Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

BIODIVERSITÀ ED EVOLUZIONE

Ciclo XXIII

Settore/i scientifico-disciplinare/i di afferenza: BIO - 05

A MOLECULAR PHYLOGENY OF BIVALVE MOLLUSKS: ANCIENT RADIATIONS AND DIVERGENCES AS REVEALED BY MITOCHONDRIAL GENES

Presentata da: **Dr Federico Plazzi**

Coordinatore Dottorato **Relatore**

Prof. Barbara Mantovani **Dr Marco Passamonti**

Esame finale anno 2011

of all marine animals, the bivalve molluscs are the most perfectly adapted for life within soft substrata of sand and mud.

Sir Charles Maurice Yonge

INDEX

FOREWORD

The main scope of my PhD is the reconstruction of the large-scale bivalve phylogeny on the basis of four mitochondrial genes, with samples taken from all major groups of the class. To my knowledge, it is the first attempt of such a breadth in Bivalvia. I decided to focus on both ribosomal and protein coding DNA sequences (two ribosomal encoding genes, *12s* and *16s*, and two protein coding ones, cytochrome *c* oxidase I and cytochrome b), since either bibliography and my preliminary results confirmed the importance of combined gene signals in improving evolutionary pathways of the group. Moreover, I wanted to propose a methodological pipeline that proved to be useful to obtain robust results in bivalves phylogenesis. Actually, best-performing taxon sampling and alignment strategies were tested, and several data partitioning and molecular evolution models were analyzed, thus demonstrating the importance of molding and implementing non-trivial evolutionary models.

In the line of a more rigorous approach to data analysis, I also proposed a new method to assess taxon sampling, by developing Clarke and Warwick statistics: taxon sampling is a major concern in phylogenetic studies, and incomplete, biased, or improper taxon assemblies can lead to misleading results in reconstructing evolutionary trees. Theoretical methods are already available to optimize taxon choice in phylogenetic analyses, but most involve some knowledge about genetic relationships of the group of interest, or even a well-established phylogeny itself; these data are not always available in general phylogenetic applications. The method I proposed measures the "phylogenetic representativeness" of a given sample or set of samples and it is based entirely on the pre-existing available taxonomy of the ingroup, which is commonly known to investigators. Moreover, it also accounts for instability and discordance in taxonomies. A Python-based script suite, called PhyRe, has been developed to implement all analyses.

Plan of the Thesis

This Thesis, after a general introduction (Chapter 1), is divided into four parts, each representing the main arguments of my research during PhD. Chapter 2 is the first attempt, with a partial dataset, to draw a phylogeny of Bivalvia, especially of deeper nodes, and to establish a methodological pipeline for further studies. This part has been already published in *Molecular Phylogenetics and Evolution* (Plazzi and Passamonti, 2010). Chapter 3 is dedicated to the abovementioned "phylogenetic representativeness" and the software PhyRe. This part has also already been published in *BMC Bioinformatics* (Plazzi et al., 2010). Chapter 4 re-analyzes Bivalvia phylogeny through a larger dataset, and better specifies phylogenetic relationships among the lower level groups of Bivalvia, whenever the dataset was suitable. Chapter 5 will address the ongoing question of the monophyly or polyphyly of Bivavia. Papers from chapters 4 and 5 will be submitted shortly for publication. Finally, Chapter 6 lists cited references in the whole Thesis, Chapter 7 is composed by Appendices, and Chapter 8 includes copies of the papers I published during my PhD, also the ones that are not directly related to the main topic of the present Thesis.

CHAPTER 1

INTRODUCTION

1.1. BIVALVE MOLLUSKS: ZOOLOGY, PHYLOGENY, AND BEYOND

The phylum Mollusca

The outstanding scientific interest for the second richest phylum in the animal kingdom – slightly less than 100,000 species known (Brusca and Brusca, 2003) – and the over time passionate work of collectors and amateur malacologists, led to a stunning abundant literature in the field of mollusk taxonomy and systematics. Georges Cuvier (1769-1832) was the first to establish the group “Mollusca” (in 1795) as something similar to the assemblage we refer to with this name. Since then, barnacles, tunicates, and brachiopods were purged from the phylum: mollusks are now bilaterally symmetrical animals, unsegmented lophotrochozoan protostomes, typically featuring a dorsal visceral mass, a mantle secreting calcareous epidermal spicules, shell plates, or a true shell, a bold muscular foot, and a radula.

Despite the lack of a complete agreement in the general classification of mollusks, the phylum can be arranged in seven or eight classes. Some of them are very poorly known, such as the unconventional grouping of Aplacophora, including Chaetodermomorpha (=Caudofoveata) and Neomeniomorpha (=Solenogastres), and the class of Monoplacophora, thought to be extinct until Lemche’s (1957) discovery of a living species, *Neopilina galatheae*. Also Chitons (class Polyplacophora) and tusk shells (class Scaphopoda) are better known to museum visitors and zoology students, rather than to non specialists. On the contrary, humans were always very familiar with the remaining three classes of mollusks, which were commonly used as popular tools, musical devices,

money, decorations, and – hence the huge economical worth – food: cowries, limpets, snails, slugs (class Gastropoda); cuttlefishes, squids, octopuses, nautiluses (class Cephalopoda); clams, cockles, oysters, quahogs, scallops, mussels (class Bivalvia).

Notwithstanding the importance they have for mankind, our knowledge of mollusks' evolutionary history is still limited. The sister group of mollusks was variably found in Sipunculida (peanut worms; Scheltema, 1993) or Ectoprocta (Haszprunar, 2000), albeit most researchers agreed to a close phylogenetic relationship between mollusks and annelids. Furthermore, molecular tools have been unexpectedly unable for long to obtain the phylum itself as a monophyletic clade. Only recently, Dunn et al. (2008) were able to obtain a solid molluscan clade in their broad phylogenomic analysis of the animal tree of life, based on 150 EST genes. Previous analyses drafted mollusks' monophyly with low statistical support (Giribet et al., 2006), or retrieved the phylum as a polyphyletic assemblage (Winnepenninckx et al., 1996).

A survey of class Bivalvia

The phylum Mollusca is notable for the great disparity of morphological adaptations it features, and bivalves are surely among the most derived classes. Following the mollusk checklist compiled by Victor Millard (2001), bivalve genera, both extant and fossil, sum up to slightly more than 3,400. They are widespread all over the world, both in seas and freshwater environments, showing adaptations to different conditions of enlightment, depth, pressure, zoocenosis, bottom, and idrology; furthermore, they share several peculiar apomorphies, which immediately distinguish them from other mollusks.

Bivalves are typically fossorial or benthonic organisms, though many uncommon features have been selected, from swimming to active predation, rock-boring to infaunal life. Fossil records are abundant, especially from the Mesozoic Era, so that we can easily investigate extinct bivalve biodiversity. The bivalve shell is perhaps the most prominent

feature of the class. Two valves are dorsally hinged: they tend to open because of an elastic ligament, and are kept close by one or two adductor muscles. The head and all related organs (including the brain) were lost: for this reason, bivalves also lack a radula, which is one of the principal diagnostic character for mollusks. Moreover, most bivalves underwent a process of modification of ancestral respiratory organs (the ctenidia), which led to the development of a filter-feeding apparatus (the gills) to convey food particles to the mouth. In many cases, mantle margins are ventrally joined to produce inhalant and exhalant siphons. Generally, the muscular foot is extensible, elongated, and laterally compressed. As a consequence, these differences make the comparison with other mollusks very difficult, as well as the identification of the sister group of bivalves (Scheltema, 1993; von Salvini-Plawen and Steiner, 1996; Haszprunar, 2000). Conversely, given all these apomorphies, the monophyletic status of bivalves as a class was never challenged from a morphological perspective. However, molecular analysis often retrieve bivalves as polyphyletic, especially when broad sampling was done.

Actually, many studies used the nuclear 18S rDNA as a phylogenetic marker, and almost invariantly the class was not supported as a valid clade (Steiner and Müller, 1996; Winnepenninckx et al., 1996; Adamkewicz et al., 1997; Canapa et al., 1999; Giribet and Wheeler, 2002; Passamaneck et al., 2004); this gene was then questioned as a good marker to resolve bivalve phylogeny (but see Giribet and Carranza, 1999; Steiner, 1999; Canapa et al., 2001; Taylor et al., 2007). Actually, it seems true that *18s* gene does not accumulate mutation at a suitable ratio to be useful for a deep phylogeny reconstruction of bivalves; nevertheless, the problem of bivalve polyphyly still persists (Giribet and Distel, 2003; Giribet et al., 2006). On one hand, Steiner and Müller (1996), Adamkewicz et al. (1997), and Canapa et al. (1999) could obtain a monophyletic bivalve clade only under few variable combinations; Passamaneck et al. (2004) could obtain it only for some datasets; the class formed a true clade in the recent work of Doucet-Beaupré et al. (2010), but

taxonomic coverage is very low in their study, whose main focus is not bivalve phylogeny itself. On the other hand, Giribet and Wheeler (2002) showed that a morphological matrix, joining molecular data in a total evidence approach, could overwhelm sequence phylogenetic signal and lead to monophyletic bivalves. Finally, Wilson et al. (2010) obtained a supported clade for the class using a wide array of eight molecular markers and 24 species. In conclusion, a complex interaction between markers' features, outgroup choice, optimality criterion, and taxon sampling must be understood and assessed before accepting or discarding bivalves' polyphyly.

Even if we accept bivalves as a monophyletic taxon, the debate about its sister group is an ongoing issue (Winnepenninckx et al., 1996; Passamaneck et al., 2004; Giribet et al., 2006; Haszprunar, 2008; Wilson et al., 2010). Probably, the most widespread scenario is the "Diasoma hypothesis" (see, f.i., Runnegar and Pojeta, 1974, 1985; Pojeta and Runnegar, 1976, 1985; Götting, 1980a, 1980b; Pojeta, 1980; von Salvini-Plawen, 1990a, 1990b, Steiner, 1992; von Salvini-Plawen and Steiner, 1996; Brusca and Brusca, 2003), which clusters bivalves together with Scaphopoda (tusk shells). Synapomorphies of the Diasoma clade, as listed by Brusca and Brusca (2003), are: head reduction, decentralized nervous system, mantle cavity basically surrounding the entire body, and the spatulate shape of the foot. The Diasoma clade would nest within the broader assemblage of "true shell-bearing mollusks" (i.e., Bivalvia, Cephalopoda, Gastropoda, Monoplacophora, Scaphopoda), the subphylum Conchifera (Götting, 1980a; von Salvini-Plawen, 1990a; Nielsen, 1995; Scheltema, 1993, 1996; von Salvini-Plawen and Steiner, 1996).

However, this view has been rejected by both morphological and molecular studies, which eventually suggested that scaphopods are better related to cephalopods and gastropods (Peel, 1991; Haszprunar, 2000; Giribet and Wheeler, 2002; Wanninger and Haszprunar, 2002; Steiner and Dreyer, 2003; Passamaneck et al., 2004). Recently, first phylogenetic analyses including monoplacophoran specimens also challenged the

traditional Aculifera grouping, by clustering together Monoplacophora and Polyplacophora in the Serialia (Giribet et al., 2006; Wilson et al., 2010), although this hypothesis was somewhat questionable (Steiner in Haszprunar, 2008; Wägele et al., 2009).

The Opponobranchia: true ctenidia for a truly vexed issue

Classical scenarios of Bivalvia phylogeny (Morton, 1996; Cope, 1996) point out that morphological convergence and homoplasy is a major issue in the evolution of the class. First bivalves emerged in the Cambrian period and they were probably shallow water burrowers. Two main evolutionary events led to the huge radiation they underwent in the following periods (Tsubaki et al., 2010). A first adaptive radiation was possible through the gain of byssus, which allowed life on hard substrates, and a “more spectacular second radiation” was triggered by mantle fusion and the emergence of siphons, which enabled dramatic novelties in bivalves’ life habits. Moreover, predation pressure was clearly identified as a major driving force of evolution (Morton, 1996). In fact, marine fossils show a sharp change in community structure along the Secondary era, which was termed “the Mesozoic Marine Revolution” (Vermeij, 1977). Although it is not clear how sudden this revolution actually was (see Hautmann, 2004), many faunal changes took place during this timespan, like the increase of durophagous predators and grazers, and the disappearance or environmental restriction of sessile animals (Vermeij, 1977, 1987, 2008; Walker and Brett, 2002; Harper, 2006). Given this framework, many bivalve evolutionary features, like the increase of shell sturdiness, some degree of infaunalisation, and other defensive mechanisms, can be strictly related to an increase of predation pressure (Stanley, 1977; Morton, 1996; Hautmann, 2004). It is also particularly interesting to link the Mesozoic Marine Revolution to the appearance of more stable ligament shapes (Hautmann, 2004, 2006; Hautmann and Golej, 2004) and the development of efficient burrowing adaptations (see Hautmann et al., 2011).

A main split is generally acknowledged in the bivalve evolutionary tree: on one side, the subclass Protobranchia, with taxodont hinge and respiratory organs (ctenidia) separated from feeding palps; on the other side, all remaining bivalves (Autobranchia), with labial palps intimately fused with gills and without palp proboscides. The two oldest known bivalves from the early Cambrian, *Pojetaia* and *Fordilla*, would represent the oldest known ancestor of both lineages, respectively (Runnegar and Bentley, 1983; Pojeta and Runnegar, 1985). Following Morton (1996), palp proboscides and feeding through palps itself are not a plesiomorphy, but an autapomorphy of Protobranchia, which subsequently radiated into deep waters as deposit feeders. Basal splits of bivalve phylogeny are differently depicted by Cope (1996), who gives more importance to the taxodont hinge and shell composition, than to respiratory system: in his view, a subclass called Palaeotaxodonta, with the only extant order Nuculoida, but comprehending both genera *Pojetaia* and *Fordilla* in the newly erected family of Fordillidae (Runnegar and Pojeta, 1992), was the common ancestor to the order Solemyoida – placed in its own subclass Lipodonta – and to other bivalves, either filibranch or eulamellibranch. Anyway, most authors agree on the difference between nuculoids and solemyoids one side (be they representatives of a single subclass or not), and autobranch bivalves the other side (Purchon, 1987; von Salvini-Plawen and Steiner, 1996; Waller, 1990, 1998; Morton, 1996; Cope, 1996, 1997). With respect to molecular phylogenetics, genera *Nucula* and *Solemya*, which are typically chosen for phylogenetic analyses, clustered in many cases with non-bivalve outgroups, thus rendering the class polyphyletic (Hoeh et al., 1998; Giribet and Wheeler, 2002; Giribet and Distel, 2003). An unexpected outcome of molecular analysis was the position of superfamily Nuculanoidea, traditionally placed among nuculoids. The homogenous shell structure of this group was thought to be derived from a prismato-nacreous shell like that of Nuculoida in post-Jurassic times (Cox, 1959; Cope, 1996); however, Giribet and Wheeler (2002) first placed Nuculanoidea as the sister group of all

Autobranchia. Nuculanoids position was somewhat unstable in the broader phylogenetic analysis of Giribet and Distel (2003), but this placement was again suggested by Bieler and Mikkelsen (2006); finally, genus *Nuculana* was firmly nested among pteriomorphians in the evolutionary tree depicted by Plazzi and Passamonti (2010). Recall that the prismato-nacreous shell is not a unique feature of palaeotaxodont and that taxondont hinge is also present among pteriomorphian families, these findings would at least lead to the paraphyly of Protobranchia *sensu* Morton (1996); therefore, Giribet (2008) proposed the name Opponobranchia for the formerly unrecognized clade Nuculoida + Solemyoida.

The Autobranchia: between tenets and question marks

From a systematic viewpoint, four high-rank monophyletic clades are generally accepted within Autobranchia: Pteriomorphia, Heterodonta, Palaeoheterodonta, and Anomalodesmata.

Mussels, scallops, oysters, arks and their kin belong to the clade Pteriomorphia; these are marine organisms typically featuring a byssus and an asymmetry in the adductor muscles, which gives the classical heteromyarian or even monomyarian shell. Gills are generally filibranch or pseudolamellibranch, with some exceptions. Clams and cockles are just few species belonging to Heterodonta, a broad taxon encompassing the highest biodiversity of the class; heterodonts are usually marine, siphonate, dimyarian, eulamellibranch filter feeders, although many exception are known throughout the group. Newell (1965) defined Palaeoheterodonta as “alike in the possession of free or incompletely fused mantle margins, an opisthodetic parivincular ligament, and prismato-nacreous shells. Posterolateral hinge teeth, where present, originate at the beaks and below the ligament”; few species of *Neotrigonia* and about 175 genera of freshwater mussels belong to this clade (Giribet, 2008). The Anomalodesmata are sometimes given the status of subclass: most are specialized bivalves, either marine or estuarine. Many of

them present the septibranchiate condition of gills, becoming strange deep-water carnivorous bivalves or notable tube dwellers.

The Autobranchia have been generally divided in two lineages, but there is lack of agreement in the basal topology of the clade: it has been described either as (Pteriomorphia + (Heterodonta + Palaeoheterodonta)) or as (Palaeoheterodonta + (Pteriomorphia + Heterodonta)). The taxon Heteroconchia, i.e. the monophyletic clade composed by Palaeoheterodonta and Heterodonta resulting from the former tree, was repeatedly proposed to be the sister group of Pteriomorphia (Waller, 1990, 1998; Giribet and Wheeler, 2002; Bieler and Mikkelsen, 2006; Giribet, 2008), but a growing body of evidence is accumulating towards the latter hypothesis (Cope, 1996, 1997; Canapa et al., 1999; Giribet and Distel, 2003; Doucet-Beaupré et al., 2010; Plazzi and Passamonti, 2010).

The Pteriomorphia were rarely challenged in their subclass status. Most palaeontologists ever accepted it, as pointed out by Newell (1965). Cox (1960) thought the Mytiloidea to have a separate origin stemming from the Modiomorphidae (Palaeoheterodonta), and Pojeta (1978) listed both mytiloids and modiomorphoids in his subclass Isofilibranchia (for a more extensive discussion and bibliography, we refer to Cope, 1996); conversely, Cope (1997) included Arcoida in their own subclass Neotaxodonta. Though first molecular studies evidenced some caveats in the group (Steiner and Müller, 1996; Winnepenninckx et al., 1996; Adamkewicz et al., 1997), recent phylogenetic work, again, almost invariantly confirms it as a monophyletic clade (Canapa et al., 1999; Campbell, 2000; Steiner and Hammer, 2000; Giribet and Wheeler, 2002; Giribet and Distel, 2003; Matsumoto, 2003; Passamaneck et al., 2004; Giribet, 2008; Doucet-Beaupré et al., 2010; Plazzi and Passamonti, 2010). Morphological characters of this subclasses were also thoroughly investigated in recent years, with special regard to ligament structure (Hautmann, 2004; Malchus, 2004).

Internal relationships within Pteriomorphia have yet to be settled; some analyses retrieved Arcoida (arks) as the sister group of remaining pteriomorphians (Cope, 1996, 1997; Giribet and Distel, 2003), whereas others had Mytiloida (mussels) in the basal position (Waller, 1998; Carter et al., 2000; Steiner and Hammer, 2000; Giribet and Wheeler, 2002; Matsumoto, 2003). Interestingly, both Steiner and Hammer (2000) and Distel (2000) found, albeit using the *18s* gene, two main lineages within pteriomorphians: Mytiloidea were the sister group of (Pinnoidea + (Ostreoidea + Pterioidea)), whereas Arcoidea were the sister group of ((Anomioidea + Plicatuloidea) + (Limoidea + Pectinoidea)); therefore, both superfamilies retained in these analyses a relatively basal position. Recalling that many analyses gave somewhat controversial results on pteriomorph branching pattern (Carter, 1990; Campbell, 2000; Steiner and Hammer, 2000; Matsumoto, 2003), Plazzi and Passamonti (2010) considered Pteriomorphia as a wide polytomy, possibly the result of a true, rapid radiation event at the Cambrian/Ordovician boundary.

As defined by Newell (1965), Heterodonta possess “non-nacreous shells [...] and more or less fused, siphonate, mantle margins. Posterolateral teeth, where present, originate some distance behind the beaks and ligament”. All these bivalves are eulamellibranch. The most ancient heterodont was identified in the genus *Babinka* dating to the early Ordovician (Babin, 1982), though Cope (1996) suggested it was rather a paleoheterodont. Pojeta (1978) supposed that the dentition of *Babinka* could proof its direct descent from a *Fordilla*-like bivalve. The extraordinary diversity crowded into the heterodonts was only recently targeted by sound molecular phylogenetics analyses. Though acknowledging the validity of the subclass, pivotal studies (Adamkewicz et al., 1997; Canapa et al., 1999) immediately pointed out the polyphyly of traditional orders Veneroida and Myoida, a suspicion that was to get more and more support in later analyses (Canapa et al., 2001; Giribet and Wheeler, 2002; Dreyer et al., 2003; Taylor et

al., 2007b). Molecular phylogenetics had had a great impact on heterodont systematic. Many studies showed that family Tridacnidae was better considered as a subfamily of family Cardiidae (Maruyama et al., 1998; Schneider and Ó Foighil, 1999); Anomalodesmata were proposed to be included as a monophyletic clade within the subclass (Giribet and Wheeler, 2002; Giribet and Distel, 2003; Taylor et al., 2007b; Dreyer et al., 2003; but see Plazzi and Passamonti, 2010); the basal phylogeny was recently modified and assessed, with special regard to the classical view of superfamily Lucinoidea (Steiner and Hammer, 2000; Giribet and Distel, 2003; Williams et al., 2004; Taylor et al., 2007a; Taylor et al., 2007b). As the subclass is currently conceived, a basal split separates two main lineages: Astartoidea, Carditoidea, and Crassatelloidea belong to the Archiheterodonta, the sister group of all remaining heterodonts – the Euheterodonta, which also include Anomalodesmata (Giribet and Distel, 2003; Taylor et al., 2007b; Giribet, 2008). Archiheterodonta do overlap with the order Carditoida *sensu* Bieler and Mikkelsen (2006) and are consistent with many observation coming from physiology (Terwilliger and Terwilliger, 1985; Taylor et al., 2005), spermiogenesis (Healy, 1995), morphology (Yonge, 1969; Purchon, 1987), molecular biology (Campbell, 2000; Park and Ó Foighil, 2000; Giribet and Wheeler, 2002; Dreyer et al., 2003; Giribet and Distel, 2003; Williams et al., 2004; Taylor et al., 2005; Harper et al., 2006; Taylor and Glover, 2006), and fossils (Carter, Campbell, and Campbell, 2006, in Giribet, 2008). The basal position of Euheterodonta is occupied by the newly-erected superfamily Thyasiroidea (Taylor et al., 2007a). Following Taylor et al. (2007b), a monophyletic clade they called Neoheterodontei clusters together most derived forms, like, among others, Pholadoidea, Myoidea, Ungulinoidea, Mactroidea, and Veneroidea; the sister group of Neoheterodontei is a clade composed by (Cardioidea + Tellinoidea).

Doubly Uniparental Inheritance

The class Bivalvia is very peculiar also because some species exhibit a unique form of mitochondrial inheritance, a feature which is unevenly scattered throughout the group. This interesting exception to the common strictly maternal descent of mitochondria is called Doubly Uniparental Inheritance (DUI; Skibinski et al., 1994a, 1994b; Zouros et al., 1994a, 1994b), as it involves two separate mitochondrial lineages, which are both uniparentally transmitted. One is called F, as it passes through mothers to the complete offspring; the other is called M, as it passes through fathers to male sons only. Therefore, female offspring tends to be omoplasmic for the F mitotype. Conversely, male offspring tends to be heteroplasmic: the M mitotype concentrate in the gonads, whereas the F one is present in the soma (Breton et al., 2007; Passamonti and Ghiselli, 2009; and reference therein). This mechanism has been found in different families of bivalves, with many variations on the general conserved scheme (Theologidis et al., 2008; Doucet-Beaupré et al., 2010; Ghiselli et al., 2011); its implications for gene orthology and evolutionary reconstruction have to be adequately assessed before starting a mitochondrial phylogeny of the class Bivalvia.

The choice of the "right" molecular marker in bivalve phylogenetics

Due to the large and still-increasing number of molecular works on the topic, several genetic markers have been employed, obtaining various degrees of affordability. Much has been written on the *18s* rDNA as a suitable phylogenetic marker for the class: it seems that it does not provide a good signal for phylogenetic inference (Steiner and Müller, 1996; Distel, 2000; Matsumoto and Hayami, 2000; Passamaneck et al., 2004), being suggested for lower taxonomic levels (Winnepenninckx et al., 1996; Adamkewicz et al., 1997; but see Giribet and Carranza, 1999; Canapa et al., 1999, 2001). The large nuclear ribosomal subunit (*28s*) was also used for phylogenetic inference and somewhat similar problems

were found (Littlewood, 1994; Ó Foighil and Taylor, 2000; Park and Ó Foighil, 2000; Giribet and Wheeler, 2002; Giribet and Distel, 2003; Kirkendale et al., 2004; Passamaneck et al., 2004; Williams et al., 2004; Taylor et al., 2007a, 2007b; Albano et al., 2009; Taylor et al., 2009; Lorion et al., 2010; Tëmkin, 2010; Tsubaki et al., 2010). Other nuclear markers were employed, such as the *5s* rDNA (López-Piñon et al., 2008), satellite DNA (Martínez-Lage et al., 2002; López-Flores et al., 2004), the histone 3 (Giribet and Distel, 2003; Kappner and Bieler, 2006; Puslednik and Serb, 2008; Tëmkin, 2010), ITS-1 (Insua et al., 2003; Lee and Ó Foighil, 2003; Shilts et al., 2007; Wang et al., 2007; Wood et al., 2007), or ITS-2 (Insua et al., 2003; Olu-Le Roy et al., 2007; Wood et al., 2007), but little was concluded on the use of these markers in phylogenetic inference. Moreover, some authors proposed other kind of approaches. For example, inasmuch bivalves exhibit an uncommon variability in the gene order on the mitochondrial genome, Serb and Lydeard (2003) showed the usefulness of mitochondrial gene order data in shaping the evolutionary tree of the class. Wang and Guo (2004) used karyotypic and chromosomal data to get data for bivalve evolution. Doucet-Beaupré et al. (2010) were the first to attempt a molecular phylogeny of bivalves using the complete mitochondrial genome sequence, although in a DUI framework (see above).

Mitochondrial sequences were the most analyzed markers. This allowed a critical assessment of their usefulness in evolutionary studies of bivalves, ranging from the possibility of sequencing single genes, to whole organellar genomes, which allows more resolution. Moreover, there is a relative certainty of avoiding paralogous sequences, as no bivalve nuclear mitochondrial pseudogenes (NUMTs) were reported to date (Bensasson, 2001; Zbawicka et al., 2007). A large number of phylogenies, therefore, are based on mitochondrial DNA: the most utilized molecular markers are the small (Barucca et al., 2004; Puslednik and Serb, 2008; Plazzi and Passamonti, 2010) and large ribosomal subunits (Canapa et al., 1996, 2000; Lydeard et al., 1996; Jozefowicz and Ó Foighil, 1998;

Schneider and Ó Foighil, 1999; Roe et al., 2001; Kirkendale et al., 2004; Therriault et al., 2004; Kappner and Bieler, 2006; Shilts et al., 2007; Puslednik and Serb, 2008; Theologidis et al., 2008; Plazzi and Passamonti, 2010; Tëmkin, 2010), cytochrome b (Theologidis et al., 2008; Plazzi and Passamonti, 2010), cytochrome oxydase I (Peek et al., 1997; Hoeh et al., 1998; Matsumoto and Hayami, 2000; Giribet et al., 2002; Giribet and Distel, 2003; Matsumoto, 2003; Kirkendale et al., 2004; Therriault et al., 2004; Kappner and Bieler, 2006; Olu-Le Roy et al., 2007; Samadi et al., 2007; Shilts et al., 2007; Wood et al., 2007; Albano et al., 2009; Lorion et al., 2010; Plazzi and Passamonti, 2010) and/or cytochrome oxydase III (Ó Foighil and Smith, 1995; Nikula et al., 2007). Moreover, recent works pointed out the importance of adding phylogenetic signals from more than one single gene (Giribet and Wheeler, 2002; Giribet and Distel, 2003; Lee and Ó Foighil, 2003; Barucca et al., 2004; Passamaneck et al., 2004; Therriault et al., 2004; Williams et al., 2004; Kappner and Bieler, 2006; Shilts et al., 2007; Taylor et al., 2007b; Wood et al., 2007; Puslednik and Serb, 2008; Plazzi and Passamonti, 2010). Many polytomies inferred from one-gene phylogenies were therefore resolved, and support values of nodes became higher; indeed, it has been pointed out that the more independent gene sequences are studied, the better the phylogeny results, while the affordability of the evolutionary tree does not necessary improve by simply increasing species number (Steiner and Müller, 1996; Winnepenninckx et al., 1996; Kappner and Bieler, 2006; Shilts et al., 2007; Wood et al., 2007; but see Adamkewicz et al., 1997; Goldman, 1998; Canapa et al., 1999; Giribet and Carranza, 1999; Giribet and Wheeler, 2002). Furthermore, Passamaneck and colleagues (2004) focused also on the interest in using protein coding data set for bivalves phylogeny.

More recently, several attempts to join information from morphology and molecules were done (Ó Foighil and Taylor, 2000; Giribet and Wheeler, 2002; Giribet and Distel, 2003; Harper et al., 2006; Mikkelsen et al., 2006; but see Graham Oliver and Järnegren,

2004). Following Giribet and Distel (2003), because morphology resolved deeper nodes better than molecules, whereas sequence data is more adequate for recent splits.

1.2. MOLECULAR EVOLUTION MODELS, MULTIGENE BAYESIAN ANALYSIS, AND PARTITION CHOICE

Maximum likelihood (ML) is a commonly used phylogenetic tool for DNA sequence data analysis. ML methods incorporates models of DNA sequence evolution better than maximum parsimony, so that they are less prone to errors due to the complexities of this process (Huelsenbeck and Crandall, 1997, and reference therein). ML methods also outperform distance methods and parsimony under several simulated conditions (Hillis et al., 1992; Huelsenbeck 1995a, 1995b; Swofford et al., 2001). Not only ML approach has been developed as an improved phylogenetic analysis, but more complex and realistic models of DNA sequence evolution have been studied as well. These allow different rates of nucleotide base substitution (Kimura, 1980), base composition (Felsenstein, 1981), and site rate heterogeneity (Yang, 1993, 1994). Classically, these are time-reversible models with four states (A, C, G, and T or U) and 12 substitutions. The most parameters-rich time-reversible model is termed GTR and was first described by Tavaré (1986), whereas JC (Jukes and Cantor, 1969) is the simplest. This is well shown by their respective rate matrix

$$Q_{GTR} = \{q_{ij}\} = \begin{pmatrix} - & r_{AC}\pi_C & r_{AG}\pi_G & r_{AT}\pi_T \\ r_{AC}\pi_A & - & r_{CG}\pi_G & r_{CT}\pi_T \\ r_{AG}\pi_A & r_{CG}\pi_C & - & r_{GT}\pi_T \\ r_{AT}\pi_A & r_{CT}\pi_C & r_{GT}\pi_G & - \end{pmatrix} \mu$$

and

$$Q_{JC} = \{q_{ij}\} = \begin{pmatrix} - & r\pi & r\pi & r\pi \\ r\pi & - & r\pi & r\pi \\ r\pi & r\pi & - & r\pi \\ r\pi & r\pi & r\pi & - \end{pmatrix} \mu$$

where r_{ij} is the $i \leftrightarrow j$ substitution rate, π_i is the frequency of the ith nucleotide and μ is the mutation rate. It is clear that JC is a special case of GTR, by constraining

$$r_{AC} = r_{AG} = r_{AT} = r_{CG} = r_{CT} = r_{GT} = r$$

and

$$\pi_A = \pi_C = \pi_G = \pi_T = \pi$$

Thus, we say they are “nested” models.

Many nucleotide substitution models have been described so far (e.g., Kimura, 1980; Felsenstein, 1981; Tamura and Nei, 1993; Posada, 2003), but many more have not yet been described. Huelsenbeck et al. (2004) described a method to determine the number of possible substitution models, based on Bell numbers (Bell, 1934). With respect only to substitution rates r_{ij}, there are 203 possible models; considering all standard parameters, the total number of models increases to 12,180.

There are also molecular evolution models which can also take into account sequence gaps (McGuire et al., 2001), secondary structure (Muse, 1995; Tillier and Collins, 1995), and codons (Goldman and Yang, 1994; Muse and Gaut, 1994).

Thus, it is nowadays possible to use well-improved, complex, and realistic evolutionary models. Despite this fact, no model can be considered “true” in a literary sense (Posada and Buckley, 2004, and reference therein). This can be especially said for data sets with multiple genes analyses and/or gene regions experiencing different selective pressures (e.g., codon positions, introns and exons). Nevertheless, standard ML analyses use a single nucleotide substitution model and associated parameter along the entire data set. This represents a compromise among the various existing partitions (hereafter defined as any homogeneous subset of the whole data set) and may be inadequate to describe the complete evolutionary history of the analyzed DNA regions. A systematic error is introduced due to this “compromised model” and the phylogenetic analysis can give wrong results (Leaché and Reeder, 2002; Reeder, 2003; Wilgenbusch and de Queiroz, 2000; Brandley, et al., 2005). Following Swofford, et al. (1996), systematic error is defined in a statistical framework as an error in a parameter’s estimate due to incorrect or violated assumptions in the method of estimation itself. This differs from random error, which is stochastic error in a parameter estimation due to a limited sample

size. It is particularly troublesome in that it may be reflected either in strong, albeit erroneous, relationships, or in decreased support for legitimate ones (Swofford et al., 1996).

In other words, the disposability of powerful models of evolution it is not necessarily a warranty of affordable results in phylogenetic studies: the most realistic one has to be individuated for each particular case and it has to account for the variability in the entire dataset. It is well known that mismodeling (the wrong choice of the model to be applied) can results in erroneous findings and that phylogenetic elaborations are especially sensible to model selection (e.g., Goldman, 1993; Sullivan et al., 1995; Posada and Crandall, 2001; Yang and Rannala, 2005). Actually, mismodeling can sometimes result in a false topology reconstruction, but it has been shown that topology is “relatively insensitive” (Alfaro and Huelsenbeck, 2006) to the choice of a model of molecular evolution (Posada and Buckley, 2004; Sullivan and Swofford, 2001); other parameters are much more sensible to mismodeling, like branch lengths (Minin et al., 2003), substitution rates (Wakeley, 1994) and, above all, bootstrap values and posterior probabilities (Alfaro and Huelsenbeck, 2006). This is very troublesome, in that one cannot know how affordable a result is.

Common examples of mismodeling involve “compromised model”. One model may be invoked to explain the evolution of a dataset with two or more partitions, best described by two or more separate and different models. A second case of mismodeling happens when multiple partitions are actually explained by the same underlying general model, but differ substantially in the specific model parameter estimates like nucleotides frequencies (e.g., Reeder, 2003). For example, Reeder (2003) found the relative rate of C ↔ T transitions was 27.2 for structural RNAs, but only 4.0 for the ND4 protein-coding gene, a sevenfold difference. The estimate of the same parameter for the combined mtDNA data was 14.7; half the best estimate for structural RNA, and over three times the estimate for

ND4. Whereas the separate data analyses used specific and seemingly appropriate models for the two individual data partitions (i.e., structural RNAs and ND4 protein-coding), the combined (single-model) mtDNA analysis did not accommodate all that was known about the partitions (i.e., specific parameter estimates). The solution to these problems would be to apply adequate models and specific parameter estimates to each single partition in the data set and subsequently merge these all into a single ML analysis. Unfortunately, this is computationally very hard and few examples are known from literature (but see Yang, 1996).

A more feasible solution involves testing for data incongruence or partitioning nodes' support. In other words, we can obtain in this case not a sum of information from separate partitions, but indications on how each partition influences and determines a topology or a node in the global tree. Three data incongruence tests are known from literature: incongruence length difference (ILD), partition homogeneity (PHT) and Templeton tests (Farris et al., 1995a, 1995b; Larson, 1994; Templeton, 1983), and some phylogenetic software packages regularly implement them, like PAUP* (Swofford, 1999). As noted by Wiens (1998) and Lambkin et al. (2002), these methods measure overall levels of agreement between the partitions in the data set; they cannot show which parts of a tree are in conflict among partitions. A partitioned Bremer support (PBS) has been introduced by Baker and DeSalle (1997) to measure the agreement of various partitions about a single node. PBS is based upon Bremer support (Bremer, 1988, 1994; Kallersjö et al., 1992) The Bremer support is very intuitive in a parsimony framework: the most parsimonious tree is found, and then a search is conducted for the most parsimonious tree lacking a particular node. The Bremer support for that particular node is given by

$$BS_i = L_i^{constrained} - L^{unconstrained}$$

where BS_i is the Bremer support for the *i*th node and *L* is the length (measured in number of steps) of the most parsimonious tree, either unconstrained or constrained to

lack the *i*th node. It is possible to compute the length of these tree based on a single partition, again either constrained or unconstrained. The PBS for that particular partition and for that particular node is given by the difference of the two. A positive PBS shows the partitions is in agreement with the node, and a negative one that it is in disagreement. The sum of PBSs for all partitions equals the BS of that node (if partitions globally comprehend the entire data set and are mutually exclusive). Although less intuitive, the same procedure can be applied to likelihood analyses (Lee and Hugall, 2003). Again, a positive partitioned likelihood support (PLS) indicates that a partition support a clade, and a negative PLS indicates that the partition contradicts the clade. Parametric bootstrapping (Huelsenbeck et al., 1996a; Huelsenbeck et al., 1996b) can be used to assess the significance of PLS, and some statistical tests are useful to the this aim (Lee and Hugall, 2003; and reference therein). However, PLS analyses are currently difficult because no widely available phylogenetic software allows such an algorithm; some approximation are needed, and a manual procedure for PLS computation has been provided by Lee and Hugall (2003). An interesting way to take into account separate partitions in a maximum-likelihood analysis is provided by Yang (1996).

Nevertheless, it is possible also to conduct a true partitioned analysis, as methods using Bayesian/Markov chain Monte Carlo (MCMC) algorithms have recently become available (MrBayes 3.1.2; Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). Bayesian tecniques generate posterior probability (PP) distributions using a likelihood function. Several models of molecular evolution can be implemented. Bayesian analyses using uniform priors should yield similar results as ML, and generally do (Huelsenbeck et al., 2002; Larget and Simon, 1999; Leaché and Reeder, 2002). Such an approach is extremely versatile, due either to the merits of the software and to the features of the method itself. PP distributions are based upon user-specified priors, that can be modeled according to several known probability distribution. Regarding partitioned

analysis, it is possible to specify priors also about single subsets of data; specific models can be applied to single partition, and the results take into account information coming from all separate partitions. The use of partition-specific modeling reduces systematic error, providing more reliable likelihood scores and more accurate PP estimates.

My study addresses these issues through partition-specific modeling in a combined analysis frame (see also Nylander et al., 2004; Brandley et al., 2005). We use partitioned Bayesian analysis to demonstrate the effect and the importance of partition choice on phylogeny reconstruction. We apply several methods to select the best partitioning strategy (Brandley et al., 2005; Shull et al., 2005; Strugnell et al., 2005; Wood et al., 2007). This is crucial because it actually provides an objective criterion for selecting the best way of partitioning data, from the traditional global analyses, through several possibilities of data subdividing, to partitioning by every character, which corresponds to the parsimony model (Tuffley and Steel, 1997). The higher the number of partitions, the smaller the amount of data contained in a single partition, thus widening the random stochastic error in model parameters estimates. Furthermore, more partitions means more parameters: this lead to more degrees of freedom. The more degrees of freedom, the bigger the variance in the results. The Bayes Factor (Kass and Raftery, 1995) is a method to overcome this issue and to evaluate a trade-off: on one side we should increase data partitioning to precisely model our data, on the other one we should avoid unjustified overparametrization and sample reduction.

CHAPTER 2

TOWARDS A MOLECULAR PHYLOGENY OF MOLLUSKS: BIVALVES' EARLY EVOLUTION AS REVEALED BY MITOCHONDRIAL GENES.

2.1. INTRODUCTION

Bivalves are among the most common organisms in marine and freshwater environments, summing up to about 8,000 species (Morton, 1996). They are characterized by a bivalve shell, filtrating gills called ctenidia, and no differentiated head and radula. Most bivalves are filter-feeders and burrowers or rock-borers, but swimming or even active predation are also found (Dreyer et al., 2003). Most commonly, they breed by releasing gametes into the water column, but some exceptions are known, including brooding (Ó Foighil and Taylor, 2000). Free-swimming planktonic larvae (veligers), contributing to species dispersion, are typically found, which eventually metamorphose to benthonic sub-adults.

Bivalve taxonomy and phylogeny are long-debated issues, and a complete agreement has not been reached yet, even if this class is well known and huge fossil records are available. In fact, bivalves' considerable morphological dataset has neither led to a stable phylogeny, nor to a truly widely accepted higher-level taxonomy. As soon as they became available, molecular data gave significant contributions to bivalve taxonomy and phylogenetics, but little consensus has been reached in literature because of a substantial lack of shared methodological approaches to retrieve and analyze bivalves' molecular data. Moreover, to improve bivalves' phylogenetics, several attempts to join morphology and molecules have also been proposed (Giribet and Wheeler, 2002; Giribet and Distel, 2003; Harper et al., 2006; Mikkelsen et al., 2006; Olu-Le Roy et al., 2007),

since, according to Giribet and Distel (2003), morphology resolves deeper nodes better than molecules, whereas sequence data are more adequate for recent splits.

Bivalves are generally divided into five extant subclasses, which were mainly established on body and shell morphology, namely Protobranchia, Palaeoheterodonta, Pteriomorphia, Heterodonta and Anomalodesmata (Millard, 2001; but see e. g. Vokes, 1980, for a slightly different taxonomy). In more detail, there is a general agreement that Protobranchia is the first emerging lineage of Bivalvia. All feasible relationships among Protobranchia superfamilies (Solemyoidea, Nuculoidea and Nuculanoidea) have been proposed on morphological approaches (Purchon, 1987b; Waller, 1990; Morton, 1996; Salvini-Plawen and Steiner, 1996; Cope, 1997; Waller, 1998), albeit some recent molecular findings eventually led to reject the monophyly of the whole subclass: while Solemyoidea and Nuculoidea do maintain their basal position, thus representing Protobranchia *sensu stricto*, Nuculanoidea is better considered closer to Pteriomorphia, placed in its own order Nuculanoida (Giribet and Wheeler, 2002; Giribet and Distel, 2003; Kappner and Bieler, 2006).

The second subclass, Palaeoheterodonta (freshwater mussels), has been considered either among the most basal (Cope, 1996) or the most derived groups (Morton, 1996). Recent molecular analyses confirm its monophyly (Giribet and Wheeler, 2002) and tend to support it as basal to other Autolamellibranchiata bivalves (Graf and Ó Foighil, 2000; Giribet and Distel, 2003).

Mussels, scallops, oysters and arks are representatives of the species-rich subclass Pteriomorphia. In literature, this subclass has been resolved as a clade within all Eulamellibranchiata (Purchon, 1987b), as a sister group of Trigonioidea (Salvini-Plawen and Steiner, 1996), of Heterodonta (Cope, 1997), of (Heterodonta+Palaeoheterodonta) (Waller, 1990, 1998), or as a paraphyletic group to Palaeoheterodonta (Morton, 1996). Moreover, some authors hypothesize its polyphyly (Carter, 1990; Starobogatov, 1992),

while others claimed that a general agreement on Pteriomorphia monophyly is emerging from molecular studies (Giribet and Distel, 2003). Such an evident lack of agreement appears to be largely due to an ancient polytomy often recovered for this group, especially in molecular analyses, which is probably the result of a rapid radiation event in its early evolution (Campbell, 2000; Steiner and Hammer, 2000; Matsumoto, 2003).

Heterodonta is the widest and most biodiversity-rich subclass, including some economically important bivalves (f.i., venerid clams). This subclass has been proposed as monophyletic (Purchon, 1987b; Carter, 1990; Starobogatov, 1992; Cope, 1996, 1997; Waller, 1990, 1998), or paraphyletic (Morton, 1996; Salvini-Plawen and Steiner, 1996), but it seems there is a growing agreement on its monophyly. At a lower taxonomic level, doubts on the taxonomic validity of its major orders, such as Myoida and Veneroida, are fully legitimate, and, in many cases, recent molecular analyses led to throughout taxonomic revisions (Maruyama et al., 1998; Williams et al., 2004; Taylor et al., 2007).

Little agreement has been reached in literature on Anomalodesmata: this subclass shows a highly derived body plan, as they are septibranchiate and some of them are also carnivore, features that possibly evolved many times (Dreyer et al., 2003). Anomalodesmata were considered as sister group of Myoida (Morton, 1996; Salvini-Plawen and Steiner, 1996), Mytiloidea (Carter, 1990), Palaeoheterodonta (Cope, 1997), or Heterodonta (Waller, 1990, 1998); alternatively, Purchon (1987b) states that they represent a monophyletic clade nested in a wide polytomy of all Bivalvia. Anomalodesmata were also considered as basal to all Autolamellibranchiata (e. g., Starobogatov, 1992). Whereas the monophyletic status of Anomalodesmata seems unquestionable on molecular data (Dreyer et al., 2003), some authors proposed that this clade should be nested within heterodonts (Giribet and Wheeler, 2002; Giribet and Distel, 2003; Bieler and Mikkelsen, 2006; Harper et al., 2006).

Molecular analyses gave clearer results at lower taxonomic levels, so that this kind of literature is more abundant: for instance, key papers have been published on Ostreidae (Littlewood, 1994; Jozefowicz and Ó Foighil, 1998; Ó Foighil and Taylor, 2000; Kirkendale et al., 2004; Shilts et al., 2007), Pectinidae (Puslednik and Serb, 2008), Cardiidae (Maruyama et al., 1998; Schneider and Ó Foighil, 1999) or former Lucinoidea group (Williams et al., 2004; Taylor et al., 2007).

In this study, we especially address bivalves' ancient phylogenetic events by using mitochondrial molecular markers, namely the *12s*, *16s*, cytochrome b (*cytb*) and cytochrome oxidase subunit 1 (*cox1*) genes. We choose mitochondrial markers since they have the great advantage to avoid problems related to multiple-copy nuclear genes (i.e. concerted evolution, Plohl et al., 2008), they have been proved to be useful at various phylogenetic levels, and, although this is not always true for bivalves, they largely experience Strict Maternal Inheritance (SMI; Gillham, 1994; Birky, 2001).

Actually, some bivalve species show an unusual mtDNA inheritance known as Doubly Uniparental Inheritance (DUI; see Breton et al., 2007; Passamonti and Ghiselli, 2009; for reviews): DUI species do have two mitochondrial DNAs, one called F as it is transmitted through eggs, the other called M, transmitted through sperm and found almost only in males' gonads. The F mtDNA is passed from mothers to complete offspring, whereas the M mtDNA is passed from fathers to sons only. Obviously, DUI sex-linked mtDNAs may result in incorrect clustering, so their possible presence must be properly taken into account. DUI has a scattered occurrence among bivalves and, until today, it has been found in species from seven families of three subclasses: palaeoheterodonts (Unionidae, Hyriidae, and Margaritiferidae), pteriomorphians (Mytilidae), and heterodonts (Donacidae, Solenidae, and Veneridae) (Theologidis et al., 2008; Fig. 2 and reference therein). In some cases, co-specific F and M mtDNAs do cluster together, and this will not significantly affect phylogeny at the level of this study: this happens, among others, for

Donax trunculus (Theologidis et al., 2008) and *Venerupis philippinarum* (Passamonti et al., 2003). In others cases, however, F and M mtDNAs cluster separately, and this might possibly result in an incorrect topology: f.i. this happens for the family of Unionidae and for *Mytilus* (Theologidis et al., 2008). All that considered, bivalves' mtDNA sequences should not be compared unless they are surely homolog, and the possible presence of two organelle genomes is an issue to be carefully evaluated (see Materials and Methods – Specimens' Collection and DNA Extraction, for further details). On the other hand, we still decided to avoid nuclear markers for two main reasons: *i*) largely used nuclear genes, like 18S rDNA, are not single-copy genes and have been seriously questioned for inferences about bivalve evolution (Littlewood, 1994; Steiner and Müller, 1996; Winnepenninckx et al., 1996; Adamkewicz et al., 1997; Steiner, 1999; Distel, 2000; Passamaneck et al., 2004); *ii*) data on single-copy nuclear markers, like β-actin or *hsp70*, lack for the class, essentially because primers often fail to amplify target sequences in Bivalvia (pers. obs.).

2.2. Materials and Methods

Specimens' collection and DNA extraction

Species name and sampling locality are given in Table 2.1. Animals were either frozen or ethanol-preserved until extraction. Total genomic DNA was extracted by DNeasy® Blood & Tissue Kit (Qiagen, Valencia, CA, USA), following manufacturer's instructions. Samples were incubated overnight at 56°C to improve tissues' lysis. Total genomic DNA was stored at -20°C in 200 µL AE Buffer, provided with the kit. DUI species are still being discovered among bivalves; nevertheless, as mentioned, a phylogenetic analysis needs comparisons between orthologous sequences, and M- or F-type genes under DUI are not. On the other hand, F-type mtDNA for DUI species and mtDNA of non-DUI species are orthologous sequences. As M-type is present mainly in sperm, we avoided sexually-mature individuals and, when possible (i.e., when the specimen was not too tiny), we did not extract DNA from gonads. If possible, DNA was obtained from foot muscle, which, among somatic tissues, carries very little M-type mtDNA in DUI species (Garrido-Ramos et al., 1998), thus reducing the possibility of spurious amplifications of the M genome. Moreover, when downloading sequences from GenBank, we paid attention in retrieving female specimens data only, whenever this information was available.

Table 2.1. Specimens used for this study, with sampling locality and taxonomy following Millard (2001). Only species whose sequences were obtained in our laboratory are shown.

Subclass	Order	Suborder	Superfamily	Family	Subfamily	Species	Provenience
Anomalodesmata	Pholadomyoida	Cuspidariina		Cuspidariidae		*Cuspidaria rostrata*	Malta
		Pholadomyina	Pandoroidea	Pandoridae		*Pandora pinna*	Trieste, Italy
				Thraciidae		*Thracia distorta*	Secche di Tor Paterno, Italy
Heterodonta	Chamida		Astartoidea	Astartidae	Astartinae	*Astarte* cfr. *castanea*	Woods Hole, MA, USA
			Mactroidea	Mactridae	Mactrinae	*Mactra corallina*	Cesenatico, Italy
						Mactra lignaria	Cesenatico, Italy
			Tellinoidea	Pharidae	Cultellinae	*Ensis directus*	Woods Hole, MA, USA
			Tridacnoidea	Tridacnidae		*Tridacna derasa*	commercially purchased
						Tridacna squamosa	commercially purchased
	Myida	Myina	Myoidea	Myidae	Myinae	*Mya arenaria*	Woods Hole, MA, USA
	Veneroida		Carditoidea	Carditidae	Carditinae	*Cardita variegata*	Nosi Bè, Madagascar
			Veneroidea	Veneridae	Gafrarinae	*Gafrarium alfredense*	Nosi Bè, Madagascar
					Gemminae	*Gemma gemma*	Woods Hole, MA, USA
Palaeheterodonta	Unionida		Unionoidea	Unionidae	Anodontinae	*Anodonta woodiana*	Po River delta, Italy
Protobranchia	Nuculoida		Nuculanoidea	Nuculanidae	Nuculaninae	*Nuculana commutata*	Malta
			Nuculoidea	Nuculidae		*Nucula nucleus*	Goro, Italy
Pteriomorphia	Arcida	Arcina	Arcoidea	Arcidae	Anadarinae	*Anadara ovalis*	Woods Hole, MA, USA
					Arcinae	*Barbatia parva*	Nosi Bè, Madagascar
						Barbatia reeveana	Galápagos Islands, Ecuador
						Barbatia cfr. *setigera*	Nosi Bè, Madagascar
	Limida		Limoidea	Limidae		*Lima pacifica galapagensis*	Galápagos Islands, Ecuador
	Ostreoida	Ostreina	Ostreoidea	Ostreidae	Pycnodonteinae	*Hyotissa hyotis*	Nosi Bè, Madagascar
		Pectinina	Anomioidea	Anomiidae		*Anomia sp.*	Woods Hole, MA, USA
			Pectinoidea	Pectinidae	Chlamydinae	*Argopecten irradians*	Woods Hole, MA, USA
						Chlamys livida	Nosi Bè, Madagascar
						Chlamys multistriata	Krk, Croatia
					Pectininae	*Pecten jacobaeus*	Montecristo Island, Italy
	Pteriida	Pinnina	Pinnoidea	Pinnidae		*Pinna muricata*	Nosi Bè, Madagascar

PCR amplification, cloning, and sequencing

PCR amplifications were carried out in a 50 µL volume, as follows: 5 or 10 µL reaction buffer, 150 nmol $MgCl_2$, 10 nmol each dNTP, 25 pmol each primer, 1-5 µL genomic DNA, 1.25 units of DNA Polymerase (Invitrogen, Carlsbad, CA, USA or ProMega, Madison, WI, USA), water up to 50 µL. PCR conditions and cycles are listed in Appendix 2.1; primers used for this study are listed in Appendix 2.2. PCR results were visualized onto a 1-2% electrophoresis agarose gel stained with ethidium bromide. PCR products were purified through Wizard® SV Gel and PCR Clean-Up System (ProMega, Madison, WI, USA), following manufacturer's instructions.

Sometimes, amplicons were not suitable for direct sequencing; thus, PCR products were inserted into a pGEM®-T Easy Vector (ProMega, Madison, WI, USA) and transformed into Max Efficiency® DH5α^{TM} Competent Cells (Invitrogen, Carlsbad, CA, USA). Positive clones were PCR-screened with M13 primers (see Appendix 2.2) and visualized onto a 1-2% electrophoresis agarose gel. However, as far as possible, we only cloned whenever it was strictly necessary; actually, as in DUI species some "leakage" of M mitotype may occur in somatic tissues of males, sensible cloning procedures could sometimes amplify such rare variants. Suitable amplicons and amplified clones were sequenced through either GeneLab (ENEA-Casaccia, Rome, Italy) or Macrogen (World Meridian Center, Seoul, South Korea) facilities.

Sequence alignment

Electropherograms were visualized by Sequence Navigator (Parker, 1997) and MEGA4 (Tamura et al., 2007) softwares. Sequences were compared to those available in GenBank through BLAST 2.2.19+ search tool (Altschul et al., 1997). Four outgroups were used for this study: the polyplacophoran *Katharina tunicata*, the scaphopod *Graptacme*

eborea and two gastropods, *Haliotis rubra* and *Thais clavigera*. Appendix 2.3 lists all DNA sequences used for this study, along with their GenBank accession number.

Alignments were edited by MEGA4 and a concatenated data set was produced; whenever only three sequences out of four were known, the fourth was coded as a stretch of missing data, since the presence of missing data does not lead to an incorrect phylogeny by itself, given a correct phylogenetic approach (as long as sufficient data are available for the analysis; see Hartmann and Vision, 2008; and reference therein). In other cases, there were not sufficient published sequences for a given species to be included in our concatenated alignment; nevertheless, we could add the genus itself by concatenating DNA sequences from different co-generic species, as this approach was already taken in other phylogenetic studies (see, f.i., Li et al., 2009). This was the case for *Donax*, *Solemya*, *Spisula*, and *Spondylus* (see Appendix 2.3 for details). Given the broad range of the analysis, which targets whole class phylogeny above the genus level, we do not think that such an approximation significantly biased our results. In any case, phylogenetic positions of such genera were taken with extreme care.

Sequences were aligned with ClustalW (Thompson et al., 1994) implemented in MEGA4. Gap opening and extension costs were set to 50/10 and 20/4 for protein- and ribosomal-coding genes, respectively. Because of the high evolutionary distance of the analyzed taxa, sequences showed high variability, and the problem was especially evident for ribosomal genes, where different selective pressures are active on different regions. These genes showed a lot of indels, which were strikingly unstable across alignment parameters; thus, we could not resolve alignment ambiguities in an objective way. The method proposed by Lutzoni et al. (2000), though very appealing, is problematic for big data sets with high variability, as shown by the authors themselves. On the other side, likelihood analyses are also problematic with the fixed character state method proposed by Wheeler (1999). Elision, as introduced by Wheeler et al. (1995), is a possibility that does

not involve particular methods of phylogenetic analyses, but only a “grand alignment”. However, variability in our ribosomal data set was so high that alignments with different parameters were almost completely different; thus, elision generated only more phylogenetic noise, whereas the original method by Gatesy et al. (1993) was not conceivable because alignment-invariant positions were less than twenty. All that considered, we preferred to use a user-assisted standard alignment-method (i.e., ClustalW) since we think this is yet the best alignment strategy for such a complex dataset. Alignment was also visually inspected searching for misaligned sites and ambiguities, and where manual optimization was not possible, alignment-ambiguous regions were excluded from the analysis. Indels were treated as a whole and converted to presence/absence data to avoid many theoretical concerns on alignments (simple indel coding; see Simmons and Ochoterena, 2000, for more details). In fact, ambiguities in alignments are mainly due to indel insertions; therefore, this technique also eliminates a large part of phylogenetic noise. We then coded indels following the rules given by Simmons and Ochoterena (2000), as implemented by the software GapCoder (Young and Healy, 2003), which considers each indel as a whole, and codes it at the end of the nucleotide matrix as presence/absence (i. e. 1/0). Possibly, a longer indel may completely overlap another across two sequences; in such cases, it is impossible to decide whether the shorter indel is present or not in the sequence presenting the longer one. Therefore, the shorter indel is coded among missing data in that sequence. Data set was then analyzed treating gaps as missing data and presence/absence data of indel events as normal binary data.

Phylogenetic analyses

A preliminary test was made on saturation: transitions and transversions uncorrected p-distances were plotted on global pairwise p-distances, as computed with PAUP* 4.0b10 (pairwise deletion of gaps; Swofford, 1999); the test was repeated on third positions only

for protein-coding genes. Linear regression and its significance were tested with PaSt 1.90 (Hammer et al., 2001).

Partitioning schemes used in this study are 10, based on 26 different partitions (Appendix 2.4), although they are not all the conceivable ones; we describe our 10 partitioning patterns in Table 2.2.

Table 2.2. Partitioning schemes. See Appendix 2.4 for details on partitions.

Partitioning scheme	Number of partitions	Partitions (see Appendix. 2.4)
t01	2	*all, all_indel*
t02[a]	4	*rib, rib_indel, prot, prot_indel*
t03	5	*rib, rib_indel, prot_12, prot_3, prot_indel*
t04	6	*rib, rib_indel, prot_1, prot_2, prot_3, prot_indel*
t05	6	*rib, rib_indel, cox1, cox1_indel, cytb, cytb_indel*
t06	8	*rib, rib_indel, cox1_12, cox1_3, cox1_indel, cytb_12, cytb_3, cytb_indel*
t07	10	*rib, rib_indel, cox1_1, cox1_2, cox1_3, cox1_indel, cytb, cytb_1, cytb_2, cytb_3, cytb_indel*
t08	8	*12s, 12s_indel, 16s, 16s_indel, prot_1, prot_2, prot_3, prot_indel*
t09	12	*12s, 12s_indel, 16s, 16s_indel, cox1_1, cox1_2, cox1_3, cox1_indel, cytb_1, cytb_2, cytb_3, cytb_indel*
t10	4	*cox1* (amminoacids), *cox1_indel, cytb* (amminoacids), *cytb_indel*

[a] *tNy98* and *tM3* were also based on this partitioning scheme.

The Bayesian Information Criterion (BIC) implemented in ModelTest 3.7 (Posada and Crandall, 1998) was used to select the best-fitting models; the graphical interface provided by MrMTgui was used (Nuin, 2008). As MrBayes 3.1.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) currently implements only models with 1, 2 or 6 substitutions, a GTR+I+Γ model (Tavaré, 1986) was chosen for all partitions. ModelTest rejected the presence of a significant proportion of invariable sites in three cases only. GTR+Γ was selected for *cox1* third positions and for *cytb* second and third positions.

Maximum Likelihood was carried out with PAUP* software at the University of Oslo BioPortal (http://www.bioportal.uio.no). Gap characters were treated as missing data and the concatenated alignment was not partitioned. Nucleotides frequencies, substitution rates, gamma shape parameter and proportion of invariable sites were set according to ModelTest results on global alignment. Outgroups were set to be paraphyletic to the monophyletic ingroup. Bootstrap with 100 replicates, using full heuristic ML searches with stepwise additions and TBR branch swapping, was performed to assess nodal support.

Machine time is a key issue in Maximum Likelihood, and, unfortunately, a parallel version of PAUP* has not been published yet. To speed up the process, we used a slightly restricted dataset and set up the analysis to simulate a parallel computation, therefore taking higher advantage of the large computational power of the BioPortal. We run 10 independent bootstrap resamplings with 10 replicates each, starting with different random seeds generated by Microsoft Excel® 2007 following PAUP* recommendations. Trees found in each run were then merged and final consensus was computed with PAUP*. A comparative analysis on a smaller but still representative dataset showed, as expected, that this strategy does not affect the topology of the tree, nor significantly changes bootstrap values (data not shown).

Although less intuitive than in the case of parsimony (Baker and DeSalle, 1997), a Partitioned Likelihood Support (PLS) can be computed for likelihood analyses (Lee and Hugall, 2003). We choose this kind of analysis because other methods (Templeton, 1983; Larson, 1994; Farris et al., 1995a, 1995b) measure overall levels of agreement between partitions in the data set, but they cannot show which parts of a tree are in conflict among partitions (Wiens, 1998; Lambkin et al., 2002). A positive PLS indicates that a partition supports a given clade, and a negative PLS indicates that the partition contradicts the clade itself. Parametric bootstrapping (Huelsenbeck et al., 1996a; Huelsenbeck et al., 1996b) and Shimodaira-Hasegawa test (Shimodaira and Hasegawa, 1999) can assess the statistical significance of PLS results (Goldman et al., 2000; Lee and Hugall, 2003; and reference therein). However, PLS analyses are currently difficult because no widely available phylogenetic software implement such an algorithm. Therefore, Partitioned Likelihood Support (PLS) was evaluated following the manual procedure described in Lee and Hugall (2003). TreeRot 3.0 (Sorenson and Franzosa, 2007) was used to produce PAUP* command file, whereas individual-site log-likelihood scores were analyzed by Microsoft Excel® 2007. Shimodaira-Hasegawa test was employed to assess confidence in

PLS, following Shimodaira and Hasegawa (1999). VBA macros implemented in Microsoft Excel® 2007 to perform PLS and Shimodaira-Hasegawa analyses are available from F. P.

MrBayes 3.1.2 software was used for Bayesian analyses, which were carried out at the BioPortal (see above). We performed a Bayesian analysis for each partitioning scheme. Except as stated elsewhere, two MC^3 algorithm runs with 4 chains were run for 10,000,000 generations; convergence was estimated through PSRF (Gelman and Rubin, 1992) and by plotting standard deviation of average split frequencies sampled every 1,000 generations. The four outgroups were constrained, trees found at convergence were retained after the burnin, and a majority-rule consensus tree was computed with the command **`sumt`**. Via the command **`sump printtofile=yes`** we could obtain the harmonic mean of the Estimated Marginal Likelihood (EML). EML was used to address model selection and partition choice.

Since there is no obvious way to define partitions in ribosomal-encoding genes and secondary structure-based alignments did not result in correct phylogenetic trees (data not shown; see also Steiner and Hammer, 2000), we first decided to test data partitioning schemes on protein-coding genes only. Therefore, after a global analysis merging all markers within the same set, we tested six different partitioning schemes for protein-coding genes, taking ribosomal ones together (Tab. 2.2; *t02-t07*). As *t04* and *t07* were selected as the most suitable ones (see Results, Bayesian Analyses), we designed two more schemes splitting *12s* and *16s* based on these datasets only (Tab. 2.2; *t08-t09*). Finally, we tested some strategies to further remove phylogenetic noise: we first constructed an amminoacid dataset (Tab. 2.2; *t10*; we were forced to completely remove ribosomal genes, as MC^3 runs could not converge in this case). However, the use of amminoacids is not directly comparable with other datasets by AIC and BF, because it not only implies a different model, but also different starting data: as a consequence, we implemented the codon model (Goldman and Yang, 1994; Muse and Gaut, 1994) on the

prot partition. This allowed us to start from an identical dataset, which makes results statistically comparable. As *t04* scheme turned out to be essentially comparable with *t09* (see Results, Bayesian analysis), we did not implement codon model also on separate *cox1* and *cytb* genes, because codon model is computationally extremely demanding. Two separate analyses were performed under such a codon model: in both cases, metazoan mitochondrial genetic code table was used; in one case Ny98 model was enforced (*tNy98*; Nielsen and Yang, 1998), whereas in the other case M3 model was used (*tM3*). Only one run of 5,000,000 generations was performed for codon models, sampling a tree every 125. Dealing with one-run analyses, codon models trees were also analytically tested for convergence via AWTY analyses (http://king2.scs.fsu.edu/ CEBProjects/ awty/ awty_start.php; Nylander et al., 2008). Moreover, our analysis on codon models allowed us to test for positive selection on protein-coding genes (see Ballard and Whitlock, 2004): MrBayes estimates the ratio of the non-synonymous to the synonymous substitution rate (ω) and implements models to accommodate variation of ω across sites using three discrete categories (Ronquist et al., 2005).

Finally, to test for the best partitioning scheme and evolutionary model, we applied Akaike Information Criterion (AIC; Akaike, 1973) and Bayes Factors (BF; Kass and Raftery, 1995). AIC was calculated, following Huelsenbeck et al. (2004), Posada and Buckley (2004), and Strugnell et al. (2005), as

$$AIC = -2EML + 2K$$

The number of free parameters *K* was computed taking into account branch number, character (nucleotide, presence/absence of an indel, amminoacid, or codon and codon-related parameters) frequencies, substitution rates, gamma shape parameter and proportion of invariable sites for each partition.

Bayes Factors were calculated, following Brandley et al. (2005), as

$$B_{ij} = \frac{EML_i}{EML_j}$$

and, doubling and turning to logarithms,

$$2\ln B_{ij} = 2\left(\ln EML_i - \ln EML_j\right)$$

where B_{ij} is the Bayes Factor measuring the strength of the *i*th hypothesis on the *j*th hypothesis. Bayes Factors were interpreted according to Kass and Raftery (1995) and Brandley et al. (2005).

All trees were graphically edited by PhyloWidget (Jordan and Piel, 2008) and Dendroscope (Huson et al., 2007) softwares. Published Maximum Likelihood and Bayesian trees, along with source data matrices, were deposited in TreeBASE under SN4787 and SN4789 Submission ID Numbers, respectively.

Taxon sampling

Taxon sampling is a crucial step in any phylogenetic analysis, and this is certainly true for bivalves (Giribet and Carranza, 1999; Puslednik and Serb, 2008). Actually, many authors claim for a bias in taxon sampling to explain some unexpected or unlikely results (Adamkewicz et al., 1997; Canapa et al., 1999; Campbell, 2000; Kappner and Bieler, 2006). As we want to find the best performing methodological pipeline for reconstructing bivalve phylogeny, we assessed taxon sampling following rigorous criteria, in order to avoid misleading results due to incorrect taxon choice. We approached this with both *a priori* and *a posteriori* perspectives, following two different (and complementary) rationales.

Quite often, taxa that are included in a phylogenetic analysis are not chosen following a formal criterion of representativeness: they are rather selected on accessibility and/or analyzer's personal choice. To avoid this, we developed a method to quantify sample representativeness with respect to the whole class. The method is based on Average Taxonomic Distinctness (AvTD) of Clarke and Warwick (1998). The mathematics of this method has been proposed in a different paper (Plazzi et al., 2010), but here we would like to mention the rationale behind it: estimating *a priori* the *phylogenetic representativeness*

of a sample is not conceptually different from estimating its *taxonomic representativeness*, i.e. testing whether our taxon sampling is representative of a given master taxonomic list, which may eventually be retrieved from bibliography. This approach does not require any specific knowledge, other than the established taxonomy of the sampled taxa; neither sequence data, nor any kind of measure are used here, which means the AvTD approach comes *before* seeing the data. Our source of reference taxonomy (master list) was obtained from Millard (2001). The AvTD was then computed for our sample and confidence limits were computed on 1,000 random resamplings of the same size from bivalve master list. If the taxon sample value is above the 95% lower confidence limit, then we can say that our dataset is representative of the whole group. We developed a software to compute this, which is available for download at www.mozoolab.net.

On the other hand, *after* seeing the data, we were interested in answering whether they were sufficient or not to accurately estimate phylogeny. For this purpose, we used the method proposed by Sullivan et al. (1999). The starting point is the tree obtained as the result of our analysis, given the correct model choice (see below). Several subtrees are obtained by pruning it without affecting branch lengths; each parameter is then estimated again from each subtree under the same model: if estimates, as size increases, converge to the values computed from the complete tree, then taxon sampling is sufficiently large to unveil optimal values of molecular parameters, such as evolutionary rates, proportion of invariable sites, and so on (Townsend, 2007). At first, we checked whether MC^3 Bayesian estimates of best model were comparable to Maximum Likelihood ones computed through ModelTest. We took into consideration all 6 mutations rates and, where present, nucleotide frequencies, invariable sites proportion and gamma-shaping parameter (which are not used into M3 codon model). In most cases (see Appendix 2.5) the Maximum Likelihood estimate fell within the 95% confidence interval as computed following Bayesian Analysis and, if not, the difference was always (except in one case) of 10^{-2} or less order of

magnitude. Therefore, we used Bayesian estimates of mean and confidence interval limits instead of bootstrapping Maximum Likelihood, as in the original method of Sullivan et al. (1999). 50 subtrees were manually generated from best tree by pruning a number of branches ranging from 1 to 50. Following Authors' suggestions, we used different pruning strategies: in some cases, we left only species very close in the original tree, whereas in others we left species encompassing the whole biodiversity of the class (Appendix 2.6). Model parameters were then estimated from each subtree for each partition (*rib* and *prot*) using original sequence data and the best model chosen by ModelTest as above. The paupblock of ModelTest was used into PAUP* to implement such specific Maximum Likelihood analyses for each partition, model, and subtree.

Dating

The r8s 1.71 (Sanderson, 2003) software was used to date the best tree we obtained. Fossil collections of bivalves are very abundant, so we could test several calibration points in our tree, but in all cases the origin of Bivalvia was constrained between 530 and 520 million years ago (Mya; Brasier and Hewitt, 1978), and no other deep node was used for calibration, as we were interested in molecular dating of ancient splits. Data from several taxa were downloaded from the Paleobiology Database on 4 November, 2009, using group names given in Table 2.3 and leaving all parameters as default. Some nodes were fixed or constrained to the given age, whereas others were left free. After the analysis, we checked whether the software was able to predict correct ages or not, i.e. whether the calibration set was reliable. The tree was re-rooted with the sole *Katharina tunicata*; for this reason, two nodes "*Katharina tunicata*" and "other outgroups" are given in Table 2.3. Rates and times were estimated following both PL and NPRS methods, which yielded very similar results. In both cases we implemented the Powell's algorithm. Several rounds of fossil-based cross-validation analysis were used to determine

the best-performing smoothing value for PL method and the penalty function was set to log. 4 perturbations of the solutions and 5 multiple starts were invoked to optimize searching in both cases. Solutions were checked through the **`checkGradient`** command. NPRS method was also used to test variability among results. 150 bootstrap replicates of original dataset were generated by the SEQBOOT program in PHYLIP (Felsenstein, 1993) and branch lengths were computed with PAUP* through r8s-bootkit scripts of Torsten Eriksson (2007). A complete NPRS analysis was performed on each bootstrap replicate tree and results were finally profiled across all replicates through the r8s command **`profile`**.

TABLE 2.3. r8s datation of *tM3* tree. If a fossil datation is shown, the clade was used for calibrating the tree using Paleobiology Database data; in bold are shown the eight calibrations point of the best-performing set, whereas the others were used as controls. Constraints enforced are shown in the fourth and fifth column; if they are identical, that node was fixed. Ages are in millions of years (Myr); rates are in substitutions per year per site and refer to the branch leading to a given node. PL, Penalized Likelihood; NPRS, Non Parametric Rate Smoothing; StDev, Standard Deviation.

	Fossil datation	Reference[a]	Constraints		PL		NPRS			
			Min	Max	Age	Local rate	Age	Local rate	Mean	StDev
Katharina tunicata					627.58		625.44			
other outgroups					561.45	1.65E-03	560.05	1.67E-03	533.95	2.67
Bivalvia	**530.0-520.0**	**5**	**520.00**	**530.00**	**529.99**	**3.46E-03**	**530.00**	**3.63E-03**	**530.00**	**0.00**
Autolamellibranchiata					520.32	2.01E-02	520.31	2.01E-02	517.04	1.70
Pteriomorphia+Heterodonta					513.59	2.26E-02	513.59	2.26E-02	508.51	1.74
Pteriomorphia					505.74	1.81E-02	505.82	1.83E-02	501.13	2.29
Heterodonta					497.83	1.51E-02	498.20	1.55E-02	490.24	3.11
traditional Pteriomorphia					496.63	1.26E-02	496.13	1.19E-02	488.88	2.38
Hiatella+Cardiidae					481.34	1.10E-02	481.61	1.09E-02	476.05	3.65
Limidae+Pectinina					474.51	1.71E-02	474.82	1.78E-02	468.49	3.49
Veneroida *sensu lato*					471.38	3.80E-03	471.87	3.82E-03	471.22	6.63
Anomioidea+Pectinoidea					464.44	1.19E-02	464.92	1.21E-02	459.25	4.26
Protobranchia					454.28	1.34E-03	455.67	1.37E-03	482.02	14.61
Arcidae	**457.5-449.5**	**29**	**449.50**	**457.50**	**449.51**	**2.35E-02**	**449.50**	**2.38E-02**	**449.50**	**0.00**
Pectinoidea	428.2-426.2	21, 27, 30			431.77	1.27E-02	433.44	1.32E-02	417.82	4.20
Anomalodesmata					431.45	3.29E-03	434.04	3.40E-03	461.87	9.59
Cardiidae	**428.2-426.2**	**18**	**427.20**	**427.20**	**427.20**	**1.18E-02**	**427.20**	**1.18E-02**	**427.20**	**0.00**
Cuspidaria clade					418.58	4.87E-03	421.63	5.04E-03	477.22	9.28
Veneroida 2					407.08	3.58E-03	407.42	3.58E-03	410.56	9.26
Ostreoida+Pteriida					393.59	3.48E-03	395.13	3.55E-03	435.47	10.95
Pectinidae	**388.1-383.7**	**2, 6, 14, 22, 26**	**385.90**	**385.90**	**385.90**	**5.18E-03**	**385.90**	**5.00E-03**	**385.90**	**0.00**
Limidae	**376.1-360.7**	**1**	**360.70**	**376.10**	**360.74**	**4.66E-03**	**360.71**	**4.65E-03**	**370.13**	**6.31**
Veneridae	**360.7-345.3**	**19, 30**	**345.30**	**360.70**	**345.33**	**3.30E-03**	**345.31**	**3.28E-03**	**347.28**	**4.57**
Pectininae					324.88	1.57E-03	327.18	1.63E-03	342.84	7.76
Unionidae	245.0-228.0	8			293.93	3.68E-03	298.00	3.74E-03	347.74	20.25
Gafrarium+Gemma					282.57	2.24E-03	283.03	2.25E-03	280.55	22.38

Ostreoida	251.0-249.7	28			264.75	3.00E-03	266.21	3.00E-03	333.04	16.09
Mactrinae	196.5-189.6	25			243.80	2.27E-03	244.76	2.28E-03	261.16	21.60
Argopecten+Pecten					220.05	1.22E-03	222.43	1.22E-03	256.84	14.94
Unioninae	**228.0-216.5**	**9, 13, 16, 20, 23**	**216.50**	**228.00**	**216.53**	**1.71E-03**	**216.51**	**1.62E-03**	**227.86**	**0.93**
Chlamys livida+Mimachlamys					190.34	1.24E-03	194.24	1.27E-03	336.20	8.12
Ensis+Sinonovacula					189.33	1.16E-03	189.83	1.16E-03	305.30	18.57
Astarte+Cardita					188.86	3.26E-03	191.12	3.25E-03	274.37	23.58
Dreissena+Mya					185.03	2.62E-03	185.82	2.62E-03	224.89	19.55
Barbatia	**167.7-164.7**	**4, 10, 24**	**166.20**	**166.20**	**166.20**	**6.93E-04**	**166.20**	**6.93E-04**	**166.20**	**0.00**
Tridacna	23.0-16.0	17			147.15	1.26E-03	149.69	1.27E-03	383.21	11.43
setigera+reeveana					77.29	2.20E-03	75.19	2.15E-03	92.77	12.17
Crassostrea	145.5-130.0	15			63.17	3.08E-03	63.52	3.07E-03	92.38	10.04
gigas+hongkongensis					23.47	2.72E-03	23.65	2.71E-03	36.93	9.36
Mactra	196.5-189.6	25			21.63	1.50E-03	21.80	1.49E-03	31.48	6.91
Mytilus	418.7-418.1	3, 7, 11, 12			1.88	2.92E-03	1.77	2.92E-03	1.79	0.60

[a] References as follows: (1) Amler et al. (1990); (2) Baird and Brett (1983); (3) Berry and Boucot (1973); (4) Bigot (1935); (5) Brasier and Hewitt (1978); (6) Brett et al. (1991); (7) Cai et al. (1993); (8) Campbell et al. (2003); (9) Chatterjee (1986); (10) Cox (1965); (11) Dou and Sun (1983); (12) Dou and Sun (1985); (13) Elder (1987); (14) Grasso (1986); (15) Hayami (1975); (16) Heckert (2004); (17) Kemp (1976); (18) Kříž (1999); (19) Laudon (1931); (20) Lehman and Chatterjee (2005); (21) Manten (1971); (22) Mergl and Massa (1992); (23) Murry (1989); (24) Palmer (1979); (25) Poulton (1991); (26) Rode and Lieberman (2004); (27) Samtleben et al. (1996); (28) Spath (1930); (29) Suarez Soruco (1976); (30) Wagner (2008).

2.3. RESULTS

Obtained sequences

Mitochondrial sequences from partial ribosomal small (*12s*) and large subunit (*16s*), cytochrome b (*cytb*) and cytochrome oxidase subunit I (*cox1*) were obtained; GenBank accession numbers are reported in Appendix 2.3. A total of 179 sequences from 57 bivalve species were used for this study: 80 sequences from 28 species were obtained in our laboratory, whereas the others were retrieved from GenBank (see App. 2.3 for details). Alignment was made by 55 taxa and 2501 sites, 592 of which, all within *12s* and *16s* genes, were excluded because they were alignment-ambiguous. After removal, 1623 sites were variable and 1480 were parsimony-informative. It is clearly impossible to show here a complete p-distance table, but the overall average value was 0.43 (computed by MEGA4, with pairwise deletion of gaps).

Quite interestingly, we found few anomalies in some of the sequences: for instance, a single-base deletion was present in *cytb* of *Hyotissa hyotis* and *Barbatia* cfr. *setigera* at position 2317 and 2450, respectively. This can suggest three possibilities: *i)* we could have amplified a mitochondrial pseudogene (NUMT); *ii)* we could have faced a real frameshift mutation, which may eventually end with a compensatory one-base insertion shortly downstream (not visible, since our sequence ends quite soon after deletion); *iii)* an error in base calling was done by the sequencer. At present no NUMTs have been observed in bivalves (Bensasson et al., 2001; Zbawicka et al., 2007) and the remaining DNA sequences are perfectly aligned with the others, which is unusual for a NUMT; therefore, we think that the second or the third hypotheses are more sound. In all subsequent analyses, we inserted missing data both in nucleotide and in amminoacid alignments. Moreover, several stop codons were found in *Anomia sp.* sequences (within *cox1*, starting at position 1796 and 1913; within *cytb*, starting at 2154, 2226, 2370, 2472 and 2484).

Again, we could have amplified two pseudogenes; however, all these stop codons are TAA and the alignment is otherwise good. A possible explanation is an exception to the mitochondrial code of this species, which surely demands further analysis, but this is beyond the scope of this paper. In any case, we kept both sequences and placed missing data in protein and codon model alignments in order to perform subsequent analyses. Of course, phylogenetic positions of all the above-mentioned species have been considered with extreme care, taking into account their sequence anomalies.

Sequence analyses

No saturation signal was observed by plotting uncorrected p-distances as described above (see Appendix 2.7), since all linear interpolations were highly significant as computed with PaSt 1.90. Moreover, deleting third codon positions we obtained a completely unresolved Bayesian tree, confirming that these sites carry some phylogenetic signal (data not shown).

Selective pressures on protein coding genes were tested through ω. In the Ny98 model (Nielsen and Yang, 1998), there are three classes with different potential ω values: $0 < \omega_1 < 1$, $\omega_2 = 1$, and $\omega_3 > 1$. The M3 model also has three classes of ω values, but these values are less constrained, in that they only have to be ordered $\omega_1 < \omega_2 < \omega_3$ (Ronquist et al., 2005). As M3 was chosen as the best model for our analysis (see below), we only considered M3 estimates about ω and its heterogeneity. Boundaries estimates for *tM3* are very far from one (Appendix 2.8) and more than 75% of codon sites fell into the first two categories. Moreover, all codon sites scored 0 as the probability of being positively selected. Therefore, we conclude that only a stabilizing pressure may be at work on these markers, which may enhance their phylogenetic relevance. This also allows to analyze protein-coding genes together.

Taxon sampling

Appendix 2.9 shows results from Average Taxonomic Distinctness test. Our sample plotted almost exactly on the mean of 1,000 same-size random subsamples from the master list of bivalve genera, thus confirming that our sample is a statistically representative subsample of the bivalves' systematics.

Appendix 2.10 shows results from *a posteriori* testing of parameter accurateness. Analysis was carried out for all main parameters describing the models, but, for clarity, only gamma shaping parameters (alpha) and invariable sites proportions (pinv) for *rib* partition are shown. In any case, all parameters behaved the same way: specifically, estimates became very close to "true" ones starting from subtrees made by 30-32 taxa. Therefore, at this size a dataset is informative about evolutionary estimates, given our approach. As we sampled nearly twice this size, this strengthens once again the representativeness of our taxon choice – this time from a molecular evolution point of view.

Maximum Likelihood

Maximum Likelihood analysis gave the tree depicted in Figure 2.1. The method could not resolve completely the phylogeny: bivalves appear to be polyphyletic, as the group corresponding to Protobranchia (*Nucula*+*Solemya*) is clustered among non-bivalve species, although with low support (BP=68). A first node (BP=100) separates Palaeoheterodonta (*Inversidens*+*Lampsilis*) from the other groups. A second weak node (BP=51) leads to two clades, one corresponding to Pteriomorphia+*Thracia* (BP=68) and the other, more supported, to Heterodonta (BP=83). A wide polytomy is evident among Pteriomorphia, with some supported groups in it, such as *Thracia*, *Mytilus*, Arcidae (all BP=100), Limidae+Pectinina (BP=87), and Pteriida+Ostreina (BP=85). Heterodonta subclass is also not well resolved, with *Astarte*+*Cardita* (BP=100) as sister group of a large

polytomy (BP=73) that includes *Donax*, *Ensis*, *Hiatella*+(*Acanthocardia*+*Tridacna*), and an heterogeneous group with Veneridae, *Spisula*, *Dreissena* and *Mya* (BP=66).

PLS tests turned out to be largely significant (Appendix 2.11). High likelihood support values were always connected with highly supported nodes, whereas the opposite is not always true (see node 11). High positive PLS values are generally showed by the *cytb* partition; good values can also be noted for *cox1* and *16s* genes, even if *16s* is sometimes notably against a given node (see nodes 23 and 24). *12s* has generally low PLS absolute values, with some notable exceptions (see nodes 15 and 16). Globally, deeper splits (see nodes 6, 13, 14, 22, 23, 24, 29) have a low likelihood support absolute value, and generally a low bootstrap score too.

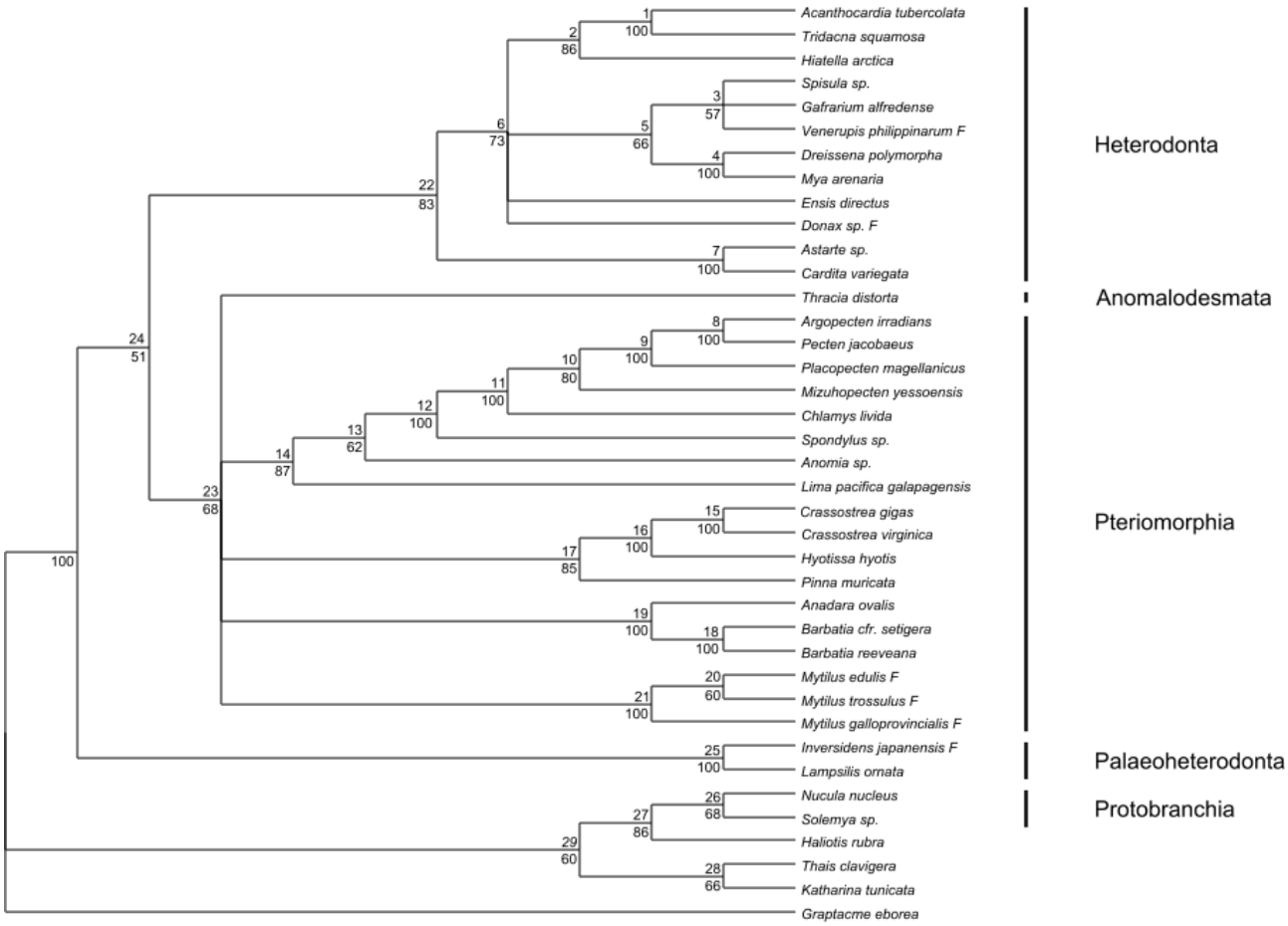


FIGURE 2.1. Majority-rule consensus tree of 100 Maximum Likelihood bootstrap replicates: node have been numbered (above branches), and numbers below the nodes are bootstrap proportions.

Bayesian Analyses

Table 2.4 shows results of model-decision statistical tests. Among classical 4by4 models (i.e., not codon models) AIC favored *t04* as best trade-off between partitions number and free parameters. However, if considered, *tM3* (a codon model) was clearly favored. As BF does not take into account the number of free parameters, *t04* is not clearly the best classical 4by4 model in this case. More complex models (with the notable exception of *t05*) turned out to be slightly favored: *t09*, the most complex model we implemented, has positive (albeit small) BF values against each simpler partition scheme. Again, when considered, *tM3* is straightforwardly the best model, with the highest BF scores in the matrix (see Tab. 2.4). It is notable that *tNy98*, even not the worst, has instead very low BF scores. Therefore, using *tM3* we obtained the best phylogenetic tree, which is shown in Figure 2.2. In this tree, several clusters agreeing with the established taxonomy are present: the first corresponds to Protobranchia (*sensu* Giribet and Wheeler, 2002) and it is basal to all the remaining bivalves (Autolamellibranchiata *sensu* Bieler and Mikkelsen, 2006; PP=1.00). A second group, which is basal to the rest of the tree, is composed by Palaeoheterodonta (PP=1.00). Sister group to Palaeoheterodonta a major clade is found (PP=1.00), in which three main groups do separate. Heterodonta constitute a cluster (PP=1.00), with two branches: *Hiatella*+Cardiidae (PP=1.00) and other heterodonts (PP=0.98). Within them, only one node remains unresolved, leading to a Veneridae+Mactridae+(*Dreissena*+*Mya*) polytomy. Another cluster (PP=0.96) is made by *Pandora*+*Thracia*, as sister group of all Pteriomorphia+*Nuculana* (both PP=1.00). A wide polytomy is evident within Pteriomorphia, with *Mytilus* species, Limidae+Pectinina, Pteriida+Ostreina, Arcidae and *Nuculana* itself as branches, all with PP=1.00. Another cluster (PP=1.00) is made by *Cuspidaria*+(*Astarte*+*Cardita*). All families have PP=1.00: Cardiidae (genera *Acanthocardia* and *Tridacna;* see Discussion, Phylogenetic Inferences about Evolution of Bivalves), Mactridae (genera *Mactra* and *Spisula*), Veneridae (genera

Gafrarium, *Gemma* and *Venerupis*), Unionidae (genera *Hyriopsis*, *Inversidens*, *Anodonta* and *Lampsilis*), Arcidae (genera *Anadara* and *Barbatia*), Limidae (genera *Acesta* and *Lima*), Ostreidae (genera *Crassostrea* and *Hyotissa*) and Pectinidae (genera *Mizuhopecten*, *Chlamys*, *Mimachlamys*, *Argopecten*, *Pecten* and *Placopecten*).

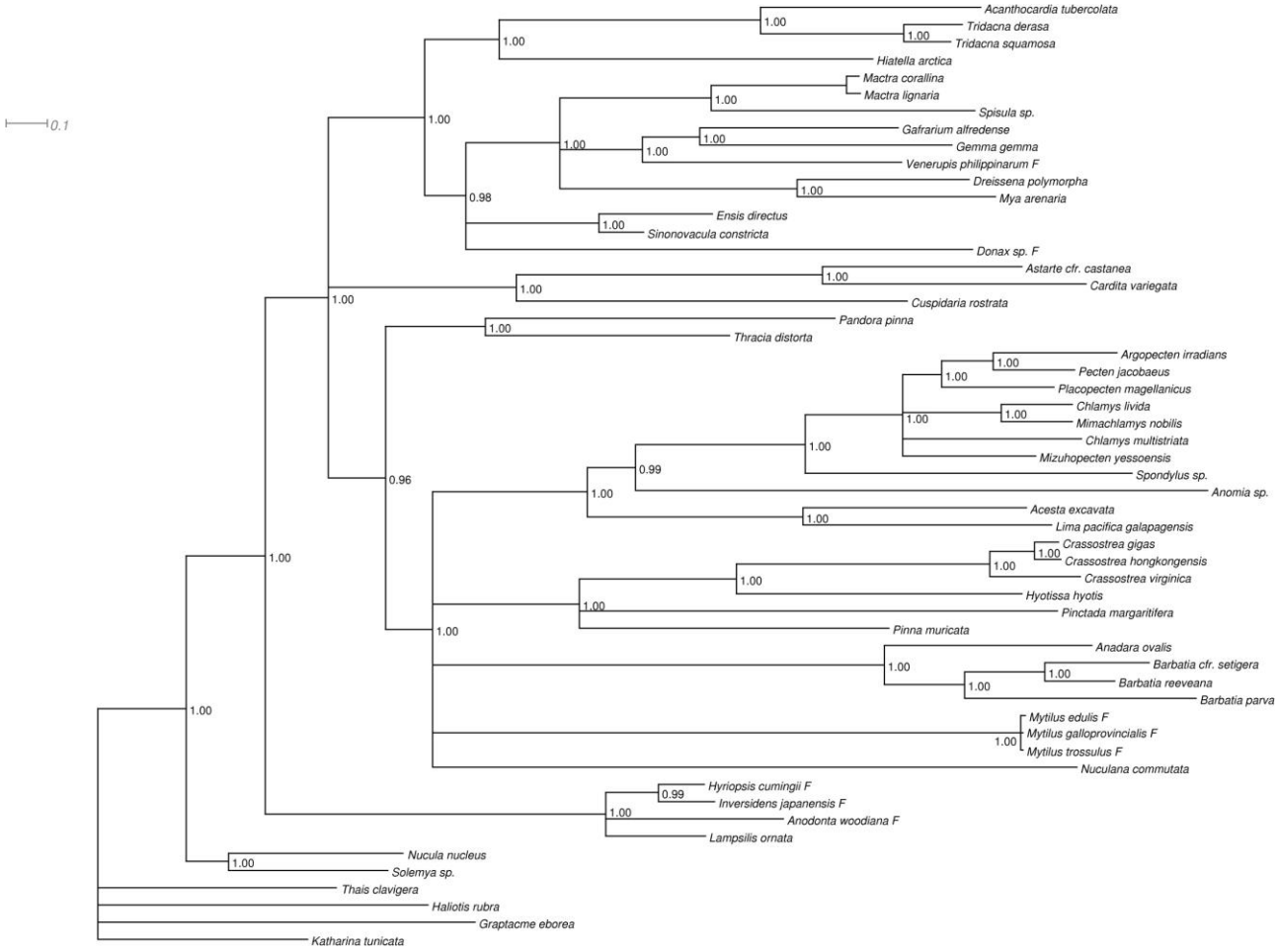


FIGURE 2.2. Majority-rule *tM3* consensus tree from the Bayesian multigene partitioned analysis. Number at the nodes are PP values. Nodes under 0.95 were collapsed. Bar units in substitution per year per site.

Table 2.4. Results from Akaike Information Criterion (AIC) and Bayes Factors (BF) tests. EML, Estimated Marginal Likelihood; p, number of partitions in the partitioning scheme; FP, Free Parameters. Partitioning schemes as in Table 2.2.

Tree	EML	p	FP	AIC	*t02*	*t03*	*t04*	*t05*	*t06*	*t07*	*t08*	*t09*	*t10*	*tNy98*	*tM3*
t01	-64,914.04	2	225	130,278.08	479.76	1,870.00	2,203.28	494.92	1,950.86	2,290.48	2,326.90	2,424.26	N/A	884.14	3,721.44
t02	-64,674.16	4	450	130,248.32		1,390.24	1,723.52	15.16	1,471.10	1,810.72	1,847.14	1,944.50	N/A	404.38	3,241.68
t03	-63,979.04	5	567	129,092.08			333.28	-1,375.08	80.86	420.48	456.90	554.26	N/A	-985.86	1,851.44
t04	-63,812.40	6	684	128,992.80				-1,708.36	-252.42	87.20	123.62	220.98	N/A	-1,319.14	1,518.16
t05	-64,666.58	6	675	130,683.16					1,455.94	1,795.56	1,831.98	1,929.34	N/A	389.22	3,226.52
t06	-63,938.61	8	907	129,691.22						339.62	376.04	473.40	N/A	-1,066.72	1,770.58
t07	-63,768.80	10	1,140	129,817.60							36.42	133.78	N/A	-1,406.34	1,430.96
t08	-63,750.59	8	909	129,319.18								97.36	N/A	-1,442.76	1,394.54
t09	-63,701.91	12	1,365	130,133.82									N/A	-1,540.12	1,297.18
t10	-13,725.38	4	450	28,350.76										N/A	N/A
tNy98	-64,471.97	4	512	129,967.94											2,837.30
tM3	-63,053.32	4	513	127,132.64											

Dating the tree

Results from r8s software are shown in Table 2.3. The relative ultrametric tree is shown in Figure 2.3 along with the geological timescale. The best-performing smoothing value for PL analysis was set to 7.26 after a fossil-based cross-validation with an increment of 0.01. The best calibration set comprises genus *Barbatia*, subfamily Unioninae, families Veneridae, Limidae, Pectinidae, Cardiidae, Arcidae, and Bivalvia; all constraints were respected. Age for many other taxa were correctly predicted with an error of always less than 50 million years (Myr), as shown in Table 2.3.

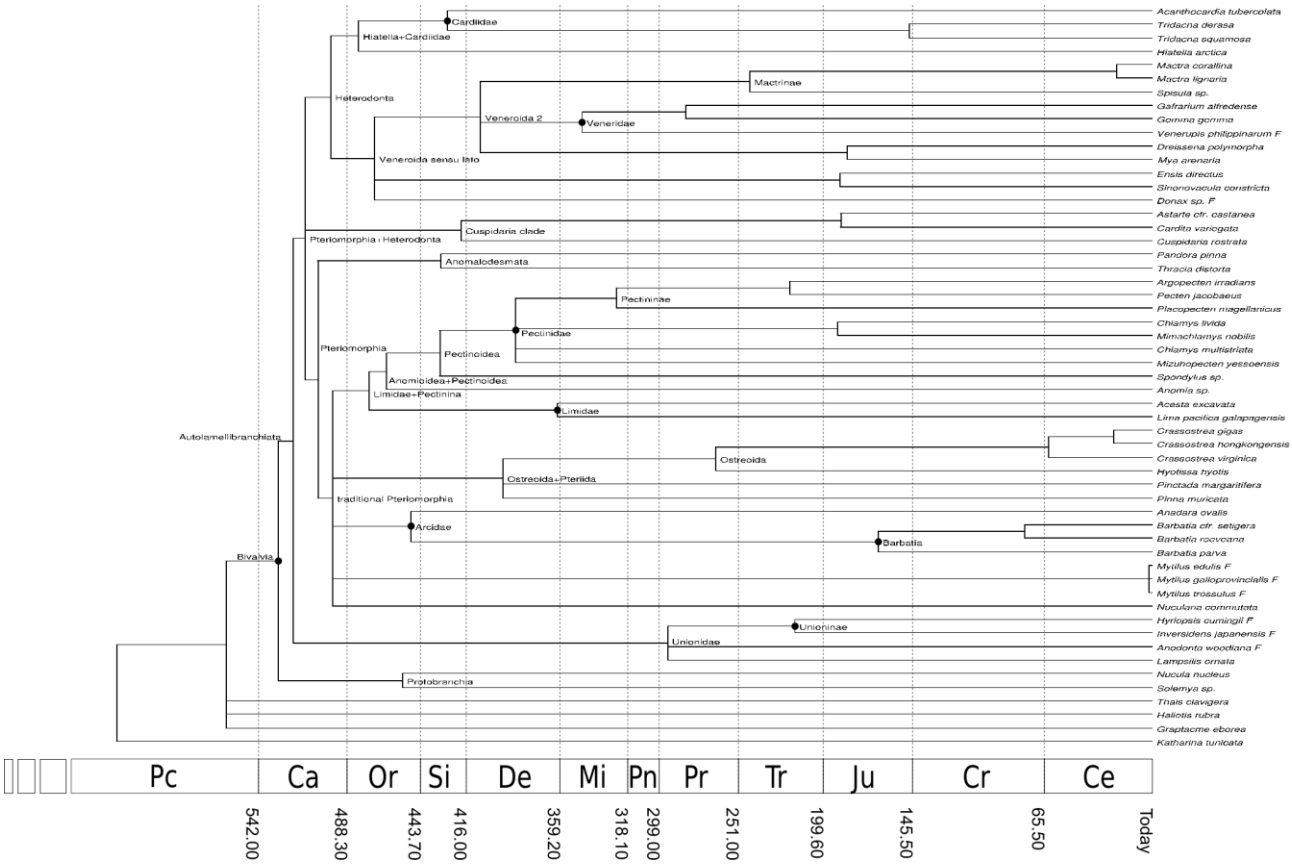


FIGURE 2.3. Results from time calibration of *tM3* tree. The ultrametric *tM3* tree computed by r8s (under Penalized Likelihood method, see text for further details) is shown along with geological time scale and major interval boundaries (ages in million years). Only deep nodes are named: for a complete survey of node datations, see Table 2.3. Geological data taken from Gradstein et al. (2004) and Ogg et al. (2008). Pc, Precambrian (partial); Ca, Cambrian; Or, Ordovician, Si, Silurian; De, Devonian; Mi, Mississippian; Pn, Pennsylvanian; Pr, Permian; Tr, Triassic; Ju, Jurassic; Cr, Cretaceous; Ce, Cenozoic.

This was not the case for genera *Mytilus*, *Mactra*, *Crassostrea*, and *Tridacna*: with the notable exception of *Tridacna*, they were predicted to be much more recent than they appeared in fossil records. This is easily explained by the fact that in all cases (except *Tridacna*) strictly related species were represented in our tree, which diverged well after the first appearance of the genus. Results from PL and NPRS were substantially identical: as in four cases NPRS analysis did not pass the **`checkGradient`** control, we will present and discuss PL results only.

Deep nodes were all dated between 530 and 450 million years ago (see Fig. 2.3): the origin of the class was dated 530 Mya, Autolamellibranchiata 520 Mya and their sister group Protobranchia 454 Mya. Within Autolamellibranchiata, the big group comprehending Heterodonta and Pteriomorphia would have arisen about 514 Mya; the radiation of Palaeoheterodonta was not computed as only specimens from Unionidae (293.93 Mya) were present. Pteriomorphia and Heterodonta originated very close in time, about 506 and 498 Mya, respectively. Within Pteriomorphia, the basal clade of Anomalodesmata is more recent (431 Mya) than the main group of traditional Pteriomorphia (497 Mya). On the other hand, the main split within Heterodonta gave rise to *Hiatella*+Cardiidae about 481 Mya, and to Veneroida *sensu lato* 471 Mya. Evolutionary rates (expressed as mutations per year per site) varied consistently, ranging from 0.000693 of branch leading to genus *Barbatia* to 0.011 of the *Hiatella*+Cardiidae group. Table 2.3 also lists the mean value of NPRS dating across 150 bootstrap replicates and its standard deviation, and it is worth noting that deeper nodes do have very little standard deviation.

2.4. DISCUSSION

The methodological pipeline

As the correct selection of suitable molecular markers was (and still is) a major concern in bivalves' phylogenetic analysis, we tested for different ways of treating the data. Our best-performing approach is based on four different mitochondrial genes, and because we obtained robust and reliable phylogenies in our analysis, we can now confirm that this choice is particularly appropriate in addressing deep phylogeny of Bivalvia, given a robust analytical apparatus.

As mentioned, our mitochondrial markers were highly informative, especially protein-coding ones and our results from model selection were straightforward. The phylogenetic signal we recovered in our dataset is complex, as different genes and different positions must have experienced different histories and selective pressures. Moreover, performed single-gene analyses yielded controversial and poorly informative trees (data not shown).

Specifically, both AIC and BF separated ribosomal and protein-coding genes for traditional 4by4 models. AIC tends to avoid overparametrization, as it presents a penalty computed on free parameters, and selected a simpler model; conversely, BF selected the most complex partitioning scheme. BF has been proposed to be generally preferable to AIC (Kass and Raftery, 1995; Alfaro and Huelsenbeck, 2006), but Nylander et al. (2004) pointed out that BF is generally consistent with other model selection methods, like AIC. Indeed, trees obtained under models *t04*, *t07*, *t08*, and *t09* are very similar (data not shown). Anyway, the *tM3* model clearly outperformed all alternatives, following both AIC and BF criteria (see Tab. 2.4). Furthermore, this was not the case for models *tNy98* and *t10*, which we used to reduce possible misleading phylogenetic noise, albeit in different ways (by a Ny98 codon model or by amminoacids, respectively). *t10* tree was similar to *tM3* one, but significantly less resolved on many nodes, thus indicating a loss of

informative signal (data not shown). M3 codon model allows lower ω categories than Ny98; on the other hand, it does not completely eliminate nucleotide information level, as amminoacid models do. All this considered, we propose that M3 codon model is the best way for investigating bivalve phylogeny.

Finally, it is quite evident that Bayesian analysis yielded the most resolved trees, when compared to Maximum Likelihood and this was especially evident for ancient nodes. The tendency of Bayesian algorithms to higher nodal support has been repeatedly demonstrated (Leaché and Reeder, 2002; Suzuki et al., 2002; Whittingham et al., 2002; Cummings et al., 2003; Douady et al., 2003; Erixon et al., 2003; Simmons et al., 2004; Cameron et al., 2007), though Alfaro et al. (2003) found that PP is usually a less biased predictor of phylogenetic accuracy than bootstrap. Anyway, it has to be noted that most of our recovered nodes are strongly supported by both methods; we therefore think that the higher support of Bayesian analysis is rather due to a great affordability of the method in shaping and partitioning models, which is nowadays impossible with Maximum Likelihood algorithms. All that considered, we suggest that a suitable methodological pipeline for bivalves' future phylogenetic reconstructions should be as such: *i)* sequence analyses for saturation and selection; *ii)* rigorous evaluation of taxon coverage; *iii)* tests for best data partitioning; *iv*) appropriate model decision statistics; *v)* Bayesian analysis; *vi)* eventual dating by cross-validation with fossil records.

The phylogeny of Bivalvia

Protobranchia Pelseneer. – Our study confirms most of the recent findings (Giribet and Wheeler, 2002; Giribet and Distel, 2003; Kappner and Bieler, 2006): Nuculoidea and Solemyoidea do maintain their basal position, thus representing Protobranchia *sensu stricto*, which is a sister group to all Autolamellibranchiata. On the contrary, Nuculanoidea, although formerly placed in Nuculoida, is better considered within Pteriomorphia, placed in

its own order Nuculanoida. The split separating *Nucula* and *Solemya* lineages is dated around the late Ordovician (454.28 Mya); since the first species of the subclass must have evolved earlier (about 500 Mya), this is a clear signal of the antiquity of this clade. In fact, based on paleontological records, the first appearance of Protobranchia is estimated around 520 Mya (early Cambrian) (He et al., 1984; Parkhaev, 2004), and our datation is only slightly different (482.02 Mya, with a standard deviation of 14.61).

Palaeoheterodonta Newell. – Freshwater mussels are basal to all the remaining Autolamellibranchiata (Heterodonta+Pteriomorphia), as supposed by Cope (1996). Therefore, there is no evidence for Heteroconchia *sensu* Bieler and Mikkelsen (2006) in our analysis. The monophyletic status of the subclass was never challenged in our Bayesian analyses, nor in traditional Maximum Likelihood ones. Finally, since we obtained sequences only from specimens from Unionoidea:Unionidae, a clear dating of the whole subclass is not sound, as shown by a relatively high difference between PL values and mean across bootstrap replicates (294 and 348 Mya, respectively). Therefore, the origin of the subclass must date back to before than 350 Mya, which is comparable to paleontological data (Morton, 1996).

Pteriomorphia Newell. – Here we obtained a Pteriomorphia *sensu novo* subclass comprising all pteriomorphians *sensu* Millard (2001), as well as Nuculanoidea and anomalodesmatans. This diverse taxon arose about 506 Mya, which makes it the first bivalve radiation in our tree, dated in the middle Cambrian, which is perfectly in agreement with paleontological data. Moreover, our results proved to be stable also with bootstrap resampling, with a standard deviation of slightly more than 2 million of years (Tab. 2.3). A wide polytomy is present within the subclass; as this polytomy is constantly present in all the analyses, and it has been found also by many other authors (see Campbell, 2000; Steiner and Hammer, 2000; Matsumoto, 2003), we consider it as a “hard polytomy”, reflecting a true rapid radiation dated about 490 Mya (Cambrian/Ordovician boundary).

Sister group to this wide polytomy is the former anomalodesmatan suborder Pholadomyina. In our estimate, the clade *Pandora*+*Thracia* seems to have originated something like 431.45 Mya, as several pteriomorphian groups, like Pectinoidea (431.77 Mya) or Arcidae (449.51 Mya). On the other hand, we failed in retrieving *Cuspidaria* within the pteriomorphian clade, while this genus is strictly associated with *Astarte*+*Cardita*. Not only the nodal support is strong, this relationship is also present across almost all trees and models. It has to be noted that the association between *Cuspidaria* and (*Astarte*+*Cardita*) has been evidenced already (Giribet and Distel, 2003). On the other side, suborder Pholadomyina is always basal to pteriomorphians (data not shown). Maybe it is worth noting that *Cuspidaria* branch is the longest among anomalodesmatans and that *Astarte* and *Cardita* branches are the longest among heterodonts (see Fig. 2.2). Moreover, this clade is somewhat unstable across bootstrap replicates (see Tab. 2.3). Maybe the large amount of mutations may overwhelm the true phylogenetic signal for such deep nodes, as also expected by their relatively high mutation rates. Hence, we see three possible alternatives: *i)* an artifact due to long-branch-attraction – all anomalodesmatans belong to Pteriomorphia, whereas *Astarte* and *Cardita* belong to Heterodonta; *ii)* anomalodesmatans do belong to Heterodonta, whose deeper nodes are not so good resolved, whereas a strong signal is present for Pteriomorphia monophyly, thus leading to some shuffling into basal positions; *iii)* anomalodesmatans are polyphyletic, and the two present-date suborders do not share a common ancestor. The two last possibilities seem unlikely to us, given our data and a considerable body of knowledge on the monophyletic status of Heterodonta and Anomalodesmata (Canapa et al., 2001; Dreyer et al., 2003; Harper et al., 2006; Taylor et al., 2007). We therefore prefer the first hypothesis, albeit an anomalodesmatan clade nested within heterodonts has also been appraised by some authors (Giribet and Wheeler, 2002; Giribet and Distel, 2003; Bieler and Mikkelsen, 2006; Harper et al., 2006). Interestingly, in *t10* tree the whole group

Cuspidaria+(*Astarte*+*Cardita*) nested within pteriomorphians species; a similar result was also yielded by a wider single-gene *cox1* dataset (data not shown). This would also account for the great difference found in *Astarte*+*Cardita* split across bootstrap replicates. A major taxonomical revision is needed for basal pteriomorphians, including also anomalodesmatans, as well as for superfamilies Astartoidea and Carditoidea.

As mentioned above, the main groups of pteriomorphians, arising in the late Cambrian, comprehend the genus *Nuculana* also. This placement was first proposed by Giribet and Wheeler (2002) on molecular bases and our data strongly support it. Its clade must have diverged from other main pteriomorphian groups at the very beginning of this large radiation. Among the main groups of Pteriomorphia, it is also worth noting the breakdown of the orders Pterioida *sensu* Vokes (1980) and Ostreoida *sensu* Millard (2001): the suborder Ostreina constitutes a net polyphyly with suborder Pectinina. The former is better related with order Pteriida *sensu* Millard (2001) (*Pinna*, *Pinctada*), whereas the latter is better related with superfamilies Limoidea (*Lima*+*Acesta*) and Anomioidea (*Anomia*). This is in agreement with most recent scientific literature about Pteriomorphia (Steiner and Hammer, 2000; Matsumoto, 2003).

Heterodonta Newell. – The subclass seems to have originated almost 500 Mya (late Cambrian) and its monophyletic status is strongly confirmed by our analysis, but a major revision of its main subdivisions is also required. The placement of *Astarte* and *Cardita* has already been discussed. At the same time, the orders Myoida and Veneroida, as well as the Chamida *sensu* Millard (2001), are no longer sustainable. A first main split separates (*Hiatella*+Cardiidae) from all remaining heterodonts. This split may correspond to two main orders in the subclass. As we sampled only 15 specimens of Heterodonta, we could only coarsely assess their phylogenetic taxonomy. However, we could precisely demonstrate the monophyly of families Veneridae and Mactridae and their sister group status. This could correspond together with *Dreissena*+*Mya* to a superfamily Veneroidea *sensu novo*,

which is stably dated around the early Devonian; however, further analyses are requested towards an affordable taxonomical revision, which is beyond the aims of this paper. Finally, recent findings about Tridacninae subfamily within Cardiidae family (Maruyama et al., 1998) are confirmed against old taxonomy based on Cardioidea and Tridacnoidea superfamilies (Millard, 2001).

Concluding, our work evidenced that all main deep events in bivalve radiation took place in a relatively short 70 Myr time during late Cambrian/early Ordovician (Fig. 2.3). Dates are stable across bootstrap replicates, especially those of deeper nodes, which were one of the main goals of this work (Tab. 2.3): most NPRS bootstrap means are indeed very close to PL estimates and standard deviations are generally low. Notable exceptions are some more recent splits on long branches (*Chlamys livida+Mimachlamys*, *Ensis+Sinonovacula*, *Astarte+Cardita*, *Tridacna*), which clearly are all artifacts of low taxon sampling for that specific branch, and Unionidae and Ostreoida. Unionidae are the only palaeoheterodonts we sampled and this could account for this anomaly; anyway, it is worth taking into account that the r8s-bootkit follows a slightly different method than *tout court* PL, therefore the results are not expected to perfectly coincide. When this happens, however, i.e. for most nodes in Figure 2.3, it accounts for a substantial stability in timing estimates.

We show in Figure 2.4 the survey on bivalve taxonomy which we described above. Given the still limited, but statistically representative, taxon sampling available, it is nowadays inconceivable to propose a rigorous taxonomy at order and superfamily level; therefore, we used in Figure 2.4 the nomenclature of Millard (2001) and Vokes (1980). More taxa and genes to be included will sharp resolution and increase knowledge on bivalves' evolutionary history.

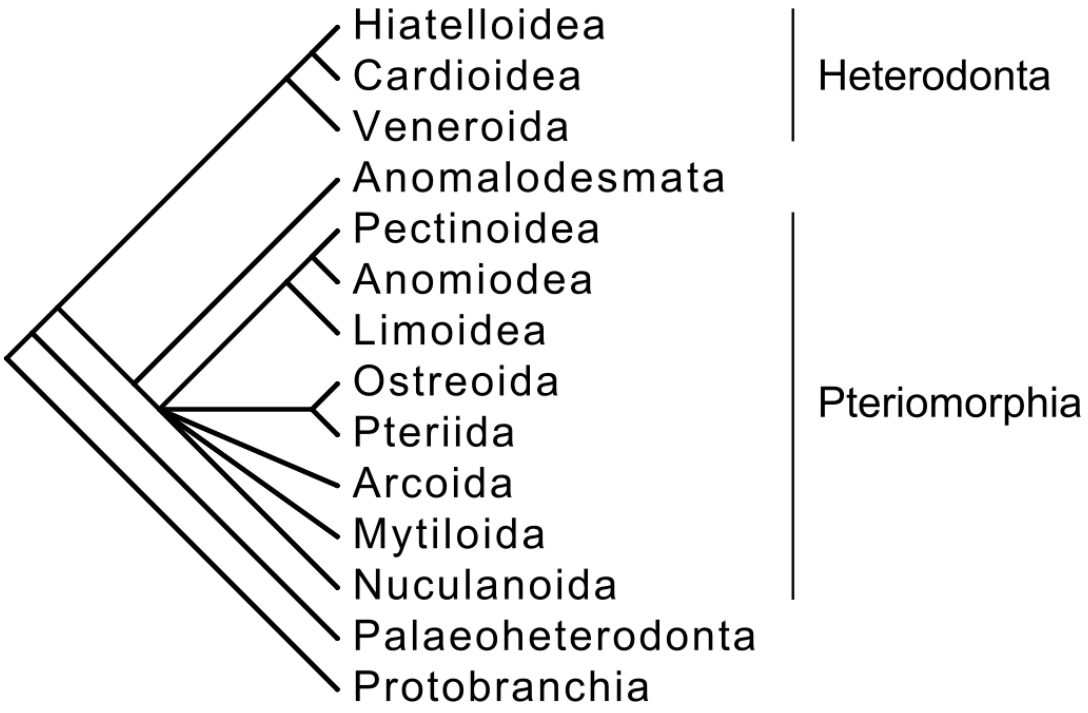


Figure 2.4. Global survey of the bivalve phylogeny.

CHAPTER 3

PHYLOGENETIC REPRESENTATIVENESS: A NEW METHOD FOR EVALUATING TAXON SAMPLING IN EVOLUTIONARY STUDIES

3.1. BACKGROUND

The study of phylogenetics has a long tradition in evolutionary biology and countless statistical, mathematical, and bioinformatic approaches have been developed to deal with the increasing amount of available data. The different statistical and computational methods reflect different ways of thinking about the phylogeny itself, but the issue of "how to treat data" has often overshadowed another question, i.e., "where to collect data from?". We are not talking about the various types of phylogenetic information, such as molecular or morphological characters, but rather we refer to which samples should be analyzed.

In phylogenetic studies, investigators generally analyze subsets of species. For example, a few species are chosen to represent a family or another high-level taxon, or a few individuals to represent a low-level taxon, such as a genus or a section. As a general practice, choices are driven by expertise and knowledge about the group; key species and taxa of interest are determined and, possibly, sampled. For example, if a biologist is choosing a group of species to represent a given class, species from many different orders and families will be included. We term the degree to which this occurs the "phylogenetic representativeness" of a given sample.

This issue is rarely formally addressed and generally treated in a rather subjective way; nevertheless, this is one of the most frequent ways incongruent phylogenetic results are accounted for. It is sufficient to browse an evolutionary biology journal to see how often

incorrect or biased taxon sampling is hypothesized to be the cause (e.g., Ilves and Taylor, 2009; Jenner et al., 2009; Palero et al., 2009; Ruiz et al., 2009; Tsui et al., 2009; Whitehead, 2009). We therefore aim to set up a rigorous taxon sampling method, which can be used alongside expertise-driven choices. Many theoretical approaches have been proposed to drive taxon sampling: see (Hillis, 1998; and reference therein) for a keystone review.

The concept of “taxonomic distinctness” was developed in the early 1990s among conservation biologists (May, 1990; Vane-Wright, 1991), who needed to measure biodiversity within a given site or sample so to assess further actions and researches. Basic measures of biodiversity take into account species richness and relative abundance (Whittaker, 1972; Peet, 1974; Taylor, 1978; Bond, 1989). However, it is clear from a conservationist point of view that not all species should be weighted the same. The presence and relative abundance of a species cannot capture all information on the variation of a given sample, and therefore a taxonomic component must also be considered in evaluating the biodiversity of a given site. This allows more realistic specification of the importance of a species in a given assemblage.

Similarly, resources for conservation biology are limited, and therefore it is important to focus on key species and ecosystems according to a formal criterion. For this purpose, several methods have recently been proposed (Ricotta and Avena, 2003; Pardi and Goldman, 2005; Pardi and Goldman, 2007; Bordewich et al., 2008). Despite recent progresses in sequencing techniques, it is still worth following a criterion of “maximizing representativeness” to best concentrate on key taxa (e.g., Bordewich et al., 2008). Nevertheless, this typically requires a well established phylogeny, or at least a genetic distance matrix, as a benchmark. These data are indeed generally available for model species or taxa with key ecological roles, but they are often unavailable in standard phylogenetic analyses. Typically, if we want to investigate a phylogeny, it has either never

been resolved before, or it has not been completely assessed at the moment we start the analysis. Further, if a reliable and widely accepted phylogenetic hypothesis were available for the studied group, we probably would not even try to attempt to formulate one at all. This means that, while the above-mentioned methods may be useful in the case of well-characterized groups, an approach using taxonomic distinctness is more powerful in general phylogenetic practice.

Our basic idea is that estimating the phylogenetic representativeness of a given sample is not conceptually different from estimating its taxonomic distinctness. A certain degree of taxonomic distinctness is required for individual samples chosen for phylogenetic analyses; again, investigators attempt to spread sampling as widely as possible over the group on which they are focusing in order to maximize the representativeness of their study. A computable measure of taxonomic distinctness is required to describe this sampling breadth.

In this article we propose a measure of phylogenetic representativeness, and we provide the software to implement it. The procedure has the great advantage of requiring only limited taxonomical knowledge, as is typically available in new phylogenetic works.

3.2. RESULTS

Algorithm

Clarke and Warwick (1999) suggest standardizing the step lengths in a taxonomic tree structure by setting the longest path (i.e., two species connected at the highest possible level of the tree) to an arbitrary number. Generally, this number is 100. Step lengths can be weighted all the same, making the standardized length measure to equal

$$l_n = \frac{100}{2(T-1)}$$

where *T* is the number of taxonomic levels considered in the tree and *n = 1*, *2*, *…* , *N*, where *N* is the number of steps connecting a pair of taxa (see Methods). All taxa in the tree belong by definition to the same uppermost taxon. Therefore, two taxa can be connected by a maximum of *2(T - 1)* steps.

However, it is also possible to set step lengths proportionally to the loss of biodiversity between two consecutive hierarchical levels, i.e., the decrease in the number of taxa contained in each one, as measured on the master list. Branch lengths are then computed as follows: we indicate $S_{(t)}$ as the number of taxa of rank *t*, with *t = 1*, *2*, … , *T* from the lowest to the highest taxonomic level. Two cases are trivial: when *t = 1*, $S_{(t)}$ equals to *S* (the number of Operational Taxonomic Units – OTUs – in the master taxonomic tree); when *t = T*, $S_{(t)}$ equals to *1* (all taxa belong to the uppermost level). The loss of biodiversity from level *t* to level *t + 1* is:

$$\Delta S_{(t)} = S_{(t)} - S_{(t+1)} \quad (1 \le t \le T-1)$$

The step length from level *t + 1* to level *t* is the same as from level *t* to level *t + 1*. Therefore, path lengths are then obtained as:

$$l_t = l_{t^*} = \frac{\frac{\Delta S_{(t)}}{\sum_{t=1}^{T-1} \Delta S_{(t)}} \times 100}{2} = \frac{\Delta S_{(t)}}{\sum_{t=1}^{T-1} \Delta S_{(t)}} \times 50 \ , \qquad t^* = N - t + 1$$

where l_t is the path length from level t to level $t + 1$ and l_{t^*} is the reverse path length.

Clarke and Warwick (1999) found the method of weighting step lengths to have little effect on final results. However, we find that standardizing path lengths improves the method in that it also complements subjectivity in taxonomies; rankings are often unrelated even across closely-related groups. To us, this is the main reason for standardizing path lengths. Moreover, adding a level in a taxonomic tree does not lead to changes in the mean or standard deviation of taxonomic distance (AvTD or VarTD) if we adopt this strategy. In addition, the insertion of a redundant subdivision cannot alter the values of the indices (Clarke and Warwick, 1999). All these analyses are carried out by our PhyRe script (downloadable at www.mozoolab.net/downloads).

Our method based on Clarke and Warwick's ecological indices has the main feature of being dependent only upon a known existing taxonomy. This leads to a key difficulty: taxonomic structures are largely subjective constructions. Nonetheless, we think that taxonomists' expertise has provided high stability to main biological classifications, at least for commonly-studied organisms, such as animals and plants. The degree of agreement which is now reached in those fields allows us to consider most systematics as stable. In our view, large-scale rearrangements are becoming more and more unlikely, so that this argument leads us to state that present taxonomies do constitute an affordable starting point for methods of phylogenetic representativeness assessment.

However, this is not sufficient to completely ensure the reliability of our method. Knowledge is growing in all fields of evolutionary biology, and the increase in data results in constant refinement of established classifications. In fact, even if large-scale changes are rare, taxonomies are frequently revised, updated, or improved. Therefore, we

implemented an algorithm that allows for testing the stability of the chosen reference taxonomy.

Essentially, our procedure can be described in two phases. In the first one, the shuffling phase, master lists are shuffled, resulting in a large number of alternative master lists. In the second, the analysis phase, a phylogenetic representativeness analysis is carried out as described above across all simulated master lists rearrangements. The shuffling phase is composed of three moves, which are repeated and combined *ad libitum* (see Methods). These moves simulate the commonest operations taxonomists do when reviewing a classification. A large number of “reviewed” master lists is then produced, repeating each time the same numbers of moves. Finally, the shuffling phase ends with a set of master lists. Standard phylogenetic representativeness analyses are performed on each master list, and all statistics are computed for each list. In this way, a set of measurements is produced for each indicator. Therefore, it is possible to compute standard 95% (two-tailed) confidence intervals for each one. This analysis phase gives an idea of the funnel plot's oscillation width upon revision. PhyloSample and PhyloAnalysis are specific scripts dealing with the shuffling analysis: the former generates the new set of master list, whereas the latter performs PhyRe operations across them all.

All scripts are available online, and a Windows executable version of the main script is also present: the software can be downloaded from the MoZoo Lab web site at http://www.mozoolab.net/index.php/software-download.html.

Testing

In order to evaluate the method, we analyze phylogenies of bivalves (Passamaneck et al., 2004), carnivores (Flynn et al., 2005), coleoids (Strugnell et al., 2005), and termites (Legendre et al., 2008). Our reference taxonomies are Millard (2001) for mollusks, the Termites of the World list hosted at the University of Toronto

(http://www.utoronto.ca/forest/termite/speclist.htm: consulted on 03/23/2009 and reference therein), and the online Checklist of the Mammals of the World compiled by Robert B. Hole, Jr. (http://www.interaktv.com/MAMMALS/Mamtitl.html: consulted on 03/11/2009 and reference therein).

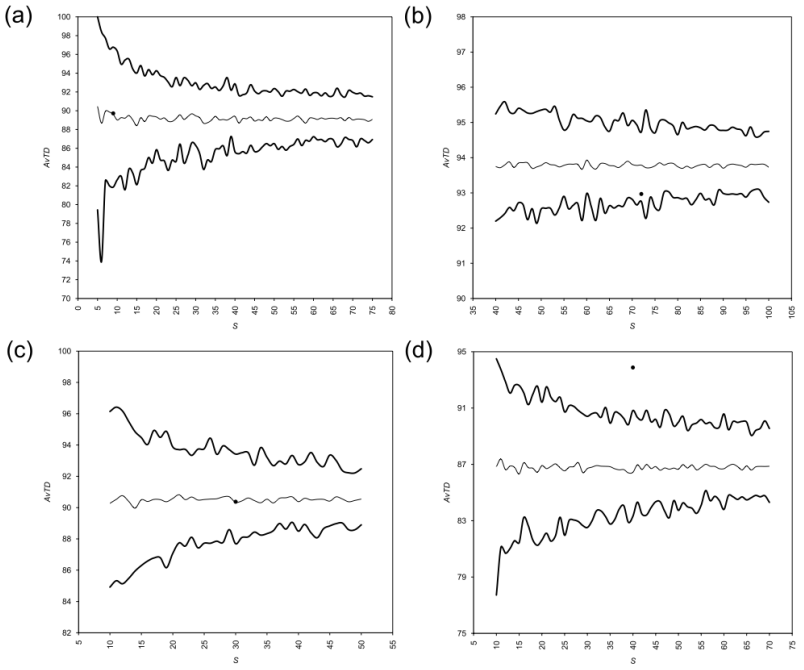


Figure 3.1. Funnel plots of Average Taxonomic Distinctness (AvTD) from (a) bivalves (Passamaneck et al., 2004), (b) carnivores (Flynn et al., 2005), (c) coleoids (Strugnell et al., 2005), and (d) termites (Legendre et al., 2008) data sets are shown. Results are from 100 random replicates. Thick lines are the highest values found across all replicates of each dimension and the lower 95% confidence limit; the thin line is the mean across all replicates; experimental samples are shown by black dots.

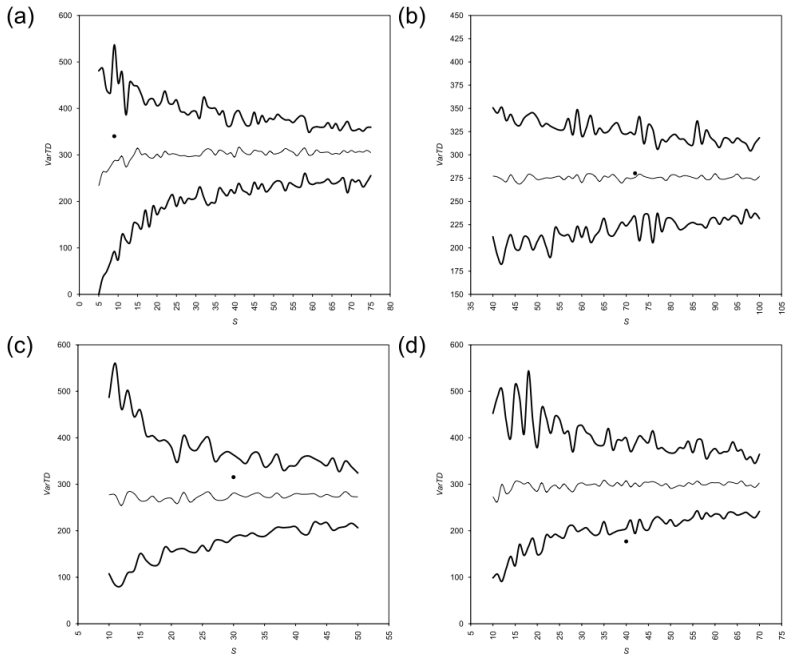


Figure 3.2. Funnel plots of Variation in Taxonomic Distinctness (VarTD) from (a) bivalves (Passamaneck et al., 2004), (b) carnivores (Flynn et al., 2005), (c) coleoids (Strugnell et al., 2005), and (d) termites (Legendre et al., 2008) data sets are shown. Results are from 100 random replicates. Thick lines are the upper 95% confidence limit and the lowest values found across all replicates of each dimension; the thin line is the mean across all replicates; experimental samples are shown by black dots. The bias towards lower values for small sample is detectable in mean.

Results from AvTD and VarTD are shown in Figures 3.1 and 3.2, respectively. Funnel plot are based arbitrarily on 100 random samplings from the master list for each sample size. Table 3.1 summarizes these results, showing also results from I_E.

Table 3.1 - Phylogenetic Representativeness analyses from four published works.

Group	Reference	Dimension	AvTD	VarTD	I_E
Bivalves	Passamaneck et al., 2004	9	89.7181	340.1874	0.0609
Carnivores	Flynn et al., 2005	72	92.9688	280.2311	0.1203
Coleoids	Strugnell et al., 2005	30	90.3758	315.3069	0.1079
Termites	Legendre et al., 2008	40	93.8788	177.1053	0.1631

Dimension, number of taxa; AvTD, Average Taxonomic Distinctness; VarTD, Variation in Taxonomic Distinctness; I_E, von Euler's (2001) Index of Imbalance.

To assess the stability of our taxonomies by performing shuffling analyses on them, we fixed the amount of "moves" to be executed according to our knowledge of each master list (see Discussion for details; Table 3.2); 1,000 new "reviewed" datasets were generated and then 100 replicates were again extracted from each master list for each

sample size. Funnel plots for AvTD and VarTD are shown in Figures 3.3 and 3.4, respectively.

We conducted additional analyses on the dataset of bivalves with real and simulated data (Appendix 3.1).

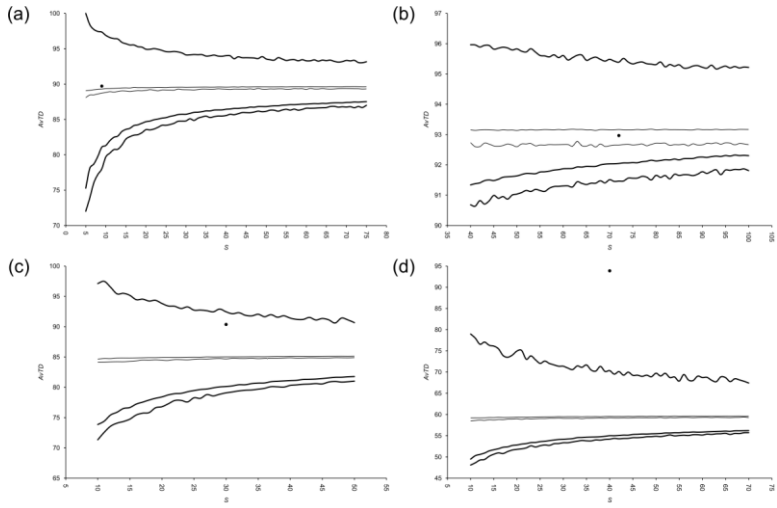


Figure 3.3. Funnel plots of Average Taxonomic Distinctness (AvTD) upon master lists' shuffling from (a) bivalves (Passamaneck et al., 2004), (b) carnivores (Flynn et al., 2005), (c) coleoids (Strugnell et al., 2005), and (d) termites (Legendre et al., 2008) data sets are shown. Results are from 1,000 shuffled master lists and 100 random replicates. Thick lines are the highest values found across all replicates and the lower 95% confidence limit (2.5% and 97.5% confidence limits); thin lines represent the mean across all replicates (2.5% and 97.5% confidence limits); experimental samples are shown by black dots. Shuffling tuning as in Table 3.2.

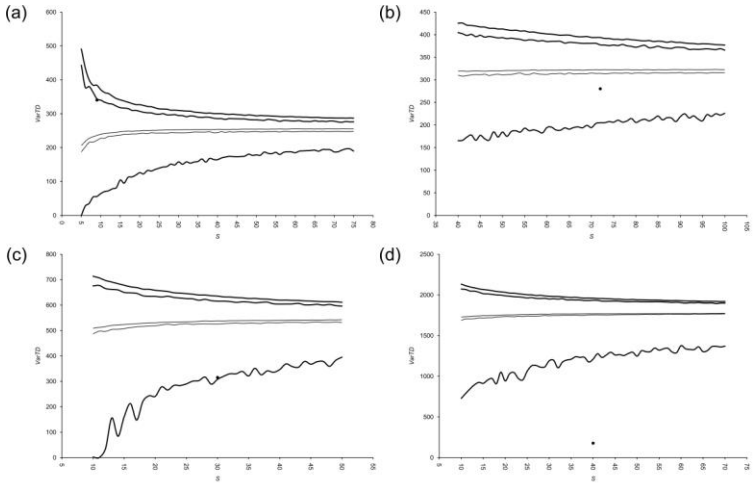


Figure 3.4. Funnel plots of Variation in Taxonomic Distinctness (VarTD) upon master lists' shuffling from (a) bivalves (Passamaneck et al., 2004), (b) carnivores (Flynn et al., 2005), (c) coleoids (Strugnell et al., 2005), and (d) termites (Legendre et al., 2008) data sets are shown. Results are from 1,000 shuffled master lists and 100 random replicates. Thick lines are the upper 95% confidence limit (2.5% and 97.5% confidence limits) and the lowest values found across all replicates; thin lines represent the mean across all replicates (2.5% and 97.5% confidence limits); experimental samples are shown by black dots. Shuffling tuning as in Table 3.2.

Table 3.2 - Shuffling moves performed on each master list

Group	Size	Level	Splits	Merges	Transfers
Bivalves	3404	Family	15	10	40
Carnivores	271	subfamily	2	1	2
Coleoids	220	Family	2	1	2
Termites	2760	species	0	0	15

Each set of splits, merges, and transfers was repeated independently 1,000 times on the relative master list. Moves were applied to the specified taxonomic level. Master list's size is reported to inform about the entity of the "reviewing" shuffle. Size in Operational Taxonomic Units (OTUs) of the global taxonomic tree.

Data from bivalve phylogenies obtained in our laboratory at different times from different samples have been tested along with imaginary samples of different known representativeness. We use the letter R to denote real data sets analyzed in our laboratory. Datasets from R1 to R4 are increasingly representative. In R1, the subclass of Protobranchia is represented by just one genus, and the subclass of Anomalodesmata is completely missing. In R2, we add one more genus to Protobranchia (*Solemya*) and one genus to Anomalodesmata (*Thracia*). In R3, the sample is expanded with several Genera from Unionidae (*Anodonta*, *Hyriopsis*), Heterodonta (*Gemma*, *Mactra*), Protobranchia (*Nuculana*; but see Giribet and Wheeler, 2002; Giribet and Distel, 2003), and more Anomalodesmata (*Pandora*, *Cuspidaria*). While all high-level taxa were already represented in R2, R3 is thus wider and more balanced in terms of sampling. R4 is identical to R3 with the exception of genus *Cerastoderma*, which was excluded due to technical problems.

Simulated data sets are indicated by the letter S. S1 is an "ideal" data set: all subclasses are represented with 4 species and 4 families, although the number of represented orders is different across the subclasses. S2 is biased towards less biodiversity-rich subclasses: it comprehends 6 anomalodesmatans, 6 palaeoheterodonts, and 7 protobranchs, along with only 1 pteriomorphian and one heterodont. S3 is strongly biased towards heterodonts, with 17 genera. Pteriomorphians, palaeoheterodonts, and protobranchs are represented by one genus each, and there are no anomalodesmatans here. S4 is an "easy-to-get" sample, with the commonest and well-known genera (e.g.,

Donax, *Chamelea*, *Teredo*, *Mytilus*, *Ostrea*), and therefore it is composed only by pteriomorphians (7 genera) and heterodonts (11 genera).

For this entire group of samples, from R1 to R4, and from S1 to S4, we conducted phylogenetic representativeness analyses to find out whether the method can describe samples following our expectations. Funnel plots were constructed on 10,000 replicates. Results are displayed in Figure 3.5 and Table 3.3.

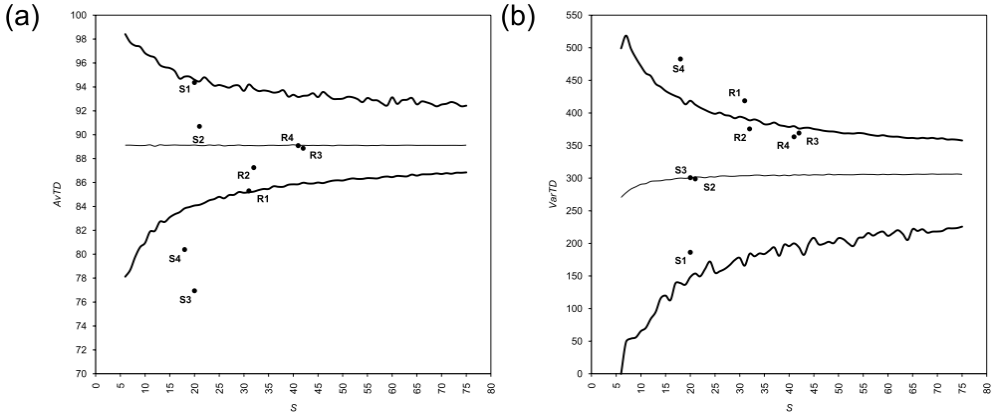


Figure 3.5. Phylogenetic Representativeness as measured by funnel plots of (a) Average Taxonomic Distinctness (AvTD) and (b) Variation in Taxonomic Distinctness (VarTD) from bivalves' master list (Millard, 2001). Results are from 10,000 random replicates. Lines are as in Figure 3.1 and 3.2 for (a) and (b), respectively. Letter S denotes simulated data sets, whereas letter R denotes real ones. See text for explanation.

Table 3.3. Phylogenetic Representativeness across real and simulated bivalve data sets.

Sample	Group	Dimension	AvTD	VarTD	I_E
real					
R1	without anomalodesmatans	31	85.3003	418.7537	0.2586
R2	+ *Solemya* and *Thracia*	32	87.2497	375.5878	0.2804
R3	increased (see text)	42	88.8653	369.2571	0.1806
R4	– *Cerastoderma*	41	89.0842	363.4391	0.1773
simulated					
S1	"ideal" (see text)	20	94.3673	186.2882	0.0476
S2	biased towards poor subclasses	21	90.6962	298.9607	0.1676
S3	biased towards heterodonts	20	76.9450	300.7505	0.7017
S4	"easy-to-get" (see text)	18	80.3913	482.7998	0.2419

Dimension, number of taxa; AvTD, Average Taxonomic Distinctness; VarTD, Variation in Taxonomic Distinctness; I_E, von Euler's (2001) Index of Imbalance.

Implementation

The distribution of AvTD from *k* random subsamples of size *S* is typically left-skewed (Clarke and Warwick 1998; Figure 3.6). This is not an effect of a low *k*, as increasing the number of subsamples the shape of distribution does not change.

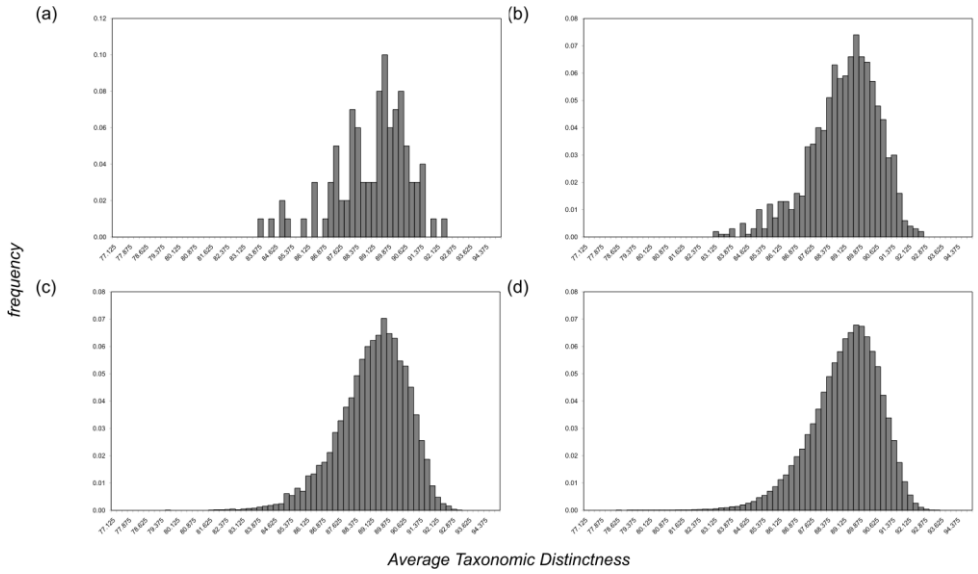


Figure 3.6. Histograms show frequencies of Average Taxonomic Distinctness (AvTD) values among k = 100 (a), 1,000 (b), 10,000 (c), and 100,000 (d) random subsamples (S = 50) from bivalves' master list by Millard (2001). The distribution shows a skeweness towards the left side.

We follow Azzalini (1985) in describing the skeweness with a parameter l. The further is l (as absolute value) from unity, the more skewed is the distribution. Using the master list of bivalves and a dimension *S* of 50, we estimated an absolute value for l which is very close to unity (~1.01, data not shown), confirming that the distribution only slightly differs from the normal one. However, this was done only for one sample, and distributions vary across different taxonomies and organisms. Similar considerations can be applied to VarTD.

We represent in our AvTD plots the lower 95% confidence limit (see figures from 3.1 to 3.5). The maximum value obtained across all replicates for that dimension is also shown because it converges to the upper absolute limit as *k* increases. Conversely, in VarTD plots the upper 95% confidence limit and minimum observed value are shown, as lower values of variation are preferable (see Methods). PhyRe produces funnel plots showing results from a range of dimensions *S.* This helps in evaluating the global situation and is very useful for comparing homogeneous samples of different sizes.

For the shuffling analysis, similar funnel plots are produced. The main difference is that for AvTD the lower 95% confidence limit is not a line: here is shown the area which comprises 95% of values for each dimension across all shuffled master lists. The same applies for the AvTD and VarTD means, and the VarTD upper 95% confidence limit.

Output from PhyRe can easily be imported into a graph editing software like Microsoft Excel®.

3.3. DISCUSSION

"Taxon sampling" is not a new topic by itself and several strategies have been proposed from different standpoints. As mentioned above, several criteria have been appraised, especially when an established phylogeny is present. Long-branch subdivision (Handy and Penny, 1989; Poe, 2003), for example, has been proposed as one strategy; see Hillis (1998) for more strategies. Much experimental interest has been focused also on outgroup sampling (see, e.g., Giribet and Carranza, 1999; Puslednik and Serb, 2008, for empirical studies) and its effects. Finally, whether it is preferable to add more characters or more taxa is a vexing question; several authors highlight the importance of adding new taxa to analyses (e.g., Pollock, 2000; Hedtke et al., 2006). However, Rokas and Carroll (2005) point out that an increase in taxon sampling does not have an improving effect *per se*. Nevertheless, they suggest several factors which may influence the accuracy of phylogenetic reconstructions, and among them the density of taxon sampling.

Rannala et al. (1998) obtained more accurate phylogenetic reconstructions when they sampled 20 taxa out of 200, rather than when 200 taxa out of 200,000 were chosen for analyses, although in the latter case the taxon number was higher. This is rather intuitive, indeed, as taxon sampling is denser in the former case. Each taxon was sampled with the same probability r in a birth-death process (see Rannala et al., 1998, for further details). Interestingly, this is somewhat similar to our random subsampling process: the more dense is a sample, the more likely is it to be representative of its master list, despite the absolute number of included taxa.

However, our approach is very different, because it is completely *a priori*. The method can always be applied to any phylogeny, given the presence of a reference taxonomy and a master list of taxa. We find useful to start from the zero point of no phylogenetic information except for the available taxonomy. Evolutionary systematics does

indeed capture some phylogenetic information, because all taxonomic categories should correspond to monophyletic clades. We employ this preliminary phylogenetic information to assess taxon sampling (but see below for further discussion on this point).

This method can be applied to every kind of analysis, from molecular to morphological ones. Furthermore, even extinct taxa can be included in a master list or in a sample: for example, the bivalve list from Millard (2001) does report fossil taxa, and we left those taxa in our reference master list, as these are part of the biodiversity of the class. In fact, a good sample aims to capture the entire diversity of the group, thus including extinct forms. Therefore, we suggest that molecular samples should be better compared to complete master lists, which comprehend both living and fossil taxa (see Figure 3.5).

Moreover, evaluating phylogenetic representativeness as described here has the great advantage of being largely size-independent: this is well shown by funnel plots of AvTD and VarTD (figures from 3.1 to 3.5). The mean is consistent across all dimensions *S* and it is very close to AvTD or VarTD values obtained from the whole master list (data not shown; see e.g., Clarke and Warwick, 1998). This fact, along with setting path lengths proportionally to biodiversity losses and rescaling their sum to 100, has a very useful and important effect: adding new taxa or new taxonomic levels does not change any parameter in the analysis. This means that more and more refined analyses can always be addressed and compared with coarser ones and with results from other data.

Most importantly, we checked the significance of both AvTD and VarTD results with one-tailed tests. The original test was two-tailed (Clarke and Warwick, 1998), and this is the greatest difference between the original test and our implementation for phylogenetic purposes. In the ecological context, these indices are used to assess environmental situations, to test for ecological stresses or pollution. In such a framework, the index must point out assemblages which are either very poor or very rich in terms of distinctness. The former will constitute signals of critically degraded habitats, whereas the latter will indicate

a pristine and particularly healthy locality, and ecologists seek explanations for both results.

In our applications, we want our sample to be representative of the studied group, so that a sample significantly higher in taxonomic distinctness than a random one of the same size can be very useful; indeed, it would be even preferred. For this reason, we state that a one-tailed test is more appropriate for our purposes.

All case studies rely on samples with good phylogenetic representativeness. Nevertheless, one sample (Passamaneck et al., 2004; Figure 3.1a and 3.2a) is relatively small to represent its master list; this is shown by quite large funnels at its size. On the other hand, one sample (Legendre et al., 2008; Figure 3.1d and 3.2d) turned out to be strikingly representative of its groups: the AvTD is higher (and the VarTD lower) than the highest (lowest) found in 100 random subsamples. We recommend the former sample be taken with care for phylogenetic inferences (in fact, see Passamaneck et al., 2004, on the polyphyly of bivalves). Conversely, the latter sample is extremely more representative than the other three. Highly representative samples are readily individuated by AvTD and VarTD funnel plots (see Figure 3.1d and 3.2d) as dots above the highest AvTD and below the lowest VarTD found across all random replicates.

This is naturally influenced by the number of such subsamples: the more subsamples that are drawn, the more likely is to find the absolute maximum (minimum) possible value. If *k* is sufficiently high, the absolute maximum (minimum) possible value is found for any dimension *S*, and no sample can appear above (below) those lines (see Figure 3.5). Therefore, we suggest to draw an intermediate number of replicates (e.g., 100 or 1,000) to avoid this widening effect and identify more optimal phylogenetic samples.

Shuffling analysis assesses the complex issue of master list subjectivity and, as such, taxonomy itself. Master lists turn out to be substantially stable upon simulated revision, as shown in Figure 3.3 and 3.4. 95% confidence areas are indeed generally

narrow and the position of experimental dots is never seriously challenged. We used 100 replicates from 1,000 master lists: this turned out to be sufficient to draw clear graphs, where borders are accurately traced.

An objective criterion to describe the amount of shuffling needed for this analysis is still lacking; however, each group of living beings has its own taxonomic history and its own open problems, therefore we think it can be very difficult to find an always-optimal criterion. An expertise-driven choice cannot be ruled out here. We suggest that, given the contingent conditions of a study, phylogeneticists choose the best degree of shuffling to describe their master list's stability. Some taxonomical situations are much more consolidated than others; in some cases higher-level taxa are well-established, whereas in others agreement has been reached on lower-level ones. A formal criterion, like moving 10% of species or merging 5% of genera, will necessarily lose this faceting and complexity.

Interestingly, the coleoid master list revealed itself to be the most sensitive to shuffling. The AvTD funnel plot places the sample of Strugnell et al. (2005) exactly across the mean line, whereas it is close to the maximum line in the shuffling analysis (see Figure 3.1c and 3.3c). This means that AvTD is globally lowered upon shuffling on the coleoid master list. In fact, whereas mean AvTD on the original master list was close to 90 for all *S*, the 95% confidence interval on shuffled master lists is always slightly under 85. Conversely, VarTD is over the mean in standard PhyRe computations, whereas it is across the minimum line in shuffling analysis (see Figure 3.2c and 3.4c): VarTD mean changes from about 300 in the former case to around 500 in the latter one. The amount of shuffling we applied (see Table 3.2) is evidently heavy in this case. Therefore, upon a taxonomic review, we would recommend to reconsider this sample and to perform a new phylogenetic representativeness analyses.

Our method is also descriptive for comparing similar samples; this is a smart way to test the improvement of a phylogenetic study while adding one or more taxa to a given sample. It is clear from our R1-R4 example (see Figure 3.5) the importance of adding just two taxa to the initial sample. The improvement is well depicted by AvTD and VarTD funnel plots: whereas R1 is just across the AvTD lower 95% confidence limit of AvTD, R2 is well above; whereas R1 is outside the VarTD upper 95% confidence limit, R2 is inside it. While VarTD remains close to the confidence limit, R3 and R4 are nevertheless even more representative in terms of AvTD, as they lie precisely on the mean of 10,000 replicates. This reflects the increase of sampled taxa with respect to several under-represented groups.

S1, the “ideal” sample, turns out to have the highest AvTD (across the maximum line) and the lowest VarTD (next to the minimum line). In this case, we have 10,000 replicates; thus, the above considerations hold true and we do not expect our dot to be neither above nor below the funnel plot for AvTD or VarTD, respectively. Sample S2, biased towards less biodiversity-rich subclasses appears to be representative: it is well inside both funnel plots. Three subclasses out of five are well represented here; this sample is therefore rather informative. However, it is clearly less preferable than sample S1; whereas the former lies always across or next to the mean line, the latter is always close to the observed extreme values. Sample S3 seems reasonable in terms of VarTD, but the AvTD funnel plot identifies it as the worst of all. Nevertheless, sample S4 (with two substantially equally-represented subclasses) turned out to be even worse than S3 (almost just one subclass included): it is below the 95% confidence limit of AvTD and above the 95% confidence limit of VarTD.

Thus, joint analysis of AvTD and VarTD provides discrimination between samples. An AvTD/VarTD plot shows that these measures are generally negatively correlated, even if

some exceptions are possible: good samples have high AvTD and low VarTD values; the opposite is true for bad samples (Figure 3.7).

Along with the two main measures, I_E can give an approximate idea of the shape of the tree. Values > 0.25 are often associated with biased samples (see Table 3.3), and thus we suggest this as a rule of thumb for directly discarding imbalanced ones. However, this cut-off value is only a rough guide in estimating phylogenetic representativeness: sample R2 has an I_E of 0.2804 (greater than R1), but funnel plots identify it as a good bivalve sample.

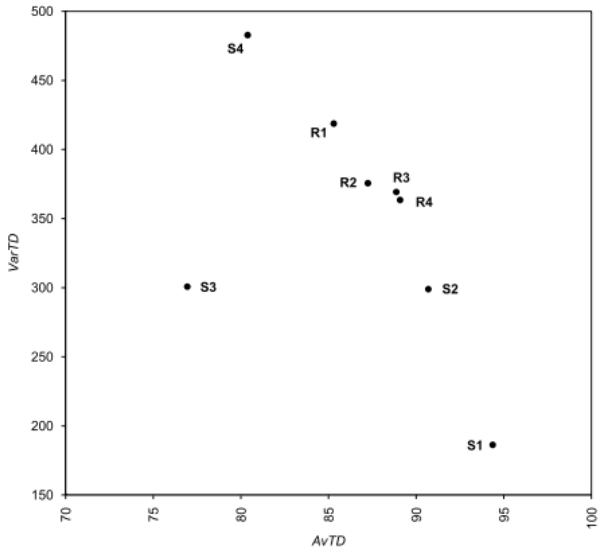


Figure 3.7. Variation in Taxonomic Distinctness (VarTD) plotted on Average Taxonomic Distinctness (AvTD) for real and simulated bivalve datasets (see Table 3.3 for further details on samples).

3.4. Conclusions

Phylogenetic representativeness analyses can be conducted at every taxonomic level, and including any taxonomic category. Moreover, inclusion or exclusion of taxonomic categories does not influence results across analyses (Clarke and Warwick, 1999; see above). Although we did not present it here, the index can also potentially take relative abundance data into account (see Warwick and Clarke, 1995, 1998; Clarke and Warwick, 1998). Thus, it may be implemented for population-level analyses as well, depicting sampling coverage among different populations from a given section, species, or subspecies.

The main strength of phylogenetic representativeness approach lies in being an *a priori* strategy of taxon selection and sampling. Therefore, it cannot take into account several empirical and experimental problems, which are not guaranteed to be avoided. For example, long-branch attraction depends essentially upon a particularly quick rate of evolution in single taxa (Felsenstein, 1978), which is only *a posteriori* identified. Moreover, topology alteration due to outgroup misspecification remains possible, as phylogenetic representativeness deals only with ingroup taxa.

Each particular study copes with specific difficulties strictly inherent to contingent conditions; for example, as a result of an unexpected selective pressure, one particular locus may turn out to be completely uninformative, even if the taxon sampling is perfectly adequate. Nevertheless, in R1-R4/S1-S4 examples (see above), our knowledge of bivalve evolution and systematics allows us to discriminate between suitable and non-suitable samples, and phylogenetic representativeness results matched perfectly with our expectations.

Moreover, being understood that expertise is always expected in planning taxon sampling, we strongly suggest to set phylogenetic representativeness alongside a formal

criterion for profiling phylogenetic informativeness of characters (e.g., Townsend, 2007). Put in other words, phylogenetic representativeness is a guarantee of a good and wise taxonomic coverage of the ingroup, but evidently it is not guarantee of a good and robust phylogeny *per se*. For this reason, we would suggest it as a springboard for every phylogenetic study, from which subsequent analyses can proceed further towards an affordable evolutionary tree.

3.5. METHODS

Average Taxonomic Distinctness (AvTD)

Mathematical aspects of this index are well explained in works by Warwick and Clarke (1995) and Clarke and Warwick (1998; 2001). However, it is useful to explain here the main points of their statistics.

AvTD is computed starting from a taxonomic tree. A taxonomic tree is merely the graphical representation of a Linnean classification, whereby OTUs are arranged hierarchically into different categories or taxa, with taxa being mutually exclusive. We use the general terms “OTUs” and “taxa” because a taxonomic tree does not necessarily include species at their tips, nor do all taxonomic trees take into account exactly the same levels of systematics.

A simple taxonomic tree is depicted in Figure 3.8. Each leaf is an OTU and each node is a taxon; for example, OTUs may correspond to species and deeper nodes to genera, families, and orders as we climb up the tree.

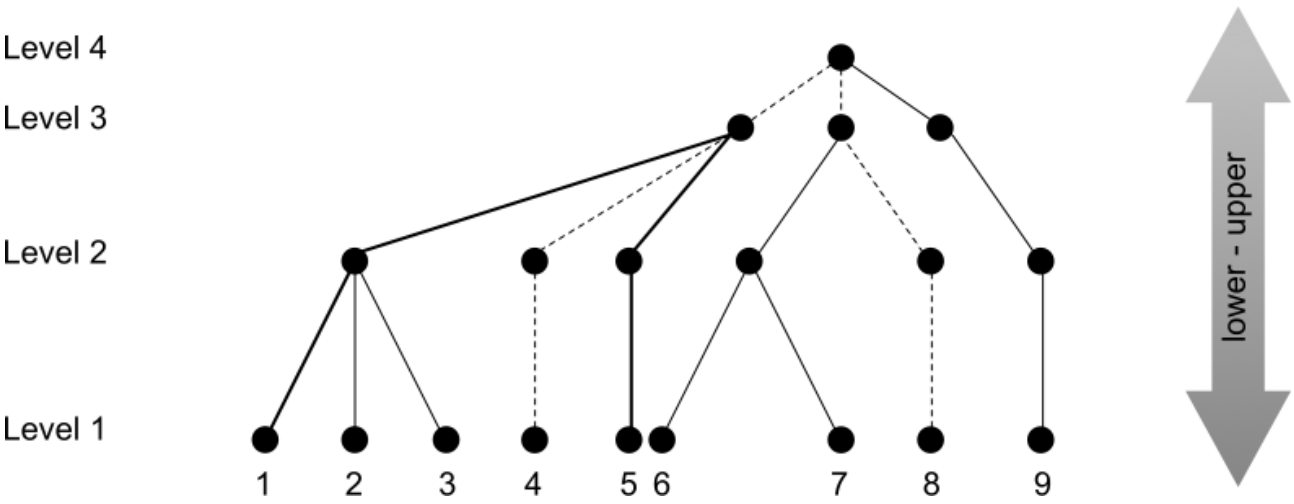


Figure 3.8. Nine Operational Taxonomic Units (OTUs) and four taxonomic levels are shown. For example, levels 1-4 could correspond to species, genera, families, and orders, respectively; in this case, species 1, 2, and 3 would belong to the same genus, species 1, 2, 3, and 4 to the same family, and so on. Taxonomic paths connecting taxa 1 and 5 (thick lines) and taxa 4 and 8 (dashed lines) are marked. See text for more details.

On a tree such as this, we can define a tree metric of taxonomic distance between any given pair of OTUs. A taxonomic tree is rooted (by definition); therefore, it is necessary to specify that our tree metric is unrooted (see Pardi and Goldman, 2007), i.e., the distance between two taxa is the shortest path on the tree that leads from one to another, and it is not required to climb up the tree from the first taxon to the root and then down to the second one, otherwise all pairs of OTUs would score the same distance.

Let us indicate with w_{ij} the taxonomic distance between OTUs *i* and *j*, which are joined by *N* steps (branches) on the tree. Now we can define:

$$\omega_{ij} = \sum_{n=1}^{N} l_n$$

where l_n is the length of the *n*th branch, *n = 1*, *2*, … , *N*. We do not want to rely on information about mutation rates nor genetic distances. If we consider that a Linnean classification is mostly arbitrary, we can set branch lengths in several ways. Further considerations on this point are given above (Results; but see also Clarke and Warwick, 1999).The simplest case is considering a length equal to 1 for all branches. Accordingly, the distance between taxa 1 and 5 in Figure 3.8 is 4, and the distance between taxa 4 and 8 is 6. Indeed, taxa 1 and 4 are more closely related than taxa 4 and 8 are. The Average Taxonomic Distinctness (AvTD) of the tree is defined as the average of all such pairwise distances:

$$AvTD = \frac{\sum_{i=1}^{S}\sum_{j>i}^{S} \omega_{ij}}{\frac{S(S-1)}{2}}$$ (modified from Clarke and Warwick, 1998)

where *S* is the number of taxa in the tree. Given the presence/absence data case, and with the distance between taxa *i* and *j*, being *i = j*, set to 0 (same taxon), we note that the formula can be reduced to the computationally simpler form:

$$AvTD = \frac{\sum_{i=1}^{S}\sum_{j=1}^{S}\omega_{ij}}{S(S-1)}$$

For example, the AvTD for the tree in Figure 3.8 would equal approximately 5.0556. The original formulation of the index considers also relative abundances of species, but here we only take into account presence/absence of OTUs.

This is the basic statistic described in this work. AvTD has been shown to be a good ecological indicator and a reliable estimator of biodiversity (Warwick and Clarke, 1998; Warwick and Light, 2002; Warwick and Turk, 2002; Leonard et al., 2006). The most appealing feature is its clear independence from sampling effort (Warwick and Clarke, 1995, 1998; see Discussion above).

Test of significance

The AvTD statistic simply gives the expected path length for a randomly selected pair of species from the set of *S* species (Clarke and Warwick, 1998). The higher the AvTD, the more taxonomically distinct is the sample. However, it is necessary to compare the AvTD of a sample to the master list from which it is taken; for example, we may be interested in the molecular phylogeny of an order and we sampled and sequenced *S* species within this order. Naturally, we wish to maximize the number of families and genera represented therein. Using the AvTD method, we can estimate this “maximization” by computing the index for our sample of *S* species, and then comparing it with one computed from the list of all species belonging to the order itself. However, comparing a pure number to another pure number is rather uninformative; therefore, a random resampling approach to test for significance is suggested here. The rationale is as follows: we must estimate whether our sample’s AvTD ($AvTD_S$) is significantly different from the master list’s one. Although the index is poorly dependent on sampling effort, we have to take into account that often the master list is consistently bigger than our sample. Thus, we draw *k* samples of size *S* from

master list. We then compute AvTD from all *k* sample and test whether $AvTD_S$ falls within the 95% confidence limits of the distribution (original two-tailed test; but see Discussion above).

Variation in Taxonomic Distinctness (VarTD)

As noted by Clarke and Warwick (2001), some differences in the structure of the taxonomic trees of samples are not fully resolved by AvTD measures. Two taxonomic trees could have very different structures, in terms of subdivision of taxa into upper-level categories, but nevertheless could have the same AvTD. Differences in taxonomic structures of samples are well described by a further index of biodiversity, the Variation in Taxonomic Distinctness (VarTD).

VarTD is computed as a standard statistical variance. It captures the distribution of taxa between levels, and should be added to AvTD in order to obtain a good measure of biodiversity. Clarke and Warwick (1998) demonstrated that VarTD can be estimated via a precise formula, but can also be obtained in the canonical statistical way from AvTD data.

Clarke and Warwick (2001) proposed to follow the same procedure as above: observed VarTD is compared with values from random resamplings of the same size. Lower values of VarTD are preferable, as they are an indication of equal subdivision of taxa among intermediate levels. Clarke and Warwick (2001) also show that VarTD is not as independent from sampling effort as AvTD is, i.e., there is a bias towards lower values for very small *S* (see Figure 3.2 and 3.4), but it can be shown (Clarke and Warwick, 2001) that this bias becomes rather negligible for $S > 10$.

Von Euler's index of imbalance

Following the idea of AvTD, von Euler (2001) proposed an index related to taxonomic distinctness, which he called an *index of imbalance*. An index of imbalance measures the

imbalance of the tree, i.e., whether and how much certain groups are under-represented and certain others are over-represented. This was not the first of such indexes (e.g., Colless, 1982; Shao and Sokal, 1990; Heard, 1992; Kirkpatrick and Slatkin, 1993); however, as noted by Mooers and Heard (1997), they do not apply to trees with polytomies, as taxonomic trees often are. Von Euler's index of imbalance (I_E) is defined as

$$I_E = \frac{AvTD_{max} - AvTD}{AvTD_{max} - AvTD_{min}}$$

where $AvTD_{max}$ and $AvTD_{min}$ are respectively the maximum and minimum possible AvTDs given a particular sample. $AvTD_{max}$ is obtained from a totally-balanced tree constructed on the given taxa, whereas $AvTD_{min}$ is obtained from a totally-imbalanced one.

Figure 3.9 depicts such trees as computed from the taxonomic tree shown in Figure 3.8; taxonomic levels are considered as orders, families, genera, and species. (i) *Obtaining a completely imbalanced tree.* The procedure is bottom-up. Each species is assigned to a different genus (left side, thick lines, species 1, 2, 3, 4, and 5), until the number of "occupied" genera equals the total number of genera minus one. Remaining species are then lumped in the last genus (right side, thick lines, species 6, 7, 8, and 9).

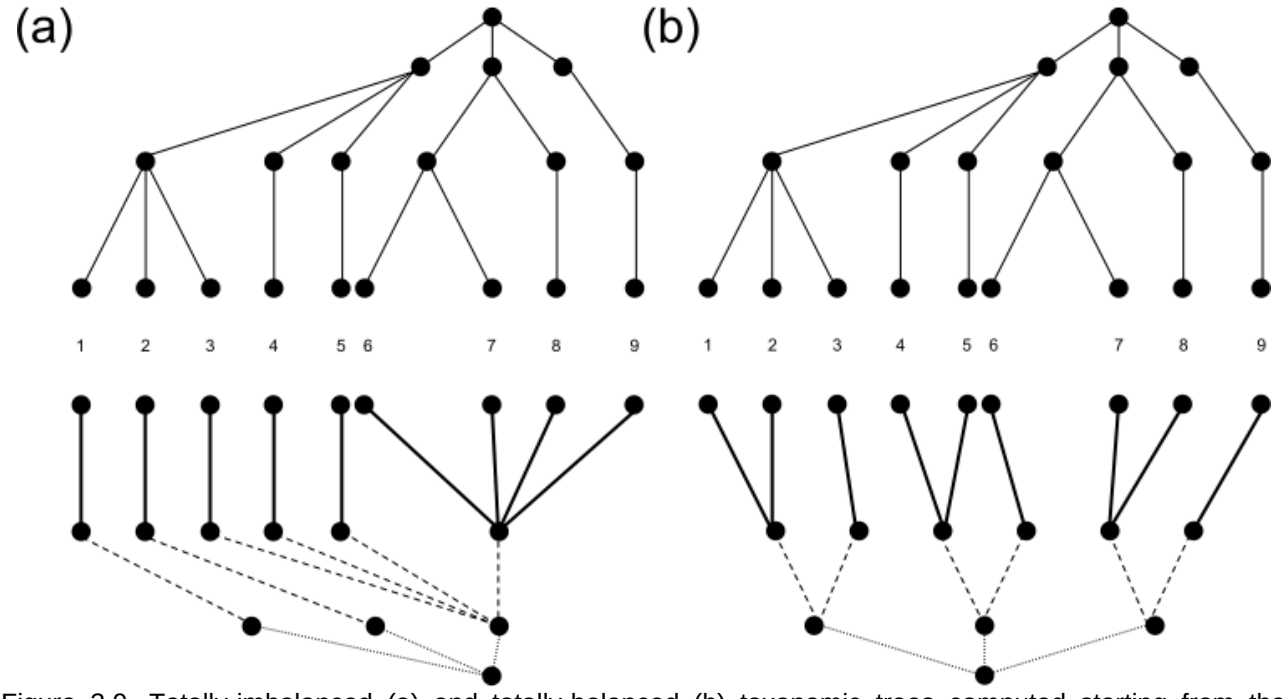


Figure 3.9. Totally-imbalanced (a) and totally-balanced (b) taxonomic trees computed starting from the taxonomic tree introduced in Figure 3.8 and shown at the top of both sides. See text for more details.

The same procedure is repeated in assigning genera to families (dashed lines). As we consider only one order, all families are lumped in it (dotted lines). More generally, the procedure is repeated until the uppermost hierarchical level is reached. (ii) *Obtaining a completely balanced tree.* The procedure is top-down. The first step is forced, as all Families must be lumped in the only present order (dotted lines). Then we proceed assigning (as far as possible) the same number of genera to each Family. In this case, we have 6 genera for 3 families, therefore it is very easy to see that the optimal distribution is 6 / 3 = 2 genera/family (dashed lines). The same step is repeated until the lowermost hierarchical level is reached. Each time we try to optimize the number of taxa which are assigned to all upper levels. We have in this case 9 species for 6 genera (thick lines). Necessarily we will have at best 3 genera with 2 species and 3 genera with 1 species (3 × 2 + 3 × 1 = 9). The optimal situation is the one depicted in the figure. For this reason, it is important to balance taxa not only with respect to the immediately upper taxon, but also with respect to all upper taxa. We note that the completely-balanced and completely-imbalanced trees may not be unique. However, differences in AvTD from different equally-balanced or equally-imbalanced trees are null or negligible.

As the original formulation of AvTD, von Euler's index of imbalance was introduced in the conservation context, since it was used to take estimates on the loss of evolutionary history, and was found to be strictly (negatively) correlated with AvTD (pers. obs.; von Euler, 2001). We introduce I_E in our topic, stating it is a useful balancing indicator for samples used in phylogenetic studies.

Shuffling analysis

Shuffling analysis concepts and purposes are extensively explained in the Results section. Here we think it is useful to report algorithms that were written to carry it out, especially for shuffling phase.

Shuffling phase

User inputs the number of shuffled master lists they want to generate. The user must also decide the number of repetitions for each kind of move. Therefore, each of the following algorithms is repeated the given number of times on the same master list. Then, the resulting file is saved to disk and a new one is produced, with same modalities.

Move: Transfer

1. user is requested to input a taxon level *t*, with $t = 1, 2, \ldots, T - 1$;
2. a taxon *a* of level *t* is randomly chosen;
3. *if* taxon *A* of level $t + 1$ containing *a* contains only *a*

   then return to 2;

   else proceed to 4;
4. a taxon *B* of level $t + 1$ is randomly chosen;
5. *if* taxon *B* = taxon *A*

   then return to 4;

   else proceed to 6;
6. taxon *a* is moved to taxon *B*.

Move: Split

1. user is requested to input a taxon level *t*, with $t = 2, \ldots, T - 1$;
2. a taxon *a* of level *t* is randomly chosen;
3. taxon *a* is split into two new taxa in the same position.

Move: Merge

1. user is requested to input a taxon level *t*, with $t = 2, \ldots, T - 1$;
2. a taxon *a* of level *t* is randomly chosen;
3. *if* taxon *A* of level $t + 1$ containing *a* contains only *a*

   then return to 2;

   else proceed to 4;

4. a taxon *b* of level *t* is randomly chosen within taxon *A*;

5. *if* *a* = *b*

then return to 4;

else proceed to 6;

6. taxa *a* and *b* are merged in a new taxon in the same position.

In all moves, downstream relationships are maintained. For example, if genus *a* containing species *a* and *b* is moved from family *A* to family *B*, species *a* and *b* will still belong to genus *a* within family *B*. The same holds true for splits and merges.

Analysis phase

In this phase, the basic phylogenetic representativeness analysis is applied on each master list. Therefore, a large number (depending upon the chosen number of master lists to be simulated) of analyses are performed and consequently six sets of measurements are obtained for each dimension *s*, namely the six parameters describing AvTD and VarTD:

lower AvTD 95% confidence limit;

mean AvTD;

mean VarTD;

upper VarTD 95% confidence limit;

maximum AvTD;

minimum VarTD;

For the first four sets of measurements, upper and lower 95% confidence limits are computed for each dimension *s* across all master lists, thus giving an idea of the stability of results. For the fifth and sixth sets of measurement, simply the maximum entry is kept for each dimension *s* as above.

CHAPTER 4

A MOLECULAR PHYLOGENY OF BIVALVE MOLLUSKS: ANCIENT RADIATIONS AND DIVERGENCES AS REVEALED BY MITOCHONDRIAL GENES

4.1. INTRODUCTION

The impressive biological success of bivalves is a perfect example of evolutionary potentials embedded in a clear-cut modification of an already successful molluscan body plan. Belonging to phylum Mollusca, first bivalves appeared in the Cambrian period: the peculiar architecture of their shell, lateral compression (and general reduction) of the foot and the complete loss of the radula allowed them to shallowly burrow into soft bottoms. Since then, bivalve phylogeny was a flourishing of branches on a wide tree.

Today's protobranchs most probably resemble those first species, with a well-developed foot, long palp proboscides to bring food to the mouth and true molluscan ctenidia only devoted to gas exchange. The modification of gills for filter feeding, with the consequent reduction and loss of palp proboscides, the gain of byssus, which made epifaunal life possible, the mantle margin fusion, with the emergence of siphons, triggered bivalves' adaptive radiations along geological eras (Stanley, 1968; Morton, 1996; Giribet, 2008; Tsubaki et al., 2010).

Nowadays, bivalve biodiversity is classified into four big clades, which are given the status of subclass. Protobranchs forms were classically divided in two clusters. Species belonging to order Nuculoida are considered among the most primitive bivalves and were included in the subclass Palaeotaxodonta by Newell (1965). The order Solemyoida was described as unrelated to nuculoids for long time, and was included in the subclass

Lipodonta (*sensu* Cope, 1996). More recently, other authors preferred to cluster together both taxa, merging them in a subclass Protobranchia (Starobogatov, 1992; Morton, 1996; von Salvini-Plawen and Steiner, 1996; Waller, 1998); indeed, molecular analyses supported a sister group relationship between the two orders (Steiner and Hammer, 2000; Passamaneck et al., 2004). Furthermore, the superfamily Nuculanoidea was removed from Protobranchia (Giribet and Wheeler, 2002; Giribet and Distel, 2003; Bieler and Mikkelsen, 2006; Plazzi and Passamonti, 2010), and Giribet (2008) proposed the name Opponobranchia referring to the subclass-rank clade Nuculoida + Solemyoida.

Sister group of the Opponobranchia are the Autobranchia (=Autolamellibranchiata *sensu* Giribet, 2008), i.e. bivalves with modified ctenidia, without palp proboscides, generally filibranch or eulamellibranch. Some authors, like Waller (1998), treat Autobranchia as a subclass itself. Following the most widely accepted taxonomy, however, three subclasses, substantially identical to the definition in Newell (1965), belong to Autobranchia: Heterodonta, Palaeoheterodonta, and Pteriomorphia.

Relationships within Autobranchia are still contentious: many studies retrieved a monophyletic clade called Heteroconchia, joining Palaeoheterodonta and Heterodonta (Waller, 1990, 1998; Giribet and Wheeler, 2002; Bieler and Mikkelsen, 2006; Giribet, 2008). Conversely, several phylogenetic analyses resulted in a close relationship between Pteriomorphia and Heterodonta (Cope, 1996, 1997; Canapa et al., 1999; Giribet and Distel, 2003; Doucet-Beaupré et al., 2010; Plazzi and Passamonti, 2010).

Eventually, Anomalodesmata (order Pholadomyoida) are generally eulamellibranch, siphonate burrowers, which developed some remarkable adaptations: some of them are septibranch and deep-water carnivorous organisms. Formerly ascribed to their own subclass (Myra Keen, 1963; Newell, 1965), they are currently considered as a basal, monophyletic clade among Heterodonta (Harper et al., 2000, 2006; Giribet and Wheeler,

2002; Dreyer et al., 2003; Giribet and Distel, 2003; Taylor et al., 2007b; Giribet, 2008; but see Plazzi and Passamonti, 2010).

Notwithstanding the animated debate about bivalve evolution (and systematics), a handful of comprehensive molecular phylogenetic studies have been released to date. After some pioneering analyses (Steiner and Müller, 1996; Adamkewicz et al., 1997; Canapa et al., 1999), and the extensive effort of Campbell (2000), most recent deep phylogenies concentrate on single subclasses: Pteriomorphia (Steiner and Hammer, 2000; Matsumoto, 2003), Anomalodesmata (Dreyer et al., 2003; Harper et al., 2006), and particularly Heterodonta, the most biodiverse group (Williams et al., 2004; Taylor et al., 2007a, 2007b, 2009). Direct optimization (Wheeler, 1996) was used for wide scale phylogenetic reconstructions, as Giribet and Wheeler (2002) and Giribet and Distel (2003) assembled a thorough total evidence matrix, the broadest ever assembled on bivalve evolution.

Finally, our previous study (Plazzi and Passamonti, 2010) was the first attempt to infer a complete evolutionary tree of the class with a robust, two-steps phylogenetic analysis. The aim of that work was to develop a sound pipeline to approach bivalve molecular phylogenetics: present paper follows this pipeline by adding more bivalve taxa, to obtain an in-depth survey of the evolutionary tree of Bivalvia. This study represents the biggest dataset to date of bivalve mollusks, characterized by four mitochondrial genes. Thanks to this improved dataset, we will address all those issues that were not possible to discuss in detail in Plazzi and Passamonti (2010), with special respect to deep relationships linking bivalve subclasses.

4.2. MATERIALS AND METHODS

Taxon sampling, PCR amplification, and sequencing

Sequences added to the bivalve mitochondrial dataset are listed in Table 4.1, along with the specimen voucher number of the MoZoo Lab collection. PCR amplification and cloning were carried out as described in Plazzi and Passamonti (2010): briefly, the Invitrogen (Carlsbad, USA) or ProMega (Madison, USA) *Taq* DNA polymerase kits were used following manufacturers' instructions to amplify target sequences (the mitochondrial genes *12s*, *16s*, *cox1*, *cytb*); a wide range of reaction conditions were used, as different species and markers needed different PCR settings. Typically, a denaturation step of 2' at 94°C was followed by 35 cycles composed by denaturation of 1' at 94°C, annealing of 30''-1' at 46-56°C, and extension of 1' at 72°C. A final extension step of 5' at 72°C was added before cooling amplicons at 4°C. We used the same primers as in Plazzi and Passamonti (2010); specific PCR conditions are available from F. P. upon request. Sequencing reactions were performed through Macrogen (World Meridian Center, Seoul, South Korea) facility. We put special care into avoiding paralogous sequences due to the presence of the DUI mechanism in some bivalve mollusks, as extensively described in Plazzi and Passamonti (2010).

Table 4.1. Species used in our laboratory for this study. All specimen vouchers refer to the bivalve collection of one of authors (M. P.), which is deposited at the Department of Experimental Evolutionary Biology of the University of Bologna, Italy.

Subclass	Order	Suborder	Superfamily	Family	Subfamily	Species	Specimen voucher	Sampling locality
Heterodonta	Chamida		Cardioidea	Cardiidae	Laevicardiinae	*Laevicardium crassum*	BES|MPB|427	41°38.13'N 16°53.24'E 135 m
			Tellinoidea	Semelidae		*Abra longicallus*	BES|MPB|348	42°50.45'N 14°49.55'E 232 m - 42°48.62'N 14°52.09'E 224 m
							BES|MPB|354	42°53.53'N 15°04.70'E 195 m - 42°55.21'N 15°04.37'E 200 m
			Veneroidea	Veneridae	Chioninae	*Clausinella brongniartii*	BES|MPB|422	42°07.34'N 15°28.86'E 32 m - 42°07.34'N 15°28.83' 31 m
						Timoclea ovata	BES|MPB|200	42°07.99'N 15°30.07'E 52 m
					Dosiniinae	*Dosinia exoleta*	BES|MPB|067	Trieste, Italy
					Pitarinae	*Pitar rudis*	BES|MPB|452	Grado, Italy
					Venerinae	*Venus casina*	BES|MPB|440	42°07.67'N 15°30.06'E 27 m
Opponobranchia	Nuculoida		Nuculoidea	Nuculidae		*Nucula decipiens*	BES|MPB|589	41°14.68'N 17°20.52'E 600 m- 41°14.67'N 17°19.50'E 293 m
						Nucula sulcata	BES|MPB|421	42°52.90'N 15°03.67'E 198 m - 42°55.24'N 15°02.33'E 187 m
Palaeoheterodonta	Unionida		Unionoidea	Unionidae	Anodontinae	*Anodonta cygnea*	BES|MPB|610	Castel dell'Alpi, Italy
Pteriomorphia	Arcida	Arcina	Arcoidea	Arcidae	Anadarinae	*Anadara diluvii*	BES|MPB|411	42°01.41'N 16°12.21'E 54 m
						Anadara transversa	BES|MPB|326	Woods Hole, USA
					Arcinae	*Asperarca nodulosa*	BES|MPB|684	Sicily Channel, Italy
						Asperarca secreta	BES|MPB|579	41°14.68'N 17°20.52'E 600 m- 41°14.67'N 17°19.50'E 293 m
						Barbatia barbata	BES|MPB|044	Scoglio del Remaiolo, Italy
				Noetiidae	Striarcinae	*Striarca lactea*	BES|MPB|132	Krk, Croatia
	Limida		Limoidea	Limidae		*Lima hians*	BES|MPB|102	Trieste, Italy
	Mytilida	Mytilina	Mytilioidea	Mytilidae	Lithophaginae	*Lithophaga lithophaga*	BES|MPB|123	Krk, Croatia
					Modiolinae	*Modiolus barbatus*	BES|MPB|446	Muggia, Italy
					Mytilinae	*Mytilaster lineatus*	BES|MPB|118	Krk, Croatia
						Mytilaster solidus	BES|MPB|120	Krk, Croatia
	Ostreoida	Ostreina	Ostreoidea	Gryphaeidae	Pycnodonteinae	*Neopycnodonte cochlear*	BES|MPB|347	42°50.45'N 14°49.55'E 232 m - 42°48.62'N 14°52.09'E 224 m
		Pectinina	Pectinoidea	Pectinidae	Chlamydinae	*Chlamys bruei*	BES|MPB|092	Vieste, Italy
						Chlamys multistriata	BES|MPB|130	Krk, Croatia
					Pectininae	*Peplum clavatum*	BES|MPB|653	35°58.29'N 14°16.28'E 184 m - 35°56.93'N 14°18.11'E 162 m
				Propeamussiidae		*Adamussium colbecki*	BES|MPB|027	Antarctica
				Spondylidae		*Spondylus gaederopus*	BES|MPB|091	Krk, Croatia
	Pteriida	Pteriina	Pterioidea	Isognomonidae		*Isognomon acutirostris*	BES|MPB|272	Nosy Be, Madagascar
				Pteriidae		*Pteria hirundo*	BES|MPB|513	Plavnik, Croatia

Assembling the dataset

Electropherograms were read through MEGA 4 (Tamura et al., 2007): sequencer files were manually checked and edited when necessary. The CLC Sequence Viewer 6.4 software (CLC bio, Aarhus, Denmark) was used to organize and to download sequences from GenBank (at December 2010). We then retrieved those taxa for which at least three on four markers were present. Four alignments were prepared with CLC Sequence Viewer and aligned with ClustalW (Thompson et al., 1994) at the EBI server (http://www.ebi.ac.uk/Tools/msa/clustalw2/; Chenna et al., 2003). For ribosomal genes, the IUB matrix was used with a 25 penalty for gap opening and a 5 penalty for gap extension; for protein-coding genes (PCGs), penalties were set to 50 and 10, respectively. When a sequence was not available for a given species, it was replaced with a stretch of missing data in that alignment; Hartmann and Vision (2008; and reference therein) showed that a large amount of missing data do not lead to incorrect phylogeny in itself, as long as sufficient data are available. In many cases, we lumped together sequences of different congeneric species to represent the genus they belong to: this is a common practice in deep phylogenetic studies and does not lead to inconsistent results at the class level, which is targeted in this study (see, f.i., Plazzi and Passamonti, 2010; Li et al., 2009). Five outgroups were selected for this study: the polyplacophoran *Katharina tunicata*, two scaphopods (*Graptacme eborea* and *Siphonodentalium lobatum*) and two gastropods (*Haliotis tuberculata* and *Thais clavigera*). Appendix 4.1 lists all sequences used for this study, both downloaded from GenBank and produced in our laboratory.

Region of ambiguous alignment for ribosomal genes were detected by GBlocks (Talavera and Castresana, 2007; Castresana, 2000) with the following parameters: minimum number of sequences for a conserved position, half + 1; minimum number of sequences for a flanking position, half + 1; maximum number of contiguous nonconserved positions, 50; minimum length of a block, 10; allowed gap positions, all. Finally, gaps were

coded following the simple indel method of Simmons and Ochoterena (2000) as described in Plazzi and Passamonti (2010); this was carried out with the software GapCoder (Young and Healy, 2003).

Evaluating phylogenetic signal

Taxon sampling was investigated through the method described in Plazzi et al. (2010), which has the property of involving only preexistent taxonomic knowledge about the target group, and does not need any preliminary genetic analysis: for this reason, this is a truly *a priori* test on taxonomic coverage. All analyses were carried out through the software PhyRe (Plazzi et al., 2010) and the bivalve checklist compiled by Millard (2001), with 100 random resamplings in all cases. Shuffling test was performed at the family level: 100 master list were generated and the number of splits, merges, and moves was set to 12, 8, and 4, respectively. We empirically showed in our previous paper (Plazzi and Passamonti, 2010) that a sample size of about 30 species is sufficient to correctly estimate all molecular evolutionary parameters from a bivalve dataset (given the four mitochondrial markers we employ here); therefore, we did not use any *a posteriori* test for taxon sampling, as the sample size is more than four times in this study.

A saturation analysis was conducted following methods recommended by Luo et al. (2011) and Roe and Sperling (2007) through the program PAUP* 4.0b10 (Swofford, 2002) using PaupUp graphical interface (Calendini and Martin, 2005). The transition/transversion (Ti/Tv) ratio was computed on the absolute number of differences; Ti/Tv ratio was then transformed to %Ti (the percentage of transition on total differences) and plotted against pairwise K2P distances. A low %Ti value was considered when less than or equal to 50% (corresponding to a Ti/Tv ratio ≤ 1; Roe and Sperling, 2007). The saturation test was conducted independently for the four markers and, about PCGs, for third codon positions only.

We performed splits-decomposition analysis with two different approaches. First, we used SplitsTree 4.6 (Dress et al., 1996; Huson and Bryant, 2006) to obtain phylogenetic networks in which more splits leading to specific clades are shown than in a strictly bifurcating tree. This method aimed to evaluate phylogenetic signal in raw data, therefore the neighbornet network was chosen (Bryant and Moulton, 2004; Wägele et al., 2009), based on either uncorrected ("*p*-") or Log-Det distances. Second, a spectral analysis was performed to investigate on split support ranking along our alignment. The software SAMS (Wägele and Mayer, 2007) identifies split-supporting positions without reference to a tree and a model choice (Lento et al., 1995; Wägele and Rödding, 1998a, 1998b). Many genera are represented in our dataset by more than one species, leading to several strong "trivial" splits, i.e. those clustering congeneric taxa, which should never be challenged at our phylogeny depth: therefore, we restricted our analysis to occurring splits where ingroups had a minimum size of 5. Bootstrap-based confidence limits were computed on 500 random replicates.

Presence and properties of phylogenetic signal were also tested with the Likelihood Mapping (LM) approach (Strimmer and von Haeseler, 1996, 1997) as implemented in the software TreePuzzle 5.2 (Schmidt et al., 2002; Schmidt and von Haeseler, 2003). The complete alignment was used as a dataset; outgroups were excluded and 1000 random quartets were drawn to produce the final result. Four-cluster Likelihood-Mapping (Strimmer and von Haeseler, 1997) analyses were conducted on each partition of our dataset (see below) independently; in each case, molecular evolutionary parameters were given as computed by ModelTest (Posada and Crandall, 1998). In this case, we excluded outgroups and Opponobranchia (given their stable basal position in all analyses) and subdivided all remaining taxa between four subclasses (Anomalodesmata, Heterodonta, Palaeoheterodonta, and Pteriomoprhia). Final plots were again constructed on 1000 randomly drawn quartets. Significance of results was tested with a Chi-Square test

assuming as a null distribution an even presence of observation in each of the three regions of the triangle.

Model decision tests and tree inference

Our dataset was arranged, according to Plazzi and Passamonti (2010), in 26 different partitions: the complete alignment (*all*), the concatenated ribosomal genes (*rib*), the concatenated PCGs (*prot*), individual genes (*12s*, *16s*, *cox1*, *cytb*), individual codon positions among the *prot* partition and single PCGs (*prot_1*, *prot_2*, *prot_3*, *cox1_1*, *cox1_2*, *cox1_3*, *cytb_1*, *cytb_2*, *cytb_3*), the concatenated first and second codon positions (*prot_12*, *cox1_12*, *cytb_12*), and the corresponding indel characters coded as 0/1, irrespective of codon positions (*all_indel*, *rib_indel*, *prot_indel*, *12s_indel*, *16s_indel*, *cox1_indel*, *cytb_indel*). These partitions were assembled in 13 different partitioning schemes, as shown in Table 4.2. The best-fitting evolutionary model was selected with ModelTest 3.7 using the graphical interface provided by MrMTgui (Nuin, 2008); the Bayesian Information Criterion (BIC) was preferred as a model decision criterion (Luo et al., 2010; and reference therein).

Table 4.2. Partitioning schemes adopted for this study. Asterisks mark schemes analyzed by both 4by4 and codon models, respectively.

Name	Number of partitions											
p01	2	*all*	*all_indel*									
*p02-p14	4	*rib*	*rib_indel*	*prot*	*prot_indel*							
*p03-p15	6	*12s*	*12s_indel*	*16s*	*16s_indel*	*prot*	*prot_indel*					
p04	5	*rib*	*rib_indel*	*prot_12*	*prot_3*	*prot_indel*						
p05	6	*rib*	*rib_indel*	*prot_1*	*prot_2*	*prot_3*	*prot_indel*					
*p06-p16	6	*rib*	*rib_indel*	*cox1*	*cox1_indel*	*cytb*	*cytb_indel*					
p07	8	*rib*	*rib_indel*	*cox1_12*	*cox1_3*	*cox1_indel*	*cytb_12*	*cytb_3*	*cytb_indel*			
p08	10	*rib*	*rib_indel*	*cox1_1*	*cox1_2*	*cox1_3*	*cox1_indel*	*cytb_1*	*cytb_2*	*cytb_3*	*cytb_indel*	
p09	7	*12s*	*12s_indel*	*16s*	*16s_indel*	*prot_12*	*prot_3*	*prot_indel*				
p10	8	*12s*	*12s_indel*	*16s*	*16s_indel*	*prot_1*	*prot_2*	*prot_3*	*prot_indel*			
*p11-p17	8	*12s*	*12s_indel*	*16s*	*16s_indel*	*cox1*	*cox1_indel*	*cytb*	*cytb_indel*			
p12	10	*12s*	*12s_indel*	*16s*	*16s_indel*	*cox1_12*	*cox1_3*	*cox1_indel*	*cytb_12*	*cytb_3*	*cytb_indel*	
p13	12	*12s*	*12s_indel*	*16s*	*16s_indel*	*cox1_1*	*cox1_2*	*cox1_3*	*cox1_indel*	*cytb_1*	*cytb_2*	*cytb_3*	*cytb_indel*

Maximum Likelihood (ML) analysis was carried out with PAUP* 4.0b10. The alignment was not partitioned and molecular evolutionary parameters computed by ModelTest 3.7 were used for likelihood calculations. Gaps were treated as missing data and binary characters were excluded from the analysis. The outgroups were forced to be paraphyletic with respect to the ingroup. Bootstrap consensus tree using full heuristic ML searches with stepwise additions and TBR branch swapping was constructed to assess nodal support. As described in Plazzi and Passamonti (2010), 150 input files were sent to the University of Oslo Bioportal facility (http://www.bioportal.uio.no) in a parallel run, each computing the maximum likelihood tree for a single bootstrap replicate. Random seed were generated according to PAUP* recommendations with Microsoft Excel® and the consensus tree was computed with Phyutility (Smith and Dunn, 2008).

All the 13 partitioning schemes were investigated in a Bayesian Analysis (BA) with the software MrBayes 3.1.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) hosted at the University of Oslo Bioportal. Initially, the so-called "4by4" nucleotide model (i.e., a traditional 4×4 substitution matrix) was used for all partitioning schemes. For 4 partitioning schemes (see Tab. 4.2), namely those containing PCG (*prot*, *cox1*, or *cytb*) partitions, we implemented for PCGs a codon model (Goldman and Yang, 1994; Muse and Gaut, 1994), the M3 model. 10,000,000 generations of two parallel MC^3 analyses of 4 chains each were run for each 4by4 partitioning scheme. Since in this analysis we are focusing on the relationships among subclasses, Bivalves were constrained to be monophyletic with respect to the five molluscan outgroups. Nucleotide partitions were treated according to ModelTest results; binary partitions were treated with the default model for restriction data enforcing the coding=variable option and a gamma heterogeneity in substitution distribution. Convergence was estimated by PSRF (Gelman and Rubin, 1992) and by plotting standard deviation of average split frequencies sampled every 1,000 generations. For each M3 analysis 4 independent run of 5,000,000

generations of one single MC^3 algorithm were run and convergence among and within runs was estimated via the AWTY tools (http://king2.scs.fsu.edu/ CEBProjects/ awty/ awty_start.php; Nylander et al., 2008). A tree was sampled every 100 (4by4 models) or every 125 (M3 models) generations and the consensus was computed at convergence after burnin removal.

The Estimated Marginal Likelihood (EML) computed by MrBayes 3.1.2 made possible to compute the Akaike Information Criterion (AIC; Akaike, 1973) and the Bayes Factor (BF; Kass and Raftery, 1995), as described in Plazzi and Passamonti (2010; and reference therein). Briefly, the AIC provides an estimate of the Kullback-Leibler distance (Kullback and Leibler, 1951), i.e. the distance of the model from the reality, considering a penalty computed on the number of free parameters; therefore, smaller values are preferable. On the other hand, the BF involves pairwise comparisons among models through the EML ratio: the larger is the BF value, the more the first model overcomes the second one.

All trees were graphically edited by PhyloWidget (Jordan and Piel, 2008) and Dendroscope (Huson et al., 2007) softwares. Optimization of morphological characters on the best evolutionary topology was carried out with Mesquite 2.74 (Maddison and Maddison, 2010): matrix was taken from Newell (1965), with the exception of hinge type, which was coded following Giribet and Wheeler (2002). The parsimony method was chosen, as in two cases multiple state characters were coded; in other cases, we tested parsimony results with ML approach, using the MK1 model as implemented by Mesquite.

4.3. RESULTS

Sequence data

A total of 60 sequences from 29 species were obtained for this study and deposited in GenBank under Accession Numbers JF496737-JF496786. Sequences of *12s*, *16s*, *cox1* and *cytb* were 19, 9, 17, and 15, respectively. Details of the concatenated alignment are listed in Table 4.3. After removal of ambiguously aligned positions and related indel characters, 2260 nucleotides and 735 indels were left for phylogenetic analyses, for a total of 2995 characters. The complete dataset comprehends 436 sequences from 122 bivalves and five outgroup species. Interestingly, we found four PCG sequences (*Chlamys multistriata*, *Neopycnodonte cochlear*, and *Spondylus gaederopus* for *cox1*; *Laevicardium crassum* for *cytb*) where single-site gaps have to be included to obtain a correct alignment. In our previous work, we noted the same for *Hyotissa hyotis* and *Barbatia* cfr. *setigera* cytochrome b sequences (Plazzi and Passamonti, 2010). The alignment, both at nucleotide and amminoacid level, is otherwise good, therefore it is unlikely we are facing a NUMT (i.e., a mitochondrial pseudogene; Sorenson and Quinn, 1998), inasmuch that no NUMTs have been reported for bivalves yet (Bensasson et al., 2001; Zbawicka et al., 2007). It is also unlikely a repeated sequencer error or a complemented frameshift mutation, as such anomalies occur in different position of the sequence. Even if we do not have empirical data on this account, single nucleotide indels in apparently functional mitochondrial genes – *cytb* being one of them – have been reported and discussed elsewhere (Mindell et al., 1998; Grant and D'Haese, 2004; Beckenbach et al., 2005; and reference therein). It is possible that we are coping with a similar situation, which surely deserves further investigation. For phylogenetic purposes, we inserted missing data instead of single-site gaps whenever they mapped in a region of the gene included in the alignment.

Table 4.3. Alignment details. Site numbers refer to the complete concatenated alignment; in the GBlocks column the number of bases retained after removal of ambiguously aligned characters is shown for *12s* and *16s* genes and indels. For further details on sequences for a specific gene alignment, see Appendix 4.1.

Marker	Start site	End site	Length	Gblocks	Number of sequences
12s	1	906	906	599	101
12s_indel	907	1545	639	344	
16s	1546	2341	796	574	112
16s_indel	2342	2950	609	362	
cox1	2951	3634	684		126
cox1_indel	3635	3655	21		
cytb	3656	4058	403		100
cytb_indel	4059	4066	8		

Evaluating phylogenetic signal

Phylogenetic Representativeness test aims to measure the degree of representativeness of a sample with respect to the group it should represent in a phylogenetic analysis (Fig. 4.1; see Plazzi and Passamonti, 2010). The measured Average Taxonomic Distinctness (AvTD) of our sample of 86 bivalve genera fell within the 95% confidence interval of AvTD computed from 100 random subsample of the same dimension. However, the Variance in Taxonomic Distinctness (VarTD) was clearly higher than its 95% confidence interval (Fig. 4.1A). Moreover, the AvTD of our sample was within the range of 95% lower confidence limit yielded by shuffling test (Fig. 4.1B). Most probably, the little sampling among Anomalodesmata taxa (which are indeed hard to obtain) is the main reason of the border-line AvTD and the high VarTD we found.

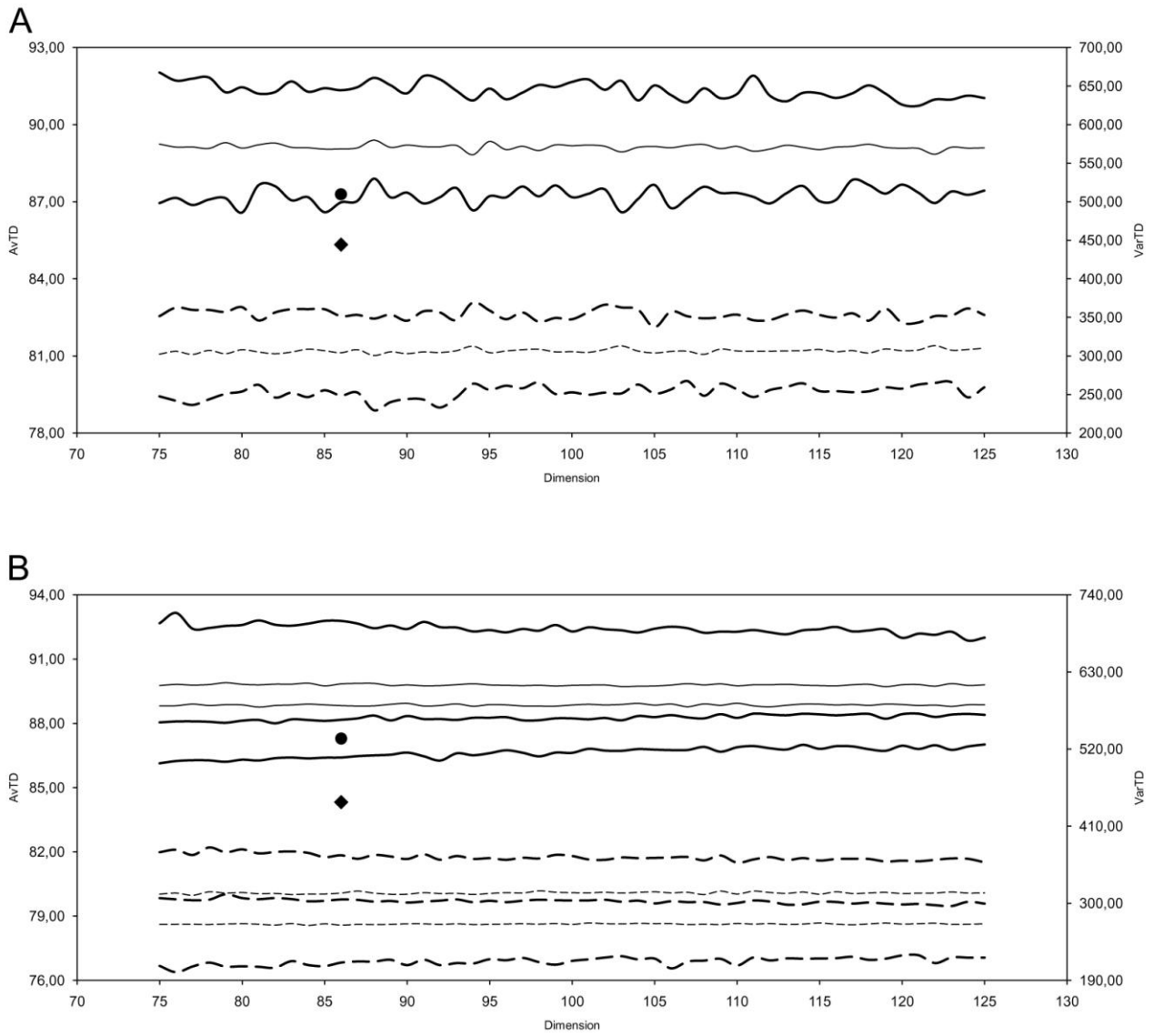


Figure 4.1. Results from Phylogenetic Representativeness test. A, AvTD and VarTD computed for the sample used for this study. AvTD is plotted on left axis: the circle represents the value obtained from the present sample, whereas continue lines indicate the lower 95% confidence limit, the maximum value for that sample dimension (thick lines), and the mean AvTD (thin line). VarTD is plotted on the right axis: the diamond represents the value obtained from the present sample, whereas dotted lines indicate the minimum value for that sample dimension, the upper 95% condifence limit (thick lines), and the mean VarTD (thin line). B, shuffling test with 100 randomly shuffled master lists (see text for details). Mean VarTD (thin dotted lines), upper 95% VarTD confidence limit (upper thick dotted lines), lower 95% AvTD confidence limit (lower thick continue lines), and mean AvTD (thin continue lines) are shown as the 95% confidence intervals across the replicates. Axes, circle, and diamond as above.

Pairwise %Ti data plotted on K2P distances showed only little saturation in substitutions along our dataset (Fig. 4.2), which is expected given the depth of this phylogeny. %Ti was slightly lower than 50% in all datasets, but this result is constant for all pairwise comparisons, even for larger K2P distances. As %Ti is used as a proxy for

saturation (Roe and Sperling, 2007), this means that saturation is not expected to increase when increasing the distance between two taxa. Eventually, larger distances values were expectedly obtained for *12s* gene, as well as for third codon positions of both PCGs.

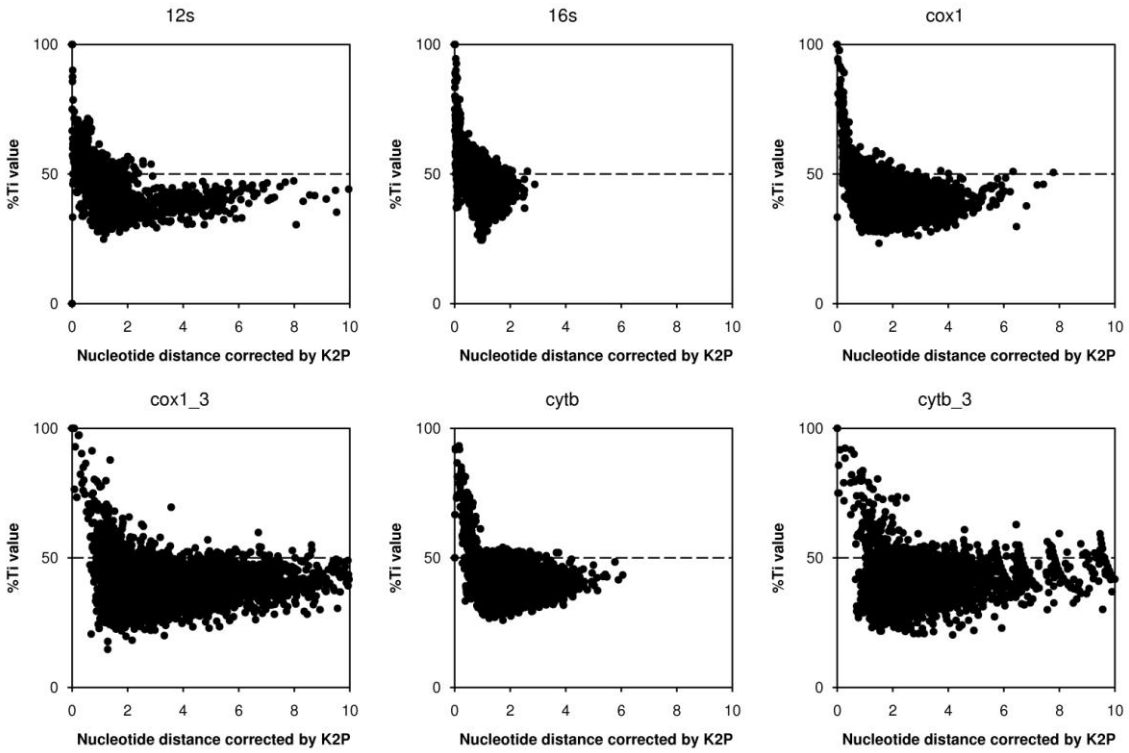


Figure 4.2. Percentage of transitions (%Ti) plotted on K2P distances to estimate saturation in our dataset. The dotted line indicates the 50% threshold for %Ti to be considered low.

Neighbornet networks of the complete alignment were produced for single genes and for the concatenated alignment, based both on uncorrected and LogDet distances. All networks are essentially similar, varying only in the positions of some taxa, like *Lucinella*, *Loripes*, *Cuspidaria*, *Nuculana*, *Astarte*, and *Cardita*. Figure 4.3 shows the LogDet neighbornet network for the complete alignment: all genera and families are retrieved as well-defined clades, with the exception of mytilids and *Chlamys*. Although the network is less clearly tree-like in deep relationships, some sharp signal is present also for major groups, like Palaeoheterodonta (the Unionidae are very well distinct in all networks). The Opponobranchia cluster often together with *Haliotis* and other outgroups. The position of

anomalodesmatans is unstable among different genes and distance methods: under LogDet model, they cluster together next to part of the family Mytilidae (*Lithophaga lithophaga*, *Mytilaster lineatus*, *Modiolus sp.*), whereas under the uncorrected method *Cuspidaria* is found close to *Loripes* and *Lucinella* between Opponobranchia and Heterodonta and *Pandora* and *Thracia* are found in a star-like region of the tree with *Cardita*, *Astarte* and *Nuculana*. These last three genera are found among pteriomorph species under the LogDet model. Single-gene networks generally are consistent with this topology, with local decreasing of resolution in some part of the graph. Long branches were individuated only in some single-gene networks (mostly those of ribosomal markers), whereas for the concatenated alignment this was only the case for the scaphopod outgroup *Siphonodentalium lobatum*.

Spectral analysis revealed the non-triviality of phylogenetic signal in our data (Fig. 4.4). The first and second most supported splits with at least 5 taxa in the ingroup do appear as monophyletic in the final evolutionary tree (see below): they correspond to the family Ostreidae and to the subfamily Mytilinae with the exception of *Mytilaster lineatus*, exactly as in the tree. In facts, only 9 out of 50 best supported splits were found as monophyletic clusters in the final tree, but they increase to 25 if we consider those splits differing for just one or two taxa from the relative cluster in the cladogram. Interestingly, most of these 25 total recovered splits refer to pteriomorph clusters. Overall, the signal was generally noisy and no binary splits were found.

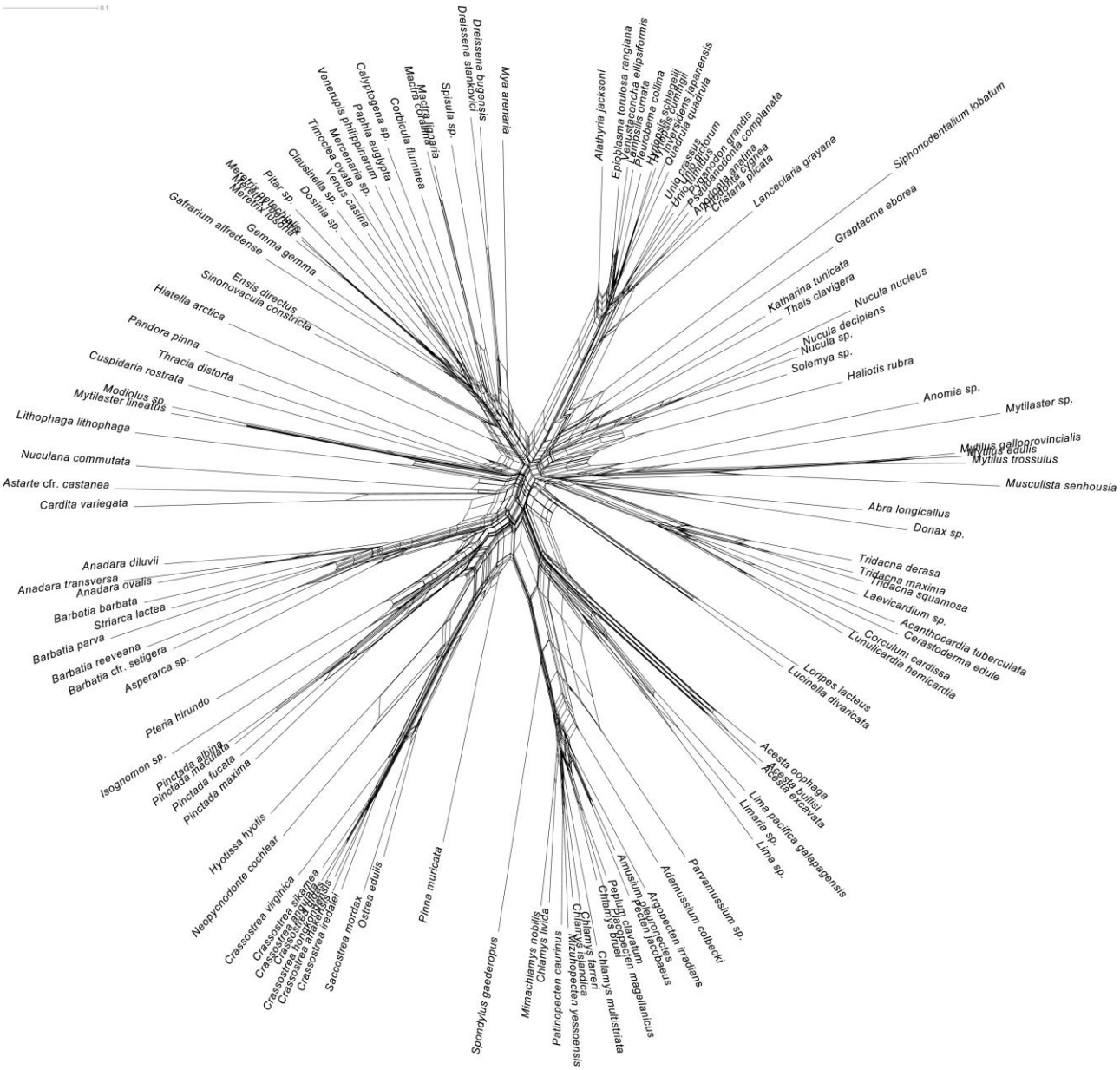
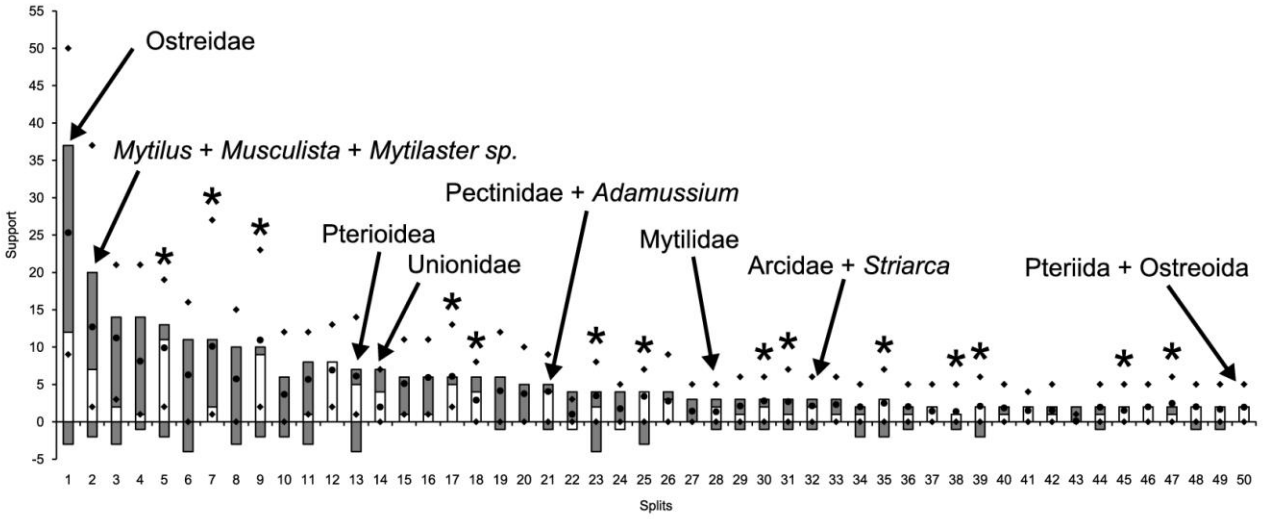


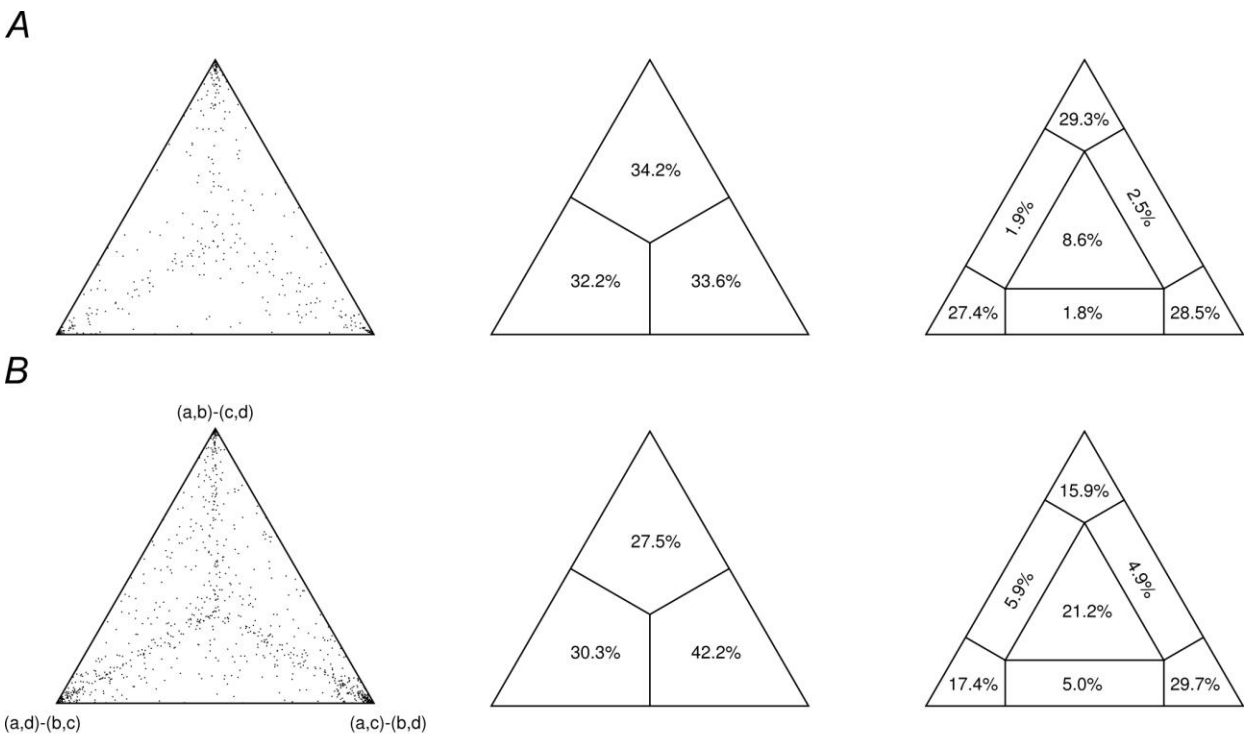
Figure 4.3. Neighbornet network based on LogDet distances.



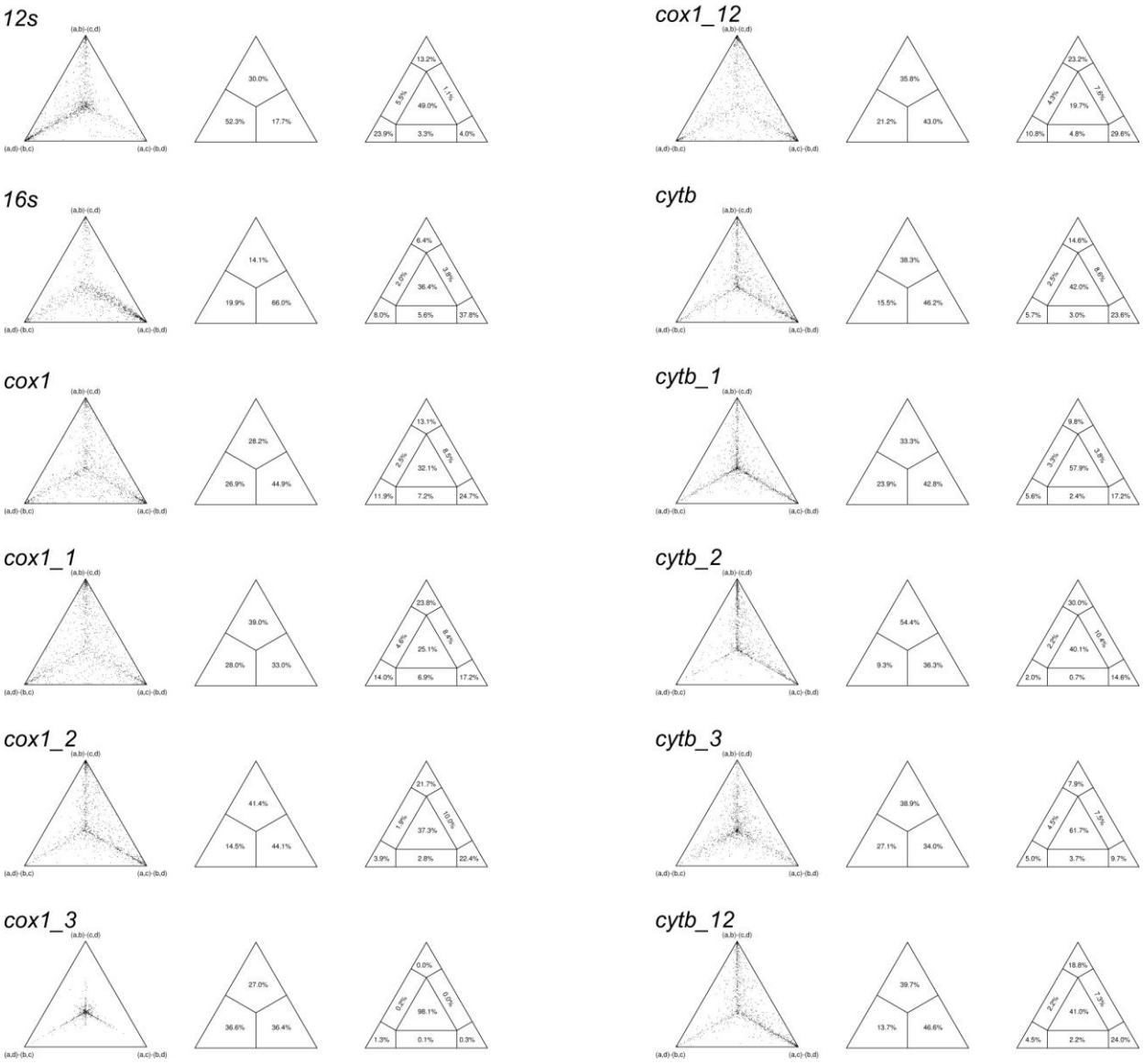
Likelihood Mapping (Fig. 4.5) allowed to estimate the amount of signal present in our data; first of all, 1000 random quartets were drawn without constraints. They are evenly ($P > 0.05$) distributed in the simplex, but only 8.6% of them do fall into the star-like tree area, while 85.2% map near one of the three vertices, indicating that in most cases a topology is strongly favored over alternative hypotheses. The concatenated alignment as well as single genes and partitions were examined, and in all cases a preferred topology was individuated (Fig. 4.5). 8 out of 13 analyses indicated the unrooted topology ((Palaeoheterodonta + Heterodonta) + (Anomalodesmata + Pteriomorphia)) as the most supported; the second most supported topology was ((Palaeoheterodonta + Anomalodesmata) + (Heterodonta + Pteriomorphia)), which was retrieved for 3 partitions. As results from all 13 analyses were significantly different from the null hypothesis ($P < 0.005$) and that more than 60% of them pointed towards the same backbone tree, it is evident that a phylogenetic signal does unveil itself in our data.

Figure 4.4 *(previous page)*. Spectral analysis. The best 50 splits with at least 5 taxa in the ingroup are shown (see text for details) on the *x*-axis. Support is shown on the *y*-axis. Positive values indicate support for the ingroup, whereas negative values indicate support for the outgroup; the ingroup was always chosen as the most supported of either clade for each split. No binary splits were found; support for a clade with noise in outgroup clade is shown in white; support for a clade with noise both in ingroup and outgroup clade is shown in gray. Dots indicate the mean ingroup support value across 500 bootstrap replicates; lower and upper 95% condifence limits are shown as diamonds. Nodes which are found on the tree are indicated; nodes which are different from those on the tree by one or two taxa are marked with asterisks.

Figure 4.5 *(next page)*. Likelihood Mapping. Each analysis was performed on 1,000 random quartets; the left simplex shows point distribution; the central one the subdivision among the three corners; the right one the subdivision among Voronoi cells (Strimmer and von Haeseler, 1997; Nieselt-Struwe and von Haeseler, 2001). A, Likelihood Mapping for the concatenated alignment without grouping. B, Likelihood Mapping for the concatenated alignment with Opponobranchia excluded and remaining taxa subdivided into Palaeoheterodonta (a), Anomalodesmata (b), Heterodonta (c), and Pteriomorphia (d). The three possible topologies are shown at vertices. C, Likelihood Mapping for single partitions with Opponobranchia excluded and remaining taxa subdivided as above.


A
34.2%
32.2%
33.6%
29.3%
1.9%
2.5%
8.6%
27.4%
1.8%
28.5%
B
(a,b)-(c,d)
(a,d)-(b,c)
(a,c)-(b,d)
27.5%
30.3%
42.2%
15.9%
5.9%
4.9%
21.2%
17.4%
5.0%
29.7%

C


12s
30.0%
52.3%
17.7%
13.2%
5.9%
1.1%
49.0%
23.9%
3.3%
4.0%
cox1_12
35.8%
21.2%
43.0%
23.2%
4.3%
7.8%
19.7%
10.8%
4.8%
29.6%
16s
14.1%
19.9%
66.0%
6.4%
2.0%
3.8%
36.4%
8.0%
5.6%
37.8%
cytb
38.3%
15.5%
46.2%
14.6%
2.5%
8.8%
42.0%
5.7%
3.0%
23.6%
cox1
28.2%
26.9%
44.9%
13.1%
2.5%
8.8%
32.1%
11.9%
7.2%
24.7%
cytb_1
33.3%
23.9%
42.8%
9.8%
3.9%
3.8%
57.9%
5.6%
2.4%
17.2%
cox1_1
39.0%
28.0%
33.0%
23.8%
4.8%
8.4%
25.1%
14.0%
6.9%
17.2%
cytb_2
54.4%
9.3%
36.3%
30.0%
2.2%
10.8%
40.1%
2.0%
0.7%
14.6%
cox1_2
41.4%
14.5%
44.1%
21.7%
1.9%
10.8%
37.3%
3.9%
2.8%
22.4%
cytb_3
38.9%
27.1%
34.0%
7.9%
4.8%
7.8%
61.7%
5.0%
3.7%
9.7%
cox1_3
27.0%
36.6%
36.4%
0.0%
0.2%
0.0%
98.1%
1.3%
0.1%
0.3%
cytb_12
39.7%
13.7%
46.6%
18.8%
2.2%
7.3%
41.0%
4.5%
2.2%
24.0%

Phylogenetic reconstructions

Results of molecular evolution models for each partition are extensively listed in Appendix 4.2. For the ML analysis the model selected for the partition *all* was implemented with PAUP*. The heuristic search with 150 bootstrap replicates yielded a well resolved consensus tree with generally high support values (Fig. 4.6).

Bivalves did not cluster in a supported monophyletic clade: the scaphopod *Siphonodentalium lobatum* was found to be the sister group of a polytomy with *Katharina*, *Haliotis*, *Thais*, genus *Nucula*, *Solemya*, and all remaining bivalves (the Autobranchia), whose monophyly has a bootstrap proportion (BP) value of 65. The first split separates Palaeoheterodonta (BP=100) and a broad assemblage of species belonging to Anomalodesmata, Heterodonta, and Pteriomorphia. This assemblage is a polytomy (BP=70); its branches are the Heterodonta bulk (BP=87), the cluster *Loripes* + *Lucinella* (BP=100), the cluster *Cardita* + *Astarte* (BP=100), *Cuspidaria rostrata*, the cluster *Pandora* + *Thracia* (BP=78), the Pteriomorphia bulk (BP=73), and the family Mytilidae (BP=100). As a consequence, neither Heterodonta or Pteriomorphia were retrieved as monophyletic, nor were anomalodesmatans.

Resolution is higher within each subclass, where most of the clades are supported by bootstrap. The only exception is a wide polytomy within heterodonts (BP=84), which is sister group of family Mactridae (BP=100): this polytomy comprehends *Calyptogena* (family Vesicomyidae), *Corbicula* (family Corbiculidae), and six branches of venerid taxa. Four main lineages can be acknowledged within Pteriomorphia: *Nuculana* (superfamily Nuculanoidea), (Anomioidea + (Limoidea + Pectionoidea)), (*Pinna* + (Ostreoidea + Pterioidea)), and Arcoidea. Families and genera are generally monophyletic, with some notable exceptions like, f.i., family Arcidae and genus *Mytilaster*.

Figure 4.6. Maximum Likelihood tree. Shown is the consensus tree of 150 bootstrap replicates, using the concatenated alignment as a single partition. Values at the nodes are Bootstrap Proportions (BP); nodes were collapsed if BP<60.

Table 4.4. Results of Akaike Information Criterion (AIC) test. Partitioning scheme details are listed in Table 4.2. K, number of free parameters used for that model; EML, Estimated Marginal Likelihood as computed by MrBayes 3.1.2; AIC, Akaike Information Criterion statistics.

	K	EML	AIC
p01	518	-121,834.76	244,705.52
p02	1,036	-121,299.29	244,670.58
p03	1,554	-121,270.99	245,649.98
p04	1,298	-119,802.75	242,201.50
p05	1,561	-119,465.02	242,052.04
p06	1,554	-121,259.23	245,626.46
p07	2,078	-119,690.34	243,536.68
p08	2,602	-119,325.67	243,855.34
p09	1,816	-119,768.83	243,169.66
p10	2,079	-119,422.14	243,002.28
p11	2,072	-121,225.15	246,594.30
p12	2,596	-119,662.18	244,516.36
p13	3,120	-119,299.99	244,839.98
p14	1,097	-118,729.10	239,652.20
p15	1,615	-118,502.26	240,234.52
p16	1,676	-118,392.57	240,137.14
p17	2,194	-118,205.79	240,799.58

Table 4.5. Bayes Factor (BF) results. Partitioning scheme details are listed in Table 4.2; Estimated Marginal Likelihood (EML) values are shown in Table 4.4.

	p01	p02	p03	p04	p05	p06	p07	p08	p09	p10	p11	p12	p13	p14	p15	p16	p17
p01		1,070.94	1,127.54	4,064.02	4,739.48	1,151.06	4,288.84	5,018.18	4,131.86	4,825.24	1,219.22	4,345.16	5,069.54	6,211.32	6,665.00	6,884.38	7,257.94
p02			56.60	2,993.08	3,668.54	80.12	3,217.90	3,947.24	3,060.92	3,754.30	148.28	3,274.22	3,998.60	5,140.38	5,594.06	5,813.44	6,187.00
p03				2,936.48	3,611.94	23.52	3,161.30	3,890.64	3,004.32	3,697.70	91.68	3,217.62	3,942.00	5,083.78	5,537.46	5,756.84	6,130.40
p04					675.46	-2,912.96	224.82	954.16	67.84	761.22	-2,844.80	281.14	1,005.52	2,147.30	2,600.98	2,820.36	3,193.92
p05						-3,588.42	-450.64	278.70	-607.62	85.76	-3,520.26	-394.32	330.06	1,471.84	1,925.52	2,144.90	2,518.46
p06							3,137.78	3,867.12	2,980.80	3,674.18	68.16	3,194.10	3,918.48	5,060.26	5,513.94	5,733.32	6,106.88
p07								729.34	-156.98	536.40	-3,069.62	56.32	780.70	1,922.48	2,376.16	2,595.54	2,969.10
p08									-886.32	-192.94	-3,798.96	-673.02	51.36	1,193.14	1,646.82	1,866.20	2,239.76
p09										693.38	-2,912.64	213.30	937.68	2,079.46	2,533.14	2,752.52	3,126.08
p10											-3,606.02	-480.08	244.30	1,386.08	1,839.76	2,059.14	2,432.70
p11												3,125.94	3,850.32	4,992.10	5,445.78	5,665.16	6,038.72
p12													724.38	1,866.16	2,319.84	2,539.22	2,912.78
p13														1,141.78	1,595.46	1,814.84	2,188.40
p14															453.68	673.06	1,046.62
p15																219.38	592.94
p16																	373.56
p17																	

Results from AIC and BF tests (Tab. 4.4 and 4.5) were straightforward in distinguishing between 4by4 and codon models: all partitioning schemes implementing the M3 codon model (i.e., p14-p17) outperformed those implementing the classical 4by4 analysis (i.e., p01-p13). The AIC test selected p14 as the best model for our dataset (EML=−118,729.10), whereas BF selected p17 (EML=−118,205.79). It has to be noted that codon-based analyses are extremely demanding in terms of computational power: therefore, as detailed in Methods section, we used single MC^3 analyses with half generations with respect to 4by4 models. Four of such analyses were run to estimate convergence within and among runs, and parameters and trees were finally summarized given the convergence evidence. In all cases, we could compute final statistics and consensus tree from 2 runs, with the exception of p17, where we could use only 3,416 generations from a single run, which is an order of magnitude lower than we did for models p14-p16. Therefore, the preference of BF for model p17 could be an effect of the low and different sample size of this specific run; moreover, AIC should be more conservative whenever these concerns are present, in that it accounts for overparametrization in the model by penalizing a high number of free parameters K (see Plazzi and Passamonti, 2010; and reference therein for further details). In conclusion, we regarded to p14 as the most supported tree of our study, which is shown in Figure 4.7.

Five monophyletic clusters with Posterior Probabililty (PP) equal to 1 were obtained, corresponding to the five traditional subclasses. Opponobranchia (here *Nucula* and *Solemya*) were retrieved as monophyletic and basal to the Autobranchia, whose topology was found to be (Palaeoheterodonta + (Anomalodesmata + (Heterodonta + Pteriomorphia))). Nodes are robustly supported along the whole tree, as most have PP=1.00.

Figure 4.7. Bayesian Inference. Shown is p14 tree, computed partitioning our dataset into ribosomal and protein coding genes; these were analyzed using the M3 codon model (see text for details). Values at the nodes are Posterior Probabilities (PP); nodes were collapsed if PP<0.95. Asterisks mark those genera formerly classified among heterodonts, here clustering with pteriomorphians.

Subclass Palaeoheterodonta is divided into two extant orders, Trigonioida and Unionoida. *Cristaria plicata* is basal to remaining Palaeoheterodonta in our tree. A polytomy separates *Lanceolaria grayana*, the genus *Unio*, the genus *Anodonta*, the cluster *Pyganodon* + *Psaudanodonta*, and a cluster with remaining unionids with *Alathyria jacksoni* (family Hyriidae). Therefore, family Unionidae is paraphyletic because of *Alathyria*, subfamily Anodontinae is paraphyletic as well, because of *Cristaria*, and subfamily Unioninae is polyphyletic. On the other hand, subfamily Ambleminae is monophyletic, and 3 out of 4 tribes are represented in our tree: only the tribe Lampsilini is represented with more than one genus (*Epioblasma*, *Lampsilis*, *Venustaconcha*), and it is monophyletic. No specimen from order Trigonioida was included in this study.

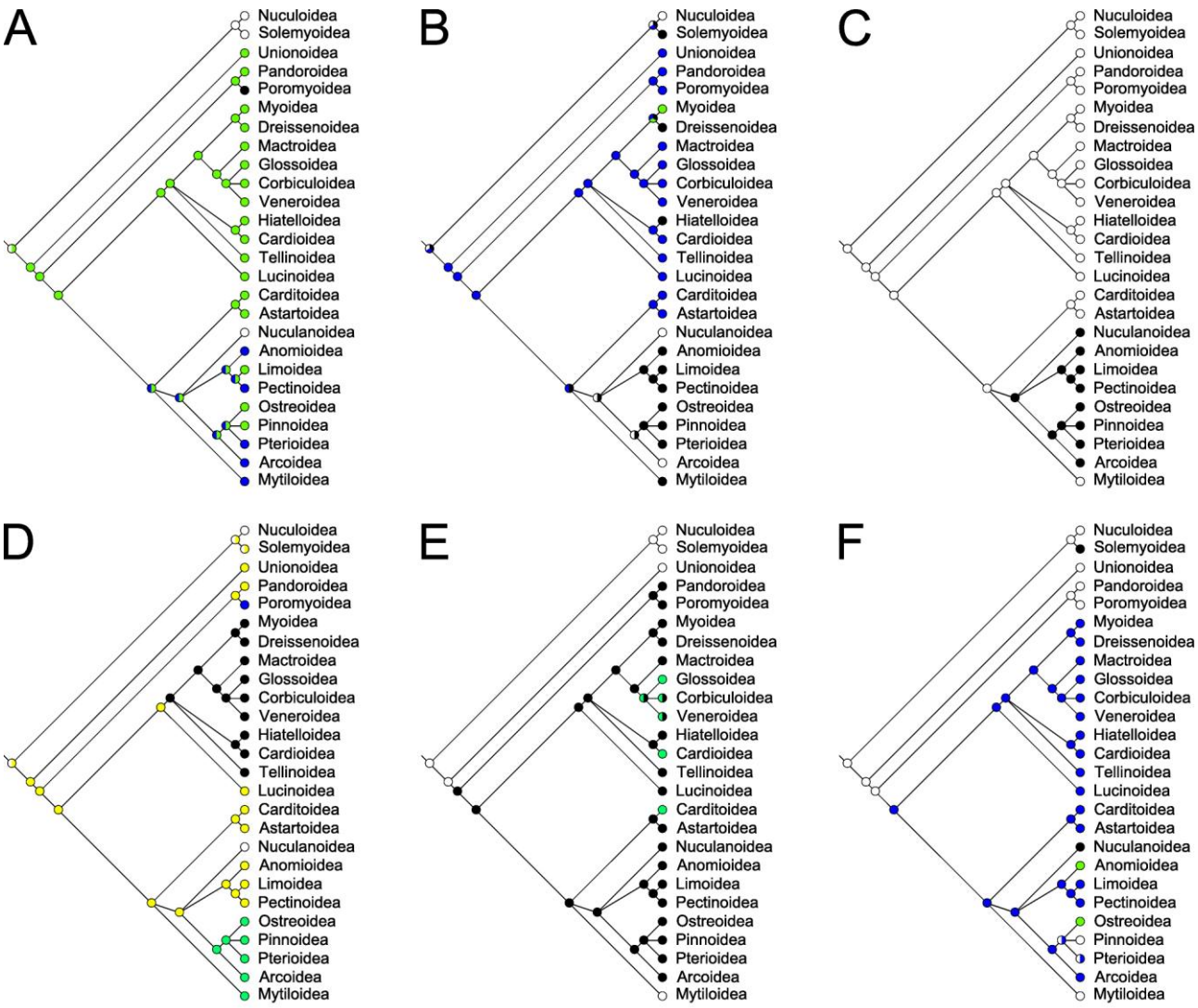
Only one order, Pholadomyoida, belong to subclass Anomalodesmata. Although the subclass is monophyletic, the internal relationships are unresolved. However, *Thracia* and *Pandora* cluster together as sister group of *Cuspidaria* with PP=0.85 in p14 and this relationship is present in all trees, being also supported with PPs>0.95 in some of them. Therefore, a signal, albeit weak, is present for the monophyly of Pandoroidea (suborder Pholadomyina).

Superfamily Lucinoidea (*Loripes* + *Lucinella*) is basal to all remaining Heterodonta. The remaining heterodont taxa (excluding *Astarte* + *Cardita*, see below*)* are arranged as a polytomy separating two big clusters and two small clades, (*Abra* + *Donax*) and (*Ensis* + *Sinonovacula*). The first big cluster can be described as ((Dreissenoidea + Myoidea) + (Mactroidea + (Corbiculoidea + Glossoidea + Veneroidea))). Genera *Dreissena* and *Mactra* are monophyletic, as are families Mactridae and Veneridae. Relationships within venerids are well resolved, and subfamily Tapetinae and Meretricinae are monophyletic; only subfamily Chioninae was not found monophyletic, because of the sister group relationship between *Clausinella* and *Venus*. The second big cluster can be described as (Hiatelloidea + Cardioidea). Subfamily Tridacninae is basal to a polytomy with Fraginae

(*Lunulicardia* + *Corculum*), Laevicardiinae (*Laevicardium*), and a cluster with Cardiinae (*Acanthocardia*) and Cerastodermatiinae (*Cerastoderma*).

Two clades are basal to the core of Pteriomorphia. The first is the cluster (*Astarte* + *Cardita*), which is generally ascribed to Heterodonta as composed by superfamilies Astartoidea and Carditoidea. The second is the monophyletic family of Mytilidae, which is divided in two sister groups: on one side, (Lithophaginae + (Modiolinae + *Mytilaster lineatus*)); on the other side, (*Mytilaster sp.* + (Crenellinae + Mytilinae)). Therefore, neither the subfamily Mytilinae nor the genus *Mytilaster* are monophyletic in this tree. Relationships within the core of Pteriomorphia are well resolved: they are subdivided into three clusters, one of them represented by *Nuculana commutata* alone, which was formerly ascribed to Palaeoheterodonta. The second cluster has *Anomia* as basal to Limoidea and Pectinoidea, both monophyletic superfamilies. Genus *Acesta* is monophyletic and sister group of the cluster (*Lima pacifica galapagensis* + (*Lima sp.* + *Limaria sp.*)), therefore genus *Lima* is paraphyletic. *Spondylus* (family Spondylidae) and *Parvamussium* (family Propeamussiidae) are basal to a heterogeneous clade of intermingled Pectinidae and Propeamussiidae (*Adamussium*, *Amusium*), where many lower taxa are found as polyphyletic: Chlamydinae, Pectininae, genus *Chlamys*. Conversely, subfamily Patinopectininae is monophyletic due to the sister group relationship between *Patinopecten* and *Mizuhopecten*. The third cluster is composed by order Arcida as sister group of (Pteriida + Ostreina). With minor exceptions, like the polyphyly of *Barbatia*, and the paraphyly of Pteriida, Pteriidae, Arcidae, and Arcinae, most taxa were recovered as monophyletic: namely, we could retrieve as highly supported clusters subfamilies Pycnodonteinae, Ostreinae, families Gryphaeidae, Ostreidae, superfamilies Ostreoidea, Arcoidea, subroders Ostreina, Pteriina, Arcina, and order Arcida.

Six morphological characters were traced and optimized on p14 tree: gill type, shell microstructure (Newell, 1965), gill cilia (Atkins, 1936-1938), stomach type (Purchon, 1958), labial palps (Stasek, 1963), and hinge (Giribet and Wheeler, 2002). Parsimony reconstructions of ancestral states are shown in Figure 4.8; ML was also implemented for all those characters where multiple states were not used, and results were in complete agreement with parsimony.



Panel	Character	White	Blue	Green	Yellow	Black
A	Gill Grade	Protobranch	Filibranch	Eulamellibranch		Septibranch
B	Hinge	Taxodont	Heterodont	Desmodont		Edentate
C	Gill Cilia	Type 1				Type 2
D	Stomach Type	Type 1	Type 2	Type 3	Type 4	Type 5
E	Labial Palps	Type 1		Type 2		Type 3
F	Shell Microstructure	Nacreous	Cross Lamellar	Foliate		Homogeneous

Figure 4.8. Optimization of six major morphological characters on bivalve phylogeny as presented in Figure 4.7. Each tree shows the parsimony reconstruction of ancestral state given the p14 topology and a matrix of morphological characters compiled following Newell (1965) and Giribet and Wheeler (2002); see text for more detail. A, gill grade; B, hinge; C, gill cilia; D, stomach type; E, labial palps; F, shell microstructure.

4.4. DISCUSSION

Phylogenetic signal

All the evidence we gathered from our dataset points towards the conclusion that abundant phylogenetic signal is available through the combined use of these four mitochondrial markers, but it is absolutely not trivial to detect it correctly.

This is expected because of the depth of this study: bivalves arose 530 million years ago (Mya), in the earliest Cambrian (Brasier and Hewitt, 1978; Morton, 1996; Plazzi and Passamonti, 2010; and reference therein). The saturation profile (see Fig. 4.2) is compatible with the old age of the class; repeated substitution events at the same site (multiple hits) were possible, which is exactly what it is expected from the old age of the class. Nevertheless, given the proximity of %Ti values to the threshold 50% value and, above all, the stability of the pattern, irrespective of sequence divergence and gene/site properties, we may conclude that the use of complex evolutionary models should account for the minor saturation occurred in the four analyzed genes.

This is further demonstrated by neighbornet networks and spectral analysis (see Fig. 4.3 and 4.4): evidence of monophyly were found for all the major groups of bivalve systematics, with special reference to pteriomorph radiation. Some groups appear to be particularly well-defined in our dataset, like Ostreidae, Unionidae, and Veneridae. Even in these cases, however, networks retain some star-likeness and no binary splits at all were found in spectral analysis shown in Fig. 4.4, which are clear indications that some noise is anyway present, and has to be treated with more complex phylogenetic analyses. The method of Likelihood Mapping implements precise and statistically tested evolutionary models, which are able to account for multiple hits along genes and for rate mutation heterogeneity. Indeed, the use of Likelihood Mapping simplex could finally demonstrate

the presence of strong phylogenetic signal in our dataset (see Fig. 4.5A) and also the evidence of one or two preferred topologies (see Fig. 4.5B).

In facts, it is sound and conservative to conclude that our dataset has a high resolving power, but the deeper is an evolutionary relationship, the more refined is expected to be a technique to unveil and exploit it. This is especially the case for the general backbone of bivalve tree, which had to be targeted with advanced BI. In this study, as in our previous preliminary analysis (Plazzi and Passamonti, 2010), selected models tend to merge in a single partition (i.e. ribosomal genes on one side and PCGs on the other), indicating that this is most likely the best trade-off between a detailed, as realistic as possible model, and overparametrization.

Bivalve phylogeny

The p14 Bayesian tree was very well resolved; the high number of taxa it included makes possible to address many evolutionary issues about bivalves.

The Opponobranchia were confirmed as separated to all Autobranchia; the reduced length of branches leading to Nuculoidea and Solemyiodea constitutes an evidence that these species tend to retain most ancestral characters, as widely hypothesized (see, f.i., Yonge, 1939; Morton and Yonge, 1964; Morton, 1996; and reference therein).

Palaeoheterodonta are confirmed to be the sister group of all remaining Autobranchia, as resulted from our previous study (Plazzi and Passamonti, 2010). This is not in agreement with other molecular and morphological studies (Waller, 1990, 1998; Giribet and Wheeler, 2002; Bieler and Mikkelsen, 2006; Giribet, 2008), which considered Palaeoheterodonta more related to Heterodonta than to Pteriomorphia, erecting a monophyletic group called Heteroconchia. However, other molecular studies retrieved Palaeoheterodonta as basal to (Heterodonta + Pteriomorphia): Canapa et al. (1999) obtained this result on the basis of the *18s* nuclear gene, whereas Giribet and Distel

(2003) used a big dataset and four molecular markers (*18s*, *28s*, *cox1*, and histone H3). Actually, it is unclear why Giribet (2008) preferred the Heteroconchia hypothesis when his most recent work was not supporting it (Giribet and Distel, 2003). Moreover, a very recent study exploiting complete mitochondrial genomes obtained Palaeoheterodonta to be basal to remaining Autobranchia (Doucet-Beaupré et al., 2010). Interestingly, the same relationship has been proposed also on morphological grounds: Cope (1996), for instance, showed that parsimonious analysis of shell microstructural types led to similar conclusions.

We here contend the monophyly of Heteroconchia *sensu* Giribet (2008) and therefore we propose the taxon “Amarsipobranchia” for the clade comprising Anomalodesmata, Heterodonta, and Pteriomorphia, as it never got a formal name. This term derives from the Greek “marsipos” (μάρσιπος) for “pouch” and means “gills not inserted into a pouch”, in reference to the relationships between anterior filaments of the inner demibranch and the oral groove. In Nuculoidea, Solemyidae, Unionoidea, and possibly Trigonioidea at least the first few anterior filaments are inserted unfused into a distal oral groove, whereas in other bivalves they are fused or not inserted at all (Yonge, 1939; Stasek, 1963; Newell, 1965; and reference therein). Although this is not a universal feature of all extant Anomalodesmata, Heterodonta, and Pteriomorphia (for example, inserted unfused anterior filaments are found also in Mytiloidea and Astartidae), this character has to be considered as a symplesiomorphy of this group and, as such, it is useful for taxonomical purposes (see below and Fig. 4.8D).

Phylogenetic relationships within Palaeoheterodonta are unclear, with special reference to subfamily Unioninae and to the position of family Hyriidae. Possibly, this is also due to the widespread presence of DUI phenomenon among Unionidae, which hampered traditional phylogenetic reconstructions. Therefore, we refer to most recent works on palaeoheterodont evolution (Graf and Ó Foighil, 2000; Roe and Hoeh, 2003; Serb et al., 2003; Huff et al., 2004; and reference therein) and, above all, to the recent

work of Breton et al. (2009) on the DUI-related comparative mitochondrial genomics of freshwater mussels. However, the monophyly of the subclass in not challenged in our study, given the high PP value (1.00) and the length of the branch separating Palaeoheterodonta from their sister group.

Anomalodesmata appear to be basal to Heterodonta and Pteriomorphia. In our previous study (Plazzi and Passamonti, 2010), we obtained anamalodesmatans to be basal to Pteriomorphia, but not monophyletic. In some other studies, anomalodesmatans were found to be a monophyletic clade among Heterodonta (Harper et al., 2000, 2006; Giribet and Wheeler, 2002; Dreyer et al., 2003; Giribet and Distel, 2003; Taylor et al., 2007b) and their subclass status was questioned (Giribet, 2008; and reference therein). Given our mitochondrial dataset, we can here suggest anomalodesmatans as a monophyletic sublclass of Bivalvia, but it is clear that more taxa have to be sampled to completely unravel this point. Anyway, this is confirmed in Giribet and Wheeler (2002). Within the subclass, we could not affordably confirm the sister group relationship between Pholadomyina and Cuspidariina. Actually, they are also very distinguishable from a morphological point of view, given the eulamellibranch gills of Pandoroidea and the septibranch condition of Cuspidariina (Newell, 1965).

As *Astarte* and *Cardita* have been included within Pteriomorphia (see below), the subclass Heterodonta corresponds here to the Euheterodonta *sensu* Giribet and Distel (2003). The basal position is occupied by Lucinoidea, confirming the work of John Taylor and colleagues (Williams et al., 2004; Taylor et al., 2007a; Taylor et al., 2007b; Taylor et al., 2009). Few conclusions can be drawn from this study on Tellinoidea and Donacoidea *sensu* Millard (2001), as the clusters (*Abra* + *Donax*) and (*Ensis* + *Sinonovacula*) were not completely resolved in p14 tree. Generally speaking, we tentatively recommend a superfamily Tellinoidea comprising Psammobiidae, Semelidae, and Donacidae, as proposed by Vokes (1980). Our tree shows three more big clusters of Heterodonta, which

could correspond to three orders. An order Cardiida *sensu novo* would contain Hiatelloidea as sister group of Cardioidea, whose only family here represented is the family Cardiidae. Subfamily Tridacninae is basal to remaining subfamilies (Fragine, Laevicardiinae, Cardiinae, Cerastodermatiinae), confirming recent studies on cardiids evolution (Maruyama et al., 1998; Schneider and Ó Foighil, 1999; Kirkendale, 2009; and reference therein). We retrieved the monophyletic group that Taylor et al. (2007b) called Neoheterodontei; we recommend the definition of two sister orders Myida and Veneroida *sensu novo*, which are represented here as (Myoidea + Dreissenoidea) and (Mactroidea + (Glossoidea + Corbiculoidea + Veneroidea), respectively. The subfamiliar taxonomy of Veneridae is probably to assess further, as already suggested by Kappner and Bieler (2006) and Taylor et al. (2007b).

Pteriomorphia are robustly monophyletic in our analysis, as repeatedly demonstrated (Steiner and Hammer, 2000; Matsumoto, 2003); in this study, however, we present the unexpected result of the inclusion of *Astarte* cfr. c*astanea* and *Cardita variegata* within this subclass as sister species. This cluster is consistent with previous molecular and morphological works (Healy, 1995; Giribet and Wheeler, 2002; Giribet and Wheeler, 2003; Taylor et al., 2007b). Superfamilies Astartoidea, Carditoidea, as well as Crassatelloidea, have generally been regarded as the most primitive heterodonts (Campbell, 2000; Park and Ó Foighil, 2000; Giribet and Wheeler, 2002; Giribet and Distel, 2003), but also different positions have been proposed (Yonge, 1969; Purchon, 1987). Specifically, Giribet and Distel (2003) also proposed Carditoidea (including *Astarte castanea*) and Crassatelloidea to be the sister group of Nuculanoidea. This is not confirmed since in our study *Nuculana commutata* is among basal Pteriomorphia (see also Giribet and Wheeler, 2002; Giribet and Wheeler, 2003), which is commonly accepted nowadays (Bieler and Mikkelsen, 2006; Giribet, 2008). All phylogenetic hypotheses about Carditoidea, Astartoidea, and Crassatelloidea (that unfortunately is not represented here) agree about

their primitive status: if the position obtained for this study will be confirmed with enlarged taxon sampling and more markers, this would lead to completely reconsider the interpretation of classical morphological characters for bivalve systematics. We prefer the ordinal name Carditoida *sensu* Bieler and Mikkelsen (2006) to indicate this clade, even if they essentially correspond to the Archiheterodonta *sensu* Taylor et al. (2007b), because this name could lead to confusion if this topology will be confirmed.

Deeper inside the pteriomorphian clade, the basal position of Mytilidae is not new, as shown by Waller (1998), Carter et al. (2000), Steiner and Hammer (2000), Giribet and Wheeler (2002), and Matsumoto (2003) with morphology and molecules (but see Cope, 1996; Morton, 1996). We also agree with Distel (2000) who found some concerns about the monophyly of some subfamilies of Mytilidae, namely Mytilinae and Modiolinae. We also note that the well known, even if not universally accepted, classification of Ostreina and Pectinina as suborders of the order Ostreoida is no longer sustainable, as already noted by Canapa et al. (1999), nor is the order Pterioida *sensu* Vokes (1980). We propose to erect an order Nuculanoida for the only superfamily Nuculanoidea (see above) and then to regard to pteriomorph systematics in terms of two big clusters. In the first, Anomioidea are basal to Limida *sensu* Millard (2001) as sister group to Pectinoidea, comprising Spondylidae, Propeamussiidae, and Pectinidae in our tree, although further investigations are deserved here, with special reference to Anomiidae (traditionally classified as Pectinina) and pectinid relationships (see, f.i., Puslednik and Serb, 2008). For instance, we suggest to consider an order Pectinida *sensu novo* which would include Anomioidea, Limoidea and Pectinoidea for what concerns our tree. In the second cluster, we individuate on p14 tree the group (Arcida + (Pinnina + Pteriina + Ostreoida *sensu novo*); this leaves unresolved the relationships within the order Pteriida, and it would exclude the possibility to elevate the suborder Pinnina *sensu* Millard (2001) to the ordinal rank. In such scenario about pteriomorph evolution, Arcida would occupy a somewhat different position with

respect to results of Distel (2000) and Steiner and Hammer (2000), albeit maintaining their basal condition.

Finally, *Striarca lactea* has been generally classified as member of the subfamily Striarcinae within family Noetiidae; however, several authors have also appraised both subfamilies Striarcinae and Noetiinae as members of the family Arcidae (Reinhart, 1935; Rost, 1955; Myra Keen, 1971), which would render Arcidae monophyletic in our tree. Moreover, genus *Asperarca* Sacco, 1898 has been occasionally considered as a synonym of *Barbatia* Gray, 1840 (see, f.i., Millard, 2011; but see also Vokes, 1980; La Perna, 1998), which would render genus *Barbatia* paraphyletic in our tree.

Tracing and optimizing major morphological characters on the evolutionary tree

Given the phylogenetic reconstruction we discussed above, the major morphological features of bivalve shell and soft parts should be re-evaluated.

Quite surprisingly, the two most used characters for bivalve taxonomy, i.e. gills and shell hinge, do not follow the evolutionary scenarios commonly accepted so far. Protobranch gills (true ctenidia) should be considered the ancestral state among Bivalvia; this is not surprising since most mollusks do have true ctenidia. The question is more puzzling when the “feeding gill” arose among Autobranchia: commonly the filibranch gill has been considered as ancestral, while the eulamellibranch one as derived. The situation, according to our tree, should be exactly the opposite: eulamellibranch gills appear to be the plesiomorphic (ancestral) state in Autobranchia (see Fig. 4.8A).

This is mainly due to the fact that all palaeoeterodonts and most anomalodesmatans, the two groups that arose first among Autobranchia according to our tree, do have an eulamellibranchiate condition (except some anomalodesmatans, which are derived septibranchs). If we accept this, then the filibranch condition of pteriomorphians seems to have evolved from an eulamellibranchiate one. Moreover, according to our tree, the

filibranch condition might be occurred at least five times among Pteriomoprhia (Anomioidea, Pectinoidea, Pterioidea, Arcoidea, and Mytiloidea), but there are three unresolved tritomies in this portion of the tree and a better resolution could result in a more parsimonious reconstruction of filibranch condition. Finally, even more surprisingly, the eulamellibranch condition seems to have reverted to the ancestral protobranchiate state in the superfamily Nuculanoidea. Of course, more studies are needed to better fit gills morphology and molecular phylogeny; nevertheless, it has to be noted that what we commonly call protobranch, filibranch or eulamellibranch gills might be artifactual assemblies of different gills types, and maybe this unexpected results might trigger further morphological studies on gills anatomy.

Similarly to gills, the heterodont hinge (once considered more derived) seems to be again the basal condition of Autobranchia (Fig. 4.8B), so that Nuculanoidea and Arcoidea independently evolved their own taxodont hinge: therefore, taxodont hinges of *Nucula*, *Nuculana*, and arks should not be considered as homologous characters. Teeth were lost in four cases: Solemyoidea, Dreissenoidea, Hiatelloidea, and all Pteriomorphia, with the exception of Astartoidea and Carditoidea which retained the ancestral condition of Autobranchia (heterodont hinge). This, as above, needs further studies, once again because different kind of hinges of different origin might possibly hide under the terms heterodont, taxodont and edentate.

On the other hand, the other characters we investigated (gill cilia, stomach type, labial palps and shell microstructure) fit better in the proposed phylogeny. F.i., Type 1 gill cilia are the plesiomorphic condition among bivalves, while Type 2 arose only once in a pteriomorphian clade, excluding Carditoidea+Astaroidea and Mytiloidea, which are therefore supported as basal among pteriomorphians (Fig. 4.8C). Stomach type (Fig. 4.8D) again follow quite well the obtained tree and only Type 3 stomach seems to appear twice independently. Labial palps of Type 1 are shared between Opponobranchia and

Palaeoheterodonta, thus supporting the basal condition of the latter. Labial palps type 3 *sensu* Stasek (1963) are symplesiomorphic for Amarsipobranchia (Fig. 4.8E), and they mutated into type 2 in three lineages: Cardioidea, Carditoidea, and Veneroida. Finally nacreous shell microstructure (Fig. 4.8F) seems to be the ancestral state of all Bivalvia, while cross lamellar shells appeared once at the arose of Amarsipobranchia.

4.5. Conclusions and Final Remarks

The phylogenetic hypothesis on bivalve evolution we extensively described in the previous paragraph is shown in Figure 4.9. Its major outcomes and new proposals are: i) mitochondrial genomes are highly informative for bivalve phylogeny, given a proper phylogenetic approach; ii) the basal subdivision in Opponobranchia and Autobranchia is confirmed; iii) Palaeoheterodonta were retrieved as sister group of a cluster comprising all remaining Autobranchia, which we propose to term Amarsipobranchia; iv) Anomalodesmata are monophyletic and maintain a basal status among Amarsipobranchia; v) three ordinal categories are proposed, namely Cardiida (Hiatelloidea and Cardioidea), Carditoida (Astartoidea and Carditoidea), and Pectinida (Anomioidea, Limoidea, and Pectinoidea); finally, vi) the heterodont hinge and eulamellibranch gills may be re-interpreted as ancestral character states in Autobranchia, and a revision of gill and hinge structures and evolution should be undertaken.

Further improvements of the present work will increase the available dataset either by exploiting more mitochondrial (or even nuclear) markers or by further enlarging the sample, with special reference to some underrepresented groups: the investigation of deep bivalve phylogeny is as just as started. Moreover, in our study, morphological characters and molecular phylogenies are generally in agreement, but sometimes do not. This is not surprising, being different kind of data under different kind of evolutionary histories. Nevertheless, an effort should be taken to better fit both kind of data in Bivalvia, and more integrated work is needed. Incidentally, the different evolutionary histories of morphological and molecular data (which are even different among genes, so that we need partitions) should advice against their use in the same phylogenetic reconstruction, as in the “total evidence” trees; however, results from either must be repeatedly compared back and forth to eventually gain a better resolution of the bivalves’ evolutionary tree.

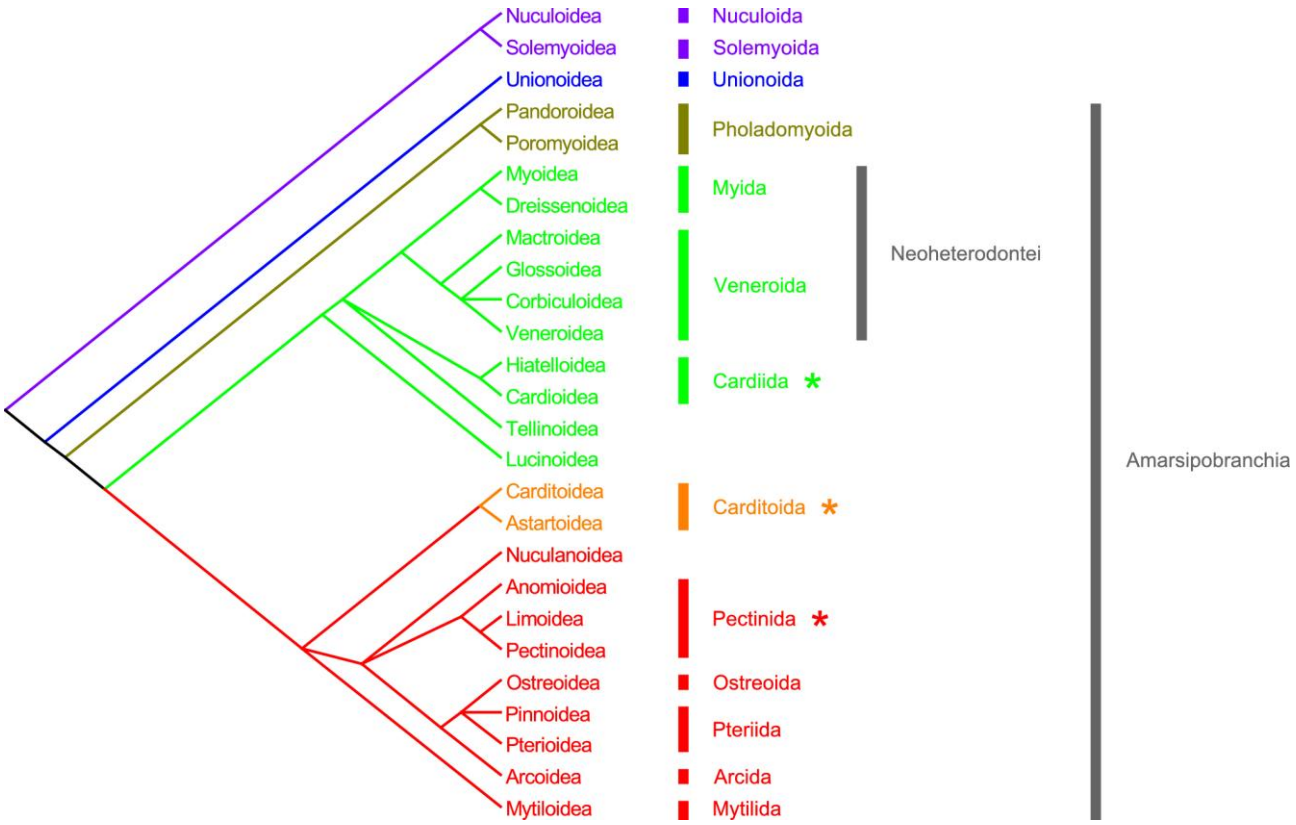


Figure 4.9. Revision of bivalve phylogeny and systematics on molecular mitochondrial bases proposed in this paper (see text for details). Superfamilial relationships are shown, with proposed ordinal classification; for anomalodesmatans, we used the nomenclature from Newell (1965) and Vokes (1980). Color codes as in Figures 4.6 and 4.7. Asterisks mark newly-proposed ordinal categories; Neoheterodontei *sensu* Taylor et al. (2007b) and Amarsipobranchia are also shown.

CHAPTER 5

A TWO-STEPS BAYESIAN PHYLOGENETIC APPROACH TO THE MONOPHYLY OF CLASS BIVALVIA (MOLLUSCA)

5.1. INTRODUCTION

One of the major challenges in bivalve phylogenetics is the apparent polyphyly of the class in many molecular analyses. This problem does not exist for morphology-based analyses, because bivalves share several autapomorphies. The unique features of Bivalvia hamper the comparison with any given molluscan outgroup, to fix ancestral character states, but conversely the monophyly of this clade as a class is generally not questioned (Scheltema, 1993; von Salvini-Plawen and Steiner, 1996; Haszprunar, 2000; Giribet, 2008). Its distinctive traits are well-known: lateral compression of the body, bivalve shell and its annexes (hinge, teeth, and ligament), reduction of head and loss of radula, modified gills for filter feeding (exception made for protobranchs), and byssus gland (Brusca and Brusca, 2003). After two decades of molecular bivalve phylogenetics, many evolutionary relationships within Bivalvia were thoroughly investigated (Giribet, 2008; Plazzi and Passamonti, 2010; and reference therein); however, concerns about the validity of the whole class came unexpectedly to light.

First molecular studies on bivalve phylogeny were mainly based on *18s* rDNA and retrieved the class as polyphyletic. Different bivalve taxa were involved in those studies, as well as different molluscan outgroups: the commonest flaw was a relationship of some veneroid genera (*Arctica*, *Mactromeris*, *Mulinia*, *Phaxas*) and/or the oyster *Crassostrea* to some gastropods (Steiner and Müller, 1996; Winnepenninckx et al., 1996; Passamaneck et al., 2004). Some Anomalodesmata (*Cuspidaria* and *Periploma*) were also linked to

gastropods in the work of Adamkewicz et al. (1997); furthermore, chitons (Polyplacophora) were often intermingled with bivalves to some extent (Winnepenninckx et al., 1996; Canapa et al., 1999; Passamaneck et al., 2004). Thus, the polyphyly of bivalves emerged under variable – and unstable – topologies. Giribet and Carranza (1999) and Steiner (1999) concluded that outgroup choice and questionable taxon sampling is the most likely causes for an artifactual polyphyly of bivalves, finding some monophyly signal for the first time (see also Canapa et al., 1999). In fact, most of those pioneering studies (Steiner and Müller, 1996; Winnepenninckx et al., 1996; Canapa et al., 1999; Giribet and Carranza, 1999; Steiner, 1999) lacked samples from protobranchiate bivalves, like *Nucula* and *Solemya*, which are universally regarded as the most primitive bivalves.

Steiner (1999) stated that the “watershed of new sequences including Protobranchia has not led to better support of bivalve monophyly” and that we “will probably have to cope with the interpretation of little-supported nodes to resolve bivalve phylogeny”. As written above, the latter statement did not come true, but, ironically, it was exactly the availability of sequence from Protobranchia that hindered the monophyly of the class in subsequent, more comprehensive studies (see also Adamkewicz et al., 1997; Passamaneck et al., 2004). The direct optimization study of Giribet and Wheeler (2002) is based on three genes – *18s*, *28s*, and *cox1* – and protobranchiate bivalves cluster with a heterogeneous assemblage of several mollusks (*Antalis*, *Rhabdus*, *Peltodoris*, *Nautilus*, *Loligo*, *Sepia*), whereas all remaining bivalves are supported as a monophyletic group. The following one-step study of Giribet and Distel (2003) yielded very similar results: one more gene was added (*h3*), but the position of genera *Solemya*, *Acila*, and *Nucula* remained essentially unchanged.

The five-gene analysis of Giribet et al. (2006) put a step forward in mollusk phylogeny by inserting for the first time sequence data from Monoplacophora and proposing the “Serialia hypothesis” (Monoplacophora + Polyplacophora); however,

bivalves were retrieved again as paraphyletic. The Heteroconchia *sensu* Bieler and Mikkelsen (2006; Heterodonta + Palaeoheterodonta) were the sister group of a big clade composed by part of Gastropoda, Serialia, and remaining bivalves (Pteriomorphia and protobranchiate species); interestingly, such a diphyletic pattern was already suggested ten years before by Winnepenninckx et al. (1996).

Finally, a monophyly of Bivalvia was firstly found by Wilson et al. (2010), who reported results from both one-step and two-steps phylogenetic analyses, by means of five molecular markers (*18s*, *28s*, *cox1*, *h3*, and *16s*). They included 24 bivalves species in their study, and the protobranchiate species *Nucula sulcata* and *Solemya velum* were sampled. The monophyly of bivalves was also an outcome of the phylogenomic analysis of Doucet-Beaupré et al. (2010), who used 12 protein-coding genes from complete mitochondrial genomes of 29 bivalve species.

In sum, after twenty years of contradictory results, bivalve monophyly was firmly supported from a molecular point of view only in these two recent studies. However, the study of Wilson et al. (2010) mainly focused on molluscan phylogeny, as the assessment of the Serialia hypothesis was the first target of that work, and only 24 taxa out of 109 (~22%) were bivalves. On the other side, Doucet-Beaupré et al. (2010) investigated bivalve phylogeny in the very peculiar context of an exception to the strictly maternal inheritance of mitochondria known as DUI (Doubly Uniparental Inheritance; Skibinski et al., 1994a, 1994b; Zouros et al., 1994a, 1994b): therefore, their taxon sampling was obviously biased towards those bivalves featuring this mechanism, and, for instance, no protobranchiate bivalve was included.

Aim of this study is to rigorously address the issue of bivalve monophyly/polyphyly, following the methodological pipeline we presented in our previous paper (Plazzi and Passamonti, 2010) to obtain (i) a robust two-steps phylogeny of mollusks, with special

reference to bivalves, and (ii) a model-decision framework to evaluate alternative topologies, by means of Bayes Factors (Kass and Raftery, 1995).

5.2. MATERIALS AND METHODS

Assembling the dataset

The first step consisted in the choice of markers and the set up of dataset. Sequences from at least one representative of each mollusk class, as well as of relevant protostome outgroups, are to date (December 2010) available in GenBank only for the large mitochondrial ribosomal subunit (*16s*), the subunit I of the cytochrome oxidase c (*cox1*), and the histone H3 (*h3*). Therefore, we selected those taxa for which all these three genes were present in GenBank, as we decided to minimize the amount of missing data in our alignments with respect to our previous paper on bivalve phylogeny (Plazzi and Passamonti, 2010). The CLC Sequence Viewer 6.4 (CLC bio, Aarhus, Denmark) environment was used to download, manage, and organize sequences we obtained from GenBank; they were arranged in three separate datasets. Suitable taxa were filtered, cross-linked and evidenced with Microsoft Excel® functions. When necessary, sequences of different congeneric species were joined together to increase coverage: this does not lead to inconsistent results at elevated phylogenetic depth, as is a phylum (see, f.i., Plazzi and Passamonti, 2010; Li et al., 2009). Seven outgroups were selected for this study: *Lumbricus terrestris* (Annelida, Oligochaeta), *Paranemertes peregrina* (Nemertea), *Platynereis dumerilii* (Annelida, Polychaeta), *Sipunculus nudus* (Sipuncula), *Symsagittifera roscoffensis* (Platyhelminthes), *Terebratulina retusa* (Brachiopoda), and *Urechis caupo* (Echiura). All sequences used for this study are listed in Appendix 5.1 with their GenBank Accession Number.

Alignments

Alignments were aligned with ClustalW (Thompson et al., 1994) at the EBI server (http://www.ebi.ac.uk/Tools/msa/clustalw2/; Chenna et al., 2003). For *16s* gene, the IUB

matrix was used with a 25 penalty for gap opening and a 5 penalty for gap extension, whereas for both protein-coding genes (PCGs), penalties were set to 50 and 10, respectively.

GBlocks software (Talavera and Castresana, 2007; Castresana, 2000) was chosen to cut ambiguously aligned regions from the *16s* alignment. The following parameters were used: minimum number of sequences for a conserved position, 38; minimum number of sequences for a flanking position, 38; maximum number of contiguous nonconserved positions, 50; minimum length of a block, 10; allowed gap positions, all. Gaps were treated as missing data and coded for their absence/presence at the end of nucleotide matrix as binary data, following the simple indel method of Simmons and Ochoterena (2000) as described in Plazzi and Passamonti (2010); this task was carried out with the software GapCoder (Young and Healy, 2003).

Preliminary analyses

Nucleotide substitution saturation was evaluated by plotting the percentage of transitions (%Ti) on corresponding K2P distance values (Roe and Sperling, 2007; Luo et al., 2011). Pairwise transitions/transversions ratios and (Ti/Tv) K2P distances were computed through the program PAUP* 4.0b10 (Swofford, 2002) using PaupUp graphical interface (Calendini and Martin, 2005). Ti/Tv ratio was obtained from the absolute number of differences, transformed to %Ti, and plotted against pairwise K2P distances. %Ti was considered low less than 50% (Ti/Tv ratio ≤ 1; Roe and Sperling, 2007). The saturation test was conducted independently for the three markers and, about PCGs, for third codon positions only, with the aim of spotting out and eliminating particularly saturated markers.

Neighbornet networks were constructed to visually inspect properties of phylogenetic signal lying in our dataset (Bryant and Moulton, 2004; Wägele et al., 2009). We used SplitsTree 4.6 (Dress et al., 1996; Huson and Bryant, 2006) to construct networks on

either uncorrected or Log-Det distances. The software TreePuzzle 5.2 (Schmidt et al., 2002; Schmidt and von Haeseler, 2003) was used to perform Likelihood Mapping (LM; Strimmer and von Haeseler, 1996, 1997). Firstly, we performed a classical LM with 5,000 randomly chosen quartets without constraint; then, the same analysis was repeated for the concatenated alignment and for single genes, but taxa were manually sorted in four groups. This technique is called Four-cluster Likelihood Mapping (Strimmer and von Haeseler, 1997); taxa were subdivided into Opponobranchia, Autobranchia, non-bivalve mollusks, and outgroups. In all cases, the best-fitting substitution model was selected with ModelTest 3.7 (Posada and Crandall, 1998) and parameters were given to TreePuzzle to compute the likelihood function. Distribution of quartets was tested for significant divergence from the null hypothesis with a Chi-Square test: the null hypothesis was an even distribution of points in the case of the three corners, while it was computed from empirical data sums in the case of Voronoi cells.

Model decision tests and tree inference

Given the three genes that were used for this study, many different partitioning ways are possible. We decided to directly follow the results we had in a preliminary study on bivalve phylogeny (Plazzi and Passamonti, 2010), which clearly showed two major ways of treating and partitioning data. The first is to limit the number of partitions and parameters by joining together genes with expected similar evolutionary properties (i.e., ribosomal genes, cytochromes, and so on) and to use different models for different codon positions; the second is to thoroughly subdivide the dataset by gene and codon positions. Therefore, we decided to test both models in this work: we subdivided our dataset in 13 different partitions: the large mitochondrial ribosomal subunit gene (*16s*), individual codon positions for the concatenated *cox1* and *h3* genes (*prot_1*, *prot_2*, *prot_3*), individual codon positions for single protein coding genes (PCGs; *cox1_1*, *cox1_2*, *cox1_3*, *h3_1*, *h3_2*,

h3_3), and the corresponding indel characters coded as 0/1, irrespective of codon positions (16s_*indel*, *prot_indel*, *cox1_indel*). Two different schemes (m01 and m02) were tested combining these partitions, as shown in Table 5.1. Evolutionary models to be implemented were selected for each partition with ModelTest 3.7 through the graphical interface provided by MrMTgui (Nuin, 2008); we used the Bayesian Information Criterion (BIC) as model-decision criterion (Luo et al., 2010; and reference therein).

Table 5.1. Partitioning schemes adopted for this study.

Name	Number of partitions									
m01	6	*16s*	*16s_indel*	*prot_1*	*prot_2*	*prot_3*	*prot_indel*			
m02	9	*16s*	*16s_indel*	*cox1_1*	*cox1_2*	*cox1_3*	*cox1_indel*	*h3_1*	*h3_2*	*h3_3*
m03	5	*16s*	*16s_indel*	*cox1*[a]	*cox1_indel*	*h3*[a]				

[a] Analyzed with the M3 codon model; see text for futher details.

A Bayesian Analysis (BA) was carried out for both m01 and m02 with the software MrBayes 3.1.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) hosted at the University of Oslo Bioportal. Parameters were those selected by ModelTest and the default analysis was chosen for restriction data, using the option coding=variable and modeling substitution occurrence with four discrete, gamma-distributed categories. Each run consisted of 10,000,000 generations of two parallel MC^3 analyses with 4 chains each. PSRF (Gelman and Rubin, 1992) and standard deviation of average split frequencies sampled every 1,000 generations were used as proxies for convergence. Trees was sampled every 100 generations and the consensus was computed after burnin removal. Each analysis was repeated using amminoacids instead of nucleotides; in this case, a "glorified" GTR+I+Γ model was used under identical MC^3 settings. Furthermore, we accounted for substitution saturation in our dataset by implementing a codon model (Goldman and Yang, 1994; Muse and Gaut, 1994). In this case (m03), the M3 codon

model was used for PCGs, which were necessarily included in two different partition because of the different translational code; 5,000,000 generations with tree sampling every 125 were run in a single analysis.

The Akaike Information Criterion (AIC; Akaike, 1973) and the Bayes Factor (BF; Kass and Raftery, 1995) were used as described in Plazzi and Passamonti (2010; and reference therein) to select best-fitting models for our dataset, with reference to partitioning strategy and monophyly constraints. In facts, four independent analyses were run for our three models. In the “b” analysis (m01b, m02b, m03b), a constraint was enforced with MrBayes on the monophyly of bivalves, without prior information on general molluscan topology; in the “bm” analysis (m01bm, m02bm, m03bm), both bivalves and mollusks were set to be monophyletic; in the “m” analysis (m01m, m02m, m03m), we fixed all mollusks as monophyletic; in the “u” analysis (m01u, m02u, m03u), no constraint was put on either clade. This method yielded 12 separate trees which were compared via the AIC/BF approach. 8 trees were also produced with amminoacids data sets (m01aab, m01aabm, m01aam, m01aau, m02aab, m02aabm, m02aam, m02aau. Trees were graphically edited by PhyloWidget (Jordan and Piel, 2008) and Dendroscope (Huson et al., 2007).

5.3. RESULTS

Preliminary analyses and phylogenetic signal

The total concatenated alignment was 1,883 bp long; 177 sites of *16s* were removed by GBlocks as ambiguously aligned. GapCoder found 486 valid indels for *16s* and 15 indels for *cox1*. No indel was present in the *h3* alignment. In sum, our alignment was finally composed by 2,207 sites, either nucleotides or binary data.

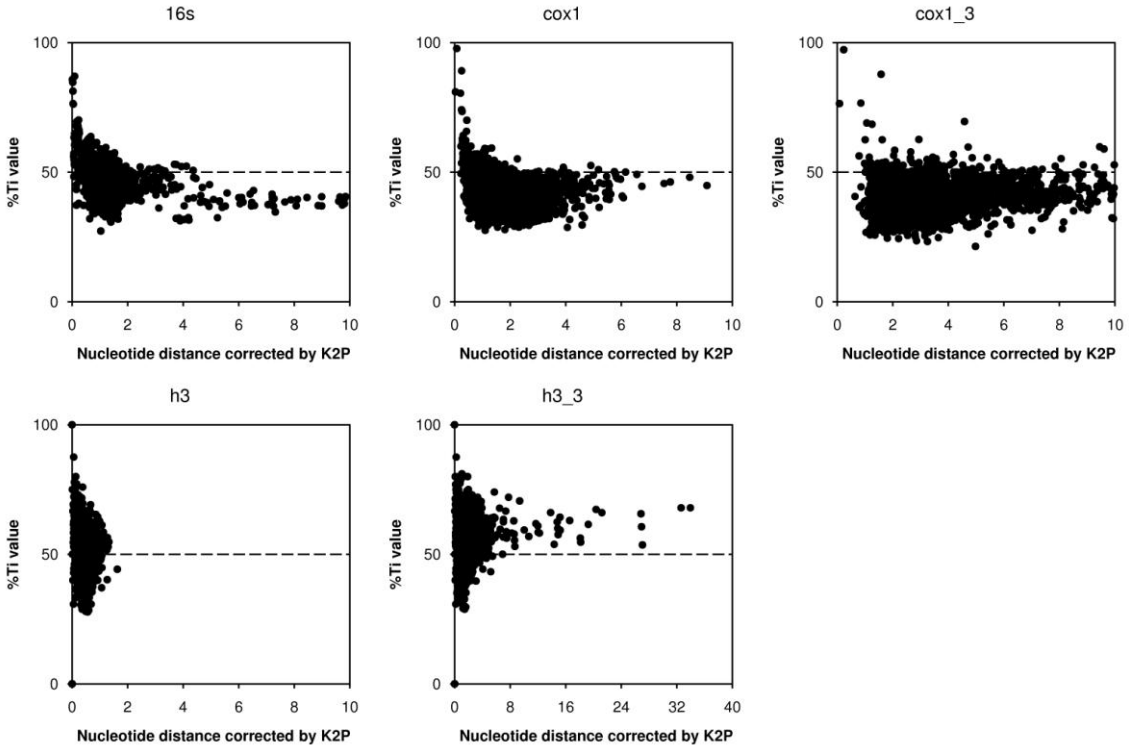


Figure 5.1. Percentage of transitions (%Ti) plotted on K2P distances to estimate saturation in our dataset. The dotted line indicates the 50% threshold for %Ti to be considered low.

Saturation plots are shown in Figure 5.1. Mitochondrial genes exhibit a different pattern with respect to *h3*. In the first two cases, percentage of trasitions tends to be somewhat low (30% ≤ %Ti ≤ 50%), but still stable even for increasing pairwise K2P distance values. For histone H3, a trend in the plots is not evident, but %Ti cloud is higher than for mitochondrial genes. Conversely, K2P distances are clearly smaller for *h3* than for

16s and *cox1*. Overall, these patterns do not change when only third codon positions are considered for PCGs. Therefore, we conclude that total saturation is our dataset is generally low and compatible with the depth of the analysis, which targets a whole phylum.

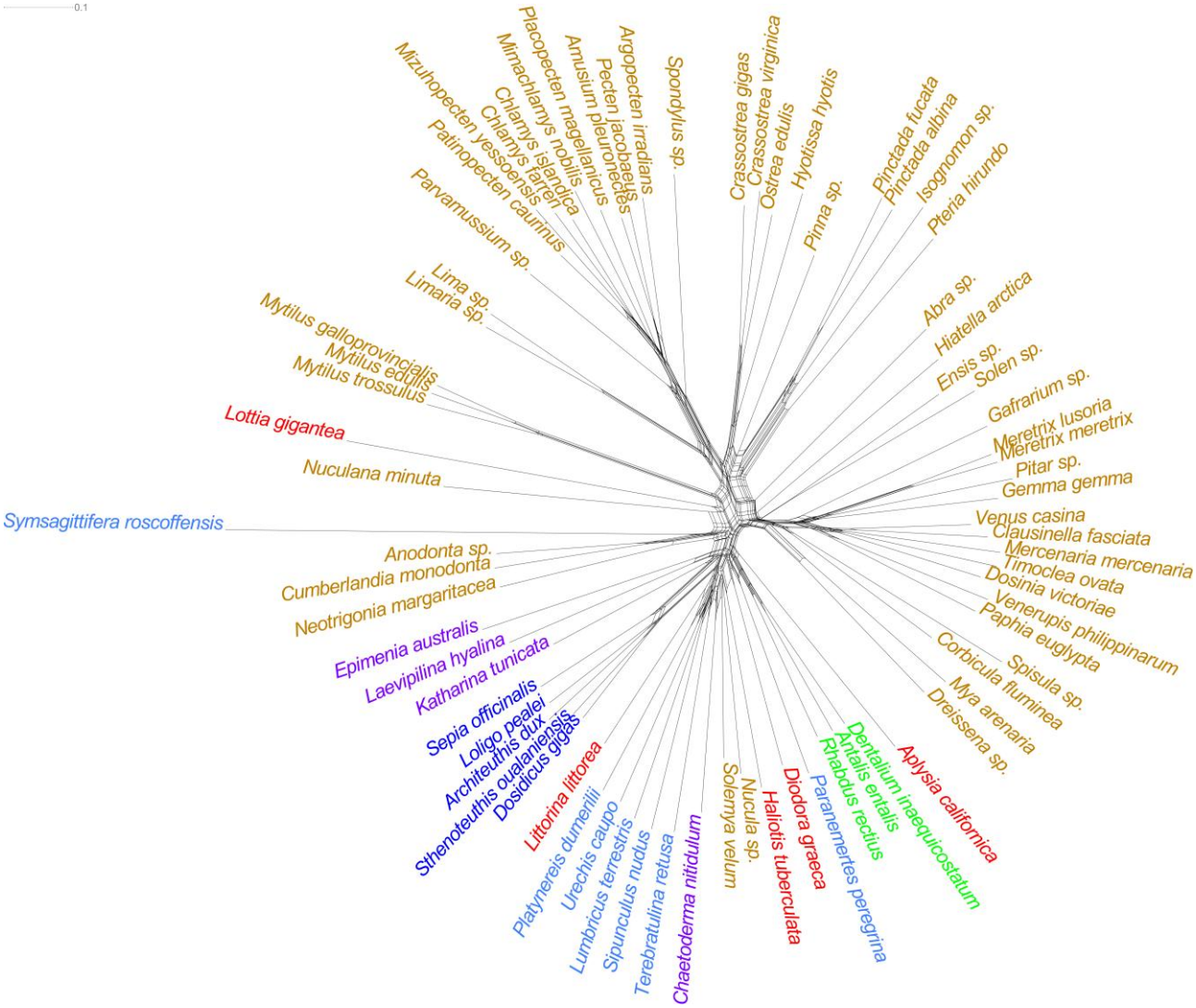


Figure 5.2. Neighbornet network based on LogDet distances computed on the whole dataset. Bivalves are shown in brown; cephalopods, gastropods, and scaphopods are shown in heavy blue, red, and green, respectively; other mollusks are shown in purple; outgroups are shown in light blue.

Neighbornet networks show some signal for bivalve monophyly, with the exception of Opponobranchia. Figure 5.2 shows the LogDet network based on the complete concatenated alignment, with all taxa included. A large portion of the network is occupied by a strong cluster of bivalves, wherein several subgroupings are also distinguishable, like (clockwise from left) mytilids, limids, pectinids, ostreids, pteriids, and venerids. *Nuculana minuta* and Palaeoheterodonta (both Unionida and Trigonioida) cluster beside other

bivalves, intermingled with the gastropod *Lottia gigantea* and *Symsagittifera roscoffensis* (Acoela). More evidently, both *Nucula sp.* and *Solemya velum* cluster distantly from other bivalves, next to *Haliotis tuberculata*, *Chaetoderma nitidulum*, and most outgroups. We could include more than one taxon from other three molluscan classes: Cephalopoda, Scaphopoda, and Gastropoda. While the first two form distinct branches, gastropods are scattered throughout the network. All outgroups cluster together with the exception of *S. roscoffensis* and *Paranemertes peregrina*. The length of branches leading to *S. roscoffensis* and *L. gigantea* could artifactually modify the topology; to address this issue, we decided to exclude these two taxa from the analysis. The reduced neighbornet network (Fig. 5.3) is very similar to the previous one, with the exception that all bivalves cluster together with the only exception of *Nucula* and *Solemya*.

The LM analysis evenly distributed quartets within the simplex and left only 5.3% of them in the central star-like tree area (Fig. 5.4), whereas 91% were distributed among the three corners ($P<0.005$). Four-cluster Likelihood Mapping show a strong preference of the concatenated alignment for the topology ((Opponobranchia + Autobranchia) + (non-bivalve mollusks + outgroups), which would suggest bivalves to be supported as a clade. Single-gene analyses unveiled a preference for different topologies when different genes are considered ($P<0.005$); the signal from *h3* strongly preferred the above topology and is therefore responsible of the concatenated overall result; however, *16s* gene favor ((Opponobranchia + non-bivalve mollusks) + (Autobranchia + outgroups)), whereas *cox1* favor ((Opponobranchia + outgroups) + (Autobranchia + non-bivalve mollusks)).

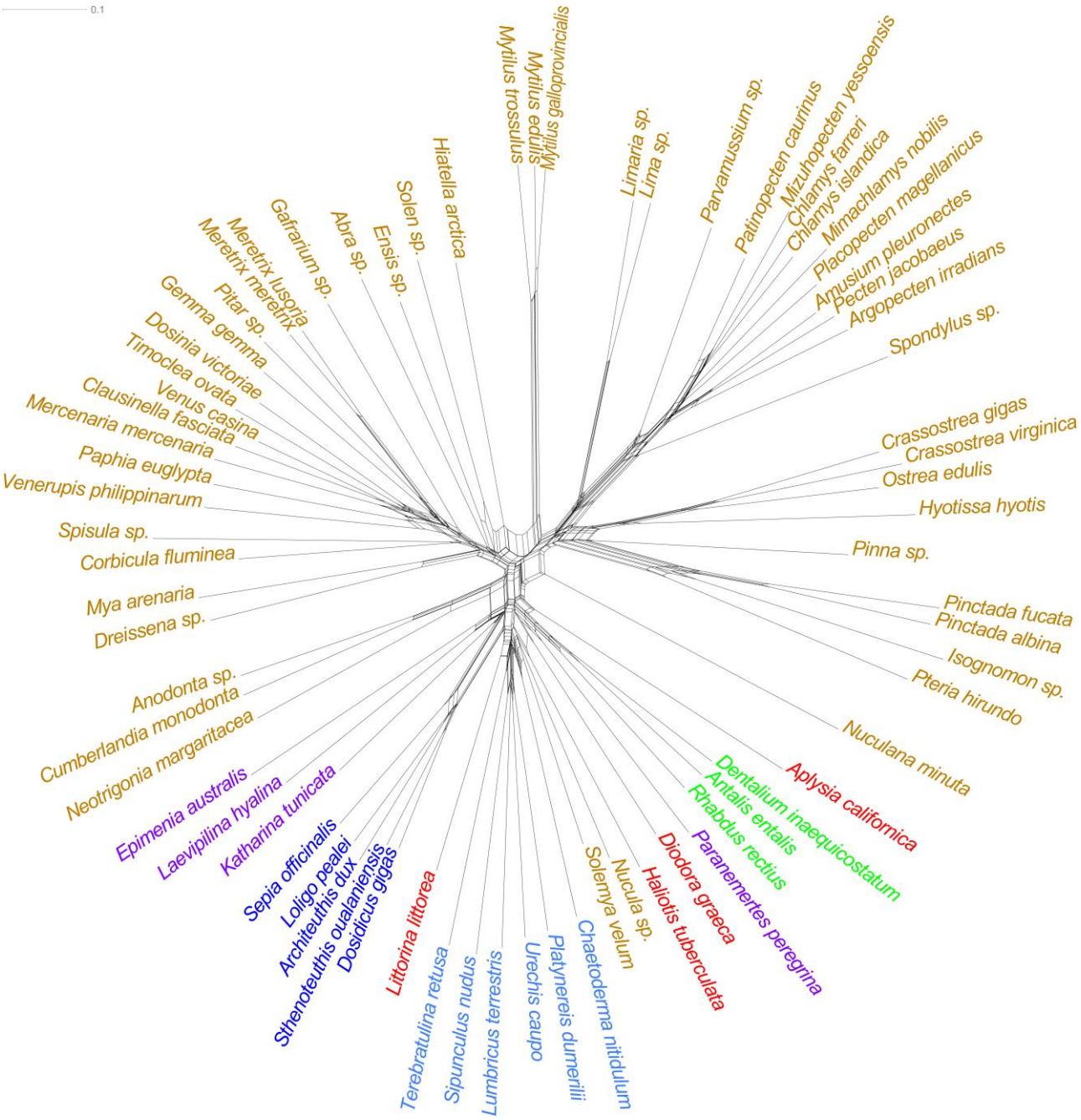


FIGURE 5.3. Neighbornet network based on LogDet distances upon the exclusion of *Symsagittifera roscoffensis* and *Lottia gigantea*. Color code as in Figure 5.2.

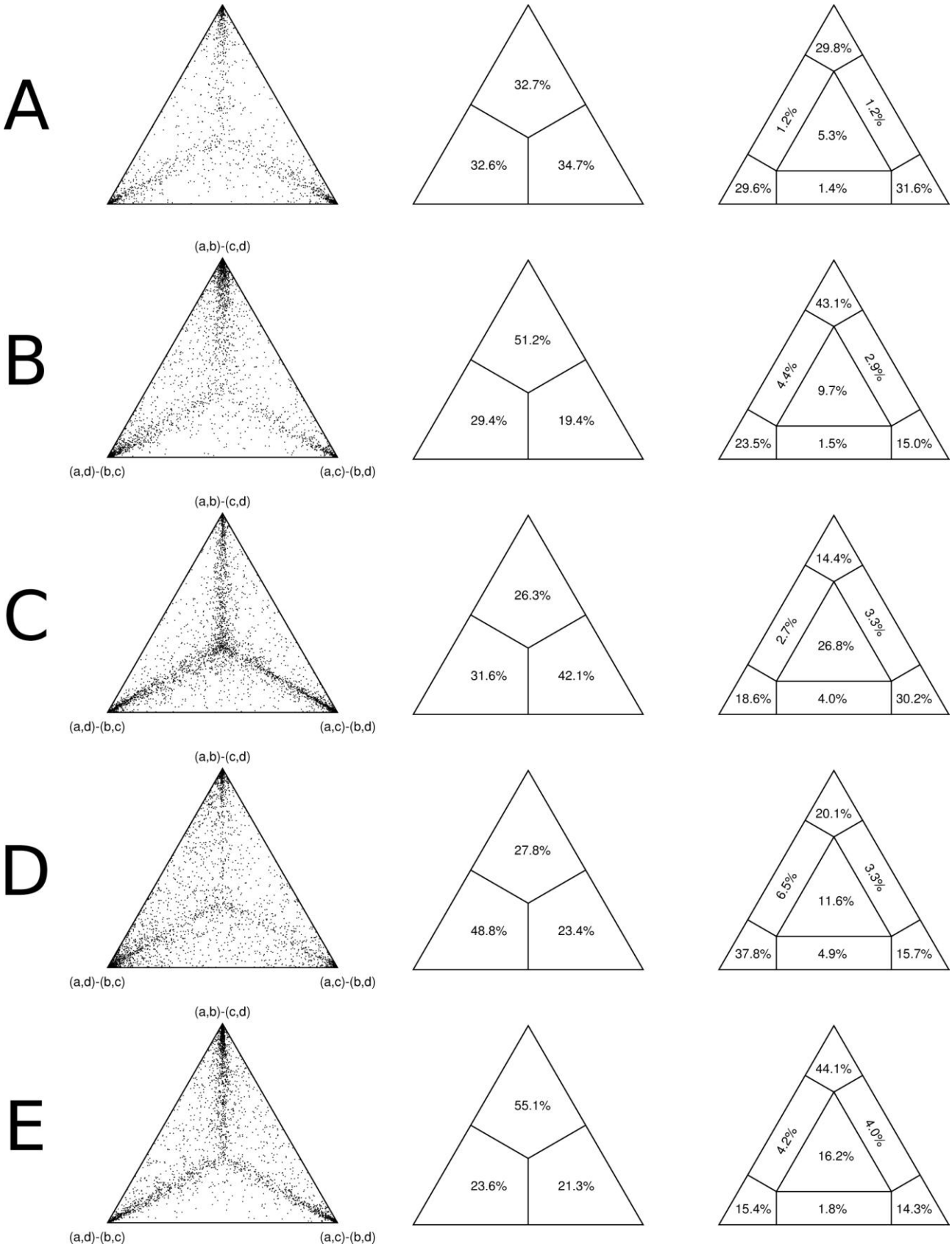


Figure 5.4. Likelihood Mapping of 5,000 random quartets from the complete concatenated dataset (A, B), *16s* (C), *cox1* (D), and *h3* (E) genes. A Four-cluster Likelihood Mapping was performed in all cases with the exception of A. Taxa were subdivided into Opponobranchia (a), Autobranchia (b), other mollusks (c), and outgroups (d). All distributions are significantly different from the null hypotheses ($P<0.005$). See text for more details.

Phylogenetic trees

Results of molecular evolution models for each partition are extensively listed in Appendix 5.2. Table 5.2 shows results from AIC test; Table 5.3 is the BF matrix. The “u” model was always chosen as the best way of treating data for classical (“4by4”) nucleotide and amminoacid analyses: both AIC and BF selected the 4by4 model m02u (EML=−60,521.36), whereas AIC selected m01aau (EML=−35,948.16) and BF selected m02aau (EML=−35,841.88) for amminoacid alignment. However, the M3 m03bm model (EML=−59,130.12) outperformed all M3 and 4by4 nucleotide models, following both AIC and BF statistics. It is not directly comparable with amminoacid analyses, as it starts from different data; however, we previously demonstrated that codon models, and specifically M3, are the best way to cope with bivalve phylogeny (Plazzi and Passamonti, 2010), therefore we regard to m03bm as the best phylogenetic tree obtained for this work (Fig. 5.5).

Table 5.2. Results of Akaike Information Criterion (AIC) test. Partitioning scheme details are listed in Table 5.1. K, number of free parameters used for that model; EML, Estimated Marginal Likelihood as computed by MrBayes 3.1.2; AIC, Akaike Information Criterion statistics.

Model	K	EML	AIC
m01b	926	-61,177.94	124,207.88
m01bm	926	-61,202.27	124,256.54
m01m	926	-61,173.30	124,198.60
m01u	926	-61,145.50	124,143.00
m02b	1,383	-60,546.17	123,858.34
m02bm	1,383	-60,579.41	123,924.82
m02m	1,383	-60,542.11	123,850.22
m02u	1,383	-60,521.36	123,808.72
m03b	891	-59,355.36	120,492.72
m03bm	891	-59,130.12	120,042.24
m03m	891	-59,134.75	120,051.50
m03u	891	-59,351.84	120,485.68
m01aab	812	-35,983.72	73,591.44
m01aabm	812	-36,009.73	73,643.46
m01aam	812	-35,982.33	73,588.66
m01aau	812	-35,948.16	73,520.32
m02aab	1,169	-35,878.05	74,094.10
m02aabm	1,169	-35,898.14	74,134.28
m02aam	1,169	-35,869.21	74,076.42
m02aau	1,169	-35,841.88	74,021.76

Table 5.3. Bayes Factor (BF) results. Partitioning scheme details are listed in Table 5.1; Estimated Marginal Likelihood (EML) values are shown in Table 5.2.

	m01b	m01bm	m01m	m01u	m02b	m02bm	m02m	m02u	m03b	m03bm	m03m	m03u
m01b		-48.66	9.28	64.88	1,263.54	1,197.06	1,271.66	1,313.16	3,645.16	4,095.64	4,086.38	3,652.20
m01bm			57.94	113.54	1,312.20	1,245.72	1,320.32	1,361.82	3,693.82	4,144.30	4,135.04	3,700.86
m01m				55.60	1,254.26	1,187.78	1,262.38	1,303.88	3,635.88	4,086.36	4,077.10	3,642.92
m01u					1,198.66	1,132.18	1,206.78	1,248.28	3,580.28	4,030.76	4,021.50	3,587.32
m02b						-66.48	8.12	49.62	2,381.62	2,832.10	2,822.84	2,388.66
m02bm							74.60	116.10	2,448.10	2,898.58	2,889.32	2,455.14
m02m								41.50	2,373.50	2,823.98	2,814.72	2,380.54
m02u									2,332.00	2,782.48	2,773.22	2,339.04
m03b										450.48	441.22	7.04
m03bm											-9.26	-443.44
m03m												-434.18
m03u												

	m01aab	m01aabm	m01aam	m01aau	m02aab	m02aabm	m02aam	m02aau
m01aab		-52.02	2.78	71.12	211.34	171.16	229.02	283.68
m01aabm			54.8	123.14	263.36	223.18	281.04	335.7
m01aam				68.34	208.56	168.38	226.24	280.9
m01aau					140.22	100.04	157.9	212.56
m02aab						-40.18	17.68	72.34
m02aabm							57.86	112.52
m02aam								54.66
m02aau								

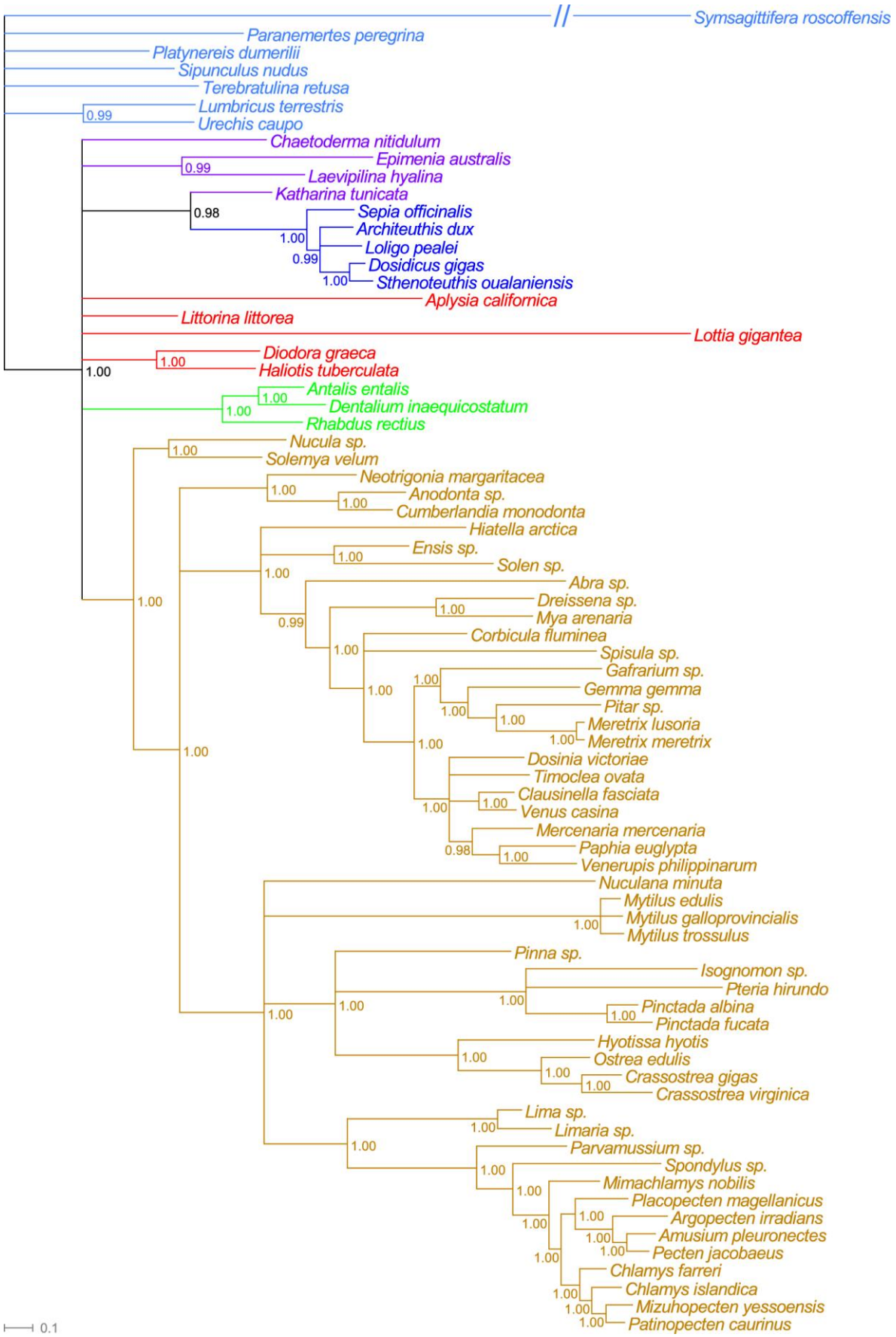


Figure 5.5. The m03u tree as computed via BI using the M3 codon model for *cox1* and *h3* partitions. Nodes with Posterior Probability (PP) <0.95 were collapsed; color code as in Figure 5.2. The long branch leading to *Symsagittifera roscoffensis* was shortened.

In the unconstrained tree m03bm, mollusks and bivalves were forced to be monophyletic. The branch leading to *Symsagittifera roscoffensis* is significantly longer than other branches in the tree. Relationships among outgroup taxa were left unresolved with the exception of the unrealistic cluster (*Lumbricus terrestris* + *Urechis caupo*), which however got a posterior probability (PP) of 0.99. Relationships of major molluscan groups are also unclear: a wide polytomy separates the aplacophoran *Chaetoderma nitidulum*, the cluster (*Laevipilina* + *Epimenia*), a clade (PP=0.98) with *Katharina tunicata* as sister group of monophyletic cephalopods (PP=1.00), the scaphopod lineage (PP=1.00) and gastropods; these were retrieved as paraphyletic, with the only sister group condition of *Diodora graeca* and *Haliotis tuberculata* (PP=1.00).

Nucula sp. and *Solemya velum* were recovered as sister taxa (PP=1.00); this cluster is basal to all Autobranchia. Relationships among Autobranchia are not completely resolved: a tritomy (PP=1.00) separate Palaeoheterodonta, Heterodonta, and Pteriomorphia, all with PP=1.00. Within Palaeoheterodonta, *Neotrigonia* is basal to unionids. Within Heterodonta, *Hiatella* and (*Ensis* + *Solen*) are basal to *Abra* as sister taxon of all remaining heterodonts. Detailed sister group conditions among *Corbicula fluminea*, *Spisula*, and venerids are not supported; moreover, Veneridae were retrieved as monophyletic (PP=1.00). Within Pteriomorphia, four major clades were obtained: *Nuculana minuta*, *Mytilus spp.*, (Pterioidea + Ostreoidea + Pinnoidea), and (Limidae + Pectinoidea).

5.4. DISCUSSION

Aim of this study was to address the bivalve monophyly/polyphyly, a long-standing issue among molecular phylogeneticists; taxa were chosen as a consequence, including a large dataset of bivalve species, and at least one representative for each molluscan class, and more for richest group (namely, cephalopods, gastropods, and scaphopods).

Despite the ancient splits here investigated, little saturation traces were recovered in our dataset. Following the Paleobiology database (http://www.paleodb.org/cgi-bin/bridge.pl?a=beginFirstAppearance, consulted on 2011/03/07), most ancient known Mollusca are dated to the earliest Cambrian (542-530 millions of years ago). Therefore, it is expected that complex methods are needed to correctly read phylogenetic signals and address evolutionary questions.

Some groups did appear clearly from neighbornet networks (see Fig. 5.2, 5.3). The presence of such clusters, as well as the emergence of a single group of bivalves, with the exception of Opponobranchia, in the reduced neighbornet network (see Fig. 5.3) is a strong evidence for the ability of these markers to resolve, albeit partially, bivalve relationships with other mollusks. In any case, the neighbornet network failed to recover bivalves as monophyletic, since Opponobranchia are far from the rest of Bivalvia. However, this method do not account for complex molecular evolution patterns of the peculiar mollusk genome; indeed, while it is actually effective in describing phylogenetic signal presence and quality, it has been repeatedly proved that more realistic models are needed to infer mollusk (or at least bivalve) phylogeny (Doucet-Beaupré et al., 2010; Plazzi and Passamonti, 2010; Plazzi et al., in preparation).

On the contrary, four-cluster Likelihood Mapping yielded evidence of bivalve monophyly, in that the topology (Opponobranchia + Autobranchia) was significantly preferred to either alternatives. However, a complex situation emerged: a strong signal

was found, as only 5.3% of quartets mapped in the star-like tree area (see Fig. 5.5A), but contrasting signals were found in support of all possible topologies by analyzing single genes. As a matter of fact, the concatenated alignment yielded the same result of the *h3* gene, which is the most conserved in our dataset. More than 50% mapped in the ((Opponobranchia + Bivalvia) + (non-bivalve mollusks + outgroups)) corner, and more than 40% in the relative Voronoi cell, about twice than in other corners ($P<0.005$); star-like tree signal is only 9.7% in the concatenated alignment Four-cluster Likelihood Mapping simplex, reflecting the affordability of the phylogenetic signal.

Finally, both AIC and BF selected as best the model in which both bivalve and mollusks were forced as monophyletic. This may be taken as an evidence that these clades should be considered monophyletic.

Summarizing the above-mentioned results, the monophyly of Bivalvia was not univocally supported in our analysis, although most data indicated them as monophyletic. The issue is tightly linked to the position that Opponobranchia have in the evolutionary tree of Mollusca. Opponobranchia *sensu* Giribet (2008) are considered basal bivalves and sister group of all Autobranchia (Purchon, 1987; von Salvini-Plawen and Steiner, 1996; Waller, 1990, 1998; Morton, 1996; Cope, 1996, 1997). Their main features are the presence of true ctenidia (i.e., respiratory organs as those of other mollusks) and of well-developed labial palps and palp proboscides for feeding (Yonge, 1939), although Stasek (1963) found a small degree of interconnection between ctenidia and palps. Interestingly, Morton (1996) considered these characters as autapomorphies of Opponobranchia (Protobranchia in his taxonomy) and not as general plesiomorphies of all bivalves. Protobranch bivalves are also unique for other features, like the stomach of type 1 (Purchon, 1958). Labial palps of type 1 (Stasek, 1963) are present in Opponobranchia and are also quite uncommon among bivalves, and always associated with primitive groups (Crassatelloidea, Mytiloidea, Palaeoheterodonta).

Protobranch bivalves underwent several systematic rearrangements. Extant representatives are subdivided into two orders, Nuculoida and Solemyoida. They have been considered either within two different subclasses (see, f.i., Newell, 1965; Cope, 1996; and reference therein) or in the same taxon, Protobranchia (see, f.i., Purchon, 1987; Bieler and Mikkelsen, 2006; and reference therein). Moreover, the superfamily Nuculanoidea has recently been moved from order Nuculoida and is currently classified among Pteriomorphia (Giribet and Wheeler, 2002; Giribet and Distel, 2003; Plazzi and Passamonti, 2010) on essentially molecular bases. The homogeneous shell microstructure is different from the nacreous one of Nuculoida (Newell, 1965) and the taxodont hinge, although quite different, is also found in some Pteriomorphia, like Arcida.

The Opponobranchia were often responsible of bivalve polyphyly in molecular studies by clustering with different non-bivalve outgroup (Adamkewicz et al., 1997; Hoeh et al., 1998; Giribet and Wheeler, 2002; Giribet and Distel, 2003; Passamaneck et al., 2004). Bivalves were retrieved as monophyletic by Doucet-Beaupré et al. (2010), but Opponobranchia were not sampled in that study; monophyly of bivalves was firstly recovered by Wilson et al. (2010), which included also *Solemya velum* and *Nucula sulcata* in their dataset.

Most evidences we gathered point towards the conclusion that bivalves are indeed monophyletic. The agreement of tree-based and tree-independent analyses, like model decision tests (see Tab. 5.2 and 5.3) and Four-cluster Likelihood Mapping (see Fig. 5.5), is particularly significant on this account. The correctness of phylogenetic relationships among autobranchiate bivalves depicted by the m03bm tree might be a further warranty to this outcome.

Although the BF for m03bm model has to be considered, according to the bibliography, as a strong evidence in favor of it and against the m03m (Kass and Raftery, 1995; Brandley et al., 2005; and reference therein), we have to mention that m03bm

outperformed only slightly the m03m (ΔEML=4.63; ΔAIC=−9.26; BF=9.26), in which bivalves were not forced as monophyletic. For this reason and for the seek of caution, we would only suggest that bivalves are monophyletic. In fact, in the slightly sub-optimal tree (not shown), a well-supported clade (PP=0.99) comprises only Autobranchia as well as other mollusks: *Lottia*, *Aplysia*, (*Katharina* + Cephalopoda), Scaphopoda, (*Laevipilina* + *Epimenia*), while Opponobranchia are nested elsewhere among different molluscan outgroups, *Chaetoderma nitidulum* and (*Haliotis* + *Diodora*).

Because of this, we would recommend a further improvement of the available molluscan sequence dataset, with special reference to bivalves, to definitely unravel this issue. The peculiar evolutionary history of bivalve genomes might heavily weaken phylogenetic signal (at least in our dataset), leading to some artifactual evidences of polyphyly using different approaches, or vice versa. For instance, the phenomenon of Doubly Uniparental Inheritance (DUI; Skibinski et al., 1994a, 1994b; Zouros et al., 1994a, 1994b; Breton et al., 2007; Passamonti and Ghiselli, 2009; and reference therein), which is scattered throughout bivalves, may constitute one of the polluters of molecular evidence, at least for mitochondrial markers.

CHAPTER 6

CITED REFERENCES

Adamkewicz, S. L., Harasewych, M. G., Blake, J., Saudek, D., and Bult, C. J. 1997. A molecular phylogeny of the bivalve mollusks. Mol. Biol. Evol. 14: 619-629.

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrox, B. N., Caski, F. (eds.), Second International Symposium on Information Theory. Akademiai Kiado, Budapest, p. 267.

Albano, P. G., Rinaldi, E., Evangelisti, F., Kuan, M., and Sabelli, B. 2009. On the identity and origin of *Anadara demiri* (Bivalvia: Arcidae). J. Mar. Biol. Ass. U. K. 89: 1289-1298.

Alfaro, M. E., and Huelsenbeck, J. P. 2006. Comparative performance of Bayesian and AIC-based measures of phylogenetic model uncertainty. Syst. Biol. 55: 89-96.

Alfaro, M. E., Zoller, S., and Lutzoni, F., 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov Chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. Mol. Biol. Evol. 20: 255-266.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J., 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acid Res. 25: 3389-3402.

Amler, M. R. W., Thomas, E, and Weber, K. M., 1990. Bivalven des hoechsten Oberdevons im Bergischen Land (Strunium; noerdliches Rheinisches Schiefergebirge). Geologica et Palaeontologica 24.

Atkins, D. 1936-1938. On the ciliary mechanisms and interrelationships of lamellibranchs. Quarterly Journal of Microscopical Science 79: 181-308, 339-445; 80: 321-436.

Azzalini, A. 1985. A class of distributions which includes the normal ones. Scand. J. Statist. 12: 171-178.

Babin, C. 1982. Mollusques bivalves et rostroconches. In Babin, C., Courtessole, S., Melou, M., Pillet, J., Vizcaino, D., and Yochelson, E. L., Brachiopodes (articulés) et mollusques (bivalves, rostroconches, monoplacophores, gastropodes) de l'Ordovicien inférieur (Trémadocien-Arenigien) de la Montaigne Noire (France méridionale). Mémoire de la Société des Études Scientifiques de l'Aude, Carcassonne, pp. 37-49.

Baird, C., and Brett, C. E., 1983. Regional variation and paleontology of two coral beds in the Middle Devonian Hamilton Group of Western New York. J. Paleontol. 57: 417-446.

Baker, R. H., and DeSalle, R., 1997. Multiple source of character information and the phylogeny of Hawaiian drosophilids. Syst. Biol. 46: 654-673.

Ballard, J. W. O., and Whitlock, M. C., 2004. The incomplete natural history of mitochondria. Mol. Ecol. 13: 729-744.

Barucca, M., Olmo, E., Schiaparelli, S., and Canapa, A. 2004. Molecular phylogeny of the family Pectinidae (Mollusca: Bivalvia) based on mitochondrial 16S and 12S rRNA genes. Mol. Phylogenet. Evol. 31: 89-95.

Beckenbach, A. T., Robson, S. K. A., and Crozier, R. H. 2005. Single Nucleotide +1 Frameshifts in an Apparently Functional Mitochondrial Cytochrome b Gene in Ants of the Genus Polyrhachis. J. Mol. Evol. 60: 141-152.

Bell, E. T. 1934. Exponential numbers. Am. Math. Monthly 41: 411-419.

Bensasson, D., Zhang, D.-X., Hartl, D. L., and Hewitt, G. M. 2001. Mitochondrial pseudogenes: Evolution's misplaced witnesses. Trends Ecol. Evol. 16: 314-321.

Berry, W. B. N., and Boucot, A. J., 1973. Correlation of the African Silurian rocks. Geological Society of America Special Paper 147: 1-83.

Bieler, R., and Mikkelsen, P. M., 2006. Bivalvia – a look at the branches. Zool. J. Linn. Soc. 148: 223-235.

Bigot, A., 1935. Les recifs bathoniens de Normandie. Bulletin de la Societe Geologique de France, ser. 5 4: 697-736.

Birky, C. W., Jr., 2001. The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. Annu. Rev. Genet. 35: 125-148.

Bond, W. J. 1989. Describing and conserving biotic diversity. In Huntley, B. J. (ed.), Biotic Diversity in Southern Africa: Concepts and Conservation, Oxford University Press, Cape Town, pp. 2-18.

Bordewich, M., Rodrigo, A. G., and Semple, C. 2008. Selecting taxa to save or sequence: Desiderable criteria and a greedy solution. Syst. Biol. 57: 825-834.

Brandley, M. C., Schmitz, A., and Reeder, T. W. 2005. Partitioned Bayesian Analyses, Partition Choice, and the Phylogenetic Relationships of Scincid Lizards. Syst. Biol. 54: 373-390.

Brasier, M. D., and Hewitt, R. A., 1978. On the Late Precambrian- Early cambrian Hartshill Formation of Warwickshire. Geol. Mag. 115: 21-36.

Bremer, K., 1988. The limits of amminoacid sequence data in angiosperm phylogenetic reconstruction. Evolution 42: 795–803.

Bremer, K., 1994. Branch support and tree stability. Cladistics 10: 295-304.

Breton, S., Doucet-Beaupré, H., Stewart, D. T., Hoeh, W. R., and Blier, P. U. 2007. The unusual system of doubly uniparental inheritance of mtDNA: isn’t one enough? Trends Genet. 23: 465-474.

Breton, S., Doucet-Beaupré, H., Stewart, D. T., Piontkivska, H., Karmakur, M., Bogan, A. E., Blier, P. U., and Hoeh, W. R. 2009. Comparative Mitochondrial Genomics of Freshwater Mussels (Bivalvia: Unionoida) With Doubly Uniparental Inheritance of mtDNA:

Gender-Specific Open Reading Frames and Putative Origins of Replication. Genetics 183: 1575-1589.

Brett, E., Dick, V. B., and Baird, G. C., 1991. Comparative Taphonomy and Paleoecology of Middle Devonian Dark Gray and Black Shale Facies from Western New York. Dynamic Stratigraphy and Depositional Environments of the Hamilton Group (Middle Devonian) in New York State, Part II. New York State Museum Bulletin 469: 5-36.

Brusca, R. C., and Brusca, G. J. 2003. Invertebrates, second ed. Sinauer Associates, Sunderland.

Bryant, D., and Moulton, V. 2004. Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks. Mol. Biol. Evol. 21: 255-265.

Cai, C. Y., Dou, Y. W., and Edwards, D., 1993. New observations on a Pridoli plant assemblage from north Xinjiang, northwest China, with comments on its evolutionary and palaeogeographical significance. Geol. Mag. 130: 155-170.

Calendini, F., and Martin, J.-F. 2005. PaupUp v1.0.3.1. A free graphical frontend for Paup* Dos software. Distributed by the authors at http://www.agro-montpellier.fr/sppe/Recherche/JFM/PaupUp/main.htm.

Cameron, S. L., Lambkin, C. L., Barker, S. C., and Whiting, M. F., 2007. A mitochondrial genome phylogeny of Diptera: whole genome sequence data accurately resolve relationships over broad timescales with high precision. Syst. Entomol. 32: 40-59.

Campbell, D. C. 2000. Molecular evidence on the evolution of the Bivalvia. In Harper, E. M., Taylor, J. D., Crame, J. A. (eds.), The Evolutionary Biology of the Bivalvia, The Geological Society of London, London, pp. 31-46.

Campbell, J. D., Coombs, D. S., and Grebneff, A., 2003. Willsher Group and geology of the Triassic Kaka Point coastal section, south-east Otago, New Zealand. J. R. Soc. New Zeal. 33: 7-38.

Canapa, A., Barucca, M., Marinelli, A., and Olmo, E. 2000. Molecular Data from the 16S rRNA Gene for the Phylogeny of Pectinidae (Mollusca: Bivalvia). J. Mol. Evol. 50: 93-97.

Canapa, A., Barucca, M., Marinelli, A., and Olmo, E. 2001. A molecular phylogeny of Heterodonta (Bivalvia) based on small ribosomal subunit RNA sequences. Mol. Phylogenet. Evol. 21: 156-161.

Canapa, A., Marota, I., Rollo, F., and Olmo, E. 1996. Phylogenetic Analysis of Veneridae (Bivalvia): Comparison of Molecular and Palaeontological Data. J. Mol. Evol. 43: 517-522.

Canapa, A., Marota, I., Rollo, F., and Olmo, E. 1999. The small-subunit rRNA gene sequences of venerids and the phylogeny of Bivalvia. J. Mol. Evol. 48: 463-468.

Carter, J. G. 1990. Evolutionary significance of shell microstructure in the Palaeotaxodonta, Pteriomorphia and Isofilibranchia. In Carter, J. G. (ed.), Skeletal Biomineralization: Patterns, Process and Evolutionary Trends volume I, Van Nostrand Reinhold, New York, pp. 135-411.

Carter, J. G., Campbell, D. C., and Campbell, M. R. 2000. Cladistic perspectives on early bivalve evolution. In Harper, E. M., Taylor, J. D., and Crame, J. A. (eds.), The Evolutionary Biology of the Bivalvia, The Geological Society of London, London, pp. 47-79.

Castresana, J., 2000. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. Mol. Biol. Evol. 17: 540-552.

Chatterjee, S., 1986. *Malerisaurus langstoni*, a new diapsid reptile from the Triassic of Texas. J. of Vertebr. Paleontol. 6: 297-312.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G., and Thompson, J. D. 2003. Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res. 31: 3497-3500.

Clarke, K. R, and Warwick, R. M. 1999. The taxonomic distinctness measure of biodiversity: Weighting of step lengths between hierarchical levels. Mar. Ecol. Prog. Ser. 184: 21-29.

Clarke, K. R., and Warwick, R. M. 1998. A taxonomic distinctness index and its statistical properties. J. Appl. Ecol. 35: 523-531.

Clarke, K. R., and Warwick, R. M. 2001. A further biodiversity index applicable to species lists: Variation in taxonomic distinctness. Mar. Ecol. Prog. Ser. 216: 265-278.

Colless, D. H. 1982. Phylogenetics: The theory and practice in phylogenetic systematic II. Book review. Syst. Zool. 31: 100-104.

Cope, J. C. W. 1996. The early evolution of the Bivalvia. In Taylor, J. D. (ed.), Origin and Evolutionary Radiation of the Mollusca. Oxford University Press, Oxford, pp. 361-370.

Cope, J. C. W. 1997. The early phylogeny of the class Bivalvia. Palaeontology 40: 713-746.

Cox, L. R. 1959. The geological history of the Protobranchia and the dual origin of the taxodont Lamellibranchia. Proceedings of the Malacological Society of London 33: 200-209.

Cox, L. R. 1960. Thoughts on the classification of the Bivalvia. Proceedings of the Malacological Society of London 34: 60-88.

Cox, L. R., 1965. Jurassic Bivalvia and Gastropoda from Tanganyika and Kenya. Bulletin of the British Museum (Natural History) Geology Supplement I.

Cummings, M. P., Handley, S. A., Myers, D. S., Reed, D. L., Rokas, A., and Winka, K. 2003. Comparing bootstrap and posterior probability values in the four-taxon case. Syst. Biol. 52: 477-487.

Distel, D. L. 2000. Phylogenetic relationships among Mytilidae (Bivalvia): 18S rRNA data suggest convergence in mytilid body plans. Mol. Phylogenet. Evol. 15: 25-33.

Dou, Y. W., Sun, Z. H. 1983. Devonian Plants. Palaeontological Atlas of Xinjiang, vol. II. Late Palaeozoic Section. Geological Publishing House, Beijing.

Dou, Y. W., Sun, Z. H. 1985. On the Late Palaeozoic plants in Northern Xinjiang. Acta Geologica Sinica 59: 1-10.

Douady, C. J., Delsuc, F., Boucher, Y., Ford Doolittle, W., and Douzery, E. J. P. 2003. Comparison of Bayesian and Maximum Likelihood bootstrap measures of phylogenetic reliability. Mol. Biol. Evol. 20: 248-254.

Doucet-Beaupré, H., Breton, S., Chapman, E. G., Blier, P. U., Bogan, A. E., Stewart, D. T., and Hoeh, W. R. 2010. Mitochondrial phylogenomics of the Bivalvia (Mollusca): searching for the origin and mitogenomic correlates of doubly uniparental inheritance of mtDNA. BMC Evolutionary Biology 10: 50.

Dress, A., Huson, D., and Moulton, V. 1996. Analyzing and Visualizing Sequence and Distance Data Using SplitsTree. Discrete Appl. Math. 71: 95-109.

Dreyer, H., Steiner, G., and Harper, E. M., 2003. Molecular phylogeny of Anomalodesmata (Mollusca: Bivalvia) inferred from 18S rRNA sequences. Zool. J. Linn. Soc. 139: 229-246.

Dunn, C. W., Hejnol, A., Matus, D. Q., Pang, K., Browne, W. E., Smith, S. A., Seaver, E., Rouse, G. W., Obst, M., Edgecombe, G. D., Sørensen, M. V., Haddock, S. H. D., Schmidt-Rhaesa, A., Okusu, A., Møbjerg Kristensen, R., Wheeler, W. C., Matindale, M. Q., and Giribet, G. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. Nature 452: 745-750.

Elder, R. L., 1987. Taphonomy and paleoecology of the Dockum Group, Howard County, Texas. Journal of the Arizona-Nevada Academy of Science 22: 85-94.

Eriksson, T., 2007. The r8s bootstrap kit. Distributed by the author.

Erixon, P., Svennblad, B., Britton, T., and Oxelman, B. 2003. Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. Syst. Biol. 52: 665-673.

Farris, J. S., Kallersjö, M., Kluge, A. G., and Bult, C., 1995a. Constructing a significance test for incongruence. Syst. Biol. 44: 570-572.

Farris, J. S., Kallersjö, M., Kluge, A. G., and Bult, C., 1995b. Testing significance of incongruence. Cladistics 10: 315-319.

Felsenstein, J. 1978. Cases in which parsimony and compatibility will be positively misleading. Syst. Zool. 27: 401-410.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17: 368-376.

Felsenstein, J., 1993. PHYLIP: phylogenetic inference package. Distributed by the author.

Flynn, J. J., Finarelli, J. A., Zehr, S., Hsu, J., and Nedbal, M. A. 2005. Molecular phylogeny of the Carnivora (Mammalia): Assessing the impact of increased sampling on resolving enigmatic relationships. Syst. Biol. 54: 317-337.

Folmer, O., Black, M., Hoeh, W. R., Lutz, R., and Vrijenhoek, R. C., 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol. Mar. Biol. Biotechnol. 3: 294-299.

Garrido-Ramos, M. S., Stewart, D. T., Sutherland, B. W., and Zouros, E. 1998. The distribution of male-transmitted and female-transmitted mitochondrial DNA types in somatic tissues of blue mussels: implications for the operation of doubly uniparental inheritance of mitochondrial DNA. Genome 41: 818-824.

Gatesy, J., DeSalle, R., and Wheeler, W., 1993. Alignment-ambiguous nucleotide sites and the exclusion of systematic data. Mol. Phylogenet. Evol. 2: 152-157.

Gelman, A., and Rubin, D. B., 1992. Inference from iterative simulation using multiple sequences. Stat. Sci. 7: 457-511.

Ghiselli, F., Milani, L., and Passamonti, M. 2011. Strict sex-specific mtDNA segregation in the germline of the DUI species *Venerupis philippinarum* (Bivalvia Veneridae). Mol. Biol. Evol. 28: 949-961.

Gillham, N. W., 1994. Transmission and compatibility of organelle genomes. In: Gillham, N. W., Organelle Genes and Genomes. Oxford University Press, Oxford, pp. 147-268.

Giribet, G. 2008. Bivalvia. In Ponder, W. F., and Lindberg, D. R., (eds.), Phylogeny and Evolution of the Mollusca. University of California Press, Berkeley, pp. 105-142.

Giribet, G., and Carranza, S. 1999. Point counter point. What can 18S rDNA da for bivalve phylogeny? J. Mol. Evol. 48: 256-258.

Giribet, G., and Distel, D. L. 2003. Bivalve phylogeny and molecular data. In Lydeard, C., and Lindberg, D. R. (eds.), Molecular systematics and phylogeography of mollusks, Smithsonian Books, Washington, pp. 45-90

Giribet, G., and Wheeler, W. 2002. On bivalve phylogeny: A high-level analysis of the Bivalvia (Mollusca) based on combined morphology and DNA sequence data. Invert. Biol. 121: 271-324.

Giribet, G., Okusu, A., Lindgren, A. R., Huff, S. W., Schrödl, M., and Nishiguchi, M. K. 2006. Evidence for a clade composed of mollusks with serially repeated structures: Monoplacophorans are related to chitons. Proc. Natl. Acad. Sci. U. S. A. 103: 7723-7728.

Goldman, N. 1993. Statistical tests of models of DNA substitution. J. Mol. Evol. 36: 182-198.

Goldman, N. 1998. Phylogenetic information and experimental design in molecular systematics. Proc. R. Soc. Lond. B 265: 1779-1786.

Goldman, N., and Yang, Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol. 11: 725-736.

Goldman, N., Anderson, J. P., Rodrigo, A. G., 2000. Likelihood-Based Tests of Topologies in Phylogenetics. Syst. Biol. 49: 652-670.

Götting, K.-J. 1980a. Origin and relationships of the Mollusca. Z. Zool. Syst. Evol. Forsch. 18: 24-27.

Götting, K.-J. 1980b. Argumente für die Deszendenz der Mollusken von metameren Antezedenten. Zool. Jb. Anat. 130: 211.218.

Gradstein, F. M., Ogg, J. G., and Smith, A. G. (eds.), 2004. A Geologic Time Scale 2004. Cambridge University Press, Cambridge.

Graf, D. L., and Ó Foighil, D. 2000. The evolution of brooding characters among the freshwater pearly mussels (Bivalvia: Unionoidea) of North America. J. Molluscan Stud. 66: 157-170.

Graham Oliver, P., and Järnegren, J. 2004. How reliable is morphology based species taxonomy in the Bivalvia? A case study on Arcopsis adamsi (Bivalvia: Arcoidea) from the Florida Keys. Malacologia 46: 327-338.

Grant, T., and D'Haese, C. 2004. Insertions and deletions in the evolution of equal-length DNA fragments. In Stevenson, D. W., Abstracts of the 22nd Annual Meeting of the Willi Hennig Society, Cladistics, 20: 84.

Grasso, T. X., 1986. Redefinition, Stratigraphy and Depostional Environments of the Mottville Member (Hamilton Group) in Central and Eastern New York. Dynamic Stratigraphy and Depositional Environments of the Hamilton Group (Middle Devonian) in New York State, Part I. New York State Museum Bulletin 457: 5-31.

Hammer, Ø., Harper, D. A. T., Ryan, P. D., 2001. PAST: Paleontological Statistics Software Package for Education and Data Anlaysis. Palaeontologia Electronica 4: 1-9.

Harper, E. M. 2006. Dissecting post-Paleozoic arms race. Paleogeogr. Paleoclimatol. Paleoecol. 232: 322-343.

Harper, E. M., Dreyer, H., and Steiner, G. 2006. Reconstructing the Anomalodesmata (Mollusca: Bivalvia): morphology and molecules. Zool. J. Linn. Soc. 148: 395-420.

Harper, E. M., Hide, E. A., and Morton, B. 2000. Relationships between the extant Anomalodesmata: a cladistic test. In Harper, E. M., Taylor, J. D., and Crame, J. A. (eds.), The Evolutionary Biology of the Bivalvia, The Geological Society of London, London, pp. 129-143.

Hartmann, S., Vision, T. J., 2008. Using ESTs for phylogenomics: Can one accurately infer a phylogenetic tree from a gappy alignment? BMC Evol. Biol. 8: 95.

Haszprunar, G. 2000. Is the Aplacophora monophyletic? A cladistic point of view. Am. Malacol. Bull. 15: 115-130.

Haszprunar, G. 2008. Monoplacophora (Tryblidia). In Ponder, W. F., Lindberg, D. R., (eds.), Phylogeny and Evolution of the Mollusca. University of California Press, Berkeley, pp. 97-104.

Hautmann, M. 2004. Early Mesozoic evolution of alivincular bivalve ligaments and its implications for the timing of the ‘Mesozoic marine revolution’. Lethaia 37: 165-172.

Hautmann, M. 2006. Shell morphology and phylogenetic origin of oysters. Paleogeogr. Paleoclimatol. Paleoecol. 240: 668-671.

Hautmann, M., Aghababalou, B., and Krystyn, L. 2011. An unusual late Triassic nuculid bivalve with divaricate shell ornamentation, and the evolutionary history of oblique ribs in Triassic bivalves. J. Paleontol. 85: 22-28.

Hautmann, M., and Golej, M. 2004. *Terquemia* (*Dentiterquemia*) *eudesdeslongchampsi* new subgenus and species, an interesting cementing bivalve from the Lower Jurassic of the Western Carpathians (Slovakia). J. Paleontol. 78: 1086-1090.

Hayami, I., 1975. A systematic survey of the Mesozoic Bivalvia from Japan. The University Museum, The University of Tokyo, Bulletin 10.

He, T., Pei, F., and Fu, G., 1984. Some small shelly fossils from the Lower Cambrian Xinji Formation in Fangcheng County, Henan Province. Acta Palaeontologica Sinica 23: 350-35.

Healy, J. M. 1995. Sperm ultrastructure in the marine bivalve families Carditidae and Crassatellidae and its bearing of unification of the Crassatelloidea with the Carditoidea. Zool. Scr. 24: 21-28.

Heard, S. B. 1992. Patterns in tree balance among cladistic, phenetic, and randomly generated phylogenetic trees. Evolution 46: 1818-1826.

Heckert, B., 2004. Late Triassic microvertebrates from the lower Chinle Group (Otischalkian-Adamanian: Carnian), southwestern USA New Mexico Museum of Natural History and Science Bulletin 27: 1-170.

Hedtke, S. M., Townsend, T. M., and Hillis, D.M. 2006. Resolution of phylogenetic conflict in large data sets by increased taxon sampling. Syst. Biol. 55: 522-529.

Hendy, M. D., and Penny, D. 1989. A framework for the quantitative study of evolutionary trees. Syst. Zool. 38: 297-309.

Hillis, D. M. 1998. Taxonomic Sampling, Phylogenetic Accuracy, and Investigator Bias. Syst. Biol. 47: 3-8.

Hillis, D. M., Bull, J. J., White, M. E., Badgett, M. R., and Molineux, I. J. 1992. Experimental phylogenetics: generation of a known phylogeny. Science 255: 589-592.

Hoeh, W. R., Black, M. B., Gustafson, R. G, Bogan, A. E., Lutz, R. A., and Vrijenhoek, R. C. 1998. Testing alternative hypotheses of *Neotrigonia* (Bivalvia: Trigonioida) phylogenetic relationships using cytochrome c oxidase subunit I DNA sequences. Malacologia 40: 267-278.

Huelsenbeck, J. P, and Ronquist, F. 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17: 754-755.

Huelsenbeck, J. P. 1995a. Performance of phylogenetic methods in simulation. Syst. Biol. 44: 17–48.

Huelsenbeck, J. P. 1995b. The robustness of two phylogenetic methods: Four taxon simulations reveal a slight superiority of maximum likelihood over neighbor joining. Mol. Biol. Evol. 12: 843–849.

Huelsenbeck, J. P., and Crandall, K. A. 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. Annu. Rev. Ecol. Syst. 28: 437-466.

Huelsenbeck, J. P., Bollback, J. P., and Levine, A.M. 2002. Inferring the root of a phylogenetic tree. Syst. Biol. 51: 32-43.

Huelsenbeck, J. P., Bull, J. J., and Cunningham, C. W. 1996a. Combining data in phylogenetic analysis. Trends Ecol. Evol. 11: 152-158.

Huelsenbeck, J. P., Bull, J. J., and Cunningham, C. W. 1996b. Reply. Trends Ecol. Evol. 11: 335.

Huelsenbeck, J. P., Hillis, D. M., and Jones, R., 1996a. Parametric bootstrapping in molecular phylogenetics: Applications and performance. In: Ferraris, J. D., Palumbi, S. R. (eds.), Molecular zoology: Advances, strategies and protocols. Wiley & Sons, New York, pp. 19-45.

Huelsenbeck, J. P., Hillis, D. M., and Nielsen, R., 1996b. A likelihood-ratio test of monophyly. Syst. Biol. 45: 546-558.

Huelsenbeck, J. P., Larget, B., and Alfaro, M. E., 2004. Bayesian Phylogenetic Model Selection Using Reversible Jump Markov Chain Monte Carlo. Mol. Biol. Evol. 21: 1123-1133.

Huff, S. W., Campbell, D., Gustafson, D. L., Lydeard, C., Altaba, C. R., and Giribet, G. 2004. Investigations into the phylogenetic relationships of the threatened freshwater pearl-mussels (Bivalvia, Unionoidea, Margaritiferidae) based on molecular data: implications for their taxonomy and biogeography. J. Molluscan Stud. 70: 379-388.

Huson, D. H., Richter, D. C., Rausch, C., Dezulian, T., Franz, M., and Rupp, R., 2007. Dendroscope – An interactive viewer for large phylogenetic trees. BMC Bioinformatics 8: 460.

Huson, D., and Bryant, D. 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23: 254-267.

Ilves, K. L., and Taylor, E. B. 2009. Molecular resolution of the systematics of a problematic group of fishes (Teleostei: Osmeridae) and evidence for morphological homoplasy. Mol. Phylogenet. Evol. 50: 163-178

Insua, A., López-Piñon, M. J., Freire, R., and Méndez, J. 2003. Sequence analysis of the ribosomal DNA internal transcribed spacer region in some scallop species (Mollusca: Bivalvia: Pectinidae). Genome 46: 595-604.

Jenner R. A., Dhubhghaill C. N., Ferla M. P., and Wills, M. A. 2009. Eumalacostracan phylogeny and total evidence: Limitations of the usual suspects. BMC Evol. Biol. 9: 21.

Jordan, G. E., and Piel, W. H., 2008. PhyloWidget: Web-based visualizations for the tree of life. Bioinformatics 15: 1641-1642.

Jozefowicz, C. J., and Ó Foighil, D., 1998. Phylogenetic analysis of Southern Hemisphere flat oysters based on partial mitochondrial 16S rDNA gene sequences. Mol. Phylogenet. Evol. 10: 426-435.

Jukes, T. H., and Cantor, C. R. 1969. Evolution of protein molecules. In Munro, H. N. (ed.), Mammalian protein metabolism, Academic Press, New York, pp 21-123.

Kallersjö, M., Farris, J.S., Kluge, A.G., and Bult, C. 1992. Skewness and permutation. Cladistics 8: 275–287.

Kappner, I., and Bieler, R., 2006. Phylogeny of venus clams (Bivalvia: Venerinae) as inferred from nuclear and mitochondrial gene sequences. Mol. Phylogenet. Evol. 40: 317-331.

Kass, R. E., and Raftery, A. E. 1995. Bayes Factors. J. Am. Stat. Assoc. 90: 773-795.

Kemp, J., 1976. Account of Excavations into the Campanile Bed (Eocene, Selsey Formation) at Stubbington, Hants. Tertiary Research 1: 41-45.

Kimura, M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16: 111-120.

Kirkendale, L. 2009. Their Day in the Sun: molecular phylogenetics and origin of photosymbiosis in the 'other' group of photosymbiotic marine bivalves (Cardiidae: Fraginae). Biol. J. Linnean Soc. 97: 448-465.

Kirkendale, L., Lee, T., Baker, P., and Ó Foighil, D. 2004. Oysters of the Conch Republic (Florida Keys): a molecular phylogenetic study of *Parahyotissa mcgintyi*, *Teskeyostrea weberi* and *Ostreola equestris*. Malacologia 46: 309-326.

Kirkpatrick, M., and Slatkin, M. 1993. Searching for evolutionary patterns in the shape of a phylogenetic tree. Evolution 47: 1171-1181.

Kocher, T. D., Thomas, W. K., Meyer, A., Edwards, S. V., Pääbo, S., Villablanca, F. X., and Wilson, A. C., 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc. Natl. Acad. Sci. USA 86: 6196-6200.

Kříž, J., 1999. Bivalvia communities of Bohemian type from the Silurian and Lower Devonian carbonate facies. In: Boucot, A. J., Lawson, J. D. (Eds.), Paleocommunities--a case study from the Silurian and Lower Devonian. Cambridge University Press, Cambridge, pp. 299-252.

Kullback, S., and Leibler, R. A. 1951. On information and sufficiency. Ann. Math. Stat. 22: 79-86.

La Perna, R. 1998. On *Asperarca* Sacco, 1898 (Bivalvia, Arcidae) and two new Mediterranean species. Bollettino Malacologico 33: 11-18.

Lambkin, C. L., Lee, M. S. Y., Winterton, S. L., and Yeates, D. K. 2002. Partitioned Bremer support and multiple trees. Cladistics 18: 436-444.

Larget, B., and Simon, D. 1999. Markov Chain Monte Carlo Algorithms for the Bayesian Analysis of Phylogenetic Trees, Mol. Biol. Evol. 16: 750-759.

Larson, A. 1994. The comparison of morphological and molecular data in phylogenetic systematics. In Schierwater, B., Street, B., Wagner, G. P., DeSalle, R. (eds.) Molecular ecology and evolution: approaches and applications. Birkhauser, Basel, pp. 371-390.

Laudon, L. R., 1931. The Stratigraphy of the Kinderhook Series of Iowa. Iowa Geological Survey 35: 333-452.

Leaché, A. D., and Reeder, T. W. 2002. Molecular systematics of the eastern fence lizard (Sceloporus undulatus): A comparison of parsimony, likelihood, and Bayesian approaches, Syst. Biol. 51: 44–68.

Lee, M. S. Y., and Hugall, A. F. 2003. Partitioned likelihood support and the evaluation of data set conflict. Syst. Biol. 52: 11-22.

Lee, T., and Ó Foighil, D. 2003. Phylogenetic structure of the Sphaeriinae, a global clade of freshwater bivalve mollusks, inferred from nuclear (ITS-1) and mitochondrial (16S) ribosomal gene sequences. Zool. J. Linn. Soc. 137: 245-260.

Legendre, F., Whiting, M. F., Bordereau, C., Cancello, E. M., and Evans, T. A. 2008. The phylogeny of termites (Dictyoptera: Isoptera) based on mitochondrial and nuclear markers: Implications for the evolution of the worker and pseudergate castes, and foraging behaviors. Mol. Phylogenet. Evol. 48: 615-627.

Lehman, T. M., and Chatterjee, S. 2005. Depositional setting and vertebrate biostratigraphy of the Triassic Dockum Group of Texas. Journal of Earth Systems Science 114: 325-351.

Lemche, H. 1957. A new living deep-sea mollusk of the Cambrio-Devonian class Monoplacophora. Nature 179: 413-416.

Lento, G. M., Hickson, R. E., Chambers, G. K., and Penny, D. 1995. Use of Spectral Analysis to test Hypotheses on the Origin of Pinnipeds. Mol. Biol. Evol. 12: 28-52.

Leonard, D. R. P., Clarke, K. R., Somerfield, P. J., and Warwick, R. M. 2006. The application of an indicator based on taxonomic distinctness for UK marine nematode assessments. J. Environ. Manag. 78: 52-62.

Li, B., Dettaï, A., Cruaud, C., Couloux, A., Desoutter-Meniger, M., and Lecointre, G., 2009. RNF213, a new nuclear marker for acanthomorph phylogeny. Mol. Phylogenet. Evol. 50: 345-363.

Littlewood, D. T. J. 1994. Molecular Phylogenetics of Cupped Oysters Based on Partial 28S rRNA Gene Sequences. Mol. Phylogenet. Evol. 3: 221-229.

López-Flores, I., de la Herrán, R., Garrido-Ramos, M. A., Boudry, P., Ruiz-Rejón, C., and Ruiz-Rejón, M. 2004. The molecular phylogeny of oysters based on a satellite DNA related to transposons. Gene 339: 181-188.

López-Piñon, M. J., Freire, R., Insua, A., and Méndez, J. 2008. Sequence characterization and phylogenetic analysis of the 5S ribosomal DNA in some scallops (Bivalvia: Pectinidae). Hereditas 145: 9-19.

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

Luo, A., Qiao, H., Zhang, Y., Shi, W., Ho, S. Y. W., Xu, W., Zhang, A., and Zhu, C. 2010. Performance of criteria for selecting evolutionary models in phylogenetics: a comprehensive study based on simulated datasets. BMC Evol. Biol. 10: 242.

Luo, A., Zhang, A., Ho, S. Y. W., Xu, W., Zhang, Y., Shi, W., Cameron, S. L., and Zhu, C. 2011. Potential efficacy of mitochondrial genes for animal DNA barcoding: a case study using eutherian mammals. BMC Genomics 12: 84.

Lutzoni, F., Wagner, P., Reeb, V., Zoller, S., 2000. Integrating ambiguously aligned regions of DNA sequences in phylogenetic analyses without violating positional homology. Syst. Biol. 49: 628-651.

Lydeard, C., Mulvey, M., and Davis, G. M. 1996. Molecular systematics and evolution of reproductive traits of North American freshwater unionacean mussels (Mollusca: Bivalvia) as inferred from 16S rRNA gene sequences. Phil. Trans. R. Soc. B 351: 1593-1603.

Maddison, W. P., and Maddison, D. R. 2010. Mesquite: A modular system for evolutionary analysis. Version 2.74. Distributed by the authors at http://mesquiteproject.org.

Malchus, N. 2004. Constraints in the ligament ontogeny and evolution of pteriomorphian Bivalvia. Paleontology 47: 1539-1574.

Manten, A., 1971. Silurian Reefs of Gotland. Developments in Sedimentology 13, 1-539.

Martínez-Lage, A., Rodríguez, F., González-Tizón, A., Prats, E., Cornudella, L., and Méndez, J. 2002. Comparative analysis of different satellite DNAs in four Mytilus species. Genome 45: 922-929.

Maruyama, T., Ishikura, M., Yamazaki, S., and Kanai, S., 1998. Molecular phylogeny of zooxanthellate bivalves. Biol. Bull. 195: 70-77.

Matsumoto, M. 2003. Phylogenetic analysis of the subclass Pteriomorphia (Bivalvia) from mtDNA COI sequences. Mol. Phylogenet. Evol. 27: 429-440.

Matsumoto, M., and Hayami, I. 2000. Phylogenetic analysis of the family Pectinidae (Bivalvia) based on mitochondrial cytochrome c oxidase subunit I. J. Moll. Stud. 66: 477-488.

May, R. M. 1990. Taxonomy as destiny. Nature 347: 129-130.

McGuire, G., Denham, M. C., and Balding, D. J. 2001. Mac5: Bayesian inference of phylogenetic trees from DNA sequences incorporating gaps. Bioinformatics 17: 479–480.

Mergl, M., and Massa, D., 1992. Devonian and Lower Carboniferous brachiopods and bivalves from western Libya. Biostratigraphie du Paleozoique 12: 1-115.

Merritt, T. J., Shi, L., Chase, M. C., Rex, M. A., Etter, R. J., and Quattro, J. M., 1998. Universal cytochrome b primers facilitate intraspecific studies in molluscan taxa. Mol. Mar. Biol. Biotechnol. 7: 7-11.

Mikkelsen, P. M., Bieler, R., Kappner, I., and Rawlings, T. A. 2006. Phylogeny of Veneroidea (Mollusca: Bivalvia) based on morphology and molecules. Zool. J. Linn. Soc. 148: 439-521.

Millard, V. 2001. Classification of Mollusca: A classification of world wide Mollusca, 2nd edition. Printed by the author, South Africa, vol. 3, pp. 915-1447.

Mindell, D. P., Sorenson, M. D., and Dimcheff, D. E. 1998. An Extra Nucleotide Is Not Translated in Mitochondrial ND3 of Some Birds and Turtles. Mol. Biol. Evol. 15: 1568-1571.

Minin, V., Abdo, Z., Joyce, P., and Sullivan, J. 2003. Performance-based selection of likelihood models for phylogeny estimation. Syst. Biol. 52: 674-683.

Mooers, A. Ø., and Heard, S. B. 1997. Inferring evolutionary process from phylogenetic tree shape. Q. Rev. Biol. 72: 31-54.

Morton, B. 1996. The evolutionary history of the Bivalvia. In Taylor, J. D. (ed.), Origin and Evolutionary Radiation of the Mollusca. Oxford University Press, Oxford, pp. 337-359.

Morton, B., and Yonge, C. M. 1964. Classification and structure of the Mollusca. In Wilbur, K. M., and Yonge, C. M. (eds.), Physiology of Mollusca, 1, Academic Press, New York, pp. 1-58.

Murry, P. A., 1989. Geology and paleontology of the Dockum Formation (Upper Triassic), west Texas and eastern New Mexico. In: Lucas, S. G., Hunt, A. P. (eds.), Dawn

of the Age of Dinosaurs in the American Southwest. New Mexico Museum of Natural History, Albuquerque, pp. 102-144.

Muse, S. V. 1995. Evolutionary analyses of DNA sequences subject to constraints on secondary structure. Genetics 139: 1429-1439.

Muse, S. V., and B. S. Gaut. 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates with application to the chloroplast genome. Mol. Biol. Evol. 11: 715-724.

Myra Keen, A. 1963. Marine molluscan genera of Western North America: an illustrated key. Stanford University Press, Stanford.

Myra Keen, A. 1971. Sea shells of Tropical West America: Marine mollusks from Baja California to Peru, 2 ed.. Stanford University Press, Stanford.

Newell, N. D. 1965. Classification of the Bivalvia. Am. Mus. Novit. 2206: 1-25.

Nielsen, C. 1995. Animal evolution, interrelationships of the living phyla. Oxford University Press, Oxford.

Nielsen, R., and Yang, Z., 1998. Likelihood models for detecting positively selected amminoacids sites and applications to the HIV-1 envelope gene. Genetics 148: 929-936.

Nieselt-Struwe, K., and von Haeseler, A. 2001. Quartet-Mapping, a Generalization of the Likelihood-Mapping Procedure. Mol. Biol. Evol. 18: 1204-1219.

Nikula, R., Strelkov, P., and Väinölä, R. 2007. Diversity and trans-arctic invasion history of mitochondrial lineages in the North Atlantic Macoma balthica complex (Bivalvia: Tellinidae). Evolution 61: 928-941.

Nuin, P. 2008. MrMTgui: Cross-platform interface for ModelTest and MrModeltest. Distributed by the author at http://www.genedrift.org/mtgui.php.

Nylander, J. A. A, Wilgenbusch, J. C., Warren, D. L., and Swofford, D. L., 2008. AWTY (are we there yet?): A system for graphical exploration of MCMC convergence in Bayesian phylogenetics. Bioinformatics 24: 581-583.

Nylander, J. A. A., Ronquist, F., Huelsenbeck, J. P., and Nieves-Aldrey, J. L. 2004. Bayesian phylogenetic analysis of combined data. Syst. Biol. 53: 47-67.

Ó Foighil, D., and Smith, M. J. 1995. Evolution of asexuality in the cosmopolitan marine clam Lasaea. Evolution 49: 140-150.

Ó Foighil, D., and Taylor, D. J. 2000. Evolution of Parental Care and Ovulation Behaviour in Oysters. Mol. Phylogenet. Evol. 15: 301-313.

Ogg, J. G., Ogg, G., Gradstein, F. M., 2008. The Concise Geologic Time Scale. Cambridge University Press, Cambridge.

Olu-Le Roy, K., von Cosel, R., Hourdez, S., Carney, S. L., and Jollivet, D. 2007. Amphi-Atlantic cold-seep Bathymodiolus species complexes across the equatorial belt. Deep-Sea Res. Pt. I 54: 1890-1911.

Palero, F., Crandall, K. A., Abelló, P., Macpherson, E., and Pascual, M. 2009. Phylogenetic relationships between spiny, slipper and coral lobsters (Crustacea, Decapoda, Achelata). Mol. Phylogenet. Evol. 50: 152-162.

Palmer, T. J., 1979. The Hampen Marly and White Limestones Formations: Florida-type carbonate lagoons in the Jurassic of central England. Palaeontology 22: 189-228.

Palumbi, S. R., Martin, A., Romano, S., McMillan, W. O., Stice, L., and Grabowski, G., 1996. The simple fool's guide to PCR. Kewalo Marine Laboratory and University of Hawaii, Hawaii.

Pardi, F., and Goldman, N. 2005. Species choice for comparative genomics: Being greedy works. PLoS Genet. 1: e71.

Pardi, F., and Goldman, N. 2007. Resource-aware taxon selection for maximizing phylogenetic diversity. Syst. Biol. 56: 431-444.

Park, J. K., and Ó Foighil, D. 2000. Sphaeriid and corbiculid clams represent separate heterodont bivalve radiations into freshwater environments. Mol. Phylogenet. Evol. 14: 75-88.

Parker, S. R., 1997. Sequence Navigator. Multiple sequence alignment software. Methods Mol. Biol. 70: 145-154.

Parkhaev, P. Y. U., 2004. Malacofauna of the Lower Cambrian Bystraya Formation of Eastern Transbaikalia. Paleontol. J. 38: 590-608.

Passamaneck, Y. J., Schander, C., and Halanych, K. M. 2004. Investigation of molluscan phylogeny using large-subunit and small-subunit nuclear rRNA sequences. Mol. Phylogenet. Evol. 32: 25-38.

Passamonti, M., 2007. An unusual case of gender-associated mitochondrial DNA heteroplasmy: the mytilid Musculista senhousia (Mollusca Bivalvia). BMC Evolutionary Biology 7, S7.

Passamonti, M., and Ghiselli, F. 2009. Doubly Uniparental Inheritance: Two Mitochondrial Genomes, One Precious Model for Organelle DNA Inheritance and Evolution. DNA Cell Biol. 28: 1-10.

Passamonti, M., Boore, J. L., Scali, V., 2003. Molecular evolution and recombination in gender-associated mitochondrial DNAs of the Manila clam *Tapes philippinarum*. Genetics 164: 603-611.

Peek, A. S., Gustafson, R. G., Lutz, R. A., and Vrijenhoek, R. C. 1997. Evolutionary relationships of deep-sea hydrothermal vent and cold-water seep clams (Bivalvia: Vesicomyidae): results from the mitochondrial cytochrome oxidase subunit I. Mar. Biol. 130: 151-161.

Peel, J. S. 1991. Functional morphology of the class Helcionelloida nov., and the early evolution of Mollusca. In Simonetta, A. M., and Conway Morris, S. (eds.), The early evolution of Metazoa and the significance of problematic taxa. Cambridge University Press, Cambridge, pp. 157-177.

Peet, R. K. 1974. The measurement of species diversity. Ann. Rev. Ecol. Syst. 5: 285-307.

Plazzi, F., and Passamonti, M. 2010. Towards a molecular phylogeny of Mollusks: Bivalves' early evolution as revealed by mitochondrial genes. Mol. Phylogenet. Evol. 57: 641-657.

Plazzi, F., Ferrucci, R. R., and Passamonti, M. 2010. Phylogenetic Representativeness: a new method for evaluating taxon sampling in evolutionary studies. BMC Bioinformatics 11: 209.

Plohl, M., Luchetti, A., Meštrović, N., and Mantovani, B., 2008. Satellite DNAs between selfishness and functionality: Structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. Gene 409: 72-82.

Poe, S. 2003. Evaluation of the strategy of Long-Branch Subdivision to improve the accuracy of phylogenetic methods. Syst. Biol. 52: 423-428.

Pojeta Jr., J. 1978. The origin and early taxonomic diversification of pelecypods. Phil. Trans. R. Soc. B 284: 225-246.

Pojeta Jr., J. 1980. Molluscan phylogeny. Tul. Stud. Geol. Paleontol. 16: 55-80.

Pojeta Jr., J., and Runnegar, B. 1976. The paleontology of rostroconch mollusks and early history of the phylum Mollusca. U.S. Geol. Surv. Prof. Paper 986: 1-88.

Pojeta Jr., J., and Runnegar, B., 1985. The early evolution of diasome Mollusca. In Trueman, E. R., and Clarke, M. R. (eds.), The Mollusca. Vol. 10: Evolution. Academic Press, New York and London, pp. 295-336.

Pollock, D. D., and Bruno, W. J. 2000. Assessing an unknown evolutionary process: Effect of increasing site-specific knowledge through taxon addition. Mol. Biol. Evol. 17: 1854-1858.

Posada, D. 2003. Using Modeltest and PAUP* to select a model of nucleotide substitution. In Baxevanis, A. D., Davison, D. B., Page, R. D. M., Petsko, G. A., Stein, L. D., and Stormo, G. D. (eds.), Current Protocols in Bioinformatics, John Wiley & Sons, Inc., pp. 6.5.1-6.5.14.

Posada, D., and Buckley, T. R. 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Syst Biol. 53: 793–808.

Posada, D., and Crandall, K. A. 1998. Modeltest: Testing the model of DNA substitution. Bioinformatics 14: 817-818.

Posada, D., and Crandall, K. A. 2001. Selecting the best-fit model of nucleotide substitution. Syst. Biol. 50: 580-601.

Poulton, T. P., 1991. Hettangian through Aalenian (Jurassic) guide fossils and biostratigraphy, Northern Yukon and adjacent Northwest Territories. Geological Survey of Canada Bulletin 410: 1-95.

Purchon, R. D. 1958. Phylogeny in the Lamellibranchia. Proc. Cent. And Bicent. Congr. Biol. Singapore 69-82.

Purchon, R. D. 1987. Classification and evolution of the Bivalvia: An analytical study. Phil. Trans. R. Soc. Lond. B 316: 277-302.

Puslednik, L., and Serb, J. M. 2008. Molecular phylogenetics of the Pectinidae (Mollusca: Bivalvia) and effect of increased taxon sampling and outgroup selection on tree topology. Mol. Phylogenet. Evol. 48: 1178-1188.

Rannala, B., Huelsenbeck, J. P., Yang, Z., and Nielsen, R. 1998. Taxon sampling and the accuracy of large phylogenies. Syst. Biol. 47: 702-710.

Reeder, T. W. 2003. A phylogeny of the Australian *Sphenomorphus* group (Scincidae: Squamata) and the phylogenetic placement of the crocodile skinks (*Tribolonotus*): A Bayesian approaches to assessing congruence and obtaining confidence in maximum likelihood inferred relationships. Mol. Phylogenet. Evol. 27: 384-397.

Reinhart, P. W. Classification of the pelecypod family Arcidae. Bulletin du Musée Royal d'Histoire Naturelle de Belgique 11: 5-68.

Ricotta, C, and Avena, G. C. 2003. An information-theoretical measure of taxonomic diversity. Acta Biotheor. 51: 35-41.

Rode, L., and Lieberman, B. S., 2004. Using GIS to unlock the interactions between biogeography, environment, and evolution in middle and Late Devonian brachiopods and bivalves. Palaeogeogr. Palaeocl. 211: 345-359.

Roe, A. D., and Sperling, F. A. 2007. Patterns of evolution of mitochondrial cytochrome c oxidase I and II DNA and implications for DNA barcoding. Mol. Phylogenet. Evol. 44: 325-345.

Roe, K. J., and Hoeh, W. R. 2003. Systematics of freshwater mussels (Bivalvia: Unionoida). In Lydeard, C., and Lindberg, D. R. (eds.), Molecular Systematics and Phylogeography of Mollusks, Smithsonian Books, Washington, pp. 91-122.

Roe, K. J., Hartfield, P. D., and Lydeard, C. 2001. Phylogeographic analysis of the threatened and endangered superconglutinate-producing mussels of the genus *Lampsilis* (Bivalvia: Unioinidae). Mol. Ecol. 10: 2225-2234.

Rokas, A., and Carroll, S. B. 2005. More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. Mol. Biol. Evol. 22: 1337-1344.

Ronquist, F., and Huelsenbeck, J. P. 2003. MRBAYES 3: Bayesian phylogenetic inference using mixed models. Bioinformatics 19: 1572-1574.

Ronquist, F., Huelsenbeck, J. P., and van der Mark, P., 2005. MrBayes 3.1 Manual. Draft 5/26/2005. Distributed with the software.

Rost, H. 1955. A report on the family Arcidae. Allan Hancock Foundation publications. Series 1. Allan Hancock Pacific expeditions 20: 177-249.

Ruiz, C., Jordal, B., and Serrano, J. 2009. Molecular phylogeny of the tribe Sphodrini (Coleoptera: Carabidae) based on mitochondrial and nuclear markers. Mol. Phylogenet. Evol. 50: 44-58.

Runnegar, B., and Bentley, C. 1983. Anatomy, ecology and affinities of the Australian Early Cambrian bivalve *Pojetaia runnegari* Jell. J. Paleontol. 57: 73-92.

Runnegar, B., and Pojeta Jr, J. 1974. Molluscan phylogeny: the paleontological viewpoint. Science 186: 311-317.

Runnegar, B., and Pojeta Jr, J. 1985. Origin and diversification of the Mollusca. In Wilbur, K. M., Trueman, E. R. (eds.), The Mollusca. Vol. 10: Evolution. Academic Press, New York and London, pp. 1-57.

Runnegar, B., and Pojeta Jr, J. 1992. The earliest bivalves and their Ordovician descendants. Am. Malacol. Bull. 9: 117-122.

Samadi, S., Quéméré, E., Lorion, J., Tillier, A., von Cosel, R., Lopoez, P., Cruaud, C., Couloux, A., and Boisselier-Dubayle M.-C. 2007. Molecular phylogeny in mytilids supports the wooden steps to deep-sea vents hypothesis. C. R. Biol. 330: 446-456.

Samtleben, C., Munnecke, A., Bickert, T., and Paetzold, J., 1996. The Silurian of Gotland (Sweden): facies interpretation based on stable isotopes in brachiopod shells. Geologische Rundschau 85: 278-292.

Sanderson, M. J., 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics 19: 301-302.

Scheltema, A. H. 1993. Aplacophora as Progenetic Aculiferans and the Coelomate Origin of Mollusks as the Sister Taxon of Sipuncula. Biol. Bull. 184: 57-78.

Scheltema, A. H. 1996. Phylogenetic position of Sipuncula, Mollusca and the progenetic Aplacophora. In Taylor, J. D. (ed.), Origin and Evolutionary Radiation of the Mollusca, Oxford University Press, Oxford, pp. 53-58.

Schmidt, H. A., and von Haeseler, A. 2003. Maximum-Likelihood Analysis Using TREE-PUZZLE. In Baxevanis, A. D, Davison, D. B., Page, R. D. M., Stormo, G., and Stein, L. (eds.), Current Protocols in Bioinformatics, Wiley and Sons, New York, pp. 6.6.1-6.6.25.

Schmidt, H. A., Strimmer, K., Vingron, M., and von Haeseler, A. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18: 502-504.

Schneider, J. A., and Ó Foighil, D., 1999. Phylogeny of giant clams (Cardiidae: Tridacninae) based on partial mitochondrial 16S rDNA gene sequences. Mol. Phylogenet. Evol. 13: 59-66.

Serb, J. M., and Lydeard, C. 2003. Complete mtDNA Sequence of the North American Freshwater Mussel, *Lampsilis ornata* (Unionidae): An Examination of the Evolution and Phylogenetic Utility of Mitochondrial Genome Organization in Bivalvia (Mollusca). Mol. Biol. Evol. 20: 1854-1866.

Serb, J. M., Buhay, J. E., and Lydeard, C. 2003. Molecular systematic of the North American freshwater bivalve genus *Quadrula* (Unionidae: Ambleminae) based on mitochondrial ND1 sequences. Mol. Phylogenet. Evol. 28: 1-11.

Shao, K.-T., and Sokal, R. R. 1990. Tree balance. Syst. Zool. 39: 266-276.

Shilts, M. H., Pascual, M. S., and Ó Foighil, D. 2007. Systematic, taxonomic and biogeographic relationships of Argentine flat oysters. Mol. Phylogenet. Evol. 44: 467-473.

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16: 1114-1116.

Shull, H. C., Pérez-Losa, D. A. M., Blair, D., Sewell, k., Sinclair, E. A., Lawler, S., Ponniah, M., and Crandall, K. A. 2005. Phylogeny and biogeography of the freshwater crayfish Euastacus (decapoda: Parastacidae) based on nuclear and mitochondrial DNA. Mol. Phylogenet. Evol. 37: 249-263.

Simmons, M. P., and Ochoterena, H., 2000. Gaps as characters in sequence-based phylogenetic analyses. Syst. Biol. 49: 369-381.

Simmons, M. P., Pickett, K. M., and Miya, M. 2004. How meaningful are Bayesian support values? Mol. Biol. Evol. 21: 188-199.

Simon, C., Buckley, T. R., Frati, F., Stewart, J. B., and Beckenbach, A. T., 2006. Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. Annu. Rev. Ecol. Evol. Syst. 37: 545-579.

Skibinski, D. O. F., Gallagher, C., and Beynon, C. M. 1994a. Mitochondrial DNA inheritance. Nature 368: 817-818.

Skibinski, D. O. F., Gallagher, C., and Beynon, C. M. 1994b.Sex-limited mitochondrial DNA transmission in the marine mussel *Mytilus edulis*. Genetics 138: 801-809.

Smith, S. A., and Dunn, C. W. 2008. Phyutility: a phyloinformatics tool for trees, alignments, and molecular data. Bioinformatics 24: 715-716.

Sorenson, M. D., and Quinn, T. W. 1998. Numts: A Challenge for Avian Systematics and Population Biology. The Auk 115: 214-221.

Sorenson, M. D., Franzosa, E. A., 2007. TreeRot, version 3. Boston University, Boston, Massachusetts, USA.

Spath, L. F., 1930. The Eotriassic invertebrate fauna of East Greenland. Meddeleser om Grønland 83: 1-90.

Stanley, S. M. 1968. Post-Paleozoic adaptive radiation of infaunal bivalve mollusks: a consequence of mantle fusion and siphon formation. J. Paleontol. 42: 214-229.

Stanley, S. M. 1977. Trends, rates, and patterns of evolution in the Bivalvia. In Hallam, A. (ed.), Patterns of evolution as illustrated by the fossil record, Elsevier, Amsterdam, pp. 209-250.

Starobogatov, Y. I. 1992. Morphological basis for phylogeny and classification of Bivalvia. Ruthenica 2: 1-25.

Stasek, C. R. 1963. Synopsis and discussion of the association of ctenidia and labial palps in the bivalve Mollusca. Veliger, 6: 91-97.

Steiner, G. 1992. Phylogeny and classification of Scaphopoda. J. Moll. Stud. 58: 385-400.

Steiner, G. 1999. Point counter point. What can 18S rDNA do for bivalve phylogeny? J. Mol. Evol. 48: 258-261.

Steiner, G., and Dreyer, H. 2003. Molecular phylogeny of Scaphopoda (Mollusca) inferred from 18S rDNA sequences: support for a Scaphopoda–Cephalopoda clade. Zool. Scripta 32: 343-356.

Steiner, G., and Hammer, S. 2000. Molecular phylogeny of the Bivalvia inferred from 18S rDNA sequences with particular reference to the Pteriomorphia. In Harper, E. M., Taylor, J. D., Crame, J. A. (eds.), The Evolutionary Biology of the Bivalvia. The Geological Society of London, London, pp. 11-29.

Steiner, G., and Müller, M. 1996. What can 18S rDNA do for bivalve phylogeny? J. Mol. Evol. 43: 58-70.

Strimmer, K. and von Haeseler, A. 1996. Quartet Puzzling: A Quartet Maximum-Likelihood Method for Reconstructing Tree Topologies. Mol. Biol. Evol. 13: 964-969.

Strimmer, K., and von Haeseler, A. 1997. Likelihood-mapping: A simple method to visualize phylogenetic content of a sequence alignment. Proc. Natl. Acad. Sci. U. S. A. 94: 6815-6819.

Strugnell, J., Norman, M., Jackson, J., Drummond, A. J., and Cooper, A. 2005. Molecular phylogeny of coleoid cephalopods (Mollusca: Cephalopoda) using a multigene approach; the effect of data partitioning on resolving phylogenies in a Bayesian framework. Mol. Phylogenet. Evol. 37: 426-441.

Suarez Soruco, R., 1976. El sistema ordovicico en Bolivia. Revista Tecnica YPF Bolivia 5: 111-123.

Sullivan, J., and Swofford, D. L. 2001. Should we use model-based methods for phylogenetic inference when we know assumptions about among-site rate variation and nucleotide substitution pattern are violated? Syst. Biol. 50: 723-729.

Sullivan, J., Holsinger, K. E., and Simon, C. 1995. Among-site rate variation and phylogenetic analysis of 12S rRNA in Sigmodontine rodents. Mol. Biol. Evol. 12: 988–1001.

Sullivan, J., Swofford, D. L., and Naylor, G. J. P., 1999. The Effect of Taxon Sampling on Estimating Rate Heterogeneity Parameters of Maximum-Likelihood Models. Mol. Biol. Evol. 16: 1347-1356.

Suzuki, Y., Glazko, G. V., and Nei, M., 2002. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. Proc. Natl. Acd. Sci. U. S. A., 99: 16138-16143.

Swofford, D. L. 1999. PAUP 4.0: Phylogenetic Analysis Using Parsimony (And Other Methods). Sinauer Associates, Inc., Sunderland.

Swofford, D. L. 2002. PAUP*. Phylogenetic analysis using parsimony (* and other methods), version 4.0b10, Sinauer Associates, Sunderland, USA.

Swofford, D. L., Olsen, G. J., Waddell, P. J., and Hillis, D. M. 1996. Phylogenetic Inference. In Hillis, D. M., Moritz, D., and Mable, B. K. (eds.), Molecular Systematics, Sinauer Associates, Sunderland, pp. 407-514.

Swofford, D. L., Waddell, P., Huelsenbeck, J., Foster, P., Lewis, P., and Rogers, J. 2001. Bias in Phylogenetic Estimation and Its Relevance to the Choice between Parsimony and Likelihood Methods. Syst. Biol. 50: 525-539.

Talavera, G., and Castresana, J. 2007. Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments. Syst. Biol. 56: 564-577.

Tamura, K., and Nei, M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10: 512-526.

Tamura, K., Dudley, J., Nei, M., and Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24: 1596-1599.

Tavaré, S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. Lectures on Mathematics in the Life Sciences, 17: 57-86.

Taylor, J. D., and Glover, E. A. 2006. Lucinidae (Bivalvia) – the most diverse group of chemosymbiotic mollusks. Zool. J. Linn. Soc. 148: 421-438.

Taylor, J. D., Glover, E. A., and Williams, S. T. 2005. Another bloody bivalve: anatomy and relationships of *Eucrassatella donacina* from south western Australia (Mollusca: Bivalvia: Crassatellidae). In Wells, F. E., Walker, D. I., and Kendrick, G. A. (eds.), The Marine Flora and Fauna of Esperance, Western Australia, Western Australian Museum, Perth, pp. 261-288.

Taylor, J. D., Glover, E. A., and Williams, S. T. 2009. Phylogenetic position of the bivalve family Cyrenoididae – removal from (and further dismantling of) the superfamily Lucinoidea. Nautilus 123: 9-13.

Taylor, J. D., Williams, S. T., and Glover, E. A., 2007a. Evolutionary relationships of the bivalve family Thyasiridae (Mollusca: Bivalvia), monophyly and superfamily status. J. Mar. Biol. Ass. UK 87: 565-574.

Taylor, J. D., Williams, S. T., Glover, E. A., and Dyal, P. 2007b. A molecular phylogeny of heterodont bivalves (Mollusca: Bivalvia: Heterodonta): new analyses of 18S and 28S rRNA genes. Zool. Scr. 36: 587-606.

Taylor, L. R. 1978. Bates, Williams, Hutchinson – A variety of diversities. Symp. R. Ent. Soc. Lond. 9: 1-18.

Tëmkin, I. 2010. Molecular phylogeny of pearl oysters and their relatives (Mollusca, Bivalvia, Pterioidea). BMC Evolutionary Biology 10: 342.

Templeton, A. R. 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. Evolution 37: 221–244.

Terwilliger, R. C., and Terwilliger, N. B. 1985. Molluscan hemoglobins. Comp. Biochem. Phys. B 81: 255-261.

Theologidis, I., Fodelianakis, S., Gaspar, M. B., and Zouros, E., 2008. Doubly uniparental inheritance (DUI) if mitochondrial DNA in *Donax trunculus* (Bivalvia: Donacidae) and the problem of its sporadic detection in Bivalvia. Evol. Int. J. Org. Evol. 62: 959-970.

Therriault, T. W., Docker, M. F., Orlova, M. I., Heath, D. D., and MacIsaac, H. J. 2004. Molecular resolution of the family Dreissenidae (Mollusca: Bivalvia) with emphasis on Ponto-Caspian species, including first report of Mytilopsis leucophaeata in the Black Sea basin. Mol. Phylogenet. Evol. 30: 479-489.

Thompson, J. D., Higgins, D. G., and Gibson, T. J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22: 4673-4680.

Tillier, E. R. M., and Collins, R. A. 1995. Neighbor-Joining and Maximum Likelihood with RNA Sequences: Addressing the Inter-dependence of Sites. Mol. Biol. Evol. 12: 7-15.

Townsend, J. P., 2007. Profiling Phylogenetic Informativeness. Syst. Biol. 56, 222-231.

Tsubaki, R., Kameda, Y., and Kato, M. 2010. Pattern and process of diversification in an ecologically diverse epifaunal bivalve group Pterioidea (Pteriomorphia, Bivalvia). Mol. Phylogenet. Evol. in press.

Tsui, C. K. M., Marshall, W., Yokoyama, R., Honda, D., Lippmeier, J. C., Craven, K. D., Peterson, P. D., and Berbee, M. L. 2009. Labyrinthulomycetes phylogeny and its implications for the evolutionary loss of chloroplasts and gain of ectoplasmic gliding. Mol. Phylogenet. Evol. 50: 129-140.

Tuffley, C., and Steel, M. 1997. Links between maximum likelihood and maximum parsimony under a simple model of site substitution. Bull. Math. Biol. 59: 581-607.

Vane-Wright, R. I., Humphries, C. J., and Williams, P. H. 1991. What to protect? Systematics and the agony of choice. Biol. Conserv. 55: 235-254.

Vermeij, G. J. 1977. The Mesozoic marine revolution: evidence form snails, predators and grazers. Paleobiology 3: 245-258.

Vermeij, G. J. 1987. Evolution and Escalation. Princeton University Press, Princeton.

Vermeij, G. J. 2008. Escalation and its role in Jurassic biotic history. Paleogeogr. Paleoclimatol. Paleoecol. 263: 3-8.

Vokes, H. E., 1980. Genera of the Bivalvia. Palaeontological Research Institution, Ithaca.

von Euler, F. 2001. Selective extinction and rapid loss of evolutionary history in bird fauna. Proc. R. Soc. Lond. B 268: 127-130.

von Salvini-Plawen, L. 1990a. Origin, phylogeny and classification of the phylum Mollusca. Iberus 9: 1-33.

von Salvini-Plawen, L. 1990b. The status of the Caudofoveata and the Solenogastres in the Mediterranean Sea. Lavori S.I.M. 23: 5-30.

von Salvini-Plawen, L., and Steiner, G. 1996. Synapomorphies and plesiomorphies in higher classification of Mollusca. In Taylor, J. D. (ed.), Origin and Evolutionary Radiation of the Mollusca, Oxford University Press, Oxford, pp. 29-51.

Wägele, J. W., and Mayer, C. 2007. Visualizing differences in phylogenetic information content of alignments and distinction of three classes of long-branch effects. BMC Evol. Biol. 7: 147.

Wägele, J. W., and Rödding, F. 1998a. Origin and phylogeny of metazoans as reconstructed with rDNA sequences. Progr. Mol. Subcell. Biol. 21: 45-70.

Wägele, J. W., and Rödding, F. 1998b. A priori estimation of phylogenetic information conserved in aligned sequences. Mol. Phylogenet. Evol. 9: 358-365.

Wägele, J. W., Letsch, H., Klussmann-Kolb, A., Mayer, C., Misof, B., and Wägele, H. 2009. Phylogenetic support values are not necessarily informative: the case of the Serialia hypothesis (a mollusk phylogeny). Frontiers in Zoology 6: 12.

Wagner, P. J., 2008. Paleozoic Gastropod, Monoplacophoran and Rostroconch Database. Paleobiology Database Online Systematics Archives 6.

Wakeley, J. 1994. Substitution-rate variation among sites and the estimation of transition bias. Mol. Biol. Evol. 11: 436-442.

Walker, S. E., and Brett, C. E. 2002. Post-Paleozoic patterns in marine predation: Was there a Mesozoic and Cenozoic marine predatory revolution? Paleontological Society Papers 8: 119-193.

Waller, T. R. 1990. The evolution of ligament systems in the Bivalvia. In Morton, B. (ed.), The Bivalvia. Hong Kong University Press, Hong Kong, pp. 49-71.

Waller, T. R. 1998. Origin of the molluscan class Bivalvia and a phylogeny of major groups. In Johnston, P. A., Haggart, J. W. (eds.), Bivalves: An Eon of Evolution. University of Calgary Press, Calgary, pp. 1-45.

Wang, S., Bao, Z., Li, N., Zhang, L., and Hu, J. 2007. Analysis of the Secondary Structure of ITS1 in Pectinidae: Implications for Phylogenetic Reconstruction and Structural Evolution. Mar. Biotechnol. 9: 231-242.

Wang, Y., and Guo, X. 2004. Chromosomal Rearrangement in Pectinidae Revealed by rRNA Loci and Implications for Bivalve Evolution. Biol. Bull. 207: 247-256.

Wanninger, A., and Haszprunar, G. 2002. Muscle Development in Antalis entalis (Mollusca, Scaphopoda) and Its Significance for Scaphopod Relationships. J. Morphol. 254: 53-64.

Warwick, R. M., and Clarke, K. R. 1995. New “biodiversity” measures reveal a decrease in taxonomic distinctness with increasing stress. Mar. Ecol. Prog. Ser. 129: 301-305.

Warwick, R. M., and Clarke, K. R. 1998. Taxonomic distinctness and environmental assessment. J. Appl. Ecol. 35: 532-543.

Warwick, R. M., and Light, J. 2002. Death assemblages of mollusks on St Martin’s Flats, Isles of Scilly: A surrogate for regional biodiversity? Biodivers. Conserv. 11: 99-112.

Warwick, R. M., and Turk, S. M. 2002. Predicting climate change effects on marine biodiversity: Comparison of recent and fossil molluscan death assemblages. J. Mar. Biol. Ass. UK 82: 847-850.

Wheeler, W. 1996. Optimization alignment: The end of multiple sequence alignment in phylogenetics? Cladistics 12: 1-9.

Wheeler, W., 1999. Fixed character state and the optimization of molecular sequence data. Cladistics 15: 379-385.

Wheeler, W., Gatesy, J., DeSalle, R., 1995. Elision: A method for accommodating multiple molecular sequence alignments with alignment-ambiguous sites. Mol. Phylogenet. Evol. 4: 1-9.

Whitehead, A. 2009. Comparative mitochondrial genomics within and among species of killfish. BMC Evol. Biol. 9: 11.

Whittaker, R. H. 1972. Evolution and measurement of species diversity. Taxon 21: 213-251.

Whittingham, L. A., Slikas, B., Winkler, D. W., and Sheldon, F. H., 2002. Phylogeny of the tree swallow genus, *Tachycineta* (Aves: Hirundinidae), by Bayesian analysis of mitochondrial DNA sequences. Mol. Phylogenet. Evol. 22: 430-441.

Wiens, J. J. 1998. Combining data sets with different phylogenetic histories. Syst. Biol. 47: 568–581.

Wilgenbusch, J. C., and de Queiroz, K. 2000. Phylogenetic relationships among the phrynosomatid sand lizards inferred from mitochondrial DNA sequences generated by heterogeneous evolutionary processes. Syst. Biol. 49: 592-612.

Williams, S. T., Taylor, J. D., and Glover, E. A., 2004. Molecular phylogeny of the Lucinoidea (Bivalvia): non-monophyly and separate acquisition of bacterial chemosymbiosis. J. Moll. Stud. 70: 187-202.

Wilson, N. G., Rouse, G. W., and Giribet, G. 2010. Assessing the molluscan hypothesis Serialia (Monoplacophora + Polyplacophora) using novel molecular data. Mol. Phylogenet. Evol. 54: 187-193.

Winnepenninckx, B., Backeljau, T., and De Wachter, R. 1996. Investigation of Molluscan Phylogeny on the Basis of 18S rRNA Sequences. Mol. Biol. Evol. 13: 1306-1317.

Wood, A. R., Apte, S., MacAvoy, E. S., and Gardner, J. P. A. 2007. A molecular phylogeny of the marine mussel genus *Perna* (Bivalvia: Mytilidae) based on nuclear (ITS1&2) and mitochondrial (COI) DNA sequences. Mol. Phylogenet. Evol. 44: 685-698.

Yang, Z. 1993. Maximum likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. Mol. Biol. Evol. 39: 105–111.

Yang, Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J. Mol. Evol. 39: 306–314.

Yang, Z. 1996. Maximum likelihood models for combined analyses of multiple sequence data. J. Mol. Evol. 42: 587-596.

Yang, Z., and Rannala, B. 2005. Branch-length prior influences Bayesian posterior probability of phylogeny. Syst. Biol. 54: 455-470.

Yonge, C. M. 1939. The protobranchiate Mollusca: a functional interpretation of their structure and evolution. Phil. Trans. R. Soc. B 230: 79-147.

Yonge, C. M. 1969. Functional morphology and evolution within the Carditacea (Bivalvia). Proceedings of the Malacological Society of London 38: 493-527.

Young, N. D., and Healy, J., 2003. GapCoder automates the use of indel characters in phylogenetic analysis. BMC Bioinformatics 4: 6.

Zbawicka, M., Burzyński, A., and Wenne, R., 2007. Complete sequences of mitochondrial genomes from the Baltic mussel Mytilus trossulus. Gene 406: 191-198.

Zouros, E. Oberhauser Ball., A., Saavedra, C., and Freeman, K. R. 1994a. Mitochondrial DNA inheritance. Nature 368: 818.

Zouros, E., Oberhauser Ball, A., Saavedra, C., and Freeman, K. R. 1994b. An unusual type of mitochondrial DNA inheritance in the blue mussel *Mytilus*. Proc. Natl. Acad. Sci. U. S. A. 91: 7463-7467.

CHAPTER 7

APPENDICES

Appendix 2.1. PCR conditions.

		12s		*16s*		*cox1*		*cytb*	
		Annealing	Primers	Annealing	Primers	Annealing	Primers	Annealing	Primers
1	*Anadara ovalis*	50°C 30''	SR-J14197÷ SR-N14745			56°C 20''	coIF÷coIR	48°C 30''	cobF÷cobR
2	*Anodonta woodiana*					48°C 1'	LCO÷HCO	48°C 1'	cobF÷cobR
3	*Anomia sp.*			48°C 1'	16SbrH(32)÷16Sar(34)	56°C-46°C 30''-1'	coIF÷coIR	48°C 30''	cobF÷cobR
4	*Argopecten irradians*	50°C 30''	SR-J14197÷ SR-N14745	48°C 1'	16SbrH(32)÷16Sar(34)	56°C-46°C 30''-1'	coIF÷coIR	55°C-45°C 30''-1'	cobF÷cobR
5	*Astarte* cfr. *castanea*	50°C 30''	SR-J14197÷ SR-N14745					48°C 30''	cobF÷cobR
6	*Barbatia parva*	50°C 30''	SR-J14197÷ SR-N14745			48°C 1'	LCO÷HCO	48°C 1'	cobF÷cobR
7	*Barbatia reeveana*	50°C 30''	SR-J14197÷ SR-N14745			52°C 20''	coIF÷coIR	53°C-43°C 30''-1'	cobF÷cobR
8	*Barbatia* cfr. *setigera*	50°C 30''	SR-J14197÷ SR-N14745			54°C 20''	coIF÷coIR	48°C 1'	cobF÷cobR
9	*Cardita variegata*	50°C 30''	SR-J14197÷ SR-N14745			48°C 1'	LCO÷HCO	48°C 1'	cobF÷cobR
10	*Chlamys livida*	50°C 30''	SR-J14197÷ SR-N14745	48°C 1'	16SbrH(32)÷16Sar(34)	52°C 20''	coIF÷coIR	48°C 1'	cobF÷cobR
11	*Chlamys multistriata*			54°C 2'	16SbrH(32)÷16SDon			48°C 1'	cobF÷cobR
12	*Cuspidaria rostrata*	50°C 30''	SR-J14197÷ SR-N14745			48°C 1'	LCO÷HCO	58°C-48°C 1'	cobF÷cobR
13	*Ensis directus*	46°C 30''	SR-J14197÷ SR-N14745	54°C 2'	16SbrH(32)÷16SDon	56°C-46°C 30''-1'	coIF÷coIR	53°C-43°C 1'	cobF÷cobR
14	*Gafrarium alfredense*	50°C 30''	SR-J14197÷ SR-N14745	48°C 1'	16SbrH(32)÷16Sar(34)			48°C 1'	cobF÷cobR
15	*Gemma gemma*			48°C 1'	16SbrH(32)÷16Sar(34)	52°C 20''	coIF÷coIR	58°C-48°C 1'	cobF÷cobR

16	*Hyotissa hyotis*	50°C 30''	SR-J14197÷ SR-N14745	48°C 1'	16SbrH(32)÷16Sar(34)	52°C 20''	coIF÷coIR	58°C-48°C 1'	cobF÷cobR
17	*Lima pacifica galapagensis*	50°C 30''	SR-J14197÷ SR-N14745	48°C 45''[a]	16SbrH(32)÷16SarL[a]	52°C 20''	coIF÷coIR	53°C-43°C 30''-1'	cobF÷cobR
18	*Mactra corallina*	48°C 1'	SR-J14197÷ SR-N14745	56°C 1'	16SbrH(32)÷16Sar(34)	48°C 1'	LCO÷HCO	48°C 1'	cobF÷cobR
19	*Mactra lignaria*	48°C 1'	SR-J14197÷ SR-N14745	56°C 1'	16SbrH(32)÷16Sar(34)	48°C 1'	LCO÷HCO		
20	*Mya arenaria*							48°C 1'	cobF÷cobR
21	*Nucula nucleus*	50°C 30''	SR-J14197÷ SR-N14745	54°C 2'	16SbrH(32)÷16SDon				
22	*Nuculana commutata*	50°C 30''	SR-J14197÷ SR-N14745			48°C 1'	LCO÷HCO	48°C 1'	cobF÷cobR
23	*Pandora pinna*	50°C 30''	SR-J14197÷ SR-N14745	53°C-43°C 1'20''	16SbrH(32)÷16SarL	48°C 1'	LCO÷HCO	53°C-43°C 1'20''	UCYTBF144F÷UCYTB272R
24	*Pecten jacobaeus*							58°C.48°C 1'	cobF÷cobR
25	*Pinna muricata*	50°C 30''	SR-J14197÷ SR-N14745	48°C 1'	16SbrH(32)÷16Sar(34)	52°C 20''	coIF÷coIR	48°C 1'	cobF÷cobR
26	*Thracia distorta*	50°C 30''	SR-J14197÷ SR-N14745			48°C 1'	LCO÷HCO	48°C 1'	cobF÷cobR
27	*Tridacna derasa*					48°C 1'	LCO÷HCO	48°C 1'	cobF÷cobR
28	*Tridacna squamosa*							48°C 1'	cobF÷cobR
	Transformed inserts	55°C 30''	M13F÷M13R	55°C 30''	M13F÷M13R	55°C 30''	M13F÷M13R	55°C 30''	M13F÷M13R

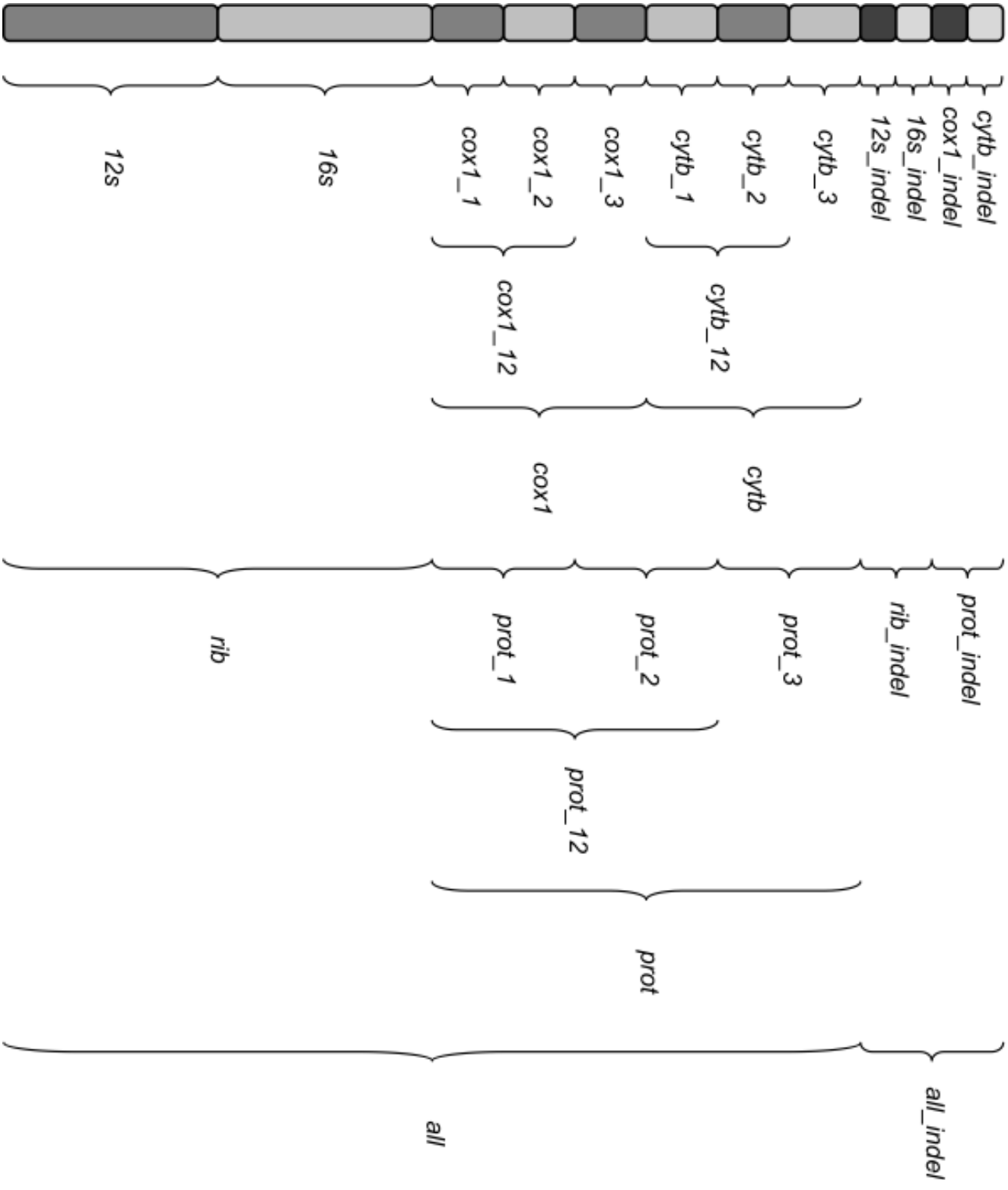
[a] This amplification was carried out with Herculase reaction kit (Stratagene, Cedar Creek, TX, USA), following manufacturer's instructions.

Appendix 2.2. Primer used in this study.

	5’-3’ sequence	Reference
SR-J14197	GTACAYCTACTATGTTACGACTT	Simon et al., 2006
SR-N14745	GTGCCAGCAGYYGCGGTTANAC	Simon et al., 2006
16SbrH(32)	CCGGTCTGAACTCAGATCACGT	Palumbi et al., 1996
16Sar(34)	CGCCTGTTTAACAAAAACAT	modified from Palumbi et al., 1996
16SarL	CGCCTGTTTATCAAAACAT	Palumbi et al., 1996
16SDon	CGCCTGTTTATCAAAAACAT	Kocher et al., 1989
LCO1490	GGTCAACAAATCATAAAGATATTGG	Folmer et al. 1994
HCO2198	TAAACTTCAGGGTGACCAAAAAATCA	Folmer et al., 1994
COIF	ATYGGNGGNTTYGGNAAYTG	Matsumoto, 2003
COIR	ATNGCRAANACNGCNCCYAT	Matsumoto, 2003
CobF	GGWTAYGTWYTWCCWTGRGGWCARAT	Passamonti, 2007
CobR	GCRTAWGCRAAWARRAARTAYCAYTCWGG	Passamonti, 2007
UCYTB144F	TGAGSNCARATGTCNTWYTG	Merritt et al., 1998
UCYTB272R	GCRAANAGRAARTACCAYTC	Merritt et al., 1998
M13F	GTAAAACGACGGCCAGT	
M13R	CAGGAAACAGCTATGAC	

Appendix 2.3. GenBank accession numbers of sequences used in this study. Bold sequences were obtained for this work.

	12s	*16s*	*cox1*	*cytb*
Acanthocardia tubercolata	DQ632743	DQ632743	DQ632743	DQ632743
Acesta excavata	AM494885	AM494899	AM494909	AM494922
Anadara ovalis	**GQ166533**		**GQ166571**	**GQ166592**
Anodonta woodiana F		DQ073815	EF440349	**GQ166594**
Anomia sp.		**GQ166557**	**GQ166573**	**GQ166595**
Argopecten irradians	**GQ166535**	**GQ166558**	**GQ166574**	**GQ166596**
Astarte castanea			AF120662	
Astarte cfr. *castanea*	**GQ166536**			**GQ166597**
Barbatia parva	**GQ166537**		**GQ166575**	**GQ166599**
Barbatia reeveana	**GQ166538**		**GQ166576**	**GQ166600**
Barbatia cfr. *setigera*	**GQ166539**		**GQ166577**	**GQ166601**
Cardita variegata	**GQ166540**		**GQ166578**	**GQ166605**
Chlamys livida	**GQ166541**	**GQ166559**	**GQ166579**	**GQ166606**
Chlamys multi striata	AJ571604	**GQ166560**		**GQ166607**
Crassostrea gigas	AF177226	AF177226	AF177226	AF177226
Crassostrea hongkongensis F	EU266073	EU266073	EU266073	EU266073
Crassostrea viriginica	AY905542	AY905542	AY905542	AY905542
Cuspidaria rostrata	**GQ166542**		**GQ166580**	**GQ166608**
Donax faba F			AB040844	
Donax trunculus F		EF417549		EF417548
Dreissena polymorpha		DQ280038	AF120663	DQ072117
Ensis directus	**GQ166543**	**GQ166561**	**GQ166581**	**GQ166610**
Gafrarium alfredense	**GQ166544**	**GQ166562**		**GQ166611**
Gemma gemma		**GQ166563**	**GQ166582**	**GQ166612**
Graptacme eborea	AY484748	AY484748	AY484748	AY484748
Haliotis rubra	AY588938	AY588938	AY588938	AY588938
Hiatella arctica	DQ632742	DQ632742	DQ632742	DQ632742
Hyotissa hyotis	**GQ166545**	**GQ166564**	**GQ166583**	**GQ166613**
Hyriopsis cumini	FJ529186	FJ529186	FJ529186	FJ529186
Inversidens japanensis F	AB055625	AB055625	AB055625	AB055625
Katharina tunicata	U09810	U09810	U09810	U09810
Lampsilis ornata	AY365193	AY365193	AY365193	AY365193
Lima pacifica galapagensis	**GQ166548**	**GQ166565**	**GQ166584**	**GQ166616**
Mactra corallina	**GQ166550**	**GQ166566**	**GQ166585**	**GQ166617**
Mactra lignaria	**GQ166551**	**GQ166567**	**GQ166586**	
Mimachlamys nobilis	FJ415225	FJ415225	FJ415225	FJ415225
Mizuhopecten yessoensis	AB271769	AB271769	AB271769	AB271769
Mya arenaria		AY377618	AF120668	**GQ166619**
Mytilus edulis F	AY484747	AY484747	AY484747	AY484747
Mytilus galloprovincialis F	AY497292	AY497292	AY497292	AY497292
Mytilus trossulus F	DQ198231	DQ198231	DQ198231	DQ198231
Nucula nucleus	**GQ166552**	**GQ166568**	AM696252	
Nuculana commutata	**GQ166553**		**GQ166587**	**GQ166622**
Pandora pinna	**GQ166554**	**GQ166569**	**GQ166588**	**GQ166623**
Pecten jacobaeus	AJ571596	AJ245394	AY377728	**GQ166624**
Pinctada margariti fera	AB250256	AB214436	AB259166	
Pinna muricata	**GQ166555**	**GQ166570**	**GQ166589**	**GQ166625**
Placopecten magellanicus	DQ088274	DQ088274	DQ088274	DQ088274
Sinonovacula constricta	EU880278	EU880278	EU880278	EU880278
Solemya velesiana				AM293670
Solemya velum		DQ280028	U56852	
Spisula solidissima				AF205083
Spisula solidissima solidissima			AY707795	
Spisula subtruncata		AJ548774		
Spondylus gaederopus	AJ571607	AJ571621		
Spondylus varius			AB076909	
Thais clavigera	DQ159954	DQ159954	DQ159954	DQ159954
Thracia distorta	**GQ166556**		**GQ166590**	**GQ166626**
Tridacna derasa		AF122976	**GQ166591**	**GQ166627**
Tridacna squamosa		AF122978	EU346361	**GQ166628**
Venerupis philippinarum F	AB065375	AB065375	AB065375	AB065375



Appendix 2.4. Partitions used in this study. Bar corresponds to the complete concatenated alignment, over both nucleotides and indels coded as 0/1.

Appendix 2.5. Comparison between Maximum Likelihood and Bayesian estimates of models' main parameters. C. I., Confidence Interval.

	Parameter	Bayesian 95% C. I.			Maximum Likelihood	
		Lower	Mean	Upper	Parameter	(deviation[a])
rib	p(A)	0.304345	0.317559	0.329482	0.332300	0.002818
	p(C)	0.137214	0.146464	0.155187	0.138200	
	p(G)	0.218883	0.230900	0.242320	0.218900	
	p(T)	0.291851	0.305077	0.318489	0.310500	
	r(A<->C)	0.076302	0.089823	0.102803	0.086376	
	r(A<->G)	0.220000	0.241830	0.263580	0.236895	
	r(A<->T)	0.122571	0.136051	0.149031	0.110587	0.011984
	r(C<->G)	0.071964	0.085993	0.101012	0.110587	0.009575
	r(C<->T)	0.331913	0.357767	0.385415	0.369179	
	r(G<->T)	0.077327	0.088536	0.100488	0.086376	
	alpha	0.824031	0.918511	1.017489	0.843200	
	pinvar	0.053833	0.074948	0.096944	0.072100	
prot	r(A<->C)	0.097619	0.114134	0.131782	0.099701	
	r(A<->G)	0.257078	0.276088	0.296525	0.227407	0.029671
	r(A<->T)	0.128236	0.140986	0.153492	0.052479	0.075757
	r(C<->G)	0.170922	0.190929	0.210337	0.202161	
	r(C<->T)	0.139429	0.149614	0.160840	0.378039	0.217199
	r(G<->T)	0.115544	0.128249	0.141076	0.040213	0.075331

[a] Deviation is shown only for estimates falling outside Bayesian confidence interval.

Appendix 2.6. Subtrees used for assessing parameter estimate accurateness.

Taxon labels:

1	*Acanthocardia tubercolata*	20	*Ensis directus*	39	*Mytilus trossulus* F
2	*Acesta excavata*	21	*Gafrarium alfredense*	40	*Nucula nucleus*
3	*Anadara ovalis*	22	*Gemma gemma*	41	*Nuculana commutata*
4	*Anodonta woodiana* F	23	*Graptacme eborea*	42	*Pandora pinna*
5	*Anomia* sp.	24	*Haliotis rubra*	43	*Pecten jacobaeus*
6	*Argopecten irradians*	25	*Hiatella arctica*	44	*Pinctada margaritifera*
7	*Astarte* cfr. *castanea*	26	*Hyotissa hyotis*	45	*Pinna muricata*
8	*Barbatia parva*	27	*Hyriopsis cumingii* F	46	*Placopecten magellanicus*
9	*Barbatia reeveana*	28	*Inversidens japanensis* F	47	*Sinonovacula constricta*
10	*Barbatia* cfr. *setigera*	29	*Katharina tunicata*	48	*Solemya* sp.
11	*Cardita variegata*	30	*Lampsilis ornata*	49	*Spisula* sp.
12	*Chlamys livida*	31	*Lima pacifica galapagensis*	50	*Spondylus* sp.
13	*Chlamys multistriata*	32	*Mactra corallina*	51	*Thais clavigera*
14	*Crassostrea gigas*	33	*Mactra lignaria*	52	*Thracia distorta*
15	*Crassostrea hongkongensis*	34	*Mimachlamys nobilis*	53	*Tridacna derasa*
16	*Crassostrea virginica*	35	*Mizuhopecten yessoensis*	54	*Tridacna squamosa*
17	*Cuspidaria rostrata*	36	*Mya arenaria*	55	*Venerupis philippinarum* F
18	*Donax* sp. F	37	*Mytilus edulis* F		
19	*Dreissena polymorpha*	38	*Mytilus galloprovincialis* F		

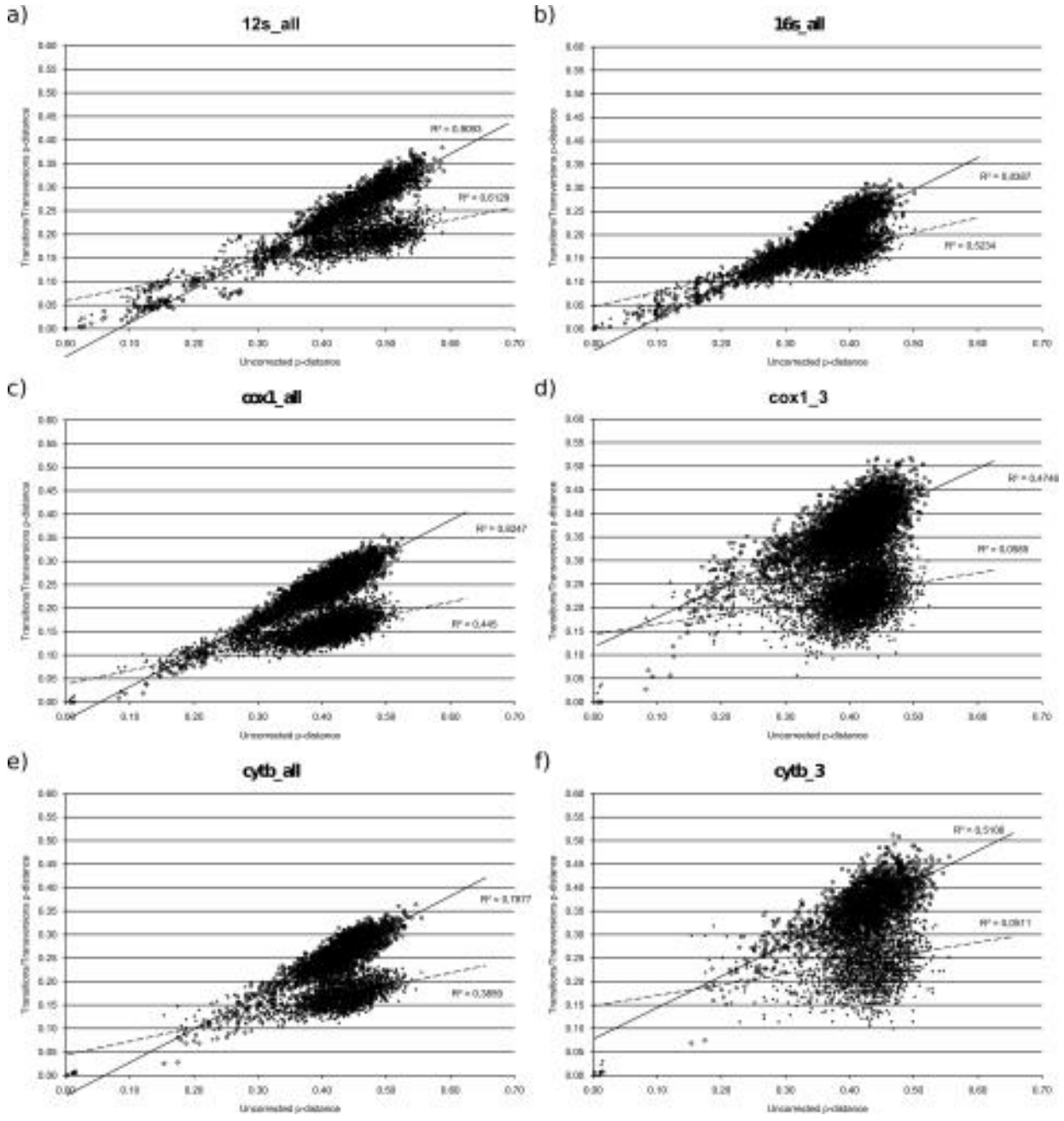
Tree *tM3*:

(51,29,24,23,(((((7,11),17),(((1,(53,54)),25),((20,47),(((32,33),49),((21,22),55),(19,36)),18)),(((37,38,39),((2,31),(5,((35,13,(12,34),((6,43),46)),50))),((((14,15),16),26),44,45),(3,((10,9),8)),41),(42,52))),((27,28),4,30)),(40,48)));

Subtrees:

1	(51,29,24,23,((((17),(((1,(53,54)),25),((20,47),(((32,33),49),((21,22),55),(19,36)),18)),(((37,38,39),((2,31),(5,((35,13,(12,34),((6,43),46)),50))),((((14,15),16),26),44,45),(3,((10,9),8)),41),(42,52))),((27,28),4,30)),(40,48)));
2	(51,29,24,23,((((((1,(53,54)),25),((20,47),(((32,33),49),((21,22),55),(19,36)),18)),(((37,38,39),((2,31),(5,((35,13,(12,34),((6,43),46)),50))),((((14,15),16),26),44,45),(3,((10,9),8)),41),(42,52))),((27,28),4,30)),(40,48)));
3	(51,29,24,23,(((((7,11),17),(((1,(53,54)),25),((20,47),(((32,33),49),((21,22),55),(19,36)),18)),(((37,38,39),((2,31),(5,((35,13,(12,34),((6,43),46)),50))),((((14,15),16),26),44,45),(3,((10,9),8)),41),(42,52)))),(40,48)));
4	(51,29,24,23,(((((7,11),17),(((1,(53,54)),25),((20,47),(((32,33),49),((21,22),55),(19,36)),18)),(((37,38,39),((2,31),(5)),((((14,15),16),26),44,45),(3,((10,9),8)),41),(42,52))),((27,28),4,30)),(40,48)));
5	(51,29,23,(((((7,11)),(((1,(53,54)),25),((20,47),(((32,33),49),((21,22),55),(19,36)))),(((37,38,39),((2,31),(5,((35,13,(12,34),((6,43),46)),50))),((((14,15),16),26),44,45),(3,((10,9),8)),41))),((27,28)))));
6	(51,29,24,23,(((((7,11),17),(((1),25),((20,47),((49),((21,22),55),(19,36)),18)),((((2,31),(5,((35,13,(12,34),((6,43),46)),50))),((((14,15),16),26),44,45),(3),41),(42,52))),((27,28),4,30)),(40,48)));
7	(51,24,23,(((((7,11),17),((((53,54)),25),((20),(((32,33),49),((21,22)),(19,36)),18)),(((38,39),((2,31),(5,((35,13,(12),((6,43),46)),50))),((((14),16),26),44,45),(3,((9),8))),(52))),((27),4,30)),(40)));
8	(23,(((((7,11),17),(((((32,33),49),((21,22),55),(19,36)),18)),(((37,38,39),((2,31),(5,((35,13,(12,34),((6,43),46)),50))),((((14,15),16),26)),(3,((10,9),8)),41),(42,52))),((28),4)),(48)));
9	(51,29,24,23,(((((7,11),17),((41,(37,38,39),((2,31),(5,((35,13,(12,34),((6,43),46)),50))),((((14,15),16),26),44,45),(3,((10,9),8))),(42,52))),((27,28),4,30)),(40,48)));
10	(51,29,(((((7),17),(((1,(54)),25),((20),(((32)),((21,22)),(19,36)))),(((37,38),((2,31),(((35,(12,34),((6,43),46)),50))),((((14,15)),26),44),(3,((10),8)),41),(52))),((27,28),30)),(40)));
11	((((7,11),17),(((1,(53,54))),((((32,33),49),((21,22),55)))),(((37,38,39),((5,((35,13,(12,34),((6,43))),50))),((((14,15),16)),44,45),(((10,9),8))))),((27,28),4,30));
12	(51,29,24,23,(((((7,11),17),(((1,(53,54)),25),((20,47),(((32,33),49),((21,22),55),(19,36)),18)),(((((10,9),8))))),((27,28),4,30)),(40,48)));
13	(29,23,(((((11)),(((1,(54))),((20),(((32),49),((22)),(19)),18)),(((38),((2),(5,((13,(34),((43))),50))),((((15)),26),45),(((10),8))),(42))),((27),4)),(40)));
14	(23,((((17),((((54))),((20,47))),((((2,31),(5,((13,(34),((6),46))))),((((14,15),16)),44,45),41),(42))),((27),4,30)),(40,48)));
15	(29,24,23,((((((1,(53,54))),((20),(((32,33),49),((22)),(19)))),(((38,39),((5,((13,(34),(46)))))),((((14))),44)))),((27)))));
16	(((((7,11),17),((25),(((36)),18)),(((37),((5)),(((16))),41),(42,52))),((27,28),4,30)),(40,48));
17	(((((53,54))),((((32,33)),(55)))),(((37,38,39),(((((12,34)))))),((((14,15),16))),(((10,9),8)))));
18	(51,24,(((((7),17),(((((33)),(19)))),((((2),(((35)))),((26))),(52))),((28),30)),(40,48)));
19	((2,31),(5,((35,13,(12,34),((6,43),46)),50)));
20	(((1,(53,54)),25),((20,47),(((32,33),49),((21,22),55),(19,36)),18));
21	(29,(((((11)),((((49)))),((((5,(50))),((8)),41),(42))),((27))),(48)));
22	(51,(((((7)),(((20))),(((37,38,39),((((14)))))))),(40)));
23	(((((((21))))),(((45)),(52))),(4)),(48));

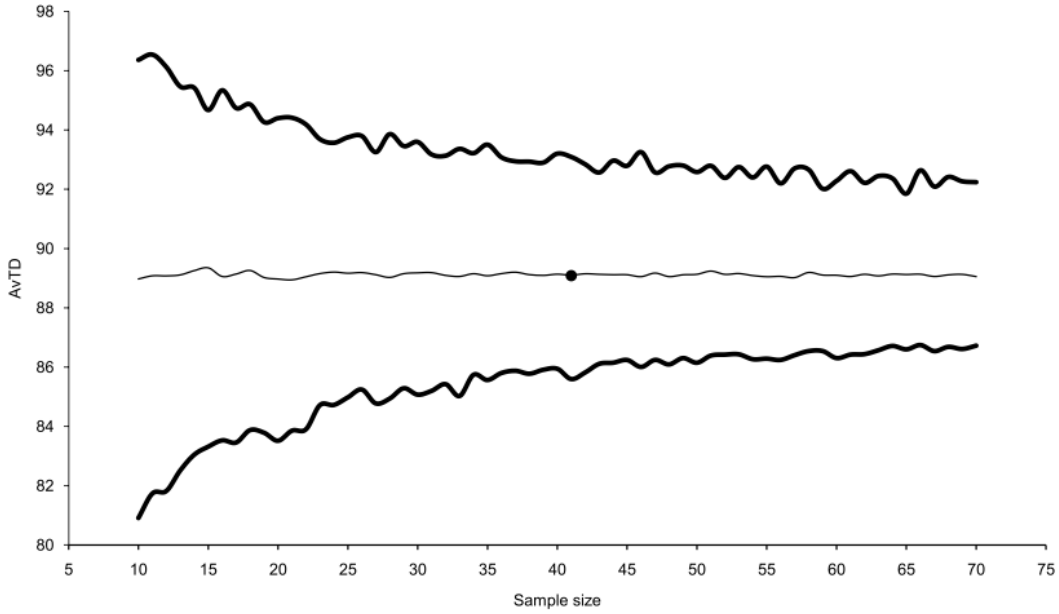
```
(51,29,24,23,(((((7,11),17),(((1,(53,54)),25),(18,(20,47),(((32,33),49),((21,22),55),(19,36)))),((41,(37,38,39),((2,31),(5,((35,13,(12,34),((6,43),46)),50))),(((14,15),26),44,45),(3,((10,9),8))),(42,52))),((27,28),4,30)),(40,48)));
(51,29,24,23,(((((7,11),17),(((1,(53,54)),25),(18,(20,47),(((32,33),49),((21,22),55),(19,36)))),((41,(37,38,39),((2,31),(5,((35,13,(12,34),((6,43),46)),50))),(44,45),(3,((10,9),8))),(42,52))),((27,28),4,30)),(40,48)));
(51,29,23,(((((7,11),17),((1,25),(18,(20,47),(((32,33),49),((21,22),55),(19,36)))),((41,((2,31),(5,((35,13,(12,34),((6,43),46)),50))),((((14,15),16),26),44,45),(3,((10,9),8))),(42,52))),((27,28),4,30)),(40,48)));
(51,29,24,23,(((((7,11),17),(((1,(53,54)),25),(18,(20,47))),((41,(37,38,39),((2,31),(5,((35,13,(12,34),((6,43),46)),50))),((((14,15),16),26),44,45),(3,((10,9),8))),(42,52))),((27,28),4,30)),(40,48)));
(51,24,23,((((11,17),(((53,1),25),(18,(20,47),((32,49),((21,22),55),(19,36)))),((3,(37,39),((2,31),(5,((46,35,13,(12,34)),50))),((((14,15),16),26),45)),(42,52))),((27,28),4)),(40,48)));
(51,29,24,23,((((((1,(53,54)),25),(18,(20,47),(((32,33),49),(19,36)))),((41,(37,38,39),((2,31),(5,((35,13,(12,34),((6,43),46)),50))),((((14,15),16),26),44,45)),(42,52))),((27,28),4,30)),(40,48)));
(51,29,24,23,((((7,17),(((1,(53,54)),25),((20,47),(36,((32,33),49),((21,22),55)))),((41,38,(31,(5,((43,35,13,(12,34)),50))),((14,26),44),(8,3)),(42,52))),(4,28,30)),(40,48)));
(51,((((7,17),(((1,(53,54)),25),((20,47),(36,((32,33),49),((21,22),55)))),((41,38,(31,(5,((43,35,13,(12,34)),50))),((14,26),44),(8,3)),(42,52))),(4,28,30)),(40,48)));
(51,29,24,23,(((18,(41,(37,38,39),((2,31),(5,((35,13,(12,34),((6,43),46)),50))),((((14,15),16),26),44,45),(3,((10,9),8))),((7,11),17)),((27,28),4,30)),(40,48)));
(51,29,23,(40,((7,(((53,1),25),((20,47),((21,22),(32,49),(19,36)))),(42,(41,((2,31),(5,((46,35,(12,34)),50))),((15,16),44,45),(9,3)))),(27,28))));
(29,23,((((7,11),(((53,54),25),((20,47),((32,33),(21,22)))),((41,(10,9),(38,39),((13,(6,43),(12,34)),(2,31)),(((14,15),26),45)),(42,52))),(27,28)),(40,48)));
(51,29,24,23,((7,(((1,(53,54)),25),(18,(((32,33),49),((21,22),55)))),(((2,31),(5,((35,13,(12,34),((6,43),46)),50))),(42,52))),((27,28),4,30)));
(40,((((7,11),17),((1,(53,54)),(18,(20,47),(36,32,(22,55)))),(42,(41,39,((15,16),26),(2,(5,((35,34,(6,46)),50))),(8,3)))),(4,27)));
(51,24,(((((1,(53,54)),25),(18,(20,47))),(((37,38,39),(26,44,45),(3,((10,9),8)),(5,((35,((6,43),46)),50))),(42,52))),(40,48)));
((((2,31),(5,((35,13,(12,34),((6,43),46)),50))),((((14,15),16),26),44,45)),((20,47),(((32,33),49),((21,22),55),(19,36))));
(48,((17,((25,54),(47,(22,19,(33,49)))),((41,39,(14,26),(31,((13,12),50)),(3,(8,9))),(42))),(4,27)));
(51,29,23,(40,(28,((7,17),(1,(47,18,(36,(21,55)))),(52,(41,(38,39),(31,(5,((34,13,6),50))),(26,45)))))));
((40,48),(((27,28),4,30),((41,(37,39),(31,(5,((34,(6,46)),50))),((14,26),45),(9,3)),(42,52))));
(51,29,24,23,(((1,(53,54)),25),(18,(20,47),(((32,33),49),((21,22),55),(19,36)))));
((40,48),((41,((2,31),(5,((35,13),50))),((((14,15),16),26),44,45)),(42,52)));
(51,(40,((11,((32,22),(25,54)),(52,(41,39,8,6,(26,45)))),(4,27))));
(29,24,23,((((42,52),((7,11),17)),((27,28),30)),(40,48)));
(((27,28),4,30),(41,3,(2,31),(26,44,45)));
(23,(40,(30,(18,((7,11),17),(42,31)))));
(51,(((27,28),4,30),(40,48)));
(((6,43),46),(12,34));
(24,(55,(37,(10,9))));
```



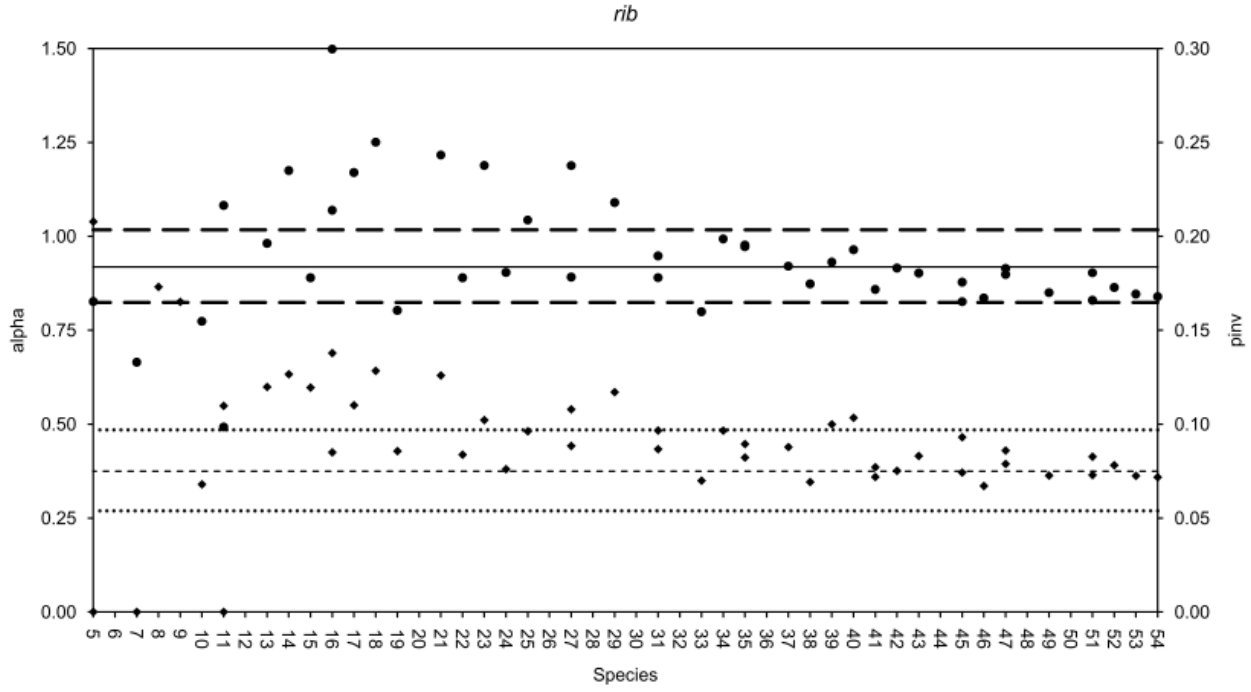
Appendix 2.7. Saturation plots. In each plot, p-distance for either transitions or transversions were plotted on global pairwise comparisons and linear regression was computed. All correlations were highly significant. Open circles, transversions; crosses, transitions; a) 12s_all, complete *12s* alignment; b) 16s_all, complete *16s* alignment; c) cox1_all, complete *cox1* alignment; d) cox1_3, saturation test only on *cox1* third codon positions; e) cytb_all, complete *cytb* alignment; f) cytb_3, saturation test only on *cytb* third codon positions.

Appendix 2.8. Codon model parameters as obtained from **sump** command in MrBayes. C. I., Confidence Interval.

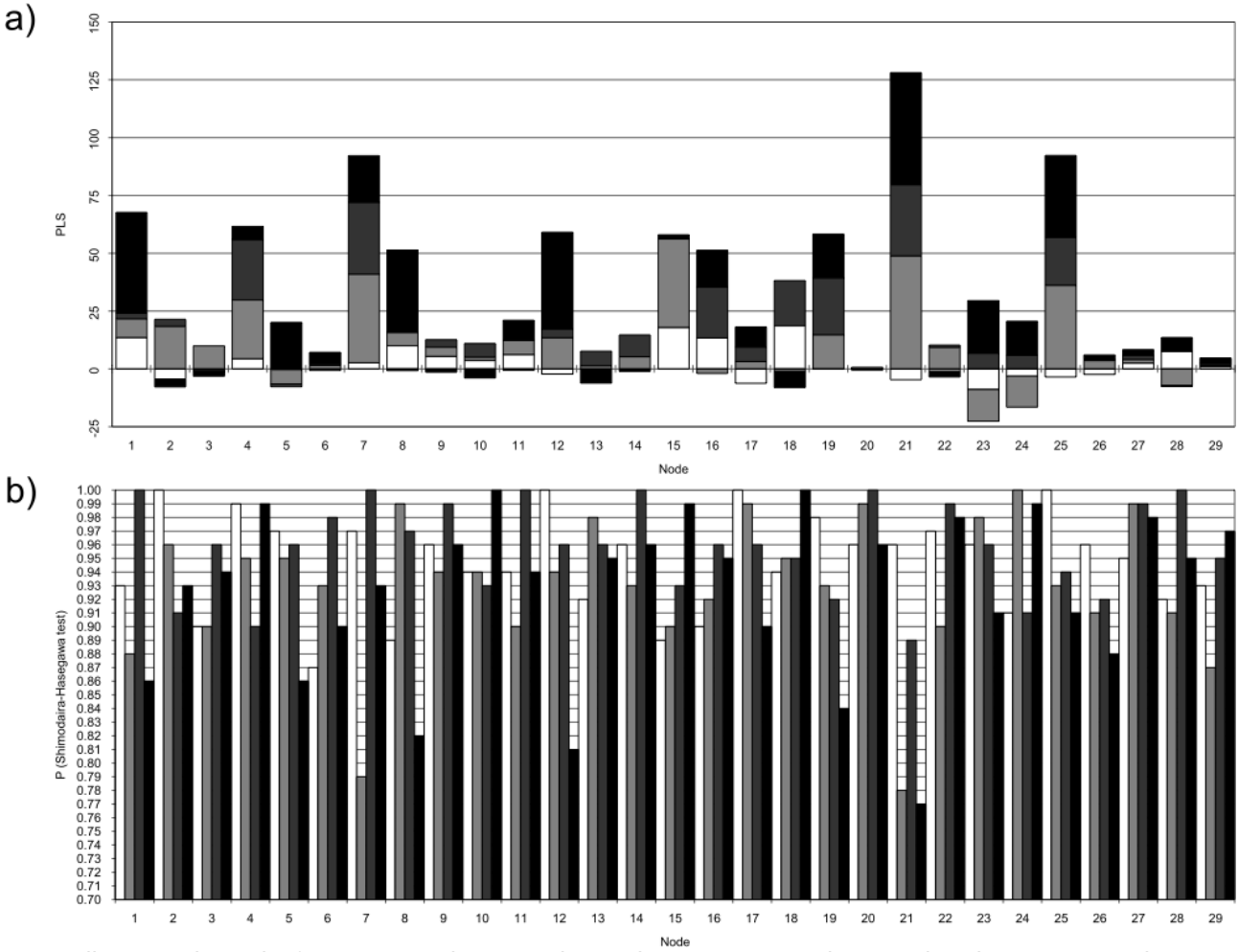
Parameter	Mean	Variance	Lower 95% C.I.	Upper 95% C.I.
ω_1	0.005319	0.000000	0.004446	0.006369
ω_2	0.044846	0.000005	0.040437	0.049470
ω_3	0.130884	0.000044	0.118953	0.144549
p(1)	0.403129	0.000877	0.346468	0.462589
p(2)	0.395454	0.000901	0.336515	0.454468
p(3)	0.201416	0.000541	0.157945	0.250160



Appendix 2.9. Average Taxonomic Distinctness (AvTD) funnel plot for the bivalve data set used in this work; analysis was performed with 1,000 replicates. Random subsample sizes are shown on *x*-axis, whereas AvTD values are shown on *y*-axis. Our sample is shown as the black dot. Thin line, AvTD mean; lower thick line, lower 95% confidence limit; upper thick line, maximum AvTD.



Appendix 2.10. Evolutionary model estimates plotted on subtree sizes. Only gamma shaping parameter (alpha, left axis; filled circles) and invariable sites proportion (pinv, right axis; filled diamonds) for *rib* partition are shown for clarity. “True” estimates from Bayesian Analysis are shown as follows. Continuous line: mean alpha; long-dashed lines: 95% alpha confidence interval; short-dashed line: mean pinv; dotted lines: 95% pinv confidence interval. All parameters are extensively listed in Appendix 2.5 and all subtrees are described in Appendix 2.6. Some extreme values are out of axis scale and are not shown.



Appendix 2.11. a) results from PLS analyses: node numbers are reported on *x*-axis, whereas PLS values are reported on *y*-axis; white, *12s*; light grey, *16s*; heavy grey, *cox1*; black, *cytb*. b) Shimodaira-Hasegawa significance test from 100 bootstrap replicates: P values are shown on *y*-axis; *x*-axis and colour code as above.

Appendix 3.1. Table showing the composition of our real and simulated samples of bivalves. Taxonomy is reported for each Genus; a plus "+" sign indicates the presence of that Genus in that sample.

Subclass	Order	Family	Genus	R1	R2	R3	R4	S1	S2	S3	S4
HETERODONTA	CHAMIDA	MACTRIDAE	*ALIOMACTRA* Stephenson, 1952 [1953]							+	
HETERODONTA	VENEROIDA	CARDITIDAE	*AMEKIGLANS* Eames, 1957							+	
HETERODONTA	VENEROIDA	CONDYLOCARDIIDAE	*AMERICUNA* Klappenbach, 1962							+	
PTERIOMORPHIA	ARCIDA	ARCIDAE	*ANADARA* Gray, 1847	+	+	+	+				+
PALAEOHETERODONTA	UNIONIDA	UNIONIDAE	*ANODONTA* Lamarck, 1799			+	+				
PTERIOMORPHIA	OSTREOIDA	ANOMIIDAE	*ANOMIA* Linnaeus, 1758	+	+	+	+				
PTERIOMORPHIA	OSTREOIDA	PECTINIDAE	*ARGOPECTEN* Monterosato, 1899	+	+	+	+	+			
HETERODONTA	CHAMIDA	ASTARTIDAE	*ASTARTE* Sowerby, 1816	+	+	+	+				
HETERODONTA	VENEROIDA	BABINKIDAE	*BABINKA* Barrande, 1881							+	
PTERIOMORPHIA	ARCIDA	ARCIDAE	*BARBATIA* Gray, 1840	+	+	+	+				
HETERODONTA	VENEROIDA	BERNARDINIDAE	*BERNARDINA* Dall, 1910							+	
HETERODONTA	VENEROIDA	FIMBRIIDAE	*BERNAYIA* Cossmann, 1887							+	
PTERIOMORPHIA	OSTREOIDA	OSTREIDAE	*BOSOSTREA* Chiplonkar & Badve, 1978					+			
PALAEOHETERODONTA	MODIOMORPHOIDA	MODIOMORPHIDAE	*BYSSODESMA* Isberg, 1934					+			
HETERODONTA	VENEROIDA	CARDITIDAE	*CARDITA* Bruguière, 1792		+	+	+				
HETERODONTA	CHAMIDA	CARDIIDAE	*CARDIUM* Linne, 1758	+	+	+	+	+			
HETERODONTA	CHAMIDA	CARDIIDAE	*CERASTODERMA* Poli, 1795	+	+	+					
HETERODONTA	CHAMIDA	VENERIDAE	*CHAMELEA* Mörch, 1853								+
HETERODONTA	VENEROIDA	CHLAMYDOCONCHIDAE	*CHLAMYDOCONCHA* Dall, 1884							+	
PTERIOMORPHIA	OSTREOIDA	PECTINIDAE	*CHLAMYS* Röding, 1798	+	+	+	+				
HETERODONTA	CHAMIDA	CORBICULIDAE	*CORBICULA* Megerle von Mühlfeld, 1811	+	+						
PTERIOMORPHIA	OSTREOIDA	OSTREIDAE	*CRASSOSTREA* Sacco, 1897	+	+	+	+				
ANOMALODESMATA	PHOLADOMYOIDA	CUSPIDARIIDAE	*CUSPIDARIA* Nardo, 1840			+	+	+			
HETERODONTA	VENEROIDA	CYRENOIDIDAE	*CYRENOIDA* de Joannis, 1835							+	
HETERODONTA	CHAMIDA	DONACIDAE	*DONAX* Linnaeus, 1758	+	+	+	+				+
HETERODONTA	CHAMIDA	DREISSENIDAE	*DREISSENA* Beneden, 1835	+	+	+	+				
HETERODONTA	CHAMIDA	PHARIDAE	*ENSIS* Schumacher, 1817	+	+	+	+	+			+
HETERODONTA	CHAMIDA	RZEHAKIIDAE	*ERGENICA* Zhizchenko, 1953							+	
HETERODONTA	CHAMIDA	VENERIDAE	*GAFRARIUM* Röding, 1798		+	+	+				
HETERODONTA	CHAMIDA	VENERIDAE	*GEMMA* Deshayes, 1853	+		+	+				
HETERODONTA	MYIDA	HIATELLIDAE	*HIATELLA* Daudin in Bosc, 1801	+	+	+	+				
PALAEOHETERODONTA	MODIOMORPHOIDA	MODIOMORPHIDAE	*HIPPOMYA* Salter, 1864						+		
PTERIOMORPHIA	OSTREOIDA	GRYPHAEIDAE	*HYOTISSA* Stenzel, 1971		+	+	+				
PALAEOHETERODONTA	UNIONIDA	UNIONIDAE	*HYRIOPSIS* Conrad, 1853			+	+				
PTERIOMORPHIA	MYTILIDA	MYTILIDAE	*IDAS* Jeffreys, 1876						+		
PTERIOMORPHIA	PTERIIDA	INOCERAMIDAE	*INOCERAMUS* J. Sowerby, 1814					+			
PALAEOHETERODONTA	UNIONIDA	UNIONIDAE	*INVERSIDENS* Haas, 1911	+	+	+	+				
PROTOBRANCHIA	NUCULOIDA	ISOARCIDAE	*ISOARCA* Münster, 1842						+		
PALAEOHETERODONTA	UNIONIDA	UNIONIDAE	*LAMPSILIS* Rafinesque, 1820	+	+	+	+				
HETERODONTA	VENEROIDA	LASAEIDAE	*LASAEA* Leach in Brown, 1827							+	
ANOMALODESMATA	PHOLADOMYOIDA	LATERNULIDAE	*LATERNULA* Röding, 1798						+		
PTERIOMORPHIA	LIMIDA	LIMIDAE	*LIMA* Bruguière, 1797	+	+	+	+				
PROTOBRANCHIA	NUCULOIDA	NUCULANIDAE	*LONGINUCULANA* Saveliev, 1958						+		
PTERIOMORPHIA	OSTREOIDA	OSTREIDAE	*LOPHA* Röding, 1798	+							
ANOMALODESMATA	PHOLADOMYOIDA	LYONSIIDAE	*LYONSIA* Turton, 1822					+	+		
HETERODONTA	CHAMIDA	TELLINIDAE	*MACOMA* Leach, 1819								+
HETERODONTA	CHAMIDA	MACTRIDAE	*MACTRA* Linne, 1767			+	+	+			+
ANOMALODESMATA	PHOLADOMYOIDA	MARGARITARIIDAE	*MARGARITARIA* Conrad, 1849						+		
PALAEOHETERODONTA	UNIONIDA	MARGARITIFERIDAE	*MARGARITIFERA* Schumacher, 1816					+	+		

ANOMALODESMATA	PHOLADOMYOIDA	MEGADESMIDAE	*MEGADESMUS* J. De Sowerby, 1838						+		
HETERODONTA	CHAMIDA	VENERIDAE	*MERCENARIA* Schumacher, 1817	+							
PTERIOMORPHIA	OSTREOIDA	PECTINIDAE	*MIMACHLAMYS* Iredale, 1929			+	+				+
PTERIOMORPHIA	OSTREOIDA	PECTINIDAE	*MIZUHOPECTEN* Masuda, 1963	+	+	+	+				
HETERODONTA	MYIDA	MYIDAE	*MYA* Linnaeus, 1758	+	+	+	+				
PALAEOHETERODONTA	UNIONIDA	MYCETOPODIDAE	*MYCETOPODA* d'Orbigny, 1835						+		
PALAEOHETERODONTA	TRIGONIOIDA	MYOPHORIIDAE	*MYOPHORIA* Bronn, 1834						+		
PTERIOMORPHIA	MYTILIDA	MYTILIDAE	*MYTILUS* Linnaeus, 1758	+	+	+	+	+			+
PALAEOHETERODONTA	TRIGONIOIDA	NAKAMURANAIIDAE	*NAKAMURANAIA* Suzuki, 1943						+		
PROTOBRANCHIA	NUCULOIDA	NUCULIDAE	*NUCULA* Lamarck, 1799	+	+	+	+	+	+		
PROTOBRANCHIA	NUCULOIDA	NUCULANIDAE	*NUCULANA* Link, 1807			+	+	+			
PROTOBRANCHIA	PRAECARDIOIDA	CARDIOLIDAE	*ONTARIA* Clarke, 1904						+		
PTERIOMORPHIA	OSTREOIDA	OSTREIDAE	*OSTREA* Linnaeus, 1758								+
PROTOBRANCHIA	NUCULOIDA	NUCULIDAE	*PALAEONUCULA* Quenstedt, 1930						+		
ANOMALODESMATA	PHOLADOMYOIDA	PANDORIDAE	*PANDORA* Bruguière, 1797			+	+		+		
PROTOBRANCHIA	PRAECARDIOIDA	PRAECARDIIDAE	*PARACARDIUM* Barrande, 1881							+	
PTERIOMORPHIA	OSTREOIDA	PECTINIDAE	*PECTEN* Müller, 1776	+	+	+	+				+
HETERODONTA	CHAMIDA	PHARIDAE	*PHARUS* Gray, 1840								+
ANOMALODESMATA	PHOLADOMYOIDA	PHOLADOMYIDAE	*PHOLADOMYA* G. B. Sowerby I, 1823					+			
PTERIOMORPHIA	PTERIIDA	PTERIIDAE	*PINCTADA* Röding, 1798	+		+	+				
PTERIOMORPHIA	PTERIIDA	PINNIDAE	*PINNA* Linnaeus, 1758	+	+	+	+				+
PTERIOMORPHIA	OSTREOIDA	PECTINIDAE	*PLACOPECTEN* Verrill, 1897	+	+	+	+				
PALAEOHETERODONTA	UNIONIDA	UNIONIDAE	*POPENAIAS* Frierson, 1927							+	
PTERIOMORPHIA	PTERIIDA	PTERIIDAE	*PTERIA* Scopoli, 1777								+
HETERODONTA	CHAMIDA	QUENSTEDTIIDAE	*QUENSTEDTIA* Morris & Lycett, 1854							+	
HETERODONTA	CHAMIDA	RZEHAKIIDAE	*RZEHAKIA* Korobkov, 1954							+	
HETERODONTA	CHAMIDA	SCROBICULARIIDAE	*SCROBICULARIA* Schumacher, 1815							+	
PTERIOMORPHIA	PTERIIDA	MYALINIDAE	*SEPTIMYALINA* Newell, 1942							+	
HETERODONTA	CHAMIDA	PSAMMOBIIDAE	*SINONOVACULA* Prashad, 1924			+	+				
PROTOBRANCHIA	SOLEMYIDA	SOLEMYIDAE	*SOLEMYA* Lamarck, 1818		+	+	+	+			
HETERODONTA	CHAMIDA	SOLENIDAE	*SOLEN* Linnaeus, 1758								+
HETERODONTA	CHAMIDA	MACTRIDAE	*SPISULA* Gray, 1837	+	+	+	+				
PTERIOMORPHIA	OSTREOIDA	SPONDYLIDAE	*SPONDYLUS* Linnaeus, 1758	+	+	+	+				
HETERODONTA	VENEROIDA	BERNARDINIDAE	*STOHLERIA* Coen, 1984							+	
HETERODONTA	CHAMIDA	TELLINIDAE	*TELLINA* Linnaeus, 1758								+
HETERODONTA	MYIDA	TEREDINIDAE	*TEREDO* Linnaeus, 1758								+
ANOMALODESMATA	PHOLADOMYOIDA	THRACIIDAE	*THRACIA* Leach in de Blainville, 1824		+	+	+	+			
HETERODONTA	CHAMIDA	VENERIDAE	*TIVELA* Link, 1807						+		
HETERODONTA	CHAMIDA	TRIDACNIDAE	*TRIDACNA* Bruguière, 1797		+	+	+				
PALAEOHETERODONTA	TRIGONIOIDA	TRIGONIIDAE	*TRIGONIA* Bruguière, 1798					+	+		
PROTOBRANCHIA	NUCULOIDA	PRAENUCULIDAE	*TRIGONOCONCHA* Sanchez, 1999						+		
PALAEOHETERODONTA	UNIONIDA	UNIONIDAE	*UNIO* Philipsson, 1788					+			
HETERODONTA	CHAMIDA	LYMNOCARDIIDAE	*UNIOCARDIUM* Capellini, 1880							+	
HETERODONTA	MYIDA	TEREDINIDAE	*UPEROTUS* Guettard, 1770							+	
HETERODONTA	CHAMIDA	VENERIDAE	*VENERUPIS* Lamarck, 1818	+	+	+	+				+
HETERODONTA	CHAMIDA	VENERIDAE	*VENUS* Linnaeus, 1758					+			+
ANOMALODESMATA	PHOLADOMYOIDA	VERTICORDIIDAE	*VERTICORDIA* Gray, 1840						+		
HETERODONTA	MYIDA	XYLOPHAGIDAE	*XYLOPHAGA* Turton, 1822							+	
PROTOBRANCHIA	NUCULOIDA	YOLDIIDAE	*YOLDIA* Möller, 1842					+			
PROTOBRANCHIA	NUCULOIDA	YOLDIIDAE	*YOLDIELLA* Verrill & Bush, 1897						+		
			Totals	31	33	42	41	20	21	20	18

Appendix 4.1. GenBank accession number of sequences used for this study. Where sequences from different congeneric species were lumped together to represent the same genus, the word "*sp.*" was written instead of specific epithets. The only exception is *Anomia sp.*: in this case, all the sequences do come from the same individual of undetermined specific designation. Bold sequences were obtained for this study.

Species	*12s*	*16s*	*cox1*	*cytb*
Abra longicallus		**JF496754**	**JF496762**	**JF496778**
Acanthocardia tuberculata	DQ632743	DQ632743	DQ632743	DQ632743
Acesta bullisi	AM494888	AM494894	AM494905	AM494916
Acesta excavata	AM494882	AM494898	AM494911	AM494920
Acesta oophaga		AM494896	AM494902	AM494918
Adamussium colbecki	EU379383	GU227001		**JF496779**
Alathyria jacksoni	AY387039	AY387021	AY386981	
Amusium pleuronectes	EU379415	DQ640830	GU120012	
Anadara diluvii	**JF496737**		**JF496763**	**JF496780**
Anadara ovalis	GQ166533		GQ166571	GQ166592
Anadara transversa	**GQ166534**		**GQ166572**	**GQ166593**
Anodonta anatina		EF571332	EU252510	GU320047
Anodonta cygnea	**JF496738**	AF232799	**JF496764**	**JF496781**
Anomia sp.		GQ166557	GQ166573	GQ166595
Argopecten irradians	GQ166535	GQ166558	GQ166574	GQ166596
Asperarca sp.	**JF496739**		**JF496765**	**JF496782**
Astarte cfr. *castanea*	GQ166536		AF120662	GQ166597
Barbatia barbata	**JF496740**		AF120645	**GQ166598**
Barbatia cfr. *setigera*	GQ166539		GQ166577	GQ166601
Barbatia parva	GQ166537		GQ166575	GQ166599
Barbatia reeveana	GQ166538		GQ166576	GQ166600
Calyptogena sp.		AF035728	AF008276	AF205081
Cardita variegata	GQ166540		GQ166578	GQ166605
Cerastoderma edule	EF520704	AF122971	AY226940	
Chlamys bruei	**JF496741**	**JF496755**	**JF496766**	
Chlamys farreri	EF473269	EF473269	EF473269	EF473269
Chlamys islandica	FJ263637	FJ263646	AB033665	EU127908
Chlamys livida	GQ166541	GQ166559	GQ166579	GQ166606
Chlamys multistriata	AJ571604	**GQ166560**	**JF496767**	**GQ166607**
Clausinella sp.		DQ459267	**JF496768**	**JF496783**
Corbicula fluminea	EF446612	AF152024	U47647	
Corculum cardissa		EU733079	FJ745336	FJ745359
Crassostrea angulata	FJ841965	FJ841965	FJ841965	FJ841965
Crassostrea ariakensis	FJ841964	FJ841964	FJ841964	FJ841964
Crassostrea gigas	EU672831	EU672831	EU672831	EU672831
Crassostrea hongkongensis	FJ841963	FJ841963	FJ841963	FJ841963
Crassostrea iredalei	FJ841967	FJ841967	FJ841967	FJ841967

Crassostrea sikamea	FJ841966	FJ841966	FJ841966	FJ841966
Crassostrea virginica	AY905542	AY905542	AY905542	AY905542
Cristaria plicata	FJ986302	FJ986302	FJ986302	FJ986302
Cuspidaria rostrata	GQ166542		GQ166580	GQ166608
Donax sp.		EF417547	AB040845	EF417548
Dosinia sp.		DQ356384	GQ855281	**GQ166609**
Dreissena bugensis		AF038996	AF096765	DQ072134
Dreissena stankovici		AY302248	DQ840108	DQ072127
Ensis directus	GQ166543	GQ166561	GQ166581	GQ166610
Epioblasma torulosa rangiana		DQ208539	DQ220724	DQ479938
Gafrarium alfredense	GQ166544	GQ166562		GQ166611
Gemma gemma		GQ166563	GQ166582	GQ166612
Graptacme eborea	AY484748	AY484748	AY484748	AY484748
Haliotis rubra	AY588938	AY588938	AY588938	AY588938
Hiatella arctica	DQ632742	DQ632742	DQ632742	DQ632742
Hyotissa hyotis	GQ166545	GQ166564	GQ166583	GQ166613
Hyriopsis cumingii	FJ529186	FJ529186	FJ529186	FJ529186
Hyriopsis schlegelii	AB250262	DQ073816	GQ360033	
Inversidens japanensis	AB055625	AB055625	AB055625	AB055625
Isognomon sp.	**GQ166546**	HQ329408	AB076926	
Katharina tunicata	U09810	U09810	U09810	U09810
Laevicardium crassum		**JF496756**	**JF496769**	**JF496784**
Lampsilis ornata	AY365193	AY365193	AY365193	AY365193
Lanceolaria grayana		GQ451847	GQ451861	GQ451874
Lima pacifica galapagensis	GQ166548	GQ166565	GQ166584	GQ166616
Lima sp.	AM494893		AM494912	**GQ166615**
Limaria sp.	EU379394	EU379448	AB076953	
Lithophaga lithophaga	**JF496742**	**JF496757**	AF120644	
Loripes lacteus	EF043341	EF043341	EF043341	EF043341
Lucinella divaricata	EF043342	EF043342	EF043342	EF043342
Lunulicardia hemicardia		EU733099	FJ745352	FJ745361
Mactra corallina	GQ166550	GQ166566	GQ166585	GQ166617
Mactra lignaria	GQ166551	GQ166567	GQ166586	
Mercenaria sp.		DQ280040	DQ184836	AF205080
Meretrix lusoria	GQ903339	GQ903339	GQ903339	GQ903339
Meretrix meretrix	GQ463598	GQ463598	GQ463598	GQ463598
Meretrix petechialis	EU145977	EU145977	EU145977	EU145977
Mimachlamys nobilis	FJ415225	FJ415225	FJ415225	FJ415225
Mizuhopecten yessoensis	FJ595959	FJ595959	FJ595959	FJ595959
Modiolus sp.	**JF496743**		FJ890501	**JF496785**
Musculista senhousia	GU001953	GU001953	GU001953	GU001953
Mya arenaria		DQ356387	AF120668	GQ166619

Mytilaster lineatus	**JF496744**		**JF496770**	**GQ166621**
Mytilaster sp.	**JF496745**	DQ836017	**JF496771**	
Mytilus edulis	AY484747	AY484747	AY484747	AY484747
Mytilus galloprovincialis	FJ890849	FJ890849	FJ890849	FJ890849
Mytilus trossulus	HM462080	HM462080	HM462080	HM462080
Neopycnodonte cochlear	**JF496746**	**JF496758**	**JF496772**	
Nucula decipiens	**JF496747**	**JF496759**	**JF496773**	
Nucula nucleus	GQ166552	GQ166568	EF211991	EF211991
Nucula sp.	**JF496748**	AY377617	AF120641	
Nuculana commutata	GQ166553		GQ166587	GQ166622
Ostrea edulis	HQ259072	AF052068	AF120651	
Pandora pinna	GQ166554	GQ166569	GQ166588	GQ166623
Paphia euglypta	GU269271	GU269271	GU269271	GU269271
Parvamussium sp.	EU379411	EU379465	AB084106	
Patinopecten caurinus	FJ263633	FJ263642	AY704170	
Pecten jacobaeus	AJ571596	FN667670	AY377728	GQ166624
Peplum clavatum	**JF496749**	**JF496760**	**JF496774**	
Pinctada albina	AB250260	AB214438	AB261165	
Pinctada fucata	AB250258	AB214444	GQ355871	
Pinctada maculata	AB250261	AB214440	AB261166	
Pinctada maxima	AB250255	AB214435	GQ355881	
Pinna muricata	GQ166555	GQ166570	GQ166589	GQ166625
Pitar sp.		AJ294951	**JF496775**	AF205082
Placopecten magellanicus	DQ088274	DQ088274	DQ088274	DQ088274
Pleurobema collina		AY655061	AY613830	EU414269
Pseudanodonta complanata		DQ060166	EU734829	GU320052
Pteria hirundo	**JF496750**	DQ280031	AF120647	
Pyganodon grandis	FJ809754	FJ809754	FJ809754	FJ809754
Quadrula quadrula	FJ809750	FJ809750	FJ809750	FJ809750
Saccostrea mordax	FJ841968	FJ841968	FJ841968	FJ841968
Sinonovacula constricta	EU880278	EU880278	EU880278	EU880278
Siphonodentalium lobatum	AY342055	AY342055	AY342055	AY342055
Solemya sp.		DQ280028	GQ280818	AM293670
Spisula sp.		AJ548774	AY707797	AF205083
Spondylus gaederopus	AJ571607	AJ571621	**JF496776**	
Striarca lactea	**JF496751**	**JF496761**	AF120646	
Thais clavigera	DQ159954	DQ159954	DQ159954	DQ159954
Thracia distorta	GQ166556		GQ166590	GQ166626
Timoclea ovata	**JF496752**	DQ459292	**JF496777**	**JF496786**
Tridacna derasa		AF122976	GQ166591	GQ166627
Tridacna maxima	EU341598	DQ115320	DQ155301	
Tridacna squamosa		AF122978	EU003615	GQ166628

Unio crassus		DQ060162	EU548052	GU320055
Unio pictorum	HM014134	HM014134	HM014134	HM014134
Unio tumidus		DQ060161	EU548053	GU320060
Venerupis philippinarum	AB065375	AB065375	AB065375	AB065375
Venus casina	**JF496753**	DQ459294	DQ458496	
Venustaconcha ellipsiformis	FJ809753	FJ809753	FJ809753	FJ809753

Appendix 4.2. Molecular evolution models selected by ModelTest 3.7.

Partition	Model
12s	TrN+I+G
16s	GTR+I+G
all	GTR+I+G
cox1	GTR+I+G
cox1_1	TrN+I+G
cox1_12	GTR+I+G
cox1_2	TVM+G
cox1_3	TrN+G
cytb	GTR+I+G
cytb_1	GTR+I+G
cytb_12	GTR+I+G
cytb_2	TVM+G
cytb_3	TrN+G
prot	GTR+I+G
prot_1	TrN+I+G
prot_12	GTR+I+G
prot_2	TVM+I+G
prot_3	TrN+G
rib	TIM+I+G

Appendix 5.1. Sequences used for this study. Whenever "*sp.*" is used instead of a specific epithet, this means that sequences from different congeneric species were joined to represent that genus. Bold sequences were obtained in our laboratory and published in separate papers (Plazzi and Passamonti, 2010; Plazzi et al., in preparation).

Phylum	Class	Species	*16s*	*cox1*	*h3*
Annelida	Clitellata	*Lumbricus terrestris*	U24570	U24570	FJ214260
	Polychaeta	*Platynereis dumerilii*	AF178678	AF178678	X53330
Brachiopoda	Rhynchonellata	*Terebratulina retusa*	AJ245743	AJ245743	DQ779768
Echiura		*Urechis caupo*	AY619711	AY619711	X58895
Mollusca	Aplacophora	*Chaetoderma nitidulum*	EF211990	EF211990	AY377763
		Epimenia australis	AY377614	AY377722	AY377767
	Bivalvia	*Abra sp.*	**JF496754**	**JF496762**	DQ280005
		Amusium pleuronectes	DQ640830	GU120012	EU379523
		Anodonta sp.	AF232799	**JF496764**	AY579132
		Argopecten irradians	**GQ166558**	**GQ166574**	EU379486
		Chlamys farreri	EF473269	EF473269	DQ407914
		Chlamys islandica	FJ263646	AB033665	FJ263666
		Clausinella fasciata	DQ459267	DQ458476	DQ458508
		Corbicula fluminea	AF152024	U47647	AY070161
		Crassostrea gigas	EU672831	EU672831	HQ009488
		Crassostrea virginica	AY905542	AY905542	HQ329250
		Cumberlandia monodonta	U72546	AF156498	AY579144
		Dosinia victoriae	DQ459271	DQ458479	DQ184854
		Dreissena sp.	AF038996	AF096765	AY070165
		Ensis sp.	**GQ166561**	**GQ166581**	AY070159
		Gafrarium sp.	**GQ166562**	EU117999	DQ184892
		Gemma gemma	**GQ166563**	**GQ166582**	DQ184894
		Hiatella arctica	DQ632742	DQ632742	AY070166
		Hyotissa hyotis	**GQ166564**	**GQ166583**	HQ329258
		Isognomon sp.	HQ329408	AB076926	HQ329266
		Lima sp.	**GQ166565**	**GQ166584**	AY070152
		Limaria sp.	EU379448	AB076953	EU379502
		Mercenaria mercenaria	DQ280040	DQ184836	DQ184887
		Meretrix lusoria	GQ903339	GQ903339	FJ429107
		Meretrix meretrix	GQ463598	GQ463598	FJ429106
		Mimachlamys nobilis	FJ415225	FJ415225	DQ407916
		Mizuhopecten yessoensis	FJ595959	FJ595959	DQ407915
		Mya arenaria	DQ356387	AF120668	AY070164
		Mytilus edulis	AY484747	AY484747	AY267749
		Mytilus galloprovincialis	FJ890849	FJ890849	AY267739
		Mytilus trossulus	HM462080	HM462080	AY267747
		Neotrigonia margaritacea	DQ280034	FJ977769	AY070155

		Nucula sp.	**JF496759**	**JF496773**	AY070147
		Nuculana minuta	DQ280030	AF120643	DQ280002
		Ostrea edulis	AF052068	AF120651	AY070151
		Paphia euglypta	GU269271	GU269271	DQ184877
		Parvamussium sp.	EU379465	AB084106	EU379519
		Patinopecten caurinus	FJ263642	AY704170	FJ263662
		Pecten jacobaeus	FN667670	AY377728	AY070153
		Pinctada albina	AB214438	AB261165	HQ329297
		Pinctada fucata	AB214444	GQ355871	HQ329300
		Pinna sp.	**GQ166570**	**GQ166589**	HQ329302
		Pitar sp.	AJ294951	**JF496775**	DQ184863
		Placopecten magellanicus	DQ088274	DQ088274	EU379506
		Pteria hirundo	DQ280031	AF120647	HQ329310
		Solemya velum	DQ280028	GQ280818	AY070146
		Solen sp.	FJ662766	FJ662781	FJ595837
		Spisula sp.	AJ548774	AY707797	M17876
		Spondylus sp.	AJ571621	**JF496776**	EU379533
		Timoclea ovata	DQ459292	**JF496777**	DQ458534
		Venerupis philippinarum	AB065375	AB065375	DQ067446
		Venus casina	DQ459294	DQ458496	DQ458537
	Cephalopoda	*Architeuthis dux*	FJ429092	FJ429092	AY557426
		Dosidicus gigas	EU068697	EU068697	EU735436
		Loligo pealei	AF110079	AF120629	AY377782
		Sepia officinalis	AB240155	AB240155	AY557415
		Sthenoteuthis oualaniensis	EU658923	EU658923	EU735433
	Gastropoda	*Aplysia californica*	AY569552	AY569552	EF457897
		Diodora graeca	DQ093476	AY923915	DQ093502
		Haliotis tuberculata	FJ599667	FJ599667	AY377775
		Littorina littorea	DQ093481	DQ093525	DQ093507
		Lottia gigantea	AB106498	AB238466	FJ977725
	Monoplacophora	*Laevipilina hyalina*	FJ445782	FJ445781	FJ445778
	Polyplacophora	*Katharina tunicata*	U09810	U09810	AY377754
	Scaphopoda	*Antalis entalis*	DQ280027	DQ280016	DQ280000
		Dentalium inaequicostatum	DQ280026	DQ280015	DQ279999
		Rhabdus rectius	AY377619	AY260826	AY377772
Nemertea	Enopla	*Paranemertes peregrina*	GU564481	GU564481	AJ436963
Platyhelminthes	Turbellaria	*Symsagittifera roscoffensis*	HM237350	HM237350	FJ555290
Sipuncula	Sipunculidea	*Sipunculus nudus*	FJ422961	FJ422961	DQ300091

Appendix 5.2. Molecular evolution models selected by ModelTest 3.7.

Partition	Model
16s	GTR+I+G
cox1_1	TIM+I+G
cox1_12	GTR+I+G
cox1_3	TIM+G
h3_1	GTR+G
h3_2	JC
h3_3	TVM+G
prot_1	GTR+I+G
prot_2	TVM+I+G
prot_3	GTR+I+G

CHAPTER 8

PUBLISHED PAPERS

Journal of the Marine Biological Association of the United Kingdom, page 1 of 12.
doi:10.1017/S0025315409991032

The bivalve mollusc *Mactra corallina*: genetic evidence of existing sibling species

I. GUARNIERO[1†], F. PLAZZI[2†], A. BONFITTO[2], A. RINALDI[3], M. TRENTINI[1] AND M. PASSAMONTI[2]

[1]Department of Veterinary Public Health and Animal Pathology, Faculty of Veterinary Medicine, University of Bologna, Via Tolara di Sopra 50, 40064 Ozzano Emilia (BO), Italy, [2]Department of Evolutionary and Experimental Biology, Faculty of Mathematical, Physical and Natural Sciences, University of Bologna, Via Selmi 3, 40126 Bologna (BO), Italy, [3]Oceanographic Structure Daphne, ARPA Emilia Romagna, Viale Vespucci, 2–47042 Cesenatico (FC), Italy; [†]these two authors equally contributed to this work

The rayed trough-shell Mactra corallina *Linnaeus 1758 is a surf clam that inhabits the Atlantic Ocean, Black Sea and Mediterranean Sea and represents a commercially important bivalve. This species is present with two different and well-defined sympatric morphotypes, which differ mainly for the colour of the shell (white in the* corallina *morph, and brown-banded in the* lignaria *morph). The aim of this work is to resolve the confused and contradictory systematics of the bivalves belonging to* M. corallina *putative species by analysing molecular and morphological features. Fifteen specimens of* M. corallina corallina *(white variant) and 19 specimens of* M. corallina lignaria *(brown variant) were collected in the North Adriatic Sea and analysed by four molecular markers (12S, 16S, 18S and COI genes, partial sequences). Genetic analyses clearly support the presence of two different species, which were previously ascribed to* M. corallina. *In addition, 35 specimens identified on a morphological basis as* M. c. corallina *and 28 specimens identified as* M. c. lignaria *collected in the same area were used for a morphometric analysis. A positive correlation was found between the maximum width of shell (W), antero-posterior length and between W and the height of specimens from umbo to ventral margin, thus adding to molecular data.*

Keywords: genetic diversity, molecular taxonomy, bivalves, *Mactra*

Submitted 8 May 2009; accepted 2 August 2009

INTRODUCTION

Surf clams (also known as duck clams or trough shells), belonging to the genus *Mactra* Linnaeus 1767, live in the surf zone of exposed beaches and are widely distributed along mud–sandy coasts of the Pacific Ocean, Atlantic Ocean, Black Sea and Mediterranean Sea (Conroy *et al.*, 1993). They represent commercially important bivalves in many countries and are extensively utilized as seafood, raw materials for manufacturing flavouring materials and live feed at various aquaculture farms (Hou *et al.*, 2006).

The rayed trough-shell *Mactra corallina* (=*M. stultorum*) Linnaeus 1758 inhabits sandy bottoms at depths between 5 and 30 m, and it is distributed along coasts of the Black Sea, Mediterranean Sea and the eastern Atlantic Ocean from Norway to Senegal. It is a medium sized marine bivalve with a very thin and delicate shell with concentric growth lines. This species is present with two different and well-defined morphotypes, which, although they live sympatrically, are generally classified as two different sub-species. These morphotypes are easily distinguishable by the colour of the shell: the white variant, named *M. corallina corallina* Linnaeus 1758, has a shell of a hyaline white with weak ivory radial bands, whereas *M. corallina lignaria* Monterosato 1878 shows brownish radiating bands (D'Angelo & Gargiulo, 1987; Fischer *et al.*, 1987).

The correct specific name for the rayed trough-shell *M. corallina* is a longstanding issue for zoologists and malacologists. As reported in the Mediterranean marine molluscs checklist (Chiarelli, 1999), three species belonging to the genus *Mactra* are present: *M. stultorum* (=*M. corallina*) Linnaeus 1758, *M. glauca* Von Born 1778 and *M. olorina* Philippi 1846. Within *M. corallina*, two taxa, *M. c. corallina* and *M. c. lignaria*, are recognized.

Nevertheless, based on analyses of partial region of 18S rDNA by PCR-SSCP, Livi *et al.* (2006) found preliminary genetic evidences that the traditional classification of *M. c. corallina* and *M. c. lignaria* as subspecies was in contrast with the high genetic distance observed between the two taxa. Besides, *M. c. corallina* formed a highly supported cluster with a further unknown genetic profile, giving evidence of a third taxon belonging to the *M. corallina* complex (Livi *et al.*, 2006).

In his handbook *Carta d'Identità delle Conchiglie del Mediterraneo* Parenzan (1976) describes five distinct phenotypes ascribable to the genus *Mactra*. But actually the most plausible hypothesis is that *M. corallina* is a complex formed by two or more species (Livi *et al.*, 2006).

The official Italian checklists of marine fauna (compiled in their latest version in 2006 and available at http://www.sibm.it/CHECKLIST/principalechecklistfauna.htm) refer to these clams as belonging to the single species *M. stultorum* whereas the FAO identification handbook of Mediterranean species

Corresponding author:
I. Guarniero
Email: ilaria.guarniero@unibo.it

(Fischer *et al.*, 1987) and Riedl (1991) indicate *M. corallina* as the valid name for this species and *M. stultorum* as a synonym.

We decided to adopt the FAO specific designation and thus we refer to the white variant as *M. c. corallina* and to the brown habitus as *M. c. lignaria* as described in D'Angelo & Gargiulo (1987).

This work represents a first attempt to resolve the confused and contradictory systematics of bivalves belonging to *M. corallina* putative species by analysing molecular and morphological characters of the two morphotypes observed. Analysed samples were collected along the north Adriatic coasts of Cesenatico (Italy). In the present study we analysed molecular data obtained by four DNA markers: a nuclear ribosomal DNA subunit (18S) and the mitochondrial genes cytochrome oxidase I (COI), small (12S) and large (16S) ribosomal subunits, in order to provide a stable and robust phylogenetic estimate of the target. In addition, a morphological analysis was carried out on the basis of five parameters of the shell.

MATERIALS AND METHODS

Sampling and DNA extraction

Samples were collected in the north Adriatic Sea in front of Cesenatico (Italy) during a single diving in September 2006 and stored at −80°C. To avoid the problem of collecting paralogous mtDNAs, as found in doubly uniparental inheritance (DUI) bivalve species (see Passamonti & Ghiselli, 2009, and references therein, for a review on the issue), foot muscle tissue was dissected from each individual using a sterile cutter and stored in 80% ethanol at 4°C for the following DNA extraction. DUI has not been searched for in *Mactra*, because of the lack of specimens with fully developed gonads, but even if it would be present, foot muscle is expected to mostly carry mtDNA of maternal origin (Garrido-Ramos *et al.*, 1998). Total genomic DNA was prepared from 25 mg of muscle tissue according to the DNeasy Tissue Kit (Quiagen) protocol.

DNA amplification, cloning and sequencing

Sequences from partial 12S, 16S, 18S and COI were obtained. PCR amplifications were carried out in a 50 μl volume, as follows: 5 μl reaction buffer, 150 nmol $MgCl_2$, 10 nmol each dNTP, 25 pmol each primer, 20 ng genomic DNA, 1.25 units of DNA polymerase (Invitrogen, Carlsbad, CA, USA), water up to 50 μl. Thermal cycling consisted of 35 cycles at 94°C for 60 seconds, the specific annealing temperature (48°C for 12S and 16S; 50°C for 18S and COI) for 60 seconds, and 72°C for 60 seconds. An initial denaturation step (94°C for 5 minutes) and a final extension holding (72°C for 7 minutes) were added to the first and last cycle, respectively. Primer pairs were SR-J14197 ÷ SR-N14745 for 12S (Simon *et al.*, 2006), 16SbrH(32) ÷ 16Sar(34) (5′–CGCCTGTTTAACAAAAACAT–3′) for 16S (modified from Palumbi *et al.*, 1996), 18SF ÷ 18SR for 18S (Livi *et al.*, 2006), and LCO1490 ÷ HCO2198 (Folmer *et al.*, 1994) for COI. Amplified DNAs were treated with Wizard® SV Gel and PCR Clean-Up System (Promega). For a single *Mactra corallina lignaria* individual it was necessary to clone the 18S rDNA gene fragment with Ultramax DH5α–Competent Cells (Invitrogen) following the manufacturer's instructions.

Purified amplifications were either cycle sequenced using the ABIPrism BigDye Terminator Cycle Sequencing kit (Applied Biosystems) and run on an ABI310 Genetic Analyser (Applied Biosystems) or sent to Macrogen (Seoul, EE Korea) for sequencing. Polymorphisms were confirmed by sequencing both strands.

Sequence analysis

Haplotypes (GenBank Accession Numbers FJ830395–FJ830446; Appendix 1) were aligned using the MAFFT multiple sequence alignment tool (Katoh *et al.*, 2002) available online at http://align.bmr.kyushu-u.ac.jp/mafft/online/server. Q-INS-i (Katoh & Toh, 2008) and G-INS-i (Katoh *et al.*, 2005) algorithms were chosen for ribosomal- and protein-coding genes, respectively. Sequences of species belonging to different families of heterodont bivalves were downloaded from the NCBI databank and added to alignment as reference data. In order to compare orthologous characters, only female mtDNA sequences from GenBank were used for DUI species. Gaps were coded as presence/absence data following the simple indel coding method of Simmons & Ochoterena (2000) with the software GapCoder (Young & Healy, 2003).

The analysis of molecular variance (AMOVA) framework (Excoffier *et al.*, 1992) implemented in Arlequin v3.11 software (Excoffier *et al.*, 2005) was used to test the overall genetic heterogeneity of surf clam samples. In this statistical method, a hierarchical AMOVA was carried out on the partitioning of molecular variability at arbitrarily chosen levels (i.e. from the individual to the group of samples level). In the present analysis, groups were obtained by pooling bivalve samples in two groups corresponding to the two subspecies *Mactra corallina corallina* and *M. c. lignaria*. Kimura 2-parameters distances (K-2-P; Kimura, 1980) were computed with MEGA4 software (Tamura *et al.*, 2007) with pairwise deletion of gaps/missing data and with a uniform mutation rate. Φ_{ST} and F_{ST} fixation indices (mitochondrial and nuclear genome respectively) as implemented in Arlequin were calculated to assess the genetic divergence. Statistical significance was estimated by comparing the observed distribution with a null distribution generated by 1000 permutations, in which individuals were randomly re-distributed into samples.

A barcoding-like approach was used to analyse genetic distances computed as formerly described. Frequencies of intra- and inter- specific distances were separately plotted in histograms to provide a visual output of genetic differentiation between the two morphs.

Phylogenetic relationships were inferred through Bayesian analyses implemented in MrBayes 3.1.2 (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003). All analyses employed a cold chain and three incrementally heated chains. Starting trees for each chain were randomly chosen and the default values of MrBayes were used for all settings (including prior distributions). Each metropolis coupled Markov Chain Monte Carlo (MCMC) was run for ten million generations, with trees sampled every 100 generations. Burn-in was visually determined for each gene fragment by plotting average standard deviation of split frequencies over generation seeking for apparent convergence. Chains had always converged to a stable average standard deviation of split frequencies values <0.01.

Posterior probabilities (PP) were used to assess clade support. Analyses were performed using the evolutionary

models selected for each gene fragment by the Bayesian information criterion of Modeltest (Posada & Crandall, 1998). Selected models were K81uf + Γ (Kimura, 1981) for 12S and 16S, K80 + Γ (Kimura, 1980) for 18S, and TVM + Γ for COI. They were implemented into MrBayes with the more similar and more complex model available in the program. *Mytilus galloprovincialis* (female) was used as outgroup to root phylogenetic trees. Nodes with PP < 0.95 were collapsed with the exception of 12S gene fragment data (PP < 0.85). Trees were graphically edited by MrEnt v2.0 (Zuccon & Zuccon, 2006).

Morphological analysis

Five morphological variables were measured: (i) shell length (antero-posterior, L); (ii) height of specimens (ventro-dorsal, H); (iii) maximum width of shell (left–right, W); (iv) distance between the points of intersection of the adductor muscles impressions and the pallial line (AP); and (v) distances between the points of intersection of the adductor muscles impressions and the apex of the umbo (UA and UP). Parameters were measured to 0.01 cm with a caliper. On the basis of such measures, the ratios H/L, W/L and W/H were obtained. Plots were graphically edited by R (Ihaka & Gentleman, 1996). Morphological data were statistically treated with Pearson's coefficient (r) to assess correlation between different sizes; ratios were examined by analysis of F test and the Welch two samples t-test to assess mean differences. The F test is a statistic used to test the hypothesis that two parameters have the same variance against the alternative hypothesis that the variances are different. Degrees of freedom were calculated taking into account number of groups (i.e. $gl_1 = 2 - 1 = 1$) and number of specimens (i.e. $gl_2 = [35 - 1] + [28 - 1] = 61$). The critical values of F with $P = 0.975$ were calculated with the function *qf(p, gl$_1$, gl$_2$)* as implemented in R statistical computing software (Ihaka & Gentleman, 1996; R Development Core Team, 2009). Welch's t-test is an adaptation of the Student's t-test intended for use with two samples having possibly unequal variances. Values of t-test were calculated using the function *t.test(x1, x2)* implemented in R software.

RESULTS

Genetic data

Twenty individuals for each morphotype were collected. A total of 34 specimens, 15 ascribed to *Mactra corallina corallina* and 19 to *M. c. lignaria*, were amplified and sequenced for the 12S, 16S, 18S and COI genes (partial sequences). Fragments of 397, 513, 516 and 571 bp respectively were obtained. Variable sites (including maximum parsimony informative sites), haplotype frequencies, specimen numbers and GenBank IDs are given in Appendix 1.

Data obtained by aligning the 12S partial sequence appeared quite soon less powerful than other gene fragments probably because of sampling artefacts. Actually, technical problems occurred during amplification and sequencing of the 12S and only four individuals of each group gave suitable PCR amplicons and electropherograms. Twenty-six repeated null amplifications were observed (11 in *M. c. corallina* and 15 in *M. c. lignaria*), accounting for the presence of point mutations in the annealing site of either primer. Further analyses will be required to unravel this latest issue.

In any case, examining sequence alignments for all the analysed gene fragments, high genetic divergences were observed between specimens of the two different morphs here considered (i.e. var. *corallina* and var. *lignaria*). Diagnostic sites were 7 out of 397 for 12S, 8 out of 513 for 16S, 25 out of 516 for 18S, and 43 out of 571 for COI (Appendix 1).

No mutation was observed at the amino acid level for the COI gene. Most point-mutations occurred at the third position of the codon. Six out of 60, however, were found at 2nd position (343, 358, 370, 412, 475 and 478).

Levels of genetic variability within the same morphotype were remarkably low and some shared haplotypes were observed (Appendix 1). A weak polymorphism was observed in the 18S fragment within both morphotypes, in the proportion of one different haplotype out of eleven in *M. c. lignaria* (sample n. 14 C2; C/T transition in position 467) and one out of six in *M. c. corallina* (sample n. 32; C/A, A/G, C/A transversion/transition in position 198, 200 and 202 respectively). Incidentally, the *M. c. lignaria* observed single 18S variant was found in a cloned sequence (see Appendix 1).

The higher proportion of overall molecular variance was always found at 'between morphotypes' hierarchical level (from 77.78%, $P < 0.05$; to 99.23%, $P < 0.01$; Table 1). All fixation indices values were high and significant or even highly significant. With the only exception of the 12S fragment ($F_{ST} = 0.778$, $P = 0.025$), fixation indices values were higher than 0.90 and ranged from 0.902 (COI) to 0.992 (18S; Table 1).

Figure 1 shows histograms obtained by plotting intra- and inter- specific K-2-P distances for the four analysed gene fragments. Intra- and inter- morphotype distances are well separated and the gap between these distances ranges from about 0.005 (16S) to about 0.064 (COI), respectively.

The Bayesian analysis performed with different combinations of data yielded differently resolved but comparable

Table 1. Analysis of partition of molecular variance (AMOVA) and fixation indices values (F_{ST} for diploid data, Φ_{ST} for haploid data). *, $P = 0.05$; **, $P = 0.01$; ***, $P = 0.001$.

Locus	Source of variation	df	Sum of squares	Variance components	Percentage of variation	Fixation index	*P* value
12S	Among morphotypes	1	7.500	1.75000 Va	77.78	$\Phi_{ST} = 0.778$	*
	Within morphotypes	6	3.000	0.50000 Vb	22.22		
16S	Among morphotypes	1	22.750	3.01339 Va	92.34	$\Phi_{ST} = 0.923$	***
	Within morphotypes	13	3.250	0.25000 Vb	7.66		
18S	Among morphotypes	1	60.797	7.82208 Va	99.23	$F_{ST} = 0.992$	***
	Within morphotypes	15	0.909	0.06061 Vb	0.77		
COI	Among morphotypes	1	108.614	18.27869 Va	90.19	$\Phi_{ST} = 0.902$	**
	Within morphotypes	10	19.886	1.98857 Vb	9.81		

Table 2. Analysis of F test with $P = 0.975$ calculated with the function *qf(p, gl₁, gl₂)* (degrees of freedom: $gl_1 = 1$ and $gl_2 = 61$) and of the Welch two samples *t*-test calculated using the function *t.test(x1, x2)* applied to H/L, W/L and W/H ratios.

Ratio	*Mactra corallina*	*Mactra lignaria*	F test	$P = 0.975$	*t* value	*P* value
H/L	0.82997 ± 0.007	0.82068 ± 0.009	0.0800	5.281162	1.5476	0.183
W/L	0.53866 ± 0.009	0.43579 ± 0.009	7.6597	5.281162	15.6507	<2.2e−16
W/H	0.64924 ± 0.011	0.53122 ± 0.011	7.0448	5.281162	14.9967	<2.2e−16

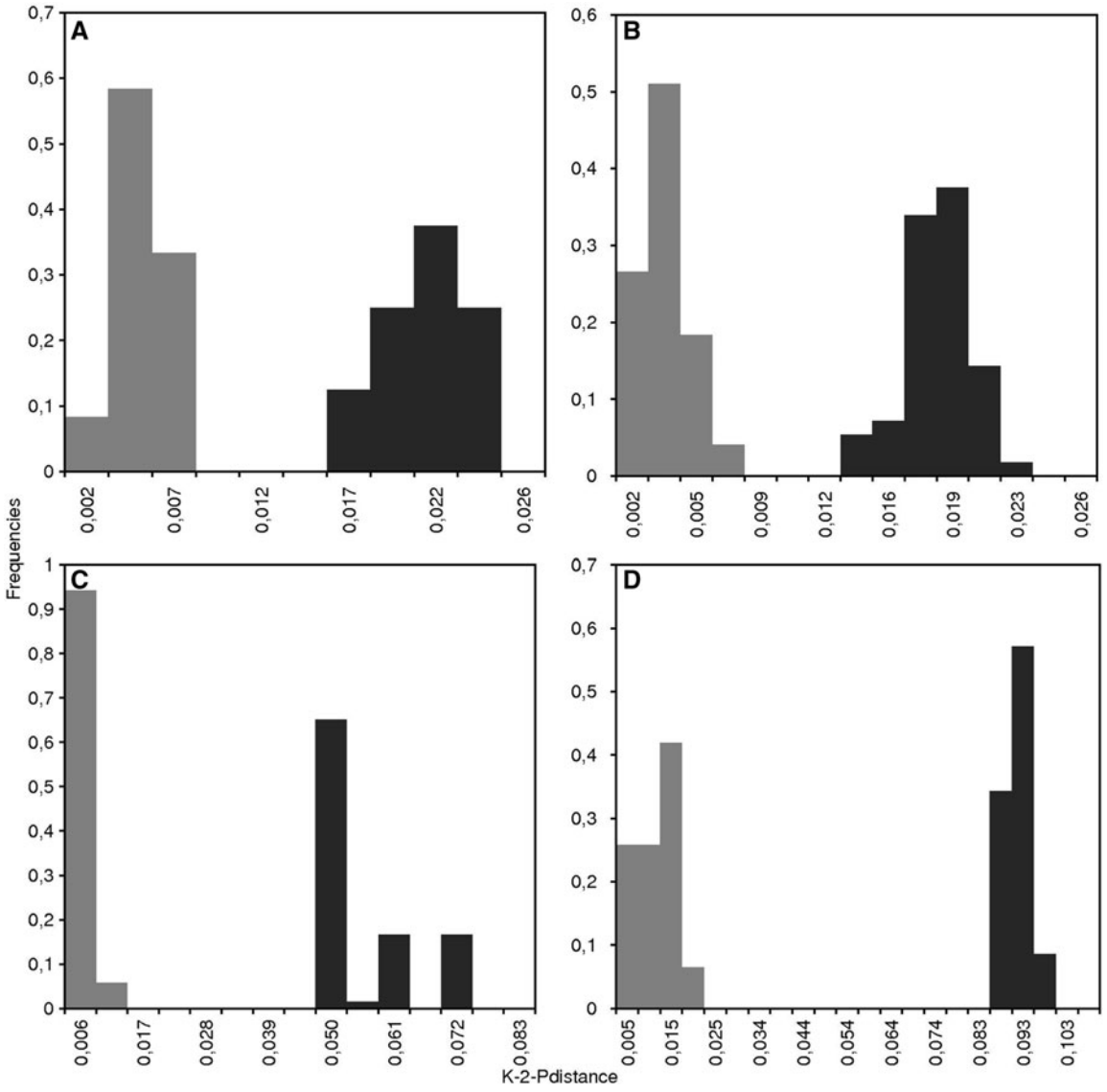


Fig. 1. Histogram illustrating K-2-P distances distribution among *Mactra corallina*/*M. lignaria* group, as resulting from the four characterized genes. K-2-P distance values are reported on x-axis, whereas their frequencies are reported on y-axis. A, 12S; B, 16S; C, 18S; D, COI; light grey: intra-specific distances; dark grey: inter-specific distances.

and well supported tree topologies (Figures 2–5). In all trees, the two morphotypes clustered separately from all other sequence data with $0.95 < PP < 1.00$. *Mactra c. corallina* was resolved as a monophyletic group for 12S ($PP = 0.88$), 18S ($PP = 1.00$) and COI ($PP = 1.00$). Similarly, *M. c. lignaria* was resolved as monophyletic for 16S ($PP = 0.96$), 18S ($PP = 1.00$) and COI ($PP = 1.00$). Both morphotypes were paraphyletic in other cases (i.e. 16S and 12S respectively). At a higher taxonomic level, the superfamily Mactroidea (= Mactracea) Lamarck 1809 (Mactridae Lamarck 1809 + Anatinellidae Gray, 1853 + Cardiliidae Fischer, 1887 + Mesodesmatidae Gray 1840) appear to be monophyletic in all obtained trees, with PP values ranging from 0.97 to 1.00, while the superfamily Veneroidea Rafinesque 1815 showed a complex situation that would require further investigations.

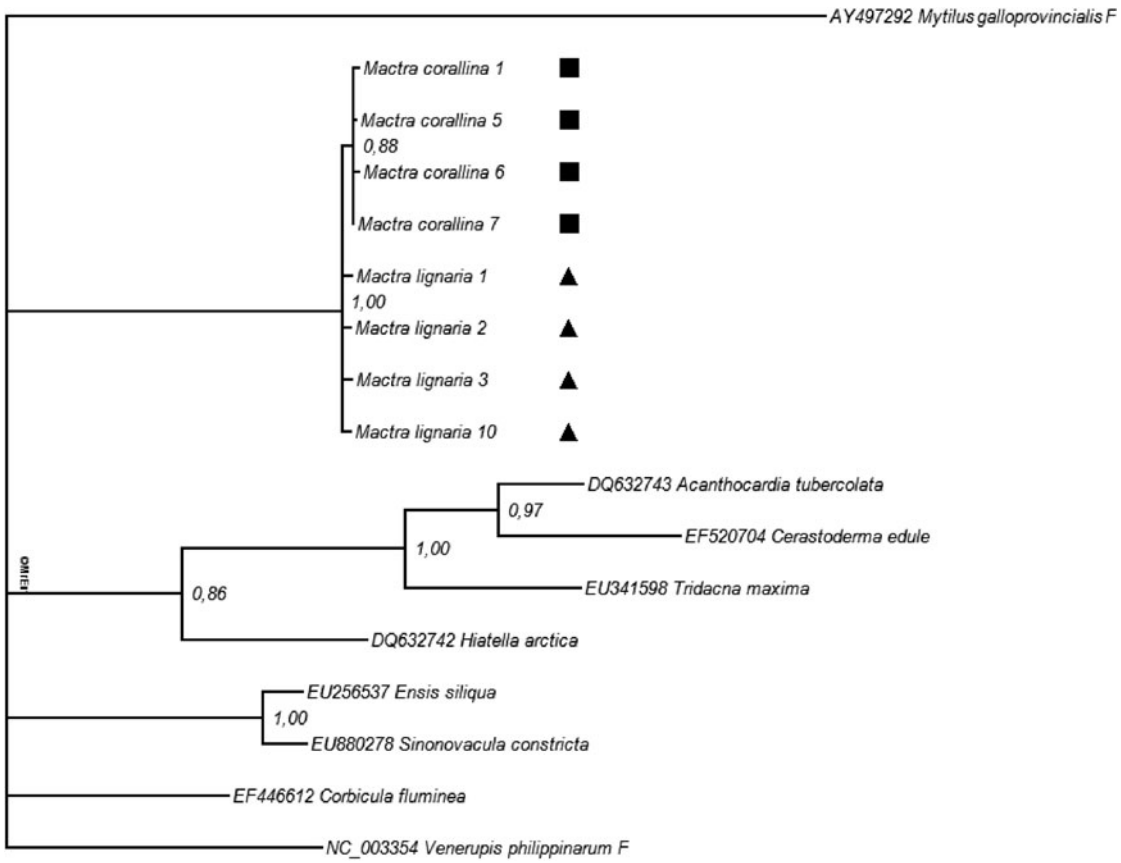


Fig. 2. Bayesian phylogeny of *Mactra corallina/M. lignaria* samples inferred by 12S sequence data. Individuals belonging to the *corallina* morphotype are marked with a square whereas individuals belonging to the *lignaria* morphotype are marked with a triangle. For correspondences to the GenBank accession number, see Appendix 1.

Morphological data

Morphological analyses showed that only three parameters (i.e. L, H and W) were statistically significant, while AP, UA and UP did not present any element of significance on discriminating the two morphotypes (data not shown). As a consequence, the last three parameters were not considered and here we will take into account ratios that only involve the former three parameters.

The analysis of Pearson's correlation reflects the degree to which two variables are related. The correlation between the considered sizes gives the following r values: in *M. c. corallina* $r_{H/L} = 0.915$, $r_{W/L} = 0.741$ and $r_{W/H} = 0.749$; in *M. c. lignaria* $r_{H/L} = 0.941$, $r_{W/L} = 0.781$ and $r_{W/H} = 0.777$.

Both in *M. c. corallina* and *M. c. lignaria*, all morphological features considered were positively correlated. In particular, high values of r were found for correlation between H and L. Morphometric ratios found are given in Figure 6.

The F test applied to W/L and W/H ratios showed statistically significant values, while for H/L the null hypothesis cannot be rejected (Table 2). Similarly, the *t*-test assessed a significant difference in W/H and W/L ratios. No significant difference was found in H/L ratio (Table 2).

DISCUSSION

The development of molecular tools for species identification scored an increased importance because of difficulties of discriminating them on the basis of morphological characters only. This is mostly true for organisms at early developmental stages and in cases of morphological stasis of adults or presence of sibling species (Øines & Heuch, 2005; Livi *et al.*, 2006).

Molecular assays presented in this paper brought to light a stable genetic divergence between *M. c. corallina* and *M. c. lignaria*. The clams analysed in this work were caught during a single dive in the very same area. The sympatric occurrence of the two morphotypes, coupled with the genetic divergence detected, is strong evidence of separate gene pools, thus supporting a reproductive isolation between the two morphs. Therefore, the taxon previously described as *M. corallina* should be rather considered as two different biological species, *M. corallina* and *M. lignaria*. A very similar experimental procedure, although based on allozyme analysis, was reported in Backeljau *et al.* (1994), who identify *Chamelea gallina* and *C. striatula*, previously considered as two subspecies of *C. gallina*, as two distinct and reproductively isolated biological species; actually, despite the probable overlap in breeding season between the two *Chamelea* morphotypes, they maintained a large genetic distance in sympatric conditions, giving evidence of two different biological species (Backeljau *et al.*, 1994).

For our *Mactra*, more genetic data obtained are consistent with two different species: the magnitude of genetic distances observed between *M. c. corallina* and *M. c. lignaria* were comparable to, if not greater than, distances detected among different genera belonging to the family Mactridae (K-2-P distance values, Figures 1 & 4B). The intra-specific pairwise

K-2-P genetic distances were an order of magnitude lower than inter-specific comparisons (Figure 1). This divergence is also clearly shown by the high and statistically supported values of fixation indices, which were close to one and indicated the presence of a sharp dichotomy between genotypes, and the unbalanced partition of molecular variance with the majority of percentage detected at the higher hierarchical level, i.e. 'among morphotypes'. In the phylogenetic trees, albeit in two cases a soft paraphyly was observed (Figures 2 & 3) we observed a separation of *M. c. corallina* clusters from *M. c. lignaria* clusters, supported by robust node values.

Finally, the observed variability in the 18S gene well falls within the range of expected variability for this locus. This gene, generally highly conserved within species, shows variability higher in bivalves than in other taxa (Adamkewicz *et al.*, 1997; Passamaneck *et al.*, 2004). Moreover, the unique different haplotype found in *M. c. lignaria* was collected from a clone, which might have brought to light a rare variant (i.e. intra-individual variability among 18S repeats within the nuclear genome).

Preliminary morphological analyses seem also concordant with genetic data, although only one shell character (other than the colour) was significantly different; in fact, the main morphological character discriminating the two morphs seems to be the W value (maximum width of shell, i.e. the convexity) which differentiates morphometrical ratios in specimens with the same length or height. According to the data, the ratios W/L and W/H assume a clear (and classic) diagnostic value and allows us to take the following value to discriminate the two groups: in *M. c. corallina* $W/L > 0.50$ and $W/H > 0.60$, while in *M. c. lignaria* $W/L < 0.50$ and $W/H < 0.60$.

The effective reproductive isolation between *M. c. corallina* and *M. c. lignaria* (and/or sterility of hybrids) has still to be directly demonstrated, but obtained data are sound enough to support the species level for both morphs. Nevertheless, an additional sampling along the Adriatic coasts has already been planned to better describe the genetic landscape of *Mactra*, which seems to represent a complex of at least two (but probably more) different species (Livi *et al.*, 2006).

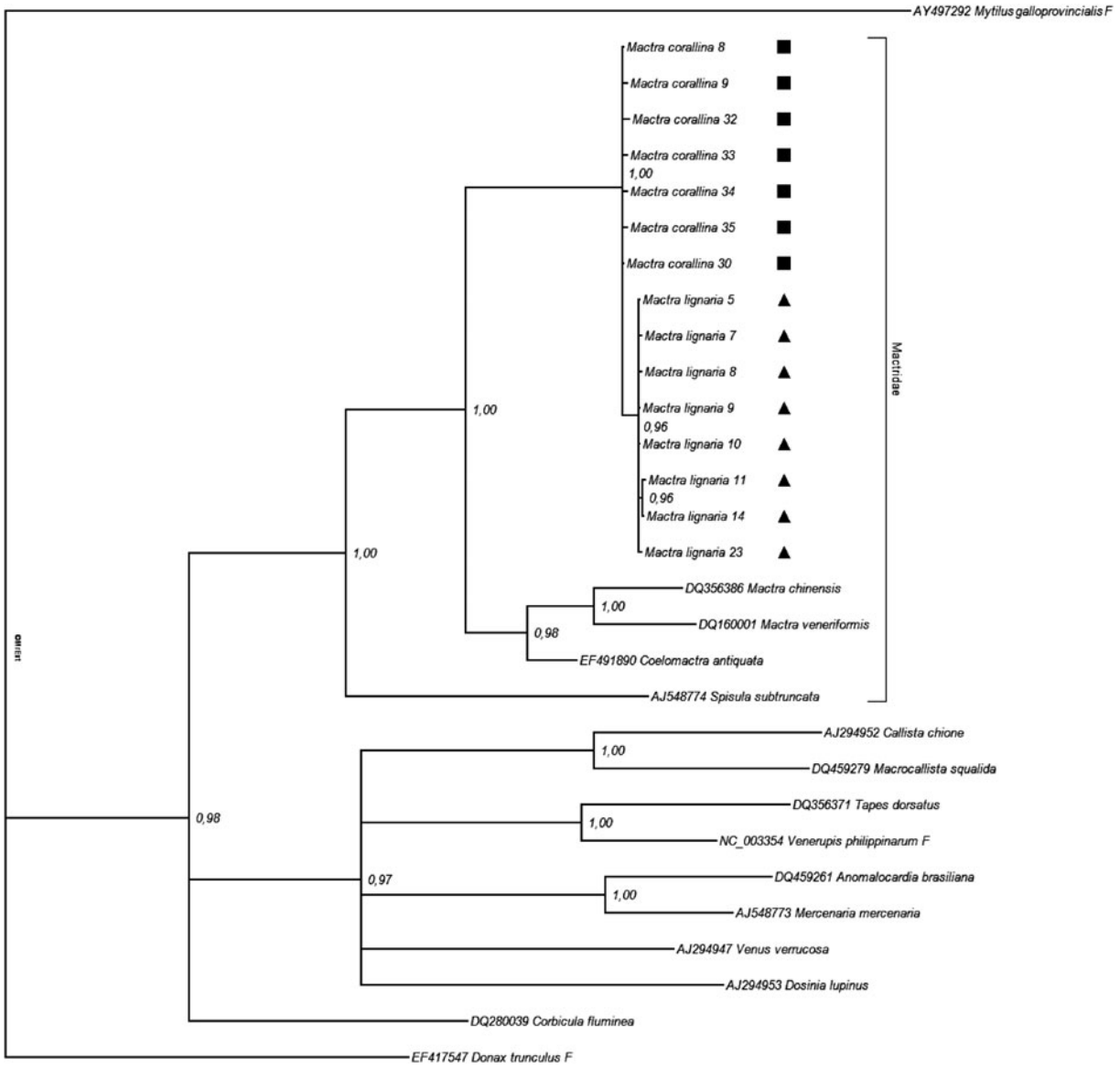


Fig. 3. Bayesian phylogeny of *Mactra corallina/M. lignaria* samples inferred by 16S sequence data. Taxon symbols as in Figure 2.

Fig. 4. (A) Bayesian phylogeny of *Mactra corallina/M. lignaria* samples inferred by 18S sequence data. Taxon symbols as in Figure 2. Grey arrow heads point to Mesodesmatidae species; (B) histogram illustrating intergeneric K-2-P distances distribution among Mactridae: K-2-P distance values are reported on x-axis, whereas their frequencies are reported on y-axis; data from established genera of Mactridae are shown in white, whereas data from inter-specific comparisons among *Mactra corallina/M. lignaria* group are shown in dark grey, as in Figure 1C.

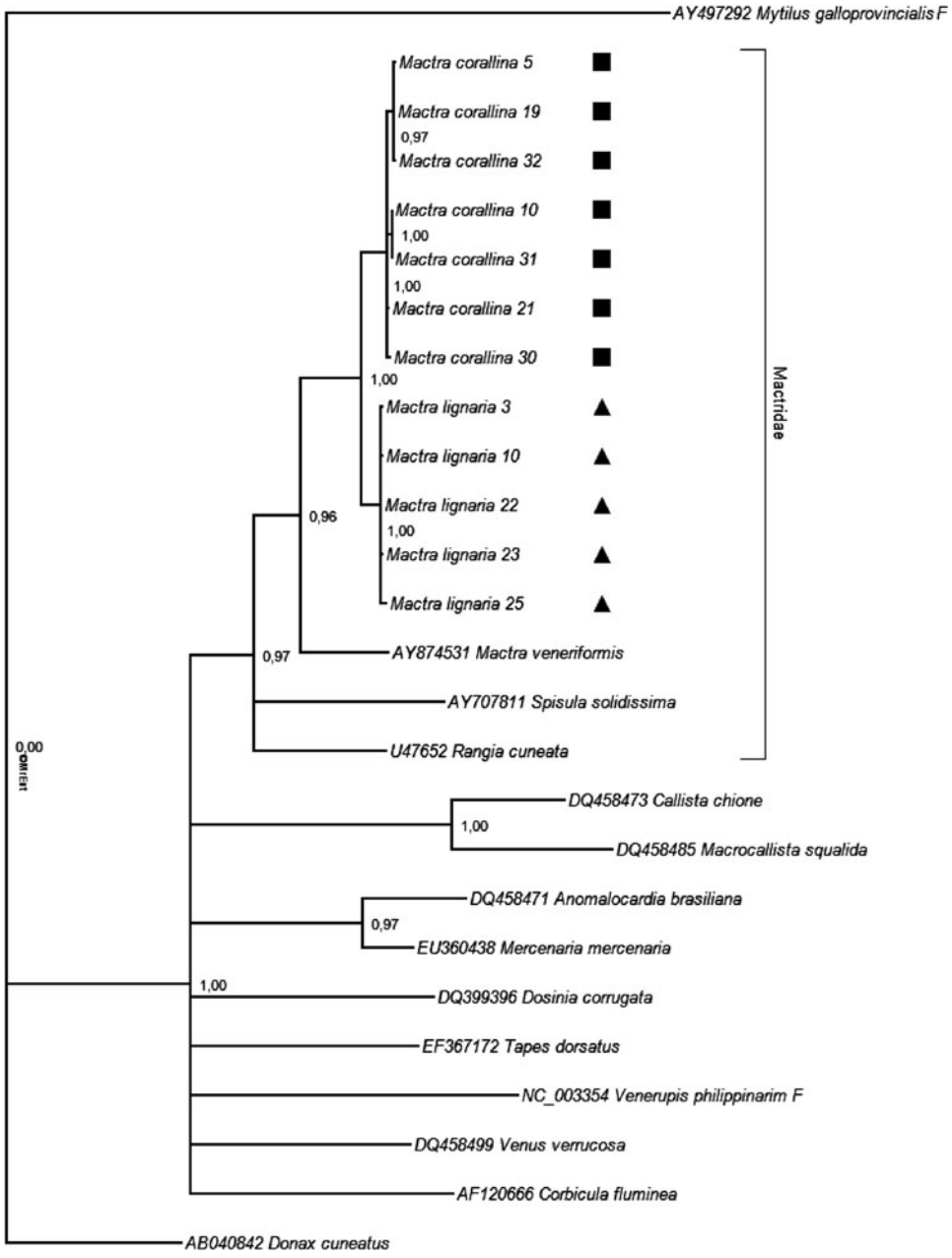


Fig. 5. Bayesian phylogeny of *Mactra corallina/M. lignaria* samples inferred by COI sequence data. Taxon symbols as in Figure 2.

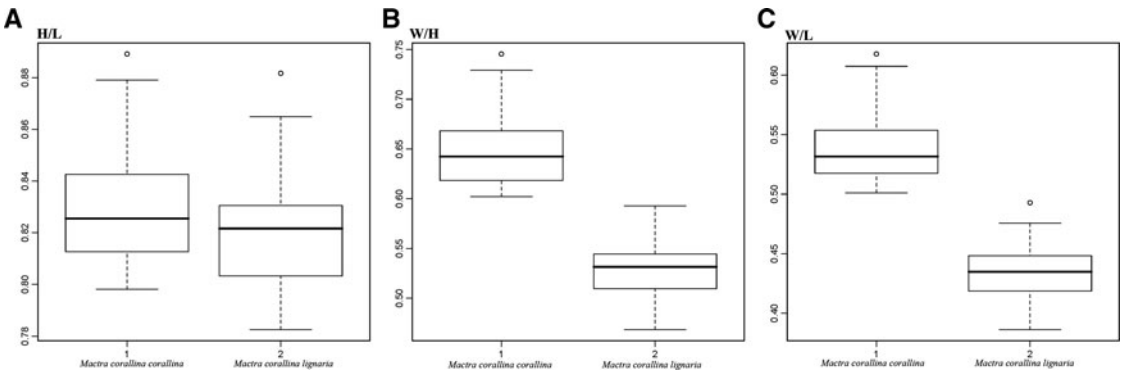


Fig. 6. Morphometric ratios in *Mactra corallina* and *M. lignaria*.

Finally, the phylogenetic position of *Mactra* was addressed in this study. On the basis of 18S and 28S rRNA genes, it was previously found that the superfamily MACTROIDEA, traditionally classified near to the superfamily CARDIOIDEA (=CARDIACEA) Lamarck 1809 with an implicit sister-group relationship, showed grater affinity to UNGULINIDAE H. & A. Adams 1857 and the group of VENERIDAE Rafinesque 1815—CORBICULARIDAE Gray 1847—ARCTIDAE Newton 1891—CHAMIDAE Blainville 1825, but no connection with CARDIOIDEA (Taylor *et al.*, 2007). In our preliminary phylogenetic analysis, the genus *Mactra* was always monophyletic, although the 16S sequence of *Coelomactra antiquata* obtained from GenBank generates a polyphyly in the clade of *Mactra* (polyphyly supported by a significant PP nodal value of 0.98). Moreover, the superfamily MACTROIDEA clustered separately in all trees and was statistically well supported. Finally, in the 18S tree, individuals belonging to families MACTRIDAE and MESODESMATIDAE were intermingled (Figure 4A). This situation suggests further investigation focused on these species to assess the monophyly of the genus *Mactra* and to validate the family status of MESODESMATIDAE.

ACKNOWLEDGEMENTS

Two anonymous referees provided valuable comments on the manuscript. Dr Federica Rongai of the Laboratory of Applied Biotechnology for Aquaculture and Fishery of the Aquaculture Institute of Cesenatico (Italy) and Dr Davide Gambarotto of the Laboratory of Molecular Zoology of the Department of Experimental Evolutionary Biology of Bologna (Italy) supported the laboratory work.

REFERENCES

Adamkewicz S.L., Harasewych M.G., Blake J., Saudek D. and Bult C. (1997) A molecular phylogeny of the bivalve mollusks. *Molecular Biology and Evolution* 14, 619–629.

Backeljau T., Bouchet P., Gofas S. and de Bruyn L. (1994) Genetic variation, systematics and distribution of the venerid clam *Chamelea gallina*. *Journal of the Marine Biological Association of the United Kingdom* 74, 211–223.

Chiarelli S. (1999) *Nuovo Catalogo delle Conchiglie Marine del Mediterraneo*. Società Italiana di Malacologia, http://www.aicon.com/sim/index.html.

Conroy A.M., Smith P.J., Michael K.P. and Stotter D.R. (1993) Identification and recruitment patterns of juvenile surf clams, *Mactra discors* and *M. murchisoni* from central New Zealand. *New Zealand Journal of Marine and Freshwater Research* 27, 279–285.

D'Angelo G. and Gargiulo S. (1987) Mactridae. In Fabbri (ed.) *Guida alle conchiglie mediterranee*. Milan: Gruppo editoriale Fabbri spa, pp. 192–193.

Excoffier L., Smouse P. and Quattro J. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131, 479–491.

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

Fischer W., Schneider M. and Bauchot M.L. (1987) *Fiches FAO d'identification des espèces pour les besoins de la pêche. Mêditerranêe et Mêr Noire. Zone de pêche 37. Végétaux et invertébrés*. Rome: FAO Publication.

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

Garrido-Ramos M.S., Stewart D.T, Sutherland B.W. and Zouros E. (1998) The distribution of male-transmitted and female-transmitted mitochondrial DNA types in somatic tissues of blue mussels: implications for the operation of doubly uniparental inheritance of mitochondrial DNA. *Genome* 41, 818–824.

Hou L., Lü H., Zou X., Bi X., Yan D. and He C. (2006) Genetic characterizations of *Mactra veneriformis* (Bivalve) along the Chinese coast using ISSR-PCR markers. *Aquaculture* 261, 865–871.

Huelsenbeck J.P. and Ronquist F. (2001) MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* 17, 754–755.

Ihaka R. and Gentleman R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 5, 299–314.

Katoh K. and Toh H. (2008) Improved accuracy of multiple ncRNA alignment by incorporating structural information into a MAFFT-based framework. *BMC Bioinformatics* 9, 212.

Katoh K., Kuma K.I., Toh H. and Miyata T. (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research* 33, 511–518.

Katoh K., Misawa K., Kuma K.I. and Miyata T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* 30, 3059–3066.

Kimura M. (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotides sequences. *Journal of Molecular Evolution* 16, 111–120.

Kimura M. (1981) Estimation of evolutionary distances between homologous nucleotide sequences. *Proceedings of the National Academy of Sciences of the USA* 78, 454–458.

Livi S., Cordisco C., Damiani C., Romanelli M. and Crosetti D. (2006) Identification of bivalve species at an early developmental stage through PCR–SSCP and sequence analysis of partial 18S rDNA. *Marine Biology* 149: 1149–1161.

Øines Ø. and Heuch P.A. (2005) Identification of sea louse species of the genus *Caligus* using mtDNA. *Journal of the Marine Biological Association of the United Kingdom* 85, 73–79.

Palumbi S.R., Martin A., Romano S., McMillan W.O., Stice L. and Grabowski G. (1996) *The simple fool's guide to PCR*. Hawaii, USA: Kewalo Marine Laboratory and University of Hawii.

Parenzan P. (1976) *Carta d'identità delle conchiglie del Mediterraneo. Volume II. Bivalvi*. Taranto, Italy: Bios Taras Editrice, 546 pp.

Passamaneck Y.J., Schandler C. and Halanych K.M. (2004). Investigation of mulluscan phylogeny using large-subunit and small-subunit nuclear rRNA sequences. *Molecular Phylogenetic Evolution* 32, 25–38.

Passamonti M. and Ghiselli F. (2009) Doubly uniparental inheritance: two mitochondrial genomes, one precious model for organelle DNA inheritance and evolution. *DNA and Cell Biology* 28, 1–10.

Posada D. and Crandall K.A. (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics* 14, 817–818.

R development Core Team (2009) *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0, URL http://www.R-project.org.

Riedl R. (1991) Mactridae. In Muzzio F. (ed.) *Fauna e flora del Mediterraneo. Dalle alghe ai mammiferi: una guida sistematica alle specie che vivono nel Mar Mediterraneo*. Trento: Legoprint srl, pp. 342–343.

Ronquist F. and Huelsenbeck J.P. (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574.

Simmons M.P. and Ochoterena H. (2000) Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* 49, 369–381.

Simon C., Buckley T.R., Frati F., Stewart J.B. and Beckenbach A.T. (2006) Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. *Annual Review of Ecology and Systematics* 37, 545–579.

Tamura K., Dudley J., Nei M. and Kumar S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* 24, 1596–1599.

Taylor J.D., Williams S.T., Glover E.A. and Dyal P. (2007) A molecular phylogeny of heterodont bivalves (Mollusca: Bivalvia: Heterodonta): new analyses of 18S and 28S rRNA genes. *Zoologica Scripta* 36, 587–606.

Young N.D. and Healy J. (2003) GapCoder automates the use of indel characters in phylogenetic analysis. *BMC Bioinformatics* 4, 6.

and

Zuccon A. and Zuccon D. (2006) MrEnt v2.0. Program distributed by the authors, Department of Vertebrate Zoology & Molecular Systematics Laboratory, Swedish Museum of Natural History, Stockholm.

Correspondence should be addressed to:
I. Guarniero
Department of Veterinary Public Health and Animal Pathology
Faculty of Veterinary Medicine, University of Bologna
Via Tolara di Sopra 50
40064 Ozzano Emilia (BO)
Italy
email: ilaria.guarniero@unibo.it

Appendix 1. Alignment of the two variants of *Mactra corallina* analysed (*lig*: *lignaria, cor*: *corallina*), related frequencies (*f*), specimen numbers as in figures 2 to 5 and GenBank accession number. Only variable sites are reported.

Locus	Variable	Variable sites	*f*	Specimen number	GenBank accession number
12S		1224444566 6			
		9660124305 8			
		9024574506 9			
	lig	TCCATATTGA T	1	1	FJ830395
	lig	C.	2	2,10	FJ830396
	lig	C. C	1	3	FJ830397
	cor	C.TGAGACAG .	1	1	FJ830399
	cor	C.TGAGAC.G .	1	5	FJ830400
	cor	CTTGAGAC.G .	1	6	FJ830401
	cor	C.TGAGA..G .	1	7	FJ830402
16S		4455566 668			
		4895601867 891			
		8004562306 479			
	lig	CCTGGAAGAT TTT	4	5,7,9,10	FJ830403
	lig	. . . . . .C.	1	8	FJ830405
	lig	T. . . . . T.	1	11	FJ830408
	lig	. . . . . T.	1	14	FJ830409
	lig	. . . . . T.	1	23	FJ830410
	cor	.T.AA..AGC G ..	2	8,30	FJ830411
	cor	.TCAA..AGC G ..	4	9,33,34,35	FJ830412
	cor	.TCAA..AGC G.G	1	32	FJ830414
18S		111111 1111222222 233334			
		2223222366 7799000112 714586			
		0694679689 0158027464 311837			
	lig	CAAGACGTGC TTGCACGACA TCGTAC	10	10,11,13,14 C1,16,17,19,21,23,31	FJ830418
	lig	. . . . . .T	1	14 C2	FJ830422
	cor	ATTTCAACAG CCC...ATTG AAACC.	5	5,6,10,30,31	FJ830430
	cor	ATTTCAACAG CCCAGAATTG AAACC.	1	32	FJ830434
COI		111111111 1112222222 2333333333 3444444444 5555555555			
		1223466778 9122334467 7880122458 8224455788 9134456777 0013345667			
		5470506587 0703584724 7097629685 8173518047 9284762158 1791708140			
	lig	GCGGTCTATA GGATCGATAT CTTGTACCAT AGCTAATTTT TCCTCTCATT AGATTCCTCG	2	3,10	FJ830435
	lig	. . . .C. . . . . .C.	1	22	FJ830435
	lig	. . . . . G.	1	23	FJ830438
	lig	....C. . . . . .A. . . . .G. . . . . .T. . . . . .T.	1	25	FJ830439

Continued

Appendix 1. Continued

Locus	Variable	Variable sites	*f*	Specimen number	GenBank accession number
	cor	ATTA.TCGCG A.GC.AG.G. TC.TCG.TGA TA.CGGCCCC CTT.T.TTCC GAGCCT..TA	1	5	FJ830440
	cor	AT.A.TCGCG A.GC.AGCG. T..TCGTTGA TA.C.GCC.C CTT.T.TTCC GAGCCTTCTA	2	10,31	FJ830441
	cor	AT.A.TCGCG A.GC.AG.G. T..TCG.TGA TA.CGGCCCC CTT.T.TTCC GAGCCT..TA	1	19	FJ830442
	cor	AT.A.TCG.G A.GC.AGCG. T.CTCG.TGA TA.C.GCC.C CTTCT.TT.C GAGCCT.CTA	1	21	FJ830443
	cor	AT.A.TCGCG A.GC.AGCGC T..TCG.TGA TA.C.GCC.C CTT.T.TTCC GAGCCT.CTA	1	30	FJ830444
	cor	AT.A.TCGCG A.GCTAG.G. T..TCG.TGA TA..GGCCCC CTT.T.TTCC GAGCCT..TA	1	32	FJ830446

BMC Bioinformatics

METHODOLOGY ARTICLE Open Access

Phylogenetic representativeness: a new method for evaluating taxon sampling in evolutionary studies

Federico Plazzi*[1], Ronald R Ferrucci[2] and Marco Passamonti[1]

Abstract

Background: Taxon sampling is a major concern in phylogenetic studies. Incomplete, biased, or improper taxon sampling can lead to misleading results in reconstructing evolutionary relationships. Several theoretical methods are available to optimize taxon choice in phylogenetic analyses. However, most involve some knowledge about the genetic relationships of the group of interest (i.e., the ingroup), or even a well-established phylogeny itself; these data are not always available in general phylogenetic applications.

Results: We propose a new method to assess taxon sampling developing Clarke and Warwick statistics. This method aims to measure the "phylogenetic representativeness" of a given sample or set of samples and it is based entirely on the pre-existing available taxonomy of the ingroup, which is commonly known to investigators. Moreover, our method also accounts for instability and discordance in taxonomies. A Python-based script suite, called PhyRe, has been developed to implement all analyses we describe in this paper.

Conclusions: We show that this method is sensitive and allows direct discrimination between representative and unrepresentative samples. It is also informative about the addition of taxa to improve taxonomic coverage of the ingroup. Provided that the investigators' expertise is mandatory in this field, phylogenetic representativeness makes up an objective touchstone in planning phylogenetic studies.

Background

The study of phylogenetics has a long tradition in evolutionary biology and countless statistical, mathematical, and bioinformatic approaches have been developed to deal with the increasing amount of available data. The different statistical and computational methods reflect different ways of thinking about the phylogeny itself, but the issue of "how to treat data" has often overshadowed another question, i.e., "where to collect data from?". We are not talking about the various types of phylogenetic information, such as molecular or morphological characters, but rather we refer to which samples should be analyzed.

In phylogenetic studies, investigators generally analyze subsets of species. For example, a few species are chosen to represent a family or another high-level taxon, or a few individuals to represent a low-level taxon, such as a genus or a section. As a general practice, choices are driven by expertise and knowledge about the group; key species and taxa of interest are determined and, possibly, sampled. For example, if a biologist is choosing a group of species to represent a given class, species from many different orders and families will be included. We term the degree to which this occurs the "phylogenetic representativeness" of a given sample.

This issue is rarely formally addressed and generally treated in a rather subjective way; nevertheless, this is one of the most frequent ways incongruent phylogenetic results are accounted for. It is sufficient to browse an evolutionary biology journal to see how often incorrect or biased taxon sampling is hypothesized to be the cause [e.g., [1-6]]. We therefore aim to set up a rigorous taxon sampling method, which can be used alongside expertise-driven choices. Many theoretical approaches have been proposed to drive taxon sampling: see [[7]; and reference therein] for a keystone review.

* Correspondence: federico.plazzi@unibo.it
[1] Department of "Biologia Evoluzionistica Sperimentale", University of Bologna, Via Selmi, 3 - 40126 Bologna, Italy
Full list of author information is available at the end of the article

The concept of "taxonomic distinctness" was developed in the early 1990s among conservation biologists [8,9], who needed to measure biodiversity within a given site or sample so to assess further actions and researches. Basic measures of biodiversity take into account species richness and relative abundance [10-13]. However, it is clear from a conservationist point of view that not all species should be weighted the same. The presence and relative abundance of a species cannot capture all information on the variation of a given sample, and therefore a taxonomic component must also be considered in evaluating the biodiversity of a given site. This allows more realistic specification of the importance of a species in a given assemblage.

Similarly, resources for conservation biology are limited, and therefore it is important to focus on key species and ecosystems according to a formal criterion. For this purpose, several methods have recently been proposed [14-17]. Despite recent progresses in sequencing techniques, it is still worth following a criterion of "maximizing representativeness" to best concentrate on key taxa [e.g., [17]]. Nevertheless, this typically requires a well established phylogeny, or at least a genetic distance matrix, as a benchmark. These data are indeed generally available for model species or taxa with key ecological roles, but they are often unavailable in standard phylogenetic analyses. Typically, if we want to investigate a phylogeny, it has either never been resolved before, or it has not been completely assessed at the moment we start the analysis. Further, if a reliable and widely accepted phylogenetic hypothesis were available for the studied group, we probably would not even try to attempt to formulate one at all. This means that, while the above-mentioned methods may be useful in the case of well-characterized groups, an approach using taxonomic distinctness is more powerful in general phylogenetic practice.

Our basic idea is that estimating the phylogenetic representativeness of a given sample is not conceptually different from estimating its taxonomic distinctness. A certain degree of taxonomic distinctness is required for individual samples chosen for phylogenetic analyses; again, investigators attempt to spread sampling as widely as possible over the group on which they are focusing in order to maximize the representativeness of their study. A computable measure of taxonomic distinctness is required to describe this sampling breadth.

In this article we propose a measure of phylogenetic representativeness, and we provide the software to implement it. The procedure has the great advantage of requiring only limited taxonomical knowledge, as is typically available in new phylogenetic works.

Results

Algorithm

Clarke and Warwick [18] suggest standardizing the step lengths in a taxonomic tree structure by setting the longest path (i.e., two species connected at the highest possible level of the tree) to an arbitrary number. Generally, this number is 100. Step lengths can be weighted all the same, making the standardized length measure to equal:

$$l_n = \frac{100}{2(T-1)}$$

where T is the number of taxonomic levels considered in the tree and $n = 1, 2, ..., N$, where N is the number of steps connecting a pair of taxa (see Methods). All taxa in the tree belong by definition to the same uppermost taxon. Therefore, two taxa can be connected by a maximum of $2(T - 1)$ steps.

However, it is also possible to set step lengths proportionally to the loss of biodiversity between two consecutive hierarchical levels, i.e., the decrease in the number of taxa contained in each one, as measured on the master list. Branch lengths are then computed as follows: we indicate $S_{(t)}$ as the number of taxa of rank t, with $t = 1, 2, ..., T$ from the lowest to the highest taxonomic level. Two cases are trivial: when $t = 1$, $S_{(t)}$ equals to S (the number of Operational Taxonomic Units - OTUs - in the master taxonomic tree); when $t = T$, $S_{(t)}$ equals to 1 (all taxa belong to the uppermost level). The loss of biodiversity from level t to level $t + 1$ is:

$$\Delta S_{(t)} = S_{(t)} - S_{(t+1)} \left(1 \le t \le T-1\right)$$

The step length from level $t + 1$ to level t is the same as from level t to level $t + 1$. Therefore, path lengths are then obtained as:

$$l_t = l_{t^*} = \frac{\frac{\Delta S_{(t)}}{T-1}\times 100}{2} = \frac{\Delta S_{(t)}}{\sum_{t=1}^{T-1} \Delta S_{(t)}} \times 50, \quad t^* = N - t + 1$$

where l_t is the path length from level t to level $t + 1$ and l_{t^*} is the reverse path length.

Clarke and Warwick [18] found the method of weighting step lengths to have little effect on final results. However, we find that standardizing path lengths improves

the method in that it also complements subjectivity in taxonomies; rankings are often unrelated even across closely-related groups. To us, this is the main reason for standardizing path lengths. Moreover, adding a level in a taxonomic tree does not lead to changes in the mean or standard deviation of taxonomic distance (AvTD or VarTD) if we adopt this strategy. In addition, the insertion of a redundant subdivision cannot alter the values of the indices [18]. All these analyses are carried out by our PhyRe script (Additional file 1).

Our method based on Clarke and Warwick's ecological indices has the main feature of being dependent only upon a known existing taxonomy. This leads to a key difficulty: taxonomic structures are largely subjective constructions. Nonetheless, we think that taxonomists' expertise has provided high stability to main biological classifications, at least for commonly-studied organisms, such as animals and plants. The degree of agreement which is now reached in those fields allows us to consider most systematics as stable. In our view, large-scale rearrangements are becoming more and more unlikely, so that this argument leads us to state that present taxonomies do constitute an affordable starting point for methods of phylogenetic representativeness assessment.

However, this is not sufficient to completely ensure the reliability of our method. Knowledge is growing in all fields of evolutionary biology, and the increase in data results in constant refinement of established classifications. In fact, even if large-scale changes are rare, taxonomies are frequently revised, updated, or improved. Therefore, we implemented an algorithm that allows for testing the stability of the chosen reference taxonomy.

Essentially, our procedure can be described in two phases. In the first one, the shuffling phase, master lists are shuffled, resulting in a large number of alternative master lists. In the second, the analysis phase, a phylogenetic representativeness analysis is carried out as described above across all simulated master lists rearrangements. The shuffling phase is composed of three moves, which are repeated and combined *ad libitum* (see Methods). These moves simulate the commonest operations taxonomists do when reviewing a classification. A large number of "reviewed" master lists is then produced, repeating each time the same numbers of moves. Finally, the shuffling phase ends with a set of master lists. Standard phylogenetic representativeness analyses are performed on each master list, and all statistics are computed for each list. In this way, a set of measurements is produced for each indicator. Therefore, it is possible to compute standard 95% (two-tailed) confidence intervals for each one. This analysis phase gives an idea of the funnel plot's oscillation width upon revision. PhyloSample and PhyloAnalysis (Additional file 1) are specific scripts dealing with the shuffling analysis: the former generates the new set of master list, whereas the latter performs PhyRe operations across them all.

All scripts are available online, and a Windows executable version of the main script is also present: the software can be downloaded from the MoZoo Lab web site at http://www.mozoolab.net/index.php/software-download.html.

Testing

In order to evaluate the method, we analyze phylogenies of bivalves [19], carnivores [20], coleoids [21], and termites [22]. Our reference taxonomies are Millard [23] for mollusks, the Termites of the World list hosted at the University of Toronto http://www.utoronto.ca/forest/termite/speclist.htm: consulted on 03/23/2009 and reference therein), and the online Checklist of the Mammals of the World compiled by Robert B. Hole, Jr. (http://www.interaktv.com/MAMMALS/Mamtitl.html: consulted on 03/11/2009 and reference therein).

Results from AvTD and VarTD are shown in Figures 1 and 2, respectively. Funnel plot are based arbitrarily on 100 random samplings from the master list for each sample size. Table 1 summarizes these results, showing also results from I_E.

To assess the stability of our taxonomies by performing shuffling analyses on them, we fixed the amount of "moves" to be executed according to our knowledge of each master list (see Discussion for details; Table 2); 1,000 new "reviewed" datasets were generated and then 100 replicates were again extracted from each master list for

Table 1: Phylogenetic Representativeness analyses from four published works.

Group	Reference	Dimension	AvTD	VarTD	IE
Bivalves	[19]	9	89.7181	340.1874	0.0609
Carnivores	[20]	72	92.9688	280.2311	0.1203
Coleoids	[21]	30	90.3758	315.3069	0.1079
Termites	[22]	40	93.8788	177.1053	0.1631

Dimension, number of taxa; AvTD, Average Taxonomic Distinctness; VarTD, Variation in Taxonomic Distinctness; I_E, von Euler's [44] Index of Imbalance.

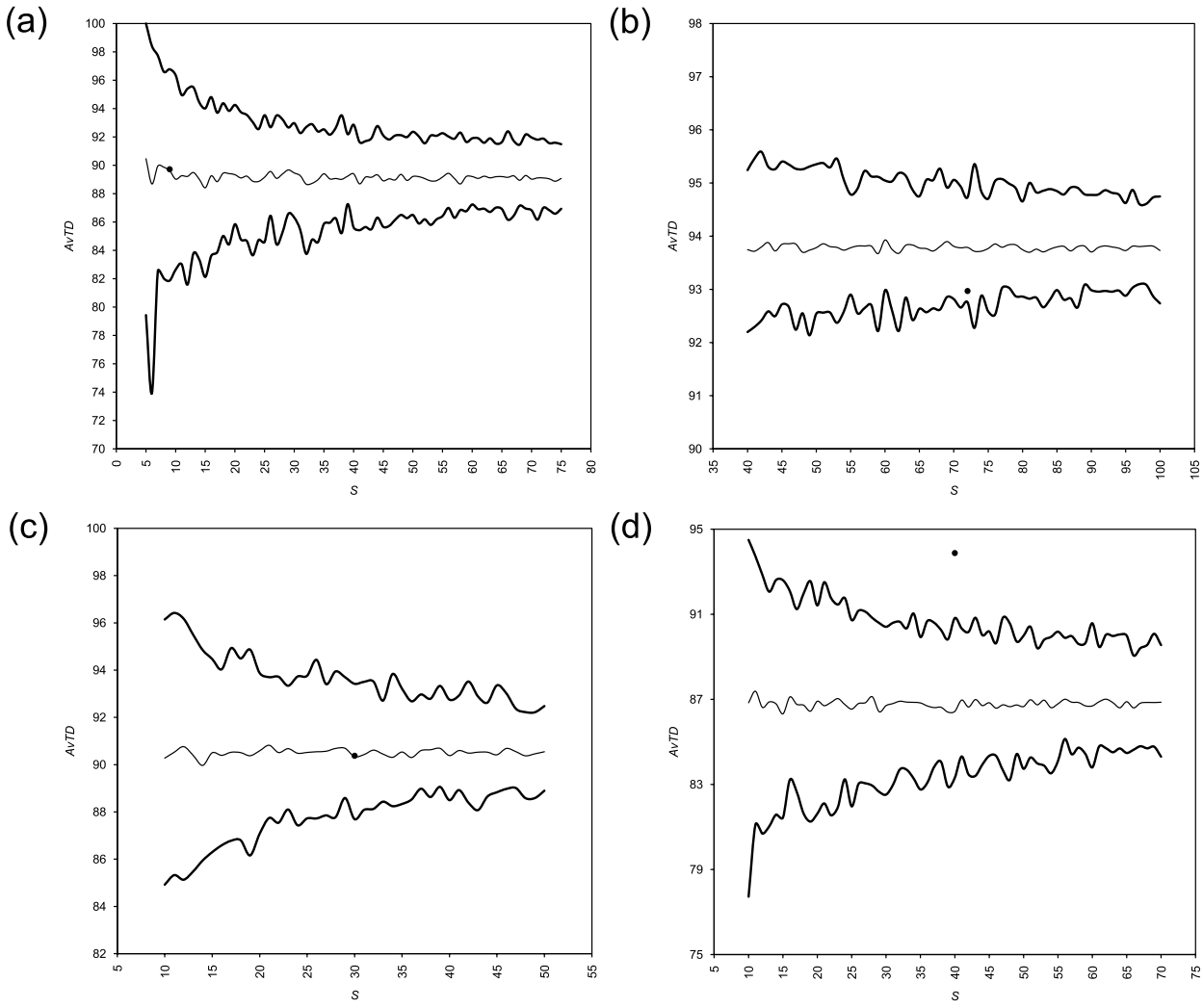


Figure 1 Funnel plots of AvTD from four published data sets. Funnel plots of Average Taxonomic Distinctness (AvTD) from (a) bivalves [19], (b) carnivores [20], (c) coleoids [21], and (d) termites [22] data sets are shown. Results are from 100 random replicates. Thick lines are the highest values found across all replicates of each dimension and the lower 95% confidence limit; the thin line is the mean across all replicates; experimental samples are shown by black dots.

each sample size. Funnel plots for AvTD and VarTD are shown in Figures 3 and 4, respectively.

We conducted additional analyses on the dataset of bivalves with real and simulated data (Additional file 2). Data from bivalve phylogenies obtained in our laboratory at different times from different samples have been tested along with imaginary samples of different known representativeness. We use the letter R to denote real data sets analyzed in our laboratory. Datasets from R1 to R4 are increasingly representative. In R1, the subclass of Protobranchia is represented by just one genus, and the subclass of Anomalodesmata is completely missing. In R2, we add one more genus to Protobranchia (Solemya) and one genus to Anomalodesmata (Thracia). In R3, the sample is expanded with several Genera from Unionidae (Anodonta, Hyriopsis), Heterodonta (Gemma, Mactra), Protobranchia (Nuculana; but see [24,25]), and more Anomalodesmata (Pandora, Cuspidaria). While all high-level taxa were already represented in R2, R3 is thus wider and more balanced in terms of sampling. R4 is identical to R3 with the exception of genus Cerastoderma, which was excluded due to technical problems.

Simulated data sets are indicated by the letter S. S1 is an "ideal" data set: all subclasses are represented with 4 species and 4 families, although the number of represented orders is different across the subclasses. S2 is biased towards less biodiversity-rich subclasses: it comprehends 6 anomalodesmatans, 6 palaeoheterodonts, and 7 protobranchs, along with only 1 pteriomorphian and one heterodont. S3 is strongly biased towards heterodonts, with

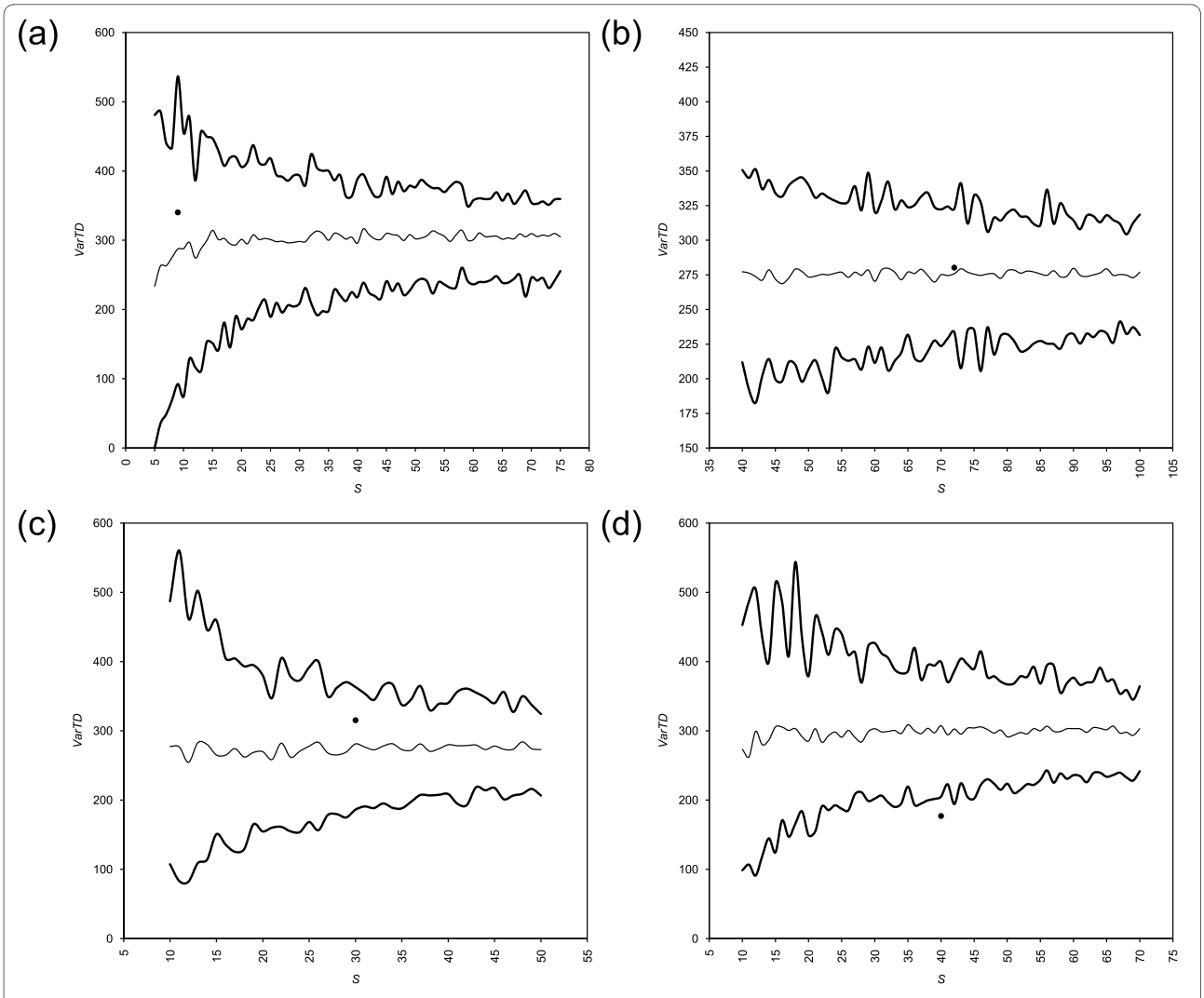


Figure 2 Funnel plots of VarTD from four published data sets. Funnel plots of Variation in Taxonomic Distinctness (VarTD) from (a) bivalves [19], (b) carnivores [20], (c) coleoids [21], and (d) termites [22] data sets are shown. Results are from 100 random replicates. Thick lines are the upper 95% confidence limit and the lowest values found across all replicates of each dimension; the thin line is the mean across all replicates; experimental samples are shown by black dots. The bias towards lower values for small sample is detectable in mean.

17 genera. Pteriomorphians, palaeoheterodonts, and protobranchs are represented by one genus each, and there are no anomalodesmatans here. S4 is an "easy-to-get" sample, with the commonest and well-known genera (e.g., *Donax*, *Chamelea*, *Teredo*, *Mytilus*, *Ostrea*), and therefore it is composed only by pteriomorphians (7 genera) and heterodonts (11 genera).

For this entire group of samples, from R1 to R4, and from S1 to S4, we conducted phylogenetic representativeness analyses to find out whether the method can describe samples following our expectations. Funnel plots were constructed on 10,000 replicates. Results are displayed in Figure 5 and Table 3.

Implementation

The distribution of AvTD from k random subsamples of size S is typically left-skewed ([26]; Figure 6). This is not an effect of a low k, as increasing the number of subsamples the shape of distribution does not change. We follow Azzalini [27] in describing the skeweness with a parameter λ. The further is λ (as absolute value) from unity, the more skewed is the distribution. Using the master list of bivalves and a dimension S of 50, we estimated an absolute value for λ which is very close to unity (~1.01, data not shown), confirming that the distribution only slightly differs from the normal one. However, this was done only for one sample, and distributions vary across different

Table 2: Shuffling moves performed on each master list.

Group	Size	Level	Splits	Merges	Transfers
Bivalves	3404	Family	15	10	40
Carnivores	271	subfamily	2	1	2
Coleoids	220	Family	2	1	2
Termites	2760	species	0	0	15

Each set of splits, merges, and transfers was repeated independently 1,000 times on the relative master list. Moves were applied to the specified taxonomic level. Master list's size is reported to inform about the entity of the "reviewing" shuffle. Size in Operational Taxonomic Units (OTUs) of the global taxonomic tree.

taxonomies and organisms. Similar considerations can be applied to VarTD.

We represent in our AvTD plots the lower 95% confidence limit (see Figures from 1 to 5). The maximum value obtained across all replicates for that dimension is also shown because it converges to the upper absolute limit as k increases. Conversely, in VarTD plots the upper 95% confidence limit and minimum observed value are shown, as lower values of variation are preferable (see Methods). PhyRe produces funnel plots showing results from a range of dimensions S. This helps in evaluating the global situation and is very useful for comparing homogeneous samples of different sizes.

For the shuffling analysis, similar funnel plots are produced. The main difference is that for AvTD the lower 95% confidence limit is not a line: here is shown the area which comprises 95% of values for each dimension across all shuffled master lists. The same applies for the AvTD

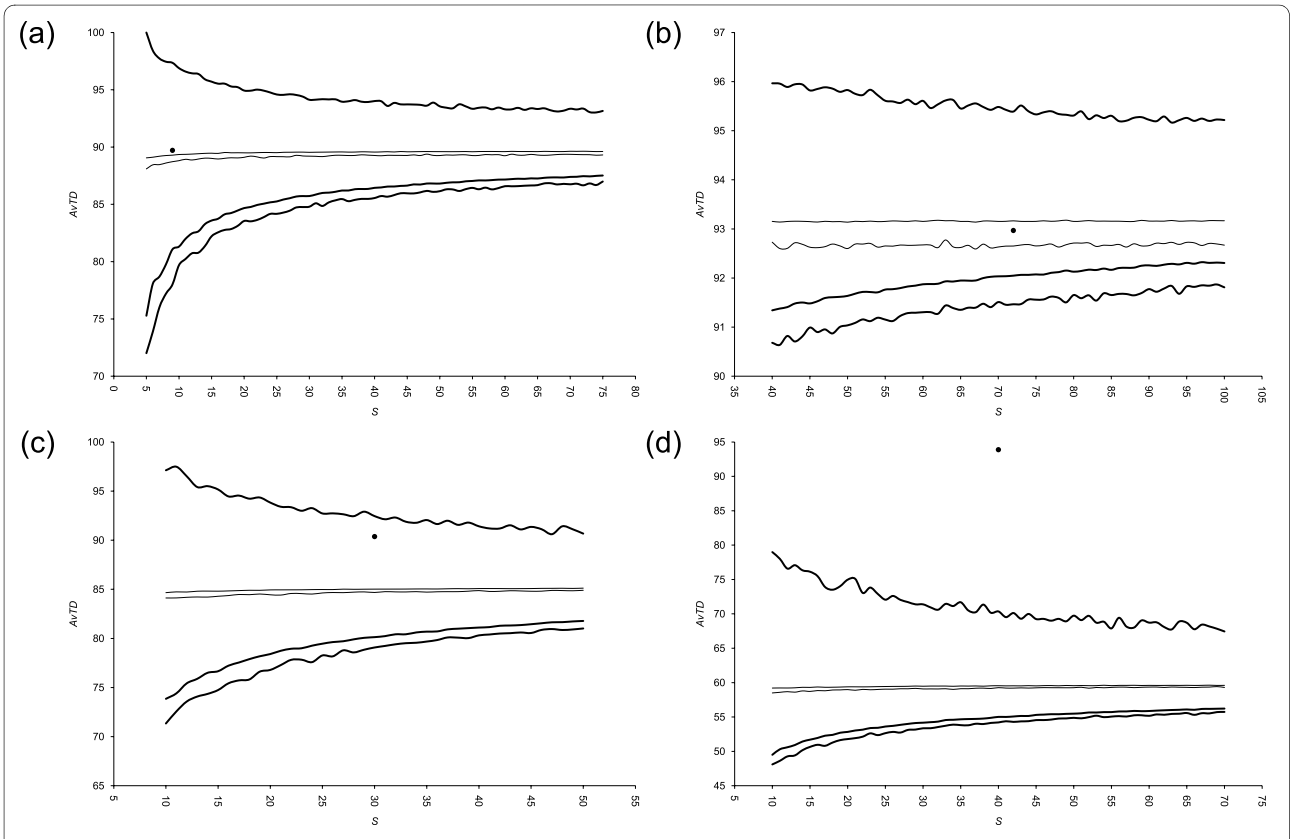


Figure 3 Funnel plots of AvTD from shuffling analyses. Funnel plots of Average Taxonomic Distinctness (AvTD) upon master lists' shuffling from (a) bivalves [19], (b) carnivores [20], (c) coleoids [21], and (d) termites [22] data sets are shown. Results are from 1,000 shuffled master lists and 100 random replicates. Thick lines are the highest values found across all replicates and the lower 95% confidence limit (2.5% and 97.5% confidence limits); thin lines represent the mean across all replicates (2.5% and 97.5% confidence limits); experimental samples are shown by black dots. Shuffling tuning as in Table 2.

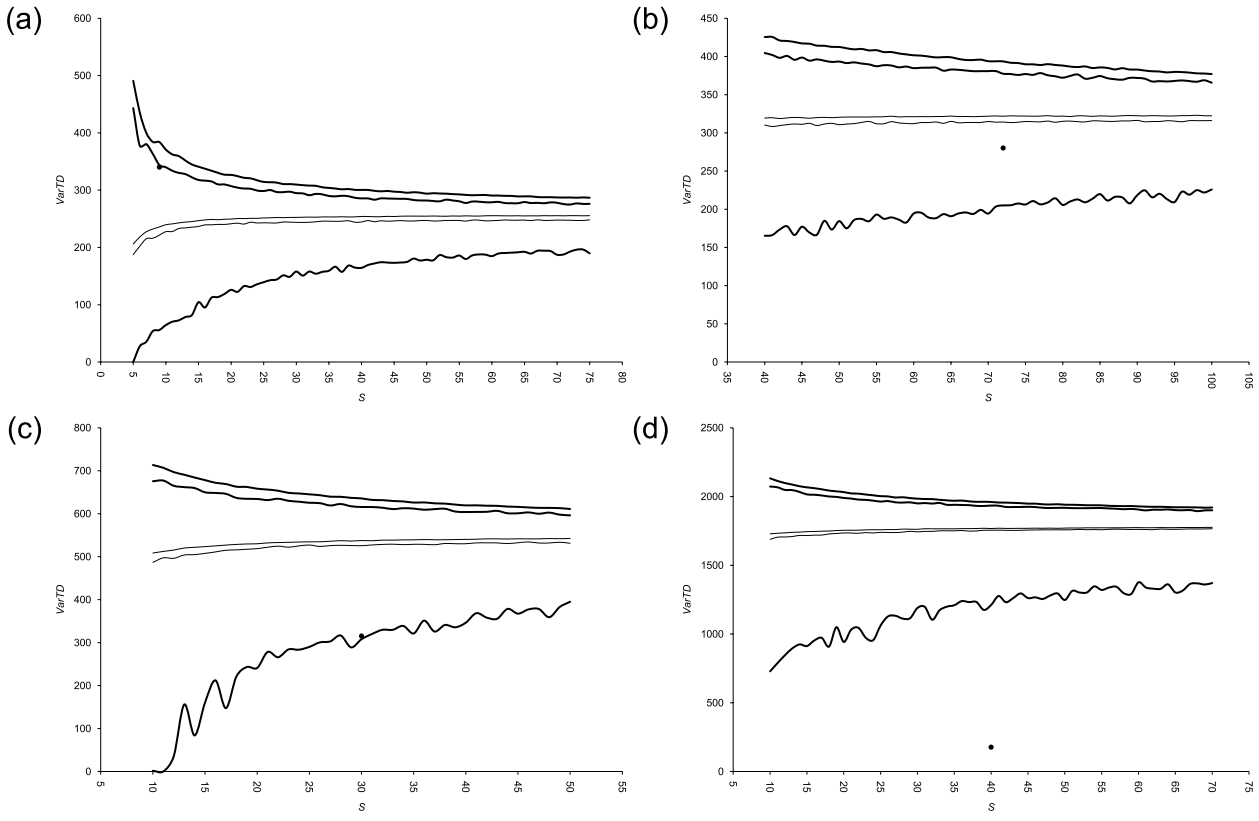


Figure 4 Funnel plots of VarTD from shuffling analyses. Funnel plots of Variation in Taxonomic Distinctness (VarTD) upon master lists' shuffling from (a) bivalves [19], (b) carnivores [20], (c) coleoids [21], and (d) termites [22] data sets are shown. Results are from 1,000 shuffled master lists and 100 random replicates. Thick lines are the upper 95% confidence limit (2.5% and 97.5% confidence limits) and the lowest values found across all replicates; thin lines represent the mean across all replicates (2.5% and 97.5% confidence limits); experimental samples are shown by black dots. Shuffling tuning as in Table 2.

and VarTD means, and the VarTD upper 95% confidence limit.

Output from PhyRe can easily be imported into a graph editing software like Microsoft Excel®.

Discussion

"Taxon sampling" is not a new topic by itself and several strategies have been proposed from different standpoints. As mentioned above, several criteria have been appraised, especially when an established phylogeny is present. Long-branch subdivision [[28,29]; and reference therein], for example, has been proposed as one strategy; see Hillis [[7]; and reference therein] for more strategies. Much experimental interest has been focused also on outgroup sampling (see, e.g., [[30,31]; and reference therein], for empirical studies) and its effects. Finally, whether it is preferable to add more characters or more taxa is a vexing question; several authors highlight the importance of adding new taxa to analyses [e.g., [32,33]]. However, Rokas and Carroll [34] point out that an increase in taxon sampling does not have an improving effect *per se*. Nevertheless, they suggest several factors which may influence the accuracy of phylogenetic reconstructions, and among them the density of taxon sampling.

Rannala et al. [35] obtained more accurate phylogenetic reconstructions when they sampled 20 taxa out of 200, rather than when 200 taxa out of 200,000 were chosen for analyses, although in the latter case the taxon number was higher. This is rather intuitive, indeed, as taxon sampling is denser in the former case. Each taxon was sampled with the same probability ρ in a birth-death process (see [35] for further details). Interestingly, this is somewhat similar to our random subsampling process: the more dense is a sample, the more likely is it to be representative of its master list, despite the absolute number of included taxa.

However, our approach is very different, because it is completely *a priori*. The method can always be applied to any phylogeny, given the presence of a reference taxonomy and a master list of taxa. We find useful to start from the zero point of no phylogenetic information except for the available taxonomy. Evolutionary systematics does

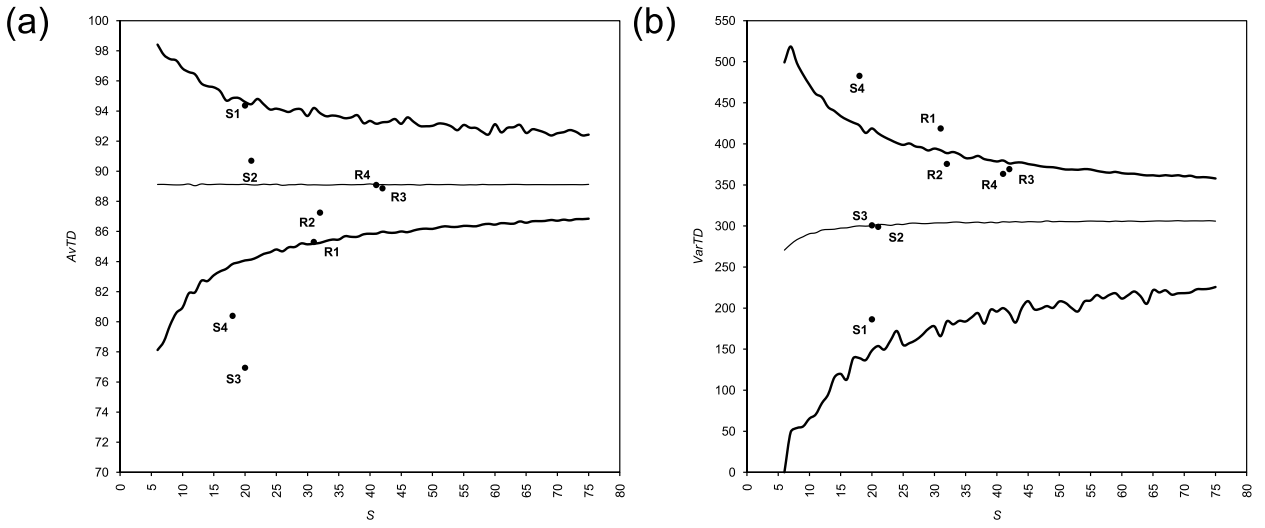


Figure 5 AvTD and VarTD from bivalve data sets. Phylogenetic Representativeness as measured by funnel plots of (a) Average Taxonomic Distinctness (AvTD) and (b) Variation in Taxonomic Distinctness (VarTD) from bivalves' master list [23]. Results are from 10,000 random replicates. Lines are as in Figure 1 and 2 for (a) and (b), respectively. Letter S denotes simulated data sets, whereas letter R denotes real ones. See text for explanation.

indeed capture some phylogenetic information, because all taxonomic categories should correspond to monophyletic clades. We employ this preliminary phylogenetic information to assess taxon sampling (but see below for further discussion on this point).

This method can be applied to every kind of analysis, from molecular to morphological ones. Furthermore, even extinct taxa can be included in a master list or in a sample: for example, the bivalve list from Millard [23] does report fossil taxa, and we left those taxa in our reference master list, as these are part of the biodiversity of the class. In fact, a good sample aims to capture the entire diversity of the group, thus including extinct forms. Therefore, we suggest that molecular samples should be better compared to complete master lists, which comprehend both living and fossil taxa (see Figure 5).

Table 3: Phylogenetic representativeness across real and simulated bivalve data sets.

Sample	Group	Dimension	AvTD	VarTD	IE
real					
R1	without anomalodesmata ns	31	85.3003	418.7537	0.2586
R2	+ *Solemya* and *Thracia*	32	87.2497	375.5878	0.2804
R3	increased (see text)	42	88.8653	369.2571	0.1806
R4	- *Cerastoderma*	41	89.0842	363.4391	0.1773
simulated					
S1	"ideal" (see text)	20	94.3673	186.2882	0.0476
S2	biased towards poor subclasses	21	90.6962	298.9607	0.1676
S3	biased towards heterodonts	20	76.9450	300.7505	0.7017
S4	"easy-to-get" (see text)	18	80.3913	482.7998	0.2419

Dimension, number of taxa; AvTD, Average Taxonomic Distinctness; VarTD, Variation in Taxonomic Distinctness; I_E, von Euler's [44] Index of Imbalance.

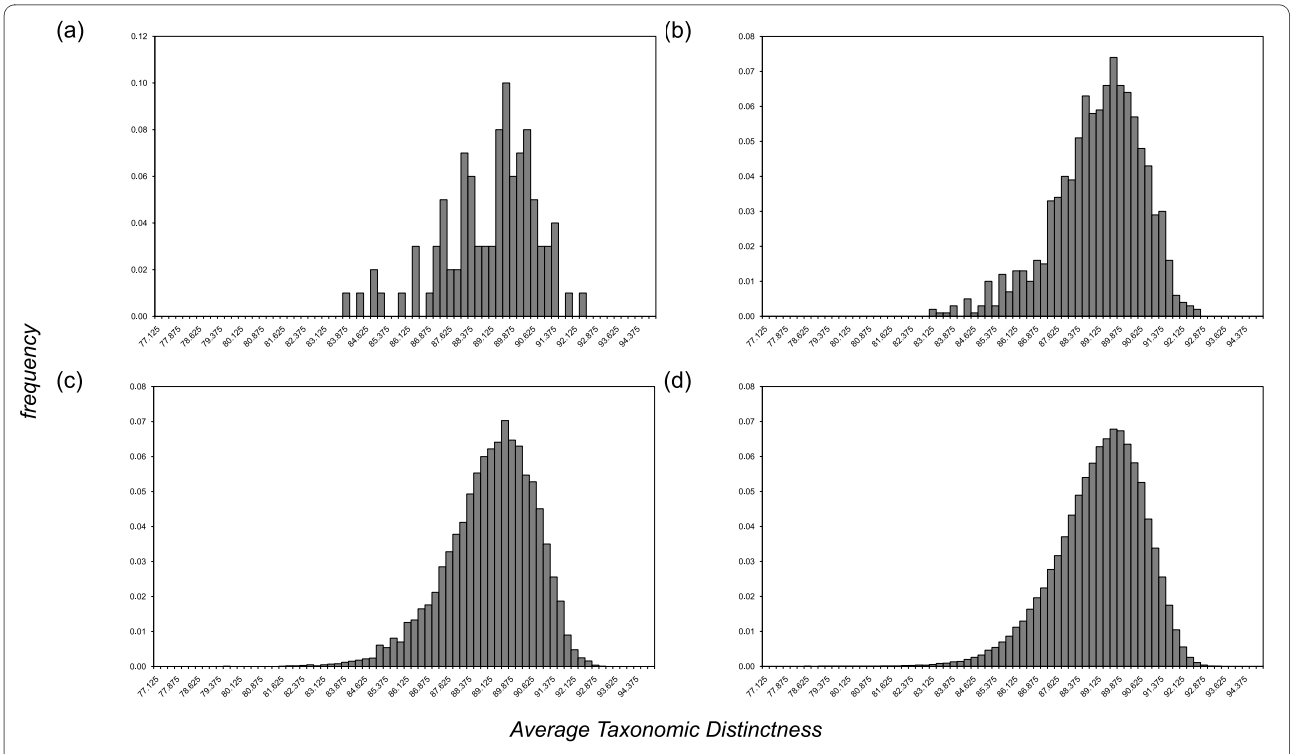


Figure 6 Average Taxonomic Distinctness distribution. Histograms show frequencies of Average Taxonomic Distinctness (AvTD) values among *k* = 100 (a), 1,000 (b), 10,000 (c), and 100,000 (d) random subsamples (*S* = 50) from bivalves' master list by Millard [23]. The distribution shows a skeweness towards the left side.

Moreover, evaluating phylogenetic representativeness as described here has the great advantage of being largely size-independent: this is well shown by funnel plots of AvTD and VarTD (Figures from 1 to 5). The mean is consistent across all dimensions *S* and it is very close to AvTD or VarTD values obtained from the whole master list (data not shown; see e.g., [26]). This fact, along with setting path lengths proportionally to biodiversity losses and rescaling their sum to 100, has a very useful and important effect: adding new taxa or new taxonomic levels does not change any parameter in the analysis. This means that more and more refined analyses can always be addressed and compared with coarser ones and with results from other data.

Most importantly, we checked the significance of both AvTD and VarTD results with one-tailed tests. The original test was two-tailed [26], and this is the greatest difference between the original test and our implementation for phylogenetic purposes. In the ecological context, these indices are used to assess environmental situations, to test for ecological stresses or pollution. In such a framework, the index must point out assemblages which are either very poor or very rich in terms of distinctness. The former will constitute signals of critically degraded habitats, whereas the latter will indicate a pristine and particularly healthy locality, and ecologists seek explanations for both results.

In our applications, we want our sample to be representative of the studied group, so that a sample significantly higher in taxonomic distinctness than a random one of the same size can be very useful; indeed, it would be even preferred. For this reason, we state that a one-tailed test is more appropriate for our purposes.

All case studies rely on samples with good phylogenetic representativeness. Nevertheless, one sample ([19]; Figure 1a and 2a) is relatively small to represent its master list; this is shown by quite large funnels at its size. On the other hand, one sample ([22]; Figure 1d and 2d) turned out to be strikingly representative of its groups: the AvTD is higher (and the VarTD lower) than the highest (lowest) found in 100 random subsamples. We recommend the former sample be taken with care for phylogenetic inferences (in fact, see [19] on the polyphyly of bivalves). Conversely, the latter sample is extremely more representative than the other three. Highly representative samples are readily individuated by AvTD and VarTD funnel plots (see Figure 1d and 2d) as dots above the highest AvTD and below the lowest VarTD found across all random replicates.

This is naturally influenced by the number of such subsamples: the more subsamples that are drawn, the more

likely is to find the absolute maximum (minimum) possible value. If *k* is sufficiently high, the absolute maximum (minimum) possible value is found for any dimension *S*, and no sample can appear above (below) those lines (see Figure 5). Therefore, we suggest to draw an intermediate number of replicates (e.g., 100 or 1,000) to avoid this widening effect and identify more optimal phylogenetic samples.

Shuffling analysis assesses the complex issue of master list subjectivity and, as such, taxonomy itself. Master lists turn out to be substantially stable upon simulated revision, as shown in Figure 3 and 4. 95% confidence areas are indeed generally narrow and the position of experimental dots is never seriously challenged. We used 100 replicates from 1,000 master lists: this turned out to be sufficient to draw clear graphs, where borders are accurately traced.

An objective criterion to describe the amount of shuffling needed for this analysis is still lacking; however, each group of living beings has its own taxonomic history and its own open problems, therefore we think it can be very difficult to find an always-optimal criterion. An expertise-driven choice cannot be ruled out here. We suggest that, given the contingent conditions of a study, phylogeneticists choose the best degree of shuffling to describe their master list's stability. Some taxonomical situations are much more consolidated than others; in some cases higher-level taxa are well-established, whereas in others agreement has been reached on lower-level ones. A formal criterion, like moving 10% of species or merging 5% of genera, will necessarily lose this faceting and complexity.

Interestingly, the coleoid master list revealed itself to be the most sensitive to shuffling. The AvTD funnel plot places the sample of [21] exactly across the mean line, whereas it is close to the maximum line in the shuffling analysis (see Figure 1c and 3c). This means that AvTD is globally lowered upon shuffling on the coleoid master list. In fact, whereas mean AvTD on the original master list was close to 90 for all *S*, the 95% confidence interval on shuffled master lists is always slightly under 85. Conversely, VarTD is over the mean in standard PhyRe computations, whereas it is across the minimum line in shuffling analysis (see Figure 2c and 4c): VarTD mean changes from about 300 in the former case to around 500 in the latter one. The amount of shuffling we applied (see Table 2) is evidently heavy in this case. Therefore, upon a taxonomic review, we would recommend to reconsider this sample and to perform a new phylogenetic representativeness analyses.

Our method is also descriptive for comparing similar samples; this is a smart way to test the improvement of a phylogenetic study while adding one or more taxa to a given sample. It is clear from our R1-R4 example (see Figure 5) the importance of adding just two taxa to the initial sample. The improvement is well depicted by AvTD and VarTD funnel plots: whereas R1 is just across the AvTD lower 95% confidence limit of AvTD, R2 is well above; whereas R1 is outside the VarTD upper 95% confidence limit, R2 is inside it. While VarTD remains close to the confidence limit, R3 and R4 are nevertheless even more representative in terms of AvTD, as they lie precisely on the mean of 10,000 replicates. This reflects the increase of sampled taxa with respect to several under-represented groups.

S1, the "ideal" sample, turns out to have the highest AvTD (across the maximum line) and the lowest VarTD (next to the minimum line). In this case, we have 10,000 replicates; thus, the above considerations hold true and we do not expect our dot to be neither above nor below the funnel plot for AvTD or VarTD, respectively. Sample S2, biased towards less biodiversity-rich subclasses appears to be representative: it is well inside both funnel plots. Three subclasses out of five are well represented here; this sample is therefore rather informative. However, it is clearly less preferable than sample S1; whereas the former lies always across or next to the mean line, the latter is always close to the observed extreme values. Sample S3 seems reasonable in terms of VarTD, but the AvTD funnel plot identifies it as the worst of all. Nevertheless, sample S4 (with two substantially equally-represented subclasses) turned out to be even worse than S3 (almost just one subclass included): it is below the 95% confidence limit of AvTD and above the 95% confidence limit of VarTD.

Thus, joint analysis of AvTD and VarTD provides discrimination between samples. An AvTD/VarTD plot shows that these measures are generally negatively correlated, even if some exceptions are possible: good samples have high AvTD and low VarTD values; the opposite is true for bad samples (Figure 7).

Along with the two main measures, I_E can give an approximate idea of the shape of the tree. Values > 0.25 are often associated with biased samples (see Table 3), and thus we suggest this as a rule of thumb for directly discarding imbalanced ones. However, this cut-off value is only a rough guide in estimating phylogenetic representativeness: sample R2 has an I_E of 0.2804 (greater than R1), but funnel plots identify it as a good bivalve sample.

Conclusions

Phylogenetic representativeness analyses can be conducted at every taxonomic level, and including any taxonomic category. Moreover, inclusion or exclusion of taxonomic categories does not influence results across analyses ([18]; see above). Although we did not present it here, the index can also potentially take relative abundance data into account [see [36,37,26]]. Thus, it may be implemented for population-level analyses as well,

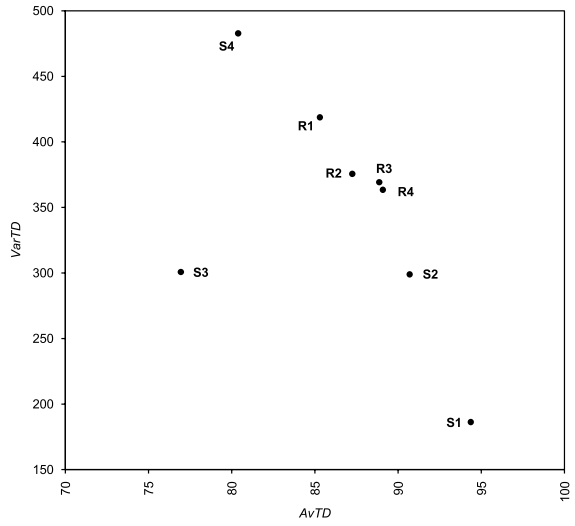


Figure 7 AvTD-VarTD plot. Variation in Taxonomic Distinctness (VarTD) plotted on Average Taxonomic Distinctness (AvTD) for real and simulated bivalve datasets (see Table 3 for further details on samples).

depicting sampling coverage among different populations from a given section, species, or subspecies.

The main strength of phylogenetic representativeness approach lies in being an *a priori* strategy of taxon selection and sampling. Therefore, it cannot take into account several empirical and experimental problems, which are not guaranteed to be avoided. For example, long-branch attraction depends essentially upon a particularly quick rate of evolution in single taxa [38], which is only *a posteriori* identified. Moreover, topology alteration due to outgroup misspecification remains possible, as phylogenetic representativeness deals only with ingroup taxa.

Each particular study copes with specific difficulties strictly inherent to contingent conditions; for example, as a result of an unexpected selective pressure, one particular locus may turn out to be completely uninformative, even if the taxon sampling is perfectly adequate. Nevertheless, in R1-R4/S1-S4 examples (see above), our knowledge of bivalve evolution and systematics allows us to discriminate between suitable and non-suitable samples, and phylogenetic representativeness results matched perfectly with our expectations.

Moreover, being understood that expertise is always expected in planning taxon sampling, we strongly suggest to set phylogenetic representativeness alongside a formal criterion for profiling phylogenetic informativeness of characters [e.g., [39]]. Put in other words, phylogenetic representativeness is a guarantee of a good and wise taxonomic coverage of the ingroup, but evidently it is not guarantee of a good and robust phylogeny *per se*. For this reason, we would suggest it as a springboard for every phylogenetic study, from which subsequent analyses can proceed further towards an affordable evolutionary tree.

Methods

Average Taxonomic Distinctness (AvTD)

Mathematical aspects of this index are well explained in works by Clarke and Warwick [36,26,40]. However, it is useful to explain here the main points of their statistics.

AvTD is computed starting from a taxonomic tree. A taxonomic tree is merely the graphical representation of a Linnean classification, whereby OTUs are arranged hierarchically into different categories or taxa, with taxa being mutually exclusive. We use the general terms "OTUs" and "taxa" because a taxonomic tree does not necessarily include species at their tips, nor do all taxonomic trees take into account exactly the same levels of systematics.

A simple taxonomic tree is depicted in Figure 8. Each leaf is an OTU and each node is a taxon; for example, OTUs may correspond to species and deeper nodes to genera, families, and orders as we climb up the tree. On a tree such as this, we can define a tree metric of taxonomic distance between any given pair of OTUs. A taxonomic tree is rooted (by definition); therefore, it is necessary to specify that our tree metric is unrooted (see [16]), i.e., the distance between two taxa is the shortest path on the tree that leads from one to another, and it is not required to climb up the tree from the first taxon to the root and then down to the second one, otherwise all pairs of OTUs would score the same distance.

Let us indicate with ω_{ij} the taxonomic distance between OTUs i and j, which are joined by N steps (branches) on the tree. Now we can define:

$$\omega_{ij} = \sum_{n=1}^{N} l_n$$

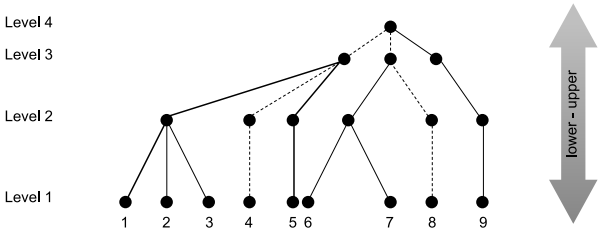


Figure 8 A hypothetical taxonomic tree. Nine Operational Taxonomic Units (OTUs) and four taxonomic levels are shown. For example, levels 1-4 could correspond to species, genera, families, and orders, respectively; in this case, species 1, 2, and 3 would belong to the same genus, species 1, 2, 3, and 4 to the same family, and so on. Taxonomic paths connecting taxa 1 and 5 (thick lines) and taxa 4 and 8 (dashed lines) are marked. See text for more details.

where l_n is the length of the nth branch, $n = 1, 2, ..., N$. We do not want to rely on information about mutation rates nor genetic distances. If we consider that a Linnean classification is mostly arbitrary, we can set branch lengths in several ways. Further considerations on this point are given above (Results; but see also [18]). The simplest case is considering a length equal to 1 for all branches. Accordingly, the distance between taxa 1 and 5 in Figure 8 is 4, and the distance between taxa 4 and 8 is 6. Indeed, taxa 1 and 4 are more closely related than taxa 4 and 8 are. The Average Taxonomic Distinctness (AvTD) of the tree is defined as the average of all such pairwise distances:

$$AvTD = \frac{\sum_{i=1}^{S}\sum_{j>i}^{S}\omega_{ij}}{\frac{S(S-1)}{2}}$$

(modified from [26])

where S is the number of taxa in the tree. Given the presence/absence data case, and with the distance between taxa i and j, being $i = j$, set to 0 (same taxon), we note that the formula can be reduced to the computationally simpler form:

$$AvTD = \frac{\sum_{i=1}^{S}\sum_{j=1}^{S}\omega_{ij}}{S(S-1)}$$

For example, the AvTD for the tree in Figure 8 would equal approximately 5.0556. The original formulation of the index considers also relative abundances of species, but here we only take into account presence/absence of OTUs.

This is the basic statistic described in this work. AvTD has been shown to be a good ecological indicator and a reliable estimator of biodiversity [37,41-43]. The most appealing feature is its clear independence from sampling effort ([36,37]; see Discussion above).

Test of significance

The AvTD statistic simply gives the expected path length for a randomly selected pair of species from the set of S species [26]. The higher the AvTD, the more taxonomically distinct is the sample. However, it is necessary to compare the AvTD of a sample to the master list from which it is taken; for example, we may be interested in the molecular phylogeny of an order and we sampled and sequenced S species within this order. Naturally, we wish to maximize the number of families and genera represented therein. Using the AvTD method, we can estimate this "maximization" by computing the index for our sample of S species, and then comparing it with one computed from the list of all species belonging to the order itself. However, comparing a pure number to another pure number is rather uninformative; therefore, a random resampling approach to test for significance is suggested here. The rationale is as follows: we must estimate whether our sample's AvTD ($AvTD_S$) is significantly different from the master list's one. Although the index is poorly dependent on sampling effort, we have to take into account that often the master list is consistently bigger than our sample. Thus, we draw k samples of size S from master list. We then compute AvTD from all k sample and test whether $AvTD_S$ falls within the 95% confidence limits of the distribution (original two-tailed test; but see Discussion above).

Variation in Taxonomic Distinctness (VarTD)

As noted by Clarke and Warwick [40], some differences in the structure of the taxonomic trees of samples are not fully resolved by AvTD measures. Two taxonomic trees could have very different structures, in terms of subdivision of taxa into upper-level categories, but nevertheless could have the same AvTD. Differences in taxonomic structures of samples are well described by a further index of biodiversity, the Variation in Taxonomic Distinctness (VarTD).

VarTD is computed as a standard statistical variance. It captures the distribution of taxa between levels, and should be added to AvTD in order to obtain a good measure of biodiversity. Clarke and Warwick [26] demonstrated that VarTD can be estimated via a precise formula, but can also be obtained in the canonical statistical way from AvTD data.

Clarke and Warwick [40] proposed to follow the same procedure as above: observed VarTD is compared with values from random resamplings of the same size. Lower values of VarTD are preferable, as they are an indication of equal subdivision of taxa among intermediate levels. Clarke and Warwick [40] also show that VarTD is not as independent from sampling effort as AvTD is, i.e., there is a bias towards lower values for very small S (see Figure 2 and 4), but it can be shown [40] that this bias becomes rather negligible for S >10.

Von Euler's index of imbalance

Following the idea of AvTD, von Euler [44] proposed an index related to taxonomic distinctness, which he called an *index of imbalance*. An index of imbalance measures the imbalance of the tree, i.e., whether and how much certain groups are under-represented and certain others are over-represented. This was not the first of such indexes [e.g., [45-48]]; however, as noted by Mooers and

Heard [49], they do not apply to trees with polytomies, as taxonomic trees often are. Von Euler's index of imbalance (I_E) is defined as:

$$I_E = \frac{AvTD_{max} - AvTD}{AvTD_{max} - AvTD_{min}}$$

where $AvTD_{max}$ and $AvTD_{min}$ are respectively the maximum and minimum possible AvTDs given a particular sample. $AvTD_{max}$ is obtained from a totally-balanced tree constructed on the given taxa, whereas $AvTD_{min}$ is obtained from a totally-imbalanced one.

Figure 9 depicts such trees as computed from the taxonomic tree shown in Figure 8; taxonomic levels are considered as orders, families, genera, and species. (i) *Obtaining a completely imbalanced tree*. The procedure is bottom-up. Each species is assigned to a different genus (left side, thick lines, species 1, 2, 3, 4, and 5), until the number of "occupied" genera equals the total number of genera minus one. Remaining species are then lumped in the last genus (right side, thick lines, species 6, 7, 8, and 9). The same procedure is repeated in assigning genera to families (dashed lines). As we consider only one order, all families are lumped in it (dotted lines). More generally, the procedure is repeated until the uppermost hierarchical level is reached. (ii) *Obtaining a completely balanced tree*. The procedure is top-down. The first step is forced, as all Families must be lumped in the only present order (dotted lines). Then we proceed assigning (as far as possible) the same number of genera to each Family. In this case, we have 6 genera for 3 families, therefore it is very easy to see that the optimal distribution is 6/3 = 2 genera/family (dashed lines). The same step is repeated until the lowermost hierarchical level is reached. Each time we try to optimize the number of taxa which are assigned to all upper levels. We have in this case 9 species for 6 genera (thick lines). Necessarily we will have at best 3 genera with 2 species and 3 genera with 1 species (3 × 2 + 3 × 1 = 9). The optimal situation is the one depicted in the figure. For this reason, it is important to balance taxa not only with respect to the immediately upper taxon, but also with respect to all upper taxa. We note that the completely-balanced and completely-imbalanced trees may not be unique. However, differences in AvTD from different equally-balanced or equally-imbalanced trees are null or negligible.

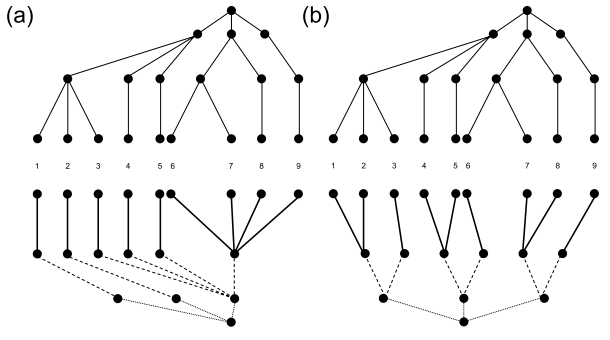


Figure 9 Totally-imbalanced and totally-balanced taxonomic trees. Totally-imbalanced (a) and totally-balanced (b) taxonomic trees computed starting from the taxonomic tree introduced in Figure 8 and shown at the top of both sides. See text for more details.

As the original formulation of AvTD, von Euler's index of imbalance was introduced in the conservation context, since it was used to take estimates on the loss of evolutionary history, and was found to be strictly (negatively) correlated with AvTD (pers. obs.; [44]). We introduce I_E in our topic, stating it is a useful balancing indicator for samples used in phylogenetic studies.

Shuffling analysis

Shuffling analysis concepts and purposes are extensively explained in the Results section. Here we think it is useful to report algorithms that were written to carry it out, especially for shuffling phase.

Shuffling phase

User inputs the number of shuffled master lists they want to generate. The user must also decide the number of repetitions for each kind of move. Therefore, each of the following algorithms is repeated the given number of times on the same master list. Then, the resulting file is saved to disk and a new one is produced, with same modalities.

Move: **Transfer**

1. user is requested to input a taxon level *t*, with *t = 1, 2, ..., T - 1*;
2. a taxon *a* of level *t* is randomly chosen;
3. *if* taxon *A* of level *t + 1* containing *a* contains only *a*
   then return to 2;
   else proceed to 4;
4. a taxon *B* of level *t + 1* is randomly chosen;
5. *if* taxon *B* = taxon *A*
   then return to 4;
   else proceed to 6;
6. taxon *a* is moved to taxon *B*.

Move: **Split**

1. user is requested to input a taxon level *t*, with *t = 2, ..., T - 1*;
2. a taxon *a* of level *t* is randomly chosen;
3. taxon *a* is split into two new taxa in the same position.

Move: **Merge**

1. user is requested to input a taxon level *t*, with *t = 2, ..., T - 1*;
2. a taxon *a* of level *t* is randomly chosen;

3. *if* taxon A of level $t + 1$ containing a contains only a
   then return to 2;
   else proceed to 4;
4. a taxon b of level t is randomly chosen within taxon A;
5. *if* $a = b$
   then return to 4;
   else proceed to 6;
6. taxa a and b are merged in a new taxon in the same position.

In all moves, downstream relationships are maintained. For example, if genus a containing species α and β is moved from family A to family B, species α and β will still belong to genus a within family B. The same holds true for splits and merges.

Analysis phase

In this phase, the basic phylogenetic representativeness analysis is applied on each master list. Therefore, a large number (depending upon the chosen number of master lists to be simulated) of analyses are performed and consequently six sets of measurements are obtained for each dimension s, namely the six parameters describing AvTD and VarTD:

lower AvTD 95% confidence limit;
mean AvTD;
mean VarTD;
upper VarTD 95% confidence limit;
maximum AvTD;
minimum VarTD;

For the first four sets of measurements, upper and lower 95% confidence limits are computed for each dimension s across all master lists, thus giving an idea of the stability of results. For the fifth and sixth sets of measurement, simply the maximum entry is kept for each dimension s as above.

Additional material

Additional file 1 PhyRe scripts and documentation. Three Python scripts constitute the PhyRe package. PhyRe script itself performs main analyses presented in this paper: AvTD, VarTD, I_E, and funnel plots parameters are computed by this script. PhyloSample generates shuffled master lists, whereas PhyloAnalysis repeats PhyRe tasks across all newly-generated master lists. All scripts have been tested under Python 2.5.4. PhyRe documentation (doc.pdf) and eight sample files referring to datasets used in the paper to validate the method [19-22] are also enclosed.

Additional file 2 Real and simulated data from bivalve data set. Real and simulated data from bivalves data set follow Millard [23] reference taxonomy. Table shows the composition of our real and simulated samples of bivalves. Taxonomy is reported for each genus; a plus "+" sign indicates the presence of that genus in that sample.

Authors' contributions
FP conceived the study and developed the Clarke and Warwick's statistics in a phylogenetic framework. RRF wrote the PhyRe software and helped to draft the manuscript. MP participated in designing and coordinating this study, and provided many essential comments. All authors read and approved the final manuscript.

Acknowledgements
We would thank professor V. U. Ceccherelli, who first drove our attention on Clarke and Warwick's statistics and suggested how they could be implemented for phylogenetic representativeness. Thanks to S. Ghesini for her invaluable statistical support and helpful comments. This work was also possible thanks to mathematical support from S. Ghesini, P. Plazzi, and C. Zucchini. Thanks also to Frank Lad, who provided useful comments on an earlier draft of this manuscript, and to anonymous reviewers for their useful comments and suggestions. This work has been supported by Italian MIUR and "Canziani Bequest" funds.

Author Details
[1]Department of "Biologia Evoluzionistica Sperimentale", University of Bologna, Via Selmi, 3 - 40126 Bologna, Italy and [2]Department of Biology and Evolution, University of Ferrara, Via Borsari, 46 - 44100 Ferrara, Italy

Received: 22 December 2009 Accepted: 27 April 2010
Published: 27 April 2010

References

1. Ilves KL, Taylor EB: **Molecular resolution of the systematics of a problematic group of fishes (Teleostei: Osmeridae) and evidence for morphological homoplasy.** *Mol Phylogenet Evol* 2009, **50:**163-178.
2. Jenner RA, Dhubhghaill CN, Ferla MP, Wills MA: **Eumalacostracan phylogeny and total evidence: Limitations of the usual suspects.** *BMC Evolutionary Biology* 2009, **9:**21.
3. Palero F, Crandall KA, Abelló P, Macpherson E, Pascual M: **Phylogenetic relationships between spiny, slipper and coral lobsters (Crustacea, Decapoda, Achelata).** *Mol Phylogenet Evol* 2009, **50:**152-162.
4. Ruiz C, Jordal B, Serrano J: **Molecular phylogeny of the tribe Sphodrini (Coleoptera: Carabidae) based on mitochondrial and nuclear markers.** *Mol Phylogenet Evol* 2009, **50:**44-58.
5. Tsui CKM, Marshall W, Yokoyama R, Honda D, Lippmeier JC, Craven KD, Peterson PD, Berbee ML: **Labyrinthulomycetes phylogeny and its implications for the evolutionary loss of chloroplasts and gain of ectoplasmic gliding.** *Mol Phylogenet Evol* 2009, **50:**129-140.
6. Whitehead A: **Comparative mitochondrial genomics within and among species of killfish.** *BMC Evolutionary Biology* 2009, **9:**11.
7. Hillis DM: **Taxonomic Sampling, Phylogenetic Accuracy, and Investigator Bias.** *Syst Biol* 1998, **47:**3-8.
8. May RM: **Taxonomy as destiny.** *Nature* 1990, **347:**129-130.
9. Vane-Wright RI, Humphries CJ, Williams PH: **What to protect? Systematics and the agony of choice.** *Biol Conserv* 1991, **55:**235-254.
10. Whittaker RH: **Evolution and measurement of species diversity.** *Taxon* 1972, **21:**213-251.
11. Peet RK: **The measurement of species diversity.** *Ann Rev Ecol Syst* 1974, **5:**285-307.
12. Taylor LR: **Bates, Williams, Hutchinson - A variety of diversities.** *Symp R Ent Soc Lond* 1978, **9:**1-18.
13. Bond WJ: **Describing and conserving biotic diversity.** In *Biotic Diversity in Southern Africa: Concepts and Conservation* Edited by: Huntley BJ. Cape Town, Oxford University Press; 1989:2-18.
14. Ricotta C, Avena GC: **An information-theoretical measure of taxonomic diversity.** *Acta Biotheor* 2003, **51:**35-41.
15. Pardi F, Goldman N: **Species choice for comparative genomics: Being greedy works.** *PLoS Genet* 2005, **1:**e71.
16. Pardi F, Goldman N: **Resource-aware taxon selection for maximizing phylogenetic diversity.** *Syst Biol* 2007, **56:**431-444.
17. Bordewich M, Rodrigo AG, Semple C: **Selecting taxa to save or sequence: Desiderable criteria and a greedy solution.** *Syst Biol* 2008, **57:**825-834.
18. Clarke KR, Warwick RM: **The taxonomic distinctness measure of biodiversity: Weighting of step lengths between hierarchical levels.** *Mar Ecol Prog Ser* 1999, **184:**21-29.
19. Passamaneck YJ, Schlander C, Halanych KM: **Investigation of molluscan phylogeny using large-subunit and small-subunit nuclear rRNA sequences.** *Mol Phylogenet Evol* 2004, **32:**25-38.

20. Flynn JJ, Finarelli JA, Zehr S, Hsu J, Nedbal MA: **Molecular phylogeny of the Carnivora (Mammalia): Assessing the impact of increased sampling on resolving enigmatic relationships.** *Syst Biol* 2005, **54:**317-337.
21. Strugnell J, Norman M, Jackson J, Drummond AJ, Cooper A: **Molecular phylogeny of coleoid cephalopods (Mollusca: Cephalopoda) using a multigene approach; the effect of data partitioning on resolving phylogenies in a Bayesian framework.** *Mol Phylogenet Evol* 2005, **37:**426-441.
22. Legendre F, Whiting MF, Bordereau C, Cancello EM, Evans TA: **The phylogeny of termites (Dictyoptera: Isoptera) based on mitochondrial and nuclear markers: Implications for the evolution of the worker and pseudergate castes, and foraging behaviors.** *Mol Phylogenet Evol* 2008, **48:**615-627.
23. Millard V: *Classification of Mollusca: A classification of world wide Mollusca Volume 3*. 2nd edition. South Africa, printed by the author; 2001:915-1447.
24. Giribet G, Wheeler W: **On bivalve phylogeny: A high-level analysis of the Bivalvia (Mollusca) based on combined morphology and DNA sequence data.** *Invert Biol* 2002, **121:**271-324.
25. Giribet G, Distel DL: **Bivalve phylogeny and molecular data.** In *Molecular systematics and phylogeography of mollusks* Edited by: Lydeard C, Lindberg DR. Washington, Smithsonian Books; 2003:45-90.
26. Clarke KR, Warwick RM: **A taxonomic distinctness index and its statistical properties.** *J Appl Ecol* 1998, **35:**523-531.
27. Azzalini A: **A class of distributions which includes the normal ones.** *Scand J Statist* 1985, **12:**171-178.
28. Hendy MD, Penny D: **A framework for the quantitative study of evolutionary trees.** *Syst Zool* 1989, **38:**297-309.
29. Poe S: **Evaluation of the strategy of Long-Branch Subdivision to improve the accuracy of phylogenetic methods.** *Syst Biol* 2003, **52:**423-428.
30. Giribet G, Carranza S: **Point counter point. What can 18S rDNA da for bivalve phylogeny?** *J Mol Evol* 1999, **48:**256-258.
31. Puslednik L, Serb JM: **Molecular phylogenetics of the Pectinidae (Mollusca: Bivalvia) and effect of increased taxon sampling and outgroup selection on tree topology.** *Mol Phylogenet Evol* 2008, **48:**1178-1188.
32. Pollock DD, Bruno WJ: **Assessing an unknown evolutionary process: Effect of increasing site-specific knowledge through taxon addition.** *Mol Biol Evol* 2000, **17:**1854-1858.
33. Hedtke SM, Townsend TM, Hillis DM: **Resolution of phylogenetic conflict in large data sets by increased taxon sampling.** *Syst Biol* 2006, **55:**522-529.
34. Rokas A, Carroll SB: **More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy.** *Mol Biol Evol* 2005, **22:**1337-1344.
35. Rannala B, Huelsenbeck JP, Yang Z, Nielsen R: **Taxon sampling and the accuracy of large phylogenies.** *Syst Biol* 1998, **47:**702-710.
36. Warwick RM, Clarke KR: **New "biodiversity" measures reveal a decrease in taxonomic distinctness with increasing stress.** *Mar Ecol Prog Ser* 1995, **129:**301-305.
37. Warwick RM, Clarke KR: **Taxonomic distinctness and environmental assessment.** *J Appl Ecol* 1998, **35:**532-543.
38. Felsenstein J: **Cases in which parsimony and compatibility will be positively misleading.** *Syst Zool* 1978, **27:**401-410.
39. Townsend JP: **Profiling phylogenetic informativeness.** *Syst Biol* 2007, **56:**222-231.
40. Clarke KR, Warwick RM: **A further biodiversity index applicable to species lists: Variation in taxonomic distinctness.** *Mar Ecol Prog Ser* 2001, **216:**265-278.
41. Warwick RM, Light J: **Death assemblages of molluscs on St Martin's Flats, Isles of Scilly: A surrogate for regional biodiversity?** *Biodivers Conserv* 2002, **11:**99-112.
42. Warwick RM, Turk SM: **Predicting climate change effects on marine biodiversity: Comparison of recent and fossil molluscan death assemblages.** *J Mar Biol Ass UK* 2002, **82:**847-850.
43. Leonard DRP, Clarke KR, Somerfield PJ, Warwick RM: **The application of an indicator based on taxonomic distinctness for UK marine nematode assessments.** *J Environ Manage* 2006, **78:**52-62.
44. von Euler F: **Selective extinction and rapid loss of evolutionary history in bird fauna.** *Proc R Soc Lond B* 2001, **268:**127-130.
45. Colless DH: **Phylogenetics: The theory and practice in phylogenetic systematic II. Book review.** *Syst Zool* 1982, **31:**100-104.
46. Shao K-T, Sokal RR: **Tree balance.** *Syst Zool* 1990, **39:**266-276.
47. Heard SB: **Patterns in tree balance among cladistic, phenetic, and randomly generated phylogenetic trees.** *Evolution* 1992, **46:**1818-1826.
48. Kirkpatrick M, Slatkin M: **Searching for evolutionary patterns in the shape of a phylogenetic tree.** *Evolution* 1993, **47:**1171-1181.
49. Mooers AØ, Heard SB: **Inferring evolutionary process from phylogenetic tree shape.** *Q Rev Biol* 1997, **72:**31-54.

doi: 10.1186/1471-2105-11-209
Cite this article as: Plazzi *et al.*, Phylogenetic representativeness: a new method for evaluating taxon sampling in evolutionary studies *BMC Bioinformatics* 2010, **11**:209

Molecular Phylogenetics and Evolution 57 (2010) 641–657

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Towards a molecular phylogeny of Mollusks: Bivalves' early evolution as revealed by mitochondrial genes

Federico Plazzi *, Marco Passamonti

Department of Biologia Evoluzionistica Sperimentale, University of Bologna, Via Selmi 3, 40126 Bologna, Italy

ARTICLE INFO

Article history:
Received 4 February 2010
Revised 31 July 2010
Accepted 27 August 2010
Available online 9 September 2010

Keywords:
Bivalvia
Codon model
Penalized likelihood
Bayesian analysis
Phylogenetics

ABSTRACT

Despite huge fossil, morphological and molecular data, bivalves' early evolutionary history is still a matter of debate: recently, established phylogeny has been mostly challenged by DNA studies, and little agreement has been reached in literature, because of a substantial lack of widely-accepted methodological approaches to retrieve and analyze bivalves' molecular data. Here we present a molecular phylogeny of the class based on four mitochondrial genes (*12s*, *16s*, *cox1*, *cytb*) and a methodological pipeline that proved to be useful to obtain robust results. Actually, best-performing taxon sampling and alignment strategies were tested, and several data partitioning and molecular evolution models were analyzed, thus demonstrating the utility of Bayesian inference and the importance of molding and implementing non-trivial evolutionary models. Therefore, our analysis allowed to target many taxonomic questions of Bivalvia, and to obtain a complete time calibration of the tree depicting bivalves' earlier natural history main events, which mostly dated in the late Cambrian.

1. Introduction

Bivalves are among the most common organisms in marine and freshwater environments, summing up to about 8000 species (Morton, 1996). They are characterized by a bivalve shell, filtrating gills called ctenidia, and no differentiated head and radula. Most bivalves are filter-feeders and burrowers or rock-borers, but swimming or even active predation are also found (Dreyer et al., 2003). Most commonly, they breed by releasing gametes into the water column, but some exceptions are known, including brooding (Ó Foighil and Taylor, 2000). Free-swimming planktonic larvae (veligers), contributing to species dispersion, are typically found, which eventually metamorphose to benthonic sub-adults.

Bivalve taxonomy and phylogeny are long-debated issues, and a complete agreement has not been reached yet, even if this class is well known and huge fossil records are available. In fact, bivalves' considerable morphological dataset has neither led to a stable phylogeny, nor to a truly widely accepted higher-level taxonomy. As soon as they became available, molecular data gave significant contributions to bivalve taxonomy and phylogenetics, but little consensus has been reached in literature because of a substantial lack of shared methodological approaches to retrieve and analyze bivalves' molecular data. Moreover, to improve bivalves' phylogenetics, several attempts to join morphology and molecules have also been proposed (Giribet and Wheeler, 2002; Giribet and Distel, 2003; Harper et al., 2006; Mikkelsen et al., 2006; Olu-Le Roy et al., 2007), since, according to Giribet and Distel (2003), morphology resolves deeper nodes better than molecules, whereas sequence data are more adequate for recent splits.

Bivalves are generally divided into five extant subclasses, which were mainly established on body and shell morphology, namely Protobranchia, Palaeoheterodonta, Pteriomorphia, Heterodonta and Anomalodesmata (Millard, 2001; but see e.g., Vokes, 1980, for a slightly different taxonomy). In more detail, there is a general agreement that Protobranchia is the first emerging lineage of Bivalvia. All feasible relationships among Protobranchia superfamilies (Solemyoidea, Nuculoidea and Nuculanoidea) have been proposed on morphological approaches (Purchon, 1987b; Waller, 1990; Morton, 1996; Salvini-Plawen and Steiner, 1996; Cope, 1997; Waller, 1998), albeit some recent molecular findings eventually led to reject the monophyly of the whole subclass: while Solemyoidea and Nuculoidea do maintain their basal position, thus representing Protobranchia *sensu stricto*, Nuculanoidea are better considered closer to Pteriomorphia, placed in their own order Nuculanoida (Giribet and Wheeler, 2002; Giribet and Distel, 2003; Kappner and Bieler, 2006).

The second subclass, Palaeoheterodonta (freshwater mussels), has been considered either among the most basal (Cope, 1996) or the most derived groups (Morton, 1996). Recent molecular analyses confirm its monophyly (Giribet and Wheeler, 2002) and tend to support it as basal to other Autolamellibranchiata bivalves (Graf and Ó Foighil, 2000; Giribet and Distel, 2003).

* Corresponding author. Fax: +39 051 20 94 173.
E-mail addresses: federico.plazzi@unibo.it (F. Plazzi), marco.passamonti@unibo.it (M. Passamonti).

1055-7903/$ - see front matter © 2010 Published by Elsevier Inc.
doi:10.1016/j.ympev.2010.08.032

Mussels, scallops, oysters and arks are representatives of the species-rich subclass Pteriomorphia. In literature, this subclass has been resolved as a clade within all Eulamellibranchiata (Purchon, 1987b), as a sister group of Trigonioidea (Salvini-Plawen and Steiner, 1996), of Heterodonta (Cope, 1997), of (Heterodonta + Palaeoheterodonta) (Waller, 1990, 1998), or as a paraphyletic group to Palaeoheterodonta (Morton, 1996). Moreover, some authors hypothesize its polyphyly (Carter, 1990; Starobogatov, 1992), while others claimed that a general agreement on Pteriomorphia monophyly is emerging from molecular studies (Giribet and Distel, 2003). Such an evident lack of agreement appears to be largely due to an ancient polytomy often recovered for this group, especially in molecular analyses, which is probably the result of a rapid radiation event in its early evolution (Campbell, 2000; Steiner and Hammer, 2000; Matsumoto, 2003).

Heterodonta is the widest and most biodiversity-rich subclass, including some economically important bivalves (f.i., venerid clams). This subclass has been proposed as monophyletic (Purchon, 1987b; Carter, 1990; Starobogatov, 1992; Cope, 1996, 1997; Waller, 1990, 1998), or paraphyletic (Morton, 1996; Salvini-Plawen and Steiner, 1996), but it seems there is a growing agreement on its monophyly. At a lower taxonomic level, doubts on the taxonomic validity of its major orders, such as Myoida and Veneroida, are fully legitimate, and, in many cases, recent molecular analyses led to throughout taxonomic revisions (Maruyama et al., 1998; Williams et al., 2004; Taylor et al., 2007a).

Little agreement has been reached in literature on Anomalodesmata: this subclass shows a highly derived body plan, as they are septibranchiate and some of them are also carnivore, features that possibly evolved many times (Dreyer et al., 2003). Anomalodesmata were considered as sister group of Myoida (Morton, 1996; Salvini-Plawen and Steiner, 1996), Mytiloidea (Carter, 1990), Palaeoheterodonta (Cope, 1997), or Heterodonta (Waller, 1990, 1998); alternatively, Purchon (1987b) states that they represent a monophyletic clade nested in a wide polytomy of all Bivalvia. Anomalodesmata were also considered as basal to all Autolamellibranchiata (e.g., Starobogatov, 1992). Whereas the monophyletic status of Anomalodesmata seems unquestionable on molecular data (Dreyer et al., 2003), some authors proposed that this clade should be nested within heterodonts (Giribet and Wheeler, 2002; Giribet and Distel, 2003; Bieler and Mikkelsen, 2006; Harper et al., 2006).

Molecular analyses gave clearer results at lower taxonomic levels, so that this kind of literature is more abundant: for instance, key papers have been published on Ostreidae (Littlewood, 1994; Jozefowicz and Ó Foighil, 1998; Ó Foighil and Taylor, 2000; Kirkendale et al., 2004; Shilts et al., 2007), Pectinidae (Puslednik and Serb, 2008), Cardiidae (Maruyama et al., 1998; Schneider and Ó Foighil, 1999) or former Lucinoidea group (Williams et al., 2004; Taylor et al., 2007b).

In this study, we especially address bivalves' ancient phylogenetic events by using mitochondrial molecular markers, namely the *12s*, *16s*, cytochrome b (*cytb*) and cytochrome oxidase subunit 1 (*cox1*) genes. We chose mitochondrial markers since they have the great advantage to avoid problems related to multiple-copy nuclear genes (i.e. concerted evolution, Plohl et al., 2008), they have been proved to be useful at various phylogenetic levels, and, although this is not always true for bivalves, they largely experience Strict Maternal Inheritance (SMI; Gillham, 1994; Birky, 2001).

Actually, some bivalve species show an unusual mtDNA inheritance known as Doubly Uniparental Inheritance (DUI; see Breton et al., 2007; Passamonti and Ghiselli, 2009; for reviews): DUI species do have two mitochondrial DNAs, one called F as it is transmitted through eggs, the other called M, transmitted through sperm and found almost only in males' gonads. The F mtDNA is passed from mothers to complete offspring, whereas the M mtDNA is passed from fathers to sons only. Obviously, DUI sex-linked mtDNAs may result in incorrect clustering, so their possible presence must be properly taken into account. DUI has a scattered occurrence among bivalves and, until today, it has been found in species from seven families of three subclasses: palaeoheterodonts (Unionidae, Hyriidae, and Margaritiferidae), pteriomorphians (Mytilidae), and heterodonts (Donacidae, Solenidae, and Veneridae) (Theologidis et al., 2008; Fig. 2 and reference therein). In some cases, co-specific F and M mtDNAs do cluster together, and this will not significantly affect phylogeny at the level of this study: this happens, among others, for *Donax trunculus* (Theologidis et al., 2008) and *Venerupis philippinarum* (Passamonti et al., 2003). In others cases, however, F and M mtDNAs cluster separately, and this might possibly result in an incorrect topology: f.i. this happens for the family of Unionidae and for *Mytilus* (Theologidis et al., 2008). All that considered, bivalves' mtDNA sequences should not be compared unless they are surely homolog, and the possible presence of two organelle genomes is an issue to be carefully evaluated (see Section 2.1, for further details). On the other hand, we still decided to avoid nuclear markers for two main reasons: (i) largely used nuclear genes, like 18S rDNA, are not single-copy genes and have been seriously questioned for inferences about bivalve evolution (Littlewood, 1994; Steiner and Müller, 1996; Winnepenninckx et al., 1996; Adamkewicz et al., 1997; Steiner, 1999; Distel, 2000; Passamaneck et al., 2004); (ii) data on putative single-copy nuclear markers, like β-actin or *hsp70*, lack for the class, essentially because primers often fail to amplify target sequences in Bivalvia (pers. obs.).

2. Materials and methods

2.1. Specimens' collection and DNA extraction

Species name and sampling locality are given in Table 1. Animals were either frozen or ethanol-preserved until extraction. Total genomic DNA was extracted by DNeasy® Blood and Tissue Kit (Qiagen, Valencia, CA, USA), following manufacturer's instructions. Samples were incubated overnight at 56 °C to improve tissues' lysis. Total genomic DNA was stored at −20 °C in 200 μL AE Buffer, provided with the kit.

DUI species are still being discovered among bivalves; nevertheless, as mentioned, a phylogenetic analysis needs comparisons between orthologous sequences, and M- or F-type genes under DUI are not. On the other hand, F-type mtDNA for DUI species and mtDNA of non-DUI species are orthologous sequences. As M-type is present mainly in sperm, we avoided sexually-mature individuals and, when possible (i.e., when the specimen was not too tiny), we did not extract DNA from gonads. If possible, DNA was obtained from foot muscle, which, among somatic tissues, carries very little M-type mtDNA in DUI species (Garrido-Ramos et al., 1998), thus reducing the possibility of spurious amplifications of the M genome. Moreover, when downloading sequences from GenBank, we paid attention in retrieving female specimen data only, whenever this information was available.

2.2. PCR Amplification, cloning, and sequencing

PCR amplifications were carried out in a 50 μL volume, as follows: 5 or 10 μL reaction buffer, 150 nmol $MgCl_2$, 10 nmol each dNTP, 25 pmol each primer, 1–5 μL genomic DNA, 1.25 units of DNA Polymerase (Invitrogen, Carlsbad, CA, USA or ProMega, Madison, WI, USA), water up to 50 μL. PCR conditions and cycles are listed in Appendix A1; primers used for this study are listed in Appendix A2. PCR results were visualized onto a 1–2% electrophoresis agarose gel stained with ethidium bromide and purified through Wizard® SV

Table 1
Specimens used for this study, with sampling locality and taxonomy following Millard (2001). Only species whose sequences were obtained in our laboratory are shown.

Subclass	Order	Suborder	Superfamily	Family	Subfamily	Species	Provenience
Anomalodesmata	Pholadomyoida	Cuspidariina		Cuspidariidae		*Cuspidaria rostrata*	Malta
		Pholadomyina	Pandoroidea	Pandoridae		*Pandora pinna*	Trieste, Italy
				Thraciidae		*Thracia distorta*	Secche di Tor Paterno, Italy
Heterodonta	Chamida		Astartoidea	Astartidae	Astartinae	*Astarte* cfr. *castanea*	Woods Hole, MA, USA
			Mactroidea	Mactridae	Mactrinae	*Mactra corallina*	Cesenatico, Italy
						Mactra lignaria	Cesenatico, Italy
			Tellinoidea	Pharidae	Cultellinae	*Ensis directus*	Woods Hole, MA, USA
			Tridacnoidea	Tridacnidae		*Tridacna derasa*	Commercially purchased
						Tridacna squamosa	Commercially purchased
	Myida	Myina	Myoidea	Myidae	Myinae	*Mya arenaria*	Woods Hole, MA, USA
	Veneroida		Carditoidea	Carditidae	Carditinae	*Cardita variegata*	Nosi Bè, Madagascar
			Veneroidea	Veneridae	Gafrarinae	*Gafrarium alfredense*	Nosi Bè, Madagascar
					Gemminae	*Gemma gemma*	Woods Hole, MA, USA
Palaeheterodonta	Unionida		Unionoidea	Unionidae	Anodontinae	*Anodonta woodiana*	Po River delta, Italy
Protobranchia	Nuculoida		Nuculanoidea	Nuculanidae	Nuculaninae	*Nuculana commutata*	Malta
			Nuculoidea	Nuculidae		*Nucula nucleus*	Goro, Italy
Pteriomorphia	Arcida	Arcina	Arcoidea	Arcidae	Anadarinae	*Anadara ovalis*	Woods Hole, MA, USA
					Arcinae	*Barbatia parva*	Nosi Bè, Madagascar
						Barbatia reeveana	Galápagos Islands, Ecuador
						Barbatia cfr. *setigera*	Nosi Bè, Madagascar
	Limida		Limoidea	Limidae		*Lima pacifica galapagensis*	Galápagos Islands, Ecuador
	Ostreoida	Ostreina	Ostreoidea	Ostreidae	Pycnodonteinae	*Hyotissa hyotis*	Nosi Bè, Madagascar
		Pectinina	Anomioidea	Anomiidae		*Anomia sp.*	Woods Hole, MA, USA
			Pectinoidea	Pectinidae	Chlamydinae	*Argopecten irradians*	Woods Hole, MA, USA
						Chlamys livida	Nosi Bè, Madagascar
						Chlamys multistriata	Krk, Croatia
					Pectininae	*Pecten jacobaeus*	Montecristo Island, Italy
	Pteriida	Pinnina	Pinnoidea	Pinnidae		*Pinna muricata*	Nosi Bè, Madagascar

Gel and PCR Clean-Up System (ProMega, Madison, WI, USA), following manufacturer's instructions.

Sometimes, amplicons were not suitable for direct sequencing; thus, PCR products were inserted into a pGEM®-T Easy Vector (ProMega, Madison, WI, USA) and transformed into Max Efficiency® DH5α™ Competent Cells (Invitrogen, Carlsbad, CA, USA). Positive clones were PCR-screened with M13 primers (see Appendix A2) and visualized onto a 1–2% electrophoresis agarose gel. However, as far as possible, we only cloned whenever it was strictly necessary; actually, as in DUI species some "leakage" of M mitotype may occur in somatic tissues of males, sensible cloning procedures could sometimes amplify such rare variants. Suitable amplicons and amplified clones were sequenced through either GeneLab (ENEA-Casaccia, Rome, Italy) or Macrogen (World Meridian Center, Seoul, South Korea) facilities.

2.3. Sequence alignment

Electropherograms were visualized by Sequence Navigator (Parker, 1997) and MEGA4 (Tamura et al., 2007) softwares. Sequences were compared to those available in GenBank through BLAST 2.2.19+ search tool (Altschul et al., 1997). Four outgroups were used for this study: the polyplacophoran *Katharina tunicata*, the scaphopod *Graptacme eborea* and two gastropods, *Haliotis rubra* and *Thais clavigera*. Appendix A3 lists all DNA sequences used for this study, along with their GenBank accession number.

Alignments were edited by MEGA4 and a concatenated data set was produced; whenever only three sequences out of four were known, the fourth was coded as a stretch of missing data, since the presence of missing data does not lead to an incorrect phylogeny by itself, given a correct phylogenetic approach (as long as sufficient data are available for the analysis; see Hartmann and Vision, 2008; and reference therein). In other cases, there were not sufficient published sequences for a given species to be included in our concatenated alignment; nevertheless, we could add the genus itself by concatenating DNA sequences from different co-generic species, as this approach was already taken in other phylogenetic studies (see, f.i., Li et al., 2009). This was the case for *Donax*, *Solemya*, *Spisula*, and *Spondylus* (see Appendix A3 for details). Given the broad range of the analysis, which targets whole class phylogeny above the genus level, we do not think that such an approximation significantly biased our results. In any case, phylogenetic positions of such genera were taken with extreme care.

Sequences were aligned with ClustalW (Thompson et al., 1994) implemented in MEGA4. Gap opening and extension costs were set to 50/10 and 20/4 for protein- and ribosomal-coding genes, respectively. Because of the high evolutionary distance of the analyzed taxa, sequences showed high variability, and the problem was especially evident for ribosomal genes, where different selective pressures are active on different regions. These genes showed a lot of indels, which were strikingly unstable across alignment parameters; thus, we could not resolve alignment ambiguities in an objective way. The method proposed by Lutzoni et al. (2000), though very appealing, is problematic for big data sets with high variability, as shown by the authors themselves. On the other side, likelihood analyses are also problematic with the fixed character state method proposed by Wheeler (1999). Elision, as introduced by Wheeler et al. (1995), is a possibility that does not involve particular methods of phylogenetic analyses, but only a "grand alignment". However, variability in our ribosomal data set was so high that alignments with different parameters were almost completely different; thus, elision generated only more phylogenetic noise, whereas the original method by Gatesy et al. (1993) was not conceivable because alignment-invariant positions were less than twenty. All that considered, we preferred to use a user-assisted standard alignment method (i.e., ClustalW) since we think this is yet the best alignment strategy for such a complex dataset. Alignment was also visually inspected searching for misaligned sites and ambiguities, and where manual optimization was not possible, alignment-ambiguous regions were excluded from the analysis. Indels were treated as a whole and converted to presence/absence data to avoid many theoretical concerns on alignments (simple indel coding; see Simmons and Ochoterena, 2000, for more details). In fact, ambiguities in alignments are mainly due to indel

insertions; therefore, this technique also eliminates a large part of phylogenetic noise. We then coded indels following the rules given by Simmons and Ochoterena (2000), as implemented by the software GapCoder (Young and Healy, 2003), which considers each indel as a whole, and codes it at the end of the nucleotide matrix as presence/absence (i.e. 1/0). Possibly, a longer indel may completely overlap another across two sequences; in such cases, it is impossible to decide whether the shorter indel is present or not in the sequence presenting the longer one. Therefore, the shorter indel is coded among missing data in that sequence. Data set was then analyzed treating gaps as missing data and presence/absence data of indel events as normal binary data.

2.4. Phylogenetic analyses

A preliminary test was made on saturation: transition and transversion uncorrected *p*-distances were plotted on global pairwise *p*-distances, as computed with PAUP* 4.0b10 (pairwise deletion of gaps; Swofford, 1999); the test was repeated on third positions only for protein-coding genes. Linear regression and its significance were tested with PaSt 1.90 (Hammer et al., 2001).

Partitioning schemes used in this study are 10, based on 26 different partitions (Supplementary Materials Fig. 1), although they are not all the conceivable ones; we describe our 10 partitioning patterns in Table 2. The Bayesian Information Criterion (BIC) implemented in ModelTest 3.7 (Posada and Crandall, 1998) was used to select the best-fitting models; the graphical interface provided by MrMTgui was used (Nuin, 2008). As MrBayes 3.1.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) currently implements only models with 1, 2 or 6 substitutions, a GTR + I + Γ model (Tavaré, 1986) was chosen for all partitions. ModelTest rejected the presence of a significant proportion of invariable sites in three cases only; GTR + Γ were selected for *cox1* third positions and for *cytb* second and third positions.

Maximum Likelihood was carried out with PAUP* software at the University of Oslo BioPortal (<http://www.bioportal.uio.no>). Gap characters were treated as missing data and the concatenated alignment was not partitioned. Nucleotides frequencies, substitution rates, gamma shape parameter and proportion of invariable sites were set according to ModelTest results on global alignment. Outgroups were set to be paraphyletic to the monophyletic ingroup. Bootstrap with 100 replicates, using full heuristic ML searches with stepwise additions and TBR branch swapping, was performed to assess nodal support.

Machine time is a key issue in Maximum Likelihood, and, unfortunately, a parallel version of PAUP* has not been published yet. To speed up the process, we used a slightly restricted dataset and set up the analysis to simulate a parallel computation, therefore taking higher advantage of the large computational power of the BioPortal. We run 10 independent bootstrap resamplings with 10 replicates each, starting with different random seeds generated by Microsoft Excel® 2007 following PAUP* recommendations. Trees found in each run were then merged and final consensus was computed with PAUP*. A comparative analysis on a smaller but still representative dataset showed, as expected, that this strategy does not affect the topology of the tree, nor significantly changes bootstrap values (data not shown).

Although less intuitive than in the case of parsimony (Baker and DeSalle, 1997), a Partitioned Likelihood Support (PLS) can be computed for likelihood analyses (Lee and Hugall, 2003). We chose this kind of analysis because other methods (Templeton, 1983; Larson, 1994; Farris et al., 1995a, 1995b) measure overall levels of agreement between partitions in the data set, but they cannot show which parts of a tree are in conflict among partitions (Wiens, 1998; Lambkin et al., 2002). A positive PLS indicates that a partition supports a given clade, and a negative PLS indicates that the partition contradicts the clade itself. Parametric bootstrapping (Huelsenbeck et al., 1996a; Huelsenbeck et al., 1996b) and Shimodaira–Hasegawa test (Shimodaira and Hasegawa, 1999) can assess the statistical significance of PLS results (Goldman et al., 2000; Lee and Hugall, 2003; and reference therein). However, PLS analyses are currently difficult because no widely available phylogenetic software implement such an algorithm. Therefore, Partitioned Likelihood Support (PLS) was evaluated following the manual procedure described in Lee and Hugall (2003). TreeRot 3.0 (Sorenson and Franzosa, 2007) was used to produce PAUP* command file, whereas individual-site log-likelihood scores were analyzed by Microsoft Excel® 2007. Shimodaira–Hasegawa test was employed to assess confidence in PLS, following Shimodaira and Hasegawa (1999). VBA macros implemented in Microsoft Excel® 2007 to perform PLS and Shimodaira–Hasegawa analyses are available from F. P.

MrBayes 3.1.2 software was used for Bayesian analyses, which were carried out at the BioPortal (see above). We performed a Bayesian analysis for each partitioning scheme. Except as stated elsewhere, two MC^3 algorithm runs with four chains were run for 10,000,000 generations; convergence was estimated through PSRF (Gelman and Rubin, 1992) and by plotting standard deviation of average split frequencies sampled every 1000 generations. The four outgroups were constrained, trees found at convergence were retained after the burnin, and a majority-rule consensus tree was computed with the command **sumt**. Via the command **sump printtofile = yes** we could obtain the harmonic mean of the Estimated Marginal Likelihood (EML). EML was used to address model selection and partition choice.

Since there is no obvious way to define partitions in ribosomal-encoding genes and secondary structure-based alignments did not result in correct phylogenetic trees (data not shown; see also Steiner and Hammer, 2000), we first decided to test data partitioning schemes on protein-coding genes only. Therefore, after a global analysis merging all markers within the same set, we tested six different partitioning schemes for protein-coding genes, taking

Table 2
Partitioning schemes. See Supplementary Materials Fig. 1 for details on partitions.

Partitioning scheme	Number of partitions	Partitions (see fig. 1)
t01	2	all, all_indel
t02[a]	4	rib, rib_indel, prot, prot_indel
t03	5	rib, rib_indel, prot_12, prot_3, prot_indel
t04	6	rib, rib_indel, prot_1, prot_2, prot_3, prot_indel
t05	6	rib, rib_indel, cox1, cox1_indel, cytb, cytb_indel
t06	8	rib, rib_indel, cox1_12, cox1_3, cox1_indel, cytb_12, cytb_3, cytb_indel
t07	10	rib, rib_indel, cox1_1, cox1_2, cox1_3, cox1_indel, cytb, cytb_1, cytb_2, cytb_3, cytb_indel
t08	8	12s, 12s_indel, 16s, 16s_indel, prot_1, prot_2, prot_3, prot_indel
t09	12	12s, 12s_indel, 16s, 16s_indel, cox1_1, cox1_2, cox1_3, cox1_indel, cytb_1, cytb_2, cytb_3, cytb_indel
t10	4	cox1 (amminoacids), cox1_indel, cytb (amminoacids), cytb_indel

[a] *tNy98* and *tM3* were also based on this partitioning scheme.

ribosomal ones together (Table 2; *t02–t07*). As *t04* and *t07* were selected as the most suitable ones (see Section 3.5), we designed two more schemes splitting *12s* and *16s* based on these datasets only (Table 2; *t08–t09*). Finally, we tested some strategies to further remove phylogenetic noise: we first constructed an amminoacid dataset (Table 2; *t10*; we were forced to completely remove ribosomal genes, as MC^3 runs could not converge in this case). However, the use of amminoacids is not directly comparable with other datasets by AIC and BF, because it not only implies a different model, but also different starting data: as a consequence, we implemented the codon model (Goldman and Yang, 1994; Muse and Gaut, 1994) on the *prot* partition. This allowed us to start from an identical dataset, which makes results statistically comparable. As *t04* scheme turned out to be essentially comparable with *t09* (see Section 3.5), we did not implement codon model also on separate *cox1* and *cytb* genes, because codon model is computationally extremely demanding. Two separate analyses were performed under such a codon model: in both cases, metazoan mitochondrial genetic code table was used; in one case Ny98 model was enforced (*tNy98*; Nielsen and Yang, 1998), whereas in the other case M3 model was used (*tM3*). Only one run of 5000,000 generations was performed for codon models, sampling a tree every 125. Dealing with one-run analyses, codon models trees were also analytically tested for convergence via AWTY analyses (<http://king2.scs.fsu.edu/CEBProjects/awty/awty_start.php>; Nylander et al., 2008). Moreover, our analysis on codon models allowed us to test for positive selection on protein-coding genes (see Ballard and Whitlock, 2004): MrBayes estimates the ratio of the non-synonymous to the synonymous substitution rate (ω) and implements models to accommodate variation of ω across sites using three discrete categories (Ronquist et al., 2005).

Finally, to test for the best partitioning scheme and evolutionary model, we applied Akaike Information Criterion (AIC; Akaike, 1973) and Bayes Factors (BF; Kass and Raftery, 1995). AIC was calculated, following Huelsenbeck et al. (2004), Posada and Buckley (2004), and Strugnell et al. (2005), as

$$AIC = -2EML + 2K$$

The number of free parameters K was computed taking into account branch number, character (nucleotide, presence/absence of an indel, amminoacid, or codon and codon-related parameters) frequencies, substitution rates, gamma shape parameter and proportion of invariable sites for each partition.

Bayes Factors were calculated, following Brandley et al. (2005), as

$$B_{ij} = \frac{EML_i}{EML_j}$$

and, doubling and turning to natural logarithms

$$2\ln B_{ij} = 2(\ln EML_i - \ln EML_j)$$

where B_{ij} is the Bayes Factor measuring the strength of the ith hypothesis on the jth hypothesis. Bayes Factors were interpreted according to Kass and Raftery (1995) and Brandley et al. (2005).

All trees were graphically edited by PhyloWidget (Jordan and Piel, 2008) and Dendroscope (Huson et al., 2007) softwares. Published Maximum Likelihood and Bayesian trees, along with source data matrices, were deposited in TreeBASE under SN4787 and SN4789 Submission ID Numbers, respectively.

2.5. Taxon sampling

Taxon sampling is a crucial step in any phylogenetic analysis, and this is certainly true for bivalves (Giribet and Carranza, 1999; Puslednik and Serb, 2008). Actually, many authors claim for a bias in taxon sampling to explain some unexpected or unlikely results (Adamkewicz et al., 1997; Canapa et al., 1999; Campbell, 2000; Kappner and Bieler, 2006). As we want to find the best performing methodological pipeline for reconstructing bivalve phylogeny, we assessed taxon sampling following rigorous criteria, in order to avoid misleading results due to incorrect taxon choice. We approached this with both *a priori* and *a posteriori* perspectives, following two different (and complementary) rationales.

Quite often, taxa that are included in a phylogenetic analysis are not chosen following a formal criterion of representativeness: they are rather selected on accessibility and/or analyzer's personal choice. To avoid this, we developed a method to quantify sample representativeness with respect to the whole class. The method is based on Average Taxonomic Distinctness (AvTD) of Clarke and Warwick (1998). The mathematics of this method has been proposed in a different paper (Plazzi et al., 2010), but here we would like to mention the rationale behind it: estimating *a priori* the *phylogenetic representativeness* of a sample is not conceptually different from estimating its *taxonomic representativeness*, i.e. testing whether our taxon sampling is representative of a given master taxonomic list, which may eventually be retrieved from bibliography. This approach does not require any specific knowledge, other than the established taxonomy of the sampled taxa; neither sequence data, nor any kind of measure are used here, which means the AvTD approach comes *before* seeing the data. Our source of reference taxonomy (master list) was obtained from Millard (2001). The AvTD was then computed for our sample and confidence limits were computed on 1000 random resamplings of the same size from bivalve master list. If the taxon sample value is above the 95% lower confidence limit, then we can say that our dataset is representative of the whole group. We developed a software to compute this, which is available for download at <www.mozoolab.net>.

On the other hand, *after* seeing the data, we were interested in answering whether they were sufficient or not to accurately estimate phylogeny. For this purpose, we used the method proposed by Sullivan et al. (1999). The starting point is the tree obtained as the result of our analysis, given the correct model choice (see below). Several subtrees are obtained by pruning it without affecting branch lengths; each parameter is then estimated again from each subtree under the same model: if estimates, as size increases, converge to the values computed from the complete tree, then taxon sampling is sufficiently large to unveil optimal values of molecular parameters, such as evolutionary rates, proportion of invariable sites, and so on (Townsend, 2007). At first, we checked whether MC^3 Bayesian estimates of best model were comparable to Maximum Likelihood ones computed through ModelTest. We took into consideration all 6 mutations rates and, where present, nucleotide frequencies, invariable sites proportion and gamma-shaping parameter (which are not used into M3 codon model). In most cases (see Supplementary Materials Table 1) the Maximum Likelihood estimate fell within the 95% confidence interval as computed following Bayesian Analysis and, if not, the difference was always (except in one case) of 10^{-2} or less order of magnitude. Therefore, we used Bayesian estimates of mean and confidence interval limits instead of bootstrapping Maximum Likelihood, as in the original method of Sullivan et al. (1999). Fifty subtrees were manually generated from best tree by pruning a number of branches ranging from 1 to 50. Following Authors' suggestions, we used different pruning strategies: in some cases, we left only species very close in the original tree, whereas in others we left species encompassing the whole biodiversity of the class (Appendix A4). Model parameters were then estimated from each subtree for each partition (*rib* and *prot*) using original sequence data and the best model chosen by ModelTest as above. The paupblock of ModelTest was used into PAUP* to implement such specific Maximum Likelihood analyses for each partition, model, and subtree.

2.6. Dating

The r8s 1.71 (Sanderson, 2003) software was used to date the best tree we obtained. Fossil collections of bivalves are very abundant, so we could test several calibration points in our tree, but in all cases the origin of Bivalvia was constrained between 530 and 520 million years ago (Mya; Brasier and Hewitt, 1978), and no other deep node was used for calibration, as we were interested in molecular dating of ancient splits. Data from several taxa were downloaded from the Paleobiology Database on 4 November, 2009, using group names given in Table 3 and leaving all parameters as default. Some nodes were fixed or constrained to the given age, whereas others were left free. After the analysis, we checked whether the software was able to predict correct ages or not, i.e. whether the calibration set was reliable. The tree was re-rooted with the sole *Katharina tunicata*; for this reason, two nodes "*Katharina tunicata*" and "other outgroups" are given in Table 3. Rates and times were estimated following both PL and NPRS methods, which yielded very similar results. In both cases we implemented the Powell's algorithm. Several rounds of fossil-based cross-validation analysis were used to determine the best-performing smoothing value for PL method and the penalty function was set to log. Four perturbations of the solutions and five multiple starts were invoked to optimize searching in both cases. Solutions were checked through the **checkGradient** command. NPRS method was also used to test variability among results. 150 bootstrap replicates of original dataset were generated by the SEQBOOT program in PHYLIP (Felsenstein, 1993) and branch lengths were computed with PAUP* through r8s-bootkit scripts of Torsten Eriksson (2007). A complete NPRS analysis was performed on each bootstrap replicate tree and results were finally profiled across all replicates through the r8s command **profile**.

3. Results

3.1. Obtained sequences

Mitochondrial sequences from partial ribosomal small (*12s*) and large (*16s*) subunit, cytochrome b (*cytb*) and cytochrome oxidase

Table 3
r8s datation of *tM3* tree. If a fossil datation is shown, the clade was used for calibrating the tree using Paleobiology Database data; in bold are shown the eight calibrations point of the best-performing set, whereas the others were used as controls. Constraints enforced are shown in the fourth and fifth column; if they are identical, that node was fixed. Ages are in millions of years (Myr); rates are in substitutions per year per site and refer to the branch leading to a given node. PL, Penalized Likelihood; NPRS, Non Parametric Rate Smoothing; StDev, Standard Deviation.

	Fossil datation	Reference[a]	Constraints		PL		NPRS			
			Min	Max	Age	Local rate	Age	Local rate	Mean	StDev
Katharina tunicata					627.58		625.44			
Other outgroups					561.45	1.65E−03	560.05	1.67E−03	533.95	2.67
Bivalvia	**530.0–520.0**	**5**	**520.00**	**530.00**	**529.99**	**3.46E−03**	**530.00**	**3.63E−03**	**530.00**	**0.00**
Autolamellibranchiata					520.32	2.01E−02	520.31	2.01E−02	517.04	1.70
Pteriomorphia + Heterodonta					513.59	2.26E−02	513.59	2.26E−02	508.51	1.74
Pteriomorphia					505.74	1.81E−02	505.82	1.83E−02	501.13	2.29
Heterodonta					497.83	1.51E−02	498.20	1.55E−02	490.24	3.11
Traditional Pteriomorphia					496.63	1.26E−02	496.13	1.19E−02	488.88	2.38
Hiatella + Cardiidae					481.34	1.10E−02	481.61	1.09E−02	476.05	3.65
Limidae + Pectinina					474.51	1.71E−02	474.82	1.78E−02	468.49	3.49
Veneroida sensu lato					471.38	3.80E−03	471.87	3.82E−03	471.22	6.63
Anomioidea + Pectinoidea					464.44	1.19E−02	464.92	1.21E−02	459.25	4.26
Protobranchia					454.28	1.34E−03	455.67	1.37E−03	482.02	14.61
Arcidae	**457.5–449.5**	**29**	**449.50**	**457.50**	**449.51**	**2.35E−02**	**449.50**	**2.38E−02**	**449.50**	**0.00**
Pectinoidea	428.2–426.2	21, 27, 30			431.77	1.27E−02	433.44	1.32E−02	417.82	4.20
Anomalodesmata					431.45	3.29E−03	434.04	3.40E−03	461.87	9.59
Cardiidae	**428.2–426.2**	**18**	**427.20**	**427.20**	**427.20**	**1.18E−02**	**427.20**	**1.18E−02**	**427.20**	**0.00**
Cuspidaria clade					418.58	4.87E−03	421.63	5.04E−03	477.22	9.28
Veneroida 2					407.08	3.58E−03	407.42	3.58E−03	410.56	9.26
Ostreoida + Pteriida					393.59	3.48E−03	395.13	3.55E−03	435.47	10.95
Pectinidae	**388.1–383.7**	**2, 6, 14, 22, 26**	**385.90**	**385.90**	**385.90**	**5.18E−03**	**385.90**	**5.00E−03**	**385.90**	**0.00**
Limidae	**376.1–360.7**	**1**	**360.70**	**376.10**	**360.74**	**4.66E−03**	**360.71**	**4.65E−03**	**370.13**	**6.31**
Veneridae	**360.7–345.3**	**19, 30**	**345.30**	**360.70**	**345.33**	**3.30E−03**	**345.31**	**3.28E−03**	**347.28**	**4.57**
Pectininae					324.88	1.57E−03	327.18	1.63E−03	342.84	7.76
Unionidae	245.0–228.0	8			293.93	3.68E−03	298.00	3.74E−03	347.74	20.25
Gafrarium + Gemma					282.57	2.24E−03	283.03	2.25E−03	280.55	22.38
Ostreoida	251.0–249.7	28			264.75	3.00E−03	266.21	3.00E−03	333.04	16.09
Mactrinae	196.5–189.6	25			243.80	2.27E−03	244.76	2.28E−03	261.16	21.60
Argopecten + Pecten					220.05	1.22E−03	222.43	1.22E−03	256.84	14.94
Unioninae	**228.0–216.5**	**9, 13, 16, 20, 23**	**216.50**	**228.00**	**216.53**	**1.71E−03**	**216.51**	**1.62E−03**	**227.86**	**0.93**
Chlamys livida + Mimachlamys					190.34	1.24E−03	194.24	1.27E−03	336.20	8.12
Ensis + Sinonovacula					189.33	1.16E−03	189.83	1.16E−03	305.30	18.57
Astarte + Cardita					188.86	3.26E−03	191.12	3.25E−03	274.37	23.58
Dreissena + Mya					185.03	2.62E−03	185.82	2.62E−03	224.89	19.55
Barbatia	**167.7–164.7**	**4, 10, 24**	**166.20**	**166.20**	**166.20**	**6.93E−04**	**166.20**	**6.93E−04**	**166.20**	**0.00**
Tridacna	23.0–16.0	17			147.15	1.26E−03	149.69	1.27E−03	383.21	11.43
Setigera + Reeveana					77.29	2.20E−03	75.19	2.15E−03	92.77	12.17
Crassostrea	145.5–130.0	15			63.17	3.08E−03	63.52	3.07E−03	92.38	10.04
Gigas + Hongkongensis					23.47	2.72E−03	23.65	2.71E−03	36.93	9.36
Mactra	196.5–189.6	25			21.63	1.50E−03	21.80	1.49E−03	31.48	6.91
Mytilus	418.7–418.1	3, 7, 11, 12			1.88	2.92E−03	1.77	2.92E−03	1.79	0.60

[a] References as follows: (1) Amler et al. (1990); (2) Baird and Brett (1983); (3) Berry and Boucot (1973); (4) Bigot (1935); (5) Brasier and Hewitt (1978); (6) Brett et al. (1991); (7) Cai et al. (1993); (8) Campbell et al. (2003); (9) Chatterjee (1986); (10) Cox (1965); (11) Dou and Sun (1983); (12) Dou and Sun (1985); (13) Elder (1987); (14) Grasso (1986); (15) Hayami (1975); (16) Heckert (2004); (17) Kemp (1976); (18) Kříž (1999); (19) Laudon (1931); (20) Lehman and Chatterjee (2005); (21) Manten (1971); (22) Mergl and Massa (1992); (23) Murry (1989); (24) Palmer (1979); (25) Poulton (1991); (26) Rode and Lieberman (2004); (27) Samtleben et al. (1996); (28) Spath (1930); (29) Suarez Soruco (1976); (30) Wagner (2008).

subunit I (*cox1*) were obtained; GenBank accession numbers are reported in Appendix A3. A total of 179 sequences from 57 bivalve species were used for this study: 80 sequences from 28 species were obtained in our laboratory, whereas the others were retrieved from GenBank (see Appendix A3 for details). Alignment was made by 55 taxa and 2501 sites, 592 of which, all within *12s* and *16s* genes, were excluded because they were alignment-ambiguous. After removal, 1623 sites were variable and 1480 were parsimony-informative. It is clearly impossible to show here a complete *p*-distance table, but the overall average value was 0.43 (computed by MEGA4, with pairwise deletion of gaps).

Quite interestingly, we found few anomalies in some of the sequences: for instance, a single-base deletion was present in *cytb* of *Hyotissa hyotis* and *Barbatia* cfr. *setigera* at position 2317 and 2450, respectively. This can suggest three possibilities: (i) we could have amplified a mitochondrial pseudogene (NUMT); (ii) we could have faced a real frameshift mutation, which may eventually end with a compensatory one-base insertion shortly downstream (not visible, since our sequence ends quite soon after deletion); (iii) an error in base calling was done by the sequencer. At present no NUMTs have been observed in bivalves (Bensasson et al., 2001; Zbawicka et al., 2007) and the remaining DNA sequences are perfectly aligned with the others, which is unusual for a NUMT; therefore, we think that the second or the third hypotheses are more sound. In all subsequent analyses, we inserted missing data both in nucleotide and in amminoacid alignments. Moreover, several stop codons were found in *Anomia sp.* sequences (within *cox1*, starting at position 1796 and 1913; within *cytb*, starting at 2154, 2226, 2370, 2472 and 2484). Again, we could have amplified two pseudogenes; however, all these stop codons are TAA and the alignment is otherwise good. A possible explanation is an exception to the mitochondrial code of this species, which surely demands further analysis, but this is beyond the scope of this paper. In any case, we kept both sequences and placed missing data in protein and codon model alignments in order to perform subsequent analyses. Of course, phylogenetic positions of all the abovementioned species have been considered with extreme care, taking into account their sequence anomalies.

3.2. Sequence analyses

No saturation signal was observed by plotting uncorrected p-distances as described above (see Supplementary Materials Fig. 2), since all linear interpolations were highly significant as computed with PaSt 1.90. Moreover, deleting third codon positions we obtained a completely unresolved Bayesian tree, confirming that these sites carry some phylogenetic signal (data not shown).

Selective pressures on protein-coding genes were tested through ω. In the Ny98 model (Nielsen and Yang, 1998), there are three classes with different potential ω values: $0 < \omega_1 < 1$, $\omega_2 = 1$, and $\omega_3 > 1$. The M3 model also has three classes of ω values, but these values are less constrained, in that they only have to be ordered $\omega_1 < \omega_2 < \omega_3$ (Ronquist et al., 2005). As M3 was chosen as the best model for our analysis (see below), we only considered M3 estimates about ω and its heterogeneity. Boundaries estimates for *tM3* are very far from one (Supplementary Materials Table 2) and more than 75% of codon sites fell into the first two categories. Moreover, all codon sites scored 0 as the probability of being positively selected. Therefore, we conclude that only a stabilizing pressure may be at work on these markers, which may enhance their phylogenetic relevance. This also allows to analyze protein-coding genes together.

3.3. Taxon sampling

Supplementary Materials Fig. 3 shows results from Average Taxonomic Distinctness test. Our sample plotted almost exactly on the mean of 1000 same-size random subsamples from the master list of bivalve genera, thus confirming that our sample is a statistically representative subsample of the bivalves' systematics.

Supplementary Materials Fig. 4 shows results from *a posteriori* testing of parameter accurateness. Analysis was carried out for all main parameters describing the models, but, for clarity, only gamma-shaping parameters (alpha) and invariable sites proportions (pinv) for *rib* partition are shown. In any case, all parameters behaved the same way: specifically, estimates became very close to "true" ones starting from subtrees made by 30–32 taxa. Therefore, at this size a dataset is informative about evolutionary estimates, given our approach. As we sampled nearly twice this size, this strengthens once again the representativeness of our taxon choice – this time from a molecular evolution point of view.

3.4. Maximum Likelihood

Maximum Likelihood analysis gave the tree depicted in Fig. 1. The method could not resolve completely the phylogeny: bivalves appear to be polyphyletic, as the group corresponding to Protobranchia (*Nucula* + *Solemya*) is clustered among non-bivalve species, although with low support (BP = 68). A first node (BP = 100) separates Palaeoheterodonta (*Inversidens* + *Lampsilis*) from the other groups. A second weak node (BP = 51) leads to two clades, one corresponding to Pteriomorphia + *Thracia* (BP = 68) and the other, more supported, to Heterodonta (BP = 83). A wide polytomy is evident among Pteriomorphia, with some supported groups in it, such as *Thracia*, *Mytilus*, Arcidae (all BP = 100), Limidae + Pectinina (BP = 87), and Pteriida + Ostreina (BP = 85). Heterodonta subclass is also not well resolved, with *Astarte* + *Cardita* (BP = 100) as sister group of a large polytomy (BP = 73) that includes *Donax*, *Ensis*, *Hiatella* + (*Acanthocardia* + *Tridacna*), and an heterogeneous group with Veneridae, *Spisula*, *Dreissena* and *Mya* (BP = 66).

PLS tests turned out to be largely significant (Supplementary Materials Fig. 5). High likelihood support values were always connected with highly supported nodes, whereas the opposite is not always true (see node 11). High positive PLS values are generally showed by the *cytb* partition; good values can also be noted for *cox1* and *16s* genes, even if *16s* is sometimes notably against a given node (see nodes 23 and 24). *12s* has generally low PLS absolute values, with some notable exceptions (see nodes 15 and 16). Globally, deeper splits (see nodes 6, 13, 14, 22, 23, 24, 29) have a low likelihood support absolute value, and generally a low bootstrap score too.

3.5. Bayesian analyses

Table 4 shows results of model-decision statistical tests. Among classical 4by4 models (i.e., not codon models) AIC favored *t04* as best trade-off between partitions number and free parameters. However, if considered, *tM3* (a codon model) was clearly favored. As BF does not take into account the number of free parameters, *t04* is not clearly the best classical 4by4 model in this case. More complex models (with the notable exception of *t05*) turned out to be slightly favored: *t09*, the most complex model we implemented, has positive (albeit small) BF values against each simpler partition scheme. Again, when considered, *tM3* is straightforwardly the best model, with the highest BF scores in the matrix (see Table 4). It is notable that *tNy98*, even not the worst, has instead very low BF scores. Therefore, using *tM3* we obtained the best phylogenetic tree, which is shown in Fig. 2. In this tree, several clusters agreeing with the established taxonomy are present: the first corresponds to Protobranchia (*sensu* Giribet and Wheeler, 2002) and it is basal to all the remaining bivalves (Autolamellibranchiata *sensu* Bieler and Mikkelsen, 2006; PP = 1.00). A second group, which is basal to the rest of the tree, is composed by Palaeoheterodonta (PP = 1.00). Sister group to Palaeoheterodonta a major clade is found (PP = 1.00), in which three

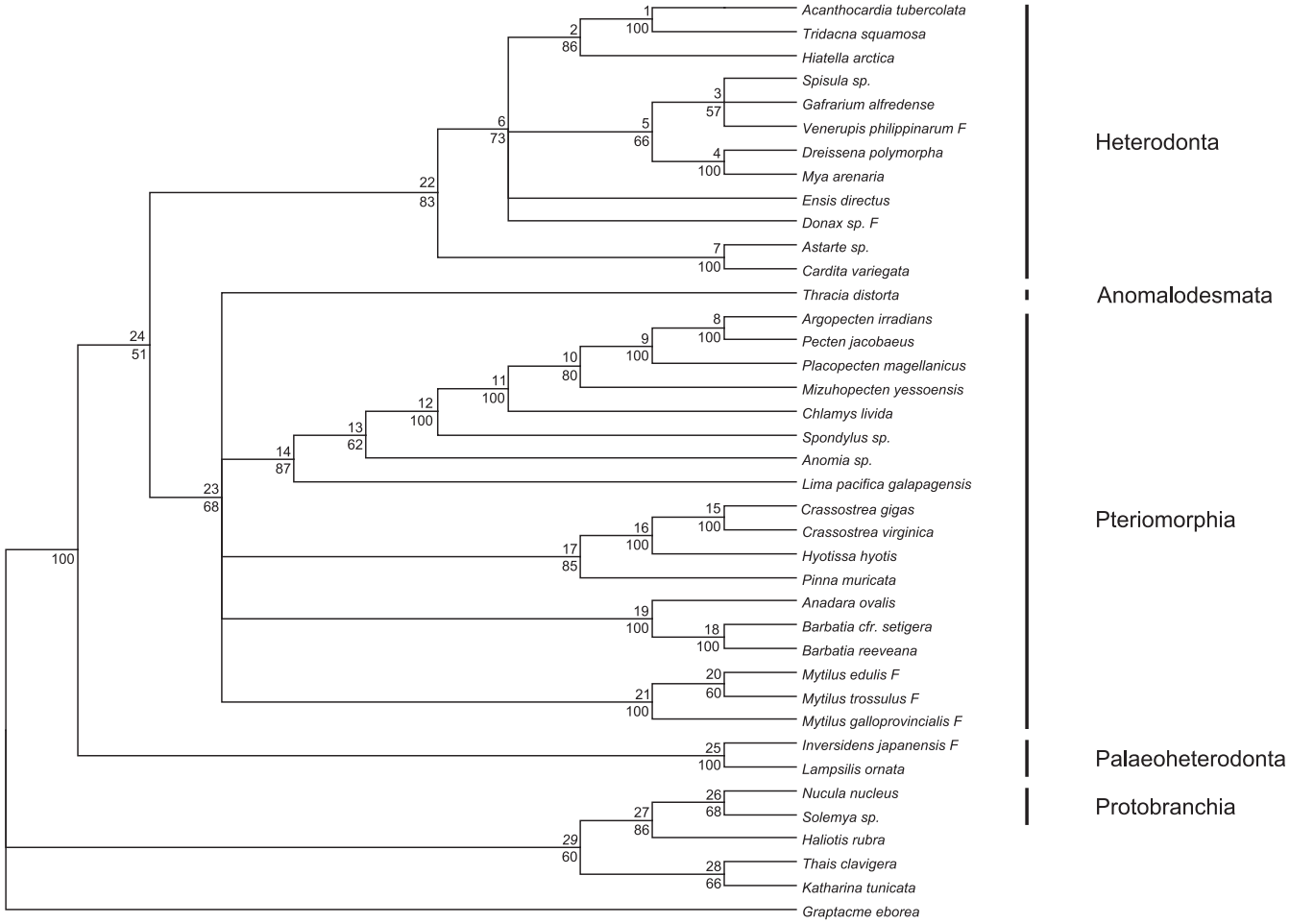


Fig. 1. Majority-rule consensus tree of 100 Maximum Likelihood bootstrap replicates: node have been numbered (above branches), and numbers below the nodes are bootstrap proportions.

Table 4
Results from Akaike Information Criterion (AIC) and Bayes Factors (BF) tests. EML, Estimated Marginal Likelihood; *p*, number of partitions in the partitioning scheme; FP, Free Parameters. Partitioning schemes as in Table 2.

Tree	EML	*p*	FP	AIC	*t02*	*t03*	*t04*	*t05*	*t06*	*t07*	*t08*	*t09*	*t10*	*tNy98*	*tM3*
t01	−64,914.04	2	225	130,278.08	479.76	1870.00	2203.28	494.92	1950.86	2290.48	2326.90	2424.26	N/A	884.14	3721.44
t02	−64,674.16	4	450	130,248.32		1390.24	1723.52	15.16	1471.10	1810.72	1847.14	1944.50	N/A	404.38	3241.68
t03	−63,979.04	5	567	129,092.08			333.28	−1375.08	80.86	420.48	456.90	554.26	N/A	−985.86	1851.44
t04	−63,812.40	6	684	128,992.80				−1708.36	−252.42	87.20	123.62	220.98	N/A	−1319.14	1518.16
t05	−64,666.58	6	675	130,683.16					1455.94	1795.56	1831.98	1929.34	N/A	389.22	3226.52
t06	−63,938.61	8	907	129,691.22						339.62	376.04	473.40	N/A	−1066.72	1770.58
t07	−63,768.80	10	1140	129,817.60							36.42	133.78	N/A	−1406.34	1430.96
t08	−63,750.59	8	909	129,319.18								97.36	N/A	−1442.76	1394.54
t09	−63,701.91	12	1365	130,133.82									N/A	−1540.12	1297.18
t10	−13,725.38	4	450	28,350.76										N/A	N/A
tNy98	−64,471.97	4	512	129,967.94											2837.30
tM3	−63,053.32	4	513	127,132.64											

main groups do separate. Heterodonta constitute a cluster (PP = 1.00), with two branches: *Hiatella* + Cardiidae (PP = 1.00) and other heterodonts (PP = 0.98). Within them, only one node remains unresolved, leading to a Veneridae + Mactridae + (*Dreissena* + *Mya*) polytomy. Another cluster (PP = 0.96) is made by *Pandora* + *Thracia*, as sister group of all Pteriomorphia + *Nuculana* (both PP = 1.00). A wide polytomy is evident within Pteriomorphia, with *Mytilus* species, Limidae + Pectinina, Pteriida + Ostreina, Arcidae and *Nuculana* itself as branches, all with PP = 1.00. Another cluster (PP = 1.00) is made by *Cuspidaria* + (*Astarte* + *Cardita*). All families have PP = 1.00: Cardiidae (genera *Acanthocardia* and *Tridacna;* see Section 4.2.4), Mactridae (genera *Mactra* and *Spisula*), Veneridae (genera *Gafrarium*, *Gemma* and *Venerupis*), Unionidae (genera *Hyriopsis*, *Inversidens*, *Anodonta* and *Lampsilis*), Arcidae (genera *Anadara* and *Barbatia*), Limidae (genera *Acesta* and *Lima*), Ostreidae (genera *Crassostrea* and *Hyotissa*) and Pectinidae (genera *Mizuhopecten*, *Chlamys*, *Mimachlamys*, *Argopecten*, *Pecten* and *Placopecten*).

3.6. Dating the tree

Results from r8s software are shown in Table 3. The relative ultrametric tree is shown in Fig. 3 along with the geological timescale. The best-performing smoothing value for PL analysis was set to 7.26 after a fossil-based cross-validation with an increment of

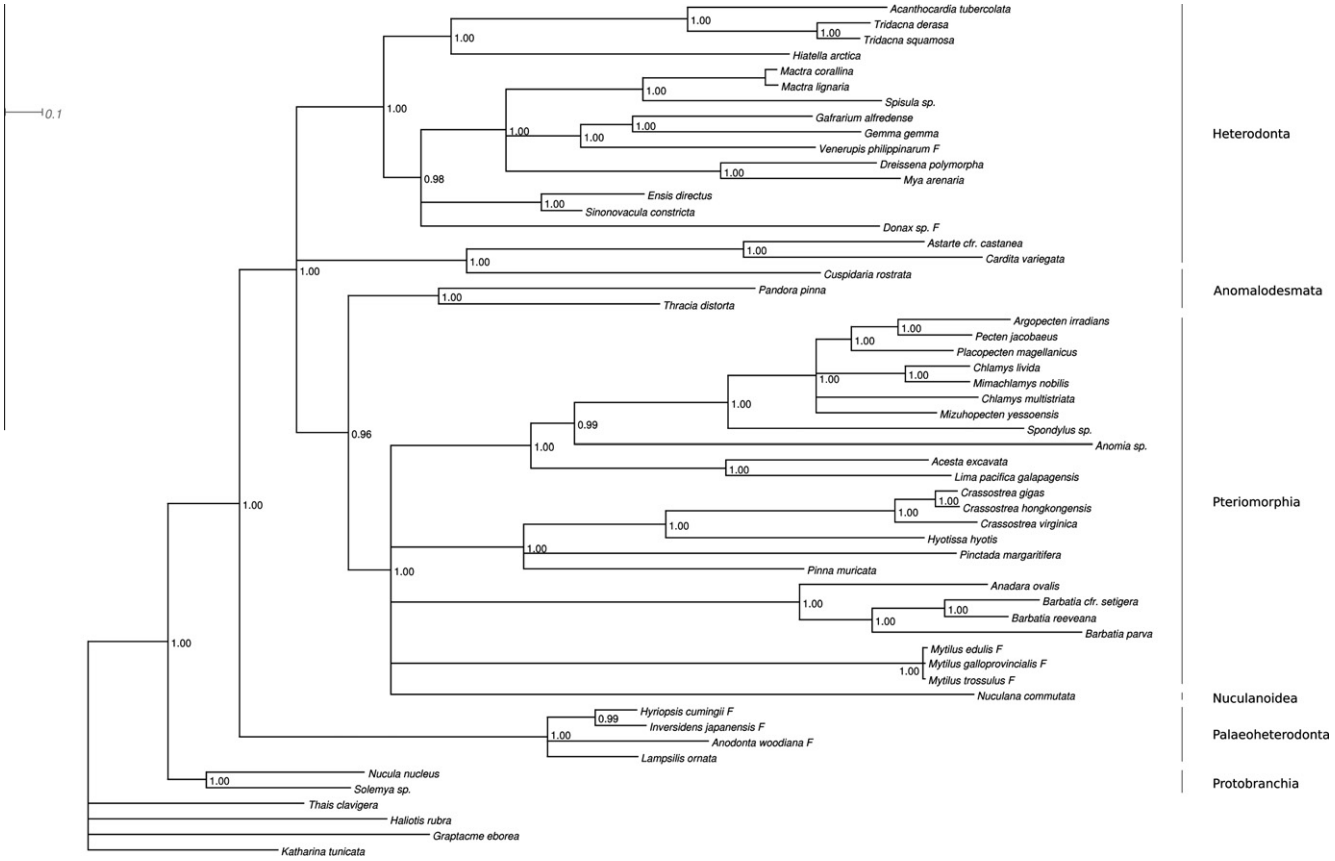


Fig. 2. Majority-rule *tM3* consensus tree from the Bayesian multigene partitioned analysis. Numbers at the nodes are PP values. Nodes under 0.95 were collapsed. Bar units in expected changes per site.

0.01. The best calibration set comprises genus *Barbatia*, subfamily Unioninae, families Veneridae, Limidae, Pectinidae, Cardiidae, Arcidae, and Bivalvia; all constraints were respected. Age for many other taxa were correctly predicted with an error of always less than 50 million years (Myr), as shown in Table 3. This was not the case for genera *Mytilus*, *Mactra*, *Crassostrea*, and *Tridacna*: with the notable exception of *Tridacna*, they were predicted to be much more recent than they appeared in fossil records. This is easily explained by the fact that in all cases (except *Tridacna*) strictly related species were represented in our tree, which diverged well after the first appearance of the genus. Results from PL and NPRS were substantially identical: as in four cases NPRS analysis did not pass the **checkGradient** control, we will present and discuss PL results only.

Deep nodes were all dated between 530 and 450 million years ago (see Fig. 3): the origin of the class was dated 530 Mya, Autolamellibranchiata 520 Mya and their sister group Protobranchia 454 Mya. Within Autolamellibranchiata, the big group comprehending Heterodonta and Pteriomorphia would have arisen about 514 Mya; the radiation of Palaeoheterodonta was not computed as only specimens from Unionidae (293.93 Mya) were present. Pteriomorphia and Heterodonta originated very close in time, about 506 and 498 Mya, respectively. Within Pteriomorphia, the basal clade of Anomalodesmata is more recent (431 Mya) than the main group of traditional Pteriomorphia (497 Mya). On the other hand, the main split within Heterodonta gave rise to *Hiatella* + Cardiidae about 481 Mya, and to Veneroida *sensu lato* 471 Mya. Evolutionary rates (expressed as mutations per year per site) varied consistently, ranging from 0.000693 of branch leading to genus *Barbatia* to 0.011 of the *Hiatella* + Cardiidae group. Table 3 also lists the mean value of NPRS dating across 150 bootstrap replicates and its standard deviation, and it is worth noting that deeper nodes do have very little standard deviation.

4. Discussion

4.1. The methodological pipeline

As the correct selection of suitable molecular markers was (and still is) a major concern in bivalves' phylogenetic analysis, we tested for different ways of treating the data. Our best-performing approach is based on four different mitochondrial genes, and because we obtained robust and reliable phylogenies in our analysis, we can now confirm that this choice is particularly appropriate in addressing deep phylogeny of Bivalvia, given a robust analytical apparatus.

As mentioned, our mitochondrial markers were highly informative, especially protein-coding ones and our results from model selection were straightforward. The phylogenetic signal we recovered in our dataset is complex, as different genes and different positions must have experienced different histories and selective pressures. Moreover, performed single-gene analyses yielded controversial and poorly informative trees (data not shown).

Specifically, both AIC and BF separated ribosomal and protein-coding genes for traditional 4by4 models. AIC tends to avoid overparametrization, as it presents a penalty computed on free parameters, and selected a simpler model; conversely, BF selected the most complex partitioning scheme. BF has been proposed to be generally preferable to AIC (Kass and Raftery, 1995; Alfaro and Huelsenbeck, 2006), but Nylander et al. (2004) pointed out that BF is generally consistent with other model selection methods, like AIC. Indeed, trees obtained under models *t04*, *t07*,

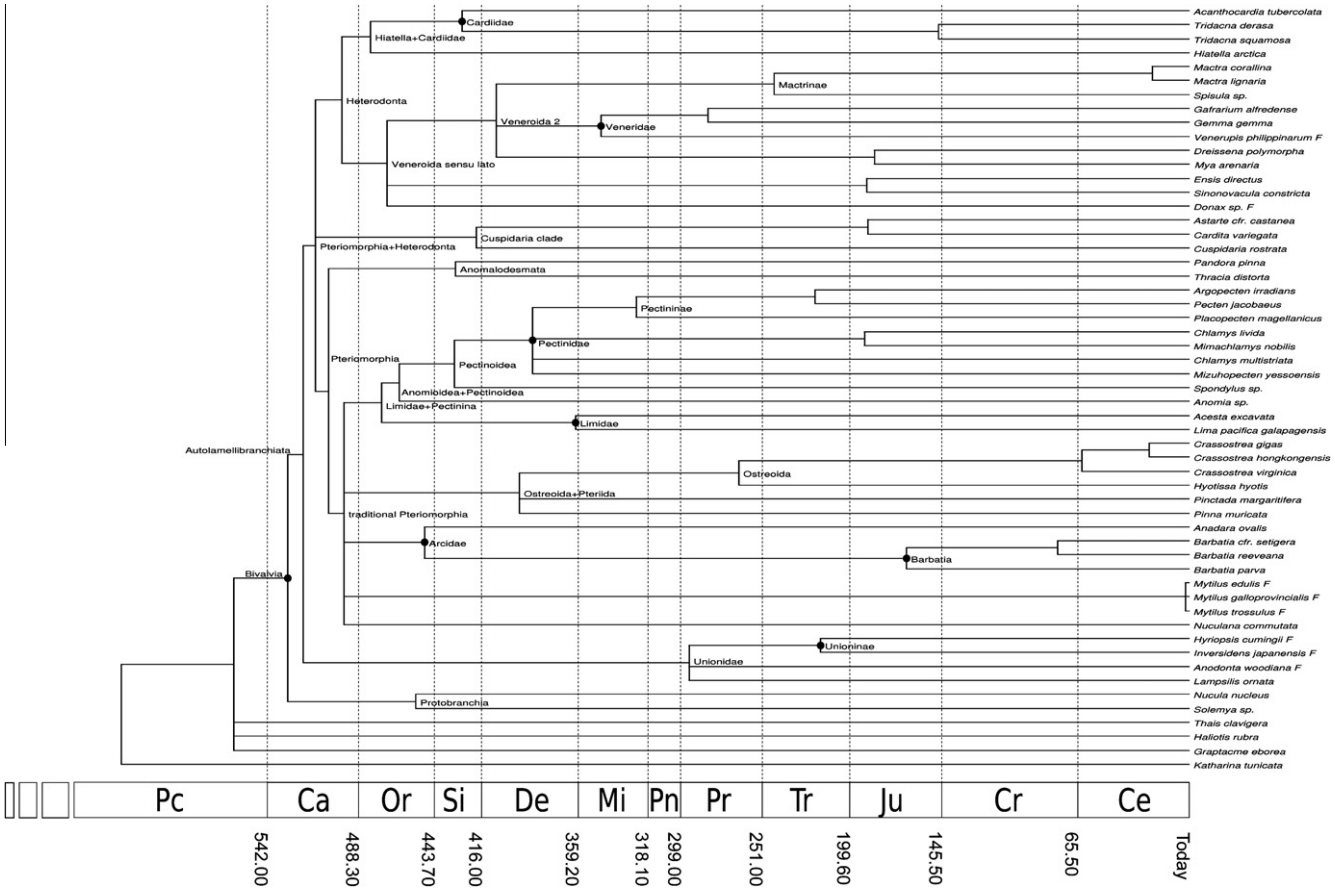


Fig. 3. Results from time calibration of *tM3* tree. The ultrametric *tM3* tree computed by r8s (under Penalized Likelihood method, see text for further details) is shown along with geological time scale and major interval boundaries (ages in million years). Only deep nodes are named: for a complete survey of node datations, see Table 3. Geological data taken from Gradstein et al. (2004) and Ogg et al. (2008). Pc, Precambrian (partial); Ca, Cambrian; Or, Ordovician; Si, Silurian; De, Devonian; Mi, Mississippian; Pn, Pennsylvanian; Pr, Permian; Tr, Triassic; Ju, Jurassic; Cr, Cretaceous; Ce, Cenozoic.

t08, and *t09* are very similar (data not shown). Anyway, the *tM3* model clearly outperformed all alternatives, following both AIC and BF criteria (see Table 4). Furthermore, this was not the case for models *tNy98* and *t10*, which we used to reduce possible misleading phylogenetic noise, albeit in different ways (by a Ny98 codon model or by amminoacids, respectively). *t10* tree was similar to *tM3* one, but significantly less resolved on many nodes, thus indicating a loss of informative signal (data not shown). M3 codon model allows lower ω categories than Ny98; on the other hand, it does not completely eliminate nucleotide information level, as amminoacid models do. All this considered, we propose that M3 codon model is the best way for investigating bivalve phylogeny.

Finally, it is quite evident that Bayesian analysis yielded the most resolved trees, when compared to Maximum Likelihood and this was especially evident for ancient nodes. The tendency of Bayesian algorithms to higher nodal support has been repeatedly demonstrated (Leaché and Reeder, 2002; Suzuki et al., 2002; Whittingham et al., 2002; Cummings et al., 2003; Douady et al., 2003; Erixon et al., 2003; Simmons et al., 2004; Cameron et al., 2007), though Alfaro et al. (2003) found that PP is usually a less biased predictor of phylogenetic accuracy than bootstrap. Anyway, it has to be noted that most of our recovered nodes are strongly supported by both methods; we therefore think that the higher support of Bayesian analysis is rather due to a great affordability of the method in shaping and partitioning models, which is nowadays impossible with Maximum Likelihood algorithms. All that considered, we suggest that a suitable methodological pipeline for bivalves' future phylogenetic reconstructions should be as such: (i) sequence analyses for saturation and selection; (ii) rigorous evaluation of taxon coverage; (iii) tests for best data partitioning; (iv) appropriate model decision statistics; (v) Bayesian analysis; (vi) eventual dating by cross-validation with fossil records.

4.2. The phylogeny of Bivalvia

4.2.1. Protobranchia Pelseneer

Our study confirms most of the recent findings (Giribet and Wheeler, 2002; Giribet and Distel, 2003; Kappner and Bieler, 2006): Nuculoidea and Solemyoidea do maintain their basal position, thus representing Protobranchia *sensu stricto*, which is a sister group to all Autolamellibranchiata. On the contrary, Nuculanoidea, although formerly placed in Nuculoida, is better considered within Pteriomorphia, placed in its own order Nuculanoida. The split separating *Nucula* and *Solemya* lineages is dated around the late Ordovician (454.28 Mya); since the first species of the subclass must have evolved earlier (about 500 Mya), this is a clear signal of the antiquity of this clade. In fact, based on paleontological records, the first appearance of Protobranchia is estimated around 520 Mya (early Cambrian) (He et al., 1984; Parkhaev, 2004), and our datation is only slightly different (482.02 Mya, with a standard deviation of 14.61).

4.2.2. Palaeoheterodonta Newell

Freshwater mussels are basal to all the remaining Autolamellibranchiata (Heterodonta + Pteriomorphia), as supposed by Cope (1996). Therefore, there is no evidence for Heteroconchia *sensu*

Bieler and Mikkelsen (2006) in our analysis. The monophyletic status of the subclass was never challenged in our Bayesian analyses, nor in traditional Maximum Likelihood ones. Finally, since we obtained sequences only from specimens from Unionoidea: Unionidae, a clear dating of the whole subclass is not sound, as shown by a relatively high difference between PL values and mean across bootstrap replicates (294 and 348 Mya, respectively). Therefore, the origin of the subclass must date back to before than 350 Mya, which is comparable to paleontological data (Morton, 1996).

4.2.3. Pteriomorphia Newell

Here we obtained a Pteriomorphia *sensu novo* subclass comprising all pteriomorphians *sensu* Millard (2001), as well as Nuculanoidea and anomalodesmatans. This diverse taxon arose about 506 Mya, which makes it the first bivalve radiation in our tree, dated in the middle Cambrian, which is perfectly in agreement with paleontological data. Moreover, our results proved to be stable also with bootstrap resampling, with a standard deviation of slightly more than 2 million of years (Table 3). A wide polytomy is present within the subclass; as this polytomy is constantly present in all the analyses, and it has been found also by many other authors (see Campbell, 2000; Steiner and Hammer, 2000; Matsumoto, 2003), we consider it as a "hard polytomy", reflecting a true rapid radiation dated about 490 Mya (Cambrian/Ordovician boundary). Sister group to this wide polytomy is the former anomalodesmatan suborder Pholadomyina. In our estimate, the clade *Pandora* + *Thracia* seems to have originated something like 431.45 Mya, as several pteriomorphian groups, like Pectinoidea (431.77 Mya) or Arcidae (449.51 Mya). On the other hand, we failed in retrieving *Cuspidaria* within the pteriomorphian clade, while this genus is strictly associated with *Astarte* + *Cardita*. Not only the nodal support is strong, this relationship is also present across almost all trees and models. It has to be noted that the association between *Cuspidaria* and (*Astarte* + *Cardita*) has been evidenced already (Giribet and Distel, 2003). On the other side, suborder Pholadomyina is always basal to pteriomorphians (data not shown). Maybe it is worth noting that *Cuspidaria* branch is the longest among anomalodesmatans and that *Astarte* and *Cardita* branches are the longest among heterodonts (see Fig. 2). Moreover, this clade is somewhat unstable across bootstrap replicates (see Table 3). Maybe the large amount of mutations may overwhelm the true phylogenetic signal for such deep nodes, as also expected by their relatively high mutation rates. Hence, we see three possible alternatives: (i) an artifact due to long-branch-attraction – all anomalodesmatans belong to Pteriomorphia, whereas *Astarte* and *Cardita* belong to Heterodonta; (ii) anomalodesmatans do belong to Heterodonta, whose deeper nodes are not so good resolved, whereas a strong signal is present for Pteriomorphia monophyly, thus leading to some shuffling into basal positions; (iii) anomalodesmatans are polyphyletic, and the two present-date suborders do not share a common ancestor. The two last possibilities seem unlikely to us, given our data and a considerable body of knowledge on the monophyletic status of Heterodonta and Anomalodesmata (Canapa et al., 2001; Dreyer et al., 2003; Harper et al., 2006; Taylor et al., 2007). We therefore prefer the first hypothesis, albeit an anomalodesmatan clade nested within heterodonts has also been appraised by some authors (Giribet and Wheeler, 2002; Giribet and Distel, 2003; Bieler and Mikkelsen, 2006; Harper et al., 2006). Interestingly, in *t10* tree the whole group *Cuspidaria* + (*Astarte* + *Cardita*) nested within pteriomorphians species; a similar result was also yielded by a wider single-gene *cox1* dataset (data not shown). This would also account for the great difference found in *Astarte* + *Cardita* split across bootstrap replicates. A major taxonomical revision is needed for basal pteriomorphians, including also anomalodesmatans, as well as for superfamilies Astartoidea and Carditoidea.

As mentioned above, the main groups of pteriomorphians, arising in the late Cambrian, comprehend the genus *Nuculana* also. This placement was first proposed by Giribet and Wheeler (2002) on molecular bases and our data strongly support it. Its clade must have diverged from other main pteriomorphian groups at the very beginning of this large radiation. Among the main groups of Pteriomorphia, it is also worth noting the breakdown of the orders Pterioida *sensu* Vokes (1980) and Ostreoida *sensu* Millard (2001): the suborder Ostreina constitutes a net polyphyly with suborder Pectinina. The former is better related with order Pteriida *sensu* Millard (2001) (*Pinna*, *Pinctada*), whereas the latter is better related with superfamilies Limoidea (*Lima* + *Acesta*) and Anomioidea (*Anomia*). This is in agreement with most recent scientific literature about Pteriomorphia (Steiner and Hammer, 2000; Matsumoto, 2003).

4.2.4. Heterodonta Newell

The subclass seems to have originated almost 500 Mya (late Cambrian) and its monophyletic status is strongly confirmed by our analysis, but a major revision of its main subdivisions is also required. The placement of *Astarte* and *Cardita* has already been discussed. At the same time, the orders Myoida and Veneroida, as well as the Chamida *sensu* Millard (2001), are no longer sustainable. A first main split separates (*Hiatella* + Cardiidae) from all remaining heterodonts. This split may correspond to two main orders in the subclass. As we sampled only 15 specimens of Heterodonta, we could only coarsely assess their phylogenetic taxonomy. However, we could precisely demonstrate the monophyly of families Veneridae and Mactridae and their sister group status. This could correspond together with *Dreissena* + *Mya* to a superfamily Veneroidea *sensu novo*, which is stably dated around the early Devonian; however, further analyses are requested towards an affordable taxonomical revision, which is beyond the aims of this paper. Finally, recent findings about Tridacninae subfamily within Cardiidae family (Maruyama et al., 1998) are confirmed against old taxonomy based on Cardioidea and Tridacnoidea superfamilies (Millard, 2001).

Concluding, our work evidenced that all main deep events in bivalve radiation took place in a relatively short 70 Myr time during late Cambrian/early Ordovician (Fig. 3). Dates are stable across bootstrap replicates, especially those of deeper nodes, which were one of the main goals of this work (Table 3): most NPRS bootstrap means are indeed very close to PL estimates and standard deviations are generally low. Notable exceptions are some more recent splits on long branches (*Chlamys livida* + *Mimachlamys*, *Ensis* + *Sinonovacula*, *Astarte* + *Cardita*, *Tridacna*), which clearly are all artifacts of low taxon sampling for that specific branch, and Unionidae and Ostreoida. Unionidae are the only palaeoheterodonts we sampled and this could account for this anomaly; anyway, it is worth taking

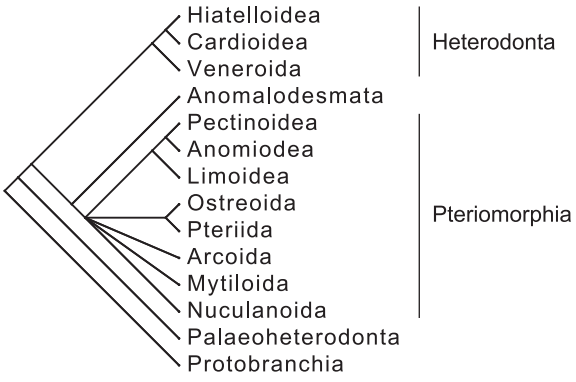


Fig. 4. Global survey of the bivalve phylogeny.

Table A1
PCR conditions.

		12s		*16s*		*cox1*		*cytb*	
		Annealing	Primers	Annealing	Primers	Annealing	Primers	Annealing	Primers
1	*Anadara ovalis*	50 °C 30″	SR-J14197 ÷ SR-N14745			56 °C 20″	coIF ÷ coIR	48 °C 30″	cobF ÷ cobR
2	*Anodonta woodiana*					48 °C 1′	LCO ÷ HCO	48 °C 1′	cobF ÷ cobR
3	*Anomia sp.*			48 °C 1′	16SbrH(32) ÷ 16Sar(34)	56–46 °C 30″–1′	coIF ÷ coIR	48 °C 30″	cobF ÷ cobR
4	*Argopecten irradians*	50 °C 30″	SR-J14197 ÷ SR-N14745	48 °C 1′	16SbrH(32) ÷ 16Sar(34)	56–46 °C 30″–1′	coIF ÷ coIR	55–45 °C 30″–1′	cobF ÷ cobR
5	*Astarte* cfr. *castanea*	50 °C 30″	SR-J14197 ÷ SR-N14745					48 °C 30″	cobF ÷ cobR
6	*Barbatia parva*	50 °C 30″	SR-J14197 ÷ SR-N14745			48 °C 1′	LCO ÷ HCO	48 °C 1′	cobF ÷ cobR
7	*Barbatia reeveana*	50 °C 30″	SR-J14197 ÷ SR-N14745			52 °C 20″	coIF ÷ coIR	53–43 °C 30″–1′	cobF ÷ cobR
8	*Barbatia* cfr. *setigera*	50 °C 30″	SR-J14197 ÷ SR-N14745			54 °C 20″	coIF ÷ coIR	48 °C 1′	cobF ÷ cobR
9	*Cardita variegata*	50 °C 30″	SR-J14197 ÷ SR-N14745			48 °C 1′	LCO ÷ HCO	48 °C 1′	cobF ÷ cobR
10	*Chlamys livida*	50 °C 30″	SR-J14197 ÷ SR-N14745	48 °C 1′	16SbrH(32) ÷ 16Sar(34)	52 °C 20″	coIF ÷ coIR	48 °C 1′	cobF ÷ cobR
11	*Chlamys multistriata*			54 °C 2′	16SbrH(32) ÷ 16SDon			48 °C 1′	cobF ÷ cobR
12	*Cuspidaria rostrata*	50 °C 30″	SR-J14197 ÷ SR-N14745			48 °C 1′	LCO ÷ HCO	58–48 °C 1′	cobF ÷ cobR
13	*Ensis directus*	46 °C 30″	SR-J14197 ÷ SR-N14745	54 °C 2′	16SbrH(32) ÷ 16SDon	56–46 °C 30″–1′	coIF ÷ coIR	53–43 °C 1′	cobF ÷ cobR
14	*Gafrarium alfredense*	50 °C 30″	SR-J14197 ÷ SR-N14745	48 °C 1′	16SbrH(32) ÷ 16Sar(34)			48 °C 1′	cobF ÷ cobR
15	*Gemma gemma*			48 °C 1′	16SbrH(32) ÷ 16Sar(34)	52 °C 20″	coIF ÷ coIR	58–48 °C 1′	cobF ÷ cobR
16	*Hyotissa hyotis*	50 °C 30″	SR-J14197 ÷ SR-N14745	48 °C 1′	16SbrH(32) ÷ 16Sar(34)	52 °C 20″	coIF ÷ coIR	58–48 °C 1′	cobF ÷ cobR
17	*Lima pacifica galapagensis*	50 °C 30″	SR-J14197 ÷ SR-N14745	48 °C 45″[a]	16SbrH(32) ÷ 16SarL[a]	52 °C 20″	coIF ÷ coIR	53–43 °C 30″–1′	cobF ÷ cobR
18	*Mactra corallina*	48 °C 1′	SR-J14197 ÷ SR-N14745	56 °C 1′	16SbrH(32) ÷ 16Sar(34)	48 °C 1′	LCO ÷ HCO	48 °C 1′	cobF ÷ cobR
19	*Mactra lignaria*	48 °C 1′	SR-J14197 ÷ SR-N14745	56 °C 1′	16SbrH(32) ÷ 16Sar(34)	48 °C 1′	LCO ÷ HCO		
20	*Mya arenaria*							48 °C 1′	cobF ÷ cobR
21	*Nucula nucleus*	50 °C 30″	SR-J14197 ÷ SR-N14745	54 °C 2′	16SbrH(32) ÷ 16SDon				
22	*Nuculana commutata*	50 °C 30″	SR-J14197 ÷ SR-N14745			48 °C 1′	LCO ÷ HCO	48 °C 1′	cobF ÷ cobR
23	*Pandora pinna*	50 °C 30″	SR-J14197 ÷ SR-N14745	53–43 °C 1′20″	16SbrH(32) ÷ 16SarL	48 °C 1′	LCO ÷ HCO	53–43 °C 1′20″	UCYTBF144F ÷ UCYTB272R
24	*Pecten jacobaeus*							58 °C.48 °C 1′	cobF ÷ cobR
25	*Pinna muricata*	50 °C 30″	SR-J14197 ÷ SR-N14745	48 °C 1′	16SbrH(32) ÷ 16Sar(34)	52 °C 20″	coIF ÷ coIR	48 °C 1′	cobF ÷ cobR
26	*Thracia distorta*	50 °C 30″	SR-J14197 ÷ SR-N14745			48 °C 1′	LCO ÷ HCO	48°C 1′	cobF ÷ cobR
27	*Tridacna derasa*					48 °C 1′	LCO ÷ HCO	48 °C 1′	cobF ÷ cobR
28	*Tridacna squamosa*							48 °C 1′	cobF ÷ cobR
	Transformed inserts	55 °C 30″	M13F ÷ M13R	55 °C 30″	M13F ÷ M13R	55 °C 30″	M13F ÷ M13R	55 °C 30″	M13F ÷ M13R

[a] This amplification was carried out with Herculase reaction kit (Stratagene, Cedar Creek, TX, USA), following manufacturer's instructions.

Table A2
Primer used in this study.

	5′–3′ Sequence	Reference
SR-J14197	GTACAYCTACTATGTTACGACTT	Simon et al. (2006)
SR-N14745	GTGCCAGCAGYYGCGGTTANAC	Simon et al. (2006)
16SbrH(32)	CCGGTCTGAACTCAGATCACGT	Palumbi et al. (1996)
16Sar(34)	CGCCTGTTTAACAAAAACAT	Modified from Palumbi et al. (1996)
16SarL	CGCCTGTTTATCAAAACAT	Palumbi et al. (1996)
16SDon	CGCCTGTTTATCAAAAACAT	Kocher et al. (1989)
LCO1490	GGTCAACAAATCATAAAGATATTGG	Folmer et al. 1994)
HCO2198	TAAACTTCAGGGTGACCAAAAAATCA	Folmer et al. (1994)
COIF	ATYGGNGGNTTYGGNAAYTG	Matsumoto (2003)
COIR	ATNGCRAANACNGCNCCYAT	Matsumoto (2003)
CobF	GGWTAYGTWYTWCCWTGRGGWCARAT	Passamonti (2007)
CobR	GCRTAWGCRAAWARRAARTAYCAYTCWGG	Passamonti (2007)
UCYTB144F	TGAGSNCARATGTCNTWYTG	Merritt et al. (1998)
UCYTB272R	GCRAANAGRAARTACCAYTC	Merritt et al. (1998)
M13F	GTAAAACGACGGCCAGT	
M13R	CAGGAAACAGCTATGAC	

into account that the r8s-bootkit follows a slightly different method than *tout court* PL, therefore the results are not expected to perfectly coincide. When this happens, however, i.e. for most nodes in Fig. 3, it accounts for a substantial stability in timing estimates.

We show in Fig. 4 the survey on bivalve taxonomy which we described above. Given the still limited, but statistically representative, taxon sampling available, it is nowadays inconceivable to propose a rigorous taxonomy at order and superfamily level; therefore, we used in Fig. 4 the nomenclature of Millard (2001) and Vokes (1980). More taxa and genes to be included will sharp resolution and increase knowledge on bivalves' evolutionary history.

Acknowledgments

We are very thankful to Paolo Giulio Albano, Mirco Bergonzoni, Jeffrey L. Boore, Alessandro Ceregato, Walter Gasperi, Ilaria Guarniero, Constantine Mifsud, Liliana Milani, Francesco Nigro, Edoardo

Table A3
GenBank accession numbers of sequences used in this study. Bold sequences were obtained for this work.

	12s	*16s*	*cox1*	*cytb*
Acanthocardia tubercolata	DQ632743	DQ632743	DQ632743	DQ632743
Acesta excavata	AM494885	AM494899	AM494909	AM494922
Anadara ovalis	**GQ166533**		**GQ166571**	**GQ166592**
Anodonta woodiana F		DQ073815	EF440349	**GQ166594**
Anomia sp.		**GQ166557**	**GQ166573**	**GQ166595**
Argopecten irradians	**GQ166535**	**GQ166558**	**GQ166574**	**GQ166596**
Astarte castanea			AF120662	
Astarte cfr. castanea	**GQ166536**			**GQ166597**
Barbatia parva	**GQ166537**		**GQ166575**	**GQ166599**
Barbatia reeveana	**GQ166538**		**GQ166576**	**GQ166600**
Barbatia cfr. setigera	**GQ166539**		**GQ166577**	**GQ166601**
Cardita variegata	**GQ166540**		**GQ166578**	**GQ166605**
Chlamys livida	**GQ166541**	**GQ166559**	**GQ166579**	**GQ166606**
Chlamys multi striata	AJ571604	**GQ166560**		**GQ166607**
Crassostrea gigas	AF177226	AF177226	AF177226	AF177226
Crassostrea hongkongensis F	EU266073	EU266073	EU266073	EU266073
Crassostrea viriginica	AY905542	AY905542	AY905542	AY905542
Cuspidaria rostrata	**GQ166542**		**GQ166580**	**GQ166608**
Donax faba F			AB040844	
Donax trunculus F		EF417549		EF417548
Dreissena polymorpha		DQ280038	AF120663	DQ072117
Ensis directus	**GQ166543**	**GQ166561**	**GQ166581**	**GQ166610**
Gafrarium alfredense	**GQ166544**	**GQ166562**		**GQ166611**
Gemma gemma		**GQ166563**	**GQ166582**	**GQ166612**
Graptacme eborea	AY484748	AY484748	AY484748	AY484748
Haliotis rubra	AY588938	AY588938	AY588938	AY588938
Hiatella arctica	DQ632742	DQ632742	DQ632742	DQ632742
Hyotissa hyotis	**GQ166545**	**GQ166564**	**GQ166583**	**GQ166613**
Hyriopsis cumini	FJ529186	FJ529186	FJ529186	FJ529186
Inversidens japanensis F	AB055625	AB055625	AB055625	AB055625
Katharina tunicata	U09810	U09810	U09810	U09810
Lampsilis ornata	AY365193	AY365193	AY365193	AY365193
Lima pacifica galapagensis	**GQ166548**	**GQ166565**	**GQ166584**	**GQ166616**
Mactra corallina	**GQ166550**	**GQ166566**	**GQ166585**	**GQ166617**
Mactra lignaria	**GQ166551**	**GQ166567**	**GQ166586**	
Mimachlamys nobilis	FJ415225	FJ415225	FJ415225	FJ415225
Mizuhopecten yessoensis	AB271769	AB271769	AB271769	AB271769
Mya arenaria		AY377618	AF120668	**GQ166619**
Mytilus edulis F	AY484747	AY484747	AY484747	AY484747
Mytilus galloprovincialis F	AY497292	AY497292	AY497292	AY497292
Mytilus trossulus F	DQ198231	DQ198231	DQ198231	DQ198231
Nucula nucleus	**GQ166552**	**GQ166568**	AM696252	
Nuculana commutata	**GQ166553**		**GQ166587**	**GQ166622**
Pandora pinna	**GQ166554**	**GQ166569**	**GQ166588**	**GQ166623**
Pecten jacobaeus	AJ571596	AJ245394	AY377728	**GQ166624**
Pinctada margariti fera	AB250256	AB214436	AB259166	
Pinna muricata	**GQ166555**	**GQ166570**	**GQ166589**	**GQ166625**
Placopecten magellanicus	DQ088274	DQ088274	DQ088274	DQ088274
Sinonovacula constricta	EU880278	EU880278	EU880278	EU880278
Solemya velesiana				AM293670
Solemya velum		DQ280028	U56852	
Spisula solidissima				AF205083
Spisula solidissima solidissima			AY707795	
Spisula subtruncata		AJ548774		
Spondylus gaederopus	AJ571607	AJ571621		
Spondylus varius			AB076909	
Thais clavigera	DQ159954	DQ159954	DQ159954	DQ159954
Thracia distorta	**GQ166556**		**GQ166590**	**GQ166626**
Tridacna derasa		AF122976	**GQ166591**	**GQ166627**
Tridacna squamosa		AF122978	EU346361	**GQ166628**
Venerupis philippinarum F	AB065375	AB065375	AB065375	AB065375

Turolla, and Diego Viola, who kindly provided some specimens for this study. Many thanks also to Dr. Andrea Ricci, who introduced one of us (F.P.) into laboratory life. This work was supported by the University and Research Italian Ministry (MIUR PRIN07, grant number 2007NSHJL8_002) and the "Canziani Bequest" fund (University of Bologna, grant number A.31.CANZELSEW). Thanks are also due to two anonymous reviewers who provided helpful comments on an earlier draft of this paper.

Appendix A

See Tables A1–A4.

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2010.08.032.

Table A4
Subtrees used for assessing parameter estimate accurateness.

Taxon labels:							
1	*Acanthocardia tubercolata*	15	*Crassostrea hongkongensis*	29	*Katharina tunicata*	43	*Pecten jacobaeus*
2	*Acesta excavata*	16	*Crassostrea virginica*	30	*Lampsilis ornata*	44	*Pinctada margaritifera*
3	*Anadara ovalis*	17	*Cuspidaria rostrata*	31	*Lima pacifica galapagensis*	45	*Pinna muricata*
4	*Anodonta woodiana* F	18	*Donax* sp. F	32	*Mactra corallina*	46	*Placopecten magellanicus*
5	*Anomia* sp.	19	*Dreissena polymorpha*	33	*Mactra lignaria*	47	*Sinonovacula constricta*
6	*Argopecten irradians*	20	*Ensis directus*	34	*Mimachlamys nobilis*	48	*Solemya* sp.
7	*Astarte* cfr. *castanea*	21	*Gafrarium alfredense*	35	*Mizuhopecten yessoensis*	49	*Spisula* sp.
8	*Barbatia parva*	22	*Gemma gemma*	36	*Mya arenaria*	50	*Spondylus* sp.
9	*Barbatia reeveana*	23	*Graptacme eborea*	37	*Mytilus edulis* F	51	*Thais clavigera*
10	*Barbatia* cfr. *setigera*	24	*Haliotis rubra*	38	*Mytilus galloprovincialis* F	52	*Thracia distorta*
11	*Cardita variegata*	25	*Hiatella arctica*	39	*Mytilus trossulus* F	53	*Tridacna derasa*
12	*Chlamys livida*	26	*Hyotissa hyotis*	40	*Nucula nucleus*	54	*Tridacna squamosa*
13	*Chlamys multistriata*	27	*Hyriopsis cumingii* F	41	*Nuculana commutata*	55	*Venerupis philippinarum* F
14	*Crassostrea gigas*	28	*Inversidens japanensis* F	42	*Pandora pinna*		

Tree tM3:

(51,29,24,23,(((((7,11),17),(((1,(53,54)),25),((20,47),(((32,33),49),((21,22),55),(19,36)),18)),(((37,38,39),((2,31),(5,((35,13,(12,34),((6,43),46)),50))),((((14,15),16),26),44,45),(3,((10,9),8)),41),(42,52))),((27,28),4,30)),(40,48)));

Subtrees:

1 (51,29,24,23,((((17),(((1,(53,54)),25),((20,47),(((32,33),49),((21,22),55),(19,36)),18)),(((37,38,39),((2,31),(5,((35,13,(12,34),((6,43),46)),50))),((((14,15),16),26),44,45),(3,((10,9),8)),41),(42,52))),((27,28),4,30)),(40,48)));

2 (51,29,24,23,(((((1,(53,54)),25),((20,47),(((32,33),49),((21,22),55),(19,36)),18)),(((37,38,39),((2,31),(5,((35,13,(12,34),((6,43),46)),50))),((((14,15),16),26),44,45),(3,((10,9),8)),41),(42,52))),((27,28),4,30)),(40,48)));

3 (51,29,24,23,(((((7,11),17),(((1,(53,54)),25),((20,47),(((32,33),49),((21,22),55),(19,36)),18)),(((37,38,39),((2,31),(5,((35,13,(12,34),((6,43),46)),50))),((((14,15),16),26),44,45),(3,((10,9),8)),41),(42,52)))),(40,48)));

4 (51,29,24,23,(((((7,11),17),(((1,(53,54)),25),((20,47),(((32,33),49),((21,22),55),(19,36)),18)),(((37,38,39),((2,31),(5)),((((14,15),16),26),44,45),(3,((10,9),8)),41),(42,52))),((27,28),4,30)),(40,48)));

5 (51,29,23,(((((7,11)),(((1,(53,54)),25),((20,47),(((32,33),49),((21,22),55),(19,36)))),(((37,38,39),((2,31),(5,((35,13,(12,34),((6,43),46)),50))),((((14,15),16),26),44,45),(3,((10,9),8)),41))),((27,28)))));

6 (51,29,24,23,(((((7,11),17),(((1),25),((20,47),((49),((21,22),55),(19,36)),18)),((((2,31),(5,((35,13,(12,34),((6,43),46)),50))),((((14,15),16),26),44,45),(3),41),(42,52))),((27,28),4,30)),(40,48)));

7 (51,24,23,(((((7,11),17),((((53,54)),25),((20),(((32,33),49),((21,22)),(19,36)),18)),(((38,39),((2,31),(5,((35,13,(12),((6,43),46)),50))),((((14),16),26),44,45),(3,((9),8))),(52))),((27),4,30)),(40)));

8 (23,(((((7,11),17),(((((32,33),49),((21,22),55),(19,36)),18)),(((37,38,39),((2,31),(5,((35,13,(12,34),((6,43),46)),50))),((((14,15),16),26)),(3,((10,9),8)),41),(42,52))),((28),4)),(48)));

9 (51,29,24,23,(((((7,11),17),((41,(37,38,39),((2,31),(5,((35,13,(12,34),((6,43),46)),50))),((((14,15),16),26),44,45),(3,((10,9),8))),(42,52))),((27,28),4,30)),(40,48)));

10 (51,29,(((((7),17),(((1,(54)),25),((20),(((32)),((21,22)),(19,36)))),(((37,38),((2,31),(((35,(12,34),((6,43),46)),50))),((((14,15)),26),44),(3,((10),8)),41),(52))),((27,28),30)),(40)));

11 ((((7,11),17),(((1,(53,54))),((((32,33),49),((21,22),55)))),(((37,38,39),((5,((35,13,(12,34),((6,43))),50))),((((14,15),16)),44,45),(((10,9),8))))),((27,28),4,30));

12 (51,29,24,23,(((((7,11),17),(((1,(53,54)),25),((20,47),(((32,33),49),((21,22),55),(19,36)),18)),(((((10,9),8)))))),((27,28),4,30)),(40,48)));

13 (29,23,(((((11)),(((1,(54)))),((20),(((32),49),((22)),(19)),18)),(((38),((2),(5,((13,(34),((43))),50))),((((15)),26),45),(((10),8))),(42))),((27),4)),(40)));

14 (23,((((17),((((54))),((20,47))),((((2,31),(5,((13,(34),((6),46))))),((((14,15),16)),44,45),41),(42))),((27),4,30)),(40,48)));

15 (29,24,23,((((((1,(53,54)))),((20),(((32,33),49),((22)),(19)))),(((38,39),((5,((13,(34),(46))))),((((14))),44)))),((27)))));

16 (((((7,11),17),((25),(((36)),18)),(((37),((5)),(((16))),41),(42,52))),((27,28),4,30)),(40,48));

17 (((((53,54))),((((32,33)),(55)))),(((37,38,39),(((((12,34)))),((((14,15),16))),(((10,9),8)))));

18 (51,24,(((((7),17),(((((33)),(19)))),((((2),(((35)))),((26))),(52))),((28),30)),(40,48)));

19 ((2,31),(5,((35,13,(12,34),((6,43),46)),50)));

20 (((1,(53,54)),25),((20,47),(((32,33),49),((21,22),55),(19,36)),18));

21 (29,(((((11)),((((49)))),((((5,(50))),((8)),41),(42))),((27))),(48)));

22 (51,(((((7)),(((20))),(((37,38,39),((((14))))))),(40)));

23 (((((((21))))),(((45)),(52))),(4)),(48));

24 (51,29,24,23,(((((7,11),17),(((1,(53,54)),25),(18,(20,47),(((32,33),49),((21,22),55),(19,36)))),((41,(37,38,39),((2,31),(5,((35,13,(12,34),((6,43),46)),50))),(((14,15),26),44,45),(3,((10,9),8))),(42,52))),((27,28),4,30)),(40,48)));

25 (51,29,24,23,(((((7,11),17),(((1,(53,54)),25),(18,(20,47),(((32,33),49),((21,22),55),(19,36)))),((41,(37,38,39),((2,31),(5,((35,13,(12,34),((6,43),46)),50))),(44,45),(3,((10,9),8))),(42,52))),((27,28),4,30)),(40,48)));

26 (51,29,23,(((((7,11),17),((1,25),(18,(20,47),(((32,33),49),((21,22),55),(19,36)))),((41,((2,31),(5,((35,13,(12,34),((6,43),46)),50))),((((14,15),16),26),44,45),(3,((10,9),8))),(42,52))),((27,28),4,30)),(40,48)));

27 (51,29,24,23,(((((7,11),17),(((1,(53,54)),25),(18,(20,47),(((32,33),49),((21,22),55),(19,36)))),((41,(37,38,39),((2,31),(5,((35,13,(12,34),((6,43),46)),50))),((((14,15),16),26),44,45),(3,((10,9),8))),(42,52))),((27,28),4,30)),(40,48)));

28 (51,24,23,((((11,17),(((53,1),25),(18,(20,47),((32,49),((21,22),55),(19,36)))),((3,(37,39),((2,31),(5,((46,35,13,(12,34)),50))),((((14,15),16),26),45)),(42,52))),((27,28),4)),(40,48)));

29 (51,29,24,23,((((((1,(53,54)),25),(18,(20,47),(((32,33),49),(19,36)))),((41,(37,38,39),((2,31),(5,((35,13,(12,34),((6,43),46)),50))),((((14,15),16),26),44,45)),(42,52))),((27,28),4,30)),(40,48)));

30 (51,29,24,23,((((7,17),(((1,(53,54)),25),((20,47),(36,((32,33),49),((21,22),55)))),((41,38,(31,(5,((43,35,13,(12,34)),50))),((14,26),44),(8,3)),(42,52))),(4,28,30)),(40,48)));

31 (51,((((7,17),(((1,(53,54)),25),((20,47),(36,((32,33),49),((21,22),55)))),((41,38,(31,(5,((43,35,13,(12,34)),50))),((14,26),44),(8,3)),(42,52))),(4,28,30)),(40,48)));

32 (51,29,24,23,(((18,(41,(37,38,39),((2,31),(5,((35,13,(12,34),((6,43),46)),50))),((((14,15),16),26),44,45),(3,((10,9),8))),((7,11),17)),((27,28),4,30)),(40,48)));

33 (51,29,23,(40,((7,(((53,1),25),((20,47),((21,22),(32,49),(19,36)))),(42,(41,((2,31),(5,((46,35,(12,34)),50))),((15,16),44,45),(9,3)))),(27,28))));

34 (29,23,((((7,11),(((53,54),25),((20,47),((32,33),(21,22)))),((41,(10,9),(38,39),((13,(6,43),(12,34)),(2,31)),(((14,15),26),45)),(42,52))),(27,28)),(40,48)));

35 (51,29,24,23,((7,(((1,(53,54)),25),(18,(((32,33),49),((21,22),55)))),(((2,31),(5,((35,13,(12,34),((6,43),46)),50))),(42,52))),((27,28),4,30)));

36 (40,((((7,11),17),((1,(53,54)),(18,(20,47),(36,32,(22,55)))),(42,(41,39,((15,16),26),(2,(5,((35,34,(6,46)),50))),(8,3)))),(4,27)));

37 (51,24,(((((1,(53,54)),25),(18,(20,47))),(((37,38,39),(26,44,45),(3,((10,9),8)),(5,((35,((6,43),46)),50))),(42,52))),(40,48)));

```
38  ((((2,31),(5,((35,13,(12,34),((6,43),46)),50))),((((14,15),16),26),44,45)),((20,47),(((32,33),49),((21,22),55),(19,36))));
39  (48,((17,((25,54),(47,(22,19,(33,49))))),((41,39,(14,26),(31,((13,12),50)),(3,(8,9))),(42))),(4,27)));
40  (51,29,23,(40,(28,((7,17),(1,(47,18,(36,(21,55))))),(52,(41,(38,39),(31,(5,((34,13,6),50))),(26,45))))));
41  ((40,48),(((27,28),4,30),((41,(37,39),(31,(5,((34,(6,46)),50))),((14,26),45),(9,3)),(42,52))));
42  (51,29,24,23,(((1,(53,54)),25),(18,(20,47),(((32,33),49),((21,22),55),(19,36)))));
43  ((40,48),((41,((2,31),(5,((35,13),50))),((((14,15),16),26),44,45)),(42,52)));
44  (51,(40,((11,((32,22),(25,54))),(52,(41,39,8,6,(26,45))))),(4,27))));
45  (29,24,23,((((42,52),((7,11),17)),((27,28),30)),(40,48)));
46  (((27,28),4,30),(41,3,(2,31),(26,44,45)));
47  (23,(40,(30,(18,((7,11),17)),(42,31)))));
48  (51,(((27,28),4,30),(40,48)));
49  (((6,43),46),(12,34));
50  (24,(55,(37,(10,9))));
```

References

Adamkewicz, S.L., Harasewych, M.G., Blake, J., Saudek, D., Bult, C.J., 1997. A molecular phylogeny of the bivalve mollusks. Mol. Biol. Evol. 14, 619–629.

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrox, B.N., Caski, F. (Eds.), Second International Symposium on Information Theory. Akademiai Kiado, Budapest, p. 267.

Alfaro, M.E., Huelsenbeck, J.P., 2006. Comparative performance of Bayesian and AIC-based measures of phylogenetic model uncertainty. Syst. Biol. 55, 89–96.

Alfaro, M.E., Zoller, S., Lutzoni, F., 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov Chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. Mol. Biol. Evol. 20, 255–266.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acid Res. 25, 3389–3402.

Amler, M.R.W., Thomas, E., Weber, K.M., 1990. Bivalven des hoechsten Oberdevons im Bergischen land (strunium; noerdliches rheinisches schiefergebirge). Geologica et Palaeontologica 24, 41–63.

Baird, C., Brett, C.E., 1983. Regional variation and paleontology of two coral beds in the Middle Devonian Hamilton Group of Western New York. J. Paleontol. 57, 417–446.

Baker, R.H., DeSalle, R., 1997. Multiple sources of character information and the phylogeny of Hawaiian drosophilids. Syst. Biol. 46, 654–673.

Ballard, J.W.O., Whitlock, M.C., 2004. The incomplete natural history of mitochondria. Mol. Ecol. 13, 729–744.

Bensasson, D., Zhang, D.-X., Hartl, D.L., Hewitt, G.M., 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends Ecol. Evol. 16, 314–321.

Berry, W.B.N., Boucot, A.J., 1973. Correlation of the African Silurian rocks. Geol. Soc. Am. Special Paper 147, 1–83.

Bieler, R., Mikkelsen, P.M., 2006. Bivalvia – a look at the branches. Zool. J. Linn. Soc. 148, 223–235.

Bigot, A., 1935. Les recifs bathoniens de Normandie. Bulletin de la Societe Geologique de France, ser. 5 (4), 697–736.

Birky Jr., C.W., 2001. The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. Annu. Rev. Genet. 35, 125–148.

Brandley, M.C., Schmitz, A., Reeder, T.W., 2005. Partitioned Bayesian analysis, partition choice, and the phylogenetic relationships of scincid lizards. Syst. Biol. 54, 373–390.

Brasier, M.D., Hewitt, R.A., 1978. On the late precambrian – early cambrian Hartshill Formation of Warwickshire. Geol. Mag. 115, 21–36.

Breton, S., Beaupre, H.D., Stewart, D.T., Hoeh, W.R., Blier, P.U., 2007. The unusual system of doubly uniparental inheritance of mtDNA: isn't one enough? Trends Genet. 23, 465–474.

Brett, E., Dick, V.B., Baird, G.C., 1991. Comparative taphonomy and paleoecology of middle Devonian dark gray and black shale facies from western New York. Dynamic stratigraphy and depositional environments of the Hamilton group (middle Devonian) in New York state, Part II. NY State Mus. Bull. 469, 5–36.

Cai, C.Y., Dou, Y.W., Edwards, D., 1993. New observations on a Pridoli plant assemblage from north Xinjiang, northwest China, with comments on its evolutionary and palaeogeographical significance. Geol. Mag. 130, 155–170.

Cameron, S.L., Lambkin, C.L., Barker, S.C., Whiting, M.F., 2007. A mitochondrial genome phylogeny of Diptera: whole genome sequence data accurately resolve relationships over broad timescales with high precision. Syst. Entomol. 32, 40–59.

Campbell, D.C., 2000. Molecular evidence on the evolution of the Bivalvia. In: Harper, E.M., Taylor, J.D., Crame, J.A. (Eds.), The Evolutionary Biology of the Bivalvia. The Geological Society of London, London, pp. 31–46.

Campbell, J.D., Coombs, D.S., Grebneff, A., 2003. Willsher group and geology of the Triassic Kaka Point coastal section, south-east Otago, New Zealand. J. R. Soc. New Zeal. 33, 7–38.

Canapa, A., Barucca, M., Marinelli, A., Olmo, E., 2001. A molecular phylogeny of Heterodonta (Bivalvia) based on small ribosomal subunit RNA sequences. Mol. Phylogenet. Evol. 21, 156–161.

Canapa, A., Marota, I., Rollo, F., Olmo, E., 1999. The small-subunit rRNA gene sequences of venerids and the phylogeny of Bivalvia. J. Mol. Evol. 48, 463–468.

Carter, J.G., 1990. Evolutionary significance of shell microstructure in the Palaeotaxadonta, Pteriomorphia and Isofilibranchia (Bivalvia: Mollusca). In: Carter, J.G. (Ed.), Skeletal Biomineralization: Patterns, Processes and Evolutionary Trends, vol. 1. Van Nostrand Reinhold, New York, pp. 135–296.

Chatterjee, S., 1986. Malerisaurus langstoni, a new diapsid reptile from the Triassic of Texas. J. Vertebr. Paleontol. 6, 297–312.

Clarke, K.R., Warwick, R.M., 1998. A taxonomic distinctness index and its statistical properties. J. Appl. Ecol. 35, 523–531.

Cope, J.C.W., 1996. The early evolution of the Bivalvia. In: Taylor, J.D. (Ed.), Origin and Evolutionary Radiation of the Mollusca. Oxford University Press, Oxford, pp. 361–370.

Cope, J.C.W., 1997. The early phylogeny of the class Bivalvia. Palaeontology 40, 713–746.

Cox, L.R., 1965. Jurassic Bivalvia and Gastropoda from Tanganyika and Kenya. Bull. Br. Mus. (Natural History) Geol. Suppl. I.

Cummings, M.P., Handley, S.A., Myers, D.S., Reed, D.L., Rokas, A., Winka, K., 2003. Comparing bootstrap and posterior probability values in the four-taxon case. Syst. Biol. 52, 477–487.

Distel, D.L., 2000. Phylogenetic relationships among Mytilidae (Bivalvia): 18S rRNA data suggest convergence in mytilid body plans. Mol. Phylogenet. Evol. 15, 25–33.

Dou, Y.W., Sun, Z.H., 1983. . Devonian Plants. Palaeontological Atlas of Xinjiang, vol. II. Late Palaeozoic Section. Geological Publishing House, Beijing.

Dou, Y.W., Sun, Z.H., 1985. On the Late Palaeozoic plants in Northern Xinjiang. Acta Geologica Sinica 59, 1–10.

Douady, C.J., Delsuc, F., Boucher, Y., Ford Doolittle, W., Douzery, E.J.P., 2003. Comparison of Bayesian and Maximum Likelihood bootstrap measures of phylogenetic reliability. Mol. Biol. Evol. 20, 248–254.

Dreyer, H., Steiner, G., Harper, E.M., 2003. Molecular phylogeny of Anomalodesmata (Mollusca: Bivalvia) inferred from 18S rRNA sequences. Zool. J. Linn. Soc. 139, 229–246.

Elder, R.L., 1987. Taphonomy and paleoecology of the Dockum Group, Howard County, Texas. J. Arizona-Nevada Acad. Sci. 22, 85–94.

Eriksson, T., 2007. The r8s bootstrap kit. Distributed by the author.

Erixon, P., Svennblad, B., Britton, T., Oxelman, B., 2003. Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. Syst. Biol. 52, 665–673.

Farris, J.S., Kallersjö, M., Kluge, A.G., Bult, C., 1995a. Constructing a significance test for incongruence. Syst. Biol. 44, 570–572.

Farris, J.S., Kallersjö, M., Kluge, A.G., Bult, C., 1995b. Testing significance of incongruence. Cladistics 10, 315–319.

Felsenstein, J., 1993. PHYLIP: phylogenetic inference package. Distributed by the author.

Folmer, O., Black, M., Hoeh, W.R., Lutz, R., Vrijenhoek, R.C., 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol. Mar. Biol. Biotechnol. 3, 294–299.

Garrido-Ramos, M.S., Stewart, D.T., Sutherland, B.W., Zouros, E., 1998. The distribution of male-transmitted and female-transmitted mitochondrial DNA types in somatic tissues of blue mussels: implications for the operation of doubly uniparental inheritance of mitochondrial DNA. Genome 41, 818–824.

Gatesy, J., DeSalle, R., Wheeler, W., 1993. Alignment-ambiguous nucleotide sites and the exclusion of systematic data. Mol. Phylogenet. Evol. 2, 152–157.

Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. Stat. Sci. 7, 457–511.

Gillham, N.W., 1994. Transmission and compatibility of organelle genomes. In: Gillham, N.W. (Ed.), Organelle Genes and Genomes. Oxford University Press, Oxford, pp. 147–268.

Giribet, G., Carranza, S., 1999. *Point counter point*. What can 18S rDNA do for bivalve phylogeny? J. Mol. Evol. 48, 256–258.

Giribet, G., Distel, D.L., 2003. Bivalve phylogeny and molecular data. In: Lydeard, C., Lindberg, D.R. (Eds.), Molecular Systematics and Phylogeography of Mollusks. Smithsonian Books, Washington, pp. 45–90.

Giribet, G., Wheeler, W., 2002. On bivalve phylogeny: a high-level analysis of the Bivalvia (Mollusca) based on combined morphology and DNA sequence data. Invert. Biol. 121, 271–324.

Goldman, N., Anderson, J.P., Rodrigo, A.G., 2000. Likelihood-based tests of topologies in phylogenetics. Syst. Biol. 49, 652–670.

Goldman, N., Yang, Z., 1994. A codon-based model of nucleotide substitution for protein coding DNA sequences. Mol. Biol. Evol. 11, 725–736.

Gradstein, F.M., Ogg, J.G., Smith, A.G. (Eds.), 2004. A Geologic Time Scale 2004. Cambridge University Press, Cambridge.

Graf, D.L., Ó Foighil, D., 2000. The evolution of brooding characters among the freshwater pearly mussels (Bivalvia: Unionoidea) of North America. J. Moll. Stud. 66, 157–170.

Grasso, T.X., 1986. Redefinition, stratigraphy and depostional environments of the mottville member (Hamilton Group) in central and eastern New York. Dynamic stratigraphy and depositional environments of the Hamilton Group (middle Devonian) in New York state, part I. NY State Mus. Bull. 457, 5–31.

Hammer, Ø., Harper, D.A.T., Ryan, P.D., 2001. PAST: paleontological statistics software package for education and data anlaysis. Palaeontologia Electronica 4, 1–9.

Harper, E.M., Dreyer, H., Steiner, G., 2006. Reconstructing the Anomalodesmata (Mollusca: Bivalvia): morphology and molecules. Zool. J. Linn. Soc. 148, 395–420.

Hartmann, S., Vision, T.J., 2008. Using ESTs for phylogenomics: can one accurately infer a phylogenetic tree from a gappy alignment? BMC Evol. Biol. 8, 95.

Hayami, I., 1975. A systematic survey of the Mesozoic Bivalvia from Japan. The University Museum, The University of Tokyo. Bulletin 10..

He, T., Pei, F., Fu, G., 1984. Some small shelly fossils from the Lower Cambrian Xinji Formation in Fangcheng County, Henan Province. Acta Palaeontologica Sinica 23, 350–355.

Heckert, B., 2004. Late Triassic microvertebrates from the lower Chinle Group (Otischalkian-Adamanian: Carnian). Southwestern USA New Mexico Mus. Nat. History Sci. Bull. 27, 1–170.

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17, 754–755.

Huelsenbeck, J.P., Hillis, D.M., Jones, R., 1996a. Parametric bootstrapping in molecular phylogenetics: applications and performance. In: Ferraris, J.D., Palumbi, S.R. (Eds.), Molecular zoology: Advances, Strategies and Protocols. Wiley and Sons, New York, pp. 19–45.

Huelsenbeck, J.P., Hillis, D.M., Nielsen, R., 1996b. A likelihood-ratio test of monophyly. Syst. Biol. 45, 546–558.

Huelsenbeck, J.P., Larget, B., Alfaro, M.E., 2004. Bayesian phylogenetic model selection using reversible jump markov chain Monte Carlo. Mol. Biol. Evol. 21, 1123–1133.

Huson, D.H., Richter, D.C., Rausch, C., Dezulian, T., Franz, M., Rupp, R., 2007. Dendroscope – an interactive viewer for large phylogenetic trees. BMC Bioinform. 8, 460.

Jordan, G.E., Piel, W.H., 2008. PhyloWidget: web-based visualizations for the tree of life. Bioinformatics 15, 1641–1642.

Jozefowicz, C.J., Ó Foighil, D., 1998. Phylogenetic analysis of Southern Hemisphere flat oysters based on partial mitochondrial 16S rDNA gene sequences. Mol. Phylogenet. Evol. 10, 426–435.

Kappner, I., Bieler, R., 2006. Phylogeny of Venus clams (Bivalvia: Venerinae) as inferred from nuclear and mitochondrial gene sequences. Mol. Phylogenet. Evol. 40, 317–331.

Kass, R.E., Raftery, A.E., 1995. Bayes Factors. J. Am. Stat. Assoc. 90, 773–795.

Kemp, J., 1976. Account of excavations into the campanile bed (Eocene, Selsey Formation) at Stubbington, Hants. Tert. Res. 1, 41–45.

Kirkendale, L., Lee, T., Baker, P., Ó Foighil, D., 2004. Oysters of the Conch Republic (Florida Keys): a molecular phylogenetic study of *Parahyotissa mcgintyi*, *Teskeyostrea weberi* and *Ostreola equestris*. Malacologia 46, 309–326.

Kocher, T.D., Thomas, W.K., Meyer, A., Edwards, S.V., Pääbo, S., Villablanca, F.X., Wilson, A.C., 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc. Natl. Acad. Sci. USA 86, 6196–6200.

Kříž, J., 1999. Bivalvia communities of Bohemian type from the Silurian and lower Devonian carbonate facies. In: Boucot, A.J., Lawson, J.D. (Eds.), Paleocommunities – A Case Study from the Silurian and Lower Devonian. Cambridge University Press, Cambridge, pp. 252–299.

Lambkin, C.L., Lee, M.S.Y., Winterton, S.L., Yeates, D.K., 2002. Partitioned bremer support and multiple trees. Cladistics 18, 436–444.

Larson, A., 1994. The comparison of morphological and molecular data in phylogenetic systematics. In: Schierwater, B., Streit, B., Wagner, G., DeSalle, R. (Eds.), Molecular Ecology and Evolution. Birkhäuser Verlag, Basel, pp. 371–390.

Laudon, L.R., 1931. The Stratigraphy of the Kinderhook Series of Iowa. Iowa Geological Survey 35, 333–452.

Leaché, A.D., Reeder, T.W., 2002. Molecular systematics of the Eastern fence lizard (*Sceloporus undulatus*): a comparison of parsimony, likelihood, and Bayesian approaches. Syst. Biol. 51, 44–68.

Lee, M.S.Y., Hugall, A.F., 2003. Partitioned likelihood support and the evaluation of data set conflict. Syst. Biol. 52, 15–22.

Lehman, T.M., Chatterjee, S., 2005. Depositional setting and vertebrate biostratigraphy of the Triassic Dockum Group of Texas. J. Earth Sys. Sci. 114, 325–351.

Li, B., Dettaï, A., Cruaud, C., Couloux, A., Desoutter-Meniger, M., Lecointre, G., 2009. RNF213, a new nuclear marker for acanthomorph phylogeny. Mol. Phylogenet. Evol. 50, 345–363.

Littlewood, D.T.J., 1994. Molecular phylogenetics of cupped oysters based on partial 28S rRNA gene sequences. Mol. Phylogenet. Evol. 3, 221–229.

Lutzoni, F., Wagner, P., Reeb, V., Zoller, S., 2000. Integrating ambiguously aligned regions of DNA sequences in phylogenetic analyses without violating positional homology. Syst. Biol. 49, 628–651.

Manten, A., 1971. Silurian reefs of gotland. Developments in Sedimentology 13, 1–539.

Maruyama, T., Ishikura, M., Yamazaki, S., Kanai, S., 1998. Molecular phylogeny of zooxanthellate bivalves. Biol. Bull. 195, 70–77.

Matsumoto, M., 2003. Phylogenetic analysis of the subclass Pteriomorphia (Bivalvia) from mtDNA COI sequences. Mol. Phylogenet. Evol. 27, 429–440.

Mergl, M., Massa, D., 1992. Devonian and Lower Carboniferous brachiopods and bivalves from western Libya. Biostratigraphie du Paleozoique 12, 1–115.

Merritt, T.J., Shi, L., Chase, M.C., Rex, M.A., Etter, R.J., Quattro, J.M., 1998. Universal cytochrome b primers facilitate intraspecific studies in molluscan taxa. Mol. Mar. Biol. Biotechnol. 7, 7–11.

Mikkelsen, P.M., Bieler, R., Kappner, I., Rawlings, T.A., 2006. Phylogeny of veneroidea (Mollusca: Bivalvia) based on morphology and molecules. Zool. J. Linn. Soc. 148, 439–521.

Millard, V., 2001. Classification of Mollusca: A Classification of World Wide Mollusca, vol. 3, second ed.. Printed by the author, South Africa. pp. 915–1447.

Morton, B., 1996. The evolutionary history of the Bivalvia. In: Taylor, J.D. (Ed.), Origin and Evolutionary Radiation of the Mollusca. Oxford University Press, Oxford, pp. 337–359.

Murry, P.A., 1989. Geology and paleontology of the Dockum formation (upper triassic), west Texas and eastern New Mexico. In: Lucas, S.G., Hunt, A.P. (Eds.), Dawn of the Age of Dinosaurs in the American Southwest. New Mexico Museum of Natural History, Albuquerque, pp. 102–144.

Muse, S.V., Gaut, B.S., 1994. A likelihood approach for comparing synonymous and nonsynonymous substitution rates, with application to the chloroplast genome. Mol. Biol. Evol. 11, 715–724.

Nielsen, R., Yang, Z., 1998. Likelihood models for detecting positively selected amino acids sites and applications to the HIV-1 envelope gene. Genetics 148, 929–936.

Nuin, P., 2008. MrMTgui: cross-platform interface for ModelTest and MrModeltest. <http://www.genedrift.org/mtgui.php>.

Nylander, J.A.A., Wilgenbusch, J.C., Warren, D.L., Swofford, D.L., 2008. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. Bioinformatics 24, 581–583.

Nylander, J.A.A., Ronquist, F., Huelsenbeck, J.P., Nieves-Aldrey, J.L., 2004. Bayesian phylogenetic analysis of combined data. Syst. Biol. 53, 47–67.

Ó Foighil, D., Taylor, D.J., 2000. Evolution of parental care and ovulation behavior in oysters. Mol. Phylogenet. Evol. 15, 301–313.

Ogg, J.G., Ogg, G., Gradstein, F.M., 2008. The Concise Geologic Time Scale. Cambridge University Press, Cambridge.

Olu-Le Roy, K., von Cosel, R., Hourdez, S., Carney, S.L., Jollivet, D., 2007. Amphi-Atlantic cold-seep *Bathymodiolus* species complexes across the equatorial belt. Deep-Sea Res. Pt. I 54, 1890–1911.

Palmer, T.J., 1979. The hampen marly and white limestones formations: Florida-type carbonate lagoons in the jurassic of central England. Palaeontology 22, 189–228.

Palumbi, S.R., Martin, A., Romano, S., McMillan, W.O., Stice, L., Grabowski, G., 1996. The simple fool's guide to PCR. Kewalo Marine Laboratory and University of Hawaii, Hawaii.

Parker, S.R., 1997. Sequence Navigator. Multiple sequence alignment software. Methods Mol. Biol. 70, 145–154.

Parkhaev, P.Y.U., 2004. Malacofauna of the Lower Cambrian Bystraya formation of eastern Transbaikalia. Paleontol. J. 38, 590–608.

Passamaneck, Y.J., Schander, C., Halanych, K.M., 2004. Investigation of molluscan phylogeny using large-subunit and small-subunit nuclear rRNA sequences. Mol. Phylogenet. Evol. 32, 25–38.

Passamonti, M., 2007. An unusual case of gender-associated mitochondrial DNA heteroplasmy: the mytilid *Musculista senhousia* (Mollusca Bivalvia). BMC Evol. Biol. 7, S7.

Passamonti, M., Boore, J.L., Scali, V., 2003. Molecular evolution and recombination in gender-associated mitochondrial DNAs of the Manila clam *Tapes philippinarum*. Genetics 164, 603–611.

Passamonti, M., Ghiselli, F., 2009. Doubly uniparental inheritance. Two mitochondrial genomes, one precious model for organelle DNA inheritance and evolution. DNA Cell Biol. 28, 1–10.

Plazzi, F., Ferrucci, R.R., Passamonti, M., 2010. Phylogenetic representativeness: a new method for evaluating taxon sampling in evolutionary studies. BMC Bioinform. 11, 209.

Plohl, M., Luchetti, A., Meštrović, N., Mantovani, B., 2008. Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. Gene 409, 72–82.

Posada, D., Buckley, T.R., 2004. Model selection and model averaging in phylogenetics: advantages of akaike information criterion and Bayesian approaches over likelihood ratio test. Syst. Biol. 53, 793–808.

Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Poulton, T.P., 1991. Hettangian through aalenian (jurassic) guide fossils and biostratigraphy, northern Yukon and adjacent Northwest Territories. Geol. Surv. Can. Bull. 410, 1–95.

Purchon, R.D., 1987b. Classification and evolution of the Bivalvia: an analytical study. Phil. Trans. R. Soc. Lond. B 316, 277–302.

Puslednik, L., Serb, J.M., 2008. Molecular phylogenetics of the Pectinidae (Mollusca: Bivalvia) and effect of increased taxon sampling and outgroup selection on tree topology. Mol. Phylogenet. Evol. 48, 1178–1188.

Rode, L., Lieberman, B.S., 2004. Using GIS to unlock the interactions between biogeography, environment, and evolution in middle and Late Devonian brachiopods and bivalves. Palaeogeogr. Palaeocl. 211, 345–359.

Ronquist, F., Huelsenbeck, J.P., 2003. MRBAYES 3: Bayesian phylogenetic inference using mixed models. Bioinformatics 19, 1572–1574.

Ronquist, F., Huelsenbeck, J.P., van der Mark, P., 2005. MrBayes 3.1 Manual. Draft 5/26/2005. Distributed with the software.

Salvini-Plawen, L., Steiner, G., 1996. Synapomorphies and plesiomorphies in higher classification of Mollusca. In: Taylor, J.D. (Ed.), Origin and Evolutionary Radiation of the Mollusca. Oxford University Press, Oxford, pp. 29–51.

Samtleben, C., Munnecke, A., Bickert, T., Paetzold, J., 1996. The Silurian of Gotland (Sweden): facies interpretation based on stable isotopes in brachiopod shells. Geologische Rundschau 85, 278–292.

Sanderson, M.J., 2003. R8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics 19, 301–302.

Schneider, J.A., Ó Foighil, D., 1999. Phylogeny of giant clams (Cardiidae: Tridacninae) based on partial mitochondrial 16S rDNA gene sequences. Mol. Phylogenet. Evol. 13, 59–66.

Shilts, M.H., Pascual, M.S., Ó Foighil, D., 2007. Systematic, taxonomic and biogeographic relationships of Argentine flat oysters. Mol. Phylogenet. Evol. 44, 467–473.

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114–1116.

Simmons, M.P., Ochoterena, H., 2000. Gaps as characters in sequence-based phylogenetic analyses. Syst. Biol. 49, 369–381.

Simmons, M.P., Pickett, K.M., Miya, M., 2004. How meaningful are Bayesian support values? Mol. Biol. Evol. 21, 188–199.

Simon, C., Buckley, T.R., Frati, F., Stewart, J.B., Beckenbach, A.T., 2006. Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. Annu. Rev. Ecol. Evol. Syst. 37, 545–579.

Sorenson, M.D., Franzosa, E.A., 2007. TreeRot, version 3. Boston University, Boston, Massachusetts, USA.

Spath, L.F., 1930. The Eotriassic invertebrate fauna of East Greenland. Meddeleser om Grønland 83, 1–90.

Starobogatov, Y.I., 1992. Morphological basis for the phylogeny and classification of Bivalvia. Ruthenica 2, 1–26.

Steiner, G., Hammer, S., 2000. Molecular phylogeny of the Bivalvia inferred from 18S rDNA sequences with particular reference to the Pteriomorphia. In: Harper, E.M., Taylor, J.D., Crame, J.A. (Eds.), The Evolutionary Biology of the Bivalvia. The Geological Society of London, London, pp. 11–29.

Steiner, G., 1999. Point counter point. What can 18S rDNA do for bivalve phylogeny? J. Mol. Evol. 48, 258–261.

Steiner, G., Müller, M., 1996. What can 18S rDNA do for bivalve phylogeny? J. Mol. Evol. 43, 58–70.

Strugnell, J., Norman, M., Jackson, J., Drummond, A.J., Cooper, A., 2005. Molecular phylogeny of coleoid cephalopods (Mollusca: Cephalopoda) using a multigene approach; the effect of data partitioning on resolving phylogenies in a Bayesian framework. Mol. Phylogenet. Evol. 37, 426–441.

Suarez Soruco, R., 1976. El sistema ordovicico en Bolivia. Revista Tecnica YPF Bolivia 5, 111–123.

Sullivan, J., Swofford, D.L., Naylor, G.J.P., 1999. The effect of taxon sampling on estimating rate heterogeneity parameters of maximum-likelihood models. Mol. Biol. Evol. 16, 1347–1356.

Suzuki, Y., Glazko, G.V., Nei, M., 2002. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. Proc. Natl. Acad. Sci. USA 99, 16138–16143.

Swofford, D., 1999. PAUP*: Phylogenetic Analysis Using Parsimony (* and other methods). Sinauer Associates, Sunderland.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol 24, 1596–1599.

Tavaré, S., 1986. Some probabilistic and statistical problems on the analysis of DNA sequences. Lect. Mathemat. Life Sci. 17, 57–86.

Taylor, J.D., Williams, S.T., Glover, E.A., 2007a. Evolutionary relationships of the bivalve family Thyasiridae (Mollusca: Bivalvia), monophyly and superfamily status. J. Mar. Biol. Ass. UK 87, 565–574.

Taylor, J.D., Williams, S.T., Glover, E.A., Dyal, P., 2007b. A molecular phylogeny of heterodont bivalves (Mollusca: Bivalvia: Heterodonta): new analyses of 18S and 28S rRNA genes. Zool. Scr. 36, 587–606.

Templeton, A., 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. Evolution 37, 221–244.

Theologidis, I., Fodelianakis, S., Gaspar, M.B., Zouros, E., 2008. Doubly uniparental inheritance (DUI) if mitochondrial DNA in *Donax trunculus* (Bivalvia: Donacidae) and the problem of its sporadic detection in Bivalvia. Evol. Int. J. Org. Evol. 62, 959–970.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

Townsend, J.P., 2007. Profiling Phylogenetic Informativeness. Syst. Biol. 56, 222–231.

Vokes, H.E., 1980. Genera of the Bivalvia. Palaeontological Research Institution, Ithaca.

Wagner, P.J., 2008. Paleozoic Gastropod, Monoplacophoran and Rostroconch Database. Paleobiology Database Online Systematics Archives 6.

Waller, T.R., 1990. The evolution of ligament systems in the Bivalvia. In: Morton, B. (Ed.), The Bivalvia. Hong Kong University Press, Hong Kong, pp. 49–71.

Waller, T.R., 1998. Origin of the molluscan class Bivalvia and a phylogeny of major groups. In: Johnston, P.A., Haggart, J.W. (Eds.), Bivalves: An Eon of Evolution. University of Calgary Press, Calgary, pp. 1–45.

Wheeler, W., 1999. Fixed character state and the optimization of molecular sequence data. Cladistics 15, 379–385.

Wheeler, W., Gatesy, J., DeSalle, R., 1995. Elision: a method for accommodating multiple molecular sequence alignments with alignment-ambiguous sites. Mol. Phylogenet. Evol. 4, 1–9.

Whittingham, L.A., Slikas, B., Winkler, D.W., Sheldon, F.H., 2002. Phylogeny of the tree swallow genus, *Tachycineta* (Aves: Hirundinidae), by Bayesian analysis of mitochondrial DNA sequences. Mol. Phylogenet. Evol. 22, 430–441.

Wiens, J.J., 1998. Combining data sets with different phylogenetic histories. Syst. Biol. 47, 568–581.

Williams, S.T., Taylor, J.D., Glover, E.A., 2004. Molecular phylogeny of the Lucinoidea (Bivalvia): non-monophyly and separate acquisition of bacterial chemosymbiosis. J. Moll. Stud. 70, 187–202.

Winnepenninckx, B., Backeljau, T., De Wachter, R., 1996. Investigation of molluscan phylogeny on the basis of 18S rRNA sequences. Mol. Biol. Evol. 13, 1306–1317.

Young, N.D., Healy, J., 2003. GapCoder automates the use of indel characters in phylogenetic analysis. BMC Bioinform. 4, 6.

Zbawicka, M., Burzyński, A., Wenne, R., 2007. Complete sequences of mitochondrial genomes from the Baltic mussel *Mytilus trossulus*. Gene 406, 191–198.

Molecular Phylogenetics and Evolution 58 (2011) 304–316

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

The mitochondrial genome of *Bacillus* stick insects (Phasmatodea) and the phylogeny of orthopteroid insects

Federico Plazzi *, Andrea Ricci, Marco Passamonti

Department of Biologia Evoluzionistica Sperimentale, University of Bologna, via Selmi 3, 40126 Bologna, Italy

ARTICLE INFO

Article history:
Received 28 July 2010
Revised 6 December 2010
Accepted 11 December 2010
Available online 16 December 2010

Keywords:
mtDNA
Phasmatodea
Bacillus
Phylogeny
Phylogenetic informativeness

ABSTRACT

The Order Phasmatodea (stick and leaf insects) includes many well-known species of cryptic phytophagous insects. In this work, we sequenced the almost complete mitochondrial genomes of two stick insect species of the genus *Bacillus*. Phasmatodea pertain to the Polyneoptera, and represent one of the major clades of heterometabolous insects. Orthopteroid insect lineages arose through rapid evolutionary radiation events, which likely blurred the phylogenetic reconstructions obtained so far; we therefore performed a phylogenetic analysis to resolve and date all major splits of orthopteroid phylogeny, including the relationships between Phasmatodea and other polyneopterans. We explored several molecular models, with special reference to data partitioning, to correctly detect any phylogenetic signal lying in rough data. Phylogenetic Informativeness analysis showed that the maximum resolving power on the orthopteroid mtDNA dataset is expected for the Upper Cretaceous, about 80 million years ago (Mya), but at least 70% of the maximum informativeness is also expected for the 150–200 Mya timespan, which makes mtDNA a suitable marker to study orthopteroid splits. A complete chronological calibration has also been computed following a Penalized Likelihood method. In summary, our analysis confirmed the monophyly of Phasmatodea, Dictyoptera and Orthoptera, and retrieved Mantophasmatodea as sister group of Phasmatodea. The origin of orthopteroid insects was also estimated to be in the Middle Triassic, while the order Phasmatodea seems to appear in the Upper Jurassic. The obtained results evidenced that mtDNA is a suitable marker to unravel the ancient splits leading to the orthopteroid orders, given a proper methodological approach.

1. Introduction

Insects (Insecta) are among the most diverse and successful terrestrial organisms, showing a great variety of shapes and life habits. Commonly, they are subdivided into two main lineages: Palaeoptera and Neoptera. The monophyly of Palaeoptera, which comprise, among the others, ephemerids, dragonflies and damselflies, has been sometimes contentious (see Wheeler et al., 2001; Whitfield and Kjer, 2008; and references therein), while Neoptera are always acknowledged as a monophyletic taxon (Wheeler et al., 2001; and references therein).

Among neopteran insects, Martynov (1925) first introduced a group named Polyneoptera, further partitioned into Blattopteroidea (nowadays known as Dictyoptera) and Orthopteroidea. The Polyneoptera, collectively referred to as "orthopteroid insects" (Bradler, 2009; Terry and Whiting, 2005; Wheeler et al., 2001), are the outcome of an ancient evolutionary radiation, leading to a heterogeneous assemblage, displaying many forms and adaptations, and about one third of the total insect diversity at the ordinal level (Terry and Whiting, 2005). They include Blattodea (roaches), Dermaptera (earwigs), Embiidina (web-spinners), Grylloblattodea (ice crawlers), Isoptera (termites), Mantodea (praying mantises), Orthoptera (grasshoppers and crickets), Plecoptera (stoneflies), Zoraptera (angel insects), and Phasmatodea (stick and leaf insects). Recently, a new polyneopteran order has been discovered and named Mantophasmatodea (gladiators) (Klass et al., 2002; Zompro, 2001). Although the monophyly of Polyneoptera is widely acknowledged by most studies (Bradler, 2009; Grimaldi and Engel, 2005; Gullan and Cranston, 2005; Wheeler et al., 2001; Willmann, 2004), others do not accept it (see Haas and Kukalová-Peck, 2001; and references therein); moreover, molecular data do not always retrieve Polyneoptera as monophyletic (Cameron et al., 2006a; Kjer, 2004; Kjer et al., 2006; Terry and Whiting, 2005; Whitfield and Kjer, 2008; Whiting, 2002).

Phylogenetic relationships within Polyneoptera are also quite controversial (Bradler, 2009; Flook and Rowell, 1998; Ishiwata et al., 2010; Wheeler et al., 2001; and references therein).

Abbreviations: AIC, Akaike Information Criterion; BF, Bayes Factor; EML, Estimated Marginal Likelihood; ML, Maximum Likelihood; mtDNA, mitochondrial DNA; Mya, million years ago; Myr, million years; PCG, protein-coding gene; PL, Penalized Likelihood.

* Corresponding author. Fax: +39 051 2094286.
E-mail address: federico.plazzi@unibo.it (F. Plazzi).

1055-7903/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2010.12.005

Boudreaux (1987) placed Embiidina + Plecoptera as a sister group to Orthopterodida, a clade including the remaining polyneopteran orders. Similarly, Hennig (1981) considered the Plecoptera basal to the newly erected group Paurometabola, composed by Embiidina as the sister group of Orthopteromorpha – again, a group which included all the remaining orders. However, many synapomorphies defining either Orthopterodida or Paurometabola were disputed (Bradler, 2009; Flook and Rowell, 1998; Kristensen, 1981; and references therein). Kristensen (1995) pointed out the lack of resolution of polyneopteran clades, which are rather a big polytomy (only Dictyoptera were retrieved as a monophyletic clade); this scenario was further embraced by Brusca and Brusca (2003) and Whitfield and Kjer (2008). Moreover, other questioned subgroups were proposed: Dictyoptera, joining termites, cockroaches and mantises (Boudreaux, 1987; Kambhampati, 1995; Kristensen, 1981; Kukalová-Peck and Peck, 1993; Thorne and Carpenter, 1992); Eukinolabia, joining Embiidina and Phasmatodea within Orthopteroidea, Haplocercata, joining earwigs and angel insects, and Xenonomia, joining ice crawlers and gladiators (all by Terry and Whiting, 2005). On the other hand, two main polyneopteran lineages are generally undisputed: one, called either Blattopteroidea (Hennig, 1981; Martynov, 1925) or Blattiformida (Boudreaux, 1987), includes most of the orders; the other, called either Orthopteroidea (Hennig, 1981) or Grylliformida (Boudreaux, 1987), includes Orthoptera and Phasmida. Further evidences led to broaden Orthopteroidea, to include Embiidina (Kukalová-Peck, 1991; Rähle, 1970; Terry and Whiting, 2005; Thomas et al., 2000; Whiting et al., 2003).

Finally, the phylogenetic placement of Phasmatodea is remarkably unstable, although, as mentioned, stick and leaf insects are included in Orthopteroidea *sensu* Hennig. Quite remarkably, Phasmatodea were hypothesized as sister group of essentially each given order within Polyneoptera (see Bradler, 2009, for an in-depth discussion on the issue). Embiidina and Orthoptera were the favorite candidates in recent years (Beutel and Gorb, 2006; Engel and Grimaldi, 2000, 2004; Grimaldi and Engel, 2005; Terry and Whiting, 2005; Wheeler et al., 2001; Whiting et al., 2003; Willmann, 2003) and the sister group relationship Embiidina + Phasmatodea (Eukinolabia *sensu* Terry and Whiting) is nowadays the most likely scenario (Beutel and Gorb, 2001; Bradler, 2003, 2009; Klug and Bradler, 2006; Ishiwata et al., 2010; Willmann, 2004).

In this paper, we target polyneopteran insect phylogeny on molecular basis, attempting to disentangle the above mentioned intricate crossing of hypotheses. We also gave special emphasis to the phylogenetic relationships of Phasmatodea. Mitochondrial DNA (mtDNA) was our marker of choice, because it is one of the most information-rich molecule in phylogenetics, it its relatively small (about 15,000 bp) and it has an almost constant gene content (37 genes). MtDNAs may differ in both nucleotide sequence and the relative position of genes within the molecule (i.e. the gene order), a character that has been profitably used as a phylogenetic marker. Unfortunately, however, Cameron et al. (2006b) clearly showed that a phylogenetic approach based on mtDNA gene order is not applicable to higher insect phylogeny, because this marker turned out to be very conservative, with most insects showing the same plesiomorphic pancrustacean groundplan (Boore et al., 1998). Therefore, sequence-based insect phylogenies are quite common (Bae et al., 2004; Cameron et al., 2004, 2006a, 2007, 2009; Dowton et al., 2009; Feng et al., 2010; Fenn et al., 2008; Flook and Rowell, 1998; Ishiwata et al., 2010; Kjer et al., 2006; Komoto et al., 2011; Nardi et al., 2001, 2003; Terry and Whiting, 2005; Wheeler et al., 2001; Whitfield and Kjer, 2008; Whiting et al., 2003). Moreover, many studies addressed the usefulness and resolving power of mitochondrial genome sequence data, and this literature especially flourished for insects and relatives (Cameron et al., 2004, 2007; Carapelli et al., 2007; Kjer and Honeycutt, 2007). These results highlighted the need of a rigorous evaluation of phylogenetic signals carried by the mitochondrial genome, to improve confidence limits of the obtained phylogenies, which should be reflective of real evolutionary histories, rather than of analytical artifacts. Different strategies of data inclusion/exclusion have been tested, from selecting some genes along the molecule to traditionally analyzing amino acid sequences (reviewed in Cameron et al., 2006b), through including all available genes, but not the control region (Castro and Dowton, 2007), or purine/pyrimidine coding (Delsuc et al., 2003). Moreover, given the complexity of mitochondrial genome data, optimality criteria and dataset compilation techniques have been explored (Cameron et al., 2004, 2007; Castro and Dowton, 2005; Kim et al., 2005; Kjer et al., 2006; Stewart and Beckenbach, 2003). For instance, Cameron et al. (2007) found out that mitochondrial genome data recover the best phylogenetic signal when all available genes are analyzed as nucleotide sequences, and different optimality criteria are used and critically evaluated. In any case, although quickly sequencing whole insect mitochondrial genomes is now a routine, questions still remain on how analyze the data at the best.

Sometimes, molecular studies facing with deeper nodes of insect phylogeny show little branch support (Whitfield and Kjer, 2008). A possible cause for this lays in rapid evolutionary radiation events, since they would result in a short divergence time for diagnostic mutations to occur. However, as noted by Whitfield and Kjer (2008), these could be easily darkened or misunderstood when poor data quality is present: it is therefore important to test whether the available data are appropriate to resolve the relationships at the given taxonomical level, and to determine eventual data biases interfering with phylogenetic signal detection. While phylogenies were efficiently resolved by mitochondrial genome data for splits ranked below the order level, as it was for Diptera (Cameron et al., 2007) and Hymenoptera (Dowton et al., 2009), more ancient splits were recovered as ambiguous and somewhat unstable (Cameron et al., 2004, 2006a; Kjer et al., 2006). Because events dating back to the Upper Triassic (225 Mya) were completely resolved, while ancient Cambrian to Devonian splits (600–360 Mya) were not, the "maximum resolving power" of complete insect mtDNA datasets might lie somewhere between these two boundaries (Fenn et al., 2008). Here we report the nearly complete mitochondrial genomes of two *Bacillus* species (*Bacillus atticus* and *Bacillus rossius*). We compared the *Bacillus* mtDNAs to the mitochondrial genome of *T. californicum* (suborder Timematodea), which is the earliest diverging stick insect (Whiting et al., 2003). The two *Bacillus* mitochondrial genomes reported here add samples to the phasmatodeans mtDNA dataset, being representatives from the Verophasmatodea suborder. The obtained results evidenced that mtDNA is a suitable marker to unravel the ancient splits leading to polyneopteran orders, given the proper methodological approach.

2. Material and methods

2.1. Sampling and mitochondrial DNA sequencing

Stick insects *B. rossius* and *B. atticus* were collected from Sardinia (Siniscola) and Israel (Golan), respectively. Field-collected specimens were stored at −80 °C. Total genomic DNA was isolated from somatic tissues with a standard phenol–chloroform protocol.

The almost complete mtDNA sequences of both *Bacillus* species were obtained in four partially overlapping mtDNA pieces via PCR using universal primers: (i) a fragment of *rrnS* gene (543 bp) was amplified using the pair of primers SR-J14197/SR-N14745 (Simon et al., 2006) via standard PCR and directly sequenced; (ii) the region from *nad2* to *cox1* genes (2100 bp) was amplified with

primers TM-J210 (Simon et al., 1994) and C1-N2329 (Simon et al., 2006) via Long PCR and directly sequenced using "primer walking" method; (iii) finally, two major fragments including the rest of the mitochondrial genome (9.0 kb and 5.5 kb) were amplified using C1-J-2195/CB-N-11367 and N4-J-8944/LR-N primers (Simon et al., 1994), respectively.

Normal PCRs were performed in a 50 μl reaction mixture consisting of 27.5 μl of sterilized water, 3 μl $MgCl_2$ 50 mM, 5 μl 10× PCR Buffer, 4 μl dNTP 2.5 mM, 2.5 μl of each primer 10 μM, 5 μl DNA template (25–50 ng), and 0.5 μl Takara Taq DNA polymerase: initial denaturation was set to 2 min at 94 °C, followed by 30 cycles of 30 s at 94 °C, 30 s at 52 °C, and 60 s at 72 °C, and a subsequent 7 min final extension step at 72 °C. Long PCR amplifications were carried out in 50 μl reaction volume composed of 31.5 μl of sterilized water, 10 μl of 5X Herculase II Fusion Reaction Buffer, 0.5 μl of dNTPs mix, 1.25 μl of each primer 10 μM, 5 μl of DNA template (25–50 ng) and 0.5 μl of Herculase II Fusion DNA Polymerase. Reaction conditions were according to supplier's recommendations: the mix was heated at 95 °C for 5 min and then incubated at 95 °C for 20 s, 50 °C for 20 s, and 68 °C for 10 min for 30 cycles and 68 °C for 8 min for a final extension. Both normal and Long PCR were performed using Gene Amp® PCR System 2720 (Applied Biosystem). PCR fragments were purified using Wizard® SV Gel and PCR Clean-Up System (Promega).

Sequencing of the two major fragments was done using a shotgun approach. Amplicons were randomly sheared to 1.2–1.5 kb DNA segments using a HydroShear device (GeneMachines). Sheared DNA was blunt end-repaired at room temperature for 60 min using 6 U of T4 DNA Polymerase (Roche), 30 U of DNA Polymerase I Klenow (NEB), 10 μl of dNTPs mix, 13 μl of 10 × NEB buffer 2 (NEB) in a 115 μl total volume and gel purified using the Wizard® SV Gel and PCR Clean-Up System (Promega). Resulting fragments were ligated into the SmaI site of a pUC18 cloning vector using the Fast-Link DNA ligation Kit (Epicentre) and electroporated into One Shot® TOP10 Electrocomp™ *E. coli* cells (Invitrogen) using standard protocols. Recombinant clones were screened by PCR using M13 universal primers. Obtained recombinant colonies were purified using Multiscreen (Millipore) according to the manufacturer's instructions. Clones were sequenced using M13 universal primers. All sequencing reactions were performed through Macrogen (World Meridian Center, Seoul, South Korea) facility. Raw sequences were manually corrected and assembled into contigs with the software Sequencher 4.6 (Gene Codes); final assemblies were based on a minimum sequence coverage of 3×.

2.2. mtDNA sequence analysis

The tRNA genes were identified by their secondary structure using tRNA-scan SE 1.21 (Lowe and Eddy, 1997) with invertebrate mitochondrial codon predictors and a cove score cut off of 1. ARWEN 1.2.3 (Laslett and Canbäck, 2008) was used to confirm tRNA-scan SE results and draw secondary structures. Open reading frames were found using ORF Finder and identified using translated BLAST searches (blastx; Altschul et al., 1997) as both implemented by the NCBI website (http://www.ncbi.nlm.nih.gov/).

To infer phylogenetic position of Verophasmatodea within pterygote insects, mtDNA sequences of 12 additional insect species were obtained from GenBank (Table 1); among them, two apterygotes, a bristletail (*Nesomachilis australica*) and a silverfish (*Tricholepidion gertschi*), were used as outgroup taxa. Annotated mitochondrial genomes were organized using MEGA 4.0 (Tamura et al., 2007) with each gene aligned separately. Protein-coding genes were translated into amino acid sequences using the invertebrate mitochondrial genetic code, and aligned using default settings in ClustalW (Thompson et al., 1994). The alignment was back-translated into the corresponding nucleotide sequences. Ribosomal and transfer RNA genes were aligned individually with MAFFT multiple sequence alignment tool (Katoh et al., 2002) available online at http://align.bmr.kyushu-u.ac.jp/mafft/online/server. Q-INS-i (Katoh and Toh, 2008) algorithm was chosen for ribosomal and transfer genes because it accounts for secondary structure. Moreover, ambiguously aligned regions in ribosomal genes were identified and excluded from the analysis through Gblocks 0.91b (Talavera and Castresana, 2007; Castresana, 2000) with the following parameters: minimum number of sequences for a conserved position, 10; minimum number of sequences for a flanking position, 10; maximum number of contiguous nonconserved positions, 22; minimum length of a block, 20; allowed gap positions, all. Finally, alignments were manually optimized and concatenated.

We coded indels following the rules given by Simmons and Ochoterena (2000) and implemented in the software GapCoder (Young and Healy, 2003): each indel is considered as a whole and coded at the end of the nucleotide matrix as present/absent (i. e. 1/0). Whenever a longer indel completely overlaps another across two sequences, it is meaningless to wonder whether the shorter indel is present or not in the sequence presenting the longer one. Therefore, the shorter indel is coded among missing data in that sequence. Finally, a saturation analysis (Xia et al., 2003) was performed on protein-coding genes using DAMBE 5.0.39 (Xia and Xie, 2001). Partitioning schemes used in this study are 33, based on 122 different partitions (Supplementary material Tables 1 and 2), although they are not all the conceivable ones. The Bayesian Information Criterion (BIC) implemented in ModelTest 3.7 (Posada and Crandall, 1998) was used to select best-fitting evolutionary model for each partition; the graphical interface provided by MrMTgui was used (Nuin, 2008).

ML analysis was carried out with PAUP* software (Swofford, 1999) at the University of Oslo BioPortal (http://www.bioportal.uio.no). Given software's limitations, the concatenated alignment was not partitioned and binary data were not included; gap characters were treated as missing data. Nucleotides frequencies, substitution rates, gamma shape parameter and proportion of invariable sites were set according to ModelTest results on global alignment. Outgroups were set to be paraphyletic to the monophyletic ingroup. Bootstrap with 500 replicates, using full heuristic ML searches with stepwise additions and TBR branch swapping, was performed to assess nodal support. Machine time is a key issue in Maximum Likelihood and unfortunately a parallel version of PAUP* has not been published yet. To speed up the process, we set up the analysis to simulate a parallel computation, therefore taking higher advantage of the large computational power of the BioPortal. We run 25 independent bootstrap resamplings with 20 replicates each, starting with different random seeds generated by Microsoft Excel® 2007 following software recommendations. Trees found in each run were then merged and final consensus was computed with PAUP*.

MrBayes 3.1.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) software was used for Bayesian analyses, which were carried out at the BioPortal as above. We performed a Bayesian analysis for each partitioning scheme listed in Supplementary material Tables 2. Schemes 26–33 involve amino acids instead of nucleotides for protein-coding genes: a "glorified" GTR + I + Γ model was used for amino acid partitions. Two MC^3 algorithm runs with four chains each were run for 10,000,000 generations; convergence was estimated through PSRF (Gelman and Rubin, 1992) and by plotting standard deviation of average split frequencies sampled every 1000 generations. The ingroup was constrained as monophyletic, trees found at convergence were retained after the burnin, and a majority-rule consensus tree was computed with the command **sumt**. Via the command **sump printtofile = yes** we could obtain the harmonic mean of the EML, which was used to address model selection and partition choice.

Table 1
Taxa and Genbank accession numbers used in this study for phylogenetic reconstructions.

Order	Species	Genbank acc no.	Reference
Archaeognatha	*Nesomachilis australica*	AY793551	Cameron et al. (2004)
Thysanura	*Tricholepidion gertschi*	AY191994	Nardi et al. (2003)
Ephemeroptera	*Parafronurus youi*	EU349015	Zhang et al. (2008)
Odonata	*Orthetrum triangolare*	AB126005	Yamauchi et al. (2004)
Blattaria	*Periplaneta fuliginosa*	AB126004	Yamauchi et al. (2004)
Isoptera	*Reticulitermes hageni*	EF206320	Cameron and Whiting (2007)
Mantodea	*Tamolanica tamolana*	DQ241797	Cameron et al. (2006a)
Orthoptera	*Locusta migratoria*	X80245	Flook et al. (1995)
	Gryllotalpa orientalis	AY660929	Kim et al. (2005)
Mantophasmatodea	*Sclerophasma paresisensis*	DQ241798	Cameron et al. (2006a)
Phasmatodea	*Timema californicum*	DQ241799	Cameron et al. (2006a)
	Bacillus rossius	**GU001956**	This study
	Bacillus atticus	GU001955	This study
Grylloblattodea	*Grylloblatta sculleni*	DQ241796	Cameron et al. (2006a)

We applied AIC (Akaike, 1973) and BF (Kass and Raftery, 1995). AIC was calculated, following Huelsenbeck et al. (2004), Posada and Buckley (2004), and Strugnell et al. (2005), as

$$AIC = -2EML + 2K$$

The number of free parameters K was computed taking into account branch number, character (nucleotide, amino acid, presence/absence of an indel) frequencies, substitution rates, gamma shape parameter and proportion of invariable sites for each partition.

Bayes Factors were calculated, following Brandley et al. (2005), as

$$B_{ij} = \frac{EML_i}{EML_j}$$

and, doubling and turning to logarithms,

$$2\ln B_{ij} = 2(\ln EML_i - \ln EML_j)$$

where B_{ij} is the Bayes Factor measuring the strength of the ith hypothesis over the jth hypothesis. Bayes Factors were interpreted according to Kass and Raftery (1995) and Brandley et al. (2005).

All trees were graphically edited by PhyloWidget (Jordan and Piel, 2008) and Dendroscope (Huson et al., 2007). Mitochondrial genomes were drawn with GenomeVx (Conant and Wolfe, 2008).

Since we obtained our best resolved and statistically supported phylogenetic tree by Bayesian analysis, we performed time calibration on this topology. The r8s 1.71 (Sanderson, 2003) software was used and three calibration points were set: the origin of winged insects, which was set between 396 and 408 Mya (Engel and Grimaldi, 2004; Grimaldi, 2010); the rise of orthopteran clade, constrained between 144.2 and 150.7 Mya (Labandeira, 1994); and the origin of genus *Bacillus*, which was estimated between 20.14 and 25.44 Mya in a previous study (Mantovani et al., 2001). Given the basal paraphyly, *N. australica* was pruned and only *T. gertschi* was used as outgroup. Rates and times were estimated following PL method by Truncated-Newton algorithm. Several rounds of cross-validation analysis were used to determine the best-performing smoothing value for PL method and the penalty function was set to log. Four perturbations of the solutions and five multiple starts were invoked to optimize searching in both cases. Solutions were checked through the **checkGradient** command.

To compute age estimate boundaries, we used PERL scripts composing the r8s-bootkit package of Torsten Eriksson (2007). This procedure involves the generation with the PHYLIP package (Felsenstein, 1993) of 100 bootstrap replicates of the original dataset to compute branch lengths. This is usually done with PAUP*. However, our best tree was obtained with MrBayes from a mixed dataset (i.e. nucleotides + amino acids + indels binary data), and PAUP* is unable to perform this optimization. On the contrary, both ML and Bayesian nucleotide-based trees were less resolved, and more prone to saturation effects, thus resulting in two polytomies (see below in Section 3). However, according to the Shimodaira–Hasegawa test (Shimodaira and Hasegawa, 1999), none (but one) of the conceivable trees resolving those polytomies is significantly better than the others (data not shown), so that we are confident that our best topology is as good estimate of the real phylogeny. For this reason, we decided to maintain the best topology, but we were forced to calculate replicate branch lengths on nucleotide alignment, upon exclusion of third codon positions and binary information on gaps. We acknowledge that this may lead to some inconsistencies between age estimates and their confidence limits, since they have been calculated with two different approaches, nevertheless both approaches start from the very same dataset and should not produce extremely different results (as actually happened, see Section 3). A complete cross-validated PL analysis was performed on each bootstrap replicate tree and age parameters (mean and confidence intervals) were computed using Microsoft Excel® 2007.

Finally, we computed Phylogenetic Informativeness following the method described by Townsend (2007). Sitewise evolutionary rates were computed by MrBayes 3.1.2 via the command **report siterates = yes**: we used the option **startingtree = user** to force the initial topology to the tree linearized by r8s and set proposal rates to 0 for those parameters influencing topology and branch lengths through the command **props**. MC^3 was kept running until stability in likelihood scores was reached. As evolutionary rates computed by MrBayes 3.1.2 represent the amount of mutation for that site across the entire tree, we divided each rate for tree height (in Myr) and obtained Phylogenetic Informativeness following Eq. (10) in Townsend (2007; p. 225). The informativeness profile was integrated by approximation with a set of rectangles having 5 Myr as base.

3. Results

3.1. The mitochondrial genomes of Bacillus stick insects

Partial mtDNA genomes, including the region downstream the *nad2* to the *rrnS* gene of *B. atticus* and *B. rossius* (order Phasmatodea, suborder Verophasmatodea), were sequenced for this study. We were unable, as it was for *T. californicum* (Cameron et al., 2006a) and other stick insects (Komoto et al., 2011), to successfully sequence the control region of *Bacillus* mtDNA. Such a failure may be either due to its extreme length or to the presence of highly repetitive AT-rich portions in this region, or both. According to the plesiomorphic pancrustacean gene arrangement, the complete control region and *trnI*, *trnQ* and *trnM* genes are therefore lacking

Table 2
Annotation of the *Bacillus atticus* mitochondrial genome (GU001955).

Start	End	Gene	Strand	Length	Start codon	Stop codon	Intergene[a]
1	999	*nad2*	H	999	ATA	TAA	−2
998	1062	*trnW*	H	65			−8
1055	1119	*trnC*	J	65			0
1120	1183	*trnY*	J	64			1
1185	2723	*coxI*	H	1539	ATG	TAA	−5
2719	2782	*trnL2*	H	64			0
2783	3449	*coxII*	H	667	ATA	T–	0
3450	3519	*trnK*	H	70			−1
3519	3584	*trnD*	H	66			0
3585	3743	*atp8*	H	159	ATA	TAA	−7
3737	4411	*atp6*	H	675	ATG	TAA	−1
4411	5197	*coxIII*	H	787	ATG	T–	0
5198	5262	*trnG*	H	65			0
5263	5614	*nad3*	H	352	ATA	T–	0
5615	5681	*trnA*	H	67			−1
5681	5746	*trnR*	H	66			0
5747	5812	*trnN*	H	66			1
5814	5878	*trnS1*	H	65			1
5880	5945	*trnE*	H	66			−2
5944	6007	*trnF*	J	64			−1
6007	7671	*nad5*	J	1665	ATA	TAA	60
7732	7795	*trnH*	J	64			2
7798	9126	*nad4*	J	1329	ATG	TAA	−7
9120	9404	*nad4L*	J	285	ATA	TAA	8
9413	9477	*trnT*	H	65			0
9478	9541	*trnP*	J	64			1
9543	10,022	*nad6*	H	480	ATA	TAA	−1
10,022	11,149	*cob*	H	1128	ATG	TAA	−2
11,148	11,214	*trnS2*	H	67			18
11,233	12,178	*nad1*	J	946	ATA	T–	3
12,182	12,248	*trnL1*	J	67			0
12,249	13,539	*rrnL*	J	1291			0
13,540	13,598	*trnV*	J	59			0
13,599	14,141	*rrnS*	J	543			

[a] Negatives numbers indicate that adjacent genes overlap.

in this study. The sequenced region include all the protein-coding genes and it is 14,141 bp long for *B. atticus* and 14,152 bp for *B. rossius*. Tables 2 and 3 show annotation of either genome and sequences are available in GenBank under accession numbers GU001955 and GU001956, respectively; the mitochondrial genome map of *B. atticus* is shown in Fig. 1.

The mtDNA genome of both *B. atticus* and *B. rossius* has the typical metazoan mitochondrial genome composition of 13 protein-coding genes, two ribosomal RNAs and 19 out of 22 transfer RNAs (lacking *trnI, trnQ* and *trnM* in our sequencing). Moreover, observed gene orders are identical to that proposed by Boore (1999) as ancestral arrangement (symplesiomorphic) for Pancrustacea. The overall AT-contents are 78.1% and 77.6% in *B. atticus* and *B. rossius*, respectively. As in typical arthropod mtDNA, there are only small non-coding regions between genes: these are between *trnY* and *coxI* (1 bp), *trnN* and *trnS1* (1 bp), *trnS1* and *trnE* (1 bp), *nad5* and *trnH* (60 bp), *nad4L* and *trnT* (8 bp), *trnP* and *nad6* (1 bp), *trnS2* and *nad1* (18 bp), and *nad1* and *trnL1* (3 bp); *B. atticus* has one more 2-bp non-coding region between *trnH* and *nad4*. The 18 bp long non-coding region between tRNA-Ser(UCR) and *nad1* shows the TACTAA box, which is also present in *T. californicum* (Cameron et al., 2006a): this motif appears to be conserved across all insects orders, with the consensus sequence DWWCYHH (Cameron and Whiting, 2008), and Taanman (1999) hypothesized it to be the binding site of a transcription attenuation factor called mtTERM.

Start and stop codons share the same pattern across the two species: start codons are either ATG (used five times) or ATA (used eight times); in both species, *coxII*, *coxIII*, *nad1* and *nad3* genes, as long as *nad4* gene in *B. rossius* only, are terminated by a T (truncated codon for TAA), whereas all the remaining stop codons are TAA. The two typical genes for ribosomal RNAs are present, one for the large and one for the small ribosomal subunit.

Finally, the sequenced tRNAs can be folded into typical cloverleaf secondary structures (see Supplementary material Figs. 1 and 2) with the only exception of tRNA-Ser(AGN), lacking stem pairings in the DHU arm. This feature has been observed in several insect orders, as well as in other metazoans (Feng et al., 2010; Kim et al., 2005; Sheffield et al., 2008; and references therein). The same feature is present in *T. californicum* (Cameron et al., 2006a), so we can confirm its presence in Phasmatodea too; in these three species, the anticodon is always GCU.

Nucleotide alignment was 15,353 bp long, and 591 indel events were added, resulting in a total of 15,944 characters. Nevertheless, Xia et al. (2003) test clearly showed significant level of saturation among third codon positions (Table 4), broadening the results of Maekawa et al. (1999) based on *cox2* gene only: therefore, these nucleotides were excluded and 12,133 characters were left for phylogenetic analysis. When PCGs were translated into amino acids, stop codons were removed from the analysis and an alignment made of 8309 characters was obtained.

3.2. Phylogenetic analysis

Fig. 2 shows ML tree computed by PAUP*. Both Phasmidae and *Bacillus* appear monophyletic, with bootstrap values of 89 and 100, respectively. The Dictyoptera are also well resolved: in fact, both splits in the *Tamolanica* + (*Periplaneta* + *Reticulitermes*) cluster have the maximum bootstrap value. On the contrary, nodes linking phasmids to *Grylloblatta* and *Sclerophasma* are not resolved, as is the orthopteran group; anyway, these are the only two polytomies to be found in this tree. Finally, the splitting of Ephemeroptera

Table 3
Annotation of the *Bacillus rossius* mitochondrial genome (GU001956).

Start	End	Gene	Strand	Length	Start codon	Stop codon	Intergene[a]
1	999	*nad2*	H	999	ATA	TAA	−2
998	1062	*trnW*	H	65			−8
1055	1129	*trnC*	J	75			−9
1121	1185	*trnY*	J	65			1
1187	2725	*coxI*	H	1539	ATG	TAA	−5
2721	2785	*trnL2*	H	65			0
2786	3452	*coxII*	H	667	ATA	T–	0
3453	3522	*trnK*	H	70			−1
3522	3586	*trnD*	H	65			0
3587	3745	*atp8*	H	159	ATA	TAA	−7
3739	4413	*atp6*	H	675	ATG	TAA	−1
4413	5199	*coxIII*	H	787	ATG	T–	0
5200	5264	*trnG*	H	65			0
5265	5616	*nad3*	H	352	ATA	T–	0
5617	5683	*trnA*	H	67			−1
5683	5747	*trnR*	H	65			0
5748	5813	*trnN*	H	66			1
5815	5879	*trnS1*	H	65			1
5881	5945	*trnE*	H	65			−2
5944	6007	*trnF*	J	64			−1
6007	7671	*nad5*	J	1665	ATA	TAA	60
7732	7795	*trnH*	J	64			0
7796	9125	*nad4*	J	1330	ATG	T–	−7
9119	9403	*nad4L*	J	285	ATA	TAA	8
9412	9476	*trnT*	H	65			−1
9476	9540	*trnP*	J	65			1
9542	10,021	*nad6*	H	480	ATA	TAA	−1
10,021	11,154	*cob*	H	1134	ATG	TAA	−2
11,153	11,219	*trnS2*	H	67			18
11,238	12,183	*nad1*	J	946	ATA	T–	3
12,187	12,253	*trnL1*	J	67			0
12,254	13,532	*rrnL*	J	1279			0
13,533	13,601	*trnV*	J	69			0
13,602	14,152	*rrnS*	J	551			

[a] Negatives numbers indicate that adjacent genes overlap.

(*Parafronurus*) from other Pterygota (i.e., Megapterygota) insects is strongly supported by bootstrap.

Supplementary material Table 3 lists results from AIC and BF statistics. EMLs from nucleotide analyses lead to different conclusion: AIC selected t19 as the best explanation of data, whereas BF selected t11. On the other hand, among amino acid analyses, both AIC and BF selected t30 as the best tree: given this agreement, and that we know from saturation test (see Table 4) that nucleotide sequences are prone to saturation in our dataset (even if it was demonstrated only for third codon positions), we consider t30 as our best estimate of orthopteroid phylogenetic tree, and it is shown in Fig. 3, along with its nucleotide counterpart. All nodes were resolved with Posterior Probabilities equal to 1 in the t30 tree.

In the t30 partitioning scheme, genes are pooled in five categories: ribosomal, *atp*, cytochrome, *nad*, and tRNA. All "splitter" models (those dividing genes within each functional category) performed worse than t30 when PCGs were translated into amino acids, whereas first and second codon positions were kept separately both in t11 and t19 models; again, in model t19 mtDNA genes were pooled in the same five categories. It is tempting to conclude that these categories correspond to real, homogeneous gene groupings that, because of different selective pressures, experienced different, discrete evolutionary pathways; however, we cannot rule out that these models simply represent the best trade-off between overparametrization in "splitter" and oversimplification in "lumper" models. Anyway, the first hypothesis seems to hold at least for ribosomal and tRNA partition, which were always preferred to single-gene subdivisions.

The obtained tree evidenced that the two *Bacillus* species are monophyletic as well as the order Phasmatodea (*Timema* + *Bacillus*). *Sclerophasma* is basal to Phasmatodea, and *Grylloblatta* to (*Sclerophasma* + Phasmatodea). Dictyopterans are also well resolved, with the praying mantis *Tamolanica* basal to *Periplaneta* (a cockroach) and *Reticulitermes* (a termite); this cluster is sister group to (*Grylloblatta* + (*Sclerophasma* + (*Timema* + *Bacillus*))). True orthopterans (*Locusta* + *Gryllotalpa*) are basal to all orthopteroid insects.

The dating of the t30 tree (Fig. 4 and Table 5) placed the origin of orthopteroid insects in the Middle Triassic (227.56 Mya), whereas most splits took place during the Jurassic period. The origins of orthopterans (150.70 Mya, as by constraints) and dictyopterans (145.76 Mya) were dated between Jurassic and Cretaceous. The split between Mantophasmatodea and Phasmatodea occurred 173.06 Mya (Middle Jurassic), and the order Phasmatodea seems to appear in the Upper Jurassic, 156.79 Mya.

Phylogenetic Informativeness plots (Fig. 5) show that the maximum resolving power of insect mtDNA is expected around the Upper Cretaceous, about 80 Mya. While grouped ribosomal genes and tRNAs substantially behave the same, different pools of PCGs show some variations in expected resolving efficiency: cytochrome and *nad* genes seem particularly apt to track more recent splits (about 60 Mya), whereas *atp* genes exhibit a more flat Phylogenetic Informativeness profile along the whole Cretaceous and the Upper Jurassic. Since most of the main nodes of this study were *a posteriori* dated between 150 and 200 Mya, we compared the Phylogenetic Informativeness under this timespan to the 50 Myr surrounding area of the optimum peak, i. e. from 55 to 105 Mya. Informativeness profiles were integrated within these intervals and the ratio between the two areas was calculated (Table 6): mtDNA conveyed in the 150–200 Mya timespan at least the 70% of the informativeness expected in the optimal period. Notably,

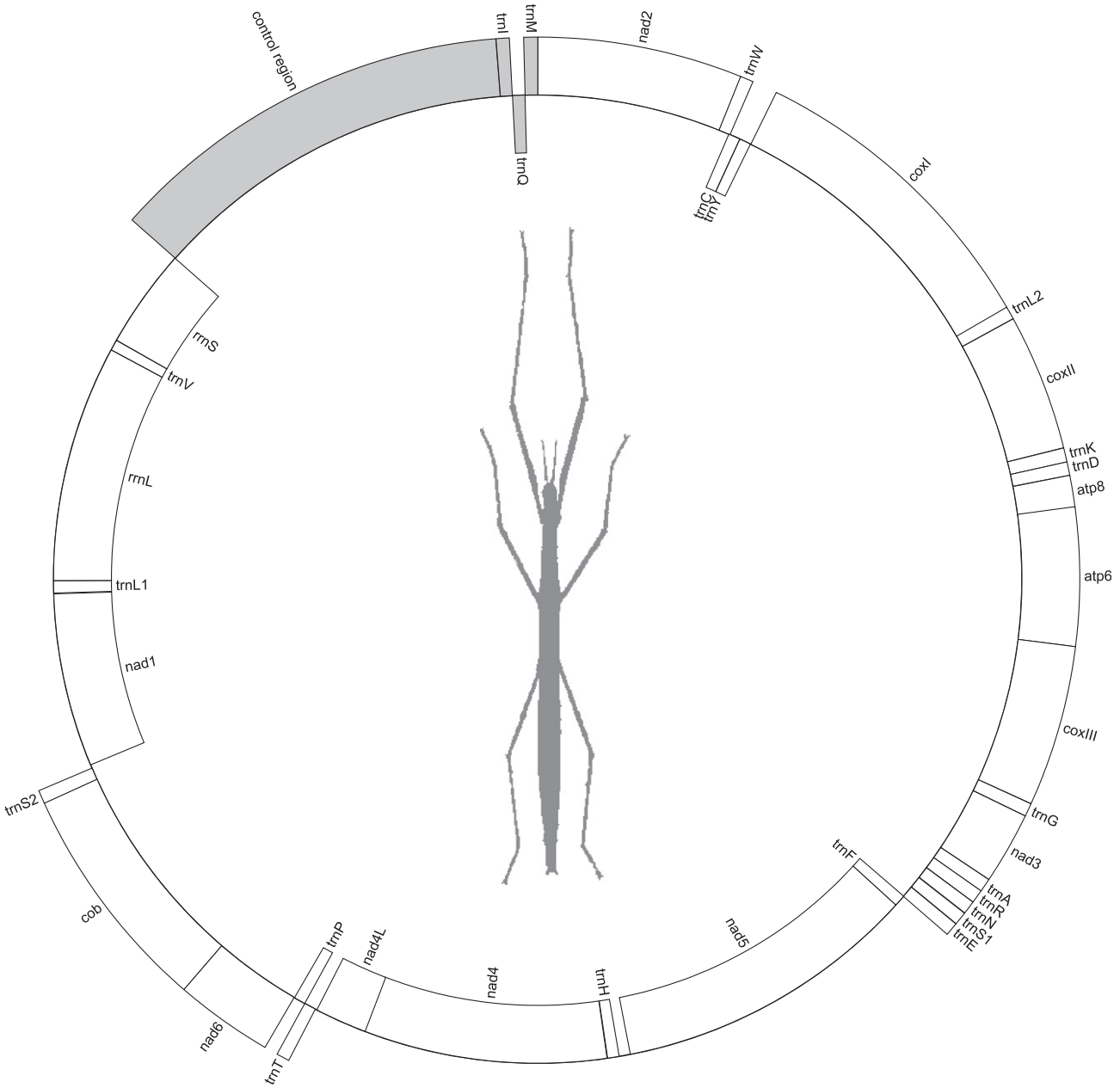


Fig. 1. Mitochondrial genome map of *Bacillus atticus*. Unsequenced regions are shaded in gray: control region was arbitrarily scaled to 2000 bp, whereas *trnI*, *trnQ*, and *trnM* were scaled to their average orthopteroid sizes of 65, 68, and 69 bp, respectively.

this ratio was down to 67% for *nad* genes and more interestingly, up to 89% for *atp6* and *atp8* genes. Phylogenetic Informativeness per nucleotide and per million years base were also computed, which allow "estimation of cost-effectiveness" of sites across different genes (Townsend, 2007).

4. Discussion

This study expands previous knowledge of mitochondrial genomes in Phasmatodea by sequencing two representatives of the suborder Verophasmatodea, *B. atticus* and *B. rossius*. As mentioned, both *Bacillus* sequenced mitochondrial genomes are similar in gene and nucleotide composition to that of *T. californicum* stick insect mitochondrial genome (Cameron et al., 2006a), as well as to the presumed ancestral hexapod (Boore, 1999; Fenn et al., 2007; Kim et al., 2005). Complete genome annotations are reported on Tables 2 and 3 and in Fig. 1.

4.1. Phylogenetic Informativeness

To our knowledge, insect mtDNA marker informativeness was never addressed before with the method proposed by Townsend (2007). When applying this to our dataset, informativeness is always higher for ribosomal genes than for tRNAs. On one hand, ribosomal genes sum up to 2500 characters (nucleotides and indel presence/absence), whereas tRNAs to 1902, thus making *trn* genes more informative on a per base criterion, which is in good agreement with Cameron et al. (2007). On the other hand, ribosomal genes are only two, whereas tRNA genes are 22, scattered throughout the whole molecule – on both strands. As in t30 model ORFs were translated into amino acids, we avoid to directly compare RNAs with PCGs. Within PCGs, *nad* genes are the most informative, with a peak around the Cretaceous–Tertiary boundary (65 Mya). However, they are made up by 2192 characters in our dataset, whereas cytochrome genes are made up by 1425 characters and

Table 4
Saturation test by Xia et al. (2003).

	95% C.I.			Iss.c[a]	Asym
	Iss[b]	Lower	Upper	Sym	
Prot	0.6192	0.6048	0.6336	0.8350	0.5823
Prot_1	0.5529	0.5263	0.5795	0.8214	0.6769
Prot_2	0.3702	0.3415	0.3990	0.8213	0.6767
Prot_12	0.4570	0.4369	0.4771	0.8268	0.6778
Prot_3	0.9234	0.9062	0.9406	0.8213	0.6767

[a] Iss.c, critical index of substitution saturation, computed for two extreme topologies: a perfectly symmetrical (Sym) and an extremely asymmetrical tree (Asym).

[b] Iss, index of substitution saturation; when this value falls above the critical threshold defined by Iss.c, level of saturation is taken as significant in the dataset. As the orthopteroid tree is expected to be somewhat asymmetric, there is some evidence of saturation (Iss > $Iss.c_{Asym}$) for the complete PCG alignment (prot), no evidence for first and second codon position nucleotides (prot_1, prot_2, and prot_12), and strong evidence for third codon position nucleotides (prot_3; Iss > $Iss.c_{Sym} \gg Iss.c_{Asym}$).

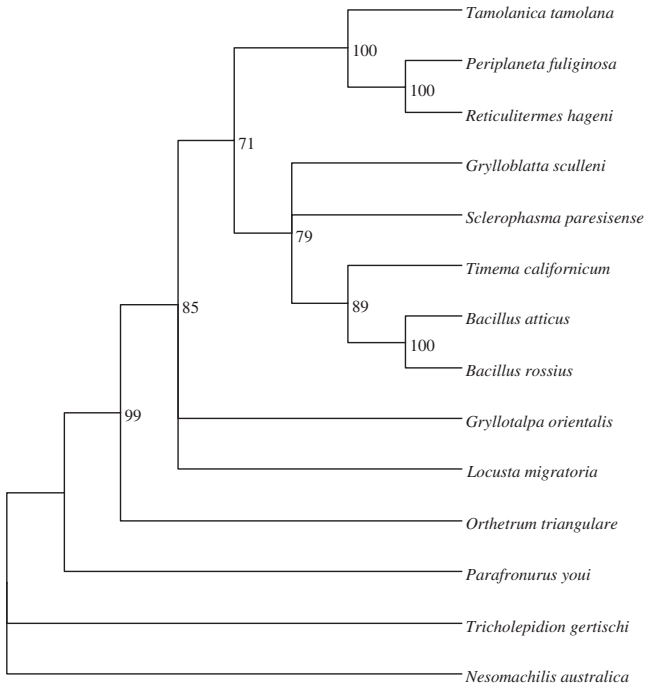


Fig. 2. ML tree based on the orthopteroid mtDNA dataset. Node numbers are bootstrap values on 500 replicates.

atp genes by 290. This makes of *atp6* and *atp8* the most per base informative genes in orthopteroid mtDNA dataset, with a maximum resolution power around more than 100 Mya; however, when complete sequence are analyzed, for recent times (0–50 Mya) cytochrome genes are preferable (see Fig. 5b). The amount of informativeness conveyed in the mtDNA is not wildly disproportionate between the optimal resolution time (around 80 Mya) and the period we focused in this study on. The informativeness we rely on to depict and date most nodes in ancient orthopteroid evolution is more than the 70% of the peak resolving power, and, for *atp6* and *atp8* genes, it is close to 90%. Phylogenetic Informativeness analysis, indeed, gives a sharp idea of phylogenetic signal presence among the mtDNA molecule and can be very useful to plan future studies on this part of insect evolution bush. Depending upon the timespan of interest and available resources, different mitochondrial markers behave differently in terms of resolving power, even if *atp* genes unexpectedly show the best cost/effectiveness ratio in any case (Table 6).

4.2. Phylogenetic inferences on orthopteroid lineages

In this study, we obtained a robust molecular phylogeny of orthopteroid insects, with nodes showing strong statistical support, especially with Bayesian analysis and given the proper model selection. It is interesting to note that our analysis, regardless the applied models, always confirmed that Timematodea and Verophasmatodea are sister groups, so that, as far as we know from our still small dataset, the order Phasmatodea should be considered as monophyletic. This was also found in previous studies on target nuclear genes (Terry and Whiting, 2005; Whiting et al., 2003), even if Kjer et al. (2006) failed to recover Phasmatodea as monophyletic. This is particularly noteworthy because the Timematodea suborder is the earliest diverging stick insect taxon: the divergence between *Timema* and Verophasmatodea (to which *Bacillus* pertain) occurred more than 95 Mya according to Buckley et al. (2009), and more than 150 Mya according to our study.

We also compared the obtained *Bacillus* sequences to other basal hexapods in order to reconstruct phylogenetic relationships between Phasmatodea and other lower pterygote insects, with special attention to orthopteroid insects. In previous studies, some molecular support was found for Plecoptera + Dermaptera, Embioptera + Phasmatodea, and Grylloblattodea + Mantophasmatodea (Ishiwata et al., 2010; Kjer et al., 2006; Terry and Whiting, 2005), while other data place Mantophasmatodea with Phasmatodea (Cameron et al., 2006a; Kjer et al., 2006). In our study a significant sister relationship between Phasmatodea and Mantophasmatodea (as well as Grylloblattodea) was found, with this clade more closely related to Dictyoptera (i. e. Mantodea + Blattodea + Isoptera), rather than to Orthoptera. Posterior probabilities were highly significant among the obtained Bayesian trees, while bootstrap values were slightly less robust. Nevertheless, the overall trend is quite stable and we are confident that our analysis evidenced a real phylogenetic signal.

This result is different from what stated by Fenn et al. (2008), who found a closer relationship between Phasmatodea and Dictyoptera, rather to Mantophasmatodea and Grylloblattodea. Interestingly, Wheeler et al. (2001) described a closer relationship between Grylloblattodea and Dictyoptera, than between Phasmida and Dictyoptera. The phylogeny we retrieved is similar to Cameron et al. (2006a) and Kjer et al. (2006), who were not able, however, to resolve deeper nodes. From our analysis, we do not have any evidence for the validity of Orthopteroidea *sensu lato*, i.e. Orthoptera + (Embiidina + Phasmatodea), but unfortunately no complete embiopteran mitochondrial genome is available at present: therefore, we cannot assess the correctness of Eukinolabia *sensu* Terry and Whiting (2005). Furthermore, Xenonomia (Grylloblattodea + Mantophasmatodea) were retrieved as paraphyletic and are not supported in our study. Finally, Whitfield and Kjer (2008) sketched a topology largely concordant with the one presented here, but nevertheless they interestingly obtained Xenonomia as monophyletic, as also Ishiwata et al. (2010) did.

Fossil Plecoptera, Orthoptera, and Dictyoptera have been found in the Permian (Whitfield and Kjer, 2008), and first neopterans (Paoliidae) in the Carboniferous (Grimaldi and Engel, 2005). This would leave only 50 Myr for the main phylogenetic events in orthopteroid evolution to occur, with lineages that are over 300 Myr old nowadays. Such branches are very long and therefore the ten neopteran lineages may constitute a 'soft polytomy', which is due to insufficient phylogenetic information rather than to actual polytomic cladogenetic events (i. e. 'hard polytomies'). Moreover, our dates confirm how quick cladogeneses were on a geological scale: timespans of 13, 7, and 16 Myr separate the first split from dictyopterans and the definitive rise of order Phasmatodea (Fig. 4). This explains why in many cases, especially with nucleotide-only data, some nodes were left unresolved, while an

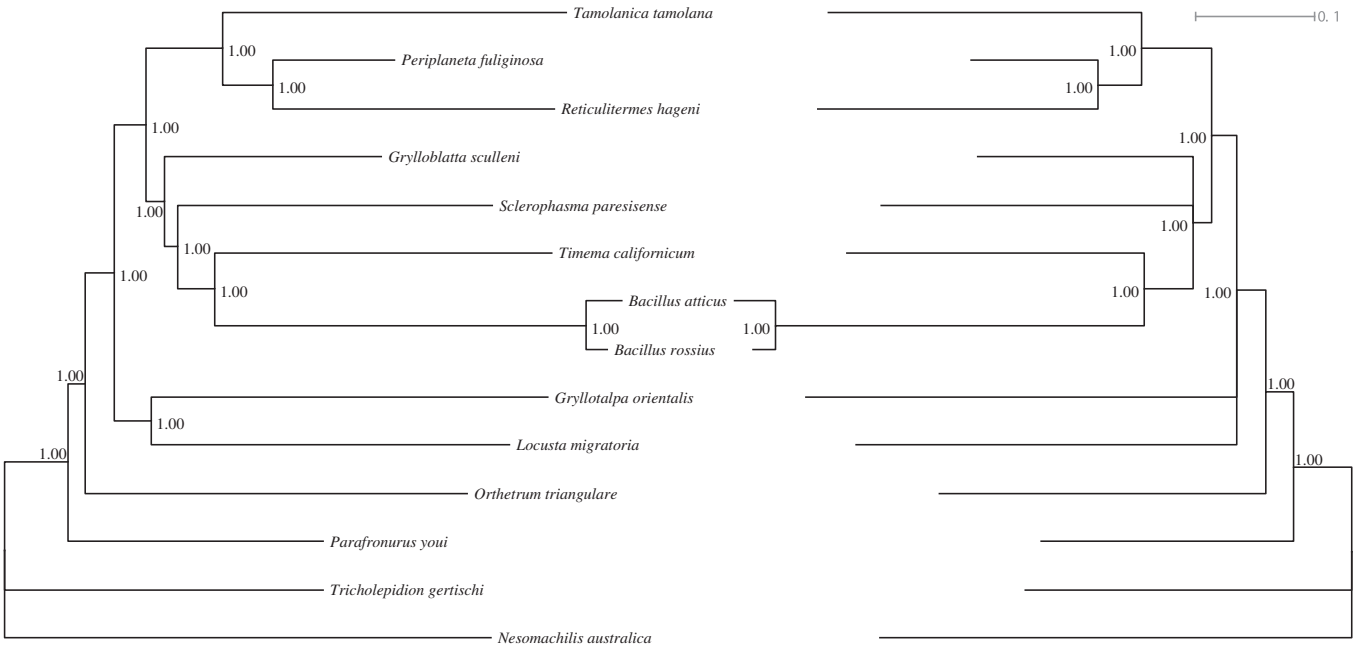


Fig. 3. Bayesian phylogenetic trees based on the orthopteroid mtDNA dataset. Node numbers are posterior probability values. On the left t30 tree is shown: PCGs were translated into amino acids and this model was chosen both by AIC and BF statistics; t19 tree, based only on nucleotides and chosen only by AIC, is shown on the right for comparison purposes.

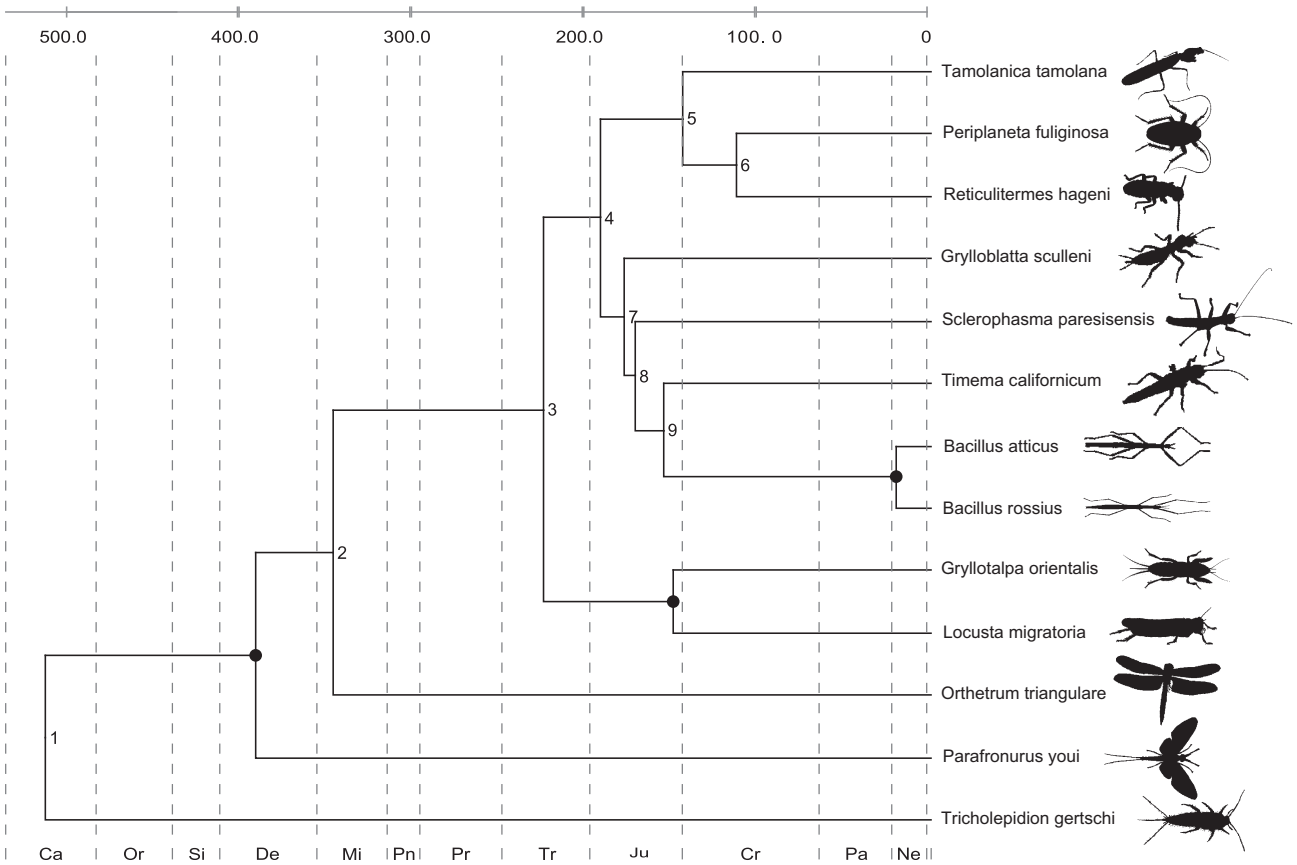


Fig. 4. Ultrametric tree computed by Penalized Likelihood on t30 tree shown in Fig. 3, left. Black dots indicate nodes used for calibration; numbers refer to node ages listed in Table 5. Geological data are taken from Gradstein et al. (2004) and Ogg et al. (2008). Ca, Cambrian; Or, Ordovician, Si, Silurian; De, Devonian; Mi, Mississippian; Pn, Pennsylvanian; Pr, Permian; Tr, Triassic; Ju, Jurassic; Cr, Cretaceous; Pa, Paleogene; Ne, Neogene. Quaternary is only shown as the timespan between the very last two bars at the bottom.

Table 5
PL age estimates.

Node[c]	Min[d]	Max[d]	t30[a]	Estimated[f]	Local[f]	t19[b]	95% C.I.	
			Age[e]			Mean	Lower	Upper
1			518.92			459.69	442.79	484.27
Pterygota	396.00	408.00	396.00	6.0936e−04	6.0928e−04	396.00	396.00	396.00
2			350.61	4.7261e−04	4.5675e−04	302.25	284.92	331.17
3			227.56	2.9538e−04	2.7691e−04	231.01	211.39	256.18
4			193.91	1.1012e−03	1.1129e−03	215.65	197.80	237.40
5			145.76	1.8785e−03	1.8810e−03	173.58	159.19	191.48
6			114.36	1.8977e−03	1.8978e−03	142.97	128.63	158.89
7			180.20	1.5687e−03	1.5778e−03	202.08	184.10	222.89
8			173.06	2.1521e−03	2.1618e−03	198.02	181.25	219.85
9			156.79	2.7446e−03	2.7542e−03	174.80	157.51	193.70
Bacillus	20.14	25.44	20.14	3.1967e−03	3.2068e−03	20.14	20.14	20.14
Orthoptera	144.20	150.70	150.70	5.8984e−04	5.5970e−04	150.70	150.70	150.70

[a] Node age, estimated, and local evolutionary rates are given for the best phylogenetic tree.
[b] Mean and confidence limits are given based upon the nucleotide alignment and the best phylogenetic tree.
[c] Node numbers refer to Fig. 4; named node are those used for tree calibration. See text for further details.
[d] Minimum and maximum age are constraints given to nodes used for tree calibration.
[e] Age in Mya.
[f] Evolutionary rates are expressed in substitution per site per time unit.

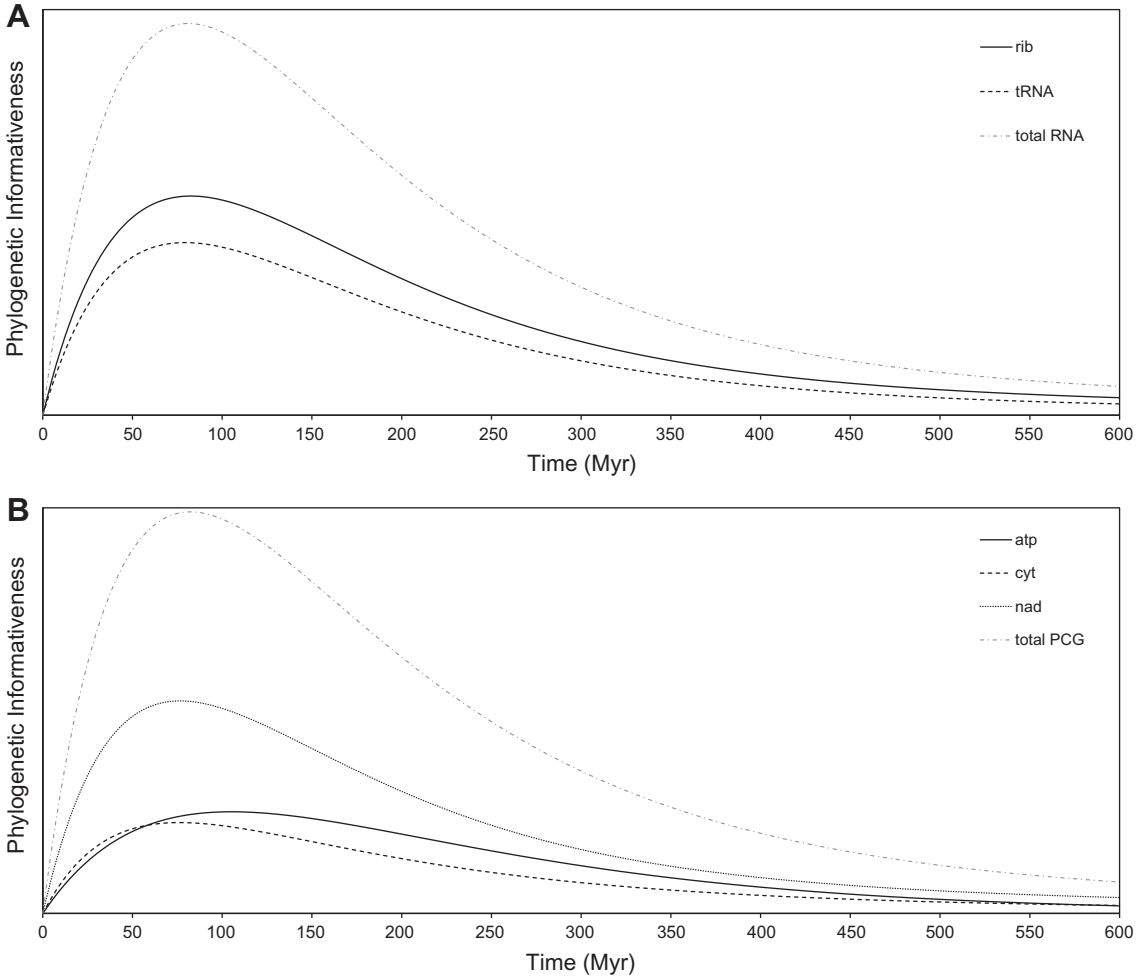


Fig. 5. Profiles of Phylogenetic Informativeness for RNA (A) and protein-coding (B) genes plotted on time (in Myr) before present. rib, ribosomal genes (*rrnL* and *rrnS*); tRNA, tRNA genes; atp, ATPase genes (*atp6* and *atp8*); cyt, cytochrome genes (*coxI*, *coxII*, *coxIII*, and *cob*); nad, NAD genes. Dot-line series show total tRNA and total PCG informativeness, respectively.

accurate model selection and shaping scored to unravel these insect relationships. Our age estimation is in a substantial agreement with the aforementioned data, even if most nodes tend be slightly younger than expected by fossils, as it is the case for orthopterans and isopterans. This may be caused by the coarse taxon sampling in our study, whose objective was the phylogenetic relationships

Table 6
The power of different mitochondrial partitions to resolve splits at different timings.

Partition[c]	Sites[d]	Ratio[e]	Overall optimum[a]		Orthopteroids[b]		Per million years optimum[a]		Orthopteroids[b]	
			Per site	Net	Per site	Net	Per site	Net	Per site	Net
Atp	280	0.89	0.12881	36.1	0.11440	32.0	0.00258	0.7	0.00229	0.6
tRNA	1721	0.70	0.08529	146.8	0.05940	102.2	0.00171	2.9	0.00119	2.0
Total RNA	3920	0.71	0.08500	333.2	0.06034	236.5	0.00170	6.7	0.00121	4.7
Rib	2199	0.72	0.08478	186.4	0.06107	134.3	0.00170	3.7	0.00122	2.7
All	7718	0.72	0.06215	479.7	0.04453	343.7	0.00124	9.6	0.00089	6.9
Prot	3798	0.73	0.03856	146.4	0.02822	107.2	0.00077	2.9	0.00056	2.1
Nad	2119	0.67	0.03647	77.3	0.02458	52.1	0.00073	1.5	0.00049	1.0
Cyt	1399	0.70	0.02366	33.1	0.01648	23.1	0.00047	0.7	0.00033	0.5

[a] Informativeness profiles for optimum were integrated from 55 to 105 Mya.
[b] 150–200 Mya.
[c] Partition nomenclature as in Supplementary material Table 1; total RNA, rib + tRNA.
[d] Only bp/amino acids were tallied here.
[e] Ratio between informativeness profiles integrated around orthopteroids main splits and around optimal peak.

among Polyneoptera with special reference to Phasmatodea, and by the use of amino acids in the linearized tree. In fact, when ages were computed using the nucleotide dataset alone (even if on the very same tree), most Mesozoic nodes turned out to be older and with rather narrow confidence intervals (Table 5): 174.80 was the mean age of the rise of Phasmatodea and 173.58 of that of Dicytoptera. The origin of orthopteroids, however, was not substantially changed (from 228 to 231 Mya). It is noteworthy to keep in mind the tendency of predating the main evolutionary events depicted in our tree, especially when agreement with fossil record is sought: for example, oldest fossil termites are known from the Lower Cretaceous (130–140 Mya; Korb, 2007; Engel et al., 2009) and, as a matter of fact, we obtained a confidence interval of 128.63–158.89 Mya for the origin of Isoptera.

Another interesting case is found for the clade Grylloblattodea + (Mantophasmatodea + Phasmatodea). t30 tree was able to resolve the node and this is noteworthy, as only this model did not yield a trichotomy for this cluster. In fact, about 7 Myr separate Grylloblattodea from Mantophasmatodea in our tree, pushing this topology towards a 'soft polytomy', which is very hard to unravel, if the model is not correctly chosen. From a morphological perspective, most outstanding Phasmatodea diagnostic apomorphies are (see a full discussion in Bradler, 2009): pear-shaped secretory appendices on the posterior part of the mesenteron; absence of mitochondria in spermatozoa; male vomer; splitting of the lateral dorsoventral musculature into isolated muscle fibres; emarginated labrum; a pair of prothoracic repellant glands (Bradler, 2003, 2009; Cameron et al., 2006a; Hennig, 1969, 1994; Jamieson, 1987; Klug and Bradler, 2006; Tilgner et al., 1999). Cameron et al. (2006a) argued that the absence of such glands in Mantophasmatodea may hamper the detection of their relationship with phasmids, and suggested two possibilities: either those glands were secondarily lost in Mantophasmatodea, or this is actually not a defining character of the whole clade, but only an autoapomorphy of a smaller set of lineages. On the other hand, as shown by Klass et al. (2003), genitalic character analysis results in clustering Phasmatodea and Mantophasmatodea together; anyway, further characters need to be examined more in detail to support our molecular conclusion.

Finally, the inclusion of a complete mitochondrial genome from a leaf insect (subfamily Phyllinae) would be of great interest, because leaf insects lie somewhere between *Timema* and *Bacillus*. In fact, as noted by Wedmann et al. (2007), the oldest known leaf insect (*Eophyllum*) dates back to 47 Myr and the maximum age of the subfamily cannot be older than the rise of flowering plants, which occurred between 125 and 90 Mya, which is in perfect agreement with our chronogram (see Fig. 4 and Table 5). In conclusion, we think this work should represent the first step towards a more stable phylogeny of orthopteroid insects and a significant methodological approach to follow, which proved to give robust phylogenetic results. Of course, further work and additional complete mitochondrial genomes (with special reference to Embiidina) will help in better shaping the branches of the Polyneoptera tree.

Acknowledgments

This work was supported by the Italian "Ministero dell'Università e della Ricerca Scientifica" (MIUR) funds and by the "Donazione Canziani" bequest; we would like to thank Sven Bradler and Valerio Scali for providing several insights into orthopteroid taxonomy, morphology, and systematics, and two anonymous reviewers whose comments greatly improved this work.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2010.12.005.

References

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrox, B.N., Caski, F. (Eds.), Second International Symposium on Information Theory. Akademiai Kiado, Budapest, p. 267.
Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.
Bae, J.S., Kim, H.D., Sohn, B.R., 2004. The mitochondrial genome of the firefly *Pyrocoelia rufa*: complete genome sequence, genome organisation and phylogenetic analysis with other insects. Mol. Phylogenet. Evol. 32, 978–985.
Beutel, R.G., Gorb, S.N., 2001. Ultrastructure of attachment specializations of hexapods (Arthropoda): evolutionary patterns inferred from a revised ordinal phylogeny. J. Zool. Syst. Evol. Res. 39, 177–207.
Beutel, R.G., Gorb, S.N., 2006. A revised interpretation of the evolution of attachment structures in Hexapoda with special emphasis on Mantophasmatodea. Arthropod Syst. Evol. 64, 3–25.
Boore, J.L., 1999. Animal mitochondrial genomes. Nucleic Acids Res. 27, 1767–1780.
Boore, J.L., Lavrov, D.V., Brown, W.M., 1998. Gene translocation links insects and crustaceans. Nature 392, 667–668.
Boudreaux, H.B., 1987. Arthropod Phylogeny with Special Reference to Insects. John Wiley, New York.
Bradler, S., 2003. Phasmatodea, Gespentschrecken. In: Dathe, H.H. (Ed.), Lehrbuch der speziellen Zoologie. I. Wirbellose Tiere. 5. Insecta. Spektrum. Heidelberg, Berlin, pp. 251–260.
Bradler, S., 2009. Die Phylogenie der Stab-und Gespenstschrecken (Insecta: Phasmatodea). S P E, vol. 2.1. Universitätsverlag Göttingen, Göttingen.
Brandley, M.C., Schmitz, A., Reeder, T.W., 2005. Partitioned Bayesian analysis, partition choice, and the phylogenetic relationships of scincid lizards. Syst. Biol. 54, 373–390.
Brusca, R.C., Brusca, G.J., 2003. Invertebrates, second ed. Sinauer Associates, Sunderland.
Buckley, T.R., Attanayake, D., Bradler, S., 2009. Extreme convergence in stick insect evolution: phylogenetic placement of the Lord Howe Island tree lobster. Proc. Biol. Sci. 22, 1055–1062.

Cameron, S.L., Whiting, M.F., 2007. Mitochondrial genomic comparisons of the subterranean termites from the Genus *Reticulitermes* (Insecta: Isoptera: Rhinotermitidae). Genome 50, 188–202.
Cameron, S.L., Whiting, M.F., 2008. The complete mitochondrial genome of the tobacco hornworm, *Manduca sexta*, (Insecta: Lepidoptera: Sphingidae), and an examination of mitochondrial gene variability within butterflies and moths. Gene 408, 112–123.
Cameron, S.L., Miller, K.B., D'Haese, C.A., Whiting, M.F., Barker, S.C., 2004. Mitochondrial genome data alone are not enough to unambiguously resolve the relationships of Entognatha, Insecta and Crustacea sensu lato (Arthropoda). Cladistics 20, 534–557.
Cameron, S.L., Barker, S.C., Whiting, M.F., 2006a. Mitochondrial genomics and the new insect order Mantophasmatodea. Mol. Phylogenet. Evol. 38, 274–279.
Cameron, S.L., Beckenbach, A.T., Dowton, M., Whiting, M.F., 2006b. Evidence from mitochondrial genomics on interordinal relationships in insects. Arthropod Syst. Phylogenet. 64, 27–34.
Cameron, S.L., Lambkin, C.L., Barker, S.C., Whiting, M.F., 2007. A mitochondrial genome phylogeny of Diptera: whole genome sequence data accurately resolve relationships over broad timescales with high precision. Syst. Entomol. 32, 40–59.
Cameron, S.L., Sullivan, J., Song, H., Miller, K.B., Whiting, M.F., 2009. A mitochondrial genome phylogeny of the Neuropterida (lace-wings, alderflies and snakeflies) and their relationship to the other holometabolous insect orders. Zool. Scr. 38, 575–590.
Carapelli, A., Liò, P., Nardi, F., Van der Wath, E., Frati, F., 2007. Phylogenetic analysis of mitochondrial protein coding genes confirms the reciprocal paraphyly of Hexapoda and Crustacea. BMC Evol. Biol. 7, S8.
Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540–552.
Castro, L.R., Dowton, M., 2005. The position of the Hymenoptera within the Holometabola as inferred from the mitochondrial genome of *Perga condei* (Hymenoptera: Symphyta: Pergidae). Mol. Phylogenet. Evol. 34, 469–479.
Castro, L.R., Dowton, M., 2007. Mitochondrial genomes in the Hymenoptera and their utility as phylogenetic markers. Syst. Entomol. 32, 60–69.
Conant, G.C., Wolfe, K.H., 2008. GenomeVx: simple web-based creation of editable circular chromosome maps. Bioinformatics 24, 861–862.
Delsuc, F., Phillips, M.J., Penny, D., 2003. Comment on "Hexapod origins: Monophyletic or Paraphyletic?". Science 301, 1482.
Dowton, M., Cameron, S.L., Austin, A.D., Whiting, M.F., 2009. Phylogenetic approaches for the analysis of mitochondrial genome sequence data in the Hymenoptera – a lineage with both rapidly and slowly evolving mitochondrial genomes. Mol. Phylogenet. Evol. 52, 512–519.
Engel, M.S., Grimaldi, D.A., 2000. A winged *Zorotypus* in Miocene amber from the Dominican Republic (Zoraptera: Zorotypidae), with the discussion on relationships of and within the order. Acta Geologica Hispanica 35, 149–164.
Engel, M.S., Grimaldi, D.A., 2004. New light shed on the oldest insect. Nature 427, 627–630.
Engel, M.S., Grimaldi, D.A., Krishna, K., 2009. Termites (Isoptera): their phylogeny, classification, and rise to ecological dominance. Am. Mus. Novit. 3650, 1–27.
Eriksson, T., 2007. The r8s Bootstrap Kit (Distributed by the Author).
Felsenstein, J., 1993. PHYLIP: Phylogenetic Inference Package (Distributed by the Author).
Feng, X., Liu, D.-F., Wang, N.-X., Zhu, C.-D., Jiang, G.-F., 2010. The mitochondrial genome of the butterfly *Papilio xuthus* (Lepidoptera: Papilionidae) and related phylogenetic analyses. Mol. Biol. Rep. 37, 3877–3888.
Fenn, J.D., Cameron, S.L., Whiting, M.F., 2007. The complete mitochondrial genome sequence of the Mormon cricket (*Anabrus simplex*: Tettigoniidae: Orthoptera) and an analysis of control region variability. Insect Mol. Biol. 16, 239–252.
Fenn, J.D., Song, H., Cameron, S.L., Whiting, M.F., 2008. A preliminary mitochondrial genome phylogeny of Orthoptera (Insecta) and approaches to maximizing phylogenetic signal found within mitochondrial genome data. Mol. Phylogenet. Evol. 49, 59–68.
Flook, P.K., Rowell, C.H.F., 1998. Inferences about orthopteroid phylogeny and molecular evolution from small subunit nuclear ribosomal DNA sequences. Insect Mol. Biol. 7, 163–178.
Flook, P.K., Rowell, C.H.F., Gellissen, G., 1995. The sequence, organization, and evolution of the *Locusta migratoria* mitochondrial genome. J. Mol. Evol. 41, 928–941.
Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. Stat. Sci. 7, 457–511.
Gradstein, F.M., Ogg, J.G., Smith, A.G. (Eds.), 2004. A Geologic Time Scale 2004. Cambridge University Press, Cambridge.
Grimaldi, D.A., 2010. 400 million years on six legs: on the origin and early evolution of Hexapoda. Arthropod Struct. Dev. 39, 191–203.
Grimaldi, D.A., Engel, M.S., 2005. Evolution of the Insects. Cambridge University Press, Cambridge.
Gullan, P.J., Cranston, P.S., 2005. The Insects. An Outline of Entomology, Auflage, vol. 3. Blackwell Publishing, Berlin.
Haas, F., Kukalová-Peck, J., 2001. Dermaptera hindwing structure and folding: new evidence for familial, ordinal and superordinal relationships within Neoptera (Insecta). Eur. J. Entomol. 98, 445–509.
Hennig, W., 1969. Die Stammesgechichte der Insekten. Kramer, Frankfurt am Main.
Hennig, W., 1981. Insect Phylogeny. John Wiley and Sons, New York (Translated and Edited by Pont, A.C. and revisionary notes by Schlee, D.).
Hennig, W., 1994. Wirbellose II. Gliedertiere. 5 AuX. (Nachdruck). Gustav Fischer, Jena.
Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17, 754–755.
Huelsenbeck, J.P., Larget, B., Alfaro, M.E., 2004. Bayesian phylogenetic model selection using reversible jump markov chain Monte Carlo. Mol. Biol. Evol. 21, 1123–1133.
Huson, D.H., Richter, D.C., Rausch, C., Dezulian, T., Franz, M., Rupp, R., 2007. Dendroscope – an interactive viewer for large phylogenetic trees. BMC Bioinf. 8, 460.
Ishiwata, K., Sasaki, G., Ogawa, J., Miyata, T., Su, Z.-H., 2010. Phylogenetic relationships among insect orders based on three nuclear protein-coding gene sequences. Mol. Phylogenet. Evol. doi:10.1016/j.ympev.2010.11.001.
Jamieson, B.G.M., 1987. The Ultrastructure and Phylogeny of Insect Spermatozoa. Cambridge Univ. Press, Cambridge.
Jordan, G.E., Piel, W.H., 2008. PhyloWidget: Web-based visualizations for the tree of life. Bioinformatics 15, 1641–1642.
Kambhampati, S., 1995. A phylogeny of cockroaches and related insects based on DNA sequence of mitochondrial ribosomal RNA genes. Proc. Natl. Acad. Sci. USA 92, 2017–2020.
Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773–795.
Katoh, K., Toh, H., 2008. Improved accuracy of multiple ncRNA alignment by incorporating structural information into a MAFFT-based framework. BMC Bioinf. 25, 212.
Katoh, K., Misawa, K., Kuma, K.I., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059–3066.
Kim, I., Cha, S.Y., Yoon, M.H., Hwang, J.S., Lee, S.M., Sohn, H.D., Jin, B.R., 2005. The complete nucleotide sequence and gene organization of the mitochondrial genome of the oriental mole cricket, *Gryllotalpa orientalis* (Orthoptera: Gryllotalpidae). Gene 353, 155–168.
Kjer, K.M., 2004. Aligned 18S and insect phylogeny. Syst. Biol. 53, 506–514.
Kjer, K.M., Honeycutt, R.L., 2007. Site specific rates of mitochondrial genome and phylogeny of eutheria. BMC Evol. Biol. 7, 8.
Kjer, K.M., Carle, F.L., Litman, J., Ware, J., 2006. A molecular phylogeny of Insecta. Arthropod Syst. Phylogenet. 64, 35–44.
Klass, K.-D., Zompro, O., Kristensen, N.P., Adis, J., 2002. Mantophasmatodea: a new insect order with extant members in the afrotropics. Science 296, 1456–1459.
Klass, K.-D., Picker, M.D., Damgaard, J., Van Noort, S., Tojo, K., 2003. The taxonomy, genitalic morphology and phylogenetic relationships of southern African Mantophasmatodea (Insecta). Entomol. Abhandl. 61, 3–67.
Klug, R., Bradler, S., 2006. The pregenital abdominal musculature in phasmids and its implications for the basal phylogeny of Phasmatodea (Insecta: Polyneoptera). Org. Divers. Evol. 6, 171–184.
Komoto, N., Yukuhiro, K., Ueda, K., Tomita, S., 2011. Exploring the molecular phylogeny of phasmids with the whole mitochondrial genome sequences. Mol. Phylogenet. Evol. 58, 43–52.
Korb, J., 2007. Termites. Curr. Biol. 17, R995–R999.
Kristensen, N.P., 1981. Phylogeny of insect orders. Annu. Rev. Entomol. 26, 135–157.
Kristensen, N.P., 1995. Forty years' insect phylogenetic systematics. Zool. Beitr. NF 36, 83–124.
Kukalová-Peck, J., 1991. Fossil history and the evolution of hexapod structures. In: CSIRO (Eds.), The Insects of Australia, second ed. Melbourne University Press, Carlton, pp. 141–179.
Kukalová-Peck, J., Peck, S.B., 1993. Zoraptera wing structures: evidence for new genera and relationship with the blattoid orders (Insecta: Blattoneoptera). Syst. Entomol. 18, 333–350.
Labandeira, C.C., 1994. A compendium of fossil insect families. In: Watkins, R. (Ed.), Contributions in Biology and Geology, vol. 88. Milwaukee Public Museum, Wisconsin, pp. 15–71.
Laslett, D., Canbäck, B., 2008. ARWEN, a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. Bioinformatics 24, 172–175.
Lowe, T.M., Eddy, S.R., 1997. TRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25, 955–964.
Maekawa, K., Kitade, O., Matsumoto, T., 1999. Molecular phylogeny of orthopteroid insects based on the mitochondrial cytochrome oxidase II gene. Zool. Sci. 16, 175–184.
Mantovani, B., Passamonti, M., Scali, V., 2001. The mitochondrial cytochrome oxidase II gene in *Bacillus* stick insects: ancestry of hybrids, androgenesis, and phylogenetic relationships. Mol. Phylogenet. Evol. 19, 157–163.
Martynov, A.B., 1925. Über zwei Grundtypen der Flügel bei den Insecten und ihre evolution. Zeitschrift zur Morphologie und Ökologie der Tiere 4, 465–501.
Nardi, F., Carapelli, A., Fanciulli, P.P., Dallai, R., Frati, F., 2001. The complete mitochondrial DNA sequence of the basal hexapod *Tetrodontophora bielanensis*: evidence for heteroplasmy and tRNA translocations. Mol. Biol. Evol. 18, 1293–1304.
Nardi, F., Spinsanti, G., Boore, J.L., Carapelli, A., Dallai, R., Frati, F., 2003. Hexapod origins: monophyletic or paraphyletic? Science 299, 1887–1889.
Nuin, P., 2008. MrMTgui: cross-platform interface for ModelTest and MrModeltest. <http://www.genedrift.org/mtgui.php>.
Ogg, J.G., Ogg, G., Gradstein, F.M., 2008. The Concise Geologic Time Scale. Cambridge University Press, Cambridge.
Posada, D., Buckley, T.R., 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio test. Syst. Biol. 53, 793–808.
Posada, D., Crandall, K.A., 1998. ModelTest: testing the best-Wt model of nucleotide substitution. Bioinformatics 14, 817–818.

Rähle, W., 1970. Untersuchungen an Kopf und Prothorax von *Embia ramburi* Rimsky-Korsakoff 1906 (Embioptera, Embiidae). Zoologischer Jahrbücher, Abteilung für Anatomie und Ontogenie der Tiere 87, 248–330.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Sanderson, M.J., 2003. R8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics 19, 301–302.

Sheffield, N.C., Song, H., Cameron, S.L., Whiting, M.F., 2008. A comparative analysis of mitochondrial genomes in Coleoptera (Arthropoda: Insecta) and genome descriptions of six new beetles. Mol. Biol. Evol. 25, 2499–2509.

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114–1116.

Simmons, M.P., Ochoterena, H., 2000. Gaps as characters in sequence-based phylogenetic analyses. Syst. Biol. 49, 369–381.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., Flook, P., 1994. Evolution, weighting and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann. Entomol. Soc. Am. 87, 651–701.

Simon, C., Buckley, T.R., Frati, F., Stewart, J.B., Beckenbach, A.T., 2006. Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. Ann. Rev. Ecol. Evol. Syst. 37, 545–579.

Stewart, J.B., Beckenbach, A.T., 2003. Phylogenetic and genomic analysis of the complete mitochondrial DNA sequence of the spotted asparagus beetle *Crioceris duodecimpunctata*. Mol. Phylogenet. Evol. 26, 513–526.

Strugnell, J., Norman, M., Jackson, J., Drummond, A.J., Cooper, A., 2005. Molecular phylogeny of coleoid cephalopods (Mollusca: Cephalopoda) using a multigene approach; the effect of data partitioning on resolving phylogenies in a Bayesian framework. Mol. Phylogenet. Evol. 37, 426–441.

Swofford, D., 1999. PAUP*: phylogenetic analysis using parsimony (*and other methods). Sinauer Associates, Sunderland.

Taanman, J.-W., 1999. The mitochondrial genome: structure, transcription, translation and replication. Biochim. Biophys. Acta 1410, 103–123.

Talavera, G., Castresana, J., 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. 56, 564–577.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24, 1596–1599.

Terry, M.D., Whiting, M.F., 2005. Mantophasmatodea and phylogeny of the lower neopterous insects. Cladistics 21, 240–257.

Thomas, M.A., Walsh, K.A., Wolf, M.R., McPheron, B.A., Marden, J.H., 2000. Molecular phylogenetic analysis of evolutionary trends in stonefly wing structure and locomotor behavior. Proc. Natl. Acad. Sci. USA 97, 13178–13183.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

Thorne, B.L., Carpenter, J.M., 1992. Phylogeny of the dictyoptera. Syst. Entomol. 17, 253–268.

Tilgner, E.H., Kiselyova, T.G., McHugh, J.V., 1999. A morphological study of *Timema cristinae* Vickery with implications for the phylogenetics of Phasmida. Dtsche. Entomol. Z. 46, 149–162.

Townsend, J.P., 2007. Profiling phylogenetic informativeness. Syst. Biol. 56, 222–231.

Wedmann, S., Bradler, S., Rust, J., 2007. The first fossil leaf insect: 47 million years of specialized cryptic morphology and behavior. Proc. Natl. Acad. Sci. USA 104, 565–569.

Wheeler, W.C., Whiting, M.F., Wheeler, Q.D., Carpenter, J.M., 2001. The phylogeny of the extant hexapod orders. Cladistics 17, 113–169.

Whitfield, J.B., Kjer, K.M., 2008. Ancient rapid radiations of insects: challenges for phylogenetic analysis. Annu. Rev. Entomol. 53, 449–472.

Whiting, M.F., 2002. Mecoptera is paraphyletic: multiple genes and phylogeny of Mecoptera and Siphonaptera. Zool. Scr. 31, 93–104.

Whiting, M.F., Bradler, S., Maxwell, T., 2003. Loss and recovery of wings in stick insects. Nature 421, 264–267.

Willmann, R., 2003. Die phylogenetischen Beziehungen der Insecta: offene Fragen und Probleme. Verhandlungen Westdeutscher Entomologentag 2001, 1–64.

Willmann, R., 2004. Phylogenetic relationships and evolution of insects. In: Cracraft, J., Donoghue, M.J. (Eds.), Assembling the Tree of Life. Oxford University Press, Oxford, pp. 330–344.

Xia, X., Xie, Z., 2001. DAMBE: software package for data analysis in molecular biology and evolution. J. Hered. 92, 371–373.

Xia, X., Xie, Z., Salemi, M., Chen, L., Wang, Y., 2003. An index of substitution saturation and its application. Mol. Phylogenet. Evol. 26, 1–7.

Yamauchi, M.M., Miya, M.U., Nishida, M., 2004. Use of a PCR-based approach for sequencing whole mitochondrial of insects: two examples (cockroach and dragonfly) based on method developed for decapod crustaceans. Insect Mol. Biol. 13, 435–442.

Young, N.D., Healy, J., 2003. GapCoder automates the use of indel characters in phylogenetic analysis. BMC Bioinf. 4, 6.

Zhang, J., Zhou, C., Gai, Y., Song, D., Zhou, K., 2008. The complete mitochondrial genome of *Parafronurus youi* (Insecta: Ephemeroptera) and phylogenetic position of the Ephemeroptera. Gene 424, 18–24.

Zompro, O., 2001. The Phasmatodea and *Raptophasma* n. gen., Orthoptera incertae sedis, in Baltic amber (Insecta: Orthoptera). Mitteilungen Geol. Paläontol. Inst. Hamburg 85, 229–261.

ACKNOWLEDGEMENTS

Hard was gathering and moulding in words and pages the content of three years of scientific work, but much harder is striving to crowd into few pages three years of laughs, tears, questions, talks, runs, flops, enthusiasms, and friends.

Sincerely thanks are really due to Dr Marco Passamonti, who allowed me to graduate, become a PhD student, work and enjoy these years of intellectual adventures; above all, he taught me the how fascinating is the world of bivalve mollusks, how many particulars have to be fine-tuned before a task is completed – especially a figure, how uncultured is my palate, and how many Middle-Italian aphorisms are there. He made these years possible, and thus a statistically significant part of what I am today is largely his fault.

Furthermore, these 1,095 days would have been completely different without the MoZoo people, with a special reference to Alessandro and Valentina, who shared with me much more than the PhD course; and to Fabrizio and Liliana, who shared with me offices, desks, trips, doubts, jokes, neighborhoods, and troubles. Thanks are also due to either Andrea, the one who taught me to hold a pipette in my hand, and the one who struggled to open bottles I tightly closed the day before; to all guys who did not fear to work with me as undergraduate students; to Marco, who immediately entered the futsal team; to Professor Valerio Scali, who always kept my digital consciousness awake; and to Professor Barbara Mantovani, who started to be a teacher for me nine years ago and never stopped.

I would like to give credits also to Thomas Bayes, Charles Darwin, Stephen Jay Gould, and the programmers of Microsoft Excel®, but I feel that this section would risk to become tedious. However, three more people have to be cited here. My parents, Cristina and Piero, grew me up feeding not only my stomach, which is notable in itself, but also my curiosity and my desire to learn. Since I was child until university, I think they never

refused a book to me, even if their mathematical minds often did not share my passion for bugs and clams, and this PhD was built up on these grounds.

Finally, even Italian words are lacking to give proper credit to that woman who was my girlfriend when this PhD started, and became my wife during these years; therefore, I am forced to borrow English acknowledgements joining two sonnets that would have been obviously beyond my skills. After all, what is true for poetry is actually true for science, too.

So long as men can breathe or eyes can see,
So long lives this, and this gives life to thee.

(William Shakespeare, Sonnet XVIII)

And yet, by heaven, I think my love as rare
As any she belied with false compare.

(William Shakespeare, Sonnet CXXX)