

Chapter 4

Problems and Questions Posed by Cryptic Species. A Framework to Guide Future Studies



Anne Chenuil, Abigail E. Cahill, Numa Délémontey,
Elrick Du Salliant du Luc, and Hadrien Fanton

Abstract Species are the currency of biology and important units of biodiversity, thus errors in species delimitations potentially have important consequences. During the last decades, owing to the use of genetic markers, many nominal species appeared to consist of several reproductively isolated entities called **cryptic species** (hereafter CS). In this chapter we explain why CS are important for practical reasons related to community and ecosystem monitoring, and for biological knowledge, particularly for understanding ecological and evolutionary processes. To find solutions to practical problems and to correct biological errors, a thorough analysis of the distinct types of CS reported in the literature is necessary and some general rules have to be identified. Here we explain how to identify CS, and we propose a rational and practical classification of CS (and putative CS), based on the crossing of distinct levels of genetic isolation with distinct levels of morphological differentiation. We also explain how to identify likely explanations for a given CS (either inherent to taxonomic processes or related to taxon biology, ecology and geography) and how to build a comprehensive database aimed at answering these practical and theoretical questions. Our pilot review of the literature in marine animals established that half of the reported cases are not CS *sensu stricto* (i.e. where morphology cannot distinguish the entities) and just need taxonomic revision. It also revealed significant associations between CS features, such as a higher proportion of diagnostic morphological differences in sympatric than in allopatric CS and more frequent ecologi-

A. Chenuil (✉) · H. Fanton
Aix Marseille Univ, Avignon Université, CNRS, IRD, IMBE, Marseille, France
e-mail: anne.chenuil@imbe.fr; hadrien.fanton@imbe.fr

A. E. Cahill
Aix Marseille Univ, Avignon Université, CNRS, IRD, IMBE, Marseille, France

Albion College, Albion, MI, USA
e-mail: acahill@albion.edu

N. Délémontey · E. D. Salliant du Luc
Aix-Marseille University, Marseille, France

cal differentiation between sympatric than allopatric CS, both observations supporting the competitive exclusion theory, thus suggesting that ignoring CS causes not only species diversity but also functional diversity underestimation.

4.1 Introduction

Quoting Agapow et al. (2004) “species are the currency of biology”. Long before the term “biodiversity” was coined and became widespread, the category of species was used as a major unit or category, not only to classify living things, but also to study ecological interactions and to assess the composition, resilience, evolution and risk of collapse of ecosystems (Gotelli and Colwell 2001). Nearly all descriptors of community assemblages and ecosystems - and their derived ecosystem functioning descriptors - require counting and separating species. The data used may contain variable amounts of information: (i) species richness (i.e. simply the number of distinct species), (ii) abundances of each species, (iii) relatedness among species and/or (iv) functional traits of species (Beauchard et al. 2017). The formal system naming the distinct species, established by Linnaeus in the eighteenth century, is the binomial nomenclature. The entities in the binominal nomenclature are called “nominal species” and are identified by a pair of Latin names, the first one corresponding to the genus to which the species belongs (e.g. *Homo sapiens*). Nominal species were described and defined exclusively from morphological characters until very recently, and are therefore sometimes called “morphospecies”. Nominal species (or groups of nominal species) were the entities considered in all the inventories of multicellular life until only a few years ago. During the last decades however, numerous nominal species appeared to be composed of separate entities which could not interbreed (Fig. 4.1), i.e. genetically isolated units. Genetic isolation for a group of individuals is the inability of its members to breed successfully with individuals from another group due to geographical, behavioral, physiological, or genetic barriers or differences. When genetic isolation is not the mere consequence of an external constraint such as geographic separation, but inherent to behavioral or genomic incompatibilities, such units constitute, by definition, distinct *biological species* (Mayr 1942). The expression “cryptic species” (hereafter CS) designates the distinct biological species that belong to one given nominal species and which were overlooked by the taxonomists who described the species initially (Knowlton 1993). This is generally, though not always, due to the absence of conspicuous diagnostic morphological differences (i.e. characters whose states allow unambiguous discrimination between species). In this chapter, “cryptic species *sensu lato*” correspond to distinct biological species within a nominal species, whatever the morphological differences or knowledge thereof. We define cryptic species *sensu stricto* as those CS where the absence of diagnostic morphological characters has been verified (and below we further explain the need to distinguish more categories of CS or putative CS). Similarly to CS (*ss* and *sl*), but less restrictively, we define **cryptic genetically isolated units** (CGI) (*ss* and *sl*) as entities that appear to be reproductively isolated *in fact* but which may potentially interbreed following range extension or after the disappearance of a geographical barrier

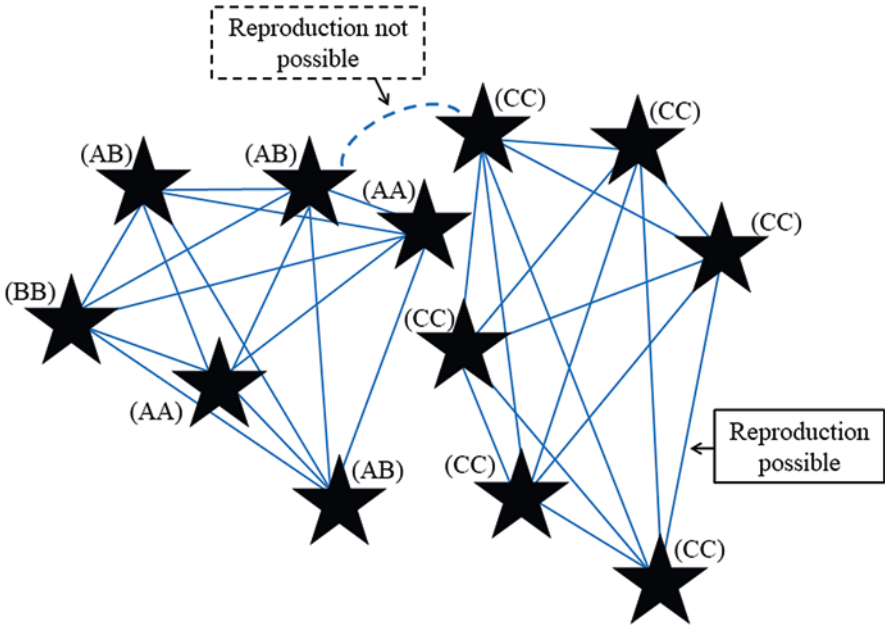


Fig. 4.1 One nominal species composed of two cryptic species: 12 individuals are represented by identical black stars (to illustrate their belonging to the same nominal species). Thin blue lines join all pairs of individuals that could potentially reproduce together and which thus belong to the same biological species. The curved dashed line joins two reproductively incompatible individuals (not all such cases are represented for clarity). Since there are two biological species, the nominal species is indeed a complex of two cryptic species (until a taxonomic revision eventually creates two nominal species). In parentheses are the individual genotypes at a codominant diagnostic locus (cf Sect. 4.3)

(Table 4.1). CS are particular cases of CGI but the problems and questions posed by CS and by the CGI that are only extrinsically isolated are essentially identical. Many reported CS (or putative CS) in the literature are indeed CGI (or *putative* CGI). We also consider *putative* CS and *putative* CGI, which are cases where the proof of genetic isolation is lacking although data suggest it may exist, because such cases are numerous and, generally, genetic isolation is confirmed when genetic information is supplemented with other types of data (cf. Sect. 4.3).

Putative CGI are being identified at an increasing rate owing to the development of genetic tools (Bickford et al. 2007; Fišer et al. 2018; Pfenninger and Schwenk 2007). Particularly in the marine realm, CS (*a fortiori* CGI) may be the rule rather than the exception ((Knowlton 1993); a seminal paper cited about 1000 times and (Nygren 2014)). One of the first marine species for which a whole genome was sequenced, the ascidian *Ciona intestinalis*, is indeed a complex of cryptic species (Nydam and Harrison 2011; Roux et al. 2013) that diverged particularly anciently (more than 10 Ma) and coexist in various regions of their distribution ranges. Interestingly, the fact that there were CS in this nominal species was ignored during the genome sequencing project and for many years despite the fact that this species was already the subject of numerous costly investigations. Our goal in this

Table 4.1 Classification of types of CGI (including putative cases) based on available knowledge and crossing the genetic isolation (GI) criteria (rows) and the morphological differentiation (MD) criteria (columns). The lower and isolated row does not belong to the classification itself but illustrates the possible causes of the origin of CGI. “BS” stands for biological species

MD \ GI	No morphological polymorphism		No morphological differentiation		Statistical morphological differentiation		Diagnostic morphological differences	
	0		1		2		3	
Biological Species	A	<i>CS sensu stricto</i>	<i>CS sensu stricto</i>	<i>CS sensu stricto</i>	<i>CS sensu stricto</i>	<i>CS sensu lato</i> Taxonomic revision should define new species		
	B	<i>CGI sensu stricto</i>	<i>CGI sensu stricto</i>	<i>CGI sensu stricto</i>	<i>CGI sensu stricto</i>	<i>CGI sensu lato</i> Should be considered at least as sub-species		
Putative Genetic Isolation	C	Independent markers needed (genetic or not) to assess GI Many such cases in the “cryptic species” literature			/		/	
No genetic differentiation	D	(Not relevant)	(Not relevant)	(Not relevant)	(Not relevant)	(Not relevant)		
Some possible causes (cf. section 4.4)		Stabilizing selection	Recently diverged entities	High effective sizes	Advantageous morphological polymorphism	Poor taxonomy		

chapter is not to participate in the debate about species concepts but to highlight problems (practical) and questions (theoretical) raised by the existence of CGI, with a particular effort to clarify the variety of causes generating CGI and CS and the features of CGI and CS that are useful to identify in order to explain their origins.

We will thus explain (i) why it is important to take CGI (and in particular CS) into account (identifying practical problems related to the assessment of biodiversity and ecosystem functioning, and theoretical problems for the understanding of community dynamics, biological evolution, etc.), (ii) how to detect CS or CGI (which is a dual task, implying both the distinction of biological species or genetically isolated entities and the characterization of morphological differentiation), (iii) how to correct inferences that are faulty due to CGI, and how to predict CGI occurrences and characteristics, which are similar questions that both require understanding of the factors responsible for the occurrence of CGI. These factors

include human factors related to science history, and biological factors, such as the geographical distribution, habitat and life history traits of the species. Finally, we will present the results of a preliminary survey of the literature on marine species.

4.2 Why It Is Important to Recognize Cryptic Species

CGI and particularly CS challenge biodiversity estimations and, potentially, biodiversity management in several important ways. Figure 4.2 illustrates some of the consequences of (ignoring) CS on fundamental biodiversity parameters. What matters is that when these parameters are erroneous, the estimation of vulnerability (of

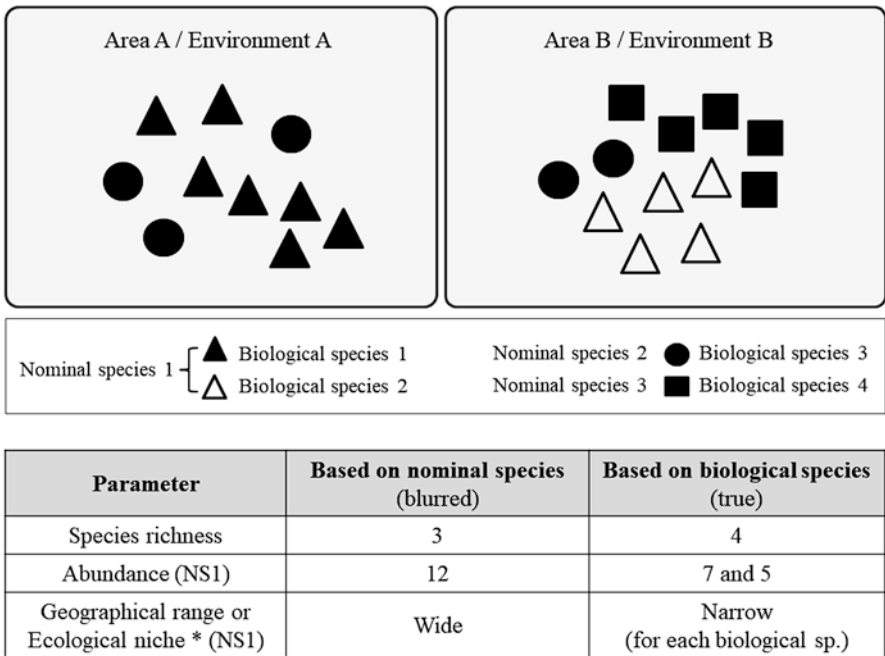


Fig. 4.2 Ignoring CGI has consequences on both assemblage parameters (e.g. species richness) and biological parameters (e.g. abundance, geographical range or ecological niche) defined for a given species. The figure represents hypothetical distributions and abundances of 3 nominal species, “nominal species 1” being a complex of two cryptic species (biological species 1 and 2). (*): The two separate zones (A and B) in which the individuals are distributed may represent either distinct geographic areas or distinct environments (i.e. habitats or ecological niches). We represented a situation where CGI have allopatric distributions or differentiated niches because these are the problematic cases, but there are situations where the CGI of a given species complex have the same geographical range or ecological niche. “NS” stands for nominal species

a species or an ecosystem) is wrong, and management measures based on these parameters may be inefficient or even deleterious.

The most conspicuous consequence of ignoring CGI is *an underestimation of species number* in a community or in an ecosystem because one nominal species is composed of several biological species. From a common biodiversity conservation point of view, this error would result in being more pessimistic than we should be about species richness in an area, species richness often being considered as a proxy of good ecological status or as a parameter to maximize. A direct corollary of the underestimation of species numbers is *the overestimation of the abundance for individual species* (by comparison to the nominal species abundance). In this case, the bias is toward undue optimism about a species' conservation status. If, instead of having one species with $2N$ individuals, there are two separate entities of N individuals, the global risk of extinction at the level of the nominal species (i.e. pooling the two biological species) may change, depending on the vulnerability component considered (e.g. genetic diversity, or inbreeding rate), for the following reason. The probability of adaptation to a change in environment is proportional to the genetic diversity within the species or the population. It is well known from population genetics theory that a metric of genetic diversity, namely nucleotide diversity (average number of nucleotide differences between two random individuals or gametes), is proportional to effective size (which, everything else being equal, is proportional to census size). Thus, in our hypothetical situation, each CGI has half the nucleotide diversity of the nominal species as a consequence of having half the number of individuals compared to the nominal species (we emphasize that this is totally compatible with the fact that most alleles at most loci may be shared among CGI). Since there are two CGI, there may be no consequences of ignoring CGI: each CGI has twice the risk of going extinct by lack of adaptive nucleotide diversity but there are two species, so globally the probability of losing the whole species complex is the same as would be estimated ignoring CGI. However, there are other components of vulnerability where small population sizes are not compensated by the number of species, such as inbreeding. In hypothetical populations of N and $2N$ individuals, the probabilities of self-reproduction are respectively $(1/N)^2$ and $(1/2N)^2$, the latter equaling $1/4 (1/N)^2$, which is a quarter of the former. Each CGI in this example therefore has a selfing probability four-fold higher than believed when ignoring that the nominal species is split in two, thus the vulnerability component is multiplied by four for each CGI which is not compensated by the presence of two (not four) CGI.

Another frequent consequence of ignoring CGI is an overestimation of the geographical range of a species: instead of a widespread (even cosmopolitan) species, there may be several geographically restricted species, allopatrically distributed or displaying partial sympatry (Egea et al. 2016; Eme et al. 2018). Again, this results in a systematic underestimation of the vulnerability of a species, particularly from a regional point of view because species with smaller geographical ranges are more vulnerable to environmental change and more threatened by extinction.

CGI may also lead to confounding numerous specialized species as a single generalist species (Morard et al. 2016), which is typically less sensitive to environmental change (Büchi and Vuilleumier 2014). More generally, functional diversity estimates may be affected depending on niche differentiation between CS: the competitive exclusion theory implies that sympatric CS may have diverged in the way they exploit limiting resources (otherwise one species would have eliminated the other by outcompeting it), with the consequence that the average niche widths of these CS may be overestimated (Van Campenhout et al. 2014) and as a result, vulnerability to perturbations would be underestimated. However, non-equilibrium situations, or more generally the neutralist theory of biodiversity, supported by many empirical studies (Hubbell 2001), prevent us from taking for granted that the ecological niches of all sympatric CS of a given complex have diverged. However, when CS share the same niche, there are also mistakes in assessing functional diversity because functional redundancy -the fact that several species ensure the same function in the ecosystem and may compensate one another in case one of them is going extinct- is underestimated when CS are ignored.

Another important element for bioconservation is the connectivity pattern of species' populations (i.e. the exchange of migrants able to reproduce with local individuals among distinct populations). The realized connectivity among populations, inferred by population genetics studies, is a key piece of information guiding the design of networks of protected areas. Inferred connectivity patterns may be erroneous when CS are ignored (Pante et al. 2015): for instance, if in two sympatric CS, samples from one area contain, by chance, only individuals of one species, and samples from another area individuals from the other species, genetic differentiation may appear very high, even if individuals migrate extensively and reproduce randomly among those areas (panmixia).

Thus far we have taken the viewpoint of community ecology, but biases induced by CGI also impact stock management of exploited species (population and range size overestimations, realized connectivity underestimations). Lastly, numerous parasites (including human parasites) are complexes of CS which may affect the efficiency of treatments (Tibayrenc 1996). CGI therefore strongly impact scientific data used by biodiversity managers and medicine.

Obviously, basic biological understanding also is challenged by CGI. Without accurate taxonomy, distributional and diversity patterns can become obscured (Paulay and Meyer 2006), and variation in taxonomic opinion can be an important source of confusion in diversification analyses (Faurby et al. 2016). For instance, ignored CGI may result in incorrectly indicating that rates of speciation have decreased toward the present (Cusimano and Renner 2010), causing false inferences of major ecological and evolutionary processes.

Beyond the erroneous inferences caused by CGI, numerous CGI are not taxonomical artefacts (i.e. morphological diagnostic differences among CGI are actually absent, not just overlooked) but they result from a significant decoupling of morphological and genetic divergence (cf below) which calls for an explanation involving evolutionary forces. Such CGI thus deserve to be studied as an important

fundamental research question, not just for practical reasons (e.g. correcting biodiversity estimates).

For all these reasons, it is necessary to undertake a thorough study of the phenomenon. Various factors may cause the presence of CGI, including human factors (e.g. the particular way in which taxonomists happened to describe and delimit the nominal species) and the habitat, biogeography and biological traits of the species. Understanding how these factors determine (i) the probability of having a CGI complex, (ii) the structure of morphological diversity in the species complex, (iii) the average number of CGI per nominal species, (iv) the probability that the CGI are ecologically differentiated or not and (v) their respective geographical ranges requires a compilation of case studies and their in-depth analysis. In Sect. 4.4, we will explain the role such factors may have in theory. Since different causes lead to different patterns of CGI, it is important to classify CGI in a relevant way. Furthermore, there are many cases of putative CGI in the literature but not as many confirmed cases; it is thus important to explain how to identify them reliably (Sect. 4.3: how to detect and classify CGI).

4.3 How to Detect and Classify Cryptic Species

There are two components in the notion of cryptic species. The first and most important component is that of genetic isolation, i.e. the presence, in a nominal species, of reproductively separated entities (though this isolation may be partial), which may correspond to distinct biological species *sensu* Mayr. The first part of this Sect. 4.3.1 presents the different levels of genetic isolation or levels of evidence of genetic isolation. In the absence of any degree of genetic isolation within a nominal species, there are no CGI, even in the wide sense (*sensu lato*). The second component is morphology (Sect. 4.3.2). Although CGI are sometimes defined as distinct biological species with similar morphology, we decided to consider as CGI (but *sensu lato*) the cases where biological species are indeed differentiated morphologically, while having the same Latin name. This choice was motivated by the fact that CGIsl as defined above pose many of the practical problems posed by CGI *sensu stricto* (where the distinct genetic entities have no diagnostic morphological differences). To avoid confusion about definitions, Table 4.1 displays our nomenclature in a 2-dimensional classification of CGI.

4.3.1 Identification of Genetic Isolation and Biological Species

The following explanations naturally only hold for taxa where “reproductive isolation” has a meaning (i.e. taxa in which there is sexual reproduction) and which also have a diploid life stage (with two copies for each marker/gene).

The most direct way to assess genetic isolation between two groups of individuals is to perform controlled crosses. However, in “non-model” species, in case of reproductive failure it is often impossible to determine whether genetic isolation or experimental conditions are responsible for the absence of offspring (or even mating). Moreover, when one does not know how to define the groups of individuals (typically the case of CGIss, due to the lack of conspicuous morphological differences), the problem has no solution. This explains why CGIss have always been discovered using genetic markers (characterized in a sufficient number of individuals).

Genetic markers may come from the nuclear genome. Since the nuclear genome is diploid, individuals have two copies for each nuclear marker, inherited from the two gametes that fused to form their first cell. There are also genetic markers that come from organellar genomes (chloroplastic or mitochondrial) which are transmitted to the (diploid) individual from a single gamete, generally the maternal gamete (oocyte) for animal mitochondrial genome, and often the paternal (pollen) for chloroplastic genomes.

When two groups are fully reproductively isolated, no genetic material is exchanged across groups (except viruses or mobile elements). There are necessarily some genetic differences among groups (otherwise they could exchange genes, if they were in contact). Diagnostic markers are those for which no allele is shared between group 1 and group 2 (yet there can be several alleles per group): if you know the allele, you can assign the organism to one of the two groups precisely. Semi-diagnostic markers are markers for which at least one allele is private to a group (absent from the other groups).

Two main types of genetic markers account for most CGI discoveries. Historically, the first type of markers which demonstrated genetic isolation within many nominal species were codominant markers, which are nuclear markers that reflect the state of both the maternal and the paternal allele of an individual. Most studies reported in the seminal review of Knowlton (1993) demonstrated CGI using such markers, in particular allozymes. By contrast, a dominant marker only provides two possible phenotypes (either presence or absence of the variant): when the variant is detected, which is often symbolized by [1], one cannot determine whether the genotype is homozygous (11) or heterozygous (10); when the phenotype is not observed [0], the genotype is necessarily (00).

A given diagnostic and codominant marker is a powerful tool to detect genetic isolation. For instance, imagine a scientist characterized 200 individuals with a marker with three alleles that are diagnostic of two biological species (alleles A and B for species 1, allele C for species 2). If the sample contains individuals from species 1 and 2, the scientist may find 4 genotypes, namely AA, AB, BB and CC. A possible distribution of the individual genotypes could be 25 individual (AA), 50 (AB), 25 (BB), and 100 (CC). Genotypes AC and BC do not exist because no genetic exchange is possible between species 1 and 2. Missing genotypes can only be explained by genetic isolation. However, to establish that the alleles are diagnostic in such a case, sample size and relative frequency among species (and also relative allele frequencies within species) matter: if only 10 individuals had been genotyped, the absence of AC and BC genotypes could have resulted by chance alone (as a

result of random sampling). If species 2 was very rare in the global sample (say 7 individuals) the absence of AC and BC would not be considered evidence of genetic isolation. So conclusions are not always straightforward and require population genetic approaches where many individuals are genotyped and analyzed using relevant (basic) statistical tests. Note that semi-diagnostic markers also produce missing genotypes which may reveal the presence of genetic isolation, but they do not allow precise species delimitation based on genotypic data because some genotypes (those composed of shared alleles) can belong to both species. With dominant markers, it is not possible to identify missing genotypes (i.e. the absence of combination of some variants in a given individual from the whole population).

During the 1990's, the use of allozymes declined in favor of approaches based on DNA. These allow field collection without refrigeration and DNA characterization was greatly facilitated by the PCR technology (Avise 1994). However, at this time, current technology did not allow routine sequencing of both alleles of many diploid individuals and the commonest data produced thus became dominant markers or sequence data from a haploid genome (mitochondrial, chloroplastic) which is represented by a single gene per individual. The distribution of haploid genotypes such as (A), (B) or (C) among individuals does not reveal anything about isolated groups, in the absence of independent information, whatever their frequencies (and whatever the divergence among these alleles). In codominant markers (example just above), genetic isolation is simply deduced by the fact that some combinations of alleles are never found associated within the same individuals, which obviously cannot be assessed with haploid markers. Among such haploid markers, however, some contributed much to the detection of putative CGI. These are the markers in which the alleles were characterized by their DNA sequences, or more generally those for which it was possible to characterize distances among alleles. Imagine now that alleles A and B are very closely related DNA sequences (differing only by one out of 500 positions), and C is very different from A and B (by 20 positions) (Fig. 4.3). The temptation is great to infer that A and B belong to one species, and C to another one. In numerous studies, alternative explanations were not even considered and the presence of CGI was inferred by such patterns. But there are alternative explanations for the observation of highly divergent alleles within a single species, even when intermediate alleles are absent. For instance, a past bottleneck in the effective size of a species (high mortality events) can lead to loss of various alleles, with only a few divergent alleles remaining (for instance 2 alleles, which may differ by 10 nucleotide positions out of 500). Then, with time, new alleles arise by mutations, which differ from their parental allele by a single mutation, leading to the presence of various (e.g. 10–15) very closely related alleles (differing by a single mutation from their parental allele, because mutations rarely hit the same nucleotide position at short time intervals) for each of the two surviving ancient alleles. The typical pattern arising from this is shown in Fig. 4.3. Note that selective sweeps, i.e. the removal of genetic diversity due to spread in the species of an advantageous allele, within a single biological species can also produce similar patterns. It looks exactly the same as the result of divergence of distinct biological species. Therefore, when a pattern like Fig. 4.3 is observed, the confirmation that there are distinct bio-

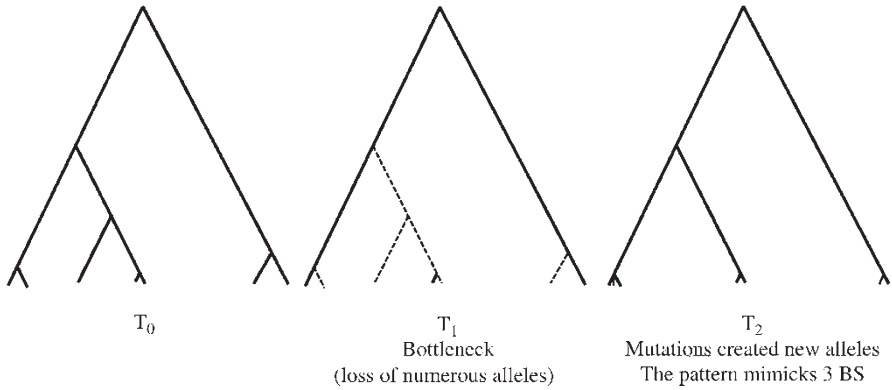


Fig. 4.3 Phylogeny of alleles may erroneously suggest the presence of several biological species. Time T_0 : Representation of a hypothetical allele phylogeny in a population of constant size, at mutation-drift equilibrium. At time T_1 a severe decrease in population size (bottleneck) causes the loss of many alleles (dashed lines). At T_2 , the population has recovered its size and mutations created new alleles closely related to the survivor alleles. The allele phylogeny mimics a pattern with 3 distinct biological species

logical species requires obtaining independent evidence supporting the genetic partition displayed by the single haploid marker, i.e. a polymorphic trait whose states appear to be linked to the marker's states. This can come from any other genetic or phenotypic (in the widest sense) marker, provided this marker is not constrained by its nature to remain tightly linked to the first marker.

As an example with genetic markers if individuals with sequences A or B (at marker 1) always have the allele X (at marker 2), and individuals with sequences C (at marker 1) have the allele Y at the independent locus (at marker 2), and if the two markers are not physically linked in the genome (which means that at each reproduction event, these two loci segregate independently and their respective alleles do not remain linked), it establishes that genes are not exchanged among groups of individuals (the first group bearing alleles A, B, and X and the other group bearing C and Y). This situation (when applied to genetic markers) corresponds to an extreme case of linkage disequilibrium. Linkage disequilibrium is defined as the non-random association between alleles at distinct loci within individuals in a population. Linkage disequilibrium, even when it is not extreme (for instance when all possible allele combinations are observed) is useful because it can detect the presence of two genetic entities (such as CGI) in a sample even when there is hybridization between them. Indeed, there are many studies reporting occasional hybridizations among distinct biological species. If such hybrids were as fertile as "pure" individuals, the two species would fuse together and after a number of generations there would be a single species. However, in most cases after long term isolation between incipient species, some incompatibility has arisen and hybrids are either sterile or less fertile. In such cases, reproductive isolation is partial, but the presence of rare hybrids does not refute the presence of reproductively isolated enti-

ties that remain genetically distinct in the long term. Even in such cases, population genetics can reveal the presence of partially isolated populations (or hybridizing species) in a sample of individuals by the detection of linkage disequilibrium between loci that are physically unlinked.

Karyotypes (shape and numbers of chromosomes), ecological characters (habitats, phenology, diet... (Johannesson 2003)) and behavior are typical phenotypic traits which can distinguish reproductively isolated units. The great majority of putative CGI detected by DNA sequences in animals were detected by mitochondrial DNA markers (haploid); thus markers from the nuclear genome (which segregates independently from the mitochondrial genome) are ideal candidates to check whether the putative biological species are true biological species (Chenuil 2012; Chenuil et al. 2010; Egea et al. 2016) as well as any phenotype not determined by the mitochondrial genome (probably more than 99.9% of phenotypes). What we called putative CGI (and putative CS), being often identified by a single molecular marker, are similar to the “Primary Species Hypotheses” of previous authors (Castelin et al. 2016; Pante et al. 2015) that need to be confirmed by independent markers or by an integrative taxonomy approach.

Apart from direct methods that are clear cut and based on a small number of markers, there is a variety of recent methods to identify and validate species delimitations using information from several independent genetic markers. Some do not require codominant markers but use DNA sequence information (Yang and Rannala 2010). For their success, some alleles must have diverged between species as a result of mutations, not only genetic drift. Other methods do not use DNA sequences but codominant markers, and can have good results even when genetic markers are not diagnostic (i.e. some alleles are shared among CGI) (Huelsenbeck et al. 2011; Jombart et al. 2010). Although these clustering methods are rarely used to assess genetic isolation, they may be the only solution for recently diverged CGI that retain ancestral shared genetic polymorphism (Weber et al. 2019). Recent methods still account for a negligible number of CGI reports.

We have thus shown how to determine genetic isolation with genetic markers and other traits recorded in samples of sufficiently numerous individuals: either using codominant markers or using distinct markers (that may be dominant) that are not inherited in a linked manner, so that their statistical association (linkage) in individuals proves that they are genetically isolated.

Let us come back to the distinction between CGI and CS (CS being particular cases of CGI). Genetic isolation may be caused by geographical isolation among groups whose genomes remain intrinsically compatible: in such cases, if individuals were put into contact (for instance by human intervention), they may be able to produce fertile offspring (thus they belong to the same biological species). We thus considered as CGI all cases where genetic isolation was established but intrinsic incompatibility was not proven. Using genetic markers exclusively, it is not possible to know whether allopatric groups are still interfertile: such groups may display

diagnostic markers as a result of genetic drift and mutation because they evolved separately for many generations. By contrast, in some (numerous) cases, genetically isolated groups detected by genetic markers are sympatric and completely intermixed in the field (Boissin et al. 2008a, b; Egea et al. 2016; Weber et al. 2014), so their reproductive incompatibility is not questioned and they deserve the status of cryptic (biological) species (CS). When the genetically isolated groups are allopatric, whether or not they kept the possibility to interbreed has few consequences for biodiversity characterization at the community level since most consequences highlighted in Sect. 4.2 still hold (e.g. range overestimation). However, the distinction is important for practical aspects of bio-conservation: in a case of strong bottlenecks endangering one geographical group, artificial introduction of individuals can be envisaged (to help restoring population size) from the other geographical group only when transplanted individual are able to reproduce with indigenous ones, thus not for actual CS.

To conclude, a practical way to classify the type of structuration within a nominal species according to genetic isolation is the following one:

Level A (biological species): True genetic isolation is shown by markers and intrinsic incompatibility is confirmed between entities (either by the observation of the genetically isolated entities in sympatry, or by controlled crosses).

Level B (genetic isolation, putative biological species): genetic isolation is confirmed (either established by a single codominant genetic marker or by an association of a genetic marker with another independent “marker”, which could be genetic, morphological, ecological or behavioral) **but** it distinguishes groups that are in allopatry, so **the status of biological species *sensu* (Mayr 1942) requiring intrinsic incompatibility** (and see Wheeler and Meier (2000)) **cannot be confirmed.**

Level C (Putative genetic isolation): putative genetic isolation that needs confirmation. These cases correspond to a high divergence among alleles in haploid or dominant markers (cf. Fig. 4.3) which has not been confirmed by any independent marker.

Level D (No genetic isolation evidence): Absence of any significant genetic differentiation within the nominal species with available genetic markers (or phenotypic characters). This does not allow rejecting the hypothesis that there are some biological species within the nominal species; we simply have no indication that there are some which need to be delimited.

This classification is a practical one which reflects available knowledge on a given nominal species. For instance, a nominal species classified as level D for genetic isolation may indeed correspond to true biological species but we lack data to confirm it. This classification will be useful when reviewing literature published on CGI because many studies report “cryptic species” while evidence of genetic isolation does not go beyond level C (i.e. genetic isolation needs to be confirmed by an independent marker, genetic or not).

4.3.2 Morphological Differentiation

Independently of the level of genetic differentiation among some groups within a nominal species, their morphological variation can be studied using various types of characters: some studies consider only very conspicuous external characters, others focus on the characters traditionally used to diagnose the species in the genus or family to which the nominal species belongs, while other ones endeavor to seek any possible character in order to find some characters corroborating groups revealed by genetic markers. For a given sample of a given nominal species, morphological differentiation and polymorphism depend on the (set of) character(s) used.

For instance, in spatangoid sea urchins, species are described and diagnosed by morphological indices from the test (i.e. the skeleton). Egea et al. (2016) revealed CS in *Echinocardium cordatum* using morphological indices from test shape: they did not find a single diagnostic character (despite the fact that morphological differentiation among CS was highly significant statistically), although sperm morphology (requiring microscopic observations) would probably reveal diagnostic differences (Drozdov and Vinnikova 2010). For taxonomists, fidelity in considering a set of characters has some justification: for example, in sea urchins, using test shape permits analyses combining extant and fossil specimens. Sperm morphology cannot be used on fossils because sperm lack hard and fossilizable structures (and also because of their microscopic size).

We propose the following classification to characterize morphological variation and differentiation among groups in a nominal species. What we name “groups” are entities which were necessarily defined independently of morphology, generally from genetic markers. This classification considers both morphological variability within groups and morphological differentiation among groups because both are relevant to interpret the nature of the evolutionary forces impinging on the evolutionary trajectory followed by the nominal species under study. As for genetic markers, the notion of diagnosticity for a morphological marker is crucial. It is useful to distinguish a situation with statistically significant morphological differentiation among groups, in the absence of diagnostic characters. For example, multivariate analyses using a set of morphological characters correctly assign more than 97% of the specimens to their genetic CS in *E. cordatum*, yet for each of the 20 morphological indices, values overlap among CS (Egea et al. 2016).

Level 0: No morphological polymorphism for this character in the nominal species, thus no differentiation among groups.

Level 1: Presence of morphological polymorphism but no differentiation among groups (not even a statistical differentiation).

Level 2: Significant morphological differentiation among groups, but no diagnostic character among groups (e.g. character values overlap for quantitative characters).

Level 3: Diagnostic morphological differences among groups.

Here again, as for the genetic component, sample sizes are crucial: it is not possible to determine if a marker is diagnostic when it was characterized in too few individuals. Beyond sample size, sample variety is important; in fact, given that individuals from a field sample may be close relatives, it is desirable to collect several field samples from reasonably distant locations. For instance, a morphological character (radial shield) appeared diagnostic of two brittle-star CS in Crete and was supported by large sample sizes (Weber et al. 2014) although this was not the case in other regions (Stohr et al. 2009).

Crossing the genetic and the morphological differentiation components, using the levels defined above, we obtain a table which provides a bi-dimensional classification of nominal species regarding the phenomenon of “cryptic species” (or CGI) (Table 4.1). Further considerations based on the different cells (or ranges of cells) from Table 4.1 rely on the assumption that the morphological differentiation status reported corresponds to the most discriminating morphological marker available in the nominal species and that such characters were investigated seriously enough. This condition is very constraining when performing a review of the literature: as shown by our preliminary survey, many studies lack sufficient detail regarding which characters were looked at and many of them do not even name any morphological character, yet conclude the absence of morphological differences among species. Therefore, rigorously establishing the absence of morphological differentiation (or diagnostic differences) within a nominal species may be impossible in the *absolute*: it is rarely possible to rule out the objections that other characters (microscopic ones, or from transitory life stages) which could have revealed stronger differentiation were dismissed/overlooked. But what is relevant for an evolutionary biology understanding of morphological evolution is to establish that the ratio of “morphological differentiation/genetic differentiation” is significantly different in the studied species than in other closely related taxa. The ideal approach to establish the morphological differentiation status in a nominal species thus requires morphological analyses of both numerous specimens from the studied nominal species as well as that of some specimens from at least one other, closely related, nominal species. This was done in (Egea et al. 2016): genetic distances between CS of the sea urchin *E. cordatum* are greater than those observed between two nominal species of another spatangoid sea urchin genus, namely *Spatagus purpureus* and *S. multispinus*.

The right-hand column in Table 4.1 (MD_3) corresponds to cases with diagnostic morphological differences. When diagnostic morphological differences confirm biological species, the possibility of having CS *sensu stricto* is ruled out, but we call such cases CS *sensu lato* because there are biological species lacking the taxonomical status of nominal species. The nominal species and its component CS are thus in need of taxonomic revision. There can be no cases in category C3 (putative genetic isolation) because, as explained above, a morphological difference diagnostic of the genetic groups (assuming this morphological character is not encoded by genes linked to the genetic marker) automatically confirms genetic isolation: this corresponds to the B3 category, in which genetic isolation is established but genetic analyses were not performed on sympatric samples so that the possibility of interbreeding, if individuals were in contact, cannot be discarded. When genetic groups are in allopatry, B2 cases correspond to sub-species.

Columns MD_0, 1 and 2 are cases without diagnostic morphological differences: these cases, when the presence of distinct biological species is confirmed (i.e. in the first row, GI_A) correspond to CS *sensu stricto* because a traditional taxonomical diagnosis of morphological species is not possible, due to lack of diagnostic morphological characters. Lower rows may also be CSs but genetic evidence is lacking to establish the presence of biological species. GI_B cases (proven genetic isolation, possible biological species), in the absence of diagnostic morphological differentiation, can be called “cryptic genetically isolated entities” (category B0 or B1). For many questions regarding biological evolution, these cases are equivalent to established biological species and should be included in meta-analyses aimed at testing hypotheses regarding the coupling of morphological and genetic divergence. Like for (C3), there are no cases in category (C2) because significant morphological differentiation among genetic groups constitutes evidence of a certain degree of genetic isolation that may only be partial (as for instance when hybridization is possible and hybrids have a lower fitness).

Two cells with putative genetic isolation and no significant morphological differentiation (C0 and C1) may be CSs but are not confirmed. Since the literature on animal CS contains many such cases, mostly from mitochondrial DNA markers, and since, when independent markers are available in addition to mitochondrial markers, they confirm genetic isolation rather frequently, we consider that such cases are worth being reported and analyzed in meta-analyses, provided their lower level of evidence of genetic isolation is recorded.

When no polymorphism at all is observed within the nominal species for the morphological character considered (left column of Table 4.1) one may just consider that information is lacking and interpretations are not possible. However, when the morphological character(s) considered is typically one that usually displays a certain amount of variability within species or that differentiates species in other, closely related nominal species, the absence of polymorphism itself can be considered informative. This leads us to part 4, where we discuss possible causes generating CGI.

4.4 Identifying the Multiple Causes of Cryptic Species

The causes of the presence of CS or CGI may be related to our taxonomic activities or to the species themselves. In the first case, they are somehow inherent to the taxonomic process (i.e. the human process of delimiting nominal species, which however may in some cases be affected by features of the species or their habitats). In the second case either they correspond to recent (young) divergences or they reflect a slow-down in the accumulation of diagnostic differences or a slow-down in morphological divergence relative to genetic divergence. After explaining possible causes and explaining how biological or habitat factors may trigger such phenomena, we explain how to determine if each of these causes is likely to explain a CS or a CGI case. The different causes and their hierarchy are summarized in Box 4.1.

Box 4.1: Classification of the Main Causes of CS1. *Taxonomic work is needed*

- 1.1. *Formal description of new nominal species is needed (for CSSl only)*
- 1.2. *Other taxonomic cause (character choice/availability, lack of samples)*
 - 1.2.1. *Technology available for observation when the nominal species was described*
 - 1.2.2. *Prevailing theories of nature and species origins when the nominal species was described*
 - 1.2.3. *Accessibility of habitats when nominal species was described*
 - 1.2.4. *Availability, quality and nature (natural selection targets / selectively neutral) of morphological characters in the group studied*

2. *Other causes than taxonomic process*

- 2.1. *Recent divergence*
 - 2.1.1. *low dispersal*
 - 2.1.2. *fragmented habitat or active landscape dynamics*
- 2.2. *True slow-down of ratio Morphological divergence/Genetic divergence*
 - 2.2.1. *natural selection*
 - 2.2.1.1. *stabilizing (in narrow niches)*
 - 2.2.1.2. *diversifying (in generalists, broadcasters...)*
 - 2.2.2. *selective neutrality of morphology (high N_e)*

4.4.1 Taxonomic Process

There are two distinct cases where taxonomic processes (i.e. the way species were delimited) are responsible for the presence of CGI. In the first case, cryptic species *sensu lato* (or CGIsl) do indeed display diagnostic morphological differences corresponding to biological species (or to units displaying genetic isolation). These cases are thus just in need of a formal description of the morphological biological species or an upgrade to the status of nominal species (or the status of sub-species, for CGI which are not CS). A second situation corresponds to cases where taxonomy failed to reveal diagnostic or differentiated morphological characters in true biological species or in CGI for various reasons discussed below which are inherent to: (1) technology available for observation when the nominal species was described, (2) prevailing theories of nature and species origins when the nominal species was

described, (3) accessibility of habitats when nominal species was described and (4) availability, quality and nature (natural selection targets/selectively neutral) of morphological characters in the group studied.

1. Technology available for observation at time of description may explain many CGI cases. Species that were described in times when (or in countries where) microscopes were not available may not have the same range of characters at their disposal to delimit morphological species. Indeed, the year in which a species was described represents a rich source of information to investigate the effects of science history in general on the presence of CGI (e.g. (Strand and Panova 2015)).
2. Nominal species of multicellular organisms correspond to the so-called “morphological species” or “morphospecies” (in more than 99.99% of nominal species) and morphological species may not correspond to biological species. Such discrepancies may lead to the presence of cryptic species *sensu stricto* but also to the opposite phenomenon (e.g. males and females, or young stages and adults, have been erroneously described as distinct species in various groups (Johnson et al. 2009)). Indeed, different species concepts may delimit species in different ways (Agapow et al. 2004). Depending on the groups, the morphological characters used to diagnose the species (and define species boundaries) may have benefitted from a cladistics approach (Hennig 1950), in which case they are more likely to reflect phylogenetic species (and also, to a lesser extent, biological species). Although the “phylogenetic species concept” includes a wide spectrum of definitions (Agapow et al. 2004; Wheeler and Meier 2000), in practice, it is often invoked (explicitly or not) to claim the presence of (cryptic) species on the basis of a phylogenetic tree inferred from a single molecular marker. Single-marker-phylogenetic-species boundaries may not delimit genetically isolated entities (cf 3–1 and Fig. 4.3), thus disagreeing with the “biological species concept”. In our Fig. 4.3 example, some widely used automatic methods of species delimitation such as the ABGD (Puillandre et al. 2012) may erroneously indicate the presence of 3 putative species. However, the formal/official description of nominal species based on molecular markers is very rare in multicellular organisms and in such cases, care is taken to use several markers (Meyer-Wachsmuth et al. 2014). Indeed, using single marker phylogenies potentially causes false reports of CGI.
3. Accessibility to an environment might limit the number of samples available for morphological analyses or cause specimen damage. Such accessibility limitations may contribute to the abundance of CGI in some environments (e.g. deep sea organism destruction by strong decrease in pressure when collected (Vacelet, 2006)). This may help to explain the high frequency of CGI in the marine environment (Barberousse and Bary 2015; Luttikhuisen et al. 2011).
4. Depending on the taxon under consideration, the morphological characters used for species diagnosis are more or less reliable. For instance, some characters may be the targets of natural selection, thus may fail to distinguish entities that have a similar niche component as a result of evolutionary convergence or stabi-

lizing selection: beak shapes in a group of birds having a similar diet may not allow species distinction, because natural selection constrains beak shapes to remain adapted to collect and grind their food. Because humans use visual information for nominal species delimitation, animals that use visual cues for mate recognition (such as vertebrates) are also much less likely to form CGI than animals that rely entirely on chemical cues for mating, such as marine invertebrates (e.g. spawning is generally triggered by chemical signals, and gametes from both sexes themselves are attracted by chemical signals (Weber et al. 2017)). Tiny organisms provide fewer characters that can be used for diagnosis, parasites often have lost many morphological characters with respect to their free-living relatives, because their bodies are simplified, having lost some major functions, etc.

4.4.2 Other Causes Besides the Taxonomic Process

Some CS or CGI are not explained by weaknesses of the taxonomic process. These are necessarily CSs or CGIss, where diagnostic characters are lacking to distinguish completely or partially genetically isolated entities.

4.4.2.1 Recent Divergence

One possible explanation for the existence of CSs (or CGIss) is the young age of divergence. Recently diverged species are more likely when speciation rates are high. Thus, factors promoting allopatric speciation may be frequently associated with CSs and more generally CGIss. Low dispersal as well as habitat fragmentation are the most conspicuous candidate factors. Thus, a review of CGIss may report dispersal ability as well as the habitat fragmentation for all cases.

4.4.2.2 Deceleration in the Accumulation of Diagnostic Morphological Differences or in Morphological Divergence Relative to Genetic Divergence

When divergence is not recent and a poor taxonomy is not involved, CSs or CGIss thus reflect an actual slow-down in the ratio of morphological over genetic diversity or divergence that persisted long enough to produce the observed pattern.

Natural selection on morphological characters may be responsible for the absence of diagnostic characters among species. (1) The cause which is most often invoked to explain such cases is stabilizing selection (Charlesworth et al. 1982; Lee and Frost 2002). When morphology is strongly constrained by natural selection, morphological variation is very low within species and following speciation, daughter species are

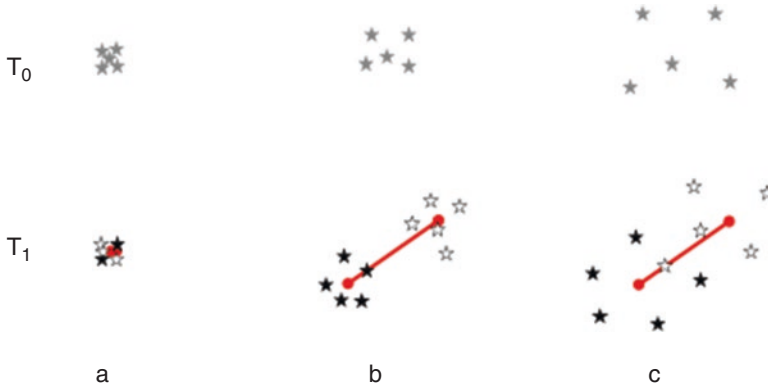


Fig. 4.4 Different patterns of morphological differentiation between genetically isolated entities. Individuals from the two entities are represented by filled or empty stars. Their relative position in the plane reflects their morphological similarity (e.g. horizontal and vertical coordinates may represent values for two continuous morphological characters). For a given divergence time (from T_0 before divergence to T_1 after several generations), the distribution of morphological variation within and among the two genetically isolated entities may correspond to one of three main patterns resulting from four main processes. (a) The absence of (or negligible) morphological diversity within species, probably resulting from stabilizing selection, impedes their divergence. (b) A standard situation without CGI *sensu stricto*. There is morphological variation within species and the genetically isolated entities diverged morphologically so there is no overlap between them (the character represented by the horizontal coordinate is diagnostic) that are therefore not CGI *sensu stricto*. (c) There is a higher level of diversity within species compared to case b, so despite identical divergence times compared to case b (represented by the same distance separating the barycenters represented by red dots as in case b), the morphological spaces of both species overlap and thus no diagnostic character distinguishes the species. Diversity may result from high effective sizes, or from natural selection favouring high morphological diversity (see text). We emphasize that pattern (a) could not be caused by low effective sizes: in such a case genetic drift would be high and lead to divergence at T_1 (length between the two barycenters would be higher than in b, instead of null)

similarly monomorphic and do not diverge in their shapes (Fig. 4.4). (2) Paradoxically, an opposite pattern of morphological diversity may also lead to CS. This pattern is that of a high morphological polymorphism within species, which is selectively advantageous for species submitted to strong spatio-temporal fluctuations (Egea et al. 2016). Morphological polymorphism may be achieved by two different mechanisms: environmental phenotypic plasticity (a single genotype may lead to a variety of morphologies) or presence of a variety of genes determining morphology (i.e. presence in the species of distinct alleles, or genetic variants, also called genetic polymorphism). Both mechanisms prevent the appearance of diagnostic morphological differences between species because, as a result of character polymorphism, the range of possible character states overlaps between sister-species (Fig. 4.4).

Recently, it has been suggested that neutral (i.e., non-adaptive) processes, may also lead to absence of diagnostic morphological differences among genetically isolated entities (Egea et al. 2016). Higher polymorphism at neutral loci is expected for taxa with larger effective population sizes. When such taxa speciate, ancestral polymorphism remains shared among daughter species for a higher number of

generations than in taxa with lower effective sizes. When the phenotypic traits used to diagnose species are selectively neutral this leads to an absence of diagnostic characters for longer temporal periods in the taxon with higher effective sizes, making the occurrence of CS more likely (because the taxonomists delimiting species cannot identify any diagnostic character). This novel neutral theory of morphological evolution provides a null model for the existence of CS, and may help to explain the abundance of marine CS because in the marine realm many species have high fecundities, abundances and range sizes.

Figure 4.4 illustrates the distribution of morphological diversity between two sister biological species corresponding to the above cases and compared to a species pair displaying diagnostic characters.

To summarize this section, five major types of causes correspond to the distinct levels of morphological differentiation of our classification (i.e. Table 4.1 columns): stabilizing selection for MD0, recent divergence for MD1, high effective sizes or advantageous morphological polymorphism for MD2, and poor taxonomy for MD3 (not excluding that various factors may interact).

4.4.3 *How to Determine If a Cause Is Likely to Explain a CGI Case*

Not all causes are possible for a given category of putative CGI (i.e. for a given cell or cell range in Table 4.1). The possible causes identified above are compiled in Table 4.2. For each cause, the “cell range” column displays the putative CGI category that can be explained by this cause, and which traits or factors are useful to assess the validity of the cause. Most causes can be assessed at two levels: for individual putative CGI or at a global level, in a higher order taxon. For instance, one may test whether the cryptic species observed in the species complex *Echinocardium cordatum* can be explained by stabilizing selection or not (Egea et al. 2016), but also whether CGI in the phylum Echinodermata are explained by stabilizing selection more often than expected at random. Testing the importance of a possible process globally (i.e. in generating CGI in a given higher-rank taxon) requires including in the (meta-) analysis not only the taxa for which CGI have been reported or suggested in the literature, but also all nominal species of the taxon for which genetic data have been published.

At this step of the analysis, we can list the different data fields that appear useful to include in a database aimed at studying the CS phenomenon. They should include both information enabling CGI characterization (both GI and morphological differentiation levels; Table 4.1) and information useful to determine the possible causes of the CS (Table 4.2; acknowledging the fact that most cases lack information in some fields). Potentially useful data fields include: (1) genetic marker type (haploid/diploid, codominant or not, number of markers), genetic structure (sample sizes, significant differentiation among groups, genetic diversity within populations/

Table 4.2 Possible causes (column 1) for different types of (putative) CGI (column 2) and traits or factors to check (column 3) to evaluate the validity of the hypothetical cause (rather than an exhaustive list, we proposed examples of the most relevant ones)

Cause or hypothetical process	Type of CGI (cell range in Table 4.1)	Traits or factors to check
T1 available technology	A-C x 0–3	Year (+place) of NS description
		Material needed for diagnosis (microscope, ...)
T2 history of science	A-C x 0–3	Year (+place) of NS description
		Higher order taxon name
T3 accessibility	A-C x 0–3	Habitat of species
T4 morphological character	A-C x 0–3	Organism size
		Mode of life (endosymbiotic)
		Selective neutrality of character
		Variability of character in higher rank taxon
Recent divergence	A-C x 0–2	Genetic divergence (CS of the NS + at least 1 pair of closely related sister species)
		Biogeography (CS sympatry/allopatry)
Low dispersal	A-C x 0–2	Life traits related to dispersal ability (but also to effective size: fecundity, reproduction mode)
High fragmentation	A-C x 0–2	Habitat of species
Stabilizing selection	A-C x 0	Morphological variability within BS
		Spatio-temporal variability of environment
Diversifying selection	A-C x 1–2	Same as above
		Check knowledge on species plasticity
Neutral (high Ne)	A-C x (1)-2	Morphological variability within BS
		Genetic diversity within populations/BS
		Life history traits (fecundity, reproduction variance)
		Size of geographic species range

“NS” stands for “nominal species” and “BS” stands for “biological species”

species and comparison with closely related taxa external to the nominal species if possible), (2) reproductive isolation among groups if tested by crosses, (3) ecological differentiation among groups, (4) any phenotypic differentiation (in the wide sense) that corresponds to genetic differentiation to confirm GI, (5) morphological variability within and among groups (and sample sizes), and also, when possible, in closely related pairs of sister species, (6) year and place of nominal species description, (7) nature of morphological characters analyzed, (8) habitat (physical fragmentation, accessibility), (9) biogeographical distribution (allopatry, sympatry among CGI, size of species range) and (10) life history and other biological traits (dispersal ability, fecundity, reproductive success variance, parasite or not, use of visual cues for mating).

4.5 Preliminary Results

A pilot study by undergraduate students (Délémontey et al. 2014) compiled articles reporting cryptic species in the marine realm and recorded information relative to some of these fields. This study collected useful data about the relative proportions of different cases of CS in the literature and revealed some associations among CS features, phyla, habitat and biological traits. For the pilot study, successive groups of search terms were used in Web of Science. We detail the different steps of the first search. The first step using « cryptic species » OR « sibling species » provided 11,416 papers (this was done in 2014). After adding «morpho* OR phenotyp*» (second step) 4417 articles remained, after adding «genetic OR molecular OR mitochondrial» (third step) we had 3055 papers, and with «marine OR sea OR ocean» (fourth step) 647 articles. To limit the number of papers while increasing the proportion of cases corresponding to validated CGI, we added the terms «nuclear marker* OR microsatellite* OR allozyme* OR intron OR ribosomal» (fifth step) to favor studies combining several molecular markers. This resulted in 222 articles. We carried out a second search identical to this one except that we replaced the fourth step (marine or sea or ocean) by the title of scientific journals dealing with marine biology («ANNU REV MAR SCI» OR «DEEP SEA RES» OR «ESTUAR COAST SHELF S» OR «HELGOLAND MAR RES» OR «ICES J MAR SCI» OR «J OCEANOGR» OR «J PLANKTON RES» OR «LIMNOL OCEANOGR» OR «MAR ECOL PROG SER» OR «OCEANOGR MAR BIOL» OR «CORAL REEFS» OR «MAR ECOL-EVOL PERSP» OR «MAR BIOL» OR «CAN J FISH AQUAT SCI» OR «J EXP MAR BIOL ECOL» OR «J FISH BIOL»). This second search provided 41 papers. For the last search we changed, again, the fourth step to select taxon names («Echinoderm*» OR «Echinoid*» OR «Asteroid*» OR «ophiuroid*» OR «bivalv*» OR «mollus*» OR «fish*» OR «sponge*» OR «porifera*» OR «cnidaria*» OR «coral*» OR «bryozoan*» OR «ascidia*» OR «mysidac*» OR «nematod*» OR «gastropod*» OR «copepod*» OR «amphipod*») which led to 264 articles. The fusion of the three searches provided 402 different articles. After abstract reading, we discarded papers that dealt with plants and algae, terrestrial and freshwater animals, endoparasites, protists and foraminiferans, and papers reporting new species but not CGI. These studies corresponded to 126 nominal species (556 CGI) from 86 families, 55 orders, 25 classes and 11 phyla. For all nominal species, putative CGI were defined based on genetic markers; for three nominal species controlled crosses were performed to determine CS; in 14 nominal species, there were some differences between CGI for at least one factor among ecology, reproduction, nutrition, hosts, gamete morphology, color and in 3 nominal species, CGI had distinct karyotypes. This preliminary survey confirmed that CS were present in a diversity of animal phyla and established an average of 4.41 CGI per nominal species (Fig. 4.5).

Out of 126 nominal species cases, 70 (56%) had been the subject of a morphological study: 37 of these display diagnostic differences among CGI (53%, thus they are not CGI *sensu stricto*), 16 display statistical morphological differences among CGI (23%), and 17 do not display morphological differences among CGI (24%).

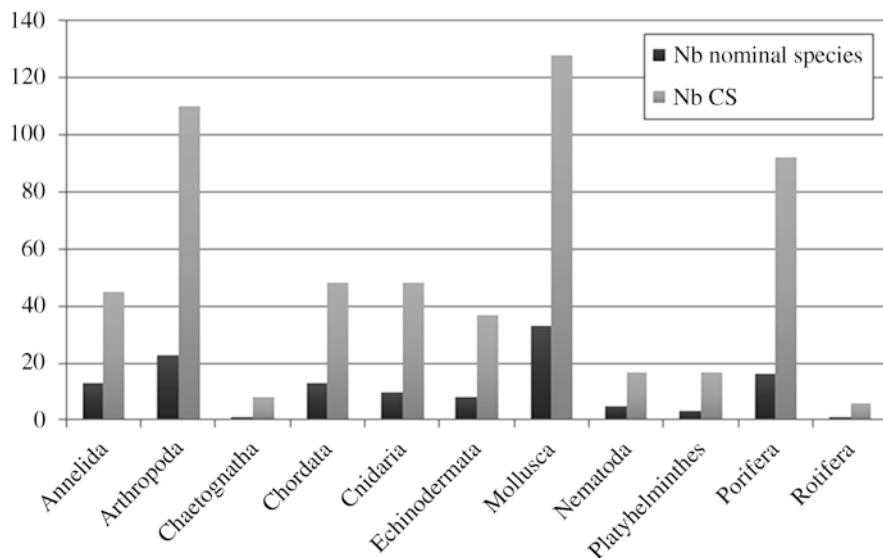


Fig. 4.5 Number of nominal species and CGI per phylum in our pilot study of 402 articles

This highlights that among reported CGI complexes, about half are just in need of taxonomic revision and may not correspond to any phenomenon of deceleration of morphological evolution. Among the 33 CGI complexes that may be CGIss, half display statistical differences in morphology and half do not display any morphological differentiation among genetic entities. These proportions are helpful to plan studies aimed at testing various hypotheses regarding the CS phenomenon. Among the hundreds of studies reporting CGI, about half may just need taxonomic revision, a quarter may be good candidates for testing hypotheses regarding natural selection, effective sizes, etc. Indeed, the categories of our classification based on crossed genetic isolation and morphological differentiation levels (Table 4.1) seem relatively well balanced. However, proportions of “diagnostic/statistically significant/not significant” morphological differences among CGI vary among phyla (these differences are statistically very significant) (Fig. 4.6).

We investigated the relative geographical distribution of CGI and their ecological differentiation and found that (i) 50% of cases have exclusively allopatric sibling-species, (ii) the ratio of cases displaying “strict allopatry” versus “sympatry” varies among phyla (this result is statistically significant), (iii) there is a higher proportion of diagnostic morphological differences in “sympatric” than in “strictly allopatric” CGI (statistically significant result), (iv) ecological differentiation within CGI is more frequent in sympatric than in allopatric CGI, supporting the competitive exclusion theory (highly significant result) which stipulates that sympatric species cannot coexist stably if they have the same niche: either they evolve distinct niches or one eliminates the other. Returning to our first section on the practical importance of CGI, *this suggests that ignoring CGI leads to underestimating not only species diversity but also local functional diversity.*

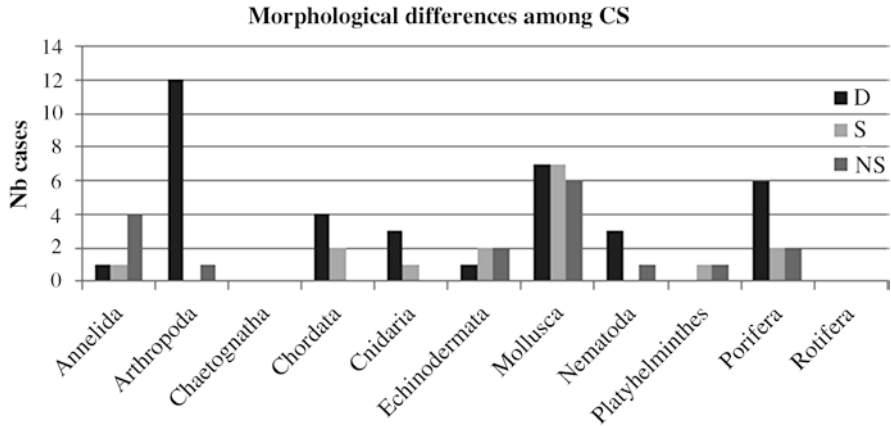


Fig. 4.6 Distribution of studies reporting CGI per phylum according to the status of morphological differentiation among CGI. Abbreviations correspondence: *D* diagnostic differences, *S* statistically significant differences, *NS* non-significant differences

To rapidly infer the ratio of morphological to genetic divergence (indirectly) we looked at (or computed) molecular phylogenies and divergences; we found that: (i) sibling species diverged more than some nominal species of the same group *in 2/3 of the cases, ruling out a “recent speciation” explanation for morphological similarity and confirming decoupling between morphological and genetic divergence* for these CGI, (ii) molecular divergence within CGI was higher for wider habitat ranges (statistically significant), and (iii) there were more diagnostic morphological differences in high dispersal taxa (statistically significant). No straightforward explanations were found for the former results. A much greater survey, also limited to marine metazoans and excluding parasites, has been carried out and its thorough analysis is ongoing (Cahill and Chenuil, unpublished). It selected 1209 studies compiled from more than 4000 titles, of which 55% report CGI, from which another 55% have morphological data, and 12% report ecological comparisons among CGI. As many studies are expected for macrophytes, perhaps more from parasites, and many additional ones would be found in terrestrial taxa. Based on these proportions, there is no doubt that scientists will be able to test many of the hypotheses raised above about factors favoring the presence of CGI in numerous phyla.

4.6 Concluding Remarks on the Use of Morphospecies for Biodiversity Assessment

Since the task is huge, one may argue that it would be more efficient to consider alternative approaches to replace the morphological identification of species in future studies of biological communities, ecosystem monitoring and conservation actions. Taxonomic sufficiency approaches, focusing on higher taxa (instead of the species level), may appear less affected by CGI. However, by lumping related

species together they often lose or bias the functionality signal (Thiault et al. 2015) which consists of the variability of ecological functions, because even closely related species frequently have distinct functional traits. Parataxonomy is another approach that eliminates the requirement of rigorous taxonomic identification: it consists of sorting samples to recognizable taxonomic units (RTU). However, the error in this approach is not predictable and depends on the sorter (Krell 2004), precluding comparisons of datasets processed by distinct persons, a big problem for monitoring programs. Neither taxonomic sufficiency nor parataxonomy allow using putative functional knowledge we may have on the entities (not necessarily “species”) recorded.

Barcoding and its derived method, metabarcoding, enable the automatic identification of species based on their DNA sequence at a given marker for which there is a huge database containing species names and their corresponding DNA sequence. Diversity estimates based on barcoding are less sensitive to CGI but have other drawbacks (Bucklin et al. 2011; Krishnamurthy and Francis 2012). Typical barcoding or metabarcoding was based on a single marker until now. The largest database is probably the 18S rDNA (and its homologous database, the 16S rDNA, for prokaryotes), which can be used in virtually all eukaryote phyla, but which sequences are not variable enough to distinguish related species within a genus and often within a family. For animals, the well-recognized “barcoding molecule” COI is much more useful than 18S due to its high variability (Chenuil 2006). Fungi and plants also have their own barcoding databases in BOLDSYSTEM (barcoding of life data system) (Ratnasingham and Hebert 2013). As explained above (Sect. 4.3), single marker data cannot establish genetic isolation. When at least another marker will have a sufficiently large database to be used in conjunction with the marker currently used for barcoding in the three main groups of living things, the identification of biological species (or GI entities) not requiring morphological identification will be possible. Another limitation of metabarcoding is its very poor representativeness of species biomass or abundances which may not be completely overcome by the use of various markers. But even with improved barcoding, understanding the discrepancy between morphological, phylogenetic, and biological species will remain necessary to validate fossil data and properly analyze the consequences of past environmental changes. This is particularly important because inferring past changes may help to predict future biodiversity responses to climate change (Condamine et al. 2013).

Once a database compiling putative CGI and containing information on GI levels, morphological differentiation, life history traits, biogeographical distribution and habitat is available, several practical questions related to bioconservation may be answered. (1) Is the error on biodiversity estimators caused by ignored CGI important or do the different errors and biases compensate each other? (2) Do barcoding approaches based on a single sequence marker represent a good solution to correct the CGI problem in common biodiversity estimates? (3) Would barcoding approaches based on *two* independent sequence markers (or more) improve biodiversity estimates? (4) Can we propose correction equations (based on meta-analysis) to solve the problem?

This study provides a robust framework to tackle the very complex question of CGI, by providing a bi-dimensional classification system, and identifying fields to be filled in a database reporting CGI cases. Our application of such a method on a pilot dataset provided promising results since the proportions of the distinct types of CGI appeared well balanced, potentially allowing the testing of all hypotheses raised in this study. Furthermore, it revealed meaningful significant associations among CGI features.

Acknowledgements A research proposal on cryptic species was selected after peer-review to be funded by the CESAB (French Centre de synthèse et d'analyse sur la biodiversité) which belongs to the FRB (Fondation pour la Recherche sur la Biodiversité). Unfortunately, no funding was provided to most laureates of this call. We nevertheless thank some of the scientists of the consortium whose interest and participation in the CRYSPIM proposal motivated us to write this chapter: Philippe Borsa, Elena Casetta, Julien Claude, Fabien Condamine, Nancy Knowlton, Pieternella Luttkhuizen, Marina Panova.

References

- Agapow, P.-M., Bininda-Emonds, O. R., Crandall, K. A., Gittleman, J. L., Mace, G. M., Marshall, J. C., & Purvis, A. (2004). The impact of species concept on biodiversity studies. *The Quarterly Review of Biology*, *79*, 161–179.
- Avise, J.C. (1994). Molecular markers, natural history and evolution. Springer Science & Business Media.
- Barberousse, A., & Bary, S. (2015). Ideal and actual inventories of biodiversity. *Riv Estet*, *59*(2), 14–31.
- Beauchard, O., Veríssimo, H., Queirós, A. M., & Herman, P. M. J. (2017). The use of multiple biological traits in marine community ecology and its potential in ecological indicator development. *Ecological Indicators*, *76*, 81–96.
- Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K., Meier, R., Winker, K., Ingram, K. K., & Das, I. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, *22*, 148–155.
- Boissin, E., Feral, J. P., & Chenuil, A. (2008a). Defining reproductively isolated units in a cryptic and syntopic species complex using mitochondrial and nuclear markers: The brooding brittle star, *Amphipholis squamata* (Ophiuroidea). *Molecular Ecology*, *17*, 1732–1744. <https://doi.org/10.1111/j.1365-294X.2007.03652.x>.
- Boissin, E., Hoareau, T. B., Feral, J. P., & Chenuil, A. (2008b). Extreme selfing rates in the cosmopolitan brittle star species complex *Amphipholis squamata*: Data from progeny-array and heterozygote deficiency. *Marine Ecology Progress Series*, *361*, 151–159. <https://doi.org/10.3354/meps07411>.
- Büchi, L., & Vuilleumier, S. (2014). Coexistence of specialist and generalist species is shaped by dispersal and environmental factors. *The American Naturalist*, *183*, 612–624. <https://doi.org/10.1086/675756>.
- Bucklin, A., Steinke, D., & Blanco-Bercial, L. (2011). DNA barcoding of marine metazoa. *Annual Review of Marine Science*, *3*, 471–508. <https://doi.org/10.1146/annurev-marine-120308-080950>.
- Castelin, M., Van Steenkiste, N., Pante, E., Harbo, R., Lowe, G., Gilmore, S. R., Therriault, T. W., & Abbott, C. L. (2016). A new integrative framework for large-scale assessments of biodiversity and community dynamics, using littoral gastropods and crabs of British Columbia, Canada. *Molecular Ecology Resources*, *16*, 1322–1339.

- Charlesworth, B., Lande, R., & Slatkin, M. (1982). A neo-darwinian commentary on macroevolution. *Evolution*, *36*, 474–498. <https://doi.org/10.2307/2408095>.
- Chenuil, A. (2012). How to infer reliable diploid genotypes from NGS or traditional sequence data: From basic probability to experimental optimization. *Journal of Evolutionary Biology*, *25*, 949–960. <https://doi.org/10.1111/j.1420-9101.2012.02488.x>.
- Chenuil, A. (2006). Choosing the right molecular genetic markers for studying biodiversity: From molecular evolution to practical aspects. *Genetica*, *127*, 101–120. <https://doi.org/10.1007/s10709-005-2485-1>.
- Chenuil, A., Hoareau, T. B., Egea, E., Penant, G., Rocher, C., Aurelle, D., Mokhtar-Jamai, K., Bishop, J. D. D., Boissin, E., Diaz, A., Krakau, M., Luttkhuizen, P. C., Patti, F. P., Blavet, N., & Mousset, S. (2010). An efficient method to find potentially universal population genetic markers, applied to metazoans. *BMC Evolutionary Biology*, *10*, 276. <https://doi.org/10.1186/1471-2148-10-276>.
- Condamine, F. L., Rolland, J., & Morlon, H. (2013). Macroevolutionary perspectives to environmental change. *Ecology Letters*, *16*, 72–85. <https://doi.org/10.1111/ele.12062>.
- Cusimano, N., & Renner, S. S. (2010). Slowdowns in diversification rates from real phylogenies may not be real. *Systematic Biology*, *59*, 458–464.
- Délémontey, N., Du Salliant du Luc, E., & Fanton, H. (2014). Recherche de facteurs associés à la présence d'espèces cryptiques en mer par analyse de données bibliographiques (encadrant: Chenuil).
- Drozhdov, A. L., & Vinnikova, V. V. (2010). Morphology of gametes in sea urchins from Peter the Great Bay, Sea of Japan. *Russian Journal of Developmental Biology*, *41*, 37–45.
- Egea, E., David, B., Chone, T., Laurin, B., Feral, J. P., & Chenuil, A. (2016). Morphological and genetic analyses reveal a cryptic species complex in the echinoid *Echinocardium cordatum* and rule out a stabilizing selection explanation. *Molecular Phylogenetics and Evolution*, *94*, 207–220. <https://doi.org/10.1016/j.ympev.2015.07.023>.
- Eme, D., Zagamajster, M., Delic, T., Fiser, C., Flot, J.-F., Konecny-Dupre, L., Palsson, S., Stoch, F., Zaksek, V., Douady, C. J., & Malarid, F. (2018). Do cryptic species matter in macroecology? Sequencing European groundwater crustaceans yields smaller ranges but does not challenge biodiversity determinants. *Ecography*, *41*, 424–436. <https://doi.org/10.1111/ecog.02683>.
- Faurby, S., Eiserhardt, W. L., & Svenning, J.-C. (2016). Strong effects of variation in taxonomic opinion on diversification analyses. *Methods in Ecology and Evolution*, *7*, 4–13.
- Fišer, C., Robinson, C. T., & Malarid, F. (2018). Cryptic species as a window into the paradigm shift of the species concept. *Molecular Ecology*, *27*, 613–635. <https://doi.org/10.1111/mec.14486>.
- Gotelli, N. J., & Colwell, R. K. (2001). Quantifying biodiversity: Procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, *4*, 379–391. <https://doi.org/10.1046/j.1461-0248.2001.00230.x>.
- Hennig, W. (1950). *Grundzüge einer Theorie der phylogenetischen Systematik*. Berlin: Deutscher Zentralverlag.
- Hubbell, S. P. (2001). *The unified neutral theory of biodiversity and biogeography*. Princeton: Princeton University Press.
- Huelsenbeck, J. P., Andolfatto, P., & Huelsenbeck, E. T. (2011). Structurama: Bayesian inference of population structure. *Evolutionary Bioinformatics*, *55*. <https://doi.org/10.4137/EBO.S6761>.
- Johannesson, K. (2003). Evolution in *Littorina*: ecology matters. *Journal of Sea Research*, *49*, 107–117. [https://doi.org/10.1016/S1385-1101\(02\)00218-6](https://doi.org/10.1016/S1385-1101(02)00218-6).
- Johnson, G. D., Paxton, J. R., Sutton, T. T., Satoh, T. P., Sado, T., Nishida, M., & Miya, M. (2009). Deep-sea mystery solved: Astonishing larval transformations and extreme sexual dimorphism unite three fish families. *Biology Letters*, *5*, 235–239. <https://doi.org/10.1098/rsbl.2008.0722>.
- Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, *11*, 94. <https://doi.org/10.1186/1471-2156-11-94>.
- Knowlton, N. (1993). Sibling species in the sea. *Annual Review of Ecology and Systematics*, *24*, 189–216.

- Krell, F.-T. (2004). Parataxonomy vs. taxonomy in biodiversity studies – pitfalls and applicability of ‘morphospecies’ sorting. *Biodiversity and Conservation*, *13*, 795–812. <https://doi.org/10.1023/B:BIOC.0000011727.53780.63>.
- Krishnamurthy, P., & Francis, R. A. (2012). A critical review on the utility of DNA barcoding in biodiversity conservation. *Biodiversity and Conservation*, *21*, 1901–1919. <https://doi.org/10.1007/s10531-012-0306-2>.
- Lee, C. E., & Frost, B. W. (2002). Morphological stasis in the Eurytemora affinis species complex (Copepoda: Temoridae). *Hydrobiologia*, *480*, 111–128. <https://doi.org/10.1023/A:1021293203512>.
- Luttikhuisen, P. C., Bol, A., Cardoso, J. F., & Dekker, R. (2011). Overlapping distributions of cryptic *Scoloplos cf. armiger* species in the western Wadden Sea. *Journal of Sea Research*, *66*, 231–237.
- Mayr, E. (1942). *Systematics and the origin of species*. New York: Columbia University Press.
- Meyer-Wachsmuth, I., Curini Galletti, M., & Jondelius, U. (2014). Hyper-cryptic marine meiofauna: Species complexes in nemertodermatida. *PLoS One*, *9*, e107688. <https://doi.org/10.1371/journal.pone.0107688>.
- Morard, R., Escarguel, G., Weiner, A. K. M., Andre, A., Douady, C. J., Wade, C. M., Darling, K. F., Ujiie, Y., Sears, H. A., Quillevère, F., de Garidel-Thoron, T., de Vargas, C., & Kucera, M. (2016). Nomenclature for the nameless: A proposal for an integrative molecular taxonomy of cryptic diversity exemplified by planktonic foraminifera. *Systematic Biology*, *65*, 925–940. <https://doi.org/10.1093/sysbio/syw031>.
- Nydam, M. L., & Harrison, R. G. (2011). Introgression despite substantial divergence in a broadcast spawning marine invertebrate. *Evolution*, *65*, 429–442.
- Nygren, A. (2014). Cryptic polychaete diversity: a review. *Zoologica Scripta*, *43*, 172–183.
- 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. *Molecular Ecology*, *24*, 525–544. <https://doi.org/10.1111/mec.13048>.
- Paulay, G., & Meyer, C. (2006). Dispersal and divergence across the greatest ocean region: do larvae matter? *Integrative and Comparative Biology*, *46*, 269–281.
- Pfenninger, M., & Schwenk, K. (2007). Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. *BMC Evolutionary Biology*, *7*, 121.
- Puillandre, N., Lambert, A., Brouillet, S., & Achaz, G. (2012). ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology*, *21*, 1864–1877.
- 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. <https://doi.org/10.1371/journal.pone.0066213>.
- Roux, C., Tsagkogeorga, G., Bierne, N., & Galtier, N. (2013). Crossing the species barrier: Genomic hotspots of introgression between two highly divergent *Ciona intestinalis* species. *Molecular Biology and Evolution*, *30*, 1574–1587.
- Stohr, S., Boissin, E., & Chenuil, A. (2009). Potential cryptic speciation in Mediterranean populations of *Ophioderma* (Echinodermata: Ophiuroidea). *Zootaxa*, *1*(20).
- Strand, M., & Panova, M. (2015). Size of genera—biology or taxonomy? *Zoologica Scripta*, *44*, 106–116.
- Thiault, L., Bevilacqua, S., Terlizzi, A., & Claudet, J. (2015). Taxonomic relatedness does not reflect coherent ecological response of fish to protection. *Biological Conservation*, *190*, 98–106. <https://doi.org/10.1016/j.biocon.2015.06.002>.
- Tibayrenc, M. (1996). Towards a unified evolutionary genetics of microorganisms. *Annual Review of Microbiology*, *50*, 401–429. <https://doi.org/10.1146/annurev.micro.50.1.401>.
- Vacelet, J. (2006). New carnivorous sponges (Porifera, Poecilosclerida) collected from manned submersibles in the deep Pacific. *Zoological Journal of the Linnean Society*, *148*, 553–584.
- Van Campenhout, J., Derycke, S., Moens, T., & Vanreusel, A. (2014). Differences in life-histories refute ecological equivalence of cryptic species and provide clues to the origin of bathyal Halomonhystera (Nematoda). *PLoS One*, *9*, e111889.

- Weber, A. A.-T., Stohr, S., & Chenuil, A. (2014). Genetic data, reproduction season and reproductive strategy data support the existence of biological species in *Ophioderma longicauda*. *Comptes Rendus Biologies*, 337, 553–560. <https://doi.org/10.1016/j.crv.2014.07.007>.
- Weber, A. A.-T., Stöhr, S., & Chenuil, A. (2019). Species delimitation in the presence of strong incomplete lineage sorting and hybridization: Lessons from *Ophioderma* (Ophiuroidea: Echinodermata). *Molecular Phylogenetics and Evolution*, 131, 138–148.
- Weber, A. A.-T., Abi-Rached, L., Galtier, N., Bernard, A., Montoya-Burgos, J. I., & Chenuil, A. (2017). Positive selection on sperm ion channels in a brooding brittle star: Consequence of life-history traits evolution. *Molecular Ecology*, 26, 3744–3759. <https://doi.org/10.1111/mec.14024>.
- Wheeler, Q., & Meier, R. (2000). *Species concepts and phylogenetic theory: A debate*. New York: Columbia University Press.
- Yang, Z., & Rannala, B. (2010). Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences*, 107, 9264–9269.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

