

Nomenclature for the Nameless: A Proposal for an Integrative Molecular Taxonomy of Cryptic Diversity Exemplified by Planktonic Foraminifera

RAPHAËL MORARD^{1,*}, GILLES ESCARGUEL², AGNES K. M. WEINER^{1,3}, AUREOLE ANDRÉ⁴, CHRISTOPHE J. DOUADY^{2,5}, CHRISTOPHER M. WADE⁶, KATE F. DARLING^{7,8}, YURIKA UJIIÉ⁹, HEIDI A. SEARS¹⁰, FRÉDÉRIC QUILLÉVÈRE¹¹, THIBAUT DE GARIDEL-THORON¹², COLOMBAN DE VARGAS^{13,14}, AND MICHAL KUCERA¹

¹MARUM Center for Marine Environmental Sciences, University of Bremen, Leobener Strasse, 28359 Bremen, Germany, ²Université de Lyon; UMR5023 Ecologie des Hydrosystèmes Naturels et Anthropisés; Université Lyon 1; ENTPE; CNRS; 6 rue Raphaël Dubois, 69622 Villeurbanne, France, ³Japan Agency for Marine Earth Science and Technology (JAMSTEC), 2-15 Natsushima-cho, Yokosuka 237-0061, Kanagawa, Japan, ⁴Université de Reims-Champagne-Ardenne, UFR Sciences Exactes et Naturelles, Campus Moulin de la Housse, Batiment 18, 51100 REIMS, France, ⁵Institut Universitaire de France, 103 Boulevard Saint-Michel, 75005 Paris, France, ⁶School of Life Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, UK, ⁷School of GeoSciences, University of Edinburgh, Edinburgh EH9 3JW, UK, ⁸School of Geography and GeoSciences, University of St Andrews, Fife KY16 9AL, UK, ⁹Department of Biology, Shinshu University, Asahi3-1-1, Matsumoto, Nagano 390-8621, Japan, ¹⁰Department of Biology, Gilmer Hall, University of Virginia, 485 McCormick Road, Charlottesville, VA 22904, USA, ¹¹Univ Lyon, Université Lyon 1, ENS de Lyon, CNRS, UMR 5276 LGL-TPE, F-69622 Villeurbanne, France; ¹²Centre Européen de Recherche et d'Enseignement de Géosciences de l'Environnement, Centre National de la Recherche Scientifique, et Aix-Marseille Université, Aix-en-Provence, France, ¹³Centre National de la Recherche Scientifique, UMR 7144, EPEP, Station Biologique de Roscoff, 29680 Roscoff, France, and ¹⁴Sorbonne Universités, UPMC Univ Paris 06, UMR 7144, Station Biologique de Roscoff, 29680 Roscoff, France
*Correspondence to be sent to: MARUM Center for Marine Environmental Sciences, University of Bremen, Leobener Strasse, 28359 Bremen, Germany; E-mail: rmorard@marum.de.

Received 1 October 2015; reviews returned 1 December 2015; accepted 4 April 2016
Associate Editor: Adrian Paterson

Abstract.—Investigations of biodiversity, biogeography, and ecological processes rely on the identification of “species” as biologically significant, natural units of evolution. In this context, morphotaxonomy only provides an adequate level of resolution if reproductive isolation matches morphological divergence. In many groups of organisms, morphologically defined species often disguise considerable genetic diversity, which may be indicative of the existence of cryptic species. The diversity hidden by morphological species can be disentangled through genetic surveys, which also provide access to data on the ecological distribution of genetically circumscribed units. These units can be identified by unique DNA sequence motifs and allow studies of evolutionary and ecological processes at different levels of divergence. However, the nomenclature of genetically circumscribed units within morphological species is not regulated and lacks stability. This represents a major obstacle to efforts to synthesize and communicate data on genetic diversity for multiple stakeholders. We have been confronted with such an obstacle in our work on planktonic foraminifera, where the stakeholder community is particularly diverse, involving geochemists, paleoceanographers, paleontologists, and biologists, and the lack of stable nomenclature beyond the level of formal morphospecies prevents effective transfer of knowledge. To circumvent this problem, we have designed a stable, reproducible, and flexible nomenclature system for genetically circumscribed units, analogous to the principles of a formal nomenclature system. Our system is based on the definition of unique DNA sequence motifs collocated within an individual, their typification (in analogy with holotypes), utilization of their hierarchical phylogenetic structure to define levels of divergence below that of the morphospecies, and a set of nomenclature rules assuring stability. The resulting molecular operational taxonomic units remain outside the domain of current nomenclature codes, but are linked to formal morphospecies as regulated by the codes. Subsequently, we show how this system can be applied to classify genetically defined units using the SSU rDNA marker in planktonic foraminifera and we highlight its potential use for other groups of organisms where similarly high levels of connectivity between molecular and formal taxonomies can be achieved. [Cryptic species; genetic diversity; planktonic foraminifera; molecular nomenclature; MOTUs.]

The Lack of Formal Taxonomy for Molecular Diversity

For over two centuries, formalized and codified nomenclature has been the cornerstone of biological research. By using simple and efficient rules for the classification of organisms into hierarchically structured units, Linnean systematics provided biologists with a common language for taxa and their names. As a consequence, the results of fundamental taxonomic research (description of taxa) could be applied in other disciplines such as evolutionary biology,

biogeography, ecology, and conservation biology. Over the last 250 years, a large and still growing body of taxonomic knowledge has been created, allowing scientists to synthesize data accumulated by generations of taxonomists across the globe (Puillandre et al. 2012a).

Current codes for biological nomenclature (ICZN 1999; ICN 2011; ICTV 2011; Garrity et al. 2015) require that each species must be associated with a formal description, but they do not specify the nature of characters on which the description is to be based. Thus, in theory, there is nothing that speaks against the

practice to describe species by molecular evidence alone (Jörger and Schrödl 2013). Nevertheless, such practice bears the risk of creating a parallel taxonomical universe detached from the existing body of knowledge (mostly based on morphological features) and associated with a potentially large instability of species delimitation. It would also isolate the nomenclature of extant species from fossils, and render the recognition of such entities dependent on access to molecular data. Even in an age of cheap-and-easy molecular analyses, it remains essential to allow quick and efficient classification and identification of extant or fossil organisms on the basis of field observations. Therefore, most species descriptions rely on phenotypic characters. Despite its practical value and stability, the traditional *de facto* phenotypically based taxonomy is challenged by the discovery that molecular data often provide greater taxonomic resolution than morphology (Bickford et al. 2007). The existence of cryptic (i.e., phenotypically unseen) species by itself does not question the validity of the current formal taxonomy. It does, however, affect the biological interpretation of phenotypically defined species and it implies that even for extant species, phenotypic taxonomy will never be comprehensive at the level of biological (reproductive) species.

A common alternative is to use a transitional system pending formal taxonomy by defining molecular operational taxonomic units (MOTUs) using DNA-based automatic delimitation (Flot 2015). However, unless such MOTUs are treated as formal taxa as for the MARine STRamenopiles "MAST" which have been first described from SSU rDNA sequencing (Massana et al. 2004), no formal rules regulate their denomination. The PhyloCode has been proposed as an alternative to the Linnean system to name phylogenetic clades (de Queiroz and Cantino 2001) but it provides names to ranks above the species level and is therefore not suitable for a nomenclature of cryptic species. The Barcoding Index Number (BIN) system (Ratnasingham and Hebert 2013) has been proposed to register all entries in the Barcode of Life Data (BOLD) system portal (Ratnasingham and Hebert 2007). The BIN system is meant to resolve to the biological species level and has been conceived to avoid synonymy in the BOLD portal. It goes some way toward stabilization of names and disambiguation of conflict, but because the system only classifies species it lacks a hierarchical structure and shares the limitations incurred by using a universal barcode marker. At present, none of the existing systems can effectively incorporate multiple hierarchies of MOTUs nested below the level of formally described species. As a result, contrary to taxa in the current formal nomenclature, it is difficult to discuss accurately the identity, occurrence, and properties of MOTUs, including potential cryptic species (Pante et al. 2014). Cryptic species, even when detected, often remain nameless. Lack of stable names of MOTUs impedes transfer of knowledge across disciplines. As a result, MOTUs are often excluded from conservation research and efforts and the lack of a stable nomenclature of

MOTUs prevents a deeper understanding of speciation processes, biogeography, and ecological interactions at the level of cryptic species.

The modalities for the development of an integrative framework to include DNA sequence motifs into taxa delimitation are currently hotly debated in the literature (Padial et al. 2010; Carstens et al. 2013; Miralles and Vences 2013; Leliaert et al. 2014; Pante et al. 2014; Flot 2015), fueled by the advent of massive environmental sequencing surveys (e. g., Logares et al. 2014; de Vargas et al. 2015). Much effort has been directed toward the evaluation of methods and concepts to delimit entities by molecular markers (Satler et al. 2013; Leavitt et al. 2015) and toward the stability of the boundaries of the circumscribed MOTUs (Carstens et al. 2013). In contrast, there have been few if any attempts to deal with the nomenclature of the resulting units (Caron 2013). The current practice is to design for each study an *ad hoc* classification scheme with arbitrary labels and levels, which are often not conserved across studies (Pante et al. 2015).

In summary, the advent of molecular ecology has added a layer of cryptic diversity to the classically described morphospecies. MOTUs often correspond better to biological (reproductive) species concepts (Amato et al. 2007) than morphologically defined OTUs, providing more appropriate units to assess ecological and evolutionary hypotheses and concepts. However, no attempt has been made to provide a nomenclature system for them, an unescapable formal prerequisite to provide a common language for discussion across disciplines and novel data sets. Such a system should be able to name MOTUs below the level of the morphospecies and should meet the criteria of an interim taxonomic system: uniqueness of the names employed, stability, and a proper definition of the given ranks (Schindel and Miller 2009). This system should be employed when the Linnean system is impractical or cannot be used. It has to provide for regular revision and integrate novel concepts in molecular taxonomy and/or automated delimitation. Finally, it should facilitate the transfer of a given MOTU from this interim system to the formal nomenclature, as soon as it is diagnosed with morphological characteristics as recently exemplified with copepods (Karanovic et al. 2016).

In attempting to export the knowledge of molecular ecology of extant planktonic foraminifera to their fossil record, we have been repeatedly confronted with both major impediments resulting from the lack of MOTU nomenclature: stability and transfer. The spatial and ecological distributions of cryptic species of extant planktonic foraminifera are narrower compared with the range of their morphospecies and the information on their present diversity and distribution could improve the accuracy of paleoceanographic reconstruction (Kucera and Darling 2002; Morard et al. 2013). Benefiting from the existence of a comprehensive morphological taxonomy of this group linked through single-cell ribosomal DNA sequencing surveys to genetic diversity (Morard et al. 2015), we are now in the position

to propose a simple, robust, and efficient nomenclature system that gives stable names (labels) to MOTUs, and links them to the formal nomenclature without compromising its rules. The purpose of this article is to describe and explain the design of the system, show how it can be applied to foraminifera, and highlight its potential for taxa with a similar degree and structure of cryptic diversity.

FROM A UNIFIED SPECIES CONCEPT TO A FORMAL NOMENCLATURE SYSTEM FOR MOTUS

Since Mayr (1942), species are viewed as groups of actually or potentially interbreeding organisms, which are reproductively isolated from each other. When integrated over generations, such reproductively defined groups constitute evolutionary lineages *sensu* Simpson (1961), that is, ancestor-descendant lineages of organisms evolving conjointly on the hereditary ground. de Queiroz (2007) advocated a unified species concept where species are treated as separately evolving metapopulation lineages, making their mutual separation and unique evolutionary history the only criteria necessary to justify their recognition. When applied to living organisms, this concept treats species as snapshots in time of such metapopulation lineages. During their evolution, metapopulation lineages acquire secondary properties (phenetic traits, different ecological niches, reciprocal monophyly, reproductive incompatibility, etc.) that differentiate them from other, more or less distantly related metapopulation lineages. In the course of this process they first occupy a “gray zone” where the species delimitation under different criteria may be ambiguous and only later reach a state of unambiguous separation (Fig. 1 in de Queiroz 2007). Thus, the ongoing process of evolution produces a spectrum of species that exist in different states of separation: from almost identical metapopulations to clearly distinct and isolated lineages. Because operational taxonomy typically requires separation by one or more of the secondary properties, species within the “gray zone” remain not differentiated in most cases.

By characterizing operational units through sequence divergence, molecular taxonomy ventures deeper into the “gray zone” than classical taxonomy. The degree to which MOTUs reach into the “gray zone” is a function of the resolution of the molecular marker (or markers) used for MOTU delineation. At some level, existing sequence divergence only describes the unique genome of non-clonal individuals and is of no taxonomic value. Therefore, the marker(s) on which a molecular nomenclature is based has/have to represent a balance between uniqueness and representativeness (Fig. 1a). For this reason, it is not necessary and perhaps even not useful to base species delimitation on whole-genome studies if the level of divergence among molecular taxa can be reduced to the divergence

of a small subset of informative loci (Amato et al. 2007). Ultimately, this procedure is the basis of the barcoding concept (Hebert et al. 2003), cataloging unique sequence patterns from a given informative “barcode” region. Although practical in its application, this approach has been criticized (e.g., Taylor and Harris 2012), because of the differences that may exist between gene trees and species trees. As a result, recent barcoding initiatives proposed the use of multiple loci (Pawlowski et al. 2012). The use of multilocus sequence data represents a balanced alternative combining the advantages of barcoding with a higher level of representativeness: allowing the investigation of both population genetics and cryptic speciation (e.g., Salerno et al. 2015).

Notwithstanding the choice of approach or marker selection, automated methods for MOTU boundary delineation have been developed to avoid arbitrary thresholds or subjective interpretation of phylogenetic trees. Such methods assume that the sequence divergence can be partitioned between divergence due to reproductive isolation and variation occurring within a population, attempting to define MOTUs such that they best correspond to species *sensu* de Queiroz (2007). These methods can be based on single or multilocus information. They either exploit a barcode gap (e.g., Puillandre et al. 2012b) or detect the boundary between population signal and divergence reflecting species branching by coalescent statistics (e.g., Pons et al. 2006), by mutual allelic exclusivity (Flot et al. 2010), or model speciation in terms of numbers of substitutions (Zhang et al. 2013). In the same way as taxonomists empirically do, some methods tend to lump, while others tend to split taxa (Miralles and Vences 2013; André et al. 2014; Fig. 1c). Splitting and lumping both bear advantages and caveats; overall, there are so many combinations of parameters that may be relevant to delineate species by individual methods that it appears impossible at this time to systematically favor any one of them (Carstens et al. 2013). When different methods return ambiguous delimitations, Miralles and Vences (2013) and Carstens et al. (2013) advocate the use of conservative decisions (i.e., lumping strategies), as such biases are easier to detect and subsequently correct through additional studies. This point is worth noting because lumping is always detrimental to ecological studies (Padial et al. 2010).

The final step in molecular taxonomy is the implementation of the resulting MOTUs in a stable but flexible nomenclature. Such nomenclature will have to account for multiple levels of genetic variability corresponding to the different stages of acquisition of secondary properties progressively differentiating a set of metapopulations into distinct morphospecies (Fig. 1d). This is best accounted for by a nested nomenclature. Placing MOTUs into a nested structure provides the advantage of a stable and informative evolutionary framework, which is more robust for successive revisions and implementation of new methods and gene markers. Hierarchical nomenclature

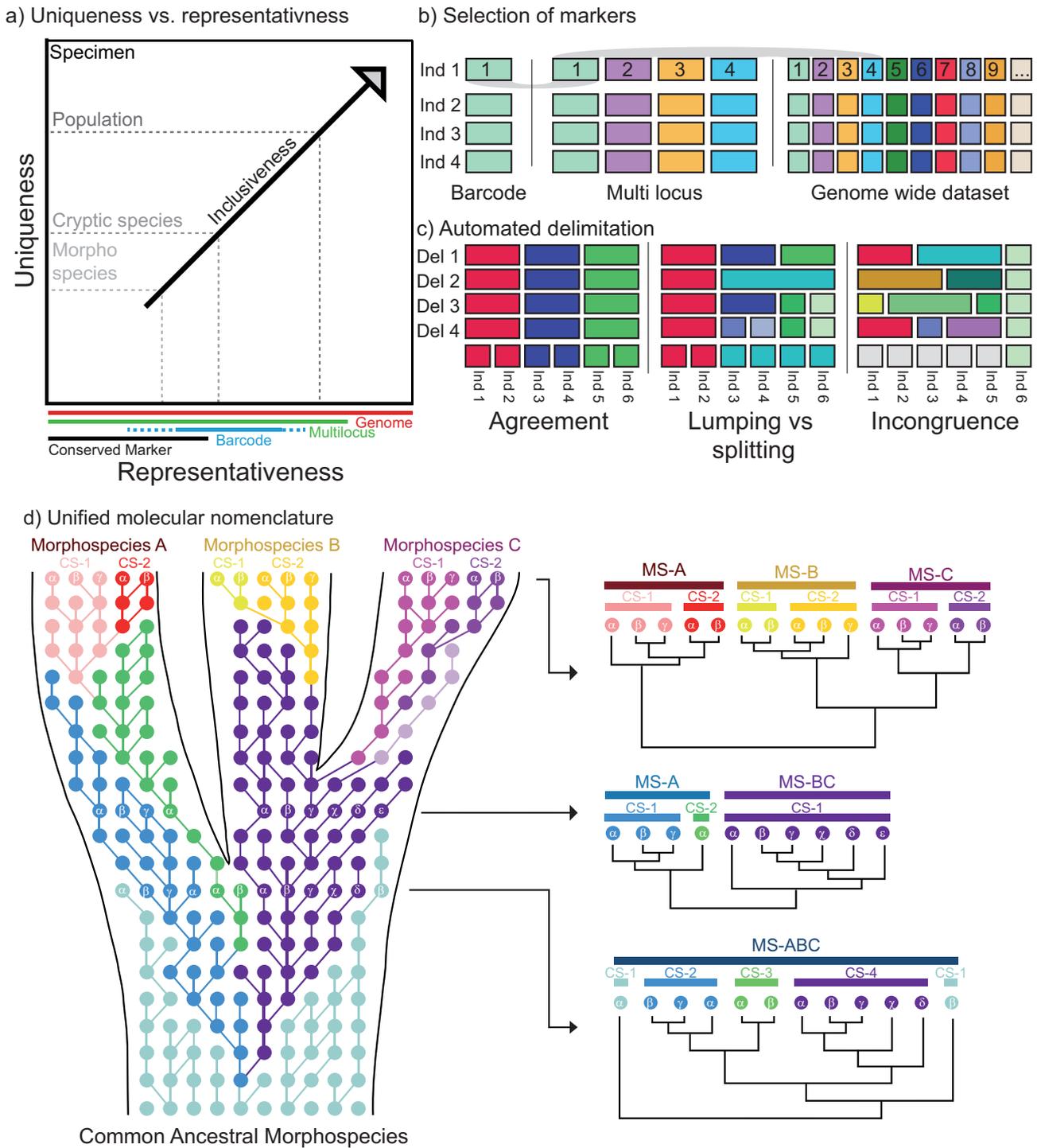


FIGURE 1. Toward a unified molecular nomenclature: Data, constraints, and consequences. a) Relationship between the potential level of resolution of molecular markers (representativeness), the biological level of uniqueness, and the inclusiveness of the resulting molecular nomenclature (increasing from morphospecies to specimens). b) Principle of contingency in MOTU delineation based on multiple markers: when the number of markers used for delineation is increased, the new markers must be linked to the initial set of markers to allow conservation of MOTU names. c) Three scenarios of changes in MOTU delineation due to new delineation methods and/or new markers (“Del” abbreviation for delimitation, “Ind” abbreviation for individual). d) An application of the proposed molecular nomenclature on a hypothetical example of an evolving lineage. MOTU delineation and nomenclature (labeling) is shown at three steps of differentiation under a coalescent process (tokogeny redrawn and modified after Leliaert et al. 2014). Dots represent metapopulations of a given cryptic species (symbolized by different color shading) which are differentiating into three morphospecies (solid black contour lines). Ancestor-descendant links are symbolized by lines between metapopulations. When the nomenclature is applied, each MOTU receives a formal label for each level below the morphospecies level (in this two-level example, an Arabic number for the cryptic species [CS] level, and a Greek letter for the metapopulation level).

can be implemented with rigid ranks or as a rank-free system (Benton 2000). A system with ranks is easier to implement in terms of having a finite number of ranks, associated with clear definition and regulated naming. On the other hand, for the same reason as the lack of a universal barcode, it seems unrealistic to prescribe the number of MOTU ranks. Thus, we propose a semiopen system with a number of ranks (levels), which is fixed within each taxon of interest but may vary among taxa. The system is bounded by morphological species on its upper level and by unique DNA substitution patterns on the lower level.

Yeates et al. (2011) proposed to make taxonomy an iterative process during which the species boundaries are tested using several lines of evidence and the process is pursued until species boundaries reach stability. Similarly, Pante et al. (2015) advocated an integrative taxonomy loop, where a Primary Species Hypothesis is proposed based on morphological or single marker delimitation, and then submitted to an integrative taxonomy loop involving the addition of new material, alternative species delimitation, or any evidence revealing the presence of a secondary species property *sensu de Queiroz* (2007). When the addition of new evidence leads to a taxonomic decision, a Secondary Species Hypothesis is proposed and then subsequently evaluated following the same process, gradually leading to a more and more stable and reliable taxonomy. However, such an iterative process can hardly be achieved within a single study, since bringing together the relevant lines of evidence requires multiple iterations. In addition, it remains unclear whether or not such a system would be applicable to cases of incipient speciation (Mittelbach et al. 2007). In any case, an efficient nomenclature system for MOTUs must be able to accommodate more or less frequent revisions.

DESCRIPTION OF THE PROPOSED ITERATIVE MOLECULAR NOMENCLATURE SYSTEM

A nomenclature system accommodating multiple lines of evidence introduced by successive studies using single or multiple markers, or even a genome-wide data set, and facilitating revisions of the Primary Species Hypotheses would provide the required connectivity between molecular and morphological taxonomy. It should be flexible enough to allow the incorporation of new evidence from automated delimitation methods and the transfer of MOTUs into formal taxonomy on the basis of morphometric analyses as well as ecological and biogeographical data. Based on the incorporation of such new evidence, the molecular nomenclature should allow the transfer of the level and nature of previously defined units without compromising the stability of names. Upon such transfer, the nomenclature should maintain the contingency of all changes and record the reasons motivating such changes.

Units and Typification

A pragmatic upper boundary for the level of inclusiveness of units in the proposed molecular taxonomy is the level of morphospecies, whose nomenclature is *de facto* regulated by formal morphotaxonomy. Conversely, the lower boundary is given by one or several specific DNA substitution patterns observed within a single marker gene (basetype). Where multiple genes are available for single individuals, each sequence pattern for each locus is a basetype. It will be essential to strive to identify all gene copies occurring within single individuals for each chosen marker (alleles, multiple copies) and use their diversity to define the lowest MOTU level—the basegroup. The basegroup is a set of basetypes, where all pairs of basetypes have been observed to co-occur within one individual. Two basegroups are considered as different as long as their basetypes are not observed to co-occur within a single individual, similar to the mutual allelic exclusivity of the haploweb of Flot et al. 2010. If only a single gene version exists for every gene marker(s) used for the nomenclature system, a basegroup will contain only one basetype. The basegroup is the basis of the nomenclature system we propose. The hierarchical nature of genetic variability between the basegroup and the morphospecies level can be used to define a number of intermediate nested levels to circumscribe units at different levels of divergence. The nature and number of such intermediate levels may differ depending on the clade under scrutiny. We stress that every level must be associated with a clear definition, and that objective delimitation methods, or clear operational criteria, should be used to circumscribe the successive levels of genetic variability.

By defining molecular units in this way, morphospecies may be (or appear) paraphyletic or even polyphyletic (Fig. 1d). This situation is likely to be common and may arise for a variety of reasons (Funk and Omland 2003). For example, paraphyly may ensue when one of several (but not all) cryptic species within a morphospecies is associated with a morphological distinction (i.e., is pseudo-cryptic) and receives its own formal name (Weiner et al. 2015), or it may reflect the existence of discrepancies between gene phylogenies during the early stages of speciation (Leliart et al. 2014). Irrespective of its origin, we consider the existence of nonholophyletic formally named morphospecies unavoidable and at least provisionally acceptable for the purpose of formal nomenclature.

As in formal nomenclature, stability in the molecular nomenclature system can only be achieved by associating MOTUs with a formal definition based on a type. A MOTU of a higher level would be typified by a unique MOTU of a lower level. Therefore, the definition of each MOTU must include a comprehensive attribution of MOTUs of a lower level, as well as a description of the data and method used to establish the distinctiveness of the molecular clade comprising the MOTUs. In case of a revision, the new definition

will have to account for continuity with existing names. The type at the level of a basetype would be a unique sequence identified by its *accession number* in a public database and a *voucher* or a label pointing unambiguously to the specimen and clone from which the sequence has been obtained. The latter is important when facing intragenomic variability, where there may be multiple basetypes found for every chosen marker(s) in the same specimen. Also, where multiple markers have been extracted from the same specimen, it will be important to establish whether the different markers have a different level of resolution or reveal conflicting signals (Fig. 1c).

The method for delineation of clades is not prescribed, but we note that when a new and promising method is developed, it should be applied to the taxa under scrutiny and associated with a comprehensive revision to achieve methodological consistency in the molecular taxonomy. In this context, addition of new data or new genes does not constitute a new method. Ultimately, the number of levels between the basegroup and the morphospecies should be tailored to the complexity of the taxonomic group of interest.

Nomenclature Rules

First, we consider the relationship between MOTUs and formal nomenclature. The first guiding principle is that MOTUs should not enter formal nomenclature, unless they can be diagnosed by some commonly accepted type of phenotypic trait(s). By diagnosis, we mean the discovery of a consistent gap in the state of one or more phenotypic characters between the MOTU of interest and all its siblings. The definition of the gap (or discontinuity) should follow the criteria typically used for morphological taxonomy of the studied group. Ecophenotypy is pervasive among animal and plant taxa. Therefore, we recommend that when a MOTU is diagnosed as a morphospecies, it should first be demonstrated that the diagnosis is applicable throughout the range and habitats of the newly named morphospecies.

In many cases, *post-hoc* assessments of cryptic genetic diversity yielded evidence for the existence of diagnostic phenotypic traits (Hebert et al. 2003), so the molecular nomenclature system must be able to accommodate a transfer of a MOTU into the formal nomenclature. Because of the often decoupled rates of molecular and phenetic divergence and the very nature of incipient speciation, the transfer would often be limited to a subset of MOTUs of a lower level within a MOTU of a higher level. In such situations, rules are needed to prevent renaming of MOTUs removed from (or left abandoned in) a morphological species and to ensure that denominations of MOTUs shifted to a morphological species will never be used again.

We propose that the *name of each MOTU* consists of the name of its parent morphospecies and a hierarchically structured denomination, without any

subspecific identifier (like subsp., var., or forma). The latter is to exclude any possibility of confusing MOTU names with formal names. Where a MOTU has to be linked to its diagnosis, its name should be associated with the publication where it was defined. An example of a MOTU name would be "*Globigerinella siphonifera* IIa1", extended to "*G. siphonifera* IIa1 *sensu* Weiner et al. (2014)" where necessary. When a hierarchical system is used, each MOTU of a lower level must be associated with exactly one MOTU of a higher level, and each MOTU of a higher level must contain at least one MOTU of a lower level.

Further, next to typification and definition of MOTUs, we borrow the *principle of priority* from formal nomenclature and establish a scheme to avoid homonymy by transfer or revision. For this, we invoke the concept of MOTU *name unavailability* in the sense of the ICZN (1999). When a new MOTU is being defined, its name must be available. Like in formal nomenclature, a delimitation of a MOTU is fixed by first description. Unlike formal nomenclature, where the species name is fixed, the names of a MOTU will change upon transfer or when their level is revised, even when their delimitation (diagnosis) remains the same.

The first case where such a situation arises is when a MOTU is transferred to the morphological species level, that is, enters formal nomenclature. In this case its name becomes unavailable, but the names of all other MOTUs remain unchanged. The MOTU that is being transferred (and its lower-level constituents) will be renamed.

Second, such a situation may arise when the concept of a basegroup is challenged by the observation of several basetypes previously assigned to different basegroups within the same individual. In this case, the new basegroup will contain all the basetypes of the former basegroups plus the new ones, leading to a new name. The names of the basegroups lumped into the new, more inclusive basegroup all become unavailable.

Third, when the definition of a MOTU changes such that some of its constituent MOTUs are removed or others are added, the newly defined MOTU will obtain a new denomination and its former name will become unavailable.

Fourth, we recommend defining a convention for the naming of the MOTUs using a succession of fields using different case types for each successive level of the nomenclature (e.g., Roman or Arabic numerals, lower or upper case roman, or Greek letters). For instance, in the case of a three-level nomenclature system, we propose to use a combination of upper case Roman numerals, followed by one or more lower case letters, and an Arabic numeral: *Genus species* IIa1. When one category of symbol is exhausted due to a large number of categories or successive renaming, we recommend doubling it: *Genus species* IIaa1. In this way, increasingly complex combinations can be easily elaborated in the run of successive revisions unveiling a deeper hierarchical structure requiring additional nested levels (e.g., Ia1 → A1a1 → A1a1α → ...). The finest level is always at the right end of the name.

Finally as progress in sequencing technologies reduce analytical costs, more marker genes will become available, especially for groups where the genome is currently unknown. Therefore, additional marker gene(s) can be included in the nomenclature when available, and will constitute additional basetype(s) keeping consistency with the available marker (Fig. 1b).

The rules above only apply to the definition of MOTUs. Attribution of unknown sequences, in contrast, can be made to any level in the nomenclature hierarchy of MOTUs. Any observation yielding sufficient genetic information unambiguously assignable to a given MOTU level can be used to investigate the ecology, biogeography, or evolution of this given MOTU. This applies to sequences only partially covering the chosen gene region, environmental sequences lacking by nature an attribution to a given specimen, or a RFLP pattern accounting for the presence of a diagnostic sequence pattern in a given specimen.

THE CASE OF PLANKTONIC FORAMINIFERA

As in many other taxa (Bickford et al. 2007) and most particularly among protists (Kosakyan et al. 2013), morphospecies in planktonic foraminifera were shown to harbor considerable cryptic diversity (de Vargas et al. 1999; Darling and Wade 2008), some of the detected genetic entities being associated with subtle but significant morphological differences (Darling et al. 2006; Morard et al. 2009, 2011; Aurahs et al. 2011; Quillévéré et al. 2013; Weiner et al. 2015). Nevertheless, the majority of the discovered genetic diversity has not been linked so far to any morphological differences and most MOTUs (cryptic species) of planktonic foraminifera remain formally nameless (André et al. 2014).

The last decade has seen an increase in genotyping studies in planktonic foraminifera carried out by different research teams. This resulted in the existence of multiple, often inconsistent molecular nomenclatures, as is the case, for example, for the successive studies of the genus *Globigerinella* (Fig. 2). In addition, the lack of clarity concerning the level of genetic and morphological divergence that represents biological species would likely lead to frequent, large-scale taxonomic revisions. For this reason, the genetically circumscribed entities, variously referred to as “clades”, “genetic types”, “genotypes”, “subtypes”, “phylotypes”, “ribotypes”, “ecotypes”, “haplotypes”, “sequence motifs”, “cryptic species”, or “genetic species”, received informal labels (e.g., Huber et al. 1997; de Vargas et al. 1999) typically reflecting the hierarchical phylogenetic structure of the classified units (Darling et al. 2003). Initially, such labels have been given without an explicit assignment of a type sequence (typification). Their definitions made use of different marker genes, and due to large heterogeneity in substitution rates (de Vargas et al. 1997), they were not based on universal thresholds.

To date, two attempts have been made to delineate MOTUs in planktonic foraminifera as a whole by an

objective analysis of sequence diversity. Göker et al. (2010) explored a clustering optimization approach, while André et al. (2014) used two complementary approaches for automated delineation of genetic clusters, namely GMYC (Pons et al. 2006) and ABGD (Puillandre et al. 2012b). In both cases, the authors concluded that a unique threshold related to a consistent level of genetic divergence could not be found to circumscribe taxa within the group as a whole. In analogy with what is observed in other taxonomic groups (Carstens et al. 2013), ABGD- and GMYC-based delineations of planktonic foraminifera species are rarely congruent, leading the authors to favor conservative (i.e., lumping) solutions to avoid over-interpretation of the results (André et al. 2014). These incongruent results between ABGD and GMYC are due to an extreme heterogeneity of substitution rates and complex evolutionary patterns of the SSU rDNA in planktonic foraminifera. For instance, an extensive survey of the morphologically diverse *Trilobatus sacculifer* did not show any sign of cryptic diversity (Fig. 3; André et al. 2013), whereas a comparable sampling and sequencing effort identified *G. siphonifera* as a hyperdiverse species complex (Weiner et al. 2014, 2015). Sharing almost identical morphology, the three distinct lineages of *Orbulina universa* seem to have diverged millions of years ago (de Vargas et al. 1999; Morard et al. 2009), whereas “genotypes” of *Globigerinoides elongatus* represent shallow divergences (Aurahs et al. 2011), if not a population signal (André et al. 2014).

Until now, there has been no evidence found for hybridization among any of the described MOTUs in planktonic foraminifera. Although these lines of evidence alone are not sufficient to unambiguously equate the observed genetic diversity with reproductive isolation, they collectively speak for the existence of reproductively isolated cryptic species in planktonic foraminifera. Protists are rarely amenable to breeding experiments, which ultimately prevents direct confirmation of the level of genetic divergence corresponding to the biological species (but see Amato et al. 2007). Therefore, the only levels of divergence in planktonic foraminifera that can be identified unambiguously with current methods are represented by the collocated sets of mutations characterizing a unique ribosomal sequence (motif), the set of motifs present within an individual, and the level of genetic divergence that is manifested morphologically and can be treated by formal taxonomy. Because it is not possible to directly test for interbreeding, we cannot determine which of the levels of the hierarchical structure in the genetic diversity of planktonic foraminifera corresponds to biological species (see also Leavitt et al. 2015). A solution to this problem may be found at some stage in the future, but for now, we consider it an additional advantage of a hierarchical taxonomy that it leaves open where the exact level of biological species occurs in each clade. It is very likely that within a given morphospecies, the level of biological species will correspond to the same level in the hierarchy.

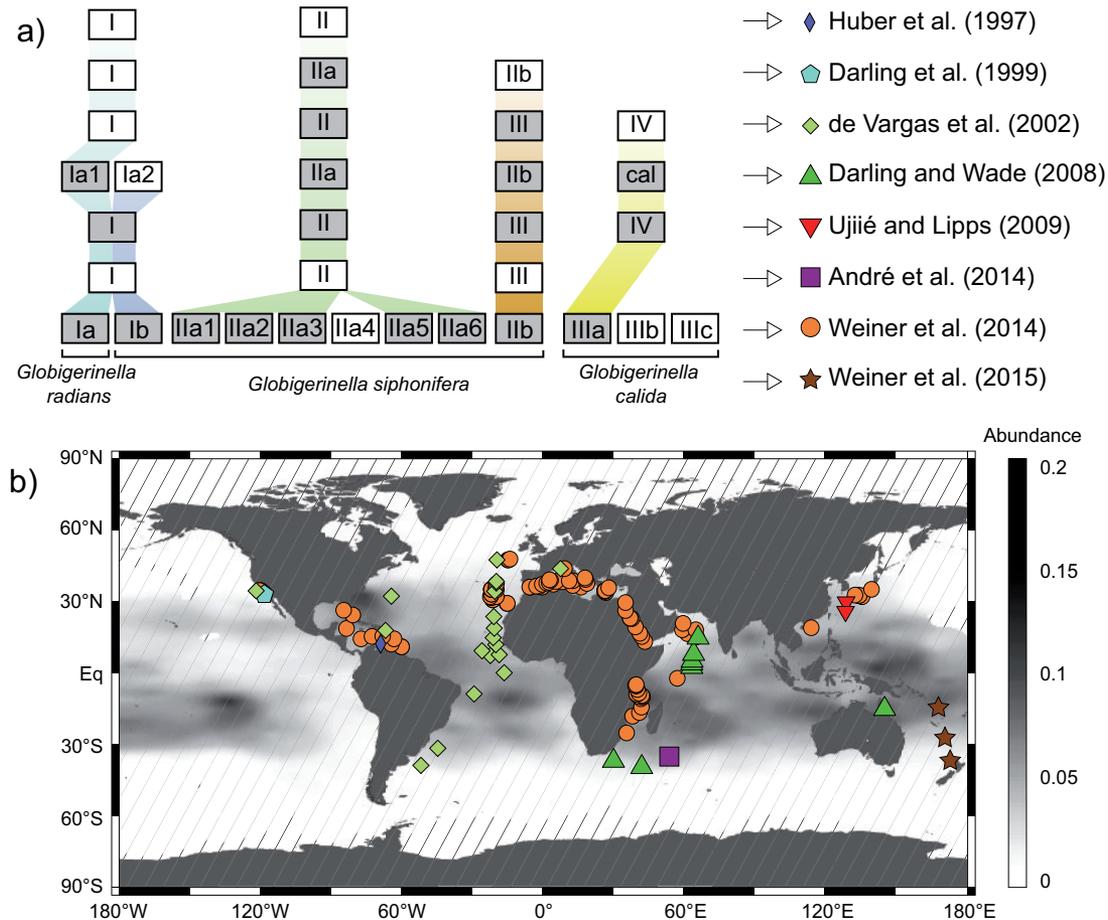


FIGURE 2. History of the molecular nomenclature of the three morphospecies within the planktonic foraminifera genus *Globigerinella*. a) Continuity of names of the different MOTUs among the successive studies listed chronologically, from the initial work of Huber et al. (1997) up to the last taxonomic revision by Weiner et al. (2015). Each rectangle represents the MOTUs described in each successive study and their connectivity is given by the strips; rectangles with a gray background highlight the renaming of a MOTU between two studies. b) Geographic origin of the sequences produced in the successive studies (same labels as in a). Background shading shows the relative abundance of the genus *Globigerinella* in surface sediments as an index of the morphospecies distribution (from Weiner et al. 2014).

IMPLEMENTATION IN PLANKTONIC FORAMINIFERA

Based on the general principles presented above, we exemplify hereafter the criteria that allow us to implement such a nomenclature scheme in the case of planktonic foraminifera. We stress that the criteria given in this work are a starting point based on currently available evidence and methodologies; these criteria may obviously evolve in the future.

Basetype Definition in Planktonic Foraminifera

We propose to define basetypes in planktonic foraminifera (Fig. 4) by using sequence patterns in the rDNA fragment located at the 3'-end of the SSU rRNA gene, between stems 32 and 50 (Wuyts et al. 2001). This fragment has been proposed as the barcode for benthic foraminifera (Pawlowski and Holzmann 2014). It captures enough phylogenetic signal (Darling and Wade 2008) and harbors six variable regions, three of them being specific to foraminifera (Pawlowski and Lecroq 2010). The sequence pattern of each region of

the sequence, as defined in Morard et al. (2015), should optimally have been observed several times (we suggest here using three times as a minimum threshold) in the available data set, either by technical or biological replicates, to validate the observed sequence pattern as a genuine biological signal. In the case of intragenomic variability, the sequencing effort must be scaled to the amount of diversity within the target morphospecies. In summary, each unique sequence pattern observed multiple times and at least once in a sequence covering the entire marker is considered a *basetype*.

Nomenclature System

The proposed planktonic foraminifera molecular nomenclature system is organized in a three-level system (Fig. 4c) deemed to be intraspecific with respect to the formal nomenclature regulated by the ICBN (1999). The nomenclature is built only on basetypes validated by multiple observations. In principle, for the reasons outlined above, any of the three levels could be

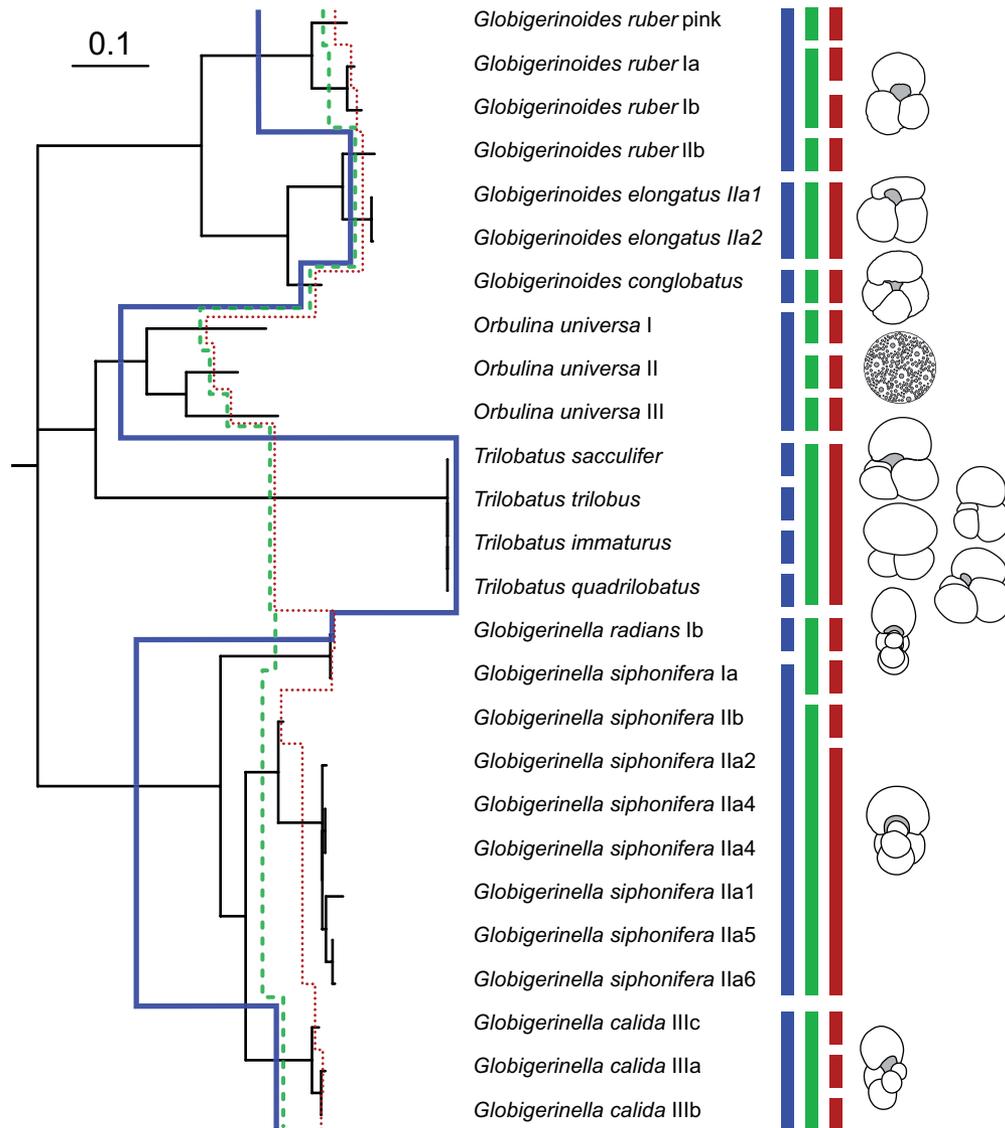


FIGURE 3. Decoupling of rates of molecular and morphological evolution in planktonic foraminifera. This molecular phylogeny shows the relationships between 11 morphospecies (illustrated by sketches of representative specimens) belonging to the genera *Globigerinoides* (Aurahs et al. 2011), *Orbulina* (de Vargas et al. 1999; Morard et al. 2009), *Trilobatus* (André et al. 2013; Spezzaferri et al. 2015) and *Globigerinella* (Weiner et al. 2014, 2015), based on their SSU rDNA sequence motifs. The solid, dashed, and dotted lines drawn on the tree show the position of morphological species, and higher and lower levels of genetic divergence as used in the cited studies, respectively. The vertical bars to the right of the taxon names highlight the incongruence between resulting molecular and morphological taxonomies.

considered a Primary Species Hypothesis *sensu* Pante et al. (2014) and represent a working hypothesis.

MOTU level 3.—The lowest level corresponds to the basegroup and represents the genetic variability potentially observable at the level of a single individual. Thus, by definition different basegroups contain basetypes that are never observed together. In the case of the absence of intragenomic variability, such as in many spinose planktonic foraminifera, the basegroup may contain only one basetype. The basegroup is the basic unit of the nomenclature system proposed here; it is the only level grounded on empirical data. The variability observed within a given basegroup

represents the intragenomic (individual) variability, whereas the variability observed among different basegroups is considered to represent at least the level of population variability (Fig. 4).

MOTU level 2.—The intermediate level most likely represents the level of biological species. It is defined as a monophyletic cluster of one or several basegroups, delineated by an objective method searching for the most likely level of divergence reflecting reproductive isolation. André et al. (2014) showed that ABGD (Puillandre et al. 2012b) or GMYC (Pons et al. 2006) can be used to objectively delineate MOTUs in planktonic foraminifera. In the case of conflicting delineations and

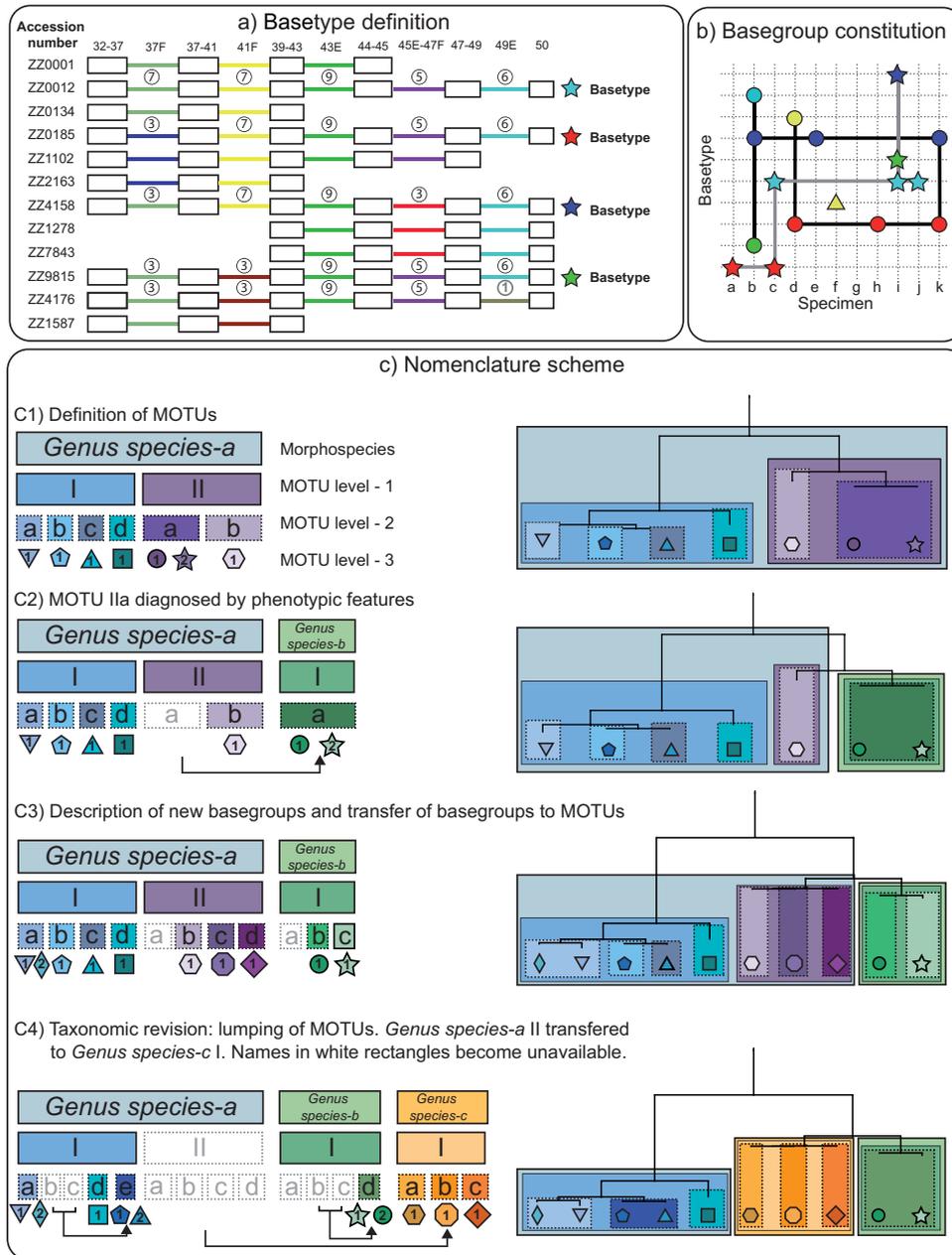


FIGURE 4. A guide for the definition of MOTUs in planktonic foraminifera. a) The selected gene for planktonic foraminifera is the fragment located at the 3'-end of the SSU rDNA, between stems 32 and 50 (Wuyts et al. 2001). The sequences are partitioned in their constitutive variable (lines) and conserved (rectangles) regions, following Morard et al. 2015. In this example, the different line colors represent unique sequence motifs found in the available data set; the position of each region is given above the first sequence. Only sequences covering the entire fragment and sequences for which each individual region has been observed at least three times in the available data set (number of observations given above the lines) are designated as basetypes. In this example, five sequences have the required length and four can be designated as basetypes (ZZ1587 is long enough but has a sequence motif for the region 49E which is observed only once in the data set). b) Occurrences of basetypes (rows) within a set of individuals (columns). The different symbols represent the different basetypes occurring within (when situated on the same column) and between (when situated on the same row) individuals. A basegroup is defined as a network of basetypes connected by pairwise co-occurrence, represented by solid lines. In this way, two basetypes can be grouped in the same basegroup even if they were not observed co-occurring in the same specimen. When intra-individual variability is discovered, we recommend generating at least 10 full-length clones for a few selected specimens in order to cover the intragenomic variability occurring within a given MOTU. c) The nomenclature scheme allows the transfer of MOTUs to higher levels, splitting and lumping of existing MOTUs, and the inclusion of new basegroups, as illustrated in a succession of theoretical nomenclature acts ((c1)–(c4)). The intra-morphospecies nomenclature we propose for planktonic foraminifera is based on a three-level hierarchical system (e.g., “Ia1”, meaning: MOTU level-1 “I” includes MOTU level-2 “a” includes MOTU level-3 “1”) (Case (c1)). The nomenclature is amended as soon as the status of the previously defined MOTUs is modified based on morphological, molecular, or ecological evidences. In the case of transfer to a higher-level taxon (Case (c2)), splitting one MOTU into two distinct ones (Case (c3)), or lumping of two MOTUs into a single one (Case (c4)), names of amended MOTUs become unavailable (dotted gray rectangles). In the case of a revision with “steps backward” (as for the MOTUs *Genus species-b* Ia1 and *Genus species-b* Ia2 in (c2), which are split into Ib1 and Ic1 in (c3) and merged again in (c4)), a new MOTU name is used. In the case of objective synonymy, we follow the same principle of priority as for the ICZN (1999).

for nomenclature purposes, we recommend the use of the most conservative boundary for species delineation in planktonic foraminifera—that is, lumping rather than splitting, following Carstens et al. (2013) and Miralles and Vences (2013). Because automated delineation is a dynamic research field, new methods may arise in the future, making the status (or level) of the second level MOTU likely to be reevaluated. In addition, sequences from new markers may be used to assess the validity of the second level MOTU based on the current SSU rRNA marker.

MOTU level 1.—Represents the highest intra-morphospecific level designed in the proposed nomenclature scheme; it is defined as a monophyletic cluster of intermediate-level MOTUs (i.e., putative biological species) reflecting a major disruption in the genetic variability observed within a given morphospecies. This level is considered to represent ancient and well-established divergences within the morphospecies evolutionary history; it is likely to represent at least the level of biological species. Optimally, this level should be statistically significantly supported as monophyletic and it should always be delineated objectively. For example, its status can be established by using a barcode gap delineation method such as ABGD, using the coarsest partitioning given by the method within a morphospecies.

We exemplify this nomenclature scheme through two cases: the *Globigerinella* species complex, which is characterized by a large genetic diversity and a disconnection between morphological and molecular evolution (Weiner et al. 2015; Fig. 5), and the morphospecies *Pulleniatina obliquiloculata*, which is characterized by a large intragenomic variability (Ujiié et al. 2012; Fig. 6). The methodology used to delineate the MOTUs is detailed in the figure captions. The *Globigerinella* species complex is composed of 10 MOTUs (MOTU level 3) structured into five putative species (MOTU level 2) constituting three genetic lineages (MOTU level 1). Attribution of these MOTUs to morphological species results in paraphyly as there is a clear disconnection between morphological and molecular evolution within this species complex. *Pulleniatina obliquiloculata* is composed of 36 MOTUs (MOTU level 3) structured into two putative species (MOTU level 2) and a single genetic lineage (MOTU level 1).

These examples show that a consistent application of the proposed system will be largely congruent with original studies, but will likely change the levels of the MOTUs. Since MOTUs have been described initially from partial sequences, the new scheme will decrease the number of MOTUs of intermediate level, unless a systematic sequencing effort is carried out to provide full-length sequences (basotypes) for all observed motifs. The application also highlights the advantage of the system of not being linked to a single threshold, thus allowing consistent classification of MOTUs with

very different substitution rates for the same marker gene.

FURTHER CONSIDERATIONS

The proposed MOTU nomenclature may be applied to any other taxa which display a hierarchically structured genetic diversity occurring below the level of morphospecies. MOTU implementation for a given group could serve as a framework to associate metadata to a consistent level of genetic divergence. Consistent application of the proposed molecular taxonomy to sequences from individuals as well as environmental sequences will allow a reproducible and reliable association of MOTUs with ecological data, geographic origin, time and depth of collection, behavior, associated microbiome, or morphology. The criteria of such nomenclature (genetic threshold between two identical levels of the same MOTUs) could in theory be propagated into environmental data sets with novel sequence motifs not linked to organisms, even in such situations where no unique threshold can be applied to the entire group. The use of a MOTU-specific threshold as a working hypothesis will lead to a more accurate interpretation of the metagenomic data. In this respect, the proposed molecular taxonomy scheme will serve both as a tool to preserve existing work and to hierarchically explore new and increasingly complex data sets. In addition, in the case of a revision of the taxonomy, metadata linked to MOTUs can be automatically transferred into the revised categories.

On the other hand, the proposed nomenclature system shows some weaknesses that should not be ignored. First, it is not primarily meant to name unknown clades from environmental sequences. Because it is tied to formal nomenclature, the morphotaxonomical identity of the specimen from which the sequences are derived must be first established or the molecular taxonomy becomes detached from formal taxonomy. In the proposed scheme, it is not possible to create artificial categories based solely on molecular characters in order to classify sequences of unknown origin. Second, the system works best when the magnitude and structure of the diversity within the group are known. Otherwise, the result of the iterative process of testing species hypotheses will be a nomenclature with many “holes”. This is unfortunate, but unavoidable, and we thus advocate for an application of the taxonomic scheme not before a sufficient amount of data covering a significant part of the ecological and biogeographical ranges of the studied taxa is already available. Clearly, the existence of a comprehensive database covering the taxonomic, biogeographic, and ecological range of the group of interest such as the one developed by the authors of this article (Morard et al. 2015) is a strong prerequisite for the establishment of a molecular nomenclature. Third, while the scheme remains reasonably flexible, we note that frequent revisions will inevitably complicate the nomenclature, especially by generating lengthy lists

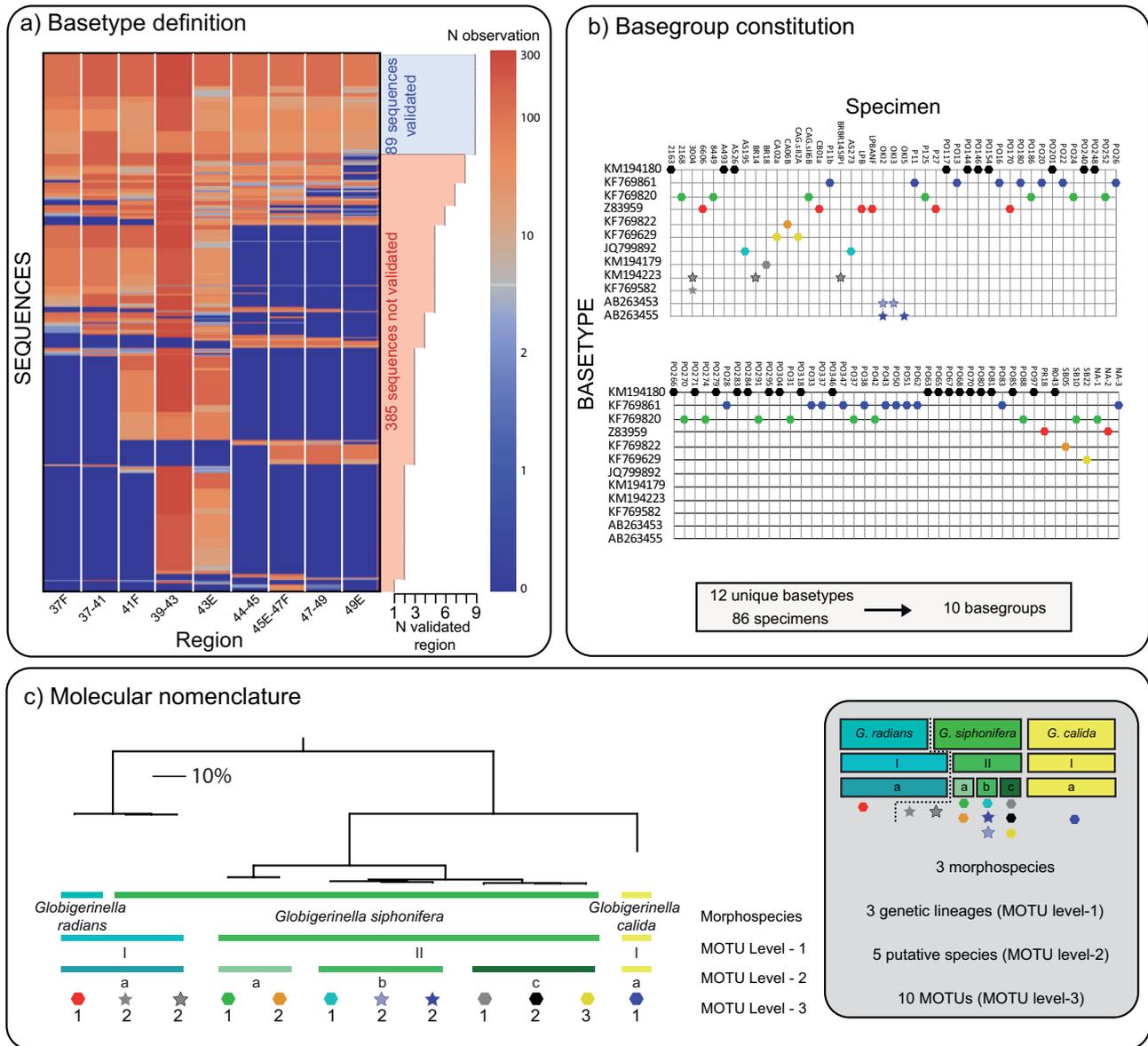


FIGURE 5. Implementation of the molecular nomenclature system for the genus *Globigerinella*. a) The heatmap represents the 474 sequences belonging to *Globigerinella* available in the PFR² database v.1 (Morard et al. 2015). Each column represents an individual variable region of the SSU rDNA gene; colors represent the number of observations of a given sequence motif. For each region, only sequence patterns (motifs) observed at least three times are validated and only the sequences for which nine regions are validated are retained for the molecular nomenclature, leading to 89 sequences validated. b) Among the 89 retained sequences, 12 unique sequence patterns are retained as basetypes. Basetypes co-occurring within the same individual(s) (marked by stars, two cases) are grouped within the same basegroup. c) The 12 basetypes were automatically aligned with MAFFT v. 7 (Kato and Standley 2013) and a phylogenetic reconstruction was performed with PhyML (Guindon et al. 2010) under the model (GTR+I) chosen with jModeltest v. 2.1.7 (Darriba et al. 2012). The same alignment was subjected to Automatic Barcode Gap Delimitation (ABGD, Puillandre et al. 2012b), using the default setting of the program and the K-80 distance to classify the basetypes into putative species. The first MOTU plateau was retained as putative species (MOTU level 2), and the lowest plateau was retained to identify the major disruption of genetic variability of the species complex (MOTU level 1). The three morphotypes occurring within *Globigerinella* identified by Weiner et al. (2015) were implemented in the taxonomy as an upper boundary. All data necessary to produce the nomenclature and resulting files are available in Supplementary Material S1, available on Dryad at <http://dx.doi.org/10.5061/dryad.64pg3>.

of unavailable names. Therefore, we recommend that a revision of the nomenclature should be attempted only when a significant new input is provided. Indeed, nothing excludes a revision when only a limited number of sequences for a given taxa are available, but a certain degree of stability is necessary between each revision

to make the nomenclature tractable and thus useful. Ultimately, it remains left to the taxonomist to evaluate the balance between stability and reliability. Last, the nomenclature system described here is currently not formally codified, making its application grounded on a voluntary basis. Therefore, it will require a

CONCLUSIONS

The nomenclatural system we propose here does not replace but complements the currently valid codes for biological nomenclature (ICZN 1999; ICN 2011; ICTV 2011; Garrity et al. 2015). In other words, it is tied to formal nomenclature as regulated by current codes at the morphospecies level, but does not overlap with them, so that no conflict can result from the use of this nomenclatural system with respect to any one of these formal codes. For planktonic foraminifera, its success and usefulness rely on the community observing the rules, albeit on a voluntary basis. For example, the nomenclature system could be linked to a registry similar in its function to the Zoobank (<http://zoobank.org/>). Unlike formal nomenclature, the proposed system is not based on universal criteria, with implementations possibly differing among the classified taxa—the reason why we think this nomenclatural system must remain informal with respect to the current codes. It is particularly useful when a certain degree of cryptic diversity has been detected in the studied group, and only works if the extent of MOTUs diversity is not excessive, enabling the nomenclature to reach a reasonable stability after a few iterations. Its consistent usage will allow seamless tracking of the described MOTUs among studies, thus enabling integration of evidence from multiple sources to discuss the biological status of the identified MOTUs. It has the advantage to be additive and hierarchically organized, following the recommendations of the integrated taxonomy by connecting with existing formal nomenclature and combining multiple lines of evidence to define taxa (Schlick-Steiner et al. 2010). The nomenclature scheme is amendable and allows the formulation of reasonable hypotheses for automated taxonomy without relying on universal thresholds. The great advantage of the system is the possibility to classify any genetic data and metadata associated to MOTUs of any given level. Therefore, it allows the synthesis of data derived from single-cell analysis with sequences of various lengths, RFLP data, or environmental sequences generated using Next-Generation Sequencing, and enlarges the amount of ecological or biogeographical data linked to a given taxon. In the case of planktonic foraminifera, this nomenclature scheme may become the wiring connecting genetic, biogeographic, and ecological data produced from molecular works since the '90s to the most recent and future data sets.

SUPPLEMENTARY MATERIAL

Data available from the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.64pg3>.

FUNDING

This work was supported by the DFG-Research Center/Cluster of Excellence "The Ocean in the Earth

System". This study is a contribution to the effort of the SCOR/IGBP Working Group 138 "Modern planktonic Foraminifera and ocean changes".

ACKNOWLEDGMENTS

The authors are thankful to Bryan Carstens and four anonymous reviewers, as well as to Adrian Paterson (Associate Editor) and Frank Anderson (Editor in chief) who provided constructive and encouraging comments that helped us to improve the present manuscript.

REFERENCES

- Amato A., Kooistra W.H.C.F., Ghiron J.H.L., Mann D.G., Pröschold T., Montresor M. 2007. Reproductive isolation among sympatric cryptic species in marine diatoms. *Protist* 158:193–207.
- André A., Quillévéré F., Morard R., Ujiie Y., Escarguel G., de Vargas C., de Garidel-Thoron T., Douady C.J. 2014. SSU rDNA divergence in planktonic foraminifera: Molecular taxonomy and biogeographic implications. *PLoS One* 9:e104641.
- André A., Weiner A., Quillévéré F., Aurahs R., Morard R., Douady C.J., de Garidel-Thoron T., Escarguel G., de Vargas C., Kucera M. 2013. The cryptic and the apparent reversed: Lack of genetic differentiation within the morphologically diverse plexus of the planktonic foraminifer *Globigerinoides sacculifer*. *Paleobiology* 39: 21–39.
- Aurahs R., Treis Y., Darling K., Kucera M. 2011. A revised taxonomic and phylogenetic concept for the planktonic foraminifer species *Globigerinoides ruber* based on molecular and morphometric evidence. *Mar. Micropaleontol.* 79:1–14.
- Benton M.J. 2000. Stems, nodes, crown clades, and rank-free lists: Is Linnaeus dead? *Biol. Rev. Camb. Phil. Soc.* 75:633–648.
- Bickford D., Lohman D.J., Sodhi N.S., Ng P.K.L., Meier R., Winker K., Ingram K.K., Das I. 2007. Cryptic species as a window on diversity and conservation. *Trends Ecol. Evol.* 22:148–155.
- Caron D.A. 2013. Towards a molecular taxonomy for protists: Benefits, risks, and applications in Plankton Ecology. *J. Eukaryot. Microbiol.* 60:407–413.
- Carstens B.C., Pelletier T.A., Reid N.M., Satler J.D. 2013. How to fail at species delimitation. *Mol. Ecol.* 22:4369–4383.
- Darling K.F., Kucera M., Kroon D., Wade C.M. 2006. A resolution for the coiling direction paradox in *Neogloboquadrina pachyderma*. *Paleoceanography* 21:PA2011.
- Darling K.F., Kucera M., Wade C.M., von Langen P., Pak D. 2003. Seasonal distribution of genetic types of planktonic foraminifer morphospecies in the Santa Barbara Channel and its paleoceanographic implications. *Paleoceanography* 18:1–10.
- Darling K.F., Wade C.M. 2008. The genetic diversity of planktic foraminifera and the global distribution of ribosomal RNA genotypes. *Mar. Micropaleontol.* 67:216–238.
- Darriba D., Taboada G.L., Doallo R., Posada D. 2012. jModelTest 2: More models, new heuristics and parallel computing. *Nat. Methods* 9:772–772.
- de Queiroz K., Cantino P. 2001. Phylogenetic nomenclature and the PhyloCode. *Bull. Zool. Nomencl.* 58:254–271.
- de Queiroz K. 2007. Species concepts and species delimitation. *Syst. Biol.* 56:879–886.
- de Vargas C., Audic S., Henry N., Decelle J., Mahé F., Logares R., Lara E., Berney C., Le Bescot N., Probert I., Carmichael M., Poulain J., Romac S., Colin S., Aury J.-M., Bittner L., Chaffron S., Dunthorn M., Engelen S., Flegontova O., Guidi L., Horák A., Jaillon O., Lima-Mendez G., Lukeš J., Malviya S., Morard R., Mulot M., Scalco E., Siano R., Vincent F., Zingone A., Dimier C., Picheral M., Searson S., Kandels-Lewis S., Coordinators T.O., Acinas S.G., Bork P., Bowler C., Gorsky G., Grimsley N., Hingamp P., Iudicone D., Not F., Ogata H., Pesant S., Raes J., Sieracki M.E., Speich S., Stemann L., Sunagawa S., Weissenbach J., Wincker P., Karsenti E. 2015. Eukaryotic plankton diversity in the sunlit ocean. *Science* 348:1–11.

- de Vargas C., Norris R., Zaninetti L., Gibb S.W., Pawlowski J. 1999. Molecular evidence of cryptic speciation in planktonic foraminifers and their relation to oceanic provinces. *Proc. Natl. Acad. Sci. USA*. 96:2864–2868.
- de Vargas C., Zaninetti L., Hilbrecht H., Pawlowski J. 1997. Phylogeny and rates of molecular evolution of planktonic foraminifera: SSU rDNA sequences compared to the fossil record. *J. Mol. Evol.* 45: 285–294.
- 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. 2015. Species delimitation's coming of age. *Syst. Biol.* 64: 897–899.
- Funk D.J., Omland K.E. 2003. Frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annu. Rev. Ecol. Evol. Syst.* 34:397–423.
- Garrity G.M., Parker C.T., Tindall B.J. 2015. International code of nomenclature of prokaryotes. *Int. J. Syst. Evol. Microbiol.*
- Göker M., Grimm G.W., Auch A.F., Aurahs R., Kučera M. 2010. A clustering optimization strategy for molecular taxonomy applied to planktonic foraminifera SSU rDNA. *Evol. Bioinform.* 6: 97–112.
- Guindon S., Dufayard J.F., Lefort V., Anisimova M., Hordijk W., Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* 59:307–321.
- Hebert P.D.N., Cywinska A., Ball S.L., DeWaard J.R. 2003. Biological identifications through DNA barcodes. *Proc. Biol. Sci.* 270: 313–321.
- Huber B.T., Bijma J., Darling K. 1997. Cryptic speciation in the living planktonic foraminifer *Globigerinella siphonifera*. *Paleobiology* 23: 33–62.
- ICN. 2011. International Code of Nomenclature for algae, fungi and plants (Melbourne Code) adopted by the Eighteenth International Botanical Congress Melbourne, Australia.
- ICTV. 2011. The International Committee on Taxonomy of Viruses (ICTV): The statutes of the ICTV. In: King A.M.Q., Carstens E.B., Lefkowitz E.J., editors. *Virus taxonomy—Ninth Report of the International Committee on Taxonomy of Viruses*. Elsevier/Academic Press. p. 1273–1277.
- ICZN. 1999. International Code of Zoological Nomenclature. 4th Edition. International Trust for Zoological Nomenclature. London. p. 335.
- Jörger K.M., Schrödl M. 2013. How to describe a cryptic species? Practical challenges of molecular taxonomy. *Front. Zool.* 10:1–27.
- Karanovic T., Djurakic M., Eberhard S.M. 2016. Cryptic Species or Inadequate Taxonomy? Implementation of 2D Geometric Morphometrics Based on Integumental Organs as Landmarks for Delimitation and Description of Copepod Taxa. *Syst. Biol.* 65: 304–327.
- Katoh K., Standley D.M. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30:772–780.
- Kosakyan A., Gomaa F., Mitchell E.A.D., Heger T.J., Lara E. 2013. Using DNA-barcoding for sorting out protist species complexes: A case study of the *Nebela tincta-collaris-bohemica* group (Amoebozoa; Arcellinida, Hyalospheniidae). *Eur. J. Protistol.* 49:222–237.
- Kucera M., Darling K.F. 2002. Cryptic species of planktonic foraminifera: Their effect on palaeoceanographic reconstructions. *Phil. Trans. A Math. Phys. Eng. Sci.* 360:695–718.
- Leavitt S.D., Moreau C.S., Thorsten Lumbsch H. 2015. The dynamic discipline of species delimitation: Progress toward effectively recognizing species boundaries in natural populations. In: Upreti D.K., Divakar P.K., Shukla V., Bajpai R., editors. *Recent advances in lichenology*. India: Springer. p. 11–44.
- Leliaert F., Verbruggen H., Vanormelingen P., Steen F., López-Bautista J.M., Zuccarello G.C., Clerck O. De. 2014. DNA-based species delimitation in algae DNA-based species delimitation in algae. *Eur. J. Phycol.* 49:179–196.
- Logares R., Audic S., Bass D., Bittner L., Boutte C., Christen R., Claverie J.-M., Decelle J., Dolan J.R., Dunthorn M., Edvardsen B., Gobet A., Kooistra W.H.C.F., Mahé F., Not F., Ogata H., Pawlowski J., Pernice M.C., Romac S., Shalchian-Tabrizi K., Simon N., Stoeck T., Santini S., Siano R., Wincker P., Zingone A., Richards T.A., de Vargas C., Massana R. 2014. Patterns of rare and abundant marine microbial eukaryotes. *Curr. Biol.* 24:813–821.
- Massana R., Castresana J., Balagué V., Romari K., Groisillier A., Valentín K., Pedrós-alió C., Balagué V., Guillou L., Pedro C. 2004. Phylogenetic and ecological analysis of novel marine stramenopiles. *Appl. Environ. Microbiol.* 70:3528–3534.
- Mayr E. 1942. *Systematics and the origin of species*. New York: Columbia University Press.
- Miralles A., Vences M. 2013. New metrics for comparison of taxonomies reveal striking discrepancies among species delimitation methods in Madascincus lizards. *PLoS One* 8:e68242.
- Mittelbach G.G., Schemske D.W., Cornell H.V., Allen A.P., Brown J.M., Bush M.B., Harrison S.P., Hurlbert A.H., Knowlton N., Lessios H.A., McCain C.M., McCune A.R., McDade L.A., McPeck M.A., Near T.J., Price T.D., Ricklefs R.E., Roy K., Sax D.F., Schluter D., Sobel J.M., Turelli M. 2007. Evolution and the latitudinal diversity gradient: Speciation, extinction and biogeography. *Ecol. Lett.* 10: 315–331.
- Morard R., Darling K.F., Mahé F., Audic S., Ujiié Y., Weiner A.K.M., André A., Seears H.A., Wade C.M., Quillévéré F., Douady C.J., Escarguel G., de Garidel-Thoron T., Siccha M., Kucera M., de Vargas C. 2015. PFR²: A curated database of planktonic Foraminifera 18S ribosomal DNA as a resource for studies of plankton ecology, biogeography, and evolution. *Mol. Ecol. Resour.* 49:1–14.
- Morard R., Quillévéré F., Douady C.J., de Vargas C., de Garidel-Thoron T., Escarguel G. 2011. Worldwide genotyping in the planktonic foraminifer *Globoconella inflata*: Implications for life history and paleoceanography. *PLoS One* 6:e26665.
- Morard R., Quillévéré F., Escarguel G., de Garidel-Thoron T., de Vargas C., Kucera M. 2013. Ecological modeling of the temperature dependence of cryptic species of planktonic Foraminifera in the Southern Hemisphere. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 391:13–33.
- Morard R., Quillévéré F., Escarguel G., Ujiié Y., de Garidel-Thoron T., Norris R.D., de Vargas C. 2009. Morphological recognition of cryptic species in the planktonic foraminifer *Orbulina universa*. *Mar. Micropaleontol.* 71:148–165.
- Padial J.M., Miralles A., De la Riva I., Vences M. 2010. The integrative future of taxonomy. *Front. Zool.* 7:1–14.
- Pante E., Puillandre N., Viricel A., Arnaud-Haond S., Aurelle D., Castelin M., Chenuil A., Destombe C., Forcioli D., Valero M., Viard F., Samadi S. 2015. Species are hypotheses: Avoid connectivity assessments based on pillars of sand. *Mol. Ecol.* 24:525–544.
- Pante E., Schoelincq C., Puillandre N. 2014. From integrative taxonomy to species description: One step beyond. *Syst. Biol.* 64:152–160.
- Pawlowski J., Audic S., Adl S., Bass D., Belbahri L., Berney C., Bowser S.S., Cepicka I., Decelle J., Dunthorn M., Fiore-Donno A.M., Gile G.H., Holzmann M., Jahn R., Jirků M., Keeling P.J., Kostka M., Kudryavtsev A., Lara E., Lukeš J., Mann D.G., Mitchell E.A.D., Nitsche F., Romeralo M., Saunders G.W., Simpson A.G.B., Smirnov A.V., Spouge J.L., Stern R.F., Stoeck T., Zimmermann J., Schindel D., de Vargas C. 2012. CBOL Protist Working Group: Barcoding Eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLoS Biol.* 10:e1001419.
- Pawlowski J., Holzmann M. 2014. A plea for DNA barcoding of foraminifera. *J. Foraminifer. Res.* 44:62–67.
- Pawlowski J., Lecroq B. 2010. Short rDNA barcodes for species identification in foraminifera. *J. Eukaryot. Microbiol.* 57:197–205.
- Pons J., Barraclough T.G., Gomez-Zurita J., Cardoso A., Duran D.P., Hazell S., Kamoun S., Suhl W.D., Vogler A.P. 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Syst. Biol.* 55:595–609.
- Puillandre N., Bouchet P., Boisselier-Dubayle M.C., Brisset J., Buge B., Castelin M., Chagnoux S., Christophe T., Corbari L., Lambourdière J., Lozouet P., Marani G., Rivasseau A., Silva N., Terryn Y., Tillier S., Utge J., Samadi S. 2012a. New taxonomy and old collections: Integrating DNA barcoding into the collection curation process. *Mol. Ecol. Resour.* 12:396–402.
- Puillandre N., Lambert A., Brouillet S., Achaz G. 2012b. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Mol. Ecol.* 21:1864–1877.

- Quillévéré F., Morard R., Escarguel G., Douady C.J., Ujiie Y., de Garidel-Thoron T., de Vargas C. 2013. Global scale same-specimen morphogenetic analysis of *Truncorotalia truncatulinoïdes*: A perspective on the morphological species concept in planktonic foraminifera. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 391:2–12.
- Ratnasingham S., Hebert P.D.N. 2007. BARCODING, BOLD?: The Barcode of life data system (www.barcodinglife.org). *Mol. Ecol. Notes* 7:355–364.
- Ratnasingham S., Hebert P.D.N. 2013. A DNA-based registry for all animal species: The Barcode Index Number (BIN) system. *PLoS One* 8:e66213.
- Salerno P.E., Señaris J.C., Rojas-Runjaic F.J.M., Cannatella D.C. 2015. Recent evolutionary history of Lost World endemics: Population genetics, species delimitation, and phylogeography of sky-island treefrogs. *Mol. Phylogenet. Evol.* 82:314–323.
- Satler J.D., Carstens B.C., Hedin M. 2013. Multilocus species delimitation in a complex of morphologically conserved trapdoor spiders (mygalomorphae, antrodiaetidae, Aliatypus). *Syst. Biol.* 62:805–823.
- Schindel D.E., Miller S.E. 2009. Provisional nomenclature: The on-ramp to taxonomic names. In: Polaszek A., editor. *Systema Naturae 250: Boca Raton (FL): The Linnaean Ark.* p. 109–115.
- Schlick-Steiner B.C., Steiner F.M., Seifert B., Stauffer C., Christian E., Crozier R.H. 2010. Integrative taxonomy: A multisource approach to exploring biodiversity. *Annu. Rev. Entomol.* 55:421–438.
- Simpson G.G. 1961. *Principles of animal taxonomy*. New York: Columbia University Press.
- Spezzaferri S., Kucera M., Pearson P.N., Wade B.S., Rappo S., Poole C.R., Morard R., Stalder C. 2015. Fossil and genetic evidence for the polyphyletic nature of the planktonic foraminifera “*Globigerinoides*”, and description of the new genus *Trilobatus*. *PLoS One*. 10:e0128108.
- Taylor H.R., Harris W.E. 2012. An emergent science on the brink of irrelevance: A review of the past 8 years of DNA barcoding. *Mol. Ecol. Resour.* 12:377–388.
- Ujiie Y., Asami T., de Garidel-Thoron T., Liu H., Ishitani Y., de Vargas C. 2012. Longitudinal differentiation among pelagic populations in a planktic foraminifer. *Ecol. Evol.* 2:1725–1737.
- Weiner A.K.M., Weinkauf M.F.G., Kurasawa A., Darling K.F., Kucera M., Grimm G.W. 2014. Phylogeography of the tropical planktonic foraminifera lineage *Globigerinella* reveals isolation inconsistent with passive dispersal by ocean currents. *PLoS One* 9:e92148.
- Weiner A.K.M., Weinkauf M.F.G., Kurasawa A., Darling K.F., Kucera M. 2015. Genetic and morphometric evidence for parallel evolution of the *Globigerinella calida* morphotype. *Mar. Micropaleontol.* 114: 19–35.
- Wuyts J., Van de Peer Y., De Wachter R. 2001. Distribution of substitution rates and location of insertion sites in the tertiary structure of ribosomal RNA. *Nucleic Acids Res.* 29:5017–5028.
- Yeates D.K., Seago A., Nelson L., Cameron S.L., Joseph L., Trueman J.W.H. 2011. Integrative taxonomy, or iterative taxonomy? *Syst. Entomol.* 36:209–217.
- Zhang J., Kapli P., Pavlidis P., Stamatakis A. 2013. A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* 29:2869–2876.