Mar Biodiv (2015) 45:433–451
DOI 10.1007/s12526-015-0319-7

SENCKENBERG

MEIOSCOOL

Guidelines for DNA taxonomy, with a focus on the meiofauna

Diego Fontaneto · Jean-François Flot · Cuong Q. Tang

Received: 31 October 2014 / Revised: 17 January 2015 / Accepted: 12 February 2015 / Published online: 28 February 2015

Abstract Describing biological diversity is a challenging endeavour, especially for the small, cryptic animals that make up the meiofauna. The field of DNA taxonomy, i.e., the use of DNA to delineate species boundaries, is rapidly growing and changing; herein we review the recent advances in the acquisition of DNA sequence data and the analytical tools for DNA-based species delimitation, with a focus on applications to the meiofauna. After providing general guidelines on the data collection and analysis steps (sampling design, sequencing, phasing of nuclear markers, and sequence alignment), we explain the rationale and usage of several widely used or promising methods developed for delineating species from single-locus data sets (distance-based DNA barcoding, automated barcode gap discovery, K/θ, generalized mixed Yule–coalescent models, Poisson tree process model, and haplowebs). As it is increasingly recognised that several loci are required to delineate species accurately, we then briefly outline multilocus species delimitation approaches (Structure, Structurama, Bayesian phylogenetics & phylogeography, SpedeSTEM, O'Meara's heuristic search, and several newly published Bayesian approaches based on the multispecies coalescent).

Communicated by D. Zeppilli

D. Fontaneto (✉)
National Research Council, Institute of Ecosystem Study,
Largo Tonolli 50, 28922 Verbania Pallanza, Italy
e-mail: d.fontaneto@ise.cnr.it

J.-F. Flot
Department of Genetics, Evolution and Environment,
University College London, Darwin Building, Gower Street,
London WC1E 6BT, UK

C. Q. Tang
Department of Life Sciences, Imperial College London,
Silwood Park Campus, Ascot, Berkshire SL5 7PY, UK

C. Q. Tang
Department of Life Sciences, The Natural History Museum, Darwin
Centre 2 - room 627, Cromwell Road, London SW7 5BD, UK

Keywords ABGD · BP&P · Biodiversity · COI · Cryptic species · GMYC · Identification · PTP

Introduction

Species are the fundamental unit of biodiversity (Mayr 1982); therefore, proper species delimitation and identification are important prerequisites to population genetic, physiological, and ecological studies (Wiens 2007; Butlin et al. 2009). However, species–level taxonomy is rife with practical issues, especially for groups whose morphology is uninformative, plastic, and/or difficult to describe (Hebert et al. 2003). This contrasts with the relative ease with which conspicuously divergent taxa can be recognized, as is often the case when dealing with large terrestrial organisms such as birds, mammals, and butterflies (Gaston and Blackburn 2000). In contrary, marine invertebrate species are notoriously difficult to delineate; extreme cases of difficulties are represented, for instance, by corals that are morphologically plastic in response to variations in environmental conditions (Todd 2008), by assemblages of planktonic larval stages (Baretta-Bekker et al. 1998), or by microscopic meiofaunal organisms in the sediments (Curini-Galletti et al. 2012).

Even in well-studied and familiar organisms, questions arise regarding the diagnosis and interpretation of morphological and/or genetic divergence as intraspecific or interspecific (Hey 2009). The situation is complicated further by the variety of alternative theoretical approaches for defining species boundaries. Using different species delimitation criteria (de Queiroz 2007) and metrics (e.g., Tang et al. 2012) can produce contrasting assessments of diversity: for example, strict phylogenetic delimitation criteria may be more prone to overestimate the number of species compared to gene-flow–based biological criteria that put emphasis on reproductive isolation (or lack thereof) as the defining property of species (Flot et al. 2010, 2011).

In addition to the ambiguities inherent to the interpretation of biological phenomena such as species, another problem is the so-called taxonomic impediment, i.e., the gap between the small number of expert taxonomists and the large number of species to describe and specimens to identify (Rodman and Cody 2003). The taxonomic impediment is especially pervasive for the meiofauna (Giere 2009), which is composed of small animals with high levels of cryptic diversity and frequent morphological stasis (e.g., Fontaneto et al. 2009). The scarcity of available taxonomic expertise is one of the reasons for the abundant undescribed diversity that is characteristic of most meiofaunal taxa (Curini-Galletti et al. 2012; Fonseca et al. 2014). Furthermore, meiofaunal organisms may inhabit areas that are difficult to sample, such as the deep or open sea, or the sediment of remote caves. Getting a better understanding of meiofaunal diversity is especially important when studying marine benthic environments, where meiofaunal organisms play key ecological roles (Zeppilli et al. 2015). Thus, we will focus our review on the methods of DNA taxonomy applicable to meiofaunal species, particularly those inhabiting marine environments. Our hope is to help systematists and non-systematists alike use DNA information to obtain reliable data on the underexplored but highly varied group of organisms that compose the meiofauna. However, the methods we outline are not specific to the meiofauna and will be useful to a broader audience.

DNA taxonomy techniques offer taxonomists and ecologists fast and objective means to assess biodiversity. Here we define DNA taxonomy as the analysis of variation in genetic data (such as DNA sequences of selected loci or complete genomes, single-nucleotide polymorphisms, microsatellites, amplified fragment length polymorphisms, etc.) to delimit species (Tautz et al. 2003). These types of studies are now more accessible than ever, and numerous sequence-based approaches have been proposed to inform species diagnosis, mostly thanks to the increasing abundance of molecular data (McCormack et al. 2013), to the existence of large sequence data sets (e.g., BOLD; Ratnasingham and Hebert 2007), and to the rise of quantitative phylogenetic methods (Sites and Marshall 2004; Carstens et al. 2013).

For animals, most studies start by (and many rely solely on) sequencing one marker of suitable variability (e.g., the "barcode" portion of the mitochondrial cytochrome oxidase *c* subunit I gene, or the internal transcribed spacer 2 located in nuclear ribosomal DNA) to delineate species, yielding species hypotheses that can be compared to morphology, ecology, cross-mating experiments, physiology, distribution, and behaviour. However, it is increasingly recognised that accurate species delineation requires a multilocus approach taking into account two or more independent markers from the same individuals, which necessarily involves sequencing nuclear genes (since all mitochondrial markers are linked together and cannot be considered independent sources of information; Moore 1995). Sequencing nuclear markers is often considered difficult and costly because of their heterozygosis: hence, multilocus species delineation methods are most often used when single-locus approaches yield results that are ambiguous or inconsistent with morphology or with the other sources of information mentioned above. Multilocus approaches to species delimitation is an active field of research and will undoubtedly predominate in future studies, given the ever-decreasing per-base cost of sequencing and the ease with which such approaches should be scalable to entire genome sequences. Although most available multilocus species delineation methods are still experimentally and computationally expensive (and are therefore difficult to apply to studies of meiofaunal species), we include them in our review, albeit briefly, given their great potential.

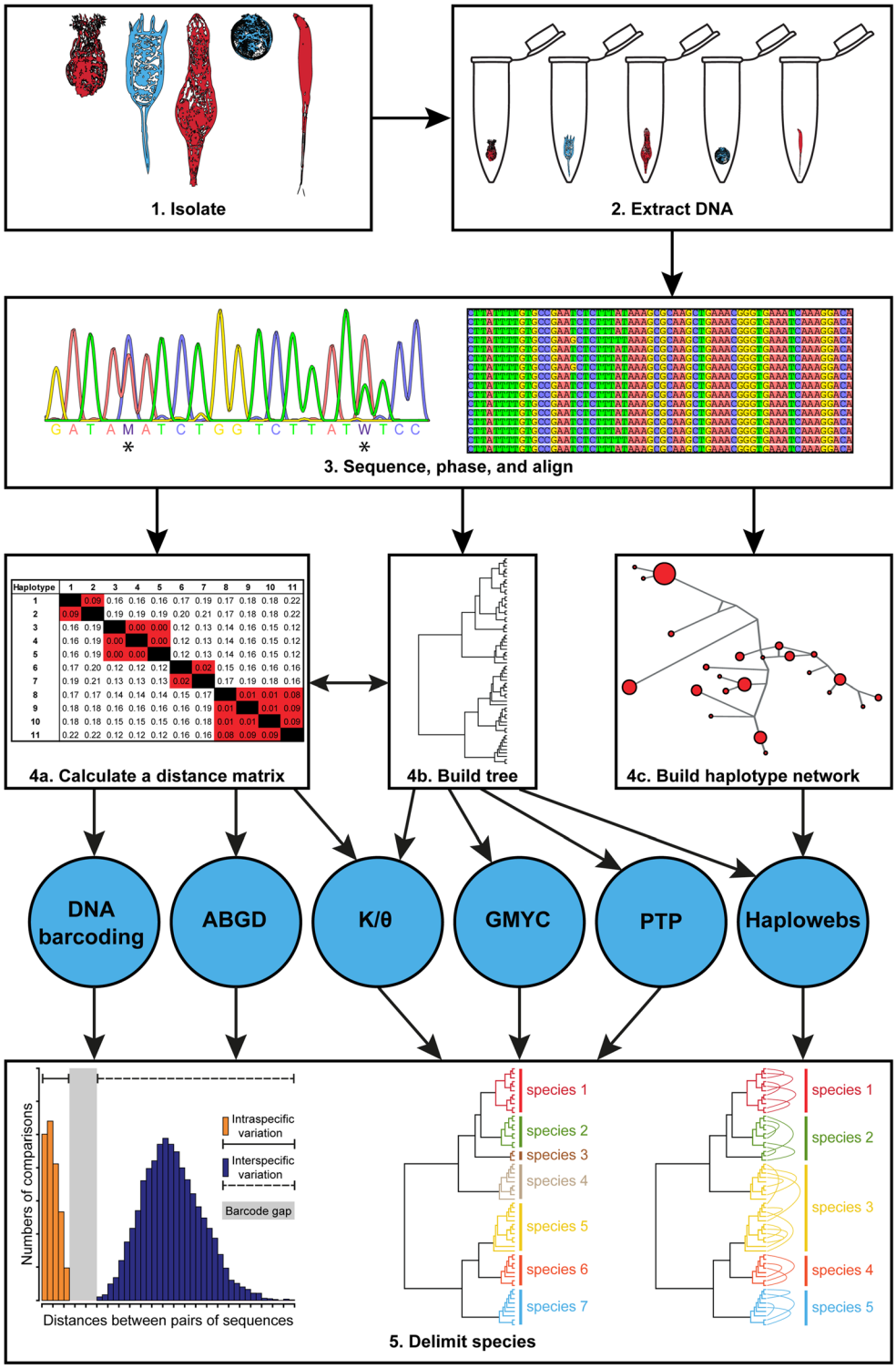
Fig. 1 Typical DNA taxonomy workflow. First row from top, laboratory procedures: (**1**) animals are isolated in the wild and identified to morphospecies level under light microscopy; (**2**) these animals are washed in double-distilled H_20 and transferred to individual tubes in which the DNA is extracted. Second and third rows, data acquisition: (**3**) specific genetic loci are PCR-amplified, Sanger-sequenced, phased to isolate individual haplotypes (in the case of nuclear loci), and aligned. The two *asterisks* denote ambiguous, potentially heterozygous, base pairs (see Fig. 2). (**4a**) The alignment is used to construct a pairwise distance matrix and/or (**4b**) a phylogenetic tree and/or (**4c**) a haplotype network. Note that phylogenetic trees can be constructed either from sequence alignments or from distance matrices, whereas it is also possible to compute a patristic distance matrix from the branch lengths of a tree. Bottom rows, data processing: (**5**) the matrix and/or phylogeny and/or network serve as a basis for various species delimitation approaches, either based on distances (*left*, DNA barcoding and ABGD), on branching rates (*centre*, K/θ, GMYC and PTP), or on heterozygosis (*right*, haplowebs). In the latter case, the curves drawn on the right side of the tree connect sequences found co-occurring in heterozygous individuals

Our review considers all the steps in DNA taxonomy, starting with data acquisition before detailing the various single-locus species delimitation methods available (Fig. 1). We then consider succinctly the more refined multilocus approaches, and conclude our review by emphasising some potential caveats and pitfalls. The various methods attempt to optimise different criteria for identifying and delimiting species: given that our review aims to provide guidelines for DNA taxonomy, we do not dwell on the differences in philosophy between these approaches, nor on their statistical properties; neither do we consider the confounding effect of interspecific gene flow, nor the possible differences between sexual and asexual organisms. Instead, we use a pragmatic approach, highlighting the previous usage of these methods in meiofaunal studies, their strengths, and their potential weaknesses.

Data acquisition

Sampling and DNA extraction

Data acquisition is the crucial first step in any species delimitation study; hence, careful consideration is required when



designing a sampling scheme, as this choice can have a strong influence on the number of species delimited (Papadopoulou et al. 2008; Lohse 2009; Bergsten et al. 2012; Talavera et al. 2013). Most DNA taxonomy methods implicitly or explicitly assume that all populations and species are adequately sampled (Lim et al. 2012). In the traditional approach, when specimens are sorted and processed individually, it is advisable to plan a redundant sampling in which each species is represented by multiple specimens, if possible collected from different locations. Collecting large numbers of specimens is rarely a problem for meiofaunal species living in undisturbed marine littoral sediments, given their small size and abundance: non-selective

methods such as aspiration of sediments followed by filtration usually yield numerous specimens, but specific methods targeting different meiofaunal groups are also available (Giere 2009; Curini-Galletti et al. 2012). Yet, given the uneven distribution of species abundances in the field, often characterised by a small number of common species and a very high number of rare species (Magurran and Henderson 2003), some clades are likely to be over-represented in any sampling strategy. This is even more problematic for rare species living in peculiar habitats such as marine caves and crevices or sediments in oligotrophic streams: collecting them may pose specific challenges, in which case a balanced sampling scheme is particularly difficult to achieve. Many different extraction methods are available to extract a reliable quantity of DNA from single individuals of the meiofauna, and the resulting DNA can be stored for years at −20 °C; however, when large numbers of minute individuals have to be processed, it is advisable to choose quick and inexpensive methods such as the HotSHOT protocol (Truett et al. 2000; Montero-Pau et al. 2008) or a combination of Chelex and proteinase K (Estoup et al. 1996), which work well even with an amount of material as small as a single rotifer egg (Montero-Pau et al. 2008).

In contrast with the traditional, specimen-per-specimen approach outlined above, an alternative that is increasingly being used is to collect samples of sediment and process them in bulk, without sorting the animals into single tubes (Creer et al. 2010; Fonseca et al. 2010, 2014). With this second approach, called "metagenetics" or "metabarcoding" (Taberlet et al. 2012), standardisation and planning of a balanced sampling design for a targeted group of species or species complexes is impossible, and no morphological information can be linked to each individual. Whereas the first method is most useful for detailed taxonomic analyses, the second one can provide quantitative genetic information suitable for species identification (provided that a reference sequence database is available) and ecological studies. Besides, analysing bulk environmental samples makes it possible to detect efficiently rare species and study their distribution (Zhan and MacIsaac 2015). Various kits are available to perform bulk DNA extraction from water or sediments, and the choice of a particular approach should be considered carefully, since different methods have been shown to yield different overall estimates of biological diversity (Knauth et al. 2013; Rees et al. 2014; Deiner et al. 2015). When the biomass of the meiofauna is very low compared to the sampled volume of soil, sediment, or water to be processed, animals are commonly isolated from their environment before pooling them and extracting bulk DNA (Creer et al. 2010; Fonseca et al. 2010, 2014).

Choice of marker(s)

Once an appropriate sampling and DNA extraction strategy has been chosen, the next important step is to decide which marker(s) will be analysed. The basic and most important feature of a marker suitable for DNA taxonomy is its variability, but other properties have to be considered as well, notably a marker's propensity to convergent evolution. Although microsatellite markers are extremely variable, they are not the most suited for delimiting species because their number of repeats is always comprised between zero and an upper boundary of at most a few tens of repeats, meaning that two individuals may present the same number of repeats by convergence instead of by descent (Garza et al. 1995; Garza and Freimer 1996). The same convergence problem plagues single-nucleotide polymorphisms (SNPs), as they can only take one out of four possible states: A, C, G and T (or even only two states for diallelic SNPs). Convergence is also frequently observed with approaches that are based on the presence/absence of bands of specific lengths, such as randomly amplified polymorphic DNAs (RAPDs), single-strand conformation polymorphisms (SSCPs), and amplified fragment length polymorphisms (AFLPs). By contrast, the probability of obtaining two identical DNA sequences of several hundred base pairs by random convergence instead of by descent is so small that it is practically negligible (Tajima 1983), which is why most DNA taxonomy studies use sequence data to delineate species. Besides, DNA sequences from previous studies can easily be obtained from GenBank (Benson et al. 2013) and reused, which greatly improves the strength of approaches based on sequences compared with other types of genetic data.

Another important property of a marker suited for DNA taxonomy is its universality: markers that are only available for a subset of species are not particularly useful when dealing with the various taxa that make up the meiofauna. Therefore, although genomes comprise thousands of genes and lots of intergenic regions that can potentially be used as markers for DNA taxonomy, it is important that the variable region be flanked by two conserved ones that can be used as priming regions to ultimately amplify and sequence the marker by the polymerase chain reaction (PCR; Saiki et al. 1988). Introns of single-copy genes are ideal from this point of view, since their sequences experience little structural constraints and the exons on each of their sides are much more conserved, making it generally possible to design excellent PCR primers. Such markers are often called EPIC (exon-primed intron-crossing) in the literature (Palumbi and Baker 1994), and can be defined with increasing ease now that complete genome sequences are available for many groups. However, given that meiofaunal species are small and therefore contain little DNA per specimen, single-copy gene introns can be difficult to amplify consistently (notably if the DNA extracts contain inhibitors; Rameckers et al. 1997). This is why single-copy gene introns are rarely a first choice in DNA taxonomy studies dealing with the meiofauna; instead, most studies use markers that are present in many copies per cell, as part of either mitochondrial or ribosomal DNA.

The most commonly used marker for the DNA taxonomy of animals is a fragment of the cytochrome *c* oxidase subunit 1 gene (abbreviated variously as *cox1* or COI), for which universal metazoan primers are available (Folmer et al. 1994) and for which other primers targeting specific groups are frequently redeveloped (e.g., Prosser et al. 2013 for freshwater microcrustaceans). This is a very variable marker in bilaterians, able to resolve not only species, but also populations within a given species (Avise et al. 1987). However, COI shows much less variation among non-bilaterian metazoans such as cnidarians (Shearer et al. 2002; Hellberg 2006; Huang et al. 2008). For these metazoans, the entire mitochondrial genome appears to be very stable: in a particularly striking example, populations of the deep-sea coral *Lophelia pertusa* collected 7,500 km apart in different oceanic basins (the Mediterranean Sea vs. the Barents Sea) were found to share near-identical complete mitochondrial genome sequences that differed by at most a single nucleotide position (Flot et al. 2013). As a result of this stability, COI sequences are not very useful for distinguishing cnidarians species (Shearer and Coffroth 2008), although other mitochondrial regions variable enough for this purpose have been found in some groups (Pont-Kingdon et al. 1995; Flot and Tillier 2007). Similar challenges are likely to occur in meiofaunal cnidarians such as the interstitial hydrozoan genus *Halammohydra*, which is prominently displayed on the logo of the International Association of Meiobenthologists, but may concern other meiofaunal groups as well.

Ribosomal DNA markers are also frequently used to delineate species. Like COI, ribosomal DNA is present in many copies per cell and can therefore be readily amplified using PCR. These copies generally evolve in a concerted fashion (Dover 1982; Hillis et al. 1991) that prevents them from diverging; as a result, although sequencing cloned PCR products of ribosomal DNA markers may yield a bewildering variety of sequences, it has been found that direct sequencing of the ITS (intergenic transcribed spacer), for example, tends to yield only one or two dominant types per specimen, rarely three (Flot et al. 2006). Some studies used a variable region of the 28S ribosomal DNA gene (e.g., Lorion et al. 2010) for which semi-universal primers are available (e.g., Verovnik et al. 2005 for isopods and amphipods), but this region is not always variable enough to distinguish closely related species; for instance, the two amphipod species *Pontoniphargus racovitzai* and *Pontoniphargus ruffoi* are morphologically distinct and reciprocally monophyletic for COI, but share the same 28S sequence (Flot et al. 2014). Hence, when sequencing a single marker, it may be preferable to target the internal transcribed spacers ITS1 and/or ITS2, located between the 18S and 28S genes in ribosomal DNA. These genes are much more variable than 28S and therefore more suitable to distinguish closely related species, although they can be difficult to amplify in some groups in which their sequences are very long; for example in niphargid amphipods, with a record length of 1159 bp for *Niphargus plateaui*, the longest ITS2 sequence ever reported for a metazoan (Kornobis and Pálsson 2013).

Sequencing

Sequencing strategies targeting a single marker or a handful of independent markers usually start with amplification of the target region(s) using PCR. Optimised PCR protocols have been developed for most taxa, the breadth of which is outside of the scope of this review; for specific protocols, it is best to consult the literature available for each group. We provide references to meiofaunal studies in the following paragraphs, and the same papers often report protocols on how the sequences were obtained. These protocols sometimes require some level of troubleshooting, with common approaches to consider. A common source of failure, notably when pooling individuals, is an excess of material during the DNA extraction step, which may result in the co-purification of various contaminants that hinder downstream PCR amplification. To avoid this problem, one should decrease the amount of material used for DNA extractions (it is perfectly fine if the amount of DNA recovered is so small that it is undetectable by spectrophotometry). For instance, when dealing with minute amphipods, it is possible to routinely use only one or two legs for DNA extractions (Flot 2010a; Flot et al. 2010); to the extreme, successful amplification of multiple markers can be obtained even from single microscopic individuals such as unicellular eukaryotes (Blin and Stafford 1976). A second important thing to check is whether the PCR amplification buffer contains dimethylsulfoxide (DMSO). Adding a final concentration of 5 % DMSO to PCR mixes helps amplifying DNA regions by preventing them from folding into secondary structures (Winship 1989), as ribosomal DNA and introns typically do. A third action to take if results are inconsistent across individuals is to increase the number of PCR cycles, to take into account the reduced PCR efficiency resulting from co-purified contaminants or DNA fragmentation (Rameckers et al. 1997): enduring 65 cycles of denaturation-annealing-elongation poses no problem to modern *Taq* polymerases. Even though increasing the number of cycles may result in high-molecular weight products visible as smears on gels (Bell and DeMarini 1991), this does negatively impact the Sanger-sequencing of these products. Finally, a failure to obtain PCR products in some or all individuals can be due to the *Taq* polymerase being too stringent, in which case trying a different, less stringent *Taq* polymerase (e.g., BioTaq or QbioTaq) may solve the problem, especially when using degenerate primers. Conversely, to get rid of multiple bands or smears, one should try a more stringent *Taq* (e.g., RedTaq). If all this fails, it probably means that the primers found in the literature are suboptimal, in which case it may be necessary to

design new ones; an excellent tool to do so, yielding robust primers pairs that do not require further optimisation, is Primer3 (Rozen and Skaletsky 2000), which is freely available as an online web tool. Otherwise, the strategies listed in Roux (2009) can be used to get suboptimal primers to work. Another problem that can be encountered when sequencing mitochondrial markers in animals is the presence of nuclear mitochondrial paralogs, called pseudogenes or numts, an abbreviation for "nuclear mitochondrial DNA" (Richly and Leister 2004). These extra copies are often co-amplified with the mitochondrial marker, or may even be the only copies that are amplified in some cases, creating problems if they are not recognised as paralogous to the other sequences in the data set (Song et al. 2008). Discarding *a posteriori* sequences likely to be pseudogenes, for instance because of stop codons or an excess of non-synonymous mutations, can alleviate the problem, but some numts may remain undetected. Thus, the most reliable solution is to take measures such as pre-PCR dilution to avoid amplifying numts in the first place, as suggested by Calvignac et al. (2011).

Most DNA taxonomy projects adopt the Sanger technology (primer elongation using a mixture of deoxynucleotides and dideoxynucleotide terminators followed by electrophoresis; Sanger et al. 1977) to sequence PCR products. An important good practice is to systematically check the chromatograms obtained from Sanger sequencing and not to blindly use the FASTA files provided with the chromatograms. Automated base-calling programs often make mistakes, such as miscalling some bases, overlooking others and including artefactual ones. If such mistakes are not detected but instead are carried on into the downstream analyses, they will inevitably affect any further inference and may bias the results. For instance, Collins and Cruickshank (2014) report how DNA barcode sequences that contain stretches of incorrect bases because of common sequencing artefacts ("dye blobs") are classified incorrectly. The best, standard way to avoid such problems is to systematically sequence each marker using several primers, so that each base is covered by at least two different chromatograms, ideally in different directions. Aligning the forward and reverse chromatograms together makes it possible to detect and correct errors automatically, in a very time-efficient way and without having to visually inspect all the bases. An open-source program for aligning forward and reverse chromatograms and checking for discrepancies between the two is SeqTrace (Stucky 2012), available online at http://seqtrace.googlecode.com/. Although sequencing PCR primers in both directions inevitably inflates the cost of DNA taxonomy, it is the best way to obtain sequences that are reliable and accurate. Spending extra time and money at the beginning of the process in order to obtain accurate sequences greatly diminishes the chances of carrying mistakes through the later stages of the analyses, thereby saving time and money overall.

As an alternative to traditional Sanger sequencing approaches, recent advances in sequencing technologies, commonly grouped under the terms next-generation sequencing (NGS) or high-throughput sequencing (HTS), have been revolutionizing DNA studies for the last decade (Kircher and Kelso 2010). These approaches can be used to analyse the complete genome of one individual at a time, but also to sequence PCR products for a fraction of the cost of traditional Sanger sequencing (Bik et al. 2012). In the most basic approach, PCR products are pooled together for sequencing; the PCR products of several individuals can be tagged using specific identification suites of nucleotides to tell them apart in the ensuing analyses (Meyer et al. 2008). It is also possible to perform hundreds of PCRs in parallel in a single test tube using microdroplets, and then simultaneously sequence all of the resulting amplicons (Tewhey et al. 2009). More refined reduced-representation approaches can be used to obtain the sequences of thousands of (supposedly) homologous loci across large numbers of individuals for a fraction of the cost of whole-genome sequencing (Van Tassell et al. 2008); notably, these approaches include RNAseq (only the messenger RNAs of genes that are expressed are sequenced; Wang et al. 2009), RADseq (only restriction-site–associated DNA regions are sequenced; Davey and Blaxter 2010), and exon capture (only exonic sequences complementary to specific oligonucleotidic probes are targeted; Hodges et al. 2007). One can already foresee that in the future, most of these approaches will be replaced by direct, full-genome sequencing of many individuals in the populations under study, generating a tsunami of data unprecedented in the history of biology. In preparation for this, Dowton et al. (2014) recently proposed a framework for next-generation DNA barcoding where multilocus data sets are coupled with coalescent-based species delimitation methods, which sparked an intriguing debate about the benefits and potential limitations of large next-generation data sets in DNA barcoding practises (Collins and Cruickshank 2014).

Data processing

Phasing heterozygous sequences

One thorny issue when sequencing nuclear markers is how to deal with heterozygosis. Indeed, markers that are variable enough to be useful for distinguishing species also present a good deal of intraspecific variations, which in the case of diploid organisms results in double peaks in the chromatograms obtained from Sanger-sequencing PCR products (Fig. 2). Although double peaks can be easily overlooked when they are few (particularly if one of the alleles gives a

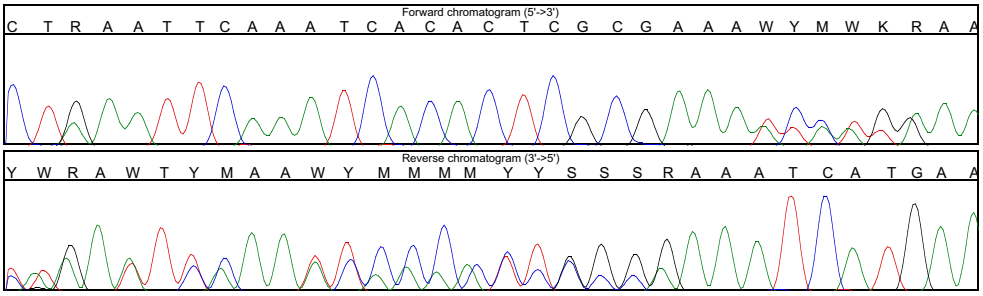


Fig. 2 Example of forward (*top*) and reverse (*bottom*) chromatograms obtained from Sanger sequencing of a length-variant heterozygote. When the two alleles of the sequenced individual differ by one or several indels in addition to SNPs, a large number of double peaks are observed; however, the double peaks in the forward and reverse chromatograms contain different, complementary information, allowing reconstruction of the two haplotypes of the individual without cloning, either by hand (Flot et al. 2006) or using Champuru (Flot 2007, accessible online at http://jfflot.mnhn.fr/champuru/). In the example shown, the two haplotype sequences differed by one substitution and a 1-base indel: T**A**AATTCAAATCACACTCGCG**AAAA**TCATGAA and T**G**AATT CAAATCACACTCGCG**AAA**TCATGAA

much stronger signal than the other one, resulting in peaks of unequal heights), a striking pattern occurs when the two alleles of an individual have different lengths: the chromatograms of such length-variant individuals display numerous double peaks in both their forward and reverse chromatograms (Flot et al. 2006). Such chromatograms were often discarded in the past as they looked hopelessly messy; however, the double peaks in the forward and reverse chromatograms are different, and it is therefore possible to combine them to reconstruct with certainty the sequences of the corresponding two alleles (Flot et al. 2006). This can be easily done by hand when the two alleles differ by only one insertion/deletion (indel), but when they differ by multiple indels it becomes quite complicated to perform the task manually, which led to the development of a web tool that automates the reconstruction process (Flot 2007; available online at http://jfflot.mnhn.fr/champuru/).

In a typical EPIC data set, about 30 % of the individuals are homozygous, 30 % are heterozygotes presenting a single double peak in their chromatogram (a case that is trivial to solve), and 30 % are length-variant heterozygotes that can be solved as outlined above. The remaining 10 % are heterozygotes with several double peaks, meaning that their two alleles have the same length but differ by more than one substitution. Phasing these individuals requires comparing their genotypes with those of other individuals sampled in order to infer the most likely haplotypes: this can be performed either by hand (Clark 1990) or computationally in a Bayesian framework using the programs SeqPHASE (Flot 2010b; available online at http://seqphase.mpg.de/seqphase) and PHASE (Stephens et al. 2001). Since the haplotypes of 90 % of the individuals in the data set are already known prior to running SeqPHASE and PHASE, the remaining 10 % are usually inferred with very high posterior probabilities. The rare individuals for whom uncertainties remain (as indicated by posterior probabilities lower than 0.9) may be solved by re-sequencing the PCR products using haplotype-specific primers (Hare and Palumbi 1999), or, as a last resort, by cloning them. Therefore, using this set of methods makes it possible to consider nuclear sequence markers as co-dominant, which was previously only the case for markers scored on gels (such as microsatellites and AFLPs).

For markers that present copy-number variations (CNVs; Freeman et al. 2006), individuals with more than two haplotypes are observed; the chromatograms of such individuals contain triple peaks if there are three haplotypes, and even quadruple peaks if more than three haplotypes are present. Paradoxically, this seems to happen more frequently with supposedly single-copy markers such as nuclear gene introns, rather than with multicopy markers such as ribosomal DNA (Flot et al. 2008). Resolving individuals with three or more haplotypes is difficult but not impossible, for instance using haplotype-specific primers or by cloning.

Aligning sequences

For protein-coding markers such as the widely used COI, alignment is straightforward since there are no single-nucleotide insertions or deletions (indels), which would result in loss-of-function frameshifts. Instead, indels always involve multiples of three bases and are typically rare among closely related species. Such sequences can easily be aligned manually using the alignment editor included in MEGA (Tamura et al. 2013) or in Mesquite (Maddison and Maddison 2014). The situation is different for markers that are not protein-coding, such as ribosomal genes and introns; non-coding markers often exhibit many indels of various sizes, even between closely related species. For small numbers of sequences with few indels, the implementation of MUSCLE (Edgar 2004) in MEGA comes in handy, but for large, complex data sets, we recommend using MAFFT (Katoh et al. 2009), which is conveniently available as a web server (http://mafft.cbrc.jp/alignment/server/). Alignments should be checked by eye for small errors in some sequences, particularly towards the ends of sequences where the quality of the chromatograms tails off; when such errors are suspected, the original chromatograms

should be consulted so that bona fide sequence differences are not mistaken for sequencing errors. Whatever the type of marker and the alignment strategy, it is a good practice to curate the sequences by removing primers and (if cloning was performed) vector sequences, so that the first base in the alignment corresponds to the first base of the marker following the forward primer and the last base in the alignment corresponds to the last base of the marker preceding the reverse primer. Sequences alignments can then be processed in programs such as MEGA (Tamura et al. 2013) to generate phylogenetic trees and/or pairwise distance matrices under various evolutionary models. To determine the best-suited model, one may use the comparison tool included in MEGA or the standalone program jModelTest (Posada 2008). Patristic distances (i.e., distances between two tips measured along the tree) can be calculated atop a phylogenetic tree using the program Patristic (Fourment and Gibbs 2006). When numerous individuals of the same species are sequenced, it may be advantageous to present the data as a network instead of as a tree (Posada and Crandall 2001); among the various approaches available, we favour the median-joining algorithm (Bandelt et al. 1999) implemented in the program Network (Fluxus Technologies), which has been shown to perform well in a simulation study (Cassens et al. 2005). Whatever the methods used, a good practice in data sharing and data quality control is to make all chromatograms, alignments and trees/networks freely available online, either as supplementary material to the paper concerned or in dedicated online repositories such as GenBank (Benson et al. 2013), BOLD (Ratnasingham and Hebert 2007), TreeBase (http://treebase.org), or Dryad (http://www.datadryad.org).

In the case of NGS data, the millions of short reads obtained are commonly aligned against a reference sequence using fast, dedicated tools such as bowtie2 (Langmead and Salzberg 2012); then other programs such as SAMtools (Li et al. 2009) are used to generate read pileups and infer consensus sequences. We will not enter into details here as this would go beyond the scope of the present review, but interested readers may refer to the studies of Creer et al. (2010) and Fonseca et al. (2014) for more information on applying metagenetic approaches to meiofaunal studies.

Data analysis

Current approaches: single-locus species discovery

Up to now, the most commonly used methods in DNA taxonomy have been designed for data sets with one single marker sequenced across several individuals (single-locus datasets), although some may also be applied on concatenated alignments of several loci. The popular methods (or methods likely to become popular in the future) described below include those that require only a matrix of pairwise genetic distances, such as DNA barcoding (Hebert et al. 2003) and ABGD (automated barcode gap discovery; Puillandre et al. 2012a); those that require both a matrix of genetic distances and a phylogenetic tree, such as K/θ (formerly known as the 4X rule; Birky et al. 2010); those that require only a phylogenetic tree, such as GMYC (generalized mixed Yule–coalescent; Pons et al. 2006; Fujisawa and Barraclough 2013) and PTP (Poisson tree process; Zhang et al. 2013); and those that require phased heterozygous markers, such as haplowebs (Flot et al. 2010) (Fig. 1).

Using several of these approaches and looking for a consensus between the results obtained may increase our confidence regarding the outcome. However, these different methods use different criteria to delineate species; therefore, one can expect some degree of incongruence (especially when delineating recent species).

DNA barcoding

We define here DNA barcoding in its strictest and narrowest meaning as the use of a fixed, a priori defined threshold in genetic distances to identify units of diversity. DNA barcoding defined in this way groups two distinct, but often lumped disciplines (Hebert et al. 2003; Vogler and Monaghan 2007; Collins and Cruickshank 2012): (1) DNA barcoding sensu stricto, which is the identification of individuals of already known species, and (2) the discovery of new species, which is a branch of the large field of DNA taxonomy. The former consists in comparing standardised stretches of DNA (barcodes) to reference databases to identify unknown specimens, and has been particularly useful for the identification of juvenile stages (e.g., Webb et al. 2006) or of processed animals in the food industry (e.g., Rasmussen et al. 2009). This method is very widely used in the applied fields and in forensic science, and has a lot of added infrastructure around it. Additional methods such as ad hoc distance thresholds (Sonet et al. 2013) have been developed to account for false positives (e.g., erroneous attribution of a specimen of a new species to an already known species), a whole new system called Barcode Index Number (BIN; Ratnasingham and Hebert 2013) has been put in place to organise and register the barcoding data for all animals, and consortia have been assembled across the world to barcode different taxa (e.g., Consortium for the Barcode of Life; CBOL). A detailed description of the rationale and the caveats of DNA barcoding can be found in Casiraghi et al. (2010).

The second aspect, DNA taxonomy through DNA barcoding, is more controversial but easy to implement, and is likely to be the first step in molecular studies of meiofaunal diversity. This approach posits an a priori nucleotide distance threshold, below which specimens are considered conspecific and above which they are considered to belong to different

species. The major assumption behind it is that intraspecific and interspecific variations do not overlap; that is to say, individuals of a given species are more similar molecularly than individuals belonging to different species. The existence of a "barcode gap" is a prerequisite for such an approach to work, but many studies employ a predefined threshold without checking whether it exists or not. Thus, a better approach is to start by plotting the distribution of pairwise distances between sequences in a data set; when this distribution reveals a clear gap, a threshold placed in this gap can be used to delineate species (Lefébure et al. 2006). However, the barcode gap is often difficult to detect or even non-existent, in which case species delimitation using this approach becomes quite arbitrary, and changing the stringency of the threshold changes the estimated diversity (e.g., Creer et al. 2010). Originally, a 3 % nucleotide divergence threshold was proposed for Lepidoptera COI sequences (Hebert et al. 2003); as an alternative, a relative threshold of ten times the mean intraspecific variation for the group under study was subsequently proposed (Hebert et al. 2004). Whether such a threshold is applicable to all groups and whether the initial design of the Lepidoptera study was representative of a natural sample has been debated and has led to suggestions that barcode gaps do not exist (Meyer and Paulay 2005; Wiemers and Fiedler 2007). The application of the DNA barcoding approach with a fixed pre-determined threshold to assess diversity in understudied organisms is certainly appealing, but its use is not as straightforward as it may seem, and the a priori decision of a cut-off threshold is ambiguous and subjective. A valid resource to analyse the results of the application of different barcoding thresholds and other parameters is jMOTU (Jones et al. 2011). Nevertheless, even this approach does not explicitly test if a barcoding gap actually exists in the data set, nor if a more likely threshold exists.

According to simulations, clear barcode gaps are only observed when species have small effective populations sizes and new species originate infrequently (Dellicour and Flot 2015). Indeed, successful studies using DNA barcoding for species discoveries are mostly reported for well-known, large organisms with relatively small effective population sizes, for which a large amount of information on putative species boundaries is already present, and when analyses are performed at a small spatial scale (Bergsten et al. 2012). This is often not the case for meiofaunal organisms since they are small, may have very large population sizes, and are understudied (Curini-Galletti et al. 2012; Fonseca et al. 2014). These features recently led to a discussion on whether delineating species of microscopic organisms using large DNA barcoding data sets is warranted (Rossberg et al. 2013, 2014; Morgan et al. 2014). Consequently, any attempt to apply DNA barcoding to the meiofauna should be crosschecked against the result of other methods (e.g., Tang et al. 2012).

Automatic barcode gap discovery

A less subjective means of defining a barcoding threshold for a given data set is the automatic barcode gap discovery tool ABGD (Puillandre et al. 2012a). Instead of using one or several predefined distance thresholds to delimit species, ABGD attempts to determine directly the threshold that is optimal for a given data set. If no satisfying threshold is detected, it concludes that all the specimens sequenced are conspecific. This method accepts an alignment as input to generate a distance matrix, either raw or corrected following the JC69 (Jukes and Cantor 1969) or K2P (Kimura 1980) models of sequence evolution; alternatively, a user-made distance matrix can be uploaded. ABGD requires users to specify one or a range of upper bounds on intraspecific genetic distances. From each of these priors and the distance matrix, it estimates a 95 % confidence interval for the population mutation rate θ (equal to $4\mu N_e$ for nuclear markers and to $2\mu N_e$ for mitochondrial ones, where μ is the mutation rate and N_e is the effective population size) using coalescent theory, and then looks for gaps in the distribution of pairwise distances that fall outside of the confidence interval for θ (Puillandre et al. 2012a). When one such gap is detected, ABGD uses it as a threshold to delimit primary species hypotheses (PSHs). Like fixed-threshold DNA barcoding methods, and in contrast to most other single-locus methods except haplowebs, ABGD does not require monophyly to delineate species (Fig. 3).

The ABGD method has been used predominantly to define metazoan PSHs. In some cases, it has been found to delimit groups identical or similar to GMYC-based approaches and K/θ (Kekkonen and Hebert 2014); however, in other cases, its results have been more divergent (e.g., Tang et al. 2012). Meiofaunal studies using ABGD have included rotifers (Leasi et al. 2013), nemerteans (Leasi and Norenburg 2014) and molluscs (Jörger et al. 2012), but also nematodes, tardigrades, gastrotrichs, acoels, and flatworms (Tang et al. 2012). A rotifer COI data set with a detailed explanation on how to apply the ABGD approach to it is available in Fontaneto (2014).

K/θ

This method, described by Birky et al. (2005; 2010), uses population genetic theory to propose that, for sister clades of a given marker, interclade divergence (K) at least four times greater than intraclade variation (θ) means that these clades have more than 95 % chances to correspond to different species. In other words, clades meeting this "4X rule" are unlikely to have arisen solely by neutral genetic drift within a single population, but probably experienced barriers to gene flow, such as physical separation (allopatry), divergent selection for adaptation to distinct niches, or both. To date, the method has been used predominantly for asexual taxa, including some

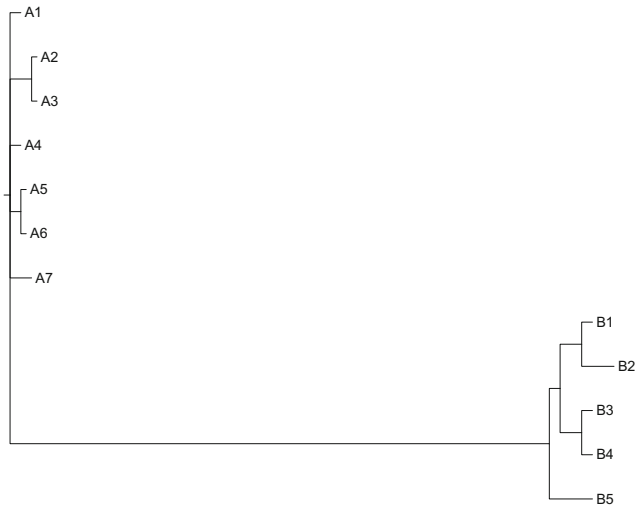


Fig. 3 Hypothetical phylogenetic tree illustrating how distance-based approaches (DNA barcoding, ABGD) can potentially delineate species that are not monophyletic; here, such approaches will group all "A" individuals (A1 to A7) in one species and all the "B" individuals (B1 to B5) in another one, even though the hypothetical species A is not monophyletic

marine meiofaunal species, as their population genetics are arguably simpler (Birky et al. 2010; Tang et al. 2012), but the principles behind the method are applicable to sexual taxa as well (Birky 2013). Although the application of this method has been limited to mitochondrial or chloroplast sequences so far, it can be used with nuclear markers as well (C. William Birky Jr., personal communication).

The K/θ approach requires a gene tree to identify putative sister clades, and distance matrices to estimate genetic variation within and between these clades. Currently, the method requires the user to manually identify which clades are to be tested. The procedure comprises seven steps: (1) generate a gene tree, typically with neighbour joining (but maximum likelihood or Bayesian inference are also acceptable); (2) identify pairs of sister clades that have high support values (> 70 % using bootstrap); then for each pair, (3) calculate the nucleotide diversity π of each clade (equal to its mean pairwise distance corrected for sample size by n/(n-1)); (4) calculate $\theta=2N_e\mu$ by $\pi/(1-4\pi/3)$; (5) calculate the mean pairwise difference between the two clades (K); (6) calculate K/θ; and (7) find the probability of the two clades being compared to be distinct species. As previously stated, ratios higher than 4 mean that sister clades have more than 95 % chances of being distinct species; the probabilities for other ratios are given in a table available from C. William Birky Jr.

This method has already been applied to several meiofaunal groups, including gastrotrichs (Kånneby et al. 2012; Kieneke et al. 2012), rotifers (Birky et al. 2011; Iakovenko et al. 2013; Leasi et al. 2013), copepods (Marrone et al. 2010, 2013), ostracods (Martens et al. 2012, 2013; Shearn et al. 2012; Schön et al. 2012), but also nematodes, tardigrades, nemerteans, acoels, and flatworms (Tang et al. 2012). Birky et al. (2010) used rotifer data sets to develop the method, and Birky (2013) provides a detailed explanation on how to perform the analyses.

GMYC-based approaches

The generalized mixed Yule–coalescent model (GMYC; Pons et al. 2006; Fujisawa and Barraclough 2013) is a coalescent-based phylogenetic method that sets a threshold between coalescent and species-level processes in order to delineate evolutionary significant units (ESUs) akin to species (Simpson 1951). Approaches based on the GMYC model rely on the expectation that intraspecific coalescent branching proceeds discernibly quicker than speciation, which is modelled as a Yule process; therefore, species can be identified in gene tree as clusters of terminals separated by longer internal branches. On an ultrametric tree (i.e., a tree whose branch lengths are proportional to time), changes in branching rates indicative of a shift from species-level processes (i.e., coalescent) to population-level processes (i.e., Yule) can be used to delimit ESUs. To do so, GMYC-based approaches cycle through each node and separately model coalescent and Yule processes, and given the observed branching processes, calculate the most likely threshold(s) between species-level and population-level branching rates.

The single-threshold model (ST-GMYC) was the first one to be proposed (Pons et al. 2006). With this approach, the most likely solution identifying Yule and coalescent processes is compared to the null hypothesis (a single branching rate within a single species) using a χ^2 test. If significant, the threshold is used to delimit ESUs. Given that likelihood values are available for all possible solutions, one may also, as a second step, assess whether other solutions are significantly less likely than the one favoured by the method. Such an approach provides very useful confidence intervals around the most likely solution, which makes it possible to determine whether the species delimitations inferred from the data are reliable. ST-GMYC can be applied using the *splits* package (Ezard et al. 2009) in R (R Core Team 2014) or using a webserver (http://species.h-its.org/gmyc/), with step-by-step guides available in Fontaneto (2014) or on Tomochika Fujisawa's webpage (http://tmfujis.wordpress.com/2013/04/23/how-to-run-gmyc/).

Several further GMYC approaches have been developed: the multiple-threshold GMYC model (MT-GMYC; Monaghan et al. 2009) allows for rate heterogeneity among species and does not assume that the same threshold applies to all the parts of the gene tree; the multimodel-averaging approach (MM-GMYC; Powell 2012) accounts for uncertainty in GMYC model selection; and a Bayesian implementation of GMYC (bGMYC; Reid and Carstens 2012) considers uncertainty in phylogenetic reconstruction. The MT-GMYC and

MM-GMYC approaches are included in the updated *splits* R package (Fujisawa and Barraclough 2013), whereas bGMYC is available for download from Noah Reid's website (https://sites.google.com/site/noahmreid/home/software). A study comparing ST-GMYC, MT-GMYC and bGMYC on simulated data sets found bGMYC to outperform its forerunners in most cases (Dellicour and Flot 2015).

The ultrametric trees needed for the GMYC methods are typically reconstructed using either maximum likelihood with post hoc branch smoothing or using BEAST (Drummond and Rambaut 2007; Bouckaert et al. 2014; Tang et al. 2014a). A recent meta-analysis shows that the ST-GMYC applied on BEAST trees provides the most robust diversity estimates in terms of both richness and identity (Tang et al. 2014a). The GMYC model has been applied to several meiofaunal groups, including gastrotrichs (Kånneby et al. 2012; Kieneke et al. 2012), rotifers (Fontaneto et al. 2007, 2011; Birky et al. 2011; Obertegger et al. 2012, 2014; Leasi et al. 2013; Tang et al. 2014b; Malekzadeh-Viayeh et al. 2014), copepods (Gollner et al. 2011; Cornils and Held 2014), ostracods (Adolfsson et al. 2010; Brandão et al. 2010; Bode et al. 2010; Martens et al. 2012; Schön et al. 2012), nemerteans (Leasi and Norenburg 2014), flatworms (Sluys et al. 2013), and molluscs (Jörger et al. 2012), as well as nematodes, tardigrades, and acoels (Tang et al. 2012). A rotifer COI data set with an explanation on how to analyse it using the single-threshold GMYC approach can be found in Fontaneto (2014).

Poisson tree process model

The Poisson tree process model (PTP; Zhang et al. 2013) is another tree-based species delimitation method that uses coalescence theory to distinguish between population-level and species-level processes. It assumes that intraspecific and interspecific substitutions follow two distinct Poisson processes, and that intraspecific substitutions are discernibly fewer than interspecific substitutions because they have less time to accumulate; this method uses substitutions directly to represent time rather than via a method that corrects for rate variation, such as GMYC. This coalescent-based method is very fast as it does not require ultrametric trees as input (as opposed to GMYC), just a regular rooted gene tree; it has been shown to produce species delimitations matching traditional taxonomic groupings (Tang et al. 2014a). This method is implemented as a standalone program and as a web server (http://species.h-its.org/). The most recent version of the method includes both maximum-likelihood and Bayesian searches for species boundaries, and returns Bayesian support values for those delimited species. Albeit very recent, this method has already found several applications in meiofaunal taxonomic studies, notably on rotifers (Tang et al. 2014a; Velasco-Castrillón et al. 2014), on nemerteans (Leasi and Norenburg 2014), and on copepods (Blanco-Bercial et al. 2014).

Haplowebs

Haplowebs rely on a different species delineation criterion, mutual allelic exclusivity (Doyle 1995; Flot et al. 2010), to delineate species of diploid organisms that have been reproductively isolated long enough to not share any identical sequence for the marker under investigation. Newly diverged species always reach mutual allelic exclusivity prior to, or at the same time as, reaching reciprocal monophyly (Flot et al. 2010); besides, the time needed for newly diverged species to reach mutual allelic exclusivity only depends on the length of the marker and its mutation rate, not on the effective size of the populations. In contrast, the time needed for newly diverged species to reach monophyly is strongly dependent on genetic drift, and therefore on the effective population size of the species. Using shared alleles to delineate species is only applicable to nuclear markers that do not exhibit homoplasy (convergence), which is why mutual allelic exclusivity was shown to perform poorly on microsatellites, RFLPs, RAPDs and AFLPs (Miller and Spooner 1999; Hausdorf and Hennig 2010)—all types of data that, in contrary to DNA sequences, exhibit frequent convergence between species.

The criterion of mutual allelic exclusivity was implemented in a graphical approach called haplowebs (short for "haplotype webs"; Flot et al. 2010): starting from a network or a tree of nuclear haplotypes (the method used for obtaining the tree or network does not really matter), connections are added between haplotypes found to co-occur in heterozygous individuals. Once all connections have been added, inspection of the graph reveals discrete pools of interconnected alleles, each of which corresponds to a group of individuals that appears to be reproductively isolated from the others; each such group is called a "field for recombination" (Carson 1957; Doyle 1995), i.e., a putative species. Since this method is based on the co-occurrence of haplotypes in heterozygous individuals, it cannot be applied to metagenetic data (in which haplotypes cannot be traced to individuals) and it requires that a sufficient number of heterozygotes be sequenced. Hence, a large data set comprising several individuals per species should be collected. Moreover, the haplowebs approach is rooted in the biological species criterion that delineates species based on the presence/absence of gene flow, and as with all approaches that follow this line of thought, it may incorrectly lump species that occasionally hybridise. This method was originally developed for cnidarian DNA taxonomy (Flot et al. 2008, 2010, 2011), and its applications have thus far dealt with cnidarians (Flot et al. 2013; Schmidt et al. 2013; Adjeroud et al. 2014; Schmidt-Roach et al. 2014), crustaceans (Flot et al. 2014) and rotifers (Li 2012).

Comparison between the different single-locus approaches

The methods mentioned above can be classified as tree-based and non–tree-based (Sites and Marshall 2003). GMYC, K/θ,

and PTP are tree-based, and as a result, only delineate species that are monophyletic in the gene trees used to run the method (Table 1); whereas DNA barcoding, ABGD and haplowebs are non–tree-based and do not require monophyly. Surveys found that 15–40 % of species of various groups of animals are not monophyletic in mitochondrial gene trees (Funk and Omland 2003; Ross 2014), and the situation is probably much worse with nuclear markers, since acquisition of monophyly is expected to be four times slower for nuclear genes than for mitochondrial ones (Moore 1995). Besides, the highest percentage of non-monophyletic species reported by Funk and Omland (2003) was for non-insect invertebrates, a category encompassing most meiofaunal taxa. Hence, DNA taxonomic results obtained using a tree-based approach should be crosschecked against a non–tree-based approach so that non-monophyletic species are not overlooked.

These methods also differ in the criterion they use to delineate species: a criterion based on genetic distances for distance-based DNA barcoding; a quantitative approach based on coalescent for ABGD; a phylogenetic criterion based on branching rates for GMYC and PTP; and a population genetic criterion based on genetic isolation for haplowebs and K/θ. When the divergence between species is large, sampling within species is comprehensive, and the effective population sizes are small, then these methods are generally congruent (Tang et al. 2012, 2014a; Carstens et al. 2013; Dellicour and Flot 2015). However, given that each of these methods uses either different criteria or inputs, incongruence between the methods is expected under certain conditions. For example, the stringency of the K/θ method in terms of the separation between lineages and the single thresholds of the ST-GMYC and PTP approaches are expected to lump potentially distinct species when lineages have recently diverged. Undersampling within and between species will likely introduce biases in the coalescent approaches; nevertheless, likelihood-based methods such as GMYC, which provide confidence intervals for the most likely solutions, are potentially able to suggest when undersampling may affect the results. Undersampling is even more problematic for haplowebs, potentially leading to oversplitting (Dellicour and Flot 2015). Furthermore, different rates of substitution among lineages make predefined thresholds inappropriate, as these are likely to either lump species in rapidly evolving lineages or split species in slowly evolving lineages.

Several studies have evaluated factors that could decrease the accuracy of some of these methods. For GMYC, simulation studies have addressed the effects of various aspects of sampling (Papadopoulou et al. 2008; Bergsten et al. 2012; Reid and Carstens 2012; Talavera et al. 2013), population size and speciation rates (Esselstyn et al. 2012; Fujisawa and Barraclough 2013; Dellicour and Flot 2015). For PTP, simulations have been used to evaluate the effect of birth rates (i.e., evolutionary distances between species) and sampling unevenness (Zhang et al. 2013). In general, it seems that GMYC based on BEAST trees provides results highly congruent with PTP (Tang et al. 2014a). A recent simulation study (Dellicour and Flot 2015) compared barcode gap detection, GMYC and haplowebs, and found a "sweet spot" (characterized by small effective population sizes and low speciation rates, resulting in large interspecific divergence and low intraspecific diversity) where all the methods tested performed well. However, none of these single-locus methods was able to delineate species properly when effective population sizes and speciation rates were both large (in which case divergence between species was small and intraspecific diversity was high), emphasizing the need for multilocus approaches to tackle such difficult cases.

Besides, single-locus approaches fail to account for possible discrepancies between markers: if the examined marker exhibits an idiosyncratic evolutionary history (for instance, because it is subjected to interspecific introgressions or gene captures, or because the data set includes paralogues and pseudogenes that obscure the signal), this will directly impact the inferred species boundaries. Most of the single-locus methods mentioned above could be used on concatenated data sets originating from independent markers, as was done for

Table 1 Comparison of popular and/or promising single-locus approaches in DNA taxonomy

Method	Main reference	Input	Type of sequence data	Requires monophyly	Suitable for metagenetics	Yields confidence intervals
DNA barcoding	Hebert et al. 2003	Matrix of genetic distances	All	No	Yes	No
ABGD	Puillandre et al. 2012b	Alignment or matrix of genetic distances	All	No	Yes	No
K/θ	Birky et al. 2010	Phylogenetic tree and matrix of genetic distances	All	Yes	Yes	No
GMYC	Fujisawa and Barraclough 2013	Phylogenetic tree	All	Yes	Yes	Yes
PTP	Zhang et al. 2013	Phylogenetic tree	All	Yes	Yes	Yes
Haplowebs	Flot et al. 2010	Alignment, phylogenetic tree, or network	Only nuclear data from diploid organisms	No	No	No

ST-GMYC on rotifers with COI+28S (Fontaneto et al. 2007) and on non-meiofaunal organisms (e.g., Williams et al. 2011; Bellati et al. 2015). Nevertheless, this cannot be considered a bona fide multilocus approach, since it does not take into account the potential discordance between the signals given by these markers. If the markers disagree with each other, it may be expected that one of them will swamp the signal from the other ones and imprint its own history on the resulting delimitation, or the contradictions to be so strong that no significant delineation will be proposed as an outcome (Bull et al. 1993). In what follows, we present some multilocus methods that make it possible to overcome this problem.

One step further: multilocus species delimitation

Some multilocus approaches looking for congruence between gene genealogies are fairly old (e.g., Koufopanou et al. 1997), but there has been a recent surge of interest in these methods, leading to the publication of several key articles in the last few months. Multilocus species discovery methods undoubtedly represent the future of DNA taxonomy, as the use of a large number of independent markers made possible by technological advances in sequencing will allow researchers to tackle several of the commonly encountered problems in species delimitation (but see Collins and Cruickshank 2014). Multilocus species delimitation methods can account for non-monophyletic species, gene tree discordance, incomplete lineage sorting, gene flow after divergence, and other confounding factors that may create problems in single-locus DNA taxonomy (Camargo et al. 2012; Fujita et al. 2012). Yet, these methods have been rarely used in meiofaunal studies so far; hence, we will just briefly mention them without assessing their strengths and weaknesses, and without providing suggestions on their use.

Structure and Structurama

The programs Structure and Structurama use Bayesian clustering algorithms (Pritchard et al. 2000; Falush et al. 2003; Huelsenbeck et al. 2011) to detect population structure in codominant genetic data such as nuclear sequences or microsatellites. They were originally developed to detect intraspecific population structure caused for example by geographic distance, but they are also frequently used to detect species boundaries (even though their suitability for this purpose is somewhat questionable). One minor drawback of Structure is that it requires the number of populations to be specified beforehand: some methods have been proposed to find the best value of this parameter (Evanno et al. 2005), or one may use the number of species suggested by DNA barcoding for instance. A different approach has been implemented in Structurama, an extension of Structure that uses a Dirichlet-process prior in order to estimate the number of populations as part of the algorithm (Huelsenbeck and Andolfatto 2007; Huelsenbeck et al. 2011). Structure is available for download from http://pritchardlab.stanford.edu/structure.html, and Structurama from http://cteg.berkeley.edu/~structurama/. An example of the application of Structure to meiofaunal studies can be found in Tulchinsky et al. (2012), which used inter-simple-sequence-repeat (ISSR) markers to delineate marine nemertean species.

Bayesian phylogenetics & phylogeography

The Bayesian phylogenetics & phylogeography (BP&P) method uses Bayesian modelling of the multispecies coalescent to generate the posterior probabilities of species assignments (Yang and Rannala 2010; Rannala and Yang 2013). It accounts for uncertainties in gene tree reconstruction and, unlike tree-based single-locus methods, is designed to deal with non-monophyletic species arising from incomplete lineage sorting. The input of the method consists of multiple gene trees, but in the classical usage of the method it also needs a user-specified guide tree to avoid integrating over all possible species delimitations; however, a new version of the program BP&P was recently published (Yang and Rannala 2014) that is able to delineate species in an unguided way. A program implementing this method is available from Ziheng Yang's webpage (BP&P: http://abacus.gene.ucl.ac.uk/software.html). BP&P has already been used in meiofaunal studies: the original description of the method used a rotifer data set to estimate its performance (Yang and Rannala 2010), and more recently, it was applied to molluscs (Jörger et al. 2012) and to several species complexes of nemertodermatids (Meyer-Wachsmuth et al. 2014).

SpedeSTEM

SpedeSTEM (Ence and Carstens 2011) uses a maximum-likelihood approach to perform species delimitation using STEM (species tree estimation; Kubatko et al. 2009). This approach calculates the probability of different models containing various numbers of evolutionary lineages, and then ranks these models according to information theory criteria. The inputs of SpedeSTEM are single gene trees (no global guide topology is needed). According to simulations performed by the authors of the method, SpedeSTEM can work using as little as five loci, but this method has yet to be applied to meiofaunal species. A program implementing this approach can be downloaded from https://spedestem.osu.edu/.

O'Meara's heuristic search

O'Meara (2010) introduced two methods (one parametric and the other non-parametric) with heuristic search strategies to delimit species using multiple trees from individual genetic

markers as input. The parametric method, also called "KC delimitation", seeks to find the delimited species tree that maximises the probability of the gene trees. The non-parametric method quantifies two metrics, called "gene tree conflict" and "excess structure", in order to minimise their costs. According to the original paper, the non-parametric method performs better, albeit inconsistencies can be found. These analyses are implemented in Brownie 2.0 (http://www.brianomeara.info/brownie), but have yet to be applied to meiofaunal studies.

New Bayesian methods using the multispecies coalescent

The multispecies coalescent was already used to obtain species trees from gene trees on multilocus data sets by approaches such as BEST (Liu et al. 2008) and *BEAST (Heled and Drummond 2010), and was successfully applied to species delimitation in BP&P (Yang and Rannala 2010). Other recent methods are continuously appearing: three articles using Bayes Factor Delimitation (BFD; Grummer et al. 2014), Bayes Factor Delimitation with genomic data (BFD*; Leaché et al. 2014) and an assignment-free Bayesian method for species discovery called DISSECT (Jones et al. 2014) were published in the last few months. Approximate Bayesian computation (ABC) can also be used to answer several topics in eco-evolutionary studies, and it has already been applied to species delimitation in animals (Camargo et al. 2012). As these methods are all very new, it is too early to write about their respective pros and cons, but this profusion of new Bayesian approaches using the multispecies coalescent suggests that other very significant advances in this field are likely to come out in the near future as well.

Caveats and perspectives

Herein we have described popular metrics and techniques that taxonomists and non-taxonomists alike can use to obtain DNA-based working hypotheses regarding species boundaries. These methods offer effective species proxies that are quick, easy to implement, and relatively robust when the assumptions of the methods used to obtain them are met by the data sets on which they are applied. Hence, DNA taxonomy is a useful springboard to gauge the diversity of groups where morphological studies are painstakingly difficult and/or where the number of species far outweighs the availability of taxonomists to investigate them, as is generally the case for the meiofauna. Still, all the methods described here should be used with caution. It is generally accepted that DNA taxonomy is not a substitute for taxonomic descriptions (Wiens and Servedio 2000; Sites and Marshall 2004), and indeed, most DNA taxonomy methods yield only primary species hypotheses (PSHs) that require further testing and validation (Puillandre et al. 2012b; Pante et al. 2015). Because different methods can provide different conclusions, it is advisable to use several approaches, look at the congruence between the results obtained from each of them (Carstens et al. 2013), and try to understand the reasons for the observed incongruences. In an optimal scenario, one may adopt an integrative, iterative approach including genetics, morphology, ecology, behaviour, geography, as well as other sources of data to support species identities (Padial et al. 2010; Schlick-Steiner et al. 2010). The drawback of including so many approaches, however, is that some level of incongruence between them will show up in most cases, and there is then no obvious, objective way to decide which results to trust and which ones to discard. This is particularly likely when dealing with meiofaunal species, since their potentially large population sizes and dispersal abilities makes them prone to incomplete lineage sorting (Rossberg et al. 2013). On the bright side, meiofaunal species that are easy to collect have often been used as test data sets when proposing new methods for species delimitation, and it is likely that this trend will continue in the future, keeping meiofaunal studies at the cutting edge of DNA taxonomy.

Acknowledgments We thank Daniela Zappilli for organising a workshop on meiofaunal studies in Brest and for inviting us to write this manuscript. We thank also Timothy G. Barraclough, William C. Birky Jr, Bryan Carstens, Simon Dellicour, Christophe Douady, Florian Malard, Nicolas Puillandre, Fabio Stoch, Ziheng Yang, and two anonymous reviewers for comments and suggestions. J.-F.F. is supported by the European Research Council (ERC-2012-AdG 322790).

References

Adjeroud M, Guérécheau A, Vidal-Dupiol J et al (2014) Genetic diversity, clonality and connectivity in the scleractinian coral *Pocillopora damicornis*: a multi-scale analysis in an insular, fragmented reef system. Mar Biol 161:531–541

Adolfsson S, Michalakis Y, Paczesniak D et al (2010) Evaluation of elevated ploidy and asexual reproduction as alternative explanations for geographic parthenogenesis in *Eucypris virens* ostracods. Evolution 64:986–997

Avise JC, Arnold J, Ball RM et al (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annu Rev Ecol Syst 18:489–522

Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48

Baretta-Bekker JG, Duursma EK, Kuipers BR (eds) (1998) Encyclopedia of marine sciences. Springer, Berlin, pp 1–357

Bell DA, DeMarini DM (1991) Excessive cycling converts PCR products to random-length higher molecular weight fragments. Nucleic Acids Res 19:5079

Bellati A, Carranza S, Garcia-Porta J, Fasola M, Sindaco R (2015) Cryptic diversity within the *Anatololacerta* species complex (Squamata: Lacertidae) in the Anatolian Peninsula: evidence from a multi-locus approach. Mol Phylogenet Evol 82:219–233

Benson DA, Cavanaugh M, Clark K et al (2013) GenBank. Nucleic Acids Res 41:D36–D42

Bergsten J, Bilton DT, Fujisawa T et al (2012) The effect of geographical scale of sampling on DNA barcoding. Syst Biol 61:851–869

Bik HM, Sung W, De Ley P et al (2012) Metagenetic community analysis of microbial eukaryotes illuminates biogeographic patterns in deep-sea and shallow water sediments. Mol Ecol 21:1048–1059

Birky CW (2013) Species detection and identification in sexual organisms using population genetic theory and DNA sequences. PLoS One 8:e52544

Birky CW, Wolf C, Maughan H et al (2005) Speciation and selection without sex. Hydrobiologia 546:29–45

Birky CW, Adams J, Gemmel M, Perry J (2010) Using population genetic theory and DNA sequences for species detection and identification in asexual organisms. PLoS One 5:e10609

Birky CW, Ricci C, Melone G, Fontaneto D (2011) Integrating DNA and morphological taxonomy to describe diversity in poorly studied microscopic animals: new species of the genus *Abrochtha* Bryce, 1910 (Rotifera: Bdelloidea: Philodinavidae). Zool J Linn Soc 161:723–734

Blanco-Bercial L, Cornils A, Copley N, Bucklin A (2014) DNA barcoding of marine copepods: assessment of analytical approaches to species identification. PLoS Curr Tree Life 6: ecurrents.tol.cdf8b74881f87e3b01d56b43

Blin N, Stafford DW (1976) A general method for isolation of high molecular weight DNA from eukaryotes. Nucleic Acids Res 3: 2303–2308

Bode SNS, Adolfsson S, Lamatsch DK et al (2010) Exceptional cryptic diversity and multiple origins of parthenogenesis in a freshwater ostracod. Mol Phylogenet Evol 54:542–552

Bouckaert R, Heled J, Kühnert D et al (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput Biol 10:e1003537

Brandão SN, Sauer J, Schön I (2010) Circumantarctic distribution in Southern Ocean benthos? A genetic test using the genus *Macroscapha* (Crustacea, Ostracoda) as a model. Mol Phylogenet Evol 55:1055–1069

Bull JJ, Huelsenbeck JP, Cunningham CW, Swofford DL, Waddell PJ (1993) Partitioning and combining data in phylogenetic analysis. Syst Biol 42:384–397

Butlin R, Bridle J, Schluter D (eds) (2009) Speciation and patterns of diversity. Cambridge University Press, Cambridge, pp 1–346

Calvignac S, Konecny L, Malard F, Douady CJ (2011) Preventing the pollution of mitochondrial datasets with nuclear mitochondrial paralogs (*numts*). Mitochondrion 11:246–254

Camargo A, Morando M, Avila LJ, Sites JW (2012) Species delimitation with ABC and other coalescent-based methods: a test of accuracy with simulations and an empirical example with lizards of the *Liolaemus darwinii* complex (Squamata: Liolaemidae). Evolution 66:2834–2849

Carson H (1957) The species as a field for recombination. In: Mayr E (ed) The species problem. American Association for the Advancement of Science, Washington, pp 23–38

Carstens BC, Pelletier TA, Reid NM, Satler JD (2013) How to fail at species delimitation. Mol Ecol 22:4369–4383

Casiraghi M, Labra M, Ferri E et al (2010) DNA barcoding: a six-question tour to improve users' awareness about the method. Brief Bioinform 11:440–453

Cassens I, Mardulyn P, Milinkovitch M (2005) Evaluating intraspecific "network" construction methods using simulated sequence data: do existing algorithms outperform the global maximum parsimony approach? Syst Biol 54:363–372

Clark AG (1990) Inference of haplotypes from PCR-amplified samples of diploid populations. Mol Biol Evol 7:111–122

Collins RA, Cruickshank RH (2012) The seven deadly sins of DNA barcoding. Mol Ecol Resour 13:969–975

Collins RA, Cruickshank RH (2014) Known knowns, known unknowns, unknown unknowns and unknown knowns in DNA barcoding: a comment on Dowton et al. Syst Biol 63:1005–1009

Cornils A, Held C (2014) Evidence of cryptic and pseudocryptic speciation in the *Paracalanus parvus* species complex (Crustacea, Copepoda, Calanoida). Front Zool 11:1–17

Creer S, Fonseca VG, Porazinska DL et al (2010) Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. Mol Ecol 19: 4–20

Curini-Galletti M, Artois TJ, Delogu V et al (2012) Patterns of diversity in soft-bodied meiofauna: dispersal ability and body size matter. PLoS One 7:e33801

Davey JL, Blaxter ML (2010) RADseq: next-generation population genetics. Brief Funct Genomics 9:416–423

de Queiroz K (2007) Species concepts and species delimitation. Syst Biol 56:879–886

Deiner K, Walser J-C, Mächler E, Altermatt F (2015) Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. Biol Conserv 183:53–63

Dellicour S, Flot J-F (2015) Delimiting species-poor datasets using single molecular markers: a study of barcode gaps, haplowebs and GMYC. Syst Biol

Dover G (1982) Molecular drive: a cohesive mode of species evolution. Nature 299:111–117

Dowton M, Meiklejohn K, Cameron SL, Wallman J (2014) A preliminary framework for DNA barcoding, incorporating the multispecies coalescent. Syst Biol 63:639–644

Doyle JJ (1995) The irrelevance of allele tree topologies for species delimitation, and a non-topological alternative. Syst Bot 20:574–588

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

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797

Ence DD, Carstens BC (2011) SpedeSTEM: a rapid and accurate method for species delimitation. Mol Ecol Resour 11:473–480

Esselstyn JA, Evans BJ, Sedlock JL, Anwarali Khan FA, Heaney LR (2012) Single-locus species delimitation: a test of the mixed Yule–coalescent model, with an empirical application to Philippine round-leaf bats. Proc R Soc Lond B 279:3678–3686

Estoup A, Largiadèr CR, Perrot E, Chourrout D (1996) Rapid one-tube DNA extraction for reliable PCR detection of fish polymorphic markers and transgenes. Mol Mar Biol Biotech 5:295–298

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

Ezard THG, Fujisawa T, Barraclough TG (2009) splits: SPecies' LImits by Threshold Statistics. http://R-Forge.R-project.org/projects/splits/

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587

Flot J-F (2007) CHAMPURU 1.0: a computer software for unravelling mixtures of two DNA sequences of unequal lengths. Mol Ecol Notes 7:974–977

Flot J-F (2010a) Vers une taxonomie moléculaire des amphipodes du genre *Niphargus*: exemples d'utilisation de séquences d'ADN pour l'identification des espèces. Bull Soc Sci Nat Ouest Fr 32:62–68

Flot J-F (2010b) SeqPHASE: a web tool for interconverting phase input/output files and fasta sequence alignments. Mol Ecol Resour 10: 162–166

Flot J-F, Tillier S (2007) The mitochondrial genome of *Pocillopora* (Cnidaria: Scleractinia) contains two variable regions: the putative D-loop and a novel ORF of unknown function. Gene 401:80–87

Flot J-F, Tillier A, Samadi S, Tillier S (2006) Phase determination from direct sequencing of length-variable DNA regions. Mol Ecol Notes 6:627–630

Flot J-F, Magalon H, Cruaud C et al (2008) Patterns of genetic structure among Hawaiian corals of the genus *Pocillopora* yield clusters of individuals that are compatible with morphology. C R Biol 331: 239–247

Flot J-F, Couloux A, Tillier S (2010) Haplowebs as a graphical tool for delimiting species: a revival of Doyle's "field for recombination" approach and its application to the coral genus *Pocillopora* in Clipperton. BMC Evol Biol 10:372

Flot J-F, Blanchot J, Charpy L et al (2011) Incongruence between morphotypes and genetically delimited species in the coral genus *Stylophora*: phenotypic plasticity, morphological convergence, morphological stasis or interspecific hybridization? BMC Ecol 11:22

Flot J-F, Dahl M, André C (2013) *Lophelia pertusa* corals from the Ionian and Barents seas share identical nuclear ITS2 and near-identical mitochondrial genome sequences. BMC Res Notes 6:144

Flot J-F, Bauermeister J, Brad T et al (2014) *Niphargus-Thiothrix* associations may be widespread in sulphidic groundwater ecosystems: evidence from southeastern Romania. Mol Ecol 23:1405–1417

Folmer O, Black M, Hoeh W et al (1994) DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. Mol Mar Biol Biotechnol 3:294–299

Fonseca VG, Carvalho GR, Sung W et al (2010) Second-generation environmental sequencing unmasks marine metazoan biodiversity. Nat Commun 1:98

Fonseca VG, Carvalho GR, Nichols B et al (2014) Metagenetic analysis of patterns of distribution and diversity of marine meiobenthic eukaryotes. Glob Ecol Biogeogr 23:1293–1302

Fontaneto D (2014) Molecular phylogenies as a tool to understand diversity in rotifers. Int Rev Hydrobiol 99:178–187

Fontaneto D, Herniou EA, Boschetti C et al (2007) Independently evolving species in asexual bdelloid rotifers. PLoS Biol 5:e87

Fontaneto D, Kaya M, Herniou EA, Barraclough TG (2009) Extreme levels of hidden diversity in microscopic animals (Rotifera) revealed by DNA taxonomy. Mol Phylogenet Evol 53:182–189

Fontaneto D, Iakovenko N, Eyres I et al (2011) Cryptic diversity in the genus *Adineta* Hudson & Gosse, 1886 (Rotifera: Bdelloidea: Adinetidae): a DNA taxonomy approach. Hydrobiologia 662:27–33

Fourment M, Gibbs M (2006) PATRISTIC: a program for calculating patristic distances and graphically comparing the components of genetic change. BMC Evol Biol 6:1

Freeman JL, Perry GH, Feuk L et al (2006) Copy number variation: new insights in genome diversity. Genome Res 16:949–961

Fujisawa T, Barraclough TG (2013) Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent (GMYC) approach: a revised method and evaluation on simulated datasets. Syst Biol 62:707–724

Fujita MK, Leaché AD, Burbrink FT, McGuire JA, Moritz C (2012) Coalescent-based species delimitation in an integrative taxonomy. Trends Ecol Evol 27:480–488

Funk DJ, Omland KE (2003) Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annu Rev Ecol Evol Syst 34:397–423

Garza JC, Freimer NB (1996) Homoplasy for size at microsatellite loci in humans and chimpanzees. Genome Res 6:211–217

Garza JC, Slatkin M, Freimer NB (1995) Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size. Mol Biol Evol 12:594–603

Gaston KJ, Blackburn TM (2000) Pattern and process in macroecology. Blackwell Publishing, Oxford, pp 1–377

Giere O (2009) Meiobenthology the microscopic motile fauna of aquatic sediments. 1–527

Gollner S, Fontaneto D, Arbizu PM (2011) Molecular taxonomy confirms morphological classification of deep-sea hydrothermal vent copepods (Dirivultidae) and suggests broad physiological tolerance of species and frequent dispersal along ridges. Mar Biol 158: 221–231

Grummer JA, Bryson RW, Reeder TW (2014) Species delimitation using Bayes factors: simulations and application to the *Sceloporus scalaris* species group (Squamata: Phrynosomatidae). Syst Biol 63:119–133

Hare MP, Palumbi SR (1999) The accuracy of heterozygous base calling from diploid sequence and resolution of haplotypes using allele-specific sequencing. Mol Ecol 8:1750–1752

Hausdorf B, Hennig C (2010) Species delimitation using dominant and codominant multilocus markers. Syst Biol 59:491–503

Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003) Biological identifications through DNA barcodes. Proc R Soc Lond B 270:313–321

Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM (2004) Identification of birds through DNA barcodes. PLoS Biol 2:e312

Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data. Mol Biol Evol 27:570–580

Hellberg ME (2006) No variation and low synonymous substitution rates in coral mtDNA despite high nuclear variation. BMC Evol Biol 6:24

Hey J (2009) On the arbitrary identification of real species. In: Butlin RK, Bridle J, Schluter D (eds) Speciat. Patterns divers. Cambridge University Press, Cambridge, pp 15–28

Hillis DM, Moritz C, Porter CA, Baker RJ (1991) Evidence for biased gene conversion in concerted evolution of ribosomal DNA. Science 251:308–310

Hodges E, Xuan Z, Balija V et al (2007) Genome-wide in situ exon capture for selective resequencing. Nat Genet 39:1522–1527

Huang D, Meier R, Todd PA, Chou LM (2008) Slow mitochondrial COI sequence evolution at the base of the metazoan tree and its implications for DNA barcoding. J Mol Evol 66:167–174

Huelsenbeck JP, Andolfatto P (2007) Inference of population structure under a Dirichlet process model. Genetics 175:1787–1802

Huelsenbeck JP, Andolfatto P, Huelsenbeck ET (2011) Structurama: Bayesian inference of population structure. Evol Bioinforma 2011: 55–59

Iakovenko NS, Kašparová E, Plewka M, Janko K (2013) *Otostephanos* (Rotifera, Bdelloidea, Habrotrochidae) with the description of two new species. Syst Biodivers 11:477–494

Jones M, Ghoorah A, Blaxter M (2011) jMOTU and taxonerator: turning DNA barcode sequences into annotated operational taxonomic units. PLoS ONE 6(4):e19259

Jones G, Aydin Z, Oxelman B (2014) DISSECT: an assignment-free Bayesian discovery method for species delimitation under the multispecies coalescent. Bioinformatics. doi:10.1093/bioinformatics/btu770

Jörger KM, Norenburg JL, Wilson NG, Schrödl M (2012) Barcoding against a paradox? Combined molecular species delineations reveal multiple cryptic lineages in elusive meiofaunal sea slugs. BMC Evol Biol 12:245

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolism, vol 3. Academic, New York, pp 21–132

Kånneby T, Todaro MA, Jondelius U (2012) A phylogenetic approach to species delimitation in freshwater Gastrotricha from Sweden. Hydrobiologia 683:185–202

Katoh K, Asimenos G, Toh H (2009) Multiple alignment of DNA sequences with MAFFT. Methods Mol Biol 537:39–64

Kekkonen M, Hebert PDN (2014) DNA barcode-based delineation of putative species: efficient start for taxonomic workflows. Mol Ecol Resour 14:706–715

Kieneke A, Martínez Arbizu PM, Fontaneto D (2012) Spatially structured populations with a low level of cryptic diversity in European marine Gastrotricha. Mol Ecol 21:1239–1254

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

Kircher M, Kelso J (2010) High-throughput DNA sequencing – concepts and limitations. Bioessays 32:524–536

Knauth S, Schmidt H, Tippkötter R (2013) Comparison of commercial kits for the extraction of DNA from paddy soils. Lett Appl Microbiol 56:222–228

Kornobis E, Pálsson S (2013) The ITS region of groundwater amphipods: length, secondary structure and phylogenetic information content in Crangonyctoids and Niphargids. J Zool Syst Evol Res 51:19–28

Koufopanou V, Burt A, Taylor JW (1997) Concordance of gene genealogies reveals reproductive isolation in the pathogenic fungus *Coccidioides immitis*. Proc Natl Acad Sci U S A 94:5478–5482

Kubatko LS, Carstens BC, Knowles LL (2009) STEM: species tree estimation using maximum likelihood for gene trees under coalescence. Bioinformatics 25:971–973

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359

Leaché AD, Fujita MK, Minin V, Bouckaert RR (2014) Species delimitation using genome-wide SNP data. Syst Biol 63:534–542

Leasi F, Norenburg JL (2014) The necessity of DNA taxonomy to reveal cryptic diversity and spatial distribution of meiofauna, with a focus on Nemertea. PLoS One 9:e104385

Leasi F, Tang CQ, De Smet WH, Fontaneto D (2013) Cryptic diversity with wide salinity tolerance in the putative euryhaline *Testudinella clypeata* (Rotifera, Monogononta). Zool J Linn Soc 168:17–28

Lefébure T, Douady CJ, Gouy M, Gibert J (2006) Relationship between morphological taxonomy and molecular divergence within Crustacea: proposal of a molecular threshold to help species delimitation. Mol Phylogenet Evol 40:435–447

Li X (2012) Molecular and evolutionary insights into sexual marine mammals and asexual bdelloid rotifers. PhD thesis: University of Namur 1–181

Li H, Handsaker B, Wysoker A et al (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079

Lim GS, Balke M, Meier R (2012) Determining species boundaries in a world full of rarity: singletons, species delimitation methods. Syst Biol 61:165–169

Liu L, Pearl DK, Brumfield RT, Edwards SV (2008) Estimating species trees using multiple-allele DNA sequence data. Evolution 62:2080–2091

Lohse K (2009) Can mtDNA barcodes be used to delimit species? A response to Pons et al. (2006). Syst Biol 58:439–442

Lorion J, Buge B, Cruaud C, Samadi S (2010) New insights into diversity and evolution of deep-sea Mytilidae (Mollusca: Bivalvia). Mol Phylogenet Evol 57:71–83

Maddison WP, Maddison DR (2014) Mesquite: a modular system for evolutionary analysis. Version 3.01 http://mesquiteproject.org

Magurran AE, Henderson PA (2003) Explaining the excess of rare species in natural species abundance distributions. Nature 422:714–716

Malekzadeh-Viayeh R, Pak-Tarmani R, Rostamkhani N, Fontaneto D (2014) Diversity of the rotifer *Brachionus plicatilis* species complex (Rotifera: Monogononta) in Iran through integrative taxonomy. Zool J Linn Soc 170:233–244

Marrone F, Brutto S, Lo AM (2010) Molecular evidence for the presence of cryptic evolutionary lineages in the freshwater copepod genus *Hemidiaptomus* G.O. Sars, 1903 (Calanoida, Diaptomidae). Hydrobiologia 644:115–125

Marrone F, Lo Brutto S, Hundsdoerfer AK, Arculeo M (2013) Overlooked cryptic endemism in copepods: systematics and natural history of the calanoid subgenus *Occidodiaptomus* Borutzky 1991 (Copepoda, Calanoida, Diaptomidae). Mol Phylogenet Evol 66:190–202

Martens K, Halse S, Schön I (2012) Nine new species of *Bennelongia* De Deckker & McKenzie, 1981 (Crustacea, Ostracoda) from Western Australia, with the description of a new subfamily. Eur J Taxon 8:1–56

Martens K, Halse S, Schön I (2013) On the *Bennelongia barangaroo* lineage (Crustacea, Ostracoda) in Western Australia, with the description of seven new species. Eur J Taxon 66:1–59

Mayr E (1982) The growth of biological thought diversity, evolution and inheritance. Belknap Press of Harvard University Press, Cambridge, pp 1–974

McCormack JE, Hird SM, Zellmer AJ et al (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. Mol Phylogenet Evol 66:526–538

Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. PLoS Biol 3:e422

Meyer M, Stenzel U, Hofreiter M (2008) Parallel tagged sequencing on the 454 platform. Nat Protoc 3:267–278

Meyer-Wachsmuth I, Curini-Galletti M, Jondelius U (2014) Hypercryptic marine meiofauna: species complexes in Nemertodermatida. PLoS One 9:e107688

Miller JT, Spooner DM (1999) Collapse of species boundaries in the wild potato *Solanum brevicaule* complex (Solanaceae, S. sect. Petota): molecular data. Plant Syst Evol 214:103–130

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

Montero-Pau J, Gómez A, Muñoz J (2008) Application of an inexpensive and high-throughput genomic DNA extraction method for the molecular ecology of zooplanktonic diapausing eggs. Limnol Oceanogr Methods 6:218–222

Moore WS (1995) Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. Evolution 49:718–726

Morgan MJ, Bass D, Bik HM et al (2014) A critique of Rossberg et al.: noise obscures the genetic signal of meiobiotal ecospecies in ecogenomic datasets. Proc R Soc Lond B 281:20133076

O'Meara BC (2010) New heuristic methods for joint species delimitation and species tree inference. Syst Biol 59:59–73

Obertegger U, Fontaneto D, Flaim G (2012) Using DNA taxonomy to investigate the ecological determinants of plankton diversity: explaining the occurrence of *Synchaeta* spp. (Rotifera, Monogononta) in mountain lakes. Freshw Biol 57:1545–1553

Obertegger U, Flaim G, Fontaneto D (2014) Cryptic diversity within the rotifer *Polyarthra dolichoptera* along an altitudinal gradient. Freshw Biol 59:2413–2427

Padial JM, Miralles A, De la Riva I, Vences M (2010) The integrative future of taxonomy. Front Zool 7:16

Palumbi S, Baker C (1994) Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. Mol Biol Evol 11:426–435

Pante E, Puillandre N, Viricel A et al. (2015) Species are hypotheses: avoid connectivity assessments based on pillars of sand. Mol Ecol 24:525–544

Papadopoulou A, Bergsten J, Fujisawa T et al (2008) Speciation and DNA barcodes: testing the effects of dispersal on the formation of discrete sequence clusters. Philos Trans R Soc Lond B Biol Sci 363:2987–2996

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

Pont-Kingdon GA, Okada NA, Macfarlane JL et al (1995) A coral mitochondrial *mutS* gene. Nature 375:109–111

Posada D (2008) jModelTest: phylogenetic model averaging. Mol Biol Evol 25:1253–1256

Posada D, Crandall KA (2001) Intraspecific gene genealogies: trees grafting into networks. Trends Ecol Evol 16:37–45

Powell JR (2012) Accounting for uncertainty in species delineation during the analysis of environmental DNA sequence data. Methods Ecol Evol 3:1–11

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Prosser SWJ, Martínez-Arce A, Elías-Gutiérrez M (2013) A new set of primers for COI amplification from freshwater microcrustaceans. Mol Ecol Resour 13:1151–1155

Puillandre N, Lambert A, Brouillet S, Achaz G (2012a) ABGD, Automatic Barcode Gap Discovery for primary species delimitation. Mol Ecol 21:1864–1877

Puillandre N, Modica MV, Zhang Y et al (2012b) Large-scale species delimitation method for hyperdiverse groups. Mol Ecol 21:2671–2691

Rameckers J, Hummel S, Herrmann B (1997) How many cycles does a PCR need? Determinations of cycle numbers depending on the number of targets and the reaction efficiency factor. Naturwissenschaften 84:259–262

Rannala B, Yang Z (2013) Improved reversible jump algorithms for Bayesian species delimitation. Genetics 194:245–253

Rasmussen RS, Morrissey MT, Hebert PDN (2009) DNA barcoding of commercially important salmon and trout species (*Oncorhynchus* and *Salmo*) from North America. J Agric Food Chem 57:8379–8385

Ratnasingham S, Hebert PDN (2007) BOLD: The barcode of life data system (www.barcodinglife.org). Mol Ecol Notes 7:355–364

Ratnasingham S, Hebert PDN (2013) A DNA-based registry for all animal species: the Barcode Index Number (BIN) system. PLoS One 8: e66213

Rees HC, Maddison BC, Middleditch DJ, Patmore JRM, Gough KC (2014) The detection of aquatic animal species using environmental DNA – a review of eDNA as a survey tool in ecology. J Appl Ecol 51:1450–1459

Reid NM, Carstens BC (2012) Phylogenetic estimation error can decrease the accuracy of species delimitation: a Bayesian implementation of the general mixed Yule-coalescent model. BMC Evol Biol 12:196

Richly E, Leister D (2004) NUMTs in sequenced eukaryotic genomes. Mol Biol Evol 21:1081–1084

Rodman JE, Cody JH (2003) The taxonomic impediment overcome: NSF's Partnerships for Enhancing Expertise in Taxonomy (PEET) as a model. Syst Biol 52:428–435

Ross HA (2014) The incidence of species-level paraphyly in animals: a re-assessment. Mol Phylogenet Evol 76:10–17

Rossberg AG, Rogers T, McKane AJ (2013) Are there species smaller than 1 mm? Proc R Soc Lond B 280:20131248

Rossberg AG, Rogers T, Mckane AJ (2014) Current noise-removal methods can create false signals in ecogenomic data. Proc R Soc Lond B 281:20140191

Roux KH (2009) Optimization and troubleshooting in PCR. Cold Spring Harb Protoc

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132:365–386

R Core Team (2014) R: A language and environment for statistical computing. R Core Team. R Foundation for Statistical Computing, Vienna

Saiki RK, Gelfand DH, Stoffel S et al (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487–491

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci 74:5463–5467

Schlick-Steiner BC, Steiner FM, Seifert B et al (2010) Integrative taxonomy: a multisource approach to exploring biodiversity. Annu Rev Entomol 55:421–438

Schmidt BR, Kéry M, Ursenbacher S et al (2013) Site occupancy models in the analysis of environmental DNA presence/absence surveys: a case study of an emerging amphibian pathogen. Methods Ecol Evol 4:646–653

Schmidt-Roach S, Miller KJ, Lundgren P, Andreakis N (2014) With eyes wide open: a revision of species within and closely related to the *Pocillopora damicornis* species complex (Scleractinia; Pocilloporidae) using morphology and genetics. Zool J Linn Soc 170:1–33

Schön I, Pinto RL, Halse S et al (2012) Cryptic species in putative ancient asexual Darwinulids (Crustacea, Ostracoda). PLoS One 7:e39844

Shearer TL, Coffroth MA (2008) Barcoding corals: limited by interspecific divergence, not intraspecific variation. Mol Ecol Resour 8:247–255

Shearer TL, Van Oppen MJH, Romano SL, Wörheide G (2002) Slow mitochondrial DNA sequence evolution in the Anthozoa (Cnidaria). Mol Ecol 11:2475–2487

Shearn R, Koenders A, Halse S et al (2012) A review of *Bennelongia* De Deckker & Mckenzie, 1981 (Crustacea, Ostracoda) species from eastern Australia, with the description of three new species. Eur J Taxon 25:1–35

Simpson GG (1951) The species concept. Evolution 5:285–298

Sites JW, Marshall JC (2003) Delimiting species: a Renaissance issue in systematic biology. Trends Ecol Evol 18:462–470

Sites JW, Marshall JC (2004) Operational criteria for delimiting species. Annu Rev Ecol Evol Syst 35:199–227

Sluys R, Solà E, Gritzalis K et al (2013) Integrative delineation of species of Mediterranean freshwater planarians (Platyhelminthes: Tricladida: Dugesiidae). Zool J Linn Soc 169:523–547

Sonet G, Jordaens K, Nagy ZT et al (2013) Adhoc: an R package to calculate ad hoc distance thresholds for DNA barcoding identification. Zookeys 365:329–335

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

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68:978–989

Stucky BJ (2012) SeqTrace: a graphical tool for rapidly processing DNA sequencing chromatograms. J Biomol Tech 23:90–93

Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. Mol Ecol 21:2045–2050

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. Genetics 105:437–460

Talavera G, Dinca V, Vila R (2013) Factors affecting species delimitations with the GMYC model: insights from a butterfly survey. Methods Ecol Evol 4:1101–1110

Tamura K, Stecher G, Peterson D et al (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729

Tang CQ, Leasi F, Obertegger U et al (2012) The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna. Proc Natl Acad Sci U S A 109:16208–16212

Tang CQ, Humphreys A, Fontaneto D, Barraclough TG (2014a) Effects of phylogenetic reconstruction method on the robustness of species delimitation using single locus data. Methods Ecol Evol 5:1086–1094

Tang CQ, Obertegger U, Fontaneto D, Barraclough TG (2014b) Sexual species are separated by larger genetic gaps than asexual species in rotifers. Evolution 68:2901–2916

Tautz D, Arctander P, Minelli A et al (2003) A plea for DNA taxonomy. Trends Ecol Evol 18:70–74

Tewhey R, Warner JB, Nakano M et al (2009) Microdroplet-based PCR enrichment for large-scale targeted sequencing. Nat Biotechnol 27: 1025–1031

Todd PA (2008) Morphological plasticity in scleractinian corals. Biol Rev 83:315–337

Truett GE, Heeger P, Mynatt RL et al (2000) Preparation of PCR-quality mouse genomic DNA with hot sodium hydroxide and Tris (HotSHOT). Biotechniques 29:52–54

Tulchinsky AY, Norenburg JL, Turbeville JM (2012) Phylogeography of the marine interstitial nemertean *Ototyphlonemertes parmula*

(Nemertea, Hoplonemertea) reveals cryptic diversity and high dispersal potential. Mar Biol 159:661–674

Van Tassell CP, Smith TPL, Matukumalli LK et al (2008) SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nat Methods 5:247–252

Velasco-Castrillón A, Page TJ, Gibson JAE, Stevens MI (2014) Surprisingly high levels of biodiversity and endemism amongst Antarctic rotifers uncovered with mitochondrial DNA. Biodiversity 15:130–142

Verovnik R, Sket B, Trontelj P (2005) The colonization of Europe by the freshwater crustacean *Asellus aquaticus* (Crustacea: Isopoda) proceeded from ancient refugia and was directed by habitat connectivity. Mol Ecol 14:4355–4369

Vogler AP, Monaghan MT (2007) Recent advances in DNA taxonomy. J Zool Syst Evol Res 45:1–10

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63

Webb KE, Barnes DKA, Clark MS, Bowden DA (2006) DNA barcoding: a molecular tool to identify Antarctic marine larvae. Deep Sea Res II Top Stud Oceanogr 53:1053–1060

Wiemers M, Fiedler K (2007) Does the DNA barcoding gap exist?—a case study in blue butterflies (Lepidoptera: Lycaenidae). Front Zool 4:8

Wiens JJ (2007) Species delimitation: new approaches for discovering diversity. Syst Biol 56:875–878

Wiens JJ, Servedio MR (2000) Species delimitation in systematics: inferring diagnostic differences between species. Proc Biol Sci 267:631–636

Williams S, Apte D, Ozawa T, Kaligis F, Nakano T (2011) Speciation and dispersal along continental coastlines and island arcs in the Indo-West Pacific turbinid gastropod genus *Lunella*. Evolution 65:1752–1771

Winship PR (1989) An improved method for directly sequencing PCR amplified material using dimethyl sulphoxide. Nucleic Acids Res 17:1266

Yang Z, Rannala B (2010) Bayesian species delimitation using multilocus sequence data. Proc Natl Acad Sci U S A 107:9264–9269

Yang Z, Rannala B (2014) Unguided species delimitation using DNA sequence data from multiple loci. Mol Biol Evol 31:3125–3135

Zeppilli D, Sarrazin J, Leduc D et al. (2015) Meiofauna as model organisms to assess global change in marine ecosystems. Mar Biodivers

Zhan A, MacIsaac HJ (2015) Rare biosphere exploration using high-throughput sequencing: research progress and perspectives. Conserv Genet

Zhang J, Kapli P, Pavlidis P, Stamatakis A (2013) A general species delimitation method with applications to phylogenetic placements. Bioinformatics 29:2869–2876