic diversity (0.995). Genetic pairwise distances for COI gene were computed using three different substitution models: (a) the p-distances, (b) the K2P distances and (c) the Tamura–Nei (TN) distances. The distribution of genetic distances, whatever the substitution model used, displayed two modes separated by a rough gap ('barcode gap') between 0.02 and 0.04 (Fig. 2A, B). However, as shown in the Fig. 2B, the number of pairwise comparisons for genetic distances that corresponded to the barcode gap was lower for the TN distance than for the K2P distance and the p-distance. Consequently, the TN distances were used thereafter for ABGD analyses.

Turridae primary species hypotheses

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

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

Overall, the partitions obtained with the ABGD and GMYC are congruent. Amongst the 103 PSHs listed in Table 1, 73 were obtained both with ABGD and the two GMYC methods (Table 1, columns 2–5). In the

Table 1 List of PSHs, as defined with the ABGD (M = more inclusive partition and L = less inclusive partition) and GMYC (S = single-threshold and M = multiple-threshold) analyses of the COI gene. Number of specimens (N) and phylogenetic support are provided for both COI and 28S genes. Geographical, bathymetrical and morphological data are also provided

	ABGD		GMYC		COI			28S			Geography			Depth			
PSH	M	L	S	M	N	Support (ML/BA)		N	Support (ML/BA)	Haplotypes shared with	Region	Distribution	Genetic structure	Range (m)	index	Morphological ID	SSH
1	x	x	x	x	1	NA		0	NA		Sol			282–327		*G.* 1	*Gemmula* sp. 1
2	x	x	x	x	1	NA		1	NA		Phil			85–88		*G.* 2	*Gemmula* sp. 2
3	x	x	x	x	5	100/1		4	Pa/Pa		Phil			35–100			*Gemmula* sp. 3
4	x	x	x	x	3	NA/1		2	99/1		Ches			345–413		*Td. armillata*	*Turridrupa armillata*
5	x	x	x	x	2	100/1		2	NA		Van			0–49		*Td. neojubata*	*Turridrupa neojubata*
6	x	x	x	x	1	NA		1	NA		Phil			6–8		*Td. Bijubata*	*Turridrupa cf. bijubata* 1
7	x	x	x	x	1	NA		1	NA		Van			0–49		*Td. albofasciata*	*Turridrupa albofasciata*
8	x	x	x	x	4	100/1		1	NA		Phil, Van			20–110		*Td. astricta*	*Turridrupa astricta*
9	x	x	x	x	1	NA		1	NA		Van			0–49		*Td. bijubata*	*Turridrupa cf. bijubata* 2
10	x	x	x	x	4	96/1		4	Po/Po		Phil, Van			0–49			*Turridrupa cf. bijubata* 3
11	x	x	x	x	1	NA		1	NA		Phil			593		*P.* 1	*Ptychosyrinx* sp. 1
12	x	x	x	x	2	100/1		1	NA		Ches			372–404		*X. gemmuloides*	*Xenuroturris gemmuloides*
13	x	x	x	x	15	89/1		11	75/1		Ches, N-C, Phil, Sol			410–741		*G. unilineata*	*Gemmula unilineata*
14	x	x	x	x	4	99/1		2	93/1		Sol			897–1057		*P.* 2	*Ptychosyrinx* sp. 2
15	x	x	x	x	24	100/1		5	76/1		Ches, N-C			440–1150		*G.* 3	*Gemmula* sp. 4
16	x	x	x	x	5	99/1		0	NA		Van			503–636		*G.* 1	*Gemmula* sp. 5
17	x	x	x	x	8	100/1		2	—	18	Van			131–308		*L. indica*	*Lophiotoma cf. indica* 1
18	x	x	x	x	8	95/1		1	NA	17	Van, Sol			131–308			
19	x	x	x	x	2	NA/1		1	NA		Van			83–339			*Lophiotoma cf. indica* 2
20	x	x	x	x	1	NA		1	NA		Phil			155–160		*L. tayabaensis*	*Lophiotoma tayabaensis*
21	x	x	x	x	1	99/1	NA	1	Po/Po	NA	Sol	allopatric		150–160	NA	*L. friedrichbonho-efferi*	*Lophiotoma cf. friedrichbonhoefferi* 1
22		x	x	x	3		100/1	3		NA	Van			106–148			*Lophiotoma cf. friedrichbonhoefferi* 2
23	x	x	x	x	7	94/1		4	Pa/Pa		Phil			72–139		*L. bisaya*	*Lophiotoma bisaya*
24	x	x	x	x	7	95/1		6	—	26	Sol, Van			83–160		*L. indica*	*Lophiotoma cf. indica* 3
25	x	x	x	x	2	76/1	99/1	1	—	NA 26	Phil	Sympatric*		42–44	25		
26		x	x	x	6		99/0.96	5		— 24, 25	Phil			42–79			
27	x	x	x	x	6	100/1		6	—	28, 29, 30, 31	Phil			219–318		*L. sikatunai*	*Lophiotoma sikatunai*
28	x	x	x	x	2	99/1	Pa/0.35	2	—	— 27, 29, 30	Ches	Sympatric*		310–400	n.s.	*L. unedo*	*Lophiotoma unedo*
29				x	6		60/0.82	4		— 27, 28, 30	Ches			267–400			
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	54/1	96/1	3	—	— 36	Phil	Sympatric*		229–400	33, 35	*L. panglaoensis*	*Lophiotoma cf. panglaoensis* 1
33			x	x	10		95/1	1		NA	Phil			182–346			*Lophiotoma cf. panglaoensis* 2
34			x	x	6		82/1	0		NA	Phil, Sol			173–400			*Lophiotoma cf. panglaoensis* 3
35			x	x	10		97/1	4		80/0.97	Sol, Van			131–600			*Lophiotoma cf. panglaoensis* 4

Table 1 *Continued*

	ABGD		GMYC		COI			28S			Geography			Depth			
PSH	M	L	S	M	N	Support (ML/BA)		N	Support (ML/BA)		Haplotypes shared with	Region	Distribution	Genetic structure	Range (m)	index	Morphological ID	SSH
36	x	x	x	x	94	99/1		75	Po/Po		32	Sol, Van		COI & 28S	350–659		*L. indica*	*Lophiotoma cf. indica* 4
37	x	x	x	x	1	NA		1	NA			Phil			8–22		*T. babylonia*	*Turris babylonia*
38	x	x	x	x	1	NA		1	NA			Van			112–148		*T. spectabilis*	*Turris spectabilis*
39	x	x	x	x	15	100/1		15	58/1			Van			0–55		*T. garnonsii*	*Turris garnonsii*
40	x	x	x	x	1	NA		1	NA			Van			0		*X. legitima*	*Xenuroturris legitima*
41	x	x	x	x	2	NA/1		2	NA			Van			20		*I. musivum*	*Iotyrris musivum*
42	x	x	x	x	3	86/1		2	55/0.90			Van			0–49		*I. cingulifera*	*Iotyrris cingulifera*
43	x	x	x	x	4	98/1		4	Pa/Pa			Van			16–20		*I. devoizei*	*Iotyrris devoizei*
44	x	x	x	x	1	NA		1	NA			Phil			85–88		(juvenile)	*Gemmula* sp. 6
45	x	x	x	x	1	NA		0	NA			Phil			120		*G.* 4	*Gemmula* sp. 7
46	x	x	x	x	2	NA/		2	68/0.99			Van			0–49		*G. lisajoni*	*Gemmula lisajoni*
47	x	x	x	x	4	100/1		3	—		60, 61	Van			0–49		*L. albina*	*Lophiotoma albina*
48	x	x	x	x	1	NA		1	NA			Van			266–281		*G.* 1	*Gemmula* sp. 8
49	x	x	x	x	1	NA		0	NA			Sol			173–379		*G.* 5	*Gemmula* sp. 9
50	x	x	x	x	1	NA		0	NA			Sol			286–423		*G.* 6	*Gemmula* sp. 10
51	x	x	x	x	1	98/1	NA	0	NA	NA		N-C	Sympatric*		386–391	NA		*Gemmula* sp. 11
52				x	2		60/0.99	0		NA		N-C			386–391			
53	x	x	x	x	8	93/1		6	90/1			Phil, Sol, Van			11–176		*G. monilifera*	*Gemmula cf. monilifera* 1
54	x	x	x	x	9	100/1		9	80/1			Van			0–118			*Gemmula cf. monilifera* 2
55	x	x	x	x	8	100/1	98/1	7	92/1	—	56	Van	Sympatric*		0–99	NA		*Gemmula cf. monilifera* 3
56				x	1		NA	1		—	55	Van			0–49			
57	x	x	x	x	1	NA		1	NA			Phil			2–3		*G. hombroni*	*Gemmula cf. hombroni* 1
58	x	x	x	x	1	NA		1	NA			Phil			85–88			*Gemmula cf. hombroni* 2
59	x	x	x	x	18	100/1		16	93/1			Van			0–99			*Gemmula cf. hombroni* 3
60	x	x	x	x	5	100/1	82/1	2	—	—	47, 61	Phil, Sol	Sympatric		410–480	60, 61	*G.* 1	*Gemmula* sp. 12
61		x	x	x	19		—/1	17		—	47, 60	Sol, Van		COI & 28S	503–773			
62	x	x	x	x	1	NA		0	NA			Van			184–271		*G.* 7	*Gemmula* sp. 13
63	x	x	x	x	8	99/1		2	90/1			Sol			150–176			*Gemmula* sp. 14
64	x	x	x	x	20	100/1		18	Pa/Pa			Phil			98–356		*G.* 8	*Gemmula* sp. 15
65	x	x	x	x	2	100/1		1	NA		66	Phil			11–20		*L. jickelli*	*Lophiotoma jickelli*
66	x	x	x	x	8	93/1		8	—		65	Van			0–58			
67	x	x	x	x	6	100/1		6	Pa/0.86			Phil			0–3		*L. polytropa*	*Lophiotoma polytropa*
68	x	x	x	x	11	100/1	53/1	8	61/0.99	54/0.97		Van	Sympatric*		0–49	NA	*L. abbreviata*	*Lophiotoma abbreviata*
69				x	23		96/1	19		Pa/Pa		Van, Phil			0–49		*L. brevicaudata*	*Lophiotoma brevicaudata*
70	x	x	x	x	3	100/1		2	91/1			Van			0–49		*L. ruthveniana*	*Lophiotoma ruthveniana*
71	x	x	x	x	14	100/1		13	79/1			Van			0–49		*L. picturata*	*Lophiotoma picturata*
72	x	x	x	x	1	97/1	NA	1	73/1	NA	73	Phil	Allopatric		2–15	NA	*L. acuta*	*Lophiotoma cf. acuta* 1
73		x		x	2		100/0.99	2		NA	72	Van			0–99			
74	x	x	x	x	101	100/1		91	81/1			Van, Phil		COI & 28S	0–99			*Lophiotoma cf. acuta* 2
75	x	x	x	x	1	NA		0	NA			N-C			418–421		*G. rarimaculata*	*Gemmula cf. rarimaculata* 1

Table 1 *Continued*

	ABGD		GMYC		COI			28S				Geography			Depth			
PSH	M	L	S	M	N	Support (ML/BA)		N	Support (ML/BA)		Haplotypes shared with	Region	Distribution	Genetic structure	Range (m)	index	Morphological ID	SSH
76	x	x	x	x	1	NA		0	NA			Phil			97–120		*G. monilifera*	*Gemmula cf. monilifera* 4
77	x	x	x	x	3	100/1		1	NA			N-C, Ches			175–370		*G. rarimaculata*	*Gemmula cf. rarimaculata* 2
78	x	x	x	x	4	100/1		4	91/1			Van, Phil			62–118		*G. hastula*	*Gemmula hastula*
79	x	x	x	x	77	100/1		64	53/1			Phil, Van		COI & 28S	35–196		*G. sogodensis*	*Gemmula cf. sogodensis* 1
80	x	x	x	x	1	NA		1	NA			Ches			330–331		*G.* 1	*Gemmula* sp. 16
81	x	x	x	x	1	NA		1	NA			Ches			627–741		*G.* 9	*Gemmula* sp. 17
82	x	x	x	x	23	97/1	—/0.93	16	NA	74/1		Van	Allopatric	COI	323–659	NA	*G.* 10	*Gemmula* sp. 18
83				x	1		NA	0		NA		Sol			381–422			
84	x	x	x	x	12	100/1	—/0.87	11	56/1	—	85	Phil, Sol, Van	Sympatric*		318–659	n.s.	*G.* 1	*Gemmula* sp. 19
85				x	30		84/0.97	24		—	84	Ches, Sol, Van		COI	345–636			
86	x	x	x	x	2	NA/1		2	—		95	Van			350–400		*G.* 11	*Gemmula* sp. 20
87	x	x	x	x	1	NA		1	NA			Phil			342–358		*G.* 5	*Gemmula* sp. 21
88	x	x	x	x	1	NA		0	NA			Sol			630–836		*G.* 12	*Gemmula* sp. 22
89	x	x	x	x	1	Po/Po	NA	1	NA	NA		Ches	Allopatric		568–570	NA	*G.* 13	*Gemmula* sp. 23
90			x	x	6		92/1	4		70/1		Phil, Sol, Van			416–786		*G.* 14	
91	x	x	x	x	2	92/1	59/0.98	0	NA	NA		Sol	Allopatric		484–836	NA	*G.* 15	*Gemmula* sp. 24
92				x	1		NA	0		NA		Ches			485–500			
93	x	x	x	x	1	69/0.96	NA	0	NA	NA		Ches	Allopatric		490–500	NA	*G.* 16	*Gemmula* sp. 25
94			x		3		100/1	1		NA		Phil			269–378		*G.* 5	
95	x	x	x	x	7	94/1		5	—		86	Van			350–600			*Gemmula* sp. 26
96	x	x	x	x	1	NA		1	NA			Phil			422–431		(juvenile)	*Gemmula* sp. 27
97	x	x	x	x	30	98/1		17	Pa/Pa			Phil		COI & 28S	219–1762		*G. diomedea*	*Gemmula diomedea*
98	x	x	x	x	9	100/1		6	NA			Phil, Sol			65–160		*G. speciosa*	*Gemmula speciosa*
99	x	x	x	x	10	96/1		9	NA			Phil			85–137		*G. kieneri*	*Gemmula kieneri*
100	x	x	x	x	1	NA		0	NA			Taiwan			157–275		*G. cosmoi*	*Gemmula cf. cosmoi* 1
101	x	x	x	x	7	96/1		2	Po/Po			Sol			300–430		*G. martini*	*Gemmula martini*
102	x	x	x	x	24	98/1		16	91/1			Van			131–444		*G. cosmoi*	*Gemmula cf. cosmoi* 2
103	x	x	x	x	71	95/1		62	82/1			Phil			72–361		*G. sogodensis*	*Gemmula cf. sogodensis* 2

NA, nonapplicable (one or no specimen or one or no haplotype); Pa, paraphyletic; Po, polyphyletic; Phil., Philippines; Sol., Solomon Islands; Ches., Chesterfield Islands; Van., Vanuatu; N-C, New Caledonia; Tai., Taiwan.
Morphological identification: *G. = Gemmula*; *P. = Ptychosyrinx*; *L. = Lophiotoma*; *T. = Turris*; *Td. = Turridrupa*; *I. = Iotyrris*; *X. = Xenuroturris*.
*Indicates that at least one specimen from each of the corresponding PSHs was collected at the same station. The 'Genetic structure' column lists the PSHs for which COI and 28S structure were tested (Table S1, Supporting information).The depth range index refers to the statistical tests explain in the Material and methods section (a PSH number indicates that the test is significant for this PSH; n.s., not significant).

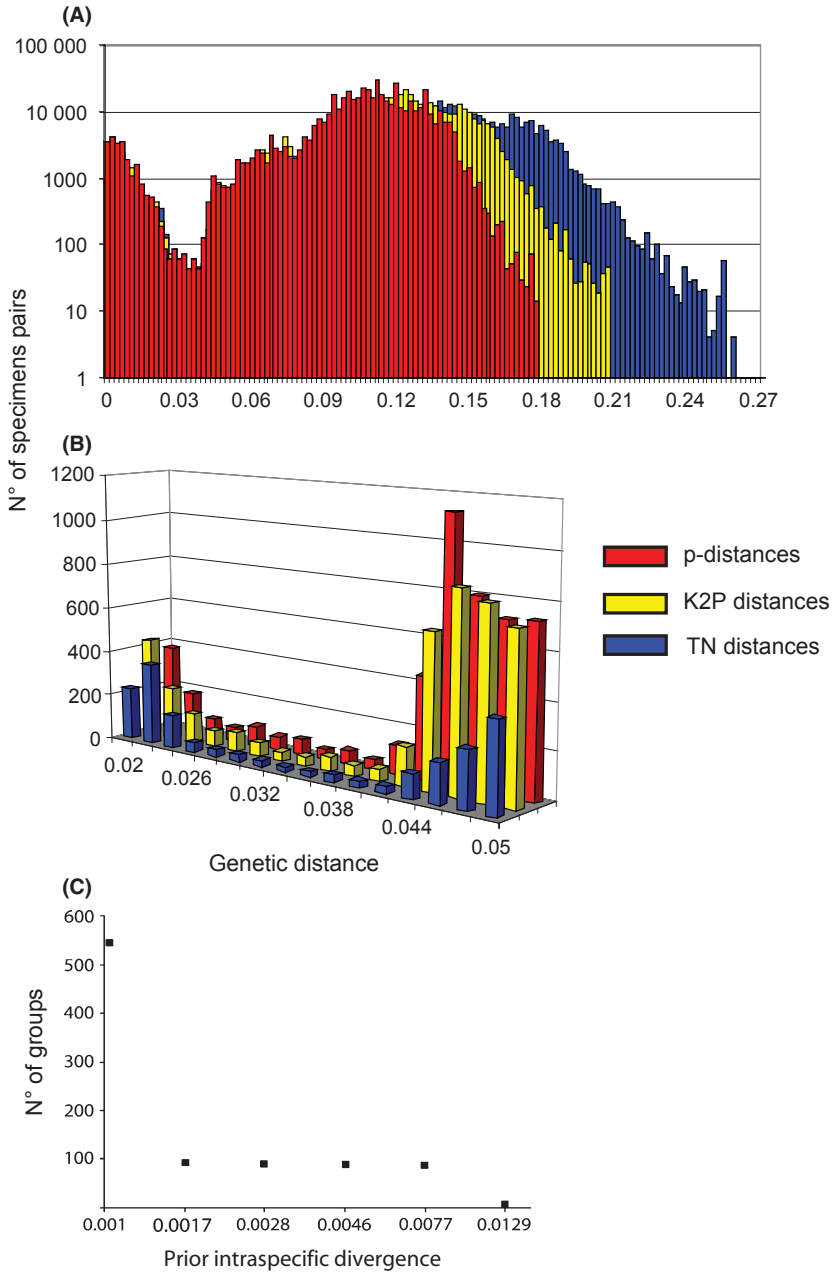

Fig. 2 Pairwise distribution for the COI gene and ABGD results. (A) Distributions of p distance, K2P distances and TN distances between each pair of specimens for the COI gene. (B) Same results, but focusing on the barcode gap zone. (C) ABGD results, with the number of PSHs obtained for each prior intraspecific divergence.

phylogenetic tree of the Fig. 3A, each of the PSHs listed in Table 1 is represented by a single branch.

Visualization of the PSHs using klee diagrams

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

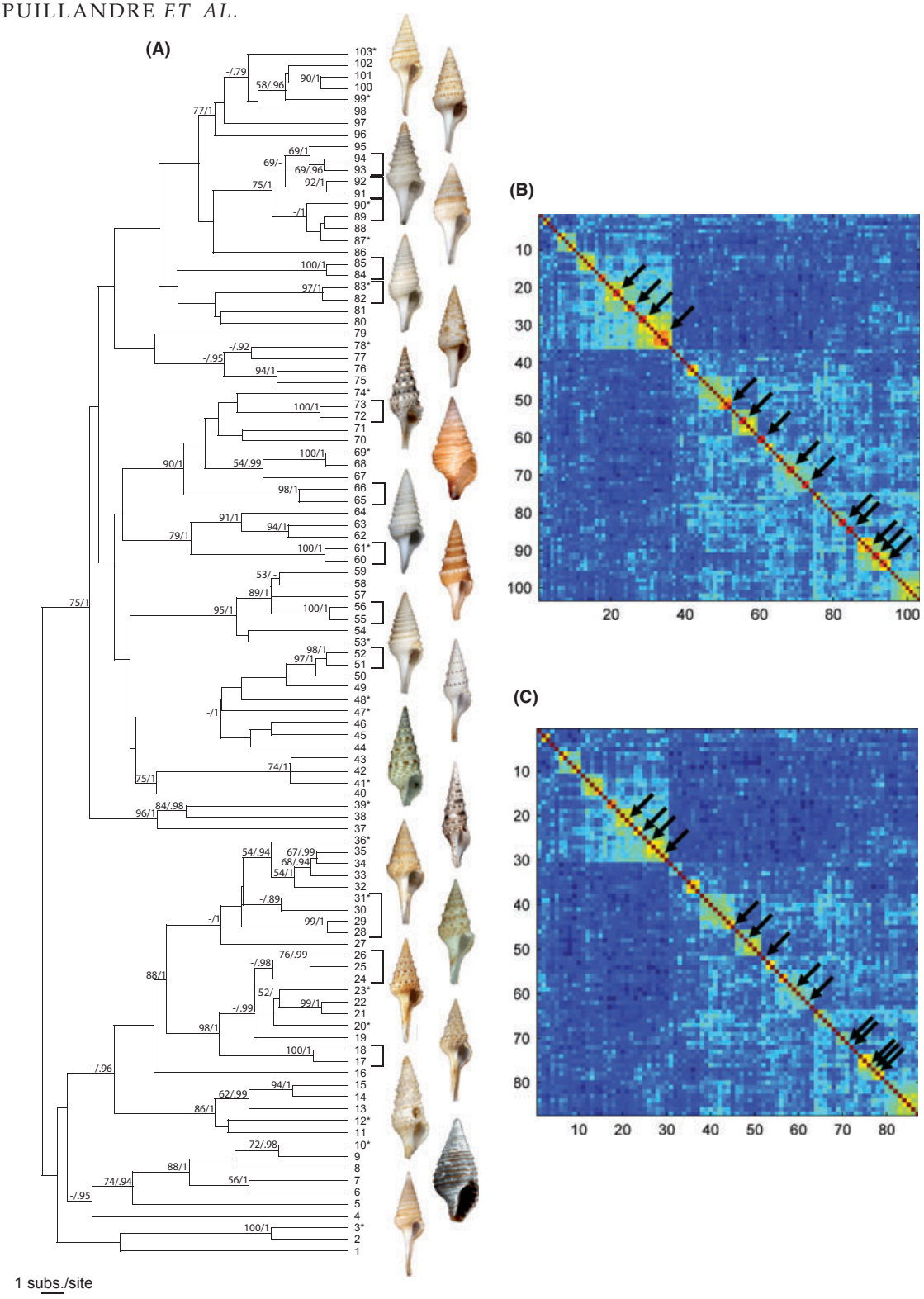

Fig. 3 COI gene results. (A) Bayesian COI gene tree with posterior probabilities (>0.8) and bootstraps (>50) indicated next to each node. The 103 PSHs listed in Table 1 (first column) are represented each by a single branch (the intra-PSH trees are not shown). Black brackets indicate the PSHs that were subsequently grouped into one SSH. *PSHs with a shell illustration. (B) Klee diagrams for the COI gene showing the correlations amongst indicator vectors for the less inclusive data set corresponding to the 103 PSHs provided by the multiple-threshold GMYC (the black arrows point to the groups of PSHs recognized as a single PSH with ABGD); gradations of red and yellow colour in the Klee diagram indicate areas of conflict. (C) For the most inclusive data set corresponding to the 87 PSHs provided by ABGD (the black arrows point to PSHs that are divided into several PSHs by the multiple-threshold GMYC method).

hypothesis. All other indicator vector analyses of ABGD and GMYC PSHs appear to be equally likely.

Phylogenetic analyses and 28S gene

Most of the PSHs defined with the COI gene, 86 of the 103 listed in Table 1, representing 708 specimens, were successfully sequenced for the 28S gene (Table 1). A 28S fragment of 908 bp after alignment displayed 228 haplotypes with 359 polymorphic sites and a haplotypic diversity of 0.979. Bootstraps and posterior probabilities are given for each PSH and each gene, COI and 28S, in Table 1. All the PSHs that included more than one specimen corresponded to highly supported clades with the COI gene (Bootstraps > 75, PP > 0.95), except in 10 cases: PSHs 28, 29, 32 + 33 + 34 + 35, 52, 61, 68, 84, 89 + 90, 91, 93 + 94. Each of these ten cases corresponded to a pair of PSHs that were alternatively recognized as a single PSH or two different PSHs with ABGD and GMYC. In the less inclusive hypothesis, when one of the two PSHs corresponded to a weakly supported clade (e.g. PSHs 28 and 29), the alternate most inclusive hypothesis systematically corresponded to a highly supported clade (PSHs 28 + 29) (Table 1).

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

Geographic distribution and genetic structure

Amongst the 103 PSHs, 80 PSHs were restricted to a single geographic region (Taiwan, Philippines, Solomon Islands, Vanuatu, Chesterfield Islands or New Caledonia), 17 in two different regions and 6 in three or more. Amongst the 14 pairs or quadruplets of PSHs recognized either as a single PSH or as two or four different PSHs depending on the method, six of them were collected in different geographic regions and were thus considered allopatric and eight were collected in at least one common area (at the same station for seven of them), and are reported as sympatric (Table 1, geographic distribution column). The genetic structure amongst different sampling sites in a single PSH was calculated for eight different PSHs with the COI gene and for five with the 28S gene (Table 1, genetic structure column). All the F_{ST} values are very low, and only one is significant (Table S1, Supporting information).

Bathymetric distribution

Amongst the 14 pairs or quadruplets of alternative PSHs, nine included at least one PSH with only one specimen and were not analysed further. Another pair included PSHs with strictly nonoverlapping bathymetric ranges (PSH 60–61), two pairs corresponded to subsamples of the association of two PSHs (28–29 and 84–85), and one pair included one PSH with bathymetric preferences (25–26). Finally, the quadruplet included two PSHs with bathymetric preferences (33–35), and two considered as a subsample of the association of the four PSHs (32, 34).

Shell morphology and attribution to species names

The shells of the specimens included in each PSH were examined. Based on the shell morphology, the PSHs were then tentatively assigned to a species name available in the literature (Table 1, morphological ID). For 28 PSHs, shells of specimens corresponded to a unique morph, and it was possible to link each of them to a unique species name; conversely, 11 species names corresponded to shell features shared by several PSHs (39 PSHs affected). Two PSHs represented by a single juvenile specimen might not be attributed to a morphospecies attached to a species name. For 11 PSHs, shells corresponded to distinct morphospecies for which no species names were available, and they were thus associated to a genus name and to a morphospecies number within each genus (*Gemmula* 3, 4, 8, 9, 11–14, 16 and *Ptychosyrinx* 1–2). Finally, the 23 remaining PSHs corresponded to three different morphospecies, not attributed to a species name: *Gemmula* 1 (PSHs 1, 16, 48, 60, 61, 80, 84, 85), 2 (PSHs 2, 3), 5 (PSHs 49, 87, 94, 95), 6 (PSHs 50–52), 7 (PSHs 62, 63), 10 (PSHs 82, 83), 15 (PSHs 90, 91).

Consolidating secondary species hypotheses

Primary species hypotheses drawn using ABGD and GMYC were converted to SSHs according to the workflow presented in step 4 of Fig. 1, and the criteria are listed in Table 1. Amongst the 103 PSHs listed in Table 1, 21 found monophyletic with the 28S gene, and 38 with unique 28S haplotypes, were converted to 59 SSHs. Of the 38 PSHs with unique 28S haplotypes, 24 were represented by specimens with identical

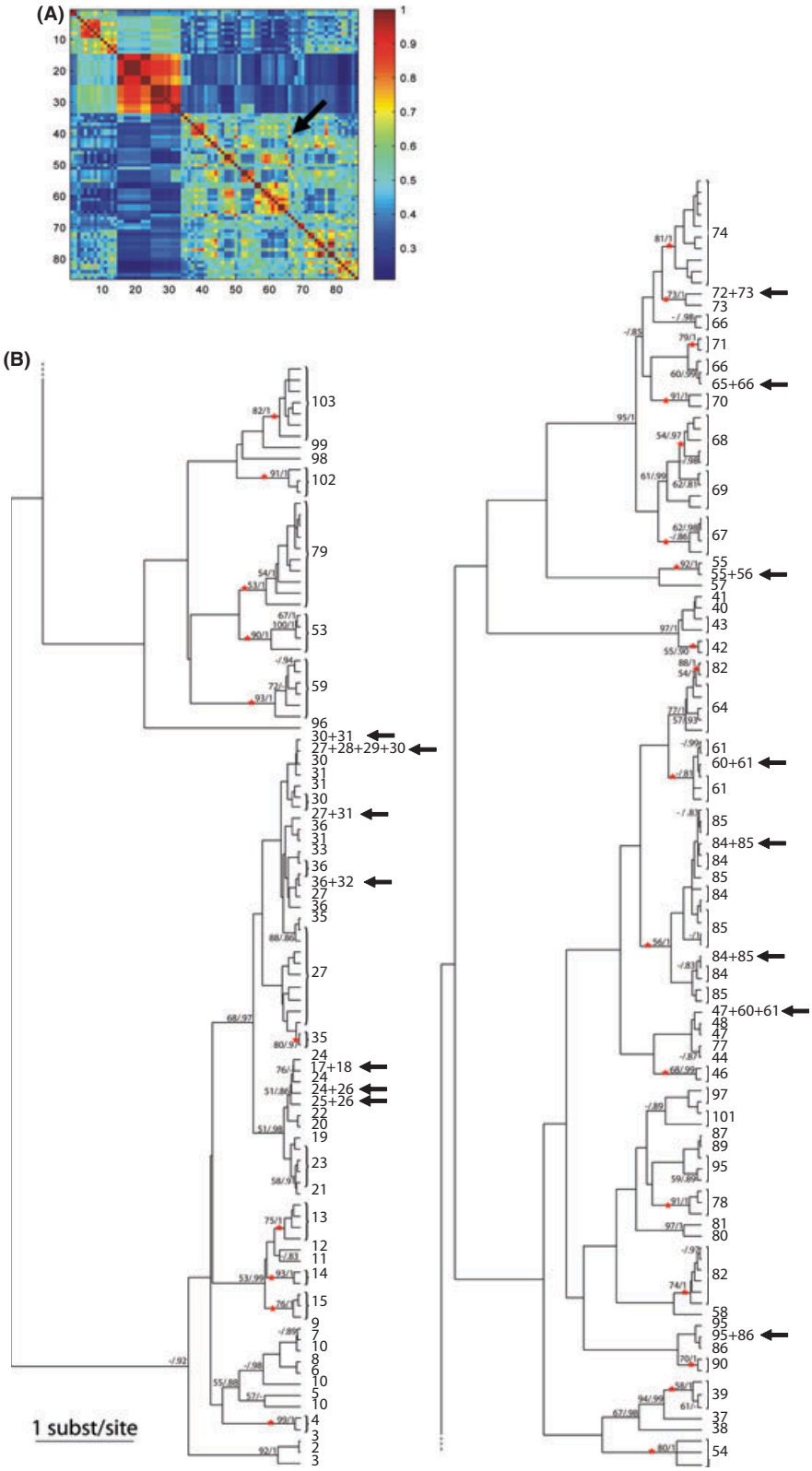

Fig. 4 28S gene results. (A) Klee diagrams for the 28S gene showing the correlations amongst indicator vectors for the 86 PSHs sequenced for this gene. (B) Phylogenetic tree obtained with the 208 28S haplotypes (Bayesian analysis). Posterior probabilities (>0.8) and bootstraps (>50) are reported for each node. Numbers at the tip of the branches refer to the PSH numbers (Table 1). Red star: monophyletic PSHs. Black arrow: haplotype shared by several PSHs.

sequences, i.e. a single haplotype was included in the phylogenetic analysis, preventing any test of the 28S monophyly for the corresponding PSH. Twenty PSHs were not sequenced for the 28S gene. Of this group, 10 PSHs were converted to SSHs after analysis of other evidence (see Table 1 for details). Following a conservative approach, the remaining 11 PSHs without 28S sequences were converted to five SSHs, as there was no comparative evidence to support additional SSH assignments. Finally, 24 PSHs sharing 28S haplotypes were converted to 13 different SSHs following guidelines in step 4 of Fig. 1. An example for which all the PSHs characters and criteria are congruent is shown in Fig. 5.

There are only four cases where the PSHs were in agreement with ABGD and GMYC analyses, but were not directly converted to SSHs. For example, the PSHs 65 and 66 were considered to correspond to a single SSH, as they shared 28S haplotypes and they were not distinguished morphologically. Similarly, 28S variability, geographical and bathymetrical ranges, dispersal abilities and morphological analysis were decisive in discussing the 14 pairs, or quadruplets, of PSHs alternatively recognized either as a single PSH or as 2–4 different PSHs by the ABGD and GMYC analyses. They were turned into 21 SSHs (see details in Table 1). Three examples, one species, two species or inconclusive species, corresponding to three different conclusions that can be obtained following step 4 of Fig. 1, are detailed as follows: (i) One species: PSHs 25 and 26 shared 28S haplotypes, were both found in the same geographic area, and in the same station for some of them. PSH 25 displayed bathymetric preferences and had weakly dispersive larvae. PSHs 25 and 26 were interpreted as a single SSH along with PSH 24, and the differences found in the COI were thought to correspond to intraspecific structure linked to the depth. Four other PSHs pairs (28–29, 55–56, 72–73 and 84–85) were similarly turned each in a single SSH. (ii) Two species: PSH 21 and 22 were interpreted as two different species as they did not share 28S haplotypes and were found in two different geographic regions (Vanuatu and Solomons) without obvious barrier between them, and their larvae are highly dispersive. Additionally, bathymetric ranges for PSHs 21 and 22 did not overlap, which can be seen as ecological differences between the two species. Two other PSHs pairs (32–35 and 68–69) were similarly converted to two or four SSHs. (iii) Inconclusive species: Following a conservative approach, PSHs 51–52 was considered as a single SSH as the supporting evidence was inconclusive. The 28S gene was not sequenced for these specimens, and they were found in the same geographic region, without bathymetric differences (see Material and methods). Five other PSHs pairs (60–61, 82–83, 89–90, 91–92 and 93–94) were similarly converted to a single SSH following a conservative inclusive approach.

Discussion

Illustrated here is a semi-automated integrative taxonomy strategy that uses a single-gene approach derived from 'DNA barcoding' to determine species as hypotheses that are consolidated using several additional lines of evidences through a process of modification and validation. The single-gene data set analysed with bioinformatics species delimitation tools, such as ABGD and GYMC, is combined with biological (life-history traits), morphological (shell characters) and ecological (bathymetric distribution, geographic barriers) data. This approach constitutes an efficient way for proposing PSHs for hyperdiverse groups especially when morphological characters are known to be problematic.

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

Key components of the integrative strategy presented here are a sampling design that covers both taxon and

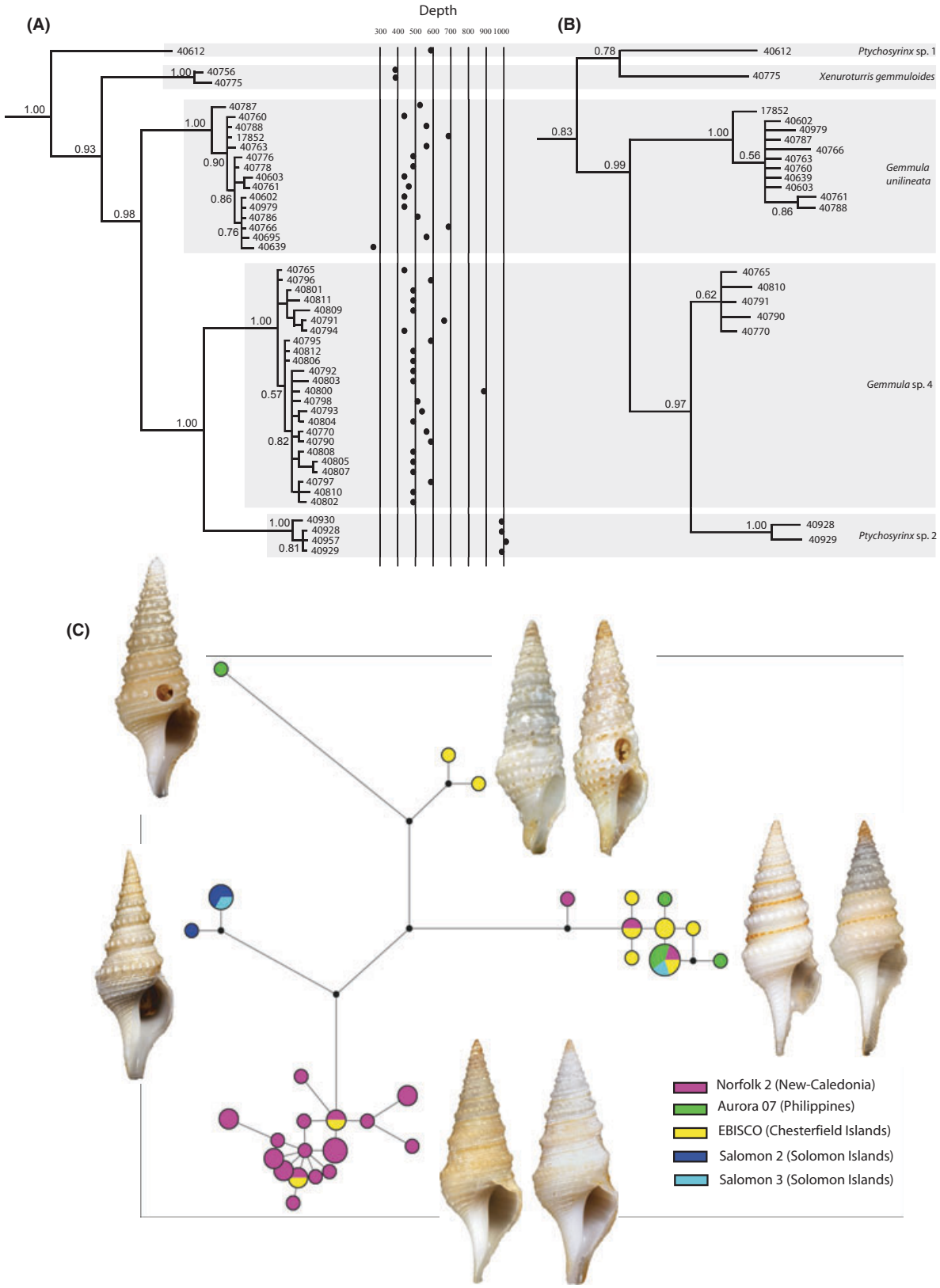

Fig. 5 Example of congruent SSHs. (A) COI tree for the SSH *Ptychosyrinx* sp. 1, *X. gemmuloides*, *G. unilineata*, *Gemmula* sp. 4 and *Ptychosyrinx* sp. 2., corresponding depth of collection for each specimen is given. (B) 28S tree for the corresponding SSHs. (C) COI haplotype network for the same SSHs. Some specimens are illustrated for each SSH by their shells.

intraspecific diversity and visualization of the PSHs proposed with ABGD and GMYC using the recently developed method of Klee diagram. For large data sets such as the 1000 specimens used for the Turridae, Klee diagrams facilitate the visualization of PSH correlation patterns and rapidly identify borderline cases.

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

Conflicting cases were detected when one PSH defined by ABGD or GMYC is split in two by the alternate method (Fig. 3). Indicator vector analysis suggests in cases where conflict is detected that neither ABGD nor GMYC can be consistently preferred (Fig. 3B, C, black arrows). In conflicting cases, ABGD and GYMC do not propose a unique threshold, but rather a range of possible partitions amongst which some PSHs are different. ABGD, by testing several a priori thresholds and by applying a recursive approach, and GMYC, with both the single- and multiple-threshold methods, are able to consider the heterogeneity amongst lineages of the rates of speciation and of coalescence that result in an overlapping distribution of the pairwise genetic distances (Fig. 2A, B). For eight of nine pairs of conflicting hypotheses obtained with the single and multiple GMYC methods, the corroboration process turned into SSHs the PSHs proposed by the single-threshold method. For the 14 conflicting cases between ABGD and GMYC methods, the final SSH were defined as a PSH only by ABGD in two cases and only by GMYC in one case (Table 1). These findings suggest ABGD and GYMC are complementary and should be used together to increase the overall robustness of the final partition to determine the set of PSHs fixed as SSHs. Compared to GMYC, ABGD may be considered as less refined in regards to underlying evolutionary processes; however, GMYC requires prior construction of a tree that must be ultrametric, which does not necessarily reflect the real divergence between species. Alternatively, ABGD is based solely on genetic distances calculated between each pair of COI sequences, allowing for the exploration of a range of thresholds and management of the heterogeneity of evolution rates. Furthermore, the short calculation time of the ABGD method, a few seconds vs several weeks to obtain a BEAST tree for the GMYC method (following the method described in Monaghan *et al.* 2009), allows a rapid comparison of different models of evolution for each data set. Thus, ABGD would be easier to apply to very large data sets. Finally, both ABGD and GYMC are problematic when species are represented with only a few specimens (Lohse 2009; Puillandre *et al.* 2011), and as underlined in the results, PSHs with less than three specimens are generally difficult to discuss with other characters and criteria. A large proportion of rare species represented by a low number of specimens is a common pattern, especially for marine gastropods (Castelin *et al.* 2011). SSH proposed for these samples are more susceptible to modification if new specimens are collected in the future.

In addition to analyses via ABGD and GMYC, and to corroborate the hypotheses drawn from single genes with criteria, we examined patterns of diversity in the same sample set to give an evolutionary meaning to the proposed hypotheses. For example, testing the reciprocal monophyly of the PSHs on several genes indicates that the proposed species represent a unique evolutionary lineage. The SSHs proposed here all correspond to PSHs for which several characters and/or lines of evidence were congruent. However, not all SSHs are equally supported. For example, SSHs based on the lack of shared haplotypes on the 28S gene should be considered more carefully than SSHs confirmed by reciprocal monophyly with the nuclear gene. Bathymetric and geographic distributions, in association with dispersal abilities, are not species delimitation criteria in themselves, but act as additional evidence towards one or another hypothesis further providing clues about the speciation process (Hyde *et al.* 2008). In Fig. 1, several patterns are interpreted as evidence for one or two species. For example, highly dispersive larvae for two PSHs in the absence of geographical barrier, or in the presence of bathymetrical differences when geographical ranges are overlapping, can be interpreted as evidence for the presence of two species, as large dispersal abilities would result in shared haplotypes if only one species was involved. However, in several cases, results are inconclusive. For example, when the two PSHs have nondispersive larvae and discontinuous geographic ranges without barrier, the differences observed with the COI could be only due to geographic structuring. Here, the decision depends on the taxonomist's choice, and we followed a conservative approach by considering only one species, even if two deep conspecific lineages (Padial *et al.* 2010), potentially corresponding to incipient species, were

revealed by the COI gene analysis. Finally, morphology, although a highly valuable character in numerous cases (Holynski 2010), is applied in the final step, where morphological differences are seen as additional evidence for the existence of different species, but knowing also that different species may share a highly similar morphology.

The borderline cases detected in the primary stage of the integrative protocol corresponded mostly to cases of recent divergence. These cases are of particular interest for understanding speciation processes. Using characters and criteria that are directly issued from evolutionary-based species criteria, such as phylogeny, reproductive isolation, phenetic divergence (Samadi & Barberousse 2006) is not only useful to propose more robust hypotheses, but also to understand what induced and drove the speciation process (Padial *et al.* 2010). For example, the distributions of most of the species illustrated in Fig. 5 are restricted to one or two geographic regions, suggesting allopatric speciation. However, *Ptychosyrinx* sp. 2 and *Gemmula* sp. 4 are also distinguished by their bathymetric ranges and could have diverged under a parapatric model. Identifying the different factors promoting the speciation event, either linked to geographical isolation and genetic drift or linked to ecological differentiation and selective forces, requires model-based studies (Crow *et al.* 2010). However, in all cases, delimitating species with clear, robust and reproducible methods remains the first step.

PSH proposed by ABGD and GMYC of species represented by large numbers of specimens or, conversely, by only a few specimens, that is, uneven sampling, is an area of concern with the integrated method described. To test the effects of uneven sampling, the number of specimens in the two largest PSHs, PSH 36 with 94 specimens and PSH 74 with 101 specimens, was reduced to 10. The PSHs defined with ABGD were unchanged as a result of the artificial minimization. Additionally, two data sets with three species each were simulated, where in the first data set, each species was represented by 37 specimens and in the second, they were represented by 1, 10 and 100 specimens, respectively. The Yule model was used to simulate the species tree, and a Kingman model in which genes from different species cannot coalesce was used for the gene tree until they reach the common ancestral species. A theta of 10 was used for the mutations, with a sequence length of 1000 to obtain an average of 1% divergence between two sequences of the same species. Under these conditions, ABGD (with a prior of 0.01) detected a mean of 3.67 species when the sampling is even, and 3.46 species when the sampling is uneven amongst the 1000 runs. The difference was subtle, but significant, with a P-value $<10^{-4}$. Similar simulations performed with GMYC suggested that GMYC may overestimate the number of species in even and uneven sampling, and further detailed exploration of the effect of uneven sampling on species delimitation with both ABGD and GMYC is clearly needed. The simulation results suggest that the Turridae PSH delimitation could be slightly influenced by the evenness of the sampling and could explain why in several cases ABGD and GMYC underestimate or overestimate, respectively, the number of PSHs compared to the number of SSHs retained at the end of the analytical process (Table 1). Increasing the sampling effort to reduce differences in specimen numbers between PSH would reduce the potential biases witnessed in ABGD and GMYC. However, this recommendation is often hardly applicable as rare species are usually present in empirical studies.

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

Acknowledgements

Key material for molecular studies originated from several expeditions to the Philippines and Vanuatu, funded via a consortium of agencies, including the Total Foundation, the French Ministry of Foreign Affairs, the Richard Lounsbery Foundation, the Philippines Bureau of Fisheries and Aquatic Research (BFAR) and the Niarchos Foundation; the Coral Sea and Solomon Islands took place on board R/V *Alis* from the Institut de Recherche pour le Développement (IRD). P. Bouchet and B Richer de Forges were P.I. for these cruises/expeditions. E. Strong, B. Buge and Y. Kantor are thanked for their role in molecular sampling during these expeditions. The phylogenetic analyses were performed on the vital-it clusters (http://www.vital-it.ch). The authors thank G. Achaz, A. Lambert and S. Brouillet for the development and the implementation of the ABGD method, A. Sysoev, B. M. Olivera and Y. Kantor for the taxonomic identifications, T. Barraclough for his assistance with the GMYC method, and P. Bouchet and G. Paulay for constructive comments on the manuscript. This work was supported by the 'Consortium National de Recherche en Génomique' and the 'Service de Systématique Moléculaire' (UMS 2700 CNRS-MNHN). It is part of the agreement no. 2005/67 between the Genoscope and the Muséum National d'Histoire Naturelle on the project 'Macrophylogeny of life' directed by Guillaume Lecointre.

This project is partially funded by an Alfred P. Sloan foundation (B2010-37), NSF (0940108) and NIH-NIGMS (GM088096) grants to M. Holford. Molecular data were obtained through the DNA-barcoding workflow established at the MNHN thanks to the MARBOL grant from the Alfred P. Sloan foundation, PI D. Steinke, Co-PI P. Bouchet and S. Samadi.

References

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

Barberousse A, Samadi S (2010) Species from Darwin onward. *Integrative Zoology*, **5**, 187–197.

Boissin E, FÉRal JP, Chenuil A (2008) Defining reproductively isolated units in a cryptic and syntopic species complex using mitochondrial and nuclear markers: the brooding brittle star, *Amphipholis squamata* (Ophiuroidea). *Molecular Ecology*, **17**, 1732–1744.

Bouchet P, Lozouet P, Sysoev AV (2009) An inordinate fondness for turrids. *Deep-Sea Research. Part II, Topical Studies in Oceanography*, **56**, 1724–1731.

Bouchet P, Kantor Y, Sysoev A, Puillandre N (2011) A new operational classification of the Conoidea (Gastropoda). *Journal of Molluscan Studies*, **77**, 273–308.

Castelin M, Puillandre N, Lozouet P, Sysoev A, Richer de Forges B, Samadi S (2011) Molluskan species richness and endemism on New Caledonian seamounts: are they enhanced compared to adjacent slopes? *Deep-Sea Research I*, **58**, 637–646.

Chase MW, Salamin N, Wilkinson M *et al.* (2005) Land plants and DNA barcodes: short-term and long-term goals. *Philosophical Transactions of the Royal Society B*, **360**, 1889–1895.

Crow KD, Munehara H, Bernardi G (2010) Sympatric speciation in a genus of marine reef fishes. *Molecular Ecology*, **19**, 2089–2105.

Dayrat B (2005) Towards integrative taxonomy. *Biological Journal of the Linnean Society*, **85**, 407–415.

De Queiroz K (2007) Species concepts and species delimitation. *Systematic Biology*, **56**, 879–886.

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

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

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

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

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

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

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

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

Goldstein PZ, DeSalle R (2011) Integrating DNA barcode data and taxonomic practice: determination, discovery, and description. *BioEssays*, **33**, 135–147.

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

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

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

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

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

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

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

Jablonski D, Lutz RA (1980) Molluscan larval shell morphology – ecological and paleontological applications. In: *Skeletal Growth of Aquatic Organisms* (eds Rhoads DC, Lutz RA), pp. 323–377. Plenum Press, New York.

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

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

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

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

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

Leliaert F, Verbruggen H, Wysor B, Clerck OD (2009) DNA taxonomy in morphologically plastic taxa: algorithmic

species delimitation in the *Boodlea* complex (Chlorophyta: Cladophorales). *Molecular Phylogenetics and Evolution*, **53**, 122–133.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Puillandre N., Holford M (2010) The Terebridae and teretoxins: combining phylogeny and anatomy for concerted discovery of bioactive compounds. *BMC Chemical Biology*, **10**, 7.

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

Puillandre N, Lambert A, Brouillet S, Achaz G (2011) ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology*, doi: 10.1111/j.1365-294X.2011.05239.x.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Wheeler QD (2009) *The New Taxonomy*. CRC Press, Boca Rotan, Florida.

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

Will KP, Mishler BD, Wheeler QD (2005) The perils of DNA Barcoding and the need for integrative taxonomy. *Systematic Biology*, **54**, 844–851.

Yeates D, Seago A, Nelson L *et al*. (2010) Integrative taxonomy, or iterative taxonomy? *Systematic Entomology*, **36**, 209–217.

P.N. is a Post-Doc student at the Museum National d'Histoire Naturelle, Paris. He is interested in the taxonomy, diversification and evolution of marine gastropods, and in particular the Conoidea. M.V.M. is a Post-Doc student at the Dip. di Biologia e Biotecnologie 'C. Darwin' of the University of Roma 'La Sapienza'. Her research focus on evolutionary biology, zoology and systematics of marine gastropods. Y.Z. is interested in applied mathematics, statistics, data mining and machine learning methods in: bioinformatics, neuroscience, genomics, healthcare, business intelligence and finance. L.S. is interested in applied mathematics, mathematical modeling and data analysis. M.C.B. is a population genetist at the Centre National de Recherche Scientifique, head of the Service de Systematique Moleculaire of the MNHN. Her scientific focuses include species delimitation and processes of speciation, especially on marine invertebrates, and barcoding. C.C. is working at the Genoscope facility and was in charge of the molecular sequencing. M.H. is as an Assistant Professor of Chemistry and Biochemistry at Hunter College, with a scientific appointment at the American Museum of Natural History. Her dual appointment reflects her interdisciplinary research, which combines biology and chemistry to investigate the evolutionary history of venomous marine snails, and to discover, characterize, and deliver novel snail peptides as tools for manipulating cell signaling in the nervous system. In 2011 she was awarded an NSF CAREER Award, and named a 21st Century Chemist in the NBC-Learn Chemistry Now series. S.S. is a evolutionary biologist at the Institut de Recherche pour le Développement; she is interest in conceptual and methodological developments in systematics.

Data accessibility

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

All samples are vouchered in the MNHN collection. They are all registered in the Barcode of Life Datasystem (BOLD), in the project 'CONO—Conoidea barcodes and taxonomy'.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Analysis of the bathymetric ranges of the PSH. It was possible to conclude from the observation of the bathymetrical ranges of two PSH if they were overlapping or not. However, the observed bathymetrical range of a PSH could be only a sampling artifact. To evaluate this hypothesis, a statistical test was performed. Two different PSH, A and B, characterized respectively by their bathymetric ranges RA and RB and numbers of collected specimens NA and NB, and their association in a single PSH C, with RC and NC, are considered. The null hypothesis is that the PSH A and B are only subsamples of the PSH C. NA bathymetrical ranges (RA') were randomly determined within RC, this operation being repeated 1000 times. The observed range RA is then compared to the distribution of the 1000 RA'. If RA is placed within the 5% of the weakest values of the distribution, the null hypothesis is rejected. The same test is also applied to the PSH B. The results are then interpreted depending on the respective position of the two ranges RA and RB.

Table S1 Genetic structure calculated with the COI and 28S genes for respectively eight and five PSH. *F*st values are calculated between each pair of stations for which at least six specimens were available. Into parentheses: number of specimens. An AMOVA was also performed for the PSH 67 between two groups (Philippines and Vanuatu) and four populations in the Vanuatu.

Table S2 List of the 1000 specimens analysed in this study.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.