PHYLOGENETIC SYSTEMATICS, TAXONOMY, AND BIOGEOGRAPHY OF JELLYFISH (CNIDARIA: MEDUSOZOA)

By

Bastian Bentlage

Submitted to the graduate degree program in Ecology and Evolutionary Biology and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

_____

Chairperson Paulyn Cartwright

_____

Allen G. Collins

_____

Mark Holder

_____

Jorge Soberón

_____

Xingong Li

Date Defended: 24 May, 2012

The Dissertation Committee for Bastian Bentlage

certifies that this is the approved version of the following dissertation:

PHYLOGENETIC SYSTEMATICS, TAXONOMY, AND BIOGEOGRAPHY OF JELLYFISH (CNIDARIA: MEDUSOZOA)

_____

Chairperson Paulyn Cartwright

Date approved: 31 May 2012

ABSTRACT

The coastal shelf inhabiting box jellyfish (Cubozoa) represent the smallest class within Cnidaria with some 50 described species. A robust phylogenetic framework had been missing for Cubozoa. Herein, a molecular phylogeny for Cubozoa is presented. This phylogeny served as the basis for several taxonomic and nomenclatural changes in a reverse taxonomic approach, striving to align the classification scheme for Cubozoa with phylogenetic history. In addition, the revised classification led to a reevaluation of morphological characters used for the delineation and identification of species and higher taxa. This information was condensed into an illustrated taxonomic key to aid the identification of box jellyfish by non-specialists. Furthermore, the utility of ecological niche models for predicting the potential geographic distributions of box jellyfish based on correlations between species occurrences and environmental data was assessed. Since box jellyfish distributions are generally poorly documented, modeling approaches that make use of the limited data available may be of much value for making predictions about species distributions. Similarly, species distributions in the open oceans, in particular in the deep sea, are poorly understood and documented. A new approach to ecological niche and species distribution modeling in three dimensions was developed that has great potential for aiding studies of open ocean fauna. This new approach was used to derive an explicit *a priori* hypothesis about the population structure of a deep-sea inhabiting jellyfish species. In order to test this hypothesis and better understand the patterns of gene-flow in open ocean environments, population genomic data was used to evaluate population structure for this species on a global scale. The combined approach of ecological niche modeling and population genomics indicated that at least the species investigated here displays panmixia on a global scale.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

Gewidmet meinen Eltern für ihre fortwährende Unterstützung.

The work presented in this thesis would not have been possible to accomplish without the help of many people at the University of Kansas, the Smithsonian Institution in DC, and other places and institutions around the world. I am especially indebted to Paulyn Cartwright and Allen Collins who served me as advisors and mentors during the last 6 years on my way to becoming an independent scientist. In addition, I would like to thank the members of my committee who helped me delve deeper into my research questions and learn the skills necessary to complete my work. In no particular order these were: Mark Holder, Xingong Li, Town Peterson, and Jorge Soberón. Furthermore, the participants of the Systematics and Niche Modeling Discussion groups provided many stimulating and interesting discussions.

I would also like to thank the many others who have assisted me in various ways throughout the years. Tim Coffer, Chad Walter, and Merrick Ekins made my visits at the National Museum of Natural History in Wasington DC and the Queensland Museum in Australia successful and productive. Steve Haddock, Dhugal Lindsay, Mike Vecchione, and Tracey Sutton generously shared specimens with me and invited me to join them on research cruises when there was a spare berth. Keith Bayha, Michael Dawson, and Liza Gomez always had an eye out for specimens to collect for me during their field trips. Jamie Seymour supplied many of the box jellyfish species that were used in my research on the phylogeny and taxonomy of Cubozoa.

Steve Cairns and Chris Mah gave me much advice on nomenclatural issues that arose during my taxonomic work.

A special thanks to my (ex-)roomate Russ Henry for getting me started with Linux; without this primer I couldn't have done much of my later work. Thanks to my wife Mariya, for patiently clarifying programming concepts to me, just so I may immediately forget again what I just learned. Lastly, my deepest gratitude to all the people I have interacted with over the years at KU and other places – I learned a lot from discussions with all of you. In particular I would like to thank Nat, Annalise, Andrea, Gemma, Cheryl, and Marcos.

Funding for my work was provided through US NSF Assembling the Cnidarian Tree of Life Grant (EF-053179, to Paulyn Cartwright, Allen G. Collins, and Daphne Fautin), US NSF Doctoral Dissertation Improvement Grant (DEB 0910237, to B.B. and P.C.), and a PADI Foundation grant (to B.B.). I wish to acknowledge generous funding through a Smithsonian Institution pre-doctoral fellowship. The Department of Ecology and Evolutionary Biology and the Biodiversity Institute provided funding to sponsor my attendance at several national and international meetings.

INTRODUCTION

This thesis is composed of five chapters that deal with the phylogenetic systematics, taxonomy, and biogeography of different groups and species of jellyfish (Cnidaria: Medusozoa). Both shallow water, coastal shelf inhabiting jellyfish and deep-sea dwelling jellyfish were studied for this dissertation using molecular genetic methods as well as traditional taxonomy and ecological modeling. In particular, a phylogenetic framework for box jellyfish (Cnidaria: Cubozoa) was established (chapter 1) that formed the basis of subsequent taxonomic revisions (chapter 1) and a taxonomic key including species descriptions (chapter 2). Box jellyfish distributions are poorly documented, even though they inhabit the coastal shelves in tropical to warm temperate waters close to the shore. To address this issue, ecological niche modeling was used in a case study to predict distributions of box jellyfish in the Indo-Pacific (chapter 3). The geographic distributions of jellyfish inhabiting the deep sea are largely undocumented. To address this issue, an approach was developed that uses existing tools for ecological niche modeling in a three-dimensional environment (chapter 4). Lastly, the newly developed three-dimensional approach to ecological niche modeling was integrated with population genomic data to illuminate the distributional patterns of a deep sea inhabiting jellyfish species (chapter 5).

Chapter 1: In the first chapter, a phylogenetic hypothesis for Cubozoa is presented that was used to revise the taxonomic classification of Cubozoa in a reverse taxonomic approach. All of the taxonomic and nomenclatural changes presented in this chapter have been made available in compliance with the 4th edition of the International Code of Zoological Nomenclature (International Commission on Zoological Nomenclature 1999) in Bentlage *et al.* (2010) which represents the authority for these changes.

Chapter 2: The second chapter takes the revised taxonomy presented in chapter 1 and highlights the morphological underpinnings of the revised taxonomy and nomenclature for one of the two major clades in Cubozoa, Carybdeida. In addition, several species are described. This work was done in collaboration with Cheryl Lewis (Department of Biology, The University of Maryland) and is currently under review. Nomenclatural acts in this chapter should not be considered available under the 4th edition of the International Code of Zoological Nomenclature (International Commission on Zoological Nomenclature, 1999).

Chapter 3: This chapter examines the potential use of GIS-based modeling methods to address the lack of species occurrence data that describe the geographic distributions of box-jellyfish and has been published in the *Marine Ecology Progress Series* (Bentlage *et al.* 2009).

Chapter 4: This chapter presents a modified approach to ecological niche (or species distribution) modeling to describe the potential distributions of species inhabiting three-dimensional habitats like the open oceans. The approach presented here uses existing software tools and is straight-forward to implement. This work is currently in review and was a collaboration with my advisor Paulyn Cartwright, Town Peterson (Biodiversity Institute, The University of Kansas), Narayani Barve (Biodiversity Institute, The University of Kansas).

Chapter 5: The last chapter uses the approach to three-dimensional ecological niche modeling presented in chapter 4 to make predictions about the distribution of the deep sea inhabiting hydrozoan jellyfish species *Halicreas minimum* (Cnidaria: Hydrozoa: Trachylina). This prediction was then used to generate a hypothesis about the population structure of the species on a global scale. This hypothesis was then tested using population genomic data generated via next-generation sequencing technologies. The dataset presented here will be

complemented by a second population genomic dataset of a different open ocean inhabiting jellyfish species prior to publication.

CHAPTER 1:

Evolution of box jellyfish (Cnidaria: Cubozoa), a group of highly toxic invertebrates

Abstract

Cubozoa (Cnidaria: Medusozoa) represents a small clade of approximately 50 described species, some of which cause serious human envenomations. Our understanding of the evolutionary history of Cubozoa has been limited by the lack of a sound phylogenetic hypothesis for the group. Here, we present a comprehensive cubozoan phylogeny based on ribosomal genes coding for near-complete nuclear 18S (small subunit) and 28S (large subunit) and partial mitochondrial 16S. We discuss the implications of this phylogeny for our understanding of cubozoan venom evolution, biogeography and life-history evolution. Our phylogenetic hypothesis suggests that: (i) the last common ancestor of Carybdeida probably possessed the mechanism(s) underlying Irukandji syndrome, (ii) deep divergences between Atlantic and Indo-Pacific clades may be explained by ancient vicariant events, and (iii) sexual dimorphism evolved a single time in concert with complex sexual behavior Furthermore, several cubozoan taxa are either para- or polyphyletic, and we address some of these taxonomic issues by designating a new family, Carukiidae, a new genus, *Copula*, and by redefining the families Tamoyidae and Tripedaliidae. Lastly, cubozoan species identities have long been misunderstood and the data presented here support many of the recent scientific descriptions of cubozoan species. However, the results of a phylogeographic analysis of *Alatina moseri* from Hawai'i and *Alatina mordens* from Australia indicate that these two nominal species represent a single species that has maintained metapopulation cohesion by natural or anthropogenic dispersal.

Introduction

Although Cubozoa is the smallest class of Cnidaria, comprising some 50 described box jellyfish species, it is well known for several remarkable attributes. From the possession of complex eyes and associated visual capabilities (e.g., Nilsson *et al.* 2005), to extraordinary courtship and mating behavior (e.g., Lewis & Long 2005), to extreme toxicity (e.g., Brinkman & Burnell 2009), there are many reasons why cubozoans catch the attention of the scientific community and public. Despite this interest, studies of cubozoan evolution have been hampered by a paucity of specimens in natural history museums preserved for both morphological and molecular investigation, as well as by their perceived lack of diversity. The last decade has seen more than a doubling in recognized cubozoan species, but so far a robust phylogenetic framework for investigating the evolution of cubozoan diversity has been missing. We present a comprehensive phylogeny for Cubozoa and use it to discuss the evolution of venom, life history and biogeography.

Materials and Methods

Phylogenetic inference. A list of specimens used for this study is provided in the Appendix (Appendix I), including museum catalog numbers where vouchers exist. Tentacle tissue was preserved in pure EtOH or saturated salt DMSO buffer (Dawson *et al.* 1998), from which DNA was extracted using organic phenol–chloroform extraction protocols according to the procedure outlined in Collins *et al.* (2008) or using the automated DNA isolation system AutoGenPrep 965 (AutoGen Inc., Holliston, MA, USA) following the manufacturer's protocol. Ribosomal genes coding for partial mitochondrial 16S (16S) and near-complete nuclear 18S (small subunit; SSU)

and 28S (large subunit; LSU) were amplified using the primers and protocols outlined in Cartwright *et al.* (2008) and Collins *et al.* (2008). PCR products were either sequenced by Cogenics (Houston, TX, USA) or at the Laboratory of Analytical Biology, Smithsonian Institution (Suitland, MD, USA). Trace fillles were assembled in SEQUENCHER (v. 4.8; Gene Codes, MI) and subsequently aligned using MUSCLE (v. 3.7; Edgar 2004). Highly variable, poorly aligned regions were removed from the final alignments using GBLOCKS (v. 0.91b; Castresana 2000) with the default parameters except that allowed gap positions were set to half. Nucleotide sequences were deposited in NCBI GenBank (Appendix I) and alignments used for analyses were deposited in TreeBASE (treebase.org). In addition to the alignments for each gene, all three were combined into a concatenated alignment. These four alignments were analyzed using maximum parsimony (MP) in PAUP* (Swofford 2003) and maximum likelihood (ML) in RAXML (v. 7.0.3; Stamatakis 2006). MRMODELTEST (v. 2.3; Nylander 2004) was used to evaluate nucleotide substitution models for ML analyses. The concatenated dataset was partitioned by gene for analyses and number of invariant sites and gamma shape parameters were calculated separately for each partition in RAXML.

To establish the root of the cubozoan phylogeny, SSU and LSU sequences of both cubozoan taxa and outgroup taxa were aligned using MUSCLE, subsequently pruned using GBLOCKS and analyzed with ML and MP using the criteria described above. Since the large divergence between ingroup and outgroup for the 16S gene does not allow for reliable alignment, we decided a priori not to analyze this marker using an outgroup.

In order to investigate the possibility of strongly supported character conflict among partitions in the combined datasets, we performed an incongruence length difference (ILD) test

(Farris *et al.* 1995a,b) as implemented in PAUP*. The ILD test has often been used as a test of combinability of datasets for phylogenetic analyses (e.g., Cunningham 1997), but interpretation of ILD test results has been the subject of debate (e.g., Barker & Lutzoni 2002).

***Alatina* phylogeography.** Mitochondrial 16S of seven specimens of *Alatina mordens* from Osprey Reef (Coral Sea, Queensland, Australia) and 19 specimens of *Alatina moseri* from Waikiki (O'ahu, HI, USA) were amplified and sequenced using the same techniques as above (GenBank nos. GQ506980–GQ507005 associated with USNM voucher specimens). All sequences were aligned using MUSCLE and the beginning and end of the alignment were trimmed to the position at which the nucleotides for every specimen are known. A statistical parsimony haplotype network was calculated in TCS (v. 1.21; Clement *et al.* 2000) using the 95 per cent connection limit criterion and gaps treated as a fifth character state.

Results

Phylogeny of Cubozoa. The partition-homogeneity test, with 100 replicates, could not refute the null hypotheses of congruence among partitions in the combined datasets LSU/SSU with outgroup ($p = 0.38$; Appendix II) and LSU/SSU/16S without outgroup ($p = 0.56$; Fig. 1), suggesting the absence of strong conflict among partitions. Phylogenetic analyses under both MP (not shown) and ML lead to highly congruent results, and node support is similar under both ML and MP (Fig. 1; Appendices II-VII). Both Chirodropida and Carybdeida are monophyletic clades with the root of Cubozoa falling in between the two in the SSU and combined LSU/SSU datasets with outgroup (Appendices II-IV). Monophyly of Carybdeida is weakly contradicted in the LSU analysis (Appendix III); several deep nodes receive lower support when an outgroup is included

in LSU analyses (compare Appendices II and V). In general, LSU and SSU analyses do not strongly disagree with one another; contradictory relationships are weakly supported (compare Appendices II-VI).

The fastest evolving marker, 16S (Appendix VII), shows much congruence with both SSU and LSU (Appendices V-VI). One point of difference involves Alatinidae and Tripedaliidae. These families group together in the 16S-based phylogeny (Appendix VII), as well as in SSU-based phylogeny without outgroup (albeit without support; Appendix VI). This putative clade uniting Alatinidae and Tripedaliidae seems surprising, but may be explained by nucleotide saturation. 16S evolves much more rapidly than both LSU and SSU, thus leading to higher degrees of nucleotide saturation that may confound signal at the deeper nodes. The addition of LSU data appears to overcome signal, artificial or otherwise, from 16S and SSU data. Indeed, a combined analysis of all three genes (Fig. 1) leads to a phylogenetic hypothesis that is most congruent with the LSU dataset. Despite the incongruence in the placement of Tripedaliidae between markers, the strongest support for its placement occurs in the combined analysis, where it is recovered as sister to Carybdeidae.

***Alatina* phylogeography.** The final alignment of *Alatina* 16S sequences contained 545 sites; of the 28 variable characters, 10 were parsimony informative. We found 20 unique haplotypes and uncorrected pairwise distances among haplotypes did not exceed 1.84 per cent. The haplotype network (Fig. 2) shows that haplotypes of both *A. moseri* and *A. mordens* are not reciprocally monophyletic and appear inter-digitated. The most common mt16S haplotype (n = 4) was found in two specimens of each *A. moseri* and *A. mordens*.

Discussion

Phylogenetic analyses. and signal Since evolutionary rates differ from gene to gene, some incongruence among topologies using different gene trees is not surprising. Combined analysis of all genes should lead to a better estimate of the evolutionary relationships of taxa compared with single-gene analyses (Gadagkar *et al.* 2005). We find alignment quality to be much improved for the ingroup when excluding highly divergent outgroup taxa. Thus, combined analysis excluding outgroup taxa (Fig. 1) should represent the best estimate of evolutionary relationships in Cubozoa to date.

Previous phylogenetic analyses of Medusozoa (Collins 2002; Collins *et al.* 2006) were limited in their sampling within Cubozoa. Increased taxon sampling lends itself to begin investigating several questions concerning the evolution of cubozoan toxicity, behavior and biogeography. Further, it becomes clear that the taxonomic framework at the family level *sensu* Daly *et al.* (2007) is inconsistent with a phylogenetic approach to taxonomy. Chiropsalmidae, Tamoyidae and Carybdeidae are probably para- or polyphyletic. Consequently, we amend the diagnoses of several taxa or designate new taxa to establish monophyly (changes reflected in figures and table). However, we choose to leave Chiropsalmidae unchanged, as we are missing several chirodropid genera in our analyses.

Toxicity. The evolution of venom in Cubozoa is of significant interest, as many cubozoans are known to be highly toxic (e.g., Williamson *et al.* 1996), resulting in major costs to public health and the tourism industry, particularly in Australia (e.g., Bailey *et al.* 2003). Efforts have led to the characterization of some venom components of a few cubozoan species (see Brinkman & Burnell 2009 for a review) and the development of an antivenom for the deadly

cubozoan Chironex fleeckeri (see Currie 2003). In order to enhance data interpretation and risk management, a historical framework providing a clear understanding of species identities and systematics is vital. In addition to retrospective interpretation of venom data, the phylogenetic framework we present here is relevant for phylogenetic forecasting. That is, close relatives of a highly toxic species are more likely than not highly toxic as well.

Toxicity varies from species to species with some being completely harmless to humans while others can cause death within minutes. The chirodropid *C. fleckeri* is considered the most lethal jellyfish known (Wiltshire *et al.* 2000). Not surprisingly, its close relative *Chironex yamaguchii* has caused human fatalities in Japan and the Philippines (Fenner & Williamson 1996; Fenner 1997 (both as *Chiropsalmus quadrigatus*); Lewis & Bentlage 2009). By contrast, *Chiropsalmus* and *Chiropsella* species are considered much less dangerous (but see Bengtson *et al.* 1991). Differences in toxicity among chirodropids may be explained by differences in the amount of tentacle surface area, and consequently, the amount of venom that can be delivered (see Nagai 2003). Interestingly, an unvouchered tissue specimen from Palau appears to be closely related to *C. yamaguchii* from Japan, raising questions about the toxicity and identity of this chirodropid.

In contrast to the notion that chirodropids represent the most lethal box jellyfishes, hemolytic activity of purified toxin proteins appears lower in *C. yamaguchii* (Nagai *et al.* 2002; as *C. quadrigatus*) than in *Alatina* sp. (Nagai *et al.* 2000a; as *Carybdea alata*) and highest in *Carybdea brevipedalia* (Nagai *et al.* 2000b; as *C. rastoni* [sic]). Similarly, lethal doses of venom appear much lower in *C. brevipedalia* when compared with *Alatina* sp. and *C. yamaguchii* (Nagai 2003). Note, however, that these hemolytic assays do not appear to have been

standardized among treatments, potentially making direct comparisons unreliable. Nonetheless, Bailey *et al.* (2005) also reported higher hemolytic activities in a species of *Carybdea* compared with two chirodropid species. However, hemolytic activity does not appear to be the lethal factor in the venoms investigated (Bailey *et al.* 2005), and hemolytic proteins represent only a fraction of the proteins present in cubozoan venom (Chung *et al.* 2001).

Sequencing of hemolytic proteins demonstrated two carybdeid and three chirodropid protein toxins to display a moderate amount of divergence (Nagai 2003; Brinkman & Burnell 2009). Despite reported differences among toxin protein sequences within Cubozoa, secondary structure models suggest at least two shared structural motifs that may be related to cytolytic activity (Brinkman & Burnell 2009 and references cited therein). Thus far no homologous protein outside of Cubozoa has been identified, suggesting that cubozoan venoms may contain a novel and unique family of proteins (Brinkman & Burnell 2009).

Several cubozoan species are known to cause a set of symptoms called Irukandji syndrome. Initially, Irukandji syndrome was attributed to *Carukia barnesi* whose sting causes a sharp prickling sensation without visible injury (Barnes 1964). Systemic effects are delayed by minutes to hours and include severe low back pain, progressing to limb cramping, nausea, vomiting, headache, restlessness and 'a feeling of impending doom' (Barnes 1964; Fenner 2006). Despite strong systemic effects Irukandji syndrome caused by *C. barnesi* is not considered life-threatening (Barnes 1964). Since its original description, the syndrome has been reported from, or attributed to, other cubozoans: *Morbakka* (Fenner *et al.* 1985), *Tamoya* (Morandini & Marques 1997), *Malo* (Gershwin 2005a 2007), *Alatina* (Yoshimoto & Yanagihara 2002 (as *C. alata*); Gershwin 2005b; Little *et al.* 2006) and *Gerongia* (Gershwin & Alderslade 2005).

Usually Irukandji syndrome in cubozoans other than *Carukia* is referred to as Irukandji-like syndrome (or as Morbakka syndrome by *Morbakka*; Fenner *et al.* 1985). Irukandji-like syndrome shares the same basic symptoms of classic Irukandji syndrome, but may be less severe in some species or even more severe in others. For example, a more severe case causing a fatality off North Queensland, Australia, was attributed to *C. barnesi*'s close relative *Malo kingi* (Fenner & Hadok 2002; Gershwin 2007; but see Bailey 2003). Thus far, very little is known about the mechanism(s) underlying Irukandji syndrome. Thorough toxicological studies of Irukandji causing species from disparate clades should clarify the function and nature of the syndrome.

While disparate clades in Cubozoa contain Irukandji causing species, all are part of Carybdeida. Irukandji syndrome is particularly well documented for species of Tamoyidae *sensu* Daly *et al.* (2007). Interestingly, *Tamoya* consistently falls outside Tamoyidae (Fig. 1; Appendices II-VII) and is easily distinguishable from its other genera, *Carukia*, *Malo*, *Gerongia* and *Morbakka*. Hence, we amend the meaning of Tamoyidae Haeckel, 1880 to contain all those carybdeid medusae that possess frown-shaped rhopaliar niche ostia lacking rhopaliar horns (type genus *Tamoya* Müller, 1859). We propose the new family Carukiidae, with type genus *Carukia* Southcott, 1967, to contain those carybdeids that lack gastric filaments and possess frown-shaped rhopaliar niche ostia with rhopaliar horns (genera *Carukia*, *Malo* Gershwin 2005, *Gerongia* Gershwin & Alderslade 2005 and *Morbakka* Gershwin 2008). Both Tamoyidae *sens. nov.* and Carukiidae branch before Carybdeidae and Tripedaliidae, while Alatinidae represents the earliest diverging carybdeid clade (?*Carybdea marsupialis* may be misidentified; see below). This topology suggests that the last common ancestor of Carybdeida probably possessed the mechanism(s) underlying Irukandji syndrome (Fig. 3). Further, the ability to cause Irukandji

syndrome may have been lost in the lineage leading to Carybdeidae and Tripedaliidae (Fig. 3; a species of *Carybdea* was linked to Irukandji syndrome [Little *et al.* 2006], but this attribution appears unconfirmed [Gershwin 2006a]). A syndrome described as Irukandji-like may be caused by a couple of non-cubozoan species (e.g., Fenner *et al.* 1996 [*Stomolophus nomurai*]; Fenner 1998 [*Gonionemus* and *Physalia*]), but homology will remain obscure until the mechanism(s) underlying the syndrome are clarified.

Even with a robust phylogeny, several problems hamper cubozoan venom studies. Difficulties in extracting venom, the use of whole tentacle tissue instead of isolated nematocysts and contradictory results among research groups (Brinkman & Burnell 2009 and references cited therein) need to be addressed. Furthermore, toxins have been reported to differ among different body parts of specimens and possibly different ontogenetic stages (Brinkman & Burnell 2009 and references cited therein). Finally, taxonomic uncertainties and resulting misidentification may impede toxicological research. For example, the number of bioactive proteins isolated from *C. marsupialis* from the Mediterranean (Rottini *et al.* 1995) differs from that of the same nominal species from the Caribbean (Sanchez-Rodriguez *et al.* 2006) leading to the interpretation of intraspecific venom variation. However, true *C. marsupialis* from the Mediterranean is easily distinguishable from its congeners in the Caribbean (i.e., *Carybdea xaymacana* and *C.* auct. *xaymacana*) by their gastric phacellae. The existence of *C. marsupialis* in the Caribbean is most probably an example of taxonomic confusion.

Courtship behavior Our working hypothesis with increased taxon sampling supports the preliminary finding that *Carybdea sivickisi* is more closely related to *Tripedalia cystophora* than it is to any species of *Carybdea* (Collins 2002; Collins *et al.* 2006). To retain monophyly of

Carybdea, we designate the new genus *Copula* to accommodate *Carybdea sivickisi* Stiasny, 1926; the name is in reference to the well-documented courtship behavior and sexual dimorphism (see below). We amend the meaning of Tripedaliidae Conant, 1897 to contain all carybdeids that display sexual dimorphism of the gonads, produce spermatophores and in which at least the males possess subgastral sacs/seminal vesicles (see Hartwick 1991a). ´ Species of *Carybdea* Peron & Lesuer, 1810, the sole genus within Carybdeidae Gegenbauer, 1857, can readily be differentiated from all other cubozoans by their possession of heart-shaped rhopaliar niche ostia (see Gershwin 2005b for diagnoses of *Carybdea* and Carybdeidae). The new genus *Copula* is defined to contain tripedaliids that possess adhesive pads on the exumbrellar apex with which they attach themselves to substrates when resting (see Hartwick 1991a); its type species is *Copula sivickisi* (Stiasny, 1926).

Tripedaliid life histories are unique among Cubozoa and Cnidaria. In *Copula sivickisi*, a mature male and female engage in sexual activity by entangling their tentacles. While swimming as a couple, the male brings its oral opening close to that of the female and produces a spermatophore that is ingested by the female (Lewis & Long 2005; Lewis *et al.* 2008). The subsequent gestation period spans some 2–3 days after which an embryo strand is released into the water column (Lewis & Long 2005; Lewis *et al.* 2008). Sexual dimorphism of medusae and similar courtship behavior were documented by Werner (1973) for *T. cystophora*, but he did not observe fertilization, gestation or embryo release. However, in contrast to the production of an embryo strand, *T. cystophora* seems to release free-swimming planulae (Conant 1898).

Species of both Carybdeidae and Alatinidae appear to be ovoviviparous and eggs are fertilized internally after female medusae have taken up sperm released into the water column by

males during spawning aggregations (Studebaker 1972; Arneson 1976). Neither courtship behavior nor sexual dimorphism appears to be present in these two families. Further, embryos are released within minutes to hours after fertilization (Studebaker 1972; Arneson 1976). To our knowledge, reproductive strategies of both Tamoyidae *sens. nov.* and Carukiidae remain undocumented. In Chirodropida, Yamaguchi & Hartwick (1980) reported external fertilization for medusae of both *C. fleckeri* and *Chiropsella bronzie* (as *C. quadrigatus*). While information from Carukiidae and Tamoyidae are needed, it appears that internal fertilization is a synapomorphy of Carybdeida. Further, we suggest that sexual dimorphism evolved a single time concomitant with complex sexual behavior (Fig. 3).

Biogeography. In general, cubozoan distributions are not well documented on intermediate geographical scales (e.g., provinces, states or countries) owing to a lack of sampling, which hampers biogeographic inquiries at this scale (see Bentlage *et al.* 2009 for a possible strategy to address this issue). However, on larger scales (e.g., ocean basins) several patterns emerge in light of our results. In particular, we uncovered numerous deep divergences among Indo-Pacific and Atlantic clades (Fig. 1). In Chirodropida, the genus *Chiropsalmus* is exclusively Atlantic, whereas the confamilial *Chiropsella* is from the Indo-Pacific groups with the exclusively Indo-Pacific *Chironex* (family Chirodropidae). Similarly, what had been recognized as Tamoyidae (Tamoyidae *sens. nov.* plus Carukiidae) can also be divided geographically: Tamoyidae is restricted to the Atlantic and the described species of Carukiidae are known from Australia and Japan, but probably range throughout the Indo-Pacific (Cleland & Southcott 1965; B. Bentlage 2009, unpublished notes). The pattern within Alatinidae and Tripedaliidae is unclear owing to limited taxon sampling. Both sampled species of Tripedaliidae,

Copula sivickisi and *T. cystophora*, can be found in all three oceans and future studies should seek to determine if these species are truly circumtropical or flocks of regional species.

Our densest sampling is in *Carybdea*, but unfortunately, relationships among Indo-Pacific and Atlantic/ Caribbean taxa lack support, so it is unclear whether there are deep divergences separating lineages into exclusively Atlantic/Caribbean and Indo-Pacific clades (Fig. 1). Nonetheless, integrating phylogeny and taxonomic investigations suggests that *Carybdea* spp. are more restricted in their geographical distributions than has been recognized by most workers. For example, *C. xaymacana* has been sampled from both the Caribbean and Western Australia, but deep divergence indicates crypticism in this nominal species. Similarly, *Carybdea rastonii* has traditionally been viewed as having a wide distribution with occurrence records from South Australia, Hawai'i and Japan among others. Our sampling shows that this is also a case of numerous species being united under the same name. Examination of the specimens suggests that *C. rastonii* can be distinguished morphologically (Gershwin & Gibbons 2009; B. Bentlage 2009, unpublished notes). Rather than having cryptic species in the sense that they are indistinguishable morphologically, this appears to be a case in which species have been proposed historically (*C. rastonii* Haacke, 1886 [South Australia], *C. brevipedalia* Kishinouye, 1891a [Japan] and *Carybdea arborifera* Maas, 1897 [Hawai'i]) but subsequently synonymized and/or disregarded.

Discovering that widespread nominal *Carybdea* spp. represent geographically isolated species assemblages indicates that these medusae do not exchange genetic material across large bodies of open water. Hence, we suppose that speciation in the genus is largely driven by vicariance. Dispersal events, however, cannot be ruled out as a source to account for

diversification. For instance, *C. arborifera* probably arose in Hawai'i after long-range dispersal; islands have existed at the present position of the Hawaiian Islands from at least the late Paleocene onwards, but were always remote (Carson & Clague 1995).

The inability to cross open ocean habitats is most likely a widespread phenomenon in Cubozoa, as most species appear to inhabit near shore habitats above the continental shelves (i.e., the neritic zone). Considering this, deep divergences between Atlantic and Indo-Pacific clades of Cubozoa may be explained by ancient vicariant events. Unfortunately fossil jellyfishes are rare, leading to uncertainty in dating cladogenetic events (Cartwright & Collins 2007).

However, fossils that possibly represent cubozoans have been discovered from the upper Jurassic (*Quadrumedusina quadrata* Haeckel, 1869), upper Carboniferous (*Anthracomedusa turnbulli* Johnson & Richardson 1968) and the middle Cambrian (Cartwright *et al.* 2007). Given their neritic habitat, cubozoans probably diversified as a result of plate movements in concert with eustatic sea-level fluctuations; the splitting of Pangea could have provided the setting for this.

In contrast to the general pattern described above, we have evidence for a pelagic cubozoan. The genus *Alatina* is represented with several nominal species in the Pacific (see Gershwin 2005b) from which we sampled *A. moseri* Mayer, 1906 and *A. mordens* Gershwin 2005 from or nearby their type localities (Hawai'i and the Coral Sea, respectively). We found no genetic divergences corresponding to geographical locality; in fact both 'species' share at least one 16S haplotype (Fig. 2). Additionally, no clear pattern differentiating the two populations exists. Rather, the haplotype network reflects a well-mixed population with regular gene flow. Inspection of specimens (including Hawaiian-type material USNM 22308, 22311 and 29632)

and study of its original description (Mayer 1906) demonstrates that *A. moseri* has been present in Hawai'i at least since the beginning of the twentieth century. The initial discovery of *Alatina* spp. in Australia seems not as well documented as is true for many marine invertebrates from this continent.

In contrast to other cubozoans, *Alatina* spp. live at or close to the edge of the continental shelf (Arneson & Cutress 1976; as *C. alata*) and have been obtained from great water depths before (e.g., Morandini 2003; as *C. alata*). It seems that *A. moseri* can only be encountered in shallow waters several days after the full moon (e.g., Thomas *et al.* 2001; Yanagihara *et al.* 2002; both as *C. alata*) when individuals congregate to spawn; the same is true for *A. mordens* (T Carrette & J Seymour 2008, personal communication). Furthermore, it has been suggested that individuals of *Alatina* spp. live up to 12 months (Arneson & Cutress 1976). Hence, it seems quite possible that *Alatina* spp. have an oceanic lifestyle and are able to maintain cohesive metapopulations across ocean basins.

Our investigation of historic specimens demonstrates that *A. moseri* was present in Hawai'i more than a century ago. An early introduction of *A. moseri* into Hawai'i is possible, but it seems unlikely that this would have occurred from the Coral Sea, given that ship traffic from Australia to Hawai'i was probably low at the time and the observed genetic signal would suggest multiple introductions rather than a single one. Considering the possible effect World War II naval traffic had on the spread of marine organisms (e.g., Coles *et al.* 1999), *A. moseri* may conversely have been introduced into the Coral Sea. Indeed, it is conceivable that *A. moseri* was introduced into the Coral Sea from Hawai'i and prior to that into Hawaiian waters from yet another location. However, considering the life cycle of *A. moseri* (and the synonymous *A.*

mordens), we find dispersal by natural means a more viable explanation of the pattern we observe. Investigation of additional *Alatina* spp. may show that some of these also represent artificial taxonomic units.

***Carybdea marsupialis*: a model organism misidentified?** We recover *C. marsupialis* together with *A. moseri* as the sister group to the remaining carybdeids. This placement appears surprising considering the stark morphological differences between *Carybdea* spp. and *Alatina* spp. (compare Gershwin & Gibbons 2009 with Gershwin 2005b). Specimens of *C. marsupialis* for this and other studies (Collins 2002; Collins *et al.* 2006) are derived from a polyp culture at the museum of the University of Hamburg, Germany. To our knowledge these polyps were originally obtained by B. Werner some 40 years ago in La Paguera, Puerto Rico (AC Arneson 2008, personal communication) and used for life cycle studies (e.g., Werner *et al.* 1971; Straehler-Pohl & Jarms 2005). *Alatina* spp. can be found in Puerto Rico, and considering the placement of *C. marsupialis* as a close relative of a member of Alatinidae, it is possible that the culture in Hamburg actually contains the polyp stage of a species of *Alatina* rather than *Carybdea*. Since this particular culture has served as the stock for several important experiments on *Carybdea* development (e.g., Stangl *et al.* 2002; Fisher & Hofmann 2004; Straehler-Pohl & Jarms 2005), it is vital to confirm the identification of the polyps by either rearing medusae to adulthood or collecting fresh material from the Caribbean for genetic comparisons. Inclusion of *C. marsupialis* from close to its type locality in Italy in future phylogenetic studies should also help shed light on this issue.

Endnote. The name *C. brevipedalia* is not in widespread usage, but its original description and type locality (Kishinouye 1891) demonstrate that it is the senior synonym of the name *C. mora*

Kishinoye, 1910, recently used as valid in Gershwin (2006b) and Gershwin & Gibbons (2009). Refer to chapter 2 of this thesis for additional details on this issue.

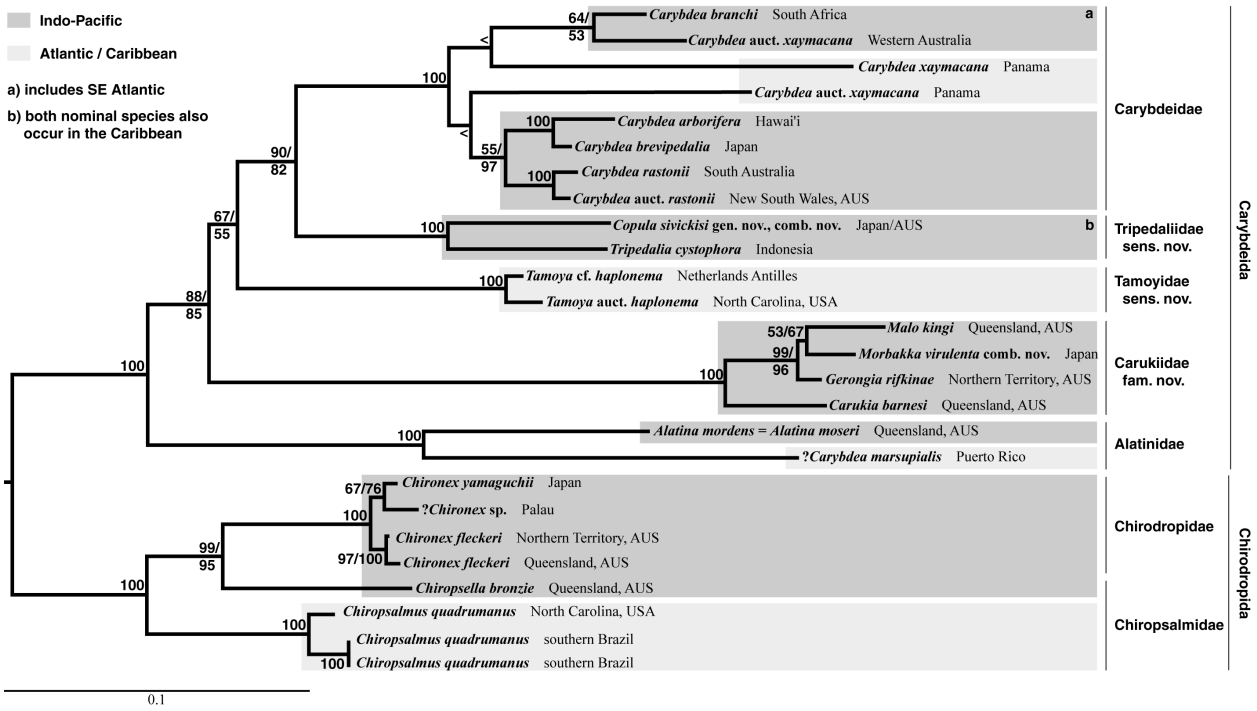


Fig. 1 Maximum likelihood topology (under GTR + I + G) of the combined nuclear LSU, SSU and mitochondrial 16S dataset. The alignment contains 5546 characters (LSU, 3292 characters; SSU, 1777 characters and 16S: 486 characters), of which 4369 are invariant and 936 parsimony informative. ML/MP parametric bootstrap support values (1000 replicates) are indicated on each node; if only one value is given it applies to both ML and MP. Less than symbol, bootstrap support less than 50; dark grey, Indo-Pacific; light grey, Atlantic/Caribbean; a, includes SE Atlantic and b, both nominal species also occur in the Caribbean.

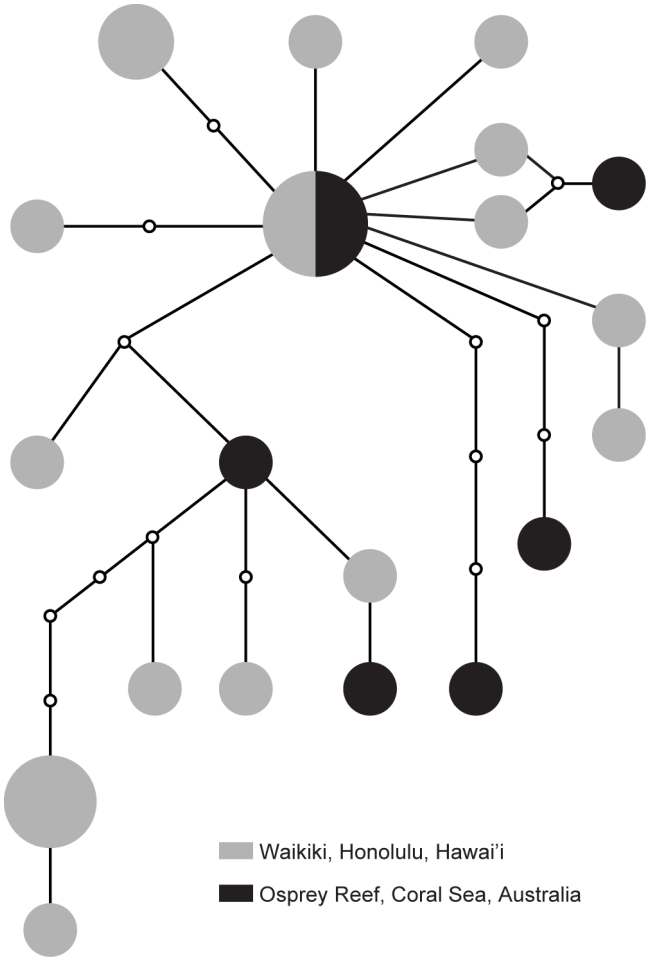


Fig. 2 Statistical parsimony network for mitochondrial 16S sequences of *Alatina* spp. from Hawai'i (*A. moseri*) and Australia (*A. mordens*) as calculated by TCS under the 95% connection limit criterion. Gaps were treated as a fifth character state. Lines represent one mutational step; small hollow circles correspond to inferred alleles that have not been sampled. The area of each respective solid circle reflects the number of alleles represented; the smallest solid circles represent a single allele. Grey circle, Waikiki, Honolulu, Hawai'i; black circle, Osprey Reef, Coral Sea, Australia.

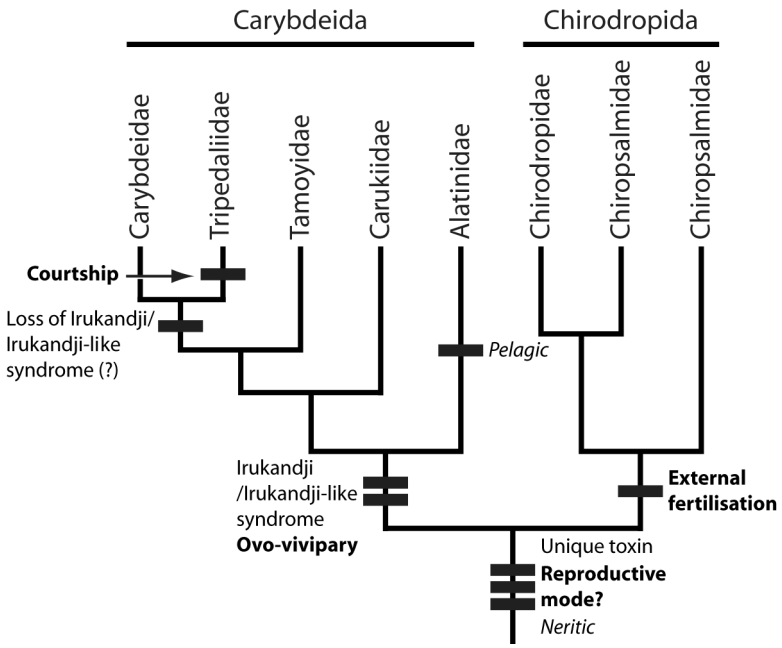


Fig. 3 Trends in toxicity and life-history evolution; phylogenetic relationships follow Fig. 1.

CHAPTER 2:

An illustrated key and synopsis of the families and genera of carybdeid box jellyfish (Cnidaria: Cubozoa: Carybdeida), with emphasis on the "Irukandji family" (Carukiidae)

Abstract

Box jellyfish (Cnidaria: Cubozoa) have a profound impact on human activities due to their highly potent venoms that may lead to severe envenomations in humans. Cubozoa is the smallest class within Cnidaria with only some fifty described species in seven families. The literature on Cubozoa is scattered and oftentimes difficult to access. In particular, comprehensive treatments of Cubozoa that present a comparative overview of the group are either non-existent or largely outdated. Here we present a synopsis of the carybdeid Cubozoa (Carybdeida) including an illustrated key to the families and genera of this order. Of particular interest is the family Carukiidae which contains the species that was originally attributed with causing a severe envenomation syndrome called Irukandji syndrome. One new species of Carukiidae, *Malo filipina* sp. nov., is described, a poorly known one, *Morbakka virulenta*, redescribed, a neotype designated, and an unidentified species of *Morbakka* is recorded from the Philippines.

Introduction

Box jellyfish comprise the smallest class of medusozoan cnidarians, Cubozoa, and are the closest relatives of the "true jellyfish" of the class Scyphozoa (see Collins 2009 for a working hypothesis of cnidarian relationships and classification). Cubozoans display a metagenetic life cycle involving a benthic polyp stage and free-swimming medusa stage similar to that found in scyphozoan jellyfish (Straehler-Pohl and Jarms 2005). The low diversity seen in the class Cubozoa may be the result of evolutionary history (e.g., a lack of opportunity for allopatric speciation) or simply an artifact stemming from lack of taxonomic study of the group (Bentlage *et al.* 2010). Regardless of diversity, cubozoans have a profound impact on human activities (e.g., Gershwin *et al.* 2010). Cubozoan medusae are particularly well-known among the general public and scientific community for their highly potent venoms. Human-box-jellyfish encounters regularly make news headlines, as these encounters have the potential to be life-threatening to the victim. Not all box jellyfish are equally venomous to humans however, making reliable identification of box jellyfish an important issue for informed risk management.

Cubozoa contains two monophyletic orders, Chirodropida and Carybdeida (Fig. 4; Collins *et al.* 2006; Collins 2009; Bentlage *et al.* 2010). The medusa stages of both chirodropids and carybdeids possess a cuboidal swimming bell. In contrast to chirodropids, carybdeids possess only a single tentacle per pedalium (Fig. 5). Chirodropida contains those species of box jellyfish that are infamous for their ability to kill a healthy, adult human within minutes (e.g., *Chironex fleckeri*). In contrast, envenomations by carybdeid cubozoans are generally considered less of a threat, even though they may still be severely painful or lethal in some cases (e.g., Gershwin *et al.* 2010). In particular, several carybdeid medusae have been attributed with the

ability to cause Irukandji syndrome in humans (e.g., see review by Bentlage *et al.* 2010). Irukandji syndrome produces systemic effects in humans, the onset of which is delayed by several minutes to hours following the initial envenomation. Among the symptoms are excruciating lower back pain, nausea, vomiting, and headaches (e.g., Barnes 1964), in most cases requiring hospitalization for treatment of symptoms (e.g., Winkel *et al.* 2003; Gershwin *et al.* 2010). Initially, Irukandji syndrome was attributed to *Carukia barnesi*, a member of the recently described cubozoan family Carukiidae, but Irukandji syndrome and Irukandji-like envenomations have been attributed to other species, genera, and families within Carybdeida since the original description of the syndrome (see review by Bentlage *et al.* 2010).

Here, we provide a synopsis and facilitate the identification of carybdeid cubozoans by providing an illustrated taxonomic key. Because much remains to be learned about species identities within Cubozoa, we generally have not provided species diagnoses, but instead provide citations of key papers with further details. Many species remain poorly described and the continued discovery of unknown cubozoans is highly likely. Therefore, it is difficult and sometimes even undesirable to identify specimens to species-level. Reliable family- or genus-level identifications, however, should be possible for non-specialists and of great value, because, for example, members of the same genus often share similar traits in terms of toxicology and/or ecology (Bentlage *et al.* 2010). The literature on cubozoans is scattered throughout various publications in several different languages over more than a century. Thus, information on cubozoans is often difficult to access and evaluate. The synthetic information presented here will facilitate identification of carybdeid medusae.

Furthermore, we clarify the identities and diagnoses of the genera of Carukiidae

(*Carukia*, *Gerongia*, *Malo*, and *Morbakka*). Thus far, all genera and species of the Carukiidae have been described from Australia (but see Bentlage *et al.* 2010). In this contribution, a new species, *Malo filipina*, is described from the Philippines, and a previously poorly known species, *Morbakka virulenta*, is redescribed from Japan and a neotype designated to stabilise its taxonomic status. In addition, specimens from the Philippines were discovered in the collections of the National Museum of Natural History, Smithsonian Institution in Washington DC that represent a species of *Morbakka* (whether this species is known to science is uncertain due to the poor condition of the specimens). These findings are particularly noteworthy, because envenomations typical of Irukandji syndrome are reported from Asia in newspapers or online blogs, and in some cases, the scientific literature (Gershwin *et al.* 2010); however, the identities of the species responsible for these envenomations, are unknown for the most part. Description of species of Carukiidae from Asia is a necessary step toward establishing the links between the species and the victims in severe jellyfish envenomations.

Materials and Methods

The taxonomic framework employed followed Daly *et al.* (2007) and modifications to that classification proposed by Bentlage *et al.* (2010) based on phylogenetic analyses. Cubozoan specimens (preserved or live) were studied from the National Museum of Natural History, Smithsonian Institution in Washington DC (USA), the Queensland Museum in Brisbane (Australia), the Museum of Comparative Zoology at Harvard (USA), the Australian Museum in Sydney (Australia), and at field sites in Japan, tropical Queensland (Australia), and on the East Coast of the USA. The taxonomic key presented here represents the synthesis of our

investigations and observations. In addition to morphology and anatomy, we used light microscopy to examine nematocysts, following the procedures and nomenclature in Collins *et al.* (2011). Nematocyst measurements are given as minimum-*mean*-maximum. Species that we considered problematic due to the absence of a reliable description were labeled as *nomen dubium* and, in most cases, should be disregarded at this time. Generalised geographic distributions are outlined for each family.

Results

Key to the families and genera of Carybdeida. Use of the following key requires access to a dissecting microscope, and specimens should be inspected while submerged in a liquid (e.g., water). The morphological and anatomical characters of carybdeid medusae are given in Fig. 6 and Table 1. Often dissection of the specimen is necessary. Several characters are located in the subumbrellar cavity of the medusa, thus cutting a specimen in half along the sagittal plane may be advisable. In addition, it may be necessary to open the stomach to examine the gastric phacellae. It is particularly easy to overlook the rhopaliar horns characteristic of the Carukiidae. In order to investigate these, the rhopaliar niche should be submerged in liquid, and light should be directed laterally at the niche. Adjusting the angle of the light may be necessary to determine the presence and shape of the horns.

Class **Cubozoa** Werner, 1973

Order **Carybdeida** Gegenbauer, 1857

Cubozoa with a single tentacle per pedalium (Fig. 5).

Families: Alatinidae, Carukiidae, Carybdeidae, Tamoyidae, and Tripedaliidae.

1. Rhopaliar niche ostium t-shaped (Fig. 6b).....Alatinidae

Rhopaliar niche ostium not t-shaped.....2

2. Rhopaliar niche ostium heart-shaped (Fig. 6c).....Carybdeidae

Rhopaliar niche ostium keyhole-shaped (Fig. 6d) or frown-shaped (horizontal oval; Figs 3e and f).....3

3. Rhopaliar niche ostium frown-shaped with rhopaliar horns (Fig. 6g); stomach lacking gastric phacellae.....Carukiidae

Rhopaliar niche ostium frown-shaped (Fig. 6f), lacking rhopaliar horns; gastric phacellae vertical (Figs 3h and i).....Tamoyidae

Rhopaliar niche ostium keyhole-shaped (Fig. 6d) or alternatively rhopaliar niche ostium frown-shaped (Fig. 6e); 2 or 3 pedalia per corner instead of 1 (Fig. 6j).....Tripedaliidae

Family **Alatinidae** Gershwin, 2005*a*

Tentacles simple.....Genus *Alatina* (Figs 2e, 3b and k)

Tentacles branching distally.....Genus *Manokia* (no Fig.)

Alatina Gershwin, 2005*a* (refer to Gershwin 2005*a*; Bentlage *et al.* 2010; Bentlage 2010 for reviews of the genus). Species: *A. alata* (Reynaud, 1830) *nomen dubium*; *A. grandis* (Agassiz & Mayer, 1902); *A. madraspanata* (Menon, 1930); *A. moseri* (Mayer, 1906); *A. mordens* Gershwin, 2005*a* = *A. moseri* (Bentlage *et al.* 2010; Bentlage 2010); *A. obeliscus* (Haeckel, 1880) *nomen dubium*; *A. philippina* Haeckel, 1880 *nomen dubium*; *A. pyramis* (Haeckel, 1880)

nomen dubium; *A. rainiensis* Gershwin 2005*a*; *A. turricula* (Haeckel, 1880) *nomen dubium*.

Manokia Gershwin, 2005*a* (refer to Gershwin 2005*a* for a review of the genus). Species: *M. stiasnyi* (Bigelow, 1938).

Geographic distribution of the family: global; tropical to temperate; neritic, and oceanic.

Alatinidae *sensu* Gershwin (2005*a*) contains all those Carybdeida with crescentic gastric phacellae, comprised of long cirri arranged (more or less) in a single plane (i.e., the genera *Alatina* and *Manokia*; Fig. 6k). The rhopaliar niche ostia of Alatinidae are t-shaped (i.e., comprised of a single upper covering scale and 2 well-developed lower covering scales; Fig. 6b). Gershwin (2005*a*) also includes "3 or 4 more or less simple velarial canals per octant" in her diagnosis. This, however, provides problems for accommodating *Alatina madraspanata*, which clearly belongs to the family Alatinidae (Gershwin 2005*a*) but possesses more than 4 velarial canals per octant. Therefore, velarial canals should be omitted from the diagnosis of the family.

Among the Alatinidae is *A. alata*, a species originally classified as *Carybdea alata* (Reynaud 1830; Gershwin 2005*a*). Over time, numerous species were synonymised with *A. alata* (Mayer 1910; Kramp 1961), but were recently considered to be separate species (Gershwin 2005*a*). Molecular genetic analysis, however, suggests that at least some of the nominal *Alatina* species may indeed be part of a more widespread species-group (Bentlage *et al.* 2010). In particular, *A. moseri* and *A. mordens* are presumed to be synonymous based on molecular genetic data (the mitochondrial ribosomal 16S gene; Bentlage *et al.* 2010), suggesting that the characters used to distinguish the two result from intra-specific rather than inter-specific variation. The biogeography of *Alatina* makes it quite possible that additional species are part of a widespread

species-group (Bentlage 2010). Therefore, the validity of many species of *Alatina* remains questionable and the taxonomic characters used for species delineation need to be reevaluated in light of molecular genetic data (Bentlage 2010; Bentlage *et al.* 2010). In particular, a neotype for *A. alata*, the type species of the genus *Alatina*, should be designated, as this species was the earliest one described in the genus *Alatina*. The name *A. alata* is prevalent throughout the literature (usually as *Carybdea alata*), but based on its original description the species is unidentifiable (Gershwin 2005*a*). We are not aware of specimens from the Atlantic (the type locality of *A. alata*) that are suitable for both molecular genetic and morphological investigation; therefore, we do not designate a neotype for *A. alata* at this time.

Family **Carybdeidae** Gegenbauer, 1857

Only one genus.....Genus *Carybdea* (Figs 2d, 3a and c)

Carybdea Péron & Lesueur, 1809 (refer to Mayer 1910; Gershwin and Gibbons 2010; Bentlage *et al.* 2010 for reviews of the genus). Species: *C. arborifera* Maas, 1897, *C. aurifera* Mayer, 1900 *nomen dubium*, *C. branchi* Gershwin, 2009, *C. brevipedalia* Kishinouye, 1891*a*, *C. latigenitalia* Kishinouye, 1891*a nomen dubium* ?= *C. brevipedalia*, *C. marsupialis* (Linnaeus, 1758), *C. mora* Kishinouye, 1910 = *C. brevipedalia*, *C. morandinii* Straehler-Pohl & Jarms, 2011 (only the polyp and early stages in medusa development are known [Straehler-Pohl and Jarms 2011] and it is possible that this species does not belong to *Carybdea* or is synonymous with another cubozoan), *C. murrayana* Haeckel, 1880, *C. prototypus* (Haeckel, 1880) *nomen dubium*, *C. rastonii* Haacke, 1886, *C. verrucosa* Hargitt, 1903 *nomen dubium*, *C. xaymacana* Conant, 1897

Geographic distribution of the family: global; tropical to temperate; neritic.

Carybdea represents the oldest described genus within Cubozoa and many cubozoan species had been classified in this genus prior to the taxonomic revisions of the last decade. Most notably *Carybdea alata* was reclassified as *Alatina alata* (Gershwin 2005*a*; see above) and *Carybdea sivickisi* was reclassified in the Tripedaliidae as *Copula sivickisi* (Bentlage *et al.* 2010; see below). At present Carybdeidae represents a monogeneric family containing the sole genus *Carybdea*. All members of this family can readily be recognised by their heart-shaped rhopaliar niche ostia (details in Gershwin and Gibbons 2009; Collins *et al.* 2011). Gershwin and Gibbons (2009) present a comparative overview of the species of *Carybdea* and the morphological characters distinguishing these species. However, some species are missing from that study (e.g., *C. arborifera*).

Another species contained in Gershwin and Gibbons' (2009) treatment is *C. mora*, which was recently considered to be the junior synonym of *C. brevipedlia* (Bentlage *et al.* 2010). Both species were described by Kishinouye (1891*a*, 1910) from Honshu, Japan. *C. brevipedlia* was described from Shima (Kishinouye 1891*a*) whereas *C. mora* was described from Tokyo Bay (Kishinouye 1910) only some 200-300 km north of Shima. The descriptions of both species are lacking considerable detail and provide little to distinguish between the two species. In fact, both species seem to overlap in their morphological characteristics. In particular, the gastric phacellae consist of rows of 10-12 brush-like filaments in each corner of the stomach in *C. brevipedalia* (Kishinouye 1891*a*) and 8-12 brush-like filaments in *C. mora* (Kishinouye 1910). We investigated several specimens of *Carybdea* collected from different locations in Japan and all

possessed similar numbers of brush-like filaments in the corners of their stomachs and did not seem to differ in any other way. Pending further evidence, molecular and morphological, it seems most prudent to consider *C. mora* to be the junior synonym of *C. brevipedalia* (cf. Bentlage *et al.* 2010); the species is commonly known as Andon Kurage in Japan.

In summary, original descriptions as well as those provided by Mayer (1910) should be consulted in addition to Gershwin and Gibbons (2009) for identifying species of *Carybdea*. It is likely that several undescribed species of *Carybdea* exist, a view supported by molecular genetic evidence (Bentlage *et al.* 2010). At the same time some species may be synonymous with each other. As such, the family Carybdeidae would benefit from a comprehensive taxonomic revision combining both morphological and molecular genetic data.

Family **Carukiidae** Bentlage *et al.* 2010

1. Rhopaliar horns straight and narrow (Fig. 7a); 2 broad, usually unbranched velarial canals per octant (Fig. 7b); perradial lappet warts lacking (Fig. 7b) or lappets with single wart on each side; tentacles with “neckerchiefs” (Fig. 7c).....Genus *Carukia* (Figs 4a–c)

Rhopaliar horns and velarial canals different from above; tentacles without “neckerchiefs”.....2

2. Rhopaliar horns short, broad, blunt or with pointed tips (Fig. 7d); 3 to 4 unbranched or mildly branched velarial canals that originate from 1 root, somewhat palmate (Fig. 7e); 2 rows of 1-4 perradial lappet warts (Fig. 7e).....Genus *Malo* (Figs 2b, 3g, 4d and e)

Velarial canals heavily branching with or without lateral diverticula; 2 rows with more than 3-4 perradial lappet warts ….....3

3. Rhopaliar horns short, broad, curved (devil-horn shaped; Fig. 7f); numerous laminar branching

velarial canals without diverticula (Fig. 7g); 2 rows of 3-6 perradial warts per row (usually 5; Fig. 7g).....Genus *Gerongia* (Figs 4f and g)

Rhopaliar horns rabbit-ear like (Figs. 7h and i); numerous heavily branching velarial canals with lateral diverticula (Fig. 7j); 2 rows of numerous perradial warts plus scattered warts (Fig. 7j).....Genus *Morbakka* (Figs 4h–j)

Carukia Southcott, 1967 (refer to Southcott 1967; Gershwin 200*5b*), Species: *C. barnesi* Southcott, 1967, *C. shinju* Gershwin 2005*b*

Gerongia Gershwin & Alderslade, 2005 (refer to Gershwin and Alderslade 2005). Species: *G. rifkinae* Gershwin & Alderslade, 2005

Malo Gerswhin, 2005*b* (refer to Gershwin 2005*b*; Gershwin 2007; see species descriptions below). Species: *M. kingi* Gerswhin, 2007, *M. filipina* sp. nov., *M. maxima* Gershwin, 2005*b*

Morbakka Gershwin, 2008 (refer to Gershwin 2008; see species descriptions below). Species: *M. fenneri* Gershwin, 2008, *M. virulenta* (Kishinouye, 1910)

Geographic distribution of the family: Indo-Pacific; tropical to (warm) temperate; neritic, and possibly oceanic.

Carukiidae contains species of box jellyfish first identified as causing a serious envenomation syndrome called Irukandji syndrome. Prior to Bentlage *et al.* (2010), the genera contained now in Carukiidae were classified in the Tamoyidae alongside *Tamoya* (see below; Gershwin and Alderslade 2005; Daly *et al.* 2007). At present, Carukiidae are only known from the Indo-Pacific ranging from New South Wales in Australia to Honshu in Japan on the North-South axis. Much

less is known about the eastern and western limits of the distribution of Carukiidae. Gershwin and Alderslade (2005) provide a tabular comparison of the carukiid genera (then called Tamoyidae). We illustrated the main characters to distinguish between the genera of Carukiidae in Fig. 7.

The close relationship and the intergrading and difficult-to-interpret characters of *Malo*, *Morbakka*, and *Gerongia* (see Gershwin 2005*b* 2008; Gershwin and Alderslade 2005; Bentlage *et al.* 2010), suggest that these three genera should be synonymised. Among the problematic characters invoked to differentiate these genera was the number of eyes per pedalium (Gershwin and Alderslade 2005); however, the eye pigment may fade in fixative and lead to inaccurate counts (Bentlage 2010). In our opinion, differences in the shape of the rhopaliar horns, the branching patterns of the velarial canals, and the number and arrangement of nematocyst warts on the perradial lappets are the most reliable morphological characters to distinguish the genera of Carukiidae. In addition, nematocyst wart counts and patterns on the velarium may be distinct among carukiid genera. In *Carukia*, velarial warts are either absent or present as a single wart in each octant. *Malo* possesses 2-4 warts per octant, whereas *Morbakka* possesses 6-8 warts, even though Gershwin (2008) notes that some specimens of *M. fenneri* lack warts altogether. Our review of the literature indicates that velarial warts might be absent in *Gerongia*. Velarial warts were neither mentioned in the literature nor visible in any of the published images; however, living specimens may possess velarial warts that were lost in preserved material due to abrasion. Additionally, medusae of each genus of Carukiidae seem to differ in size, albeit sizes may overlap. Maximum bell heights of *Carukia* range from 1-2 cm, of *Malo* from 2-5 cm, of *Gerongia* about 6 cm, and of *Morbakka* from 10-15cm (Gershwin and Alderslade 2005;

unpublished observations).

Records of Carukiidae species are fairly rare and are mostly limited to the Australian continent. To add to the knowledge of the diversity and distribution of Carukiidae, we describe a new species of *Malo* from the Philippines, redescribe *Morbakka virulenta* from Japan, and document the discovery of an unknown/unidentified species of *Morbakka* from the Philippines. *Morbakka virulenta* was originally described as *Tamoya virulenta* by Kishinouye (1910). The species, however, lacks the vertical gastric phacellae characteristic of *Tamoya* (see Collins *et al.* 2011), but possesses the distinctive "rabbit-ear" shaped rhopaliar horns of *Morbakka*. The original type material appears lost and our inquiries among Japanese colleagues did not reveal where potential type specimens may be located. It is likely that the material investigated for the original description of *M. virulenta* was lost circa WWII. *Morbakka virulenta* is very similar in appearance to *M. fenneri* from Australia. To aid future taxonomic studies, we designate a neotype for *M. virulenta* and provide a description of the material examined in another section of the manuscript below.

Family **Tamoyidae** Haeckel, 1880

Only one genus.....Genus *Tamoya* (Figs 2a, 3f, h and i)

Tamoya Müller, 1859 (refer to Collins *et al.* 2011 for a review of the genus). Species: *T. bursaria* (Lesson, 1829) *nomen dubium*, *T. haeckeli* Southcott, 1967 = *T. gargantua*, *T. gargantua* (Lesson, 1829) *nomen dubium*, *T. ohboya* Collins *et al.*, 2011, *T. prismatica* Haeckel, 1880 = *T. haplonema*, *T. haplonema* Müller, 1859

Geographic distribution of the family: Atlantic Ocean and Caribbean, with most records

from the western Atlantic and the Caribbean; tropical to temperate; mostly neritic.

Gershwin and Alderslade (2005) considered the genera of Carukiidae to be part of the family Tamoyidae. The phylogenetic and taxonomic treatment of Bentlage *et al.* (2010) shows that Tamoyidae is a monogeneric family containing only the genus *Tamoya*. Morphologically, Tamoyidae can be distinguished from the Carukiidae based on the lack of rhopaliar horns. In addition, *Tamoya* species possess vertical rows of gastric cirri running along the perradial sides of the stomach (Collins *et al.* 2011). Collins *et al.* (2011) review the Tamoyidae with details of its history and morphology/anatomy.

Some species of *Tamoya* that currently are unrecognizable were originally described from the Indo-Pacific (e.g., *T. bursaria*). Additionally, cubozoans collected from the Indo-Pacific have regularly been identified as species of *Tamoya*. When inspection of such specimens was possible, we determined that the specimens belong to the Carukiidae (e.g., *Morbakka virulenta* from Japan that was described, and is regularly referred to, as *Tamoya virulenta*). All records of *Tamoya* originating from the Indo-Pacific should be treated as suspect; these records most likely refer to specimens of *Morbakka* or *Gerongia* that were erroneously identified as *Tamoya*, because the lack of gastric phacellae and presence of rhopaliar horns are often overlooked.

Family **Tripedaliidae** Conant, 1897

A single pedalium per bell corner.....Genus *Copula* (Figs 2c and3d)

2 or 3 pedalia per bell corner.....Genus *Tripedalia* (Figs 3e and j)

Copula Bentlage *et al.*, 2010. Species: *C. sivickisi* (Stiasny, 1926)

Tripedalia Conant, 1897. Species: *T. binata* Moore, 1988; *T. cystophora* Conant, 1897

Geographic distribution of the family: Global; tropical; neritic.

Tripedaliidae previously comprised only one genus, *Tripedalia*, and was immediately recognizable because both species of *Tripedalia* possess more than one pedalium per corner of the swimming bell (two in *T. binata* and three in *T. cystophora*). Bentlage *et al.* (2010), however, determined that *Copula sivickisi* (previously *Carybdea sivickisi*) forms a monophyletic clade with *Tripedalia* and thus changed the diagnosis of the family to accommodate *C. sivickisi*. The present diagnosis of the family pertains to characters that are mainly visible in mature specimens. Tripedaliidae, according to Bentlage *et al.* (2010), contains all those carybdeids that display sexual dimorphism of the gonads, produce spermatophores, and in which at least the males possess sub-gastric sacs/seminal vesicles (details in Hartwick 1991). These characters are important in the reproduction of tripedaliids. Species of Tripedaliidae engage in courtship behaviour, unusual among cnidarians, that involves coupling and the active transfer of a spermatophore from the male to the female (Werner 1973; Hartwick 1991; Lewis and Long 2005; Lewis *et al.* 2008). Tripedaliids have wide geographic distributions ranging from the Indo-Pacific (e.g., Moore 1988; Hartwick 1991; Lewis *et al.* 2008) to the Atlantic, including the Gulf of Mexico and Caribbean (e.g., Marques *et al.* 1997; Migotto *et al.* 2002; Coates 2003; Orellana and Collins 2011). Because of the small size of these species (< 1 cm to about 2 cm bell height) and their neritic habitat, it is likely that several cryptic species remain to be discovered in these genera.

Non-Australian members of the "Irukandji family" (Carukiidae). Here we describe specimens of carukiid species that were collected from the Philippines and Japan. The following abbreviations were used throughout the descriptions. National Museum of Natural History, Smithsonian Institution, USA: USNM; Bell height in mm: BH (measured from velarial turn-over to the apex of the swimming bell); interradial bell width in mm: IRW; nematocyst capsule length: L; nematocyst capsule width: W.

Malo filipina, sp. nov. from the Philippines

Figs 7d and e, 8, 9

Synonymies. Carybdea rastonii: Mayer (1910: pp 508-509 later became USNM 27937*); Mayer (1915: p. 170; USNM 27937* and USNM 28714); Mayer (1917: pp.187–188; USNM 27935, USNM 27936, USNM 27937, USNM 28714); *: we were unable to locate and inspect USNM 27937.

Material examined. Holotype: USNM 27935, female 32 mm BH, 13 mm IRW, Nasugbu, Luzon, The Philippines. Paratypes: USNM 28714, male 30 mm BH, 16 mm IRW, Taal Anchorage, Luzon, The Philippines; USNM 1150373, male, 40 mm BH, 15 mm IRW, Nasugbu, Luzon, The Philippines; USNM 27936, male 37 mm BH, 18 mm IRW, Mansalaya, Mindoro, The Philippines.

Type locality. Nasugbu, Luzon, The Philippines.

Etymology. The species name indicates the geographic origin of the species.

Diagnosis. Malo of 30-40 mm BH; bell densely covered with nematocyt warts. Pedalia with thorn-like extension at pedalial canal bend. Tentacular cnidome consisting of microbasic p-

mastigophores and microbasic p-euryteles.

Description. Carybdeid medusa, bell taller than wide with leaf-life gonads in mature individuals (Fig. 8a). Exumbrella with regularly spaced nematocyst warts (Fig. 8a). Maximum BH about 40 mm (observed range from 30 to 40 mm in mature individuals) and maximum IRW about 18 mm (observed range from 13 to 18 mm in mature individuals). Stomach reaching deep into subumbrella, suspended with well-developed mesenteries (Fig. 8b); gastric phacellae absent. Upper half of mesenteries well developed while lower half extends cord-like to rhopaliar window (Fig. 8b). Manubrium short, extending 1/3 of bell height into subumbrellar cavity; with smooth and somewhat rounded lips (Fig. 8b). Four muscular brackets (frenulae) brace the right-angle connection from tip of rhopaliar window to 3/4 the distance between velarial turnover and its margin on each perradius; each frenulum consisting of 1 solid gelatinous sheet (Fig. 8c). Perradial lappets broad, triangular, not reaching subumbrellar edge of velarium, with 1 row of 2-4 nematocyst warts on each side (Figs 4e, 5d and e). 3-4 velarial canals per octant; velarial canals appear as digitiform projections (some branched) from a single root giving a palmate appearance to velarial canals (Figs 5d and e). 4-6 velarial warts per octant. Pedalia with scalpel shaped inner keel and distal overhang (Fig. 8f); pedalial canal thorn-shaped (Fig. 8g). Nematocyst warts present on abaxial portion of pedalia (Fig. 8f), but most warts appear to have rubbed off. Rhopaliar niche ostium frown-shaped with short, broad, and blunt rhopaliar horns (Figs 4d and 5h). Each of the four rhopalia bears two median lens eyes; lateral eyes not visible due to poor condition of specimens. It is unclear if the lateral pigment and slit eyes are not present or if the pigment has faded over time, leaving only the two central lens eyes discernible.

Cnidome. Nematocysts collected and measured from USNM 28714. Distal tentacle tip:

rod-shaped, microbasic p-mastigophore (L 27.2-*31.5*-36.4 µm, W 11.6-*13*-14.2 µm, n=20; Fig. 9a). Proximal tentacle/tentacle base: rod-shaped, microbasic p-mastigophore (L 38.8-*40*-43.8 µm, W 12.8-*15.6*-17.7 µm, n=20; Figs 6b and c); oval, microbasic p-eurytele (L 29.4-*30.5*-31.3 µm, W 16.5-*16.8*-17.4 µm, n=5; Figs 6d and e). Exumbrellar warts: large, spherical isorhiza (L 28.5-*32*-40.5 µm, W 27.1-*29.4*-32.9 µm, n=20; Fig. 9f); oval, microbasic p-eurytele (L 30.6-*34.3*-36.6 µm, W 20.4-*22.1*-24.3 µm, n=8; Figs. 9g and h); ?rod-shaped, microbasic p-mastigophore (L 36.7-*38.6*-40.1 µm, W 15.3-*16*-17.2 µm, n=10; Fig. 9i).

Differential Diagnosis. M. filipina is most likely to be confused with *M. maxima* from Western Australia, especially because both species are of similar size. *M. filipina*'s tentacular cnidome contains both microbasic p-mastigophores and microbasic p-euryteles whereas *M. maxima* possesses only the former. We observed the following nematocysts in tentacles of a specimen of *M. kingi* (USNM 1125368), the third species in the genus: spherical isorhizas (average L 26 µm, average W 21 µm), microbasic p-mastigophores (average L 36 µm, average W 13 µm), and amastigophores (average L 7 µm, average W 5 µm). Furthermore, the shape of the pedalial canal bend allows distinction among the species of *Malo*. *M. maxima* lacks a spike or thorn-like extension at the proximal bend of its pedalial canal bend whereas *M. filipina* possesses a spike at the pedalial canal bend (Fig. 8g). *M. kingi* also lacks the thorn-like extension of *M. filipina*. Several specimens of *M. kingi* seem to display halo-like bands on their tentacles (Gershwin 2007); no such structures were observed in *M. filipina*.

Morbakka virulenta (Kishinouye, 1910) from Japan

Figs 7i and j, 10, 11

Synonymies: Tamoya virulenta: Kishinouye (1910: p. 7, Pl. I, fig. 10); Uchida (1947: p. 316); Williamson *et al.* (1996: p. 414). *Hikurage*: Kishinouye (1891b: pp. 508 – 509). *Tamoya alata*: Uchida (1929: p. 172, Figs 8–88). non *Tamoya alata* Reynaud [refers to *Copula sivickisi*]: Uchida (1929: pp. 178-180; Figs 86–87). *Tamoya bursaria*: Stiasny (1929); Uchida (1947: pp. 314–316, figs. 2, 3); Uchida (1954, pp. 209-219);

Tamoya haplonema: Uchida, T (1970; pp 289, 293-294, Figs 3 and 4); Yamaguchi, M (1982: p. 249); Kubota, S (1998: p. 33); Iwama (2001; pg 109); Yamasu & Yoshida (1976: pp. 325–326). *Tamoya bursaria* (?*gargantua*): Williamson *et al.* (1996: p. 414). *Tamoya gargantua*: Williamson *et al.* (1996: p. 414)

Material examined. Neotype: USNM 1124253, female, 150 mm BH, 60 mm IRW, Hiroshima Bay, Japan. *Other material:* USNM 1124251, 150 mm BH, 68 mm IRW, Hiroshima Bay, Japan; USNM 1124252, 140 mm BH, 50 mm IRW, Hiroshima Bay, Japan.

Type locality. Hiroshima Bay, Honshu, Japan.

Common name. Hikurage ("fire jellyfish" in Japanese).

Diagnosis. Large carybdeid medusa (up to 150 mm BH) lacking gastric phacellae; robust bell densely covered with nematocyst warts (Fig. 10a). *M. virulenta* possesses "rabbit-ear" shaped rhopaliar horns and differs from its congener, *M. fenneri*, in that *M. virulenta*'s rhopaliar horns are swollen rather than pointed at the tips, and project from the top center of the rhopaliar niche at a more oblique angle (Fig. 7h vs. 7i). Further, *M. virulenta* lacks the nematocyst wart on the rhopaliar stalk that is characteristic of *M. fenneri.*

Description. Carybdeid medusa, bell taller than wide, rectangular, with flat apex, and leaf-like gonads in mature individuals (Figs 7a and b); extended tentacles flat, cord-like in live

specimens (Fig. 10b). Exumbrella densely covered with nematocyst warts (Fig. 10a). Maximum BH about 150 mm (observed range from 140 to 150 mm in mature individuals) and maximum IRW about 68 mm (observed range from 50 to 68 mm in mature individuals). Gastric phacellae absent; stomach with well-developed musculature (area corrugata; Fig. 10c). Manubrium with rounded, smooth-edged lips (Fig. 10d) extending to about 1/3 of BH from subumbrellar ostium. Mesenteries well developed in upper 1/2 of the subumbrella, extending cord-like to rhopaliar window (Fig. 10d). Frenulum consisting of a single sheet that splits longitudinally near the rhopaliar niche, extending onto the lower half of the rhopaliar window (Fig. 10e); frenulum covering 3/4 of the distance between velarial turnover and velarial margin (Fig. 10e). Perradial lappets broad, triangular, not reaching subumbrellar edge of velarium (Fig. 10f); approximately 7-10 nematocyst warts on each side of the perradial lappet, mostly arranged in single rows but some scattered (Figs 7f and 4j). Velarium with 6-8 complex dendritic canals per octant with lateral diverticula per octant (Figs 7f and 4j). Pedalium keeled on both sides of lateral median line; inner keel with overhang (Fig. 10g). Outer keel with warts (Fig. 10g), that appear to have mostly rubbed off after collection. Pedalial canal with thorn-like extension (Fig. 10h). Rhopaliar niche ostium frown-shaped with rabbit-ear-like rhopaliar horns (Figs 7i and 4i).

Cnidome. Tentacles were truncated in the preserved specimens, but nematocysts were sampled from both the distal end of the truncated tentacles as well as the proximal base of the tentacles near the pedalia. Distal tentacle tip: rod-shaped, microbasic p-mastigophores (L 58.2-*66.2*-71.2 µm, W 14.1-*18.2*-16 µm, n=6; Figs. 11a and b); large, oval, holotrichous isorhizas (L 59.1-*60.1*-64.8 µm, W 31.2-*38.6*-39.3 µm, n=20; Figs. 11c and d); oval, microbasic p-eurytele (L 18-*21.8*-23.6 µm, W 13.5-*15.5*-17.6 µm, n=5; Figs 8e and f); small, oval amastigophore (L 10.6

µm, W 6.4 µm, n=1; Fig. 11g). Proximal tentacle/tentacle base: rod-shaped, microbasic p-mastigophore (L 69.4-*71.7*-74.6 µm, W 14.9-*16.7*-19.1 µm, n=20; Fig. 11h); large, oval, holotrichous isorhiza (L 61.8-*66.9*-77.6 µm, W 30.1-*35.9*-43.3 µm, n=20; Figs 8i and j); oval, microbasic p-eurytele (L 43.3-*43.3*-44.5 µm, W 18.6-*18.6*-21.1 µm, n=3; Fig. 11k). Manubrium: ?oval, microbasic p-eurytele (L 20.7 µm, W 12.8 µm, n=1; Fig. 11l). Pedalial warts: large, oval, holotrichous isorhizas (L 60.9-*64.5*-69.5 µm, W 37.8-*40.6*-43.2 µm, n=20; Figs 8m and n); rod-shaped, microbasic p-mastigophore (L 61.7-*67.2*-71 µm, W 15.7-*18.9*-21.3 µm, n=13; Figs 8o and p). Exumbrellar warts: large, oval, holotrichous isorhizas (L 42.9-*45.9*-50.2 µm, W 27.6-*31*-33.6 µm, n=20; Fig. 11q); rod-shaped, microbasic p-mastigophore (L 62.8-*70*-79.4 µm, W 14.7-*17.2*-19.1 µm, n=20; Fig. 11r). Apex warts: oval, holotrichous isorhiza (L 33.6-*36.8*-40.4 µm, W 22.8-*25.5*-27.6 µm, n=20; Figs. 11s and t); rod-shaped, microbasic p-mastigophore (L 67.9 & 68 µm, W 16.8 & 19.9 µm, n=2; Fig. 11u); ?large, oval, holotrichous isorhizas (only empty capsules observed, L 48.3 & 53 µm, W 27.6 & 31.6 µm, n=2).

Remarks. A neotype is designated for *M. virulenta* to stabilise its taxonomic status and clarify its identity, with particular emphasis on differentiating it from its congener *M. fenneri*. The description and drawing of *M. virulenta* in the original description (as *Tamoya haplonema*; Kishinouye 1910) agree well with the material examined by us, but lack the characters typical of *Morbakka*, in particular the rhopaliar horns. We believe Kishinouye (1910) overlooked these characters, as did his contemporaries, because the potential importance of these characters had not been recognised. The original description does not mention the presence of vertical gastric phacellae that are characteristic of *Tamoya*, which suggests that the specimen(s) investigated by Kishinouye (1910) did not possess gastric phacellae – a trait characteristic of Carukiidae.

Additionally, the common name used by Kishinouye (1910) for *T. virulenta* is “Hikurage”, the same name applied to the specimens we studied. The original type locality was the Inland Sea off Kagoshima and Innoshima; Hiroshima Bay is about 50 km east of Innoshima.

Morbakka sp. from the Philippines

Fig. 12

Synonymies. Carybdea alata, var. grandis: Mayer (1915: p. 171 later became USNM 28713; only specimen from Manila Bay, Luzon); Mayer (1917: pp.189 as USNM 28713); Light (1921: pp 29-30). non *Carybdea alata, var. grandis* [refers to *Alatina grandis*]: Mayer (1915: p. 171; specimen from Mount Putri, Borneo); Mayer (1917: p. 171; specimen from Mount Putri, Borneo).

Material examined. USNM 28713, 2 individuals, ~100 mm and ~50 mm IRW, and ~80 mm BH and ~40 mm IRW; Manila Bay, The Philippines.

Description. Lot USNM 28713 contains two large carybdeid specimens (Fig. 12a and b) collected by the steamer Albatross in 1909 during the Pacific Expedition of the US Fish Commission. The specimens were originally identified as *Carybdea grandis* (now *Alatina grandis*) by A. G. Mayer. However, we determined that both specimens belong to the Carukiidae. In particular, both specimens appear to be species of *Morbakka*. The material is in poor condition and it is unclear if the specimens belong to one of the two described species of *Morbakka* (*M. fenneri* or *M. virulenta*), or if they represent an undescribed species. Both specimens display morphological characters that demonstrate their affinity to the genus *Morbakka*. In particular, both lack gastric phacellae, but possess rabbit-ear- like rhopaliar horns (Fig. 12c), thorn-shaped

extensions at the bases of their pedalial canals (Figs 9d and e), and numerous branching velarial canals with lateral diverticula (Figs 9f and g). Nematocyst warts on the perradial lappets, the exumbrella, or the pedalia were not observed, but this is likely a result of the poor preservation of the specimens.

Cnidome. Nematocysts were sampled from the tentacles, but all capsules sampled were discharged and lacked any interpretable features on the discharged shaft (Figs 9h and i). The following measurements were obtained from the unidentified nematocysts: L 51.6-*55.8*-59.3 μm, W 27.5-*29.8*-31.2 μm (n=9).

Discussion

Bentlage *et al.* (2010) recently revised the taxonomic framework of Cubozoa while focusing on the evolution of cubozoan toxins, behavior, and biogeographic patterns. The material presented here extends that work by highlighting the morphological basis of the current, phylogenetics-based taxonomy of Cubozoa. The synopsis and key to the carybdeid Cubozoa is a tool to enable others to distinguish among and identify families and genera of carybdeid cubozoans. Currently, the taxonomic framework of the second order (Chirodropida) within the Cubozoa is somewhat ambiguous, because at least one family (Chiropsalmidae) is likely to be paraphyletic (Bentlage *et al.* 2010). Several chirodropid genera are missing from the phylogenetic analyses of Cubozoa to date. A reevaluation of the taxonomy and the characters used to distinguish among the chirodropid families, genera, and species may be necessary as gaps in the sampling for molecular phylogenetic analyses are filled. Currently, Gershwin (2006c) presents the most up to date and comprehensive treatment of the Chirodropida, complementing the treatment of Carybdeida

presented here.

Table 1 Glossary of terminology of morphology and anatomy in alphabetical order.

adradial	Lying halfway between perradial and interradial plane.
area corrugata	Area of muscular tissue that displays alternating ridges and grooves.
diverticula	Sacs or pouches stemming from velarial canals.
exumbrella	Surface of the bell of a medusa.
frenulum	Perradial sheet of muscular tissue bracing the right-angle connection of velarium and subumbrella.
gastric phacellae	Part of the digestive system of box jellyfish; usually comprised of bundles of cirri in the corners of the stomach.
gonad	Interradial sheet of tissue containing eggs or sperm.
interradial	In box jellyfish the planes that are marked by the presence of pedalia and tentacles.
manubrium	Cruciform or quadrangular tube of varying length or thickness projecting downward from the stomach.
mesentery	Perradial sheet of tissue that attaches the stomach to the subumbrella.
nematocyst	Capsule containing an oftentimes barbed tubule that delivers venom to predator or prey. Commonly referred to as “stinging cell”.
nematocyst wart	Dense accumulation of numerous cells containing nematocysts; visible to the naked eye as freckles or warts on the body of the medusa.
pedalium	Interradial muscular structures inserting at the base of each corner of the bell. Pedalia bear the tentacles of box jellyfish. In Carybdeida each

	pedalium bears a single tentacle; in Chirodropida pedalia branch into numerous “fingers” distally, each bearing a single tentacle.
pedalial canal	Canal in the pedalium that connects the gastro-vascular system of the bell to the hollow tentacles.
perradial	In box jellyfish the planes that are marked by the presence of rhopalia.
perradial lappet	Muscular, triangular lappet of tissue in the perradius of the velarium.
rhopaliar horn	Blind ending canal that possesses an opening to the inside of the rhopaliar niche; function unknown.
rhopaliar niche	Cavity that is open to the environment at the exumbrellar side of the bell; contains rhopalium.
rhopaliar niche ostium	The exumbrellar opening of the rhopaliar niche; usually bearing covering scales.
rhopaliar window	Tissue covering the rhopaliar niche on the subumbrellar side.
rhopalium	Sensory structure bearing eyes and statocyst. In box jellyfish rhophalia possess a muscular stalk that allows for active movement of the rhopalium.
stomach	Sac-like enlargement of the gastro-vascular system in the upper portion of the subumbrellar cavity.
submubrella	Underside of the bell of a medusa.
subumbrellar cavity	Cavity formed by the subumbrella.
velarium	A circular flap of muscular tissue forming the opening of the subumbrellar cavity.
velarial canal	Canal in the velarium that is connected to the gastro-vascular system.

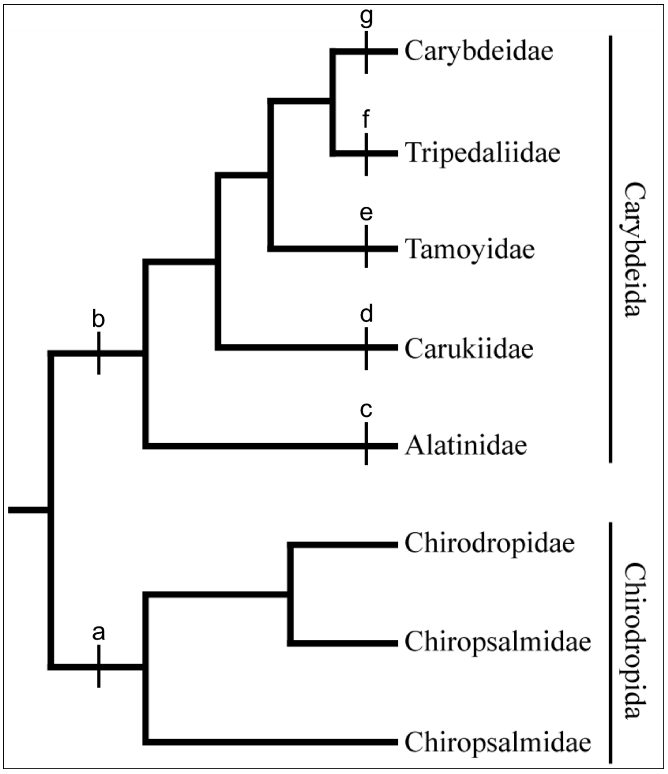


Fig. 4 Family level relationships within Cubozoa following Bentlage *et al.* (2010). Key characters to the orders and families are indicated on the branches of the cladogram. Detailed explanations of the characters can be found in the dichotomous key and associated figures. Chirodropida possess multiple tentacles perpedalium (a) whereas Carybdeida possess a single tentacle per pedalium (b). In the Tripedaliidae multiple pedalia each bearing a single tentacle may be present that can be mistaken for multiple tentacles arising from a single pedalium. The families of the Carybdeida can be identified using the following characters. Alatinidae (c): rhopaliar niche ostium t-shaped. Carukiidae (d): rhopaliar niche ostium with rhopaliar horns, stomach lacking gastric phacellae. Tamoyidae (e): rhopaliar niche ostium frown-shaped, lacking rhopaliar horns, gastric phacellae vertical. Tripedaliidae (f): rhopaliar niche ostium keyhole-shaped, or alternatively rhopaliar niche ostium frown-shaped and 2 or 3 pedalia per corner instead of 1. Carybdeidae (g): rhopaliar niche ostium heart-shaped.

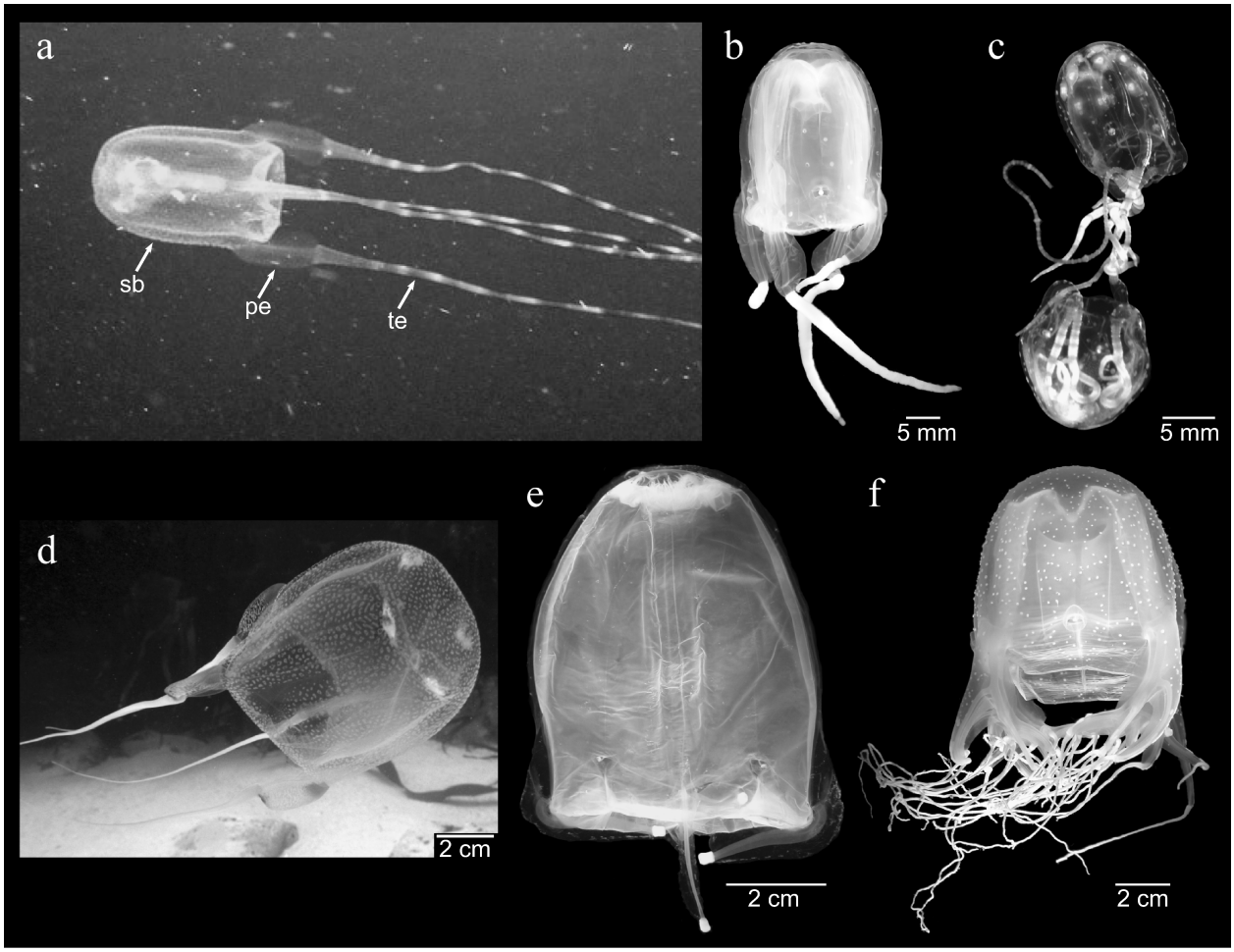


Fig. 5 Representative species of box jellyfish from each family of Carybdeida (a-e) and one from Chirodropida (f). Note that the chirodropid possesses multiple tentacles per pedalium (f) whereas the carybdeid species possess only a single tentacle per pedalium (a-e). a) *Tamoya* (Tamoyidae; image courtesy of Tim Lowry), b) *Malo kingi* (Carukiidae), c) male and female of *Copula sivickisi* coupling during courtship (Tripedaliidae; image courtesy of Alvaro Migotto), d) *Carybdea branchi* (Carybdeidae; image courtesy of Brent Viljoen), e) *Alatina* sp. (Alatinidae), f) *Chiropsalmus quadrumanus* (Chiropsalmidae). sb: swimming bell, pe: pedalium, te: tentacle

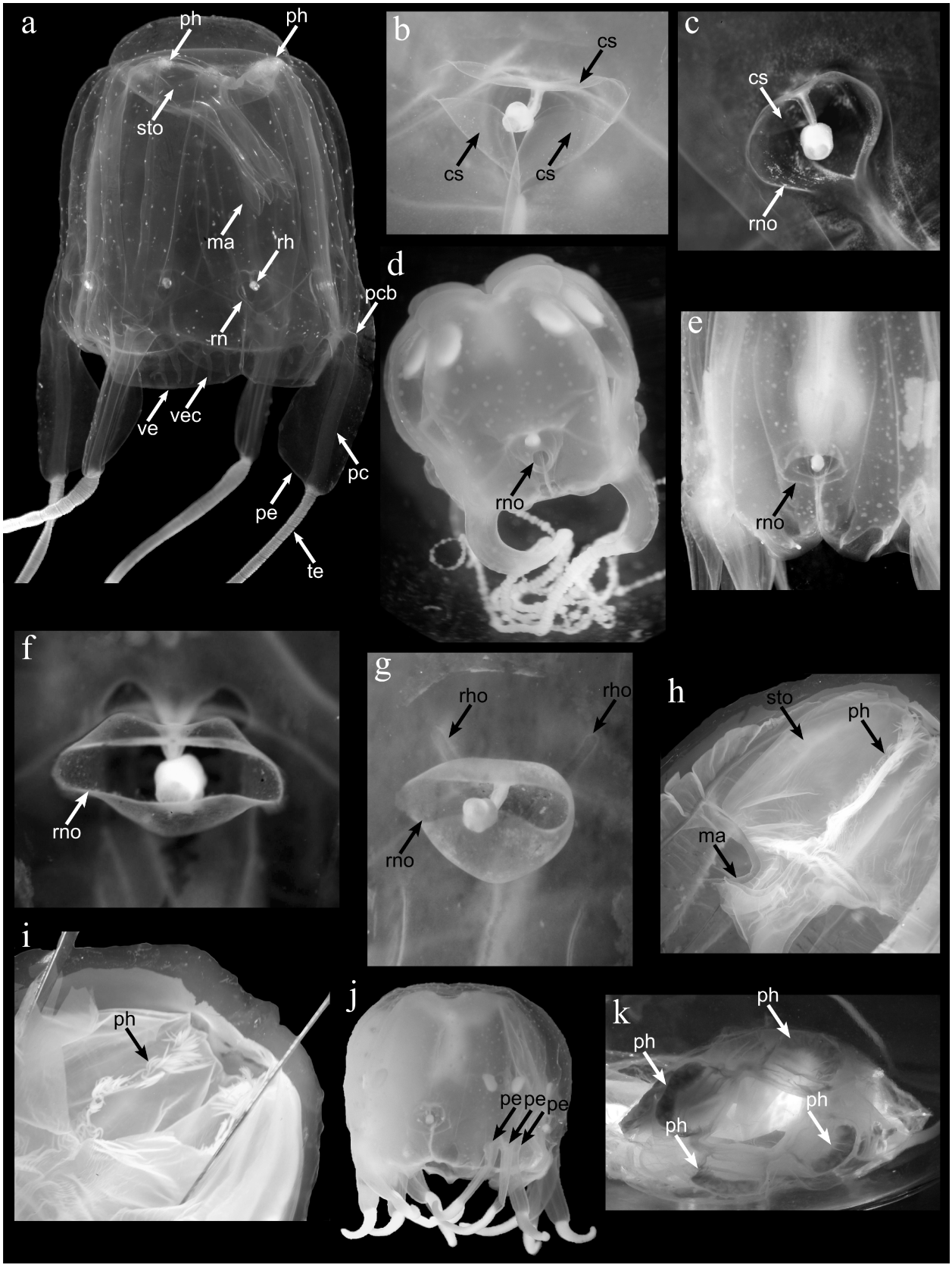

a
ph
ph
sto
ma
rh
pcb
rn
ve
vec
pc
pe
te
b
cs
cs
cs
c
cs
rno
d
rno
e
rno
f
rno
g
rho
rho
rno
h
sto
ph
ma
i
ph
j
pe pe pe
k
ph
ph
ph
ph
ph

Fig. 6 (previous page) Key characters for identification of the Carybdeida: a) location of taxonomic characters (*Carybdea brevipedalia*; image courtesy of Alvaro Migotto), b) t-shaped rhopaliar niche ostium; note the 3 scales covering the niche, thus leaving a t-shaped ostium, c) heart-shaped rhopaliar niche ostium; niche ostium with only 1 upper covering scale, d) key-hole shaped rhopaliar niche ostium in *Copula sivickisi*, e) frown-shaped rhopaliar niche ostium in *Tripedalia*, f) frown-shaped rhopaliar niche ostium in *Tamoya*, g) frown-shaped rhopaliar niche ostium with rhopaliar horns in *Malo*, h & i) vertical gastric phacellae in the stomach of *Tamoya*, j) multiple pedalia on each corner of the swimming bell in *Tripedalia*, k) crescentric gastric phacellae in *Alatina* (apical view). cs: covering scale, ma: manubrium, ph: gastric phacellae (comprises gastric cirri) in the stomach cavity, pc: pedalial canal, pcb: pedalial canal bend, pe: pedalium, rh: rhopalium, rn: rhopaliar niche that opens on the exumbrellar side of the bell with the rhopaliar niche ostium, rno: rhopaliar niche ostium, rho: rhopaliar horn, sto: stomach, te: tentacle, ve: velarium, vec: velarial canal

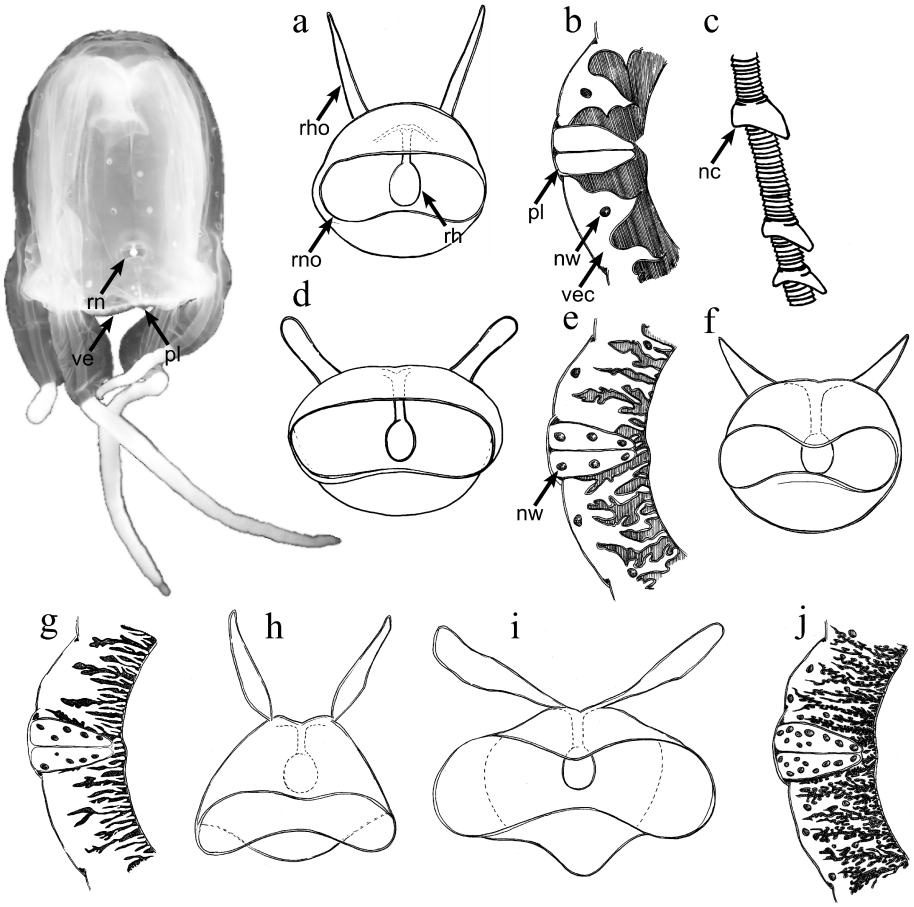


Fig. 7 Characters distinguishing the genera of Carukiidae. The location of the characters illustrated in a-j are indicated on the specimen in the upper left hand corner. The rhopaliar niche ostium is always horizontally oriented and more or less oval in the Carukiidae; however, its shape may vary among specimens. *Carukia*: a) rhopaliar niche with horns, b) velarium, c) tentacle with "neckerchiefs"; *Malo*: d) rhopaliar niche with horns, e) velarium; *Gerongia*: f) rhopaliar niche with horns, g) velarium; *Morbakka*: h and i) differences of rhopaliar niche and horns between *M. fenneri* (h) and *M. virulenta* (i), in both cases the rhopaliar horns are "rabbit-ear" shaped but the angle and breadth of the horns vary, j) velarium. nc: "neckerchief", nw: nematocyst wart, pl: perradial lappet, rn: rhopaliar niche, rno: rhopaliar niche ostium, rho: rhopaliar horn, ve: velarium, vec: velarial canal (line drawings by Cheryl Lewis)

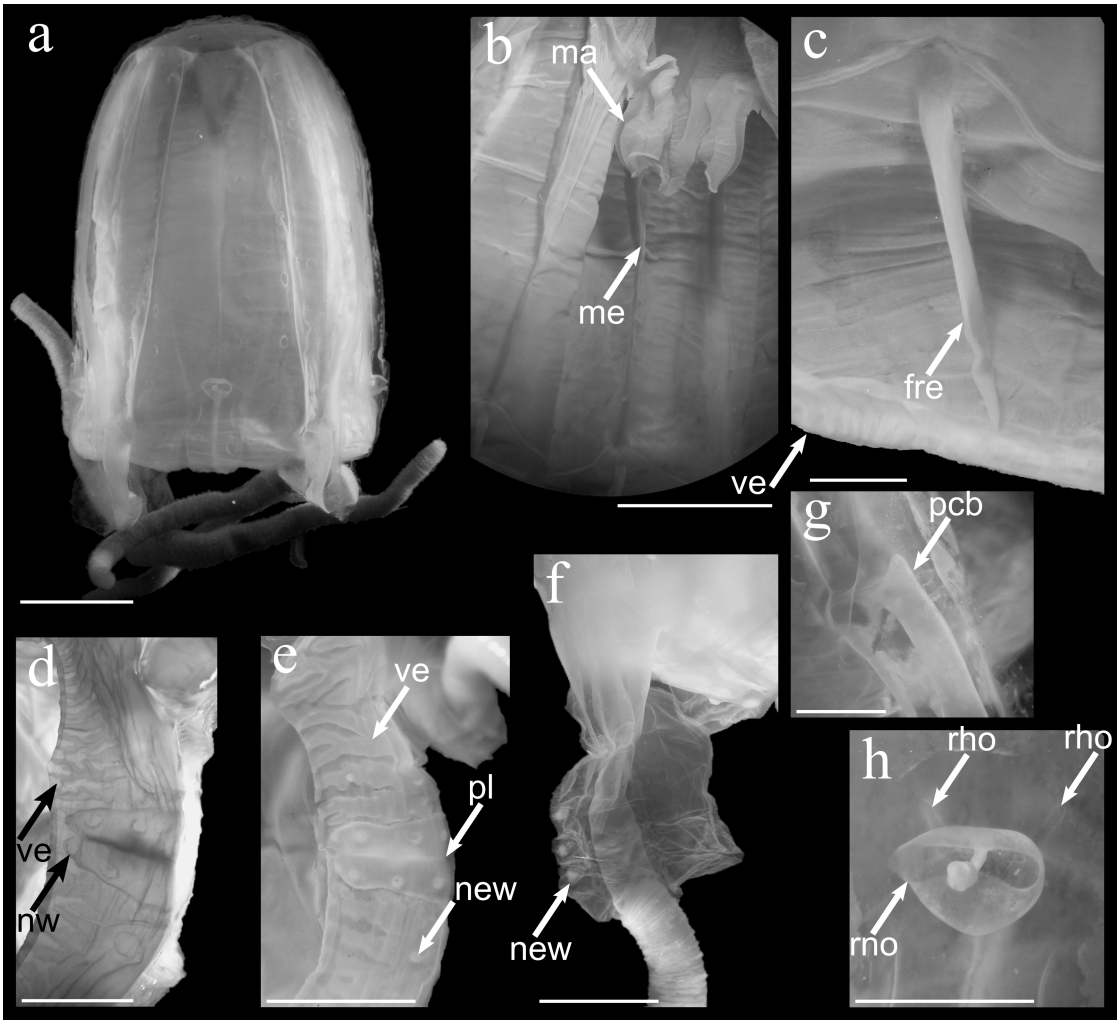


Fig. 8 *Malo filipina*: a) whole specimen, USNM 27936, b) view from subumbrellar opening into subumbrellar cavity; the manubrium was cut open showing the inside of the lower portion of the stomach (USNM 27936), c) frenulum 27936, d) velarium with perradial lappet (USNM 27936), e) velarium with perradial lappet (USNM 27935), f) pedalium (USNM 28714), g) pedalial canal with proximal pedalial canal bend (USNM 27936), h) rhopaliar niche (USNM 27936). fre: frenulum, ma: manubrium, me: mesenterium, nw: nematocyst wart, pcb: pedalial canal bend, pl: perradial lappet, rho: rhopaliar horn, rno: rhopaliar niche ostium, ve: velarium. Scale bars: 10 mm (a, b, d, e and f), 5 mm (c, g and h)

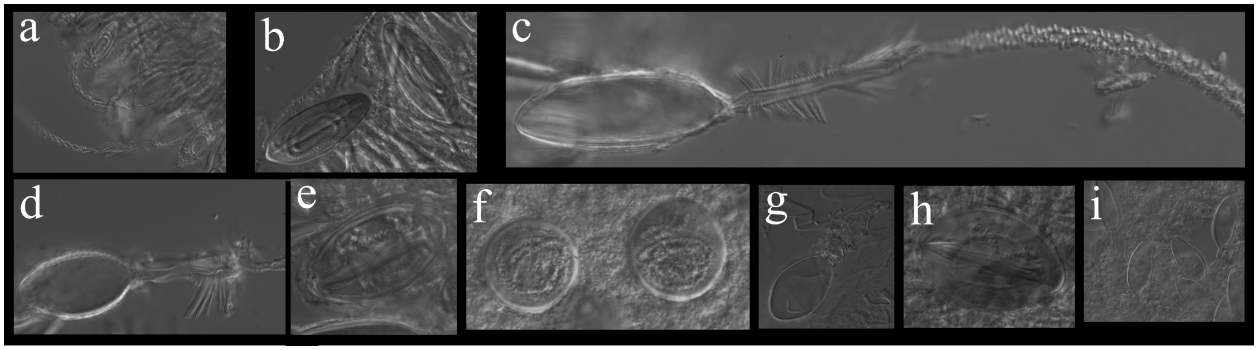


Fig. 9 Nematocysts of *Malo filipina*; for identifications and sizes see cnidome description.

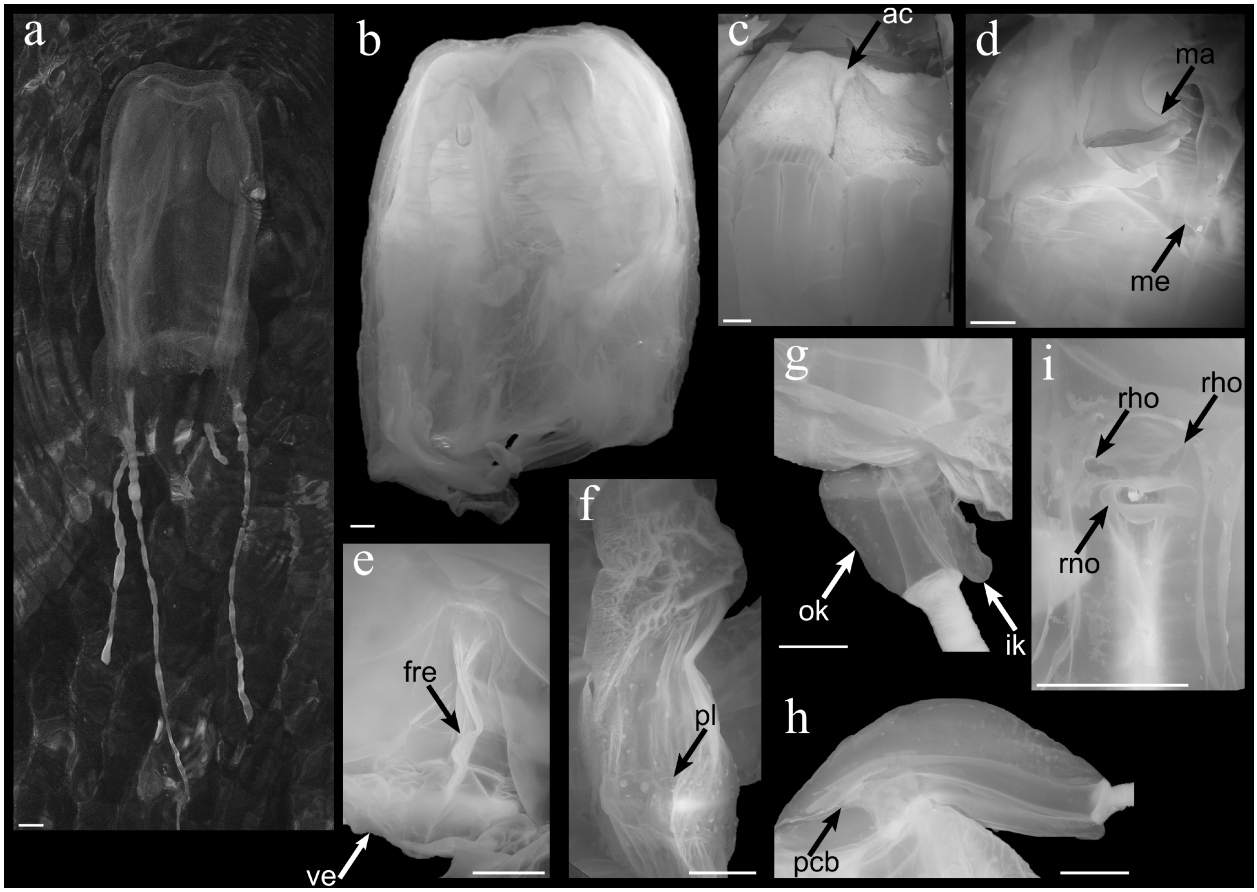


Fig. 10 *Morbakka virulenta*: a) specimen in its natural environment in Hiroshima Bay (image courtesy of Masato Kawahara), b) whole specimen of the neotype USNM 1124253, c) interior of stomach with area corrugata, d) subumrellar view with manubrium and mesentery, e) Rhopaliar niche ostium t-shaped mm (d, e and f), and 50 m (e and d).frenulum, f) velarium with perradial lappet, g) pedalium with nematocyst warts on outer keel and overhang on inner keel, h) pedalial canal with pedalial canal bend, i) rhopaliar niche. b–c and i) USNM 1124253; h) USNM 114252. ac: area corrugata, fre: frenulum, ik: inner keel, ma: manubrium, me: mesentery, ok: outer keel, pl: perradial lappet, rho: rhopaliar horn, rno: rhopaliar niche ostium, ve: velarium. Scale bars: 10 mm

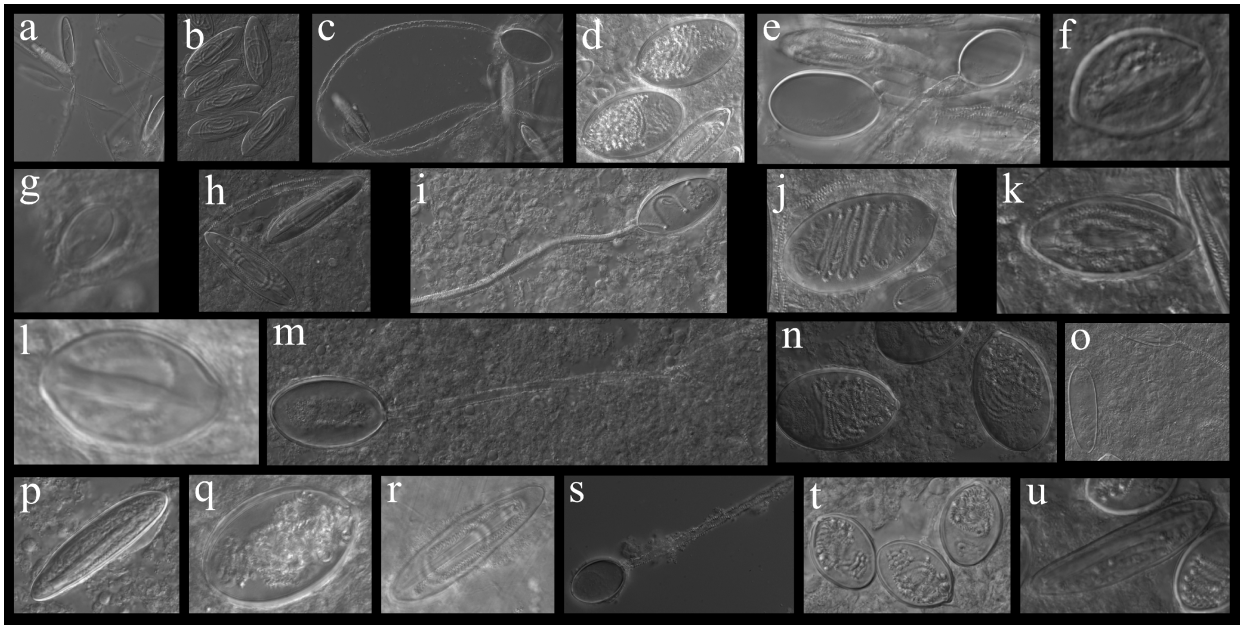


Fig. 11 Nematocysts of Morbakka virulenta; for identifications and sizes see cnidome description.

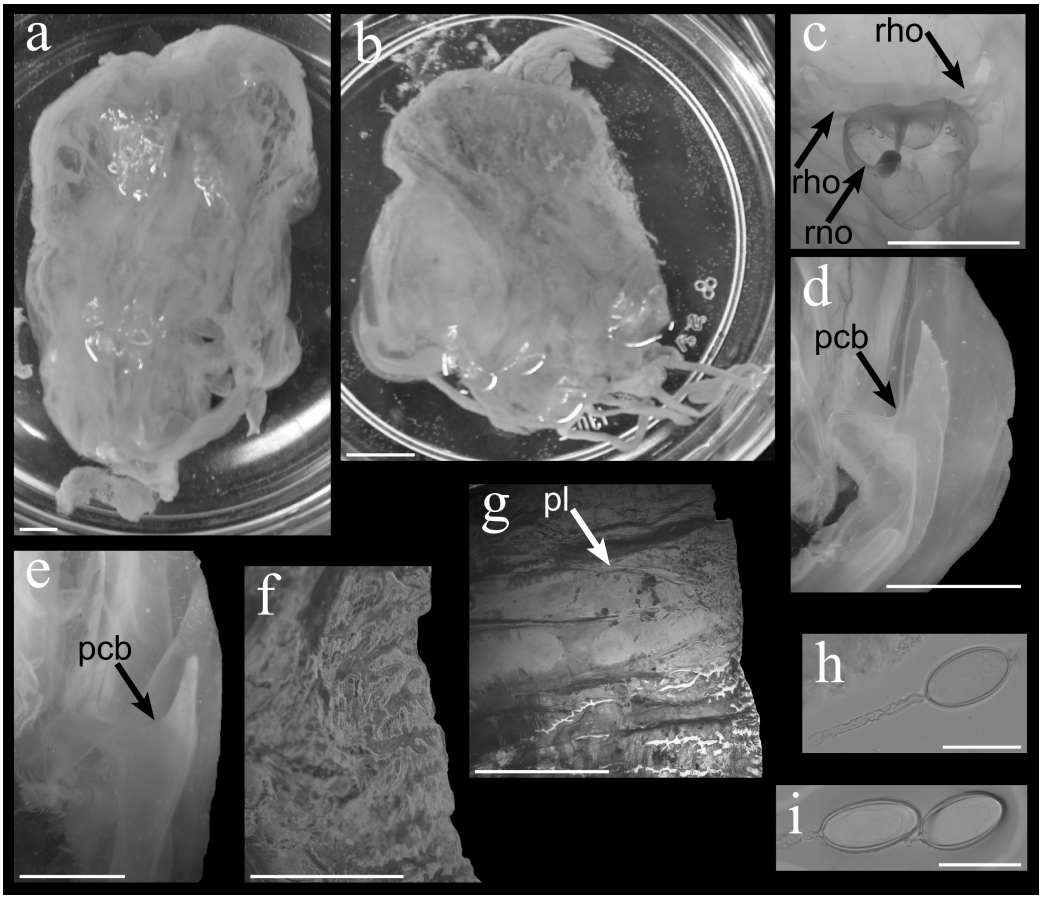


Fig. 12 *Morbakka* sp. from the Philippines (USNM 28713): a) whole specimen of the large individual, b) whole specimen of the small individual, c) rhopaliar niche ostium, d) pedalial canal of the large specimen, e) pedalial canal bend of the small specimen, f) velarial canals of the small specimen, g) velarial canals of the large specimen, h and I) unidentified nematocyst capsules from the tentacles. rho: rhopaliar horn, rno: rhopaliar niche ostium, pcb: pedalial canal bend, pl: perradial lappet. Scale bars: 10 mm (a, b and g), 5 mm (c, d, e and f), and 50 m (h and i)

CHAPTER 3:

Inferring distributions of chirodropid box-jellyfishes (Cnidaria: Cubozoa) in geographic and ecological space using ecological niche modeling

Abstract

Geographic distributions of many marine species are poorly documented or understood, which is particularly true for marine invertebrates. Ecological niche modeling (ENM) offers a means to address this issue, but to date most studies using ENM have focused on terrestrial taxa. In general, ENM relates environmental information to species' occurrence data to estimate the ecological niche of a species, rather than just interpolating a geographic distribution. This process leads to predictions of suitable habitat that generally exceed the range actually inhabited by a single species: such areas of geographic over-prediction (commission) may be inhabited by closely related species, and the model thus offers the inferential power to predict the potential distributions of these species as well. We explored the utility of ENM to investigate potential distributions of chirodropid box-jellyfishes (Cnidaria: Cubozoa), a group of highly toxic invertebrates whose biogeography is poorly understood. We were able to predict reported occurrences of box-jellyfishes throughout the Indo-Pacific from data of closely related species. By doing so, we demonstrate that geographic over-prediction in ENM can be desirable when concerned with predictions beyond current knowledge of species' distributions. Several methods are used for ENM; here, we compared the 2 most commonly used methods, the Genetic Algorithm for Rule-Set Predictions (GARP) and a maximum entropy approach (Maxent). Our comparison shows that Maxent may be more prone to overfitting, whereas GARP tends to

produce broader predictions. Transforming continuous Maxent predictions into binary predictions remedies problems of overfitting, and allows for effective extrapolation into unsampled geographic space.

Introduction

Understanding geographic ranges of species is crucial in studying biogeographic as well as large-scale ecological questions in the marine realm. A particular problem in studying marine species' distributions arises from the paucity of reliable occurrence records that frequently represent only a fraction of the species' actual ranges (Guinotte *et al.* 2006), providing an incomplete picture of distributions and factors affecting them. Ecological niche modeling (ENM) offers a means of addressing the problem of limited occurrence data on species' distributions by providing predictions about species potential ranges based on environmental parameters (Soberón & Peterson 2004).

In ENM, known occurrences of species are related to raster environmental data layers in an evolutionarycomputing environment to reconstruct species' distributions in environmental dimensions. ENMs can be projected across broader landscapes than the original sampling area to identify a potential geographic distribution for the species. ENMs can then be used for a variety of applications, including design of sampling strategies (e.g., Guisan *et al.* 2006, Pearson *et al.* 2007) and identification of areas vulnerable to species invasions (Peterson 2003). ENM is also being employed as a forecasting tool for biotic responses to changing climates (e.g., Martínez-Meyer & Peterson 2006) and can be incorporated into analyses of evolution of species' autecology (e.g., Martínez-Meyer *et al.* 2003).

In particular, theoretical studies predict a high degree of conservation of ecological niches among closely related species over evolutionary timescales (Holt & Gaines 1992), which has subsequently been demonstrated empirically (e.g., Peterson *et al.* 1999, MartínezMeyer & Peterson 2006, Kambhampati & Peterson 2007, Peterson & Nyári 2008, but see Rice *et al.* 2003

and Graham *et al.* 2004 for counterexamples). In essence, conservatism is expected since mortality rates are higher under conditions outside the niche (Holt & Gaines 1992). This property of ecological niches may allow prediction of species' distributions and ecology from information regarding their relatives. Specifically, in a phylogenetic context, ENMs can be used to test hypotheses about evolution of species' ecology. For example, investigating patterns of overlap of niche requirements of closely related species may allow untangling the roles of ecology and historical contingencies in shaping of species' distributions and the process of speciation (see Swenson 2008 for a review).

ENM studies to date have focused almost exclusively on terrestrial fauna and flora and, although ENM has been demonstrated to produce robust predictions of geographic distributions of marine species (Wiley *et al.* 2003, Kluza & McNyset 2005, Guinotte *et al.* 2006, Bryan & Metaxas 2007, Therriault & Herborg 2008), it has seen only limited application by marine biologists thus far. Our aim was to expand ENM applications to evaluate the utility of these methods for understanding potential distributions and large-scale ecological properties of 3 species of chirodropid cubozoans.

Cubozoa encompasses approximately 40 described species in 2 orders, Carybdeida and Chirodropida (see Daly *et al.* 2007). Like many other medusozoan cnidarians, cubozoans are characterized by alternation of benthic polyp and free-swimming medusa generations. Here, we focus on potential distributions of the medusa generation in chirodropid box-jellyfishes; polyps have been observed in the wild only once (Hartwick 1991b), so knowledge of this life stage is limited. Chirodropid medusae can frequently be encountered in coastal waters in the tropics and subtropics and may display species-specific patterns of seasonal occurrence and abundance.

Many cubozoan species contain potent toxins in their stinging capsules (called nematocysts), and their high toxicity regularly leads to severe stings that can cause cardiac dysfunction or cardiac arrest in humans (reviewed by Winkel *et al.* 2003). For this reason, cubozoans have received considerable attention from scientists and authorities in Australia, where they represent a public health threat and cause financial losses in the tourism industry (Bailey *et al.* 2003). In particular, the Australian chirodropid Chironex fleckeri is considered the world's most lethal jellyfish (Fenner & Williamson 1996). Given its practical importance, its distribution in Australian waters is better documented than those of other cubozoan species. Using data from museum records, we investigated the degree to which ENM approaches are able to predict the species' native geographic range.

Outside of Australia, knowledge of chirodropid distributions is mainly anecdotal, but assuming that niche conservatism holds for chirodropids, it should be possible to extrapolate niche predictions based on Australian *Chironex fleckeri* to predict the potential distribution of related species. We also modeled niche distributions of 2 Asian chirodropids (*Chiropsoides buitendijki* sensu lato and *Chironex yamaguchii*), using what limited occurrence information was available. By doing so, we modeled possible geographic ranges of chirodropids throughout Oceania and Asia. We used these models to investigate niche overlap in geographic and ecological space among the 3 chirodropid species. This study demonstrates that ENM can improve understanding of distributions and ecological properties of marine organisms, even when very few distributional data are available.

Materials and Methods

Study species: taxonomy and nomenclature. Of the 3 species studied, *Chironex fleckeri* Southcott, 1956 is the only one with a stable taxonomic history. *Chiropsoides buitendijki* (van der Horst, 1907) and *Chiropsoides quadrigatus* (Haeckel, 1880) have seen historical and current taxonomic debate (Mayer 1917, Thiel 1928, Stiasny 1937, Gershwin 2006, Lewis & Bentlage 2009). *C. quadrigatus* was originally described from Burmese waters (Haeckel 1880) and later redescribed based on material from the Philippines (Mayer 1910). The original description, however, was based on a single damaged, juvenile specimen that differs strikingly from Philippine and Japanese chirodropids (Thiel 1928, Stiasny 1937, Gershwin 2006, Lewis & Bentlage 2009). In particular, the few interpretable characters of Haeckel's C. quadrigatus closely resemble *C. buitendijki*, which was originally described from Java (van der Horst 1907), suggesting that they may be synonymous (Gershwin 2006). Hence, we included Haeckel's *C. quadrigatus* specimen from Burma in our *C. buitendijki* modeling experiments. The taxonomic confusion surrounding *C. quadrigatus* misled others to identify the chirodropid species that can be found in the Philippines and Japan as *C. quadrigatus*. A recent study (Lewis & Bentlage 2009) demonstrated this case of mistaken identity, and described the Philippine/ Japanese chirodropid as *Chironex yamaguchii* Lewis & Bentlage, 2009.

Input data. Occurrence data were accumulated from catalogued museum specimens and the scientific literature (Fig. 13; Table 2). Spatial accuracy of data points varied from source to source; some specimens were georeferenced with reasonably precise latitude and longitude, whereas other locality information consisted of textual descriptions of collection sites only. The latter were georeferenced for the present study, and hence may be subject to greater uncertainty

(Table 2).

Environmental data included raster geographic information system grid layers summarizing yearly averages of remotely sensed environmental parameters and ocean colors from the Aqua-MODIS satellite (Table 3; Savtchenko *et al.* 2004). Aqua-MODIS ocean color satellite imagery for 2003 to 2006 was averaged across the binned yearly composite images. We also included seafloor topographic grids derived from a combination of satellite measurements of marine gravity field anomalies and ship depth soundings (Smith & Sandwell 1997). All environmental grids were resampled to a resolution of 2.52' for analysis using nearestneighbor sampling because ENM algorithms require all grids to be at the same resolution and extent. All grids except bathymetry were already on the same resolution and scale, but the native resolution of the bathymetry grid was 2'.

Ecological niche modeling. Several approaches have been used and explored for ENM (reviewed by Elith *et al.* 2006). From these algorithms, we employed 2 commonly used evolutionary-computing algorithms, the Genetic Algorithm for Rule-Set Prediction (GARP; Stockwell & Peters 1999) and a maximum entropy approach (Maxent; Phillips *et al.* 2004, 2006). Both GARP and Maxent are designed specifically to manage presence-only data, which offers advantages over other algorithms that require both presence and absence data for model development. Absence information could be used to allow the modeling algorithm to evaluate commission error (i.e., the area predicted present but not actually inhabited by the species). However, absence data are particularly problematic because non-observation of a species at a locality may result from lack of sampling, lack of opportunity for dispersal to sites with suitable habitat, vicariant speciation, or real absence of suitable conditions (Anderson *et al.* 2003),

particularly in poorly-known organisms like chirodropids. Below, we outline the implementation of each algorithm, but Stockwell & Peters (1999) and Phillips *et al.* (2004, 2006) provide more indepth explanations of the respective software packages.

GARP is a heuristic algorithm that iteratively 'evolves' rule-sets (if-then statements regarding environmental conditions) that are combined into final predictions. First, input occurrence data are divided into training data for model development and extrinsic testing data for model selection. GARP then works in an iterative process, in which it first selects a method from a set of possibilities (logistic regression, bioclimatic rules, range rules, negated range rules), and applies this method to the training data. The resulting rule is evaluated for its significance, based on a set of 1250 presence points generated by resampling from a subset of the input training points with replacement and an equal number of pseudoabsence points (i.e., points generated at random from grid pixels at which the species has not been recorded). In subsequent iterations, rules are changed (evolved) randomly, based on a set of change operators designed to mimic chromosomal evolution (e.g., crossingover, deletion), thus introducing elements of stochasticity in the rulechanging process. Significance of changes in the predictive values (presence or absence of the species) before and after the application of the new rule is evaluated using a significance parameter outlined by Stockwell & Peters (1999). As GARP is a stochastic approach, alternate runs using the same input data lead to different results. Hence, following best-practice approaches (Anderson *et al.* 2003), we generated 200 models for each species (previous tests showed that 1000 replicates do not improve the final model compared to 100 replicates). From these 200 models we retained 20 models (in our cross-validation experiments; see below) or 40 models (for our large-scale projections) that showed the lowest omission error,

and then discarded the 50% that predicted the most extreme values of areas present ('commission error index' of Anderson *et al.* 2003). This 'best subset' of replicate models was then summed, pixel by pixel, to create a final prediction.

Maxent addresses the issue of presence-only modeling differently compared to GARP, as it does not rely as explicitly on pseudoabsence data. The basic idea behind Maxent is to model a probability distribution for species occurrence that is most spread out (i.e., exhibits the highest degree of entropy), given certain constraints. In terms of ENM, the constraints are represented by the values of environmental parameters observed at known occurrence points (Phillips *et al.* 2004, 2006). Maximum entropy is invoked in Maxent because it gives the least biased estimate possible on the given information when characterizing unknown events with a statistical model (Jaynes 1957). In addition to using raw environmental data values (features) as input, Maxent also uses transformations (i.e., feature vectors) of the variables when the input features are continuous variables, which may be the variables squared, multiplications of 2 or more variables, categorizations, and other types of transformations. Feature vectors are included in the modeling process to improve the model-to-data fit. The auto feature option that we used lets Maxent find the feature-vector combination most suitable to the particular input data set based on sample sizes available (for further information see Phillips *et al.* 2006; Phillips & Dudík 2008).

Rather than assigning a probability from 0 to 1 to each pixel, the Maxent probability distribution sums to 1 over the entire grid to produce a 'raw' output. Such individual grid cells are assigned very small probability values, making a prediction more difficult to interpret on a map. This problem is addressed by offering 2 additional output formats: cumulative output in which the value assigned to each pixel is the sum of probabilities assigned to all pixels that have

equal or smaller raw probability values, and logistic output, which is a logistic transformation of the raw values, providing a probability of species' presence (or habitat suitability) for every grid cell ranging from 0 to 1 (Phillips & Dudík 2008). We use the logistic output format throughout this contribution, but note that all 3 output formats are monotonically related to each other. In contrast to GARP, Maxent is deterministic, so every run with the same dataset returns the same answer.

We performed all experiments using Desktop GARP (ver. 1.1.6; www.nhm.ku.edu/desktopgarp) and Maxent (ver. 3.0; www.cs.princeton.edu/~schapire/maxent). For model building, we considered only spatially unique occurrence points, and models were trained on the same geographic area over which they were projected. Models were built with a random selection of 15% of occurrence points set aside for internal model evaluation by the algorithms. Maxent models were imported into ArcView 3.2 (ESRI) as floating-point grids and then multiplied by 1000 and converted into integer grids, since the former are difficult to manipulate.

***K*-fold cross-validation.** To investigate the ability of our ENM techniques to predict known occurrences, we divided the Australian study area (i.e., *Chironex fleckeri*) into 6 sectors containing roughly similar numbers of spatially unique occurrence points. We restricted this cross-validation approach to Australia since the other 2 areas (= species) held only very limited occurrence points. Models were built (1) using occurrence points falling in 5 of the 6 sectors to predict the distribution of points in the sixth, and (2) using points falling in half of the sectors to predict the other half. The significance test we employed to evaluate model performance follows Anderson *et al.* (2002, 2003): given the proportion of pixels predicted present versus absent in each evaluation sector, we tested whether occurrence points fell into pixels predicted present

more often than would be expected at random. Cumulative, 1-tailed binomial probabilities of observed proportions of test points falling in pixels of predicted presence and absence at different threshold levels of model output probabilities were calculated to test for significance. In Maxent, because the area of predicted occurrence becomes extremely small in the upper part of the output probability distribution, we reclassified the continuous probability distribution of the output model into a discrete one using several thresholds (i.e., > 0, > 0.01, > 0.1, > 0.3, > 0.5, and > 0.8) for each of which we calculated binomial probabilities. For GARP, binomial probabilities were calculated for each congruence level of the 10 best models.

Large-scale predictions and visualizations in ecological space. To generate broad-scale predictions of predicted distributions for each species, we set a threshold for prediction of presence as the lowest prediction score, which represents the model output values at points of known occurrence (Pearson *et al.* 2007). In a few cases, training points fell into grid cells that had been assigned 0 probability of occurrence by the model; as these pixels were adjacent to pixels predicted at a higher threshold, we assigned these occurrences the threshold that was immediately adjacent. To explore more restrictive thresholds, we also calculated the average probability value assigned by the model to grid cells containing occurrence points and set this value as a more stringent threshold for prediction of presence.

Finally, we explored the distribution of each modeled niche in environmental dimensions. We used the 'combine grids' option in the Grid Transformation Tools extension of ArcView to create a composite of all environmental parameter grids with the final ecological niche model grids. We exported the attributes table associated with this grid and then imported a random subset of 100 000 pixel values for visualization in SPlus (Insightful Corporation). To reduce the

dimensionality of this space, we employed principal components analysis based on the correlation matrix. Niche models were transformed into binary predictions using the same thresholds as described above, and areas of modeled niche space were then related to overall availability of environmental conditions in 2-dimensional scatterplots. We also related the environmental conditions within a 250 nautical mile (nmi) (= 463 km) buffer zone around each occurrence point to overall environmental conditions. By this means, we could evaluate which environmental conditions are likely to be encountered by each species in the immediate vicinity of its known populations.

Results

Effects of missing data and sampling biases. Models generated performed well in validation exercises in which distributional data from 5 regions were used to train a model that predicts distributions in a sixth region: in all cases, occurrence points set aside during model building were predicted significantly better than random ($p < 0.01$; Fig. 14a). In Maxent, under these conditions, the high-probability values clustered around occurrence points used in model building (Fig. 14b). Maxent models nonetheless performed well in extrapolating into 'unsampled' sectors at lower thresholds (i.e., around 0.8; all $p < 0.01$).

More extreme validation experiments, in which half of the regions were set aside for model testing (a total of 15 experiments in each GARP and Maxent), revealed similar differences between GARP and Maxent (Fig. 14c–f). GARP performed well in most cases, predicting points falling into sectors left out from model building with high precision ($p < 0.01$). However, in 2 cases presenting the most extreme spatial biases (i.e., training points all from 1 extreme of the

geographic distribution), GARP failed to predict points in the 'unsampled' sectors (both $p > 0.05$; e.g., Fig. 14e). Maxent models behaved similarly and failed in the same 2 scenarios (both $p > 0.05$; e.g., Fig. 14f).

Large-scale predictions and niche overlap among species. Although GARP and Maxent models differed in several aspects, they displayed largely congruent patterns (Fig. 15). In general, GARP predicted broader areas for Chiropsoides buitendijki and Chironex yamaguchii than the corresponding Maxent models, but more stringent GARP thresholds and less stringent Maxent thresholds yielded geographic predictions that were closely similar. The GARP model differs from the Maxent model in this comparison in predicting areas offshore of southwestern and southern Australia, as well as near New Zealand. Niche models for *C. fleckeri* and *C. buitendijki* were largely congruent using both algorithms (Fig. 15). The models for *C. yamaguchii*, however, showed only limited overlap with the other 2 species, regardless of modeling method: in fact, the models for this species versus the other 2 seemed to be almost the 'negative' of one another.

Considering only areas of agreement between the 2 modeling techniques, predictions for each species were as follows. Niche models for *Chironex fleckeri* predicted suitable habitat along the northern Australian coast from approximately Exmouth, Western Australia, to central and southern Queensland, including southern Papua New Guinea. In addition, large coastal areas in Southeast Asia are predicted, ranging roughly from Vietnam to Pakistan. The prediction for *Chiropsoides buitendijki* is very similar as regards the geographic extent of the prediction, but with more area predicted in Asia and less in Australia. In both cases, the Philippines are predicted as unsuitable, except for a few pixels. By contrast, niche models for *C. yamaguchii* are distinct

from those of the other species, and Philippine distributional areas, and areas of Sulawesi and the lesser Sunda Islands are predicted as suitable. As noted above, this model is nearly the inverse of the other 2, predicting coasts opposite to the ones predicted by the other models (e.g., the northern coast of Papua New Guinea).

In ecological space, 79% of the variation in the 10-dimensional environmental dataset are explained by its first 2 principal components (Table 3). Projecting the niche models for each species from geographic space into environmental dimensions makes the overlap and differentiation among the 3 niche models especially apparent (Fig. 16). The pattern is similar to that observed in geographic space: overlap among *Chironex fleckeri* and *Chiropsoides buitendijki* niche models is extensive, but the *C. yamaguchii* model occupies a region of environmental space adjacent to, but distinct from, the other 2 species. The only conditions under which C. yamaguchii overlaps extensively with the other 2 species is when the lenient GARP prediction is used, which appears to show excessive commission error in both geographic and environmental space. This differentiation is mirrored in the environments that are 'accessible' (i.e., with a 250 nmi buffer zone around each occurrence point) for each species. Whereas the buffered regions for the former 2 appear quite similar, the environment surrounding *C. yamaguchii* in the Philippines and Japan overlaps only on one extreme with those of *C. fleckeri* and *C. buitendijki*.

Discussion

ENM offers a promising avenue by which to investigate the biogeography of marine invertebrates for which distributional data are often depauperate. In particular, the ready

availability of ENM algorithms facilitates exploration of the technique for the GIS user. However, some degree of care is necessary when interpreting ENM results. For example, when challenged to predict species' occurrences across broad, unsampled regions, differences between GARP and Maxent emerge. In general, both algorithms predict testing data considerably better than random expectations. The major difference between GARP and Maxent predictions parallel those documented previously for terrestrial taxa (Peterson *et al.* 2007): GARP models are more general than those generated by Maxent, thus performing better in extrapolation challenges like the partitioning experiment we conducted, but are more prone to overprediction. In particular, the highest probabilities of predicted occurrence in Maxent closely associate with the points on which the model was trained, which leads to predictive failures of the higher probability levels of many Maxent models. In essence, the number of pixels that are assigned higher suitability scores is so small that numbers of occurrence data points omitted from predictions become extremely large. As such, thresholding becomes a critical step in considering predictions derived from Maxent models: above a certain probability level, Maxent models tend to reflect overfitting (i.e., they appear to be more accurate in fitting known data than predicting new data, thus hampering extrapolation; Peterson *et al.* 2007). Recent changes to Maxent output (i.e., the 'logistic' output), which we employed here, represent a rescaling of the probability distribution to deemphasize this distributional skew (Phillips & Dudík 2008) but do not seem to completely remedy the problems pointed out by Peterson *et al.* (2007).

Our results support the idea that the broader 'ramp' of probability values in Maxent models is less meaningful, and should be interpreted with caution (Peterson *et al.* 2007). Since all Maxent outputs are related to one another monotonically (Phillips & Dudík 2008) this is true

for both output options. A post hoc transformation of continuous Maxent results into binary values by thresholding, however, seems to produce reasonable predictions that allow effective extrapolation. Interestingly, although application of different strategies for thresholding niche models produces results that are much more congruent between GARP and Maxent, GARP models generally cover broader areas when applying the same thresholding criteria to both GARP and Maxent models. Prediction beyond present knowledge of species' distributions, in general, is desirable in ENM, but the question is how much of such overprediction or commission 'error' is biologically informative? Here, ENM 'overprediction' permitted us to predict distributional patterns of species across broad, unsampled areas with considerable accuracy, and to compare patterns of ecological niche occupancy. We were able to predict among species, as well as predict chirodropid occurrences throughout the Indo-Pacific. In this sense, as has been pointed out previously (Soberón & Peterson 2005), commission 'error' is quite desirable in ENM exercises. It is important to note, however, that no algorithm is capable of effective extrapolation if the input data for modeling are biased. In our most extreme spatial stratification experiments (Fig. 14e–f), extrapolation exercises failed most likely because the occurrence data for model training were sampled from one extreme of an ecological gradient (see Stockwell & Peterson 2002, Peterson 2005).

Ecological niche comparisons. One may argue that, instead of modeling ecological niches, it may be easier and more desirable simply to interpolate among known occurrence points (Bahn & McGill 2007). For well-sampled species, this approach might be a viable means of estimating realized distributions. ENM, as noted earlier, goes beyond reconstructing species' distributions to assemble a predictive model of their ecological requirements, and as such offers

considerably more inferential power.

Among the 3 species analyzed, the geographic distribution of the Australian *Chironex fleckeri* is best known (i.e., from Exmouth, Western Australia, to Gladstone, Queensland). However, current distributional maps (e.g., Fenner & Williamson 1996) do not account for the possibility that the width of the species' distribution along the coastline is unlikely to be uniform. ENMs offer predictions that create a more detailed and explicit picture. For example, 2 occurrence points off the coast of Queensland were not included in any prediction from either algorithms (not discernible in our figures); these specimens may indeed have been sampled that far offshore and identified correctly, but the models suggest that those sites are atypical of the species' occurrence. In particular, these records may represent samples from sink populations, which would be consistent with the idea that *C. fleckeri* does not venture far off the coast into deep waters (e.g., Hooper 2000). In other areas, however, the models suggest that the species may occur farther offshore than the actual occurrence points would indicate (e.g., at the Northern Territory and Queensland boundary), which is a hypothesis that can be tested via additional sampling efforts.

Further, we obtained a single record from Moreton Bay, Queensland, for *Chironex fleckeri*, some 400 km south of Gladstone. In almost all modeling experiments, Moreton Bay was predicted suitable for the species. The Moreton Bay area predicted by the ENM is disjunct from the rest of the C. fleckeri potential distribution in Maxent models, and only loosely connected in GARP models (Fig. 15). Hence, it may well be that the bay represents a place to which individuals are dispersed by the predominantly southward-flowing coastal currents, and then may be viable. Clearly, these predictions await additional sampling effort to be iteratively confirmed,

rejected, refined, and reinterpreted. Nonetheless, ENMs present a means of offering up hypotheses of geographic distributions of species within their native ranges, which can be tested through further sampling.

That ENM captures the ecological niche, rather than simply the geographic distribution, leads to predictions of suitable habitat for a given species in geographic space that generally exceeds the range it actually inhabits. In some cases, given prevalent niche conservatism (Peterson *et al.* 1999), sampling areas of 'overprediction' will lead to discovery of related species (Raxworthy *et al.* 2003), as shown by broad niche overlap between *Chironex fleckeri* and *Chiropsoides buitendijki* throughout Oceania and Asia. Indeed, in several places predicted by these models, chirodropid occurrences have been noted (Fig. 13). Only in a few of these cases has the animal been captured (Dawydoff 1936, Stiasny 1937, Ranson 1945, Nair 1951); at most of these sites, however, characteristic sting marks and clinical symptoms provide the only evidence of chirodropid occurrences (Fenner & Williamson 1996, Suntrarachun *et al.* 2001). Similarly, the *C. yamaguchii* ENM predicts several areas from which chirodropids have been reported, but for which no specific determination has been possible (Fenner & Williamson 1996).

Niche differentiation. Somewhat surprisingly, the *Chironex yamaguchii* niche model was distinct from the models for *C. fleckeri* and *Chiropsoides buitendijki*. In particular, ecological theory would lead one to expect that the congeners *C. fleckeri* and *C. yamaguchii* should share more ecological characteristics than either does with *C. buitendijki*. It could be argued that the taxonomic framework is wrong, but this seems unlikely (Lewis & Bentlage 2009; B Bentlage, P Cartwright & AG Collins unpubl.). Niche conservatism is an expectation, but its violation does not imply that model fit is erroneous. Rather, organisms evolve new

characteristics, and instances in which the expectation of discovering niche conservation is not met represent interesting avenues for further inquiry.

Niche differentiation (sensu Kambhampati & Peterson 2007) can be real (i.e., 2 species inhabit different niches, despite having access to the same set of environmental conditions) or apparent (i.e., 2 species inhabit different niches because the environments accessible to each are different). Given these considerations, it may be argued that the differentiation we observe is apparent rather than real. The environment accessible to *Chironex yamaguchii* is simply very different from the environments surrounding the other 2 species (Fig. 16; 250 nmi buffer). Similar conditions do exist in all 3 environments, but the area of overlap is limited to a few pixels (Fig. 16).

Niche differentiation may thus have taken place by ancestral species establishing themselves in the rare, shared environments (which may have been more extensive at the time) with dispersal into more common environments occurring subsequently. Alternatively, all environments were more equal in the past, and changing climates affected different localities in different ways, leading to niche differentiation. A comprehensive study integrating ENMs with phylogenetic analyses and paleooceanography will allow testing such scenarios.

The exact ecological nature of this differentiation is somewhat complex since a full interpretation of the loadings of each principal component does not seem feasible at this point. Ocean color, such as waterleaving radiances at different wavelengths, which strongly contributes to principal components 1 and 2, has been used previously to distinguish among marine habitats (e.g., Mishra *et al.* 2005). We suggest that the ocean color can be interpreted as a proxy for habitat structure. However, large-scale ground-truthing will be required to fully interpret which

habitats the models predict.

Concluding remarks. We have outlined the types of questions that can be addressed with ENM concerning the distribution, large scale ecology, and evolution of marine taxa. ENMs offer the opportunity to address questions on broad spatial scales that may not be easily addressed using other approaches. Undoubtedly, the predictions derived through modeling approaches will benefit much from evaluation using experimental approaches to test the models' ecological components as well as sampling programs to evaluate their geographic predictions.

Table 2 Sources of occurrence data used to model species' ecological niches. Italicized latitude and longitude values indicate records that were georeferenced by us based on a textual description of the sampling locality. Where applicable, museum catalog numbers or references to the scientific literature are provided. NCL: National Chemical Laboratory India; USNM: US National Museum of Natural History; MAGNT: Museum and Art Galleries of the Northern Territory; QM: Queensland Museum; SAM: South Australian Museum. AUS: Australia; PH: Philippines; JP: Japan; ID: Indonesia; LK: Sri Lanka; IN: India; BUR: Burma; PK: Pakistan

Species	Source	Latitude	Longitude	Place of Origin
Chironex fleckeri	SAM	-27.29	153.26	Moreton Bay, Qld, AUS
Chironex fleckeri	SAM	-23.14	150.82	Yeppoon, Qld, AUS
Chironex fleckeri	QM G317061	-21.20	149.27	Mackay, Qld, AUS
Chironex fleckeri	SAM	-21.12	149.26	Mackay, Qld, AUS
Chironex fleckeri	QM G317015	-20.90	149.05	Mackay, Qld, AUS
Chironex fleckeri	QM G3569	-20.58	148.75	Repulse Bay, Qld, AUS
Chironex fleckeri	SAM	*-20.50*	*148.81*	Prosperpine, Qld, AUS
Chironex fleckeri	SAM	-20.07	148.35	Bowen, Qld, AUS
Chironex fleckeri	SAM	*-19.14*	*146.78*	Townsville, Qld, AUS
Chironex fleckeri	SAM	-19.02	146.46	Townsville, Qld, AUS
Chironex fleckeri	SAM	-18.73	146.84	Ross River, Qld, AUS
Chironex fleckeri	SAM	*-18.23*	*146.09*	Cardwell, Qld, AUS
Chironex fleckeri	SAM	*-18.07*	*146.09*	Tully Head, Qld, AUS
Chironex fleckeri	SAM	*-17.95*	*122.21*	Broome, WA, AUS
Chironex fleckeri	SAM	*-17.48*	*146.15*	Thompson's Creek, WA, AUS
Chironex fleckeri	QM G322298	-17.15	139.60	Sweers Island, Qld, AUS
Chironex fleckeri	QM G4140	-16.86	145.81	Cairns, Qld, AUS
Chironex fleckeri	SAM	*-16.83*	*145.77*	Cairns, Qld, AUS
Chironex fleckeri	SAM	-16.80	145.76	Machan's Beach, Qld, AUS
Chironex fleckeri	SAM	-16.68	145.64	Gladstone, Qld, AUS
Chironex fleckeri	QM G317043	-16.48	145.47	Port Douglas, Qld, AUS
Chironex fleckeri	NMM F81653	-16.22	145.90	Kurrimine Beach, Qld, AUS
Chironex fleckeri	SAM	-15.46	145.35	Townsville, Qld, AUS
Chironex fleckeri	MAGNT C011739	-15.38	136.25	McArthur River, NT, AUS
Chironex fleckeri	SAM	-14.22	143.97	Bathurst Head, Qld, AUS
Chironex fleckeri	SAM	*-13.60*	*141.51*	Aurukun, Qld, AUS
Chironex fleckeri	SAM	*-12.64*	*141.84*	Weipa, Qld, AUS
Chironex fleckeri	SAM	*-12.49*	*130.84*	Darwin, NT, AUS
Chironex fleckeri	SAM	-12.48	130.84	Darwin, NT, AUS
Chironex fleckeri	MAGNT C012108	-12.48	130.83	Darwin, NT, AUS
Chironex fleckeri	MAGNT C011617	-12.48	130.84	Darwin, NT, AUS
Chironex fleckeri	MAGNT C005479	-12.48	130.84	Darwin, NT, AUS

Chironex fleckeri	SAM	-12.47	130.87	Darwin, NT, AUS
Chironex fleckeri	MAGNT C014843	-12.47	130.85	Darwin, NT, AUS
Chironex fleckeri	MAGNT C013719	-12.42	130.85	Darwin, NT, AUS
Chironex fleckeri	MAGNT C011147	-12.42	130.83	Darwin, NT, AUS
Chironex fleckeri	MAGNT C003862	-12.40	130.80	Darwin, NT, AUS
Chironex fleckeri	MAGNT C004706	-12.40	130.81	Darwin, NT, AUS
Chironex fleckeri	MAGNT C011641	*-12.35*	*130.83*	Berry Springs, NT, AUS
Chironex fleckeri	MAGNT C011185	-12.35	130.88	Darwin, NT, Australia
Chironex fleckeri	MAGNT C011186	-12.34	130.88	Darwin, NT, Australia
Chironex fleckeri	MAGNT C011975	-12.32	130.97	Shoal Bay, NT, AUS
Chironex fleckeri	MAGNT C014953	-12.19	132.34	Kakadu ntl. park, NT, AUS
Chironex fleckeri	MAGNT C014936	-12.15	135.01	Milingimbi, NT, AUS
Chironex fleckeri	MAGNT C014934	-12.15	132.24	Kakadu ntl. park, NT, Australia
Chironex fleckeri	MAGNT C014933	-12.06	132.39	Kakadu ntl. park, NT, AUS
Chironex fleckeri	MAGNT C014952	-12.03	134.93	Milingimbi, NT, AUS
Chironex fleckeri	MAGNT C014951	-11.97	134.23	Maningrida, NT, AUS
Chironex fleckeri	MAGNT C014954	-11.85	134.08	Maningrida, NT, AUS
Chironex fleckeri	SAM	-11.60	130.17	Bathurst Island, NT, AUS
Chironex fleckeri	MAGNT C015246	-11.27	132.12	Port Essington, NT, AUS
Chironex fleckeri	SAM	-11.20	132.14	Garik Gunak Baru ntl. park, NT, AUS
Chironex fleckeri	SAM	-10.74	142.38	Simpson Point, Qld, AUS
Chironex yamaguchii	USNM 27914	6.91	126.24	Pujada Bay, Mindanao, PH
Chironex yamaguchii	USNM 27916	7.70	122.00	Panabutan Bay, Mindanao, PH
Chironex yamaguchii	USNM 28691	10.14	118.78	Ulugan Bay, Palawan, PH
Chironex yamaguchii	USNM 28699	10.80	119.40	Malampaya Sound, Palawan, PH
Chironex yamaguchii	USNM 38016	10.88	119.54	Taytay, Palawan, PH
Chironex yamaguchii	USNM 28692	11.60	119.80	Malcochin, Linapacan, PH
Chironex yamaguchii	USNM 27913	12.02	124.03	Cataingan Bay, Masbate, PH
Chironex yamaguchii	USNM 27917	12.30	121.50	Mansalay Bay, Mindoro, PH
Chironex yamaguchii	USNM 28698	13.90	123.11	Tilik Bay, Lubang, PH
Chironex yamaguchii	USNM 28695	13.90	123.11	San Miguel Bay, Luzon, PH
Chironex yamaguchii	USNM 28700	*14.15*	*120.56*	Hamilo Point, Luzon, PH
Chironex yamaguchii	USNM 27911	14.74	120.22	Subic Bay, Luzon, PH
Chironex yamaguchii	USNM 28697	16.60	119.90	Bolinao Bay, Luzon, PH
Chironex yamaguchii	USNM 1121556	26.33	127.75	Okinawa Island, JP
Chironex yamaguchii	QM G317050	26.45	127.83	Nakagusuka Bay, Okinawa Island, JP
Chironex yamaguchii	QM G317051	26.47	127.97	Kana Beach, Okinawa Island, JP
Chironex yamaguchii	QM G317064	*26.30*	*127.73*	Chatan, Okinawa Island, JP
Chironex yamaguchii	Lewis & Bentlage (2009)	26.28	127.72	Ginowan, Okinawa Island, JP
Chironex yamaguchii	Lewis & Bentlage (2009)	*26.30*	*127.73*	Chatan Beach, Okinawa Island, JP
Chironex yamaguchii	Lewis & Bentlage (2009)	26.65	127.88	Motobu Port, Okinawa Island, JP
Chironex yamaguchii	Lewis & Bentlage (2009)	*24.44*	*124.12*	Sukuji Beach, Ishigaki Island, JP

Chiropsoides buitendijki	van der Horst (1907)	*-6.10*	*106.87*	Jakarta, Java, ID
Chiropsoides buitendijki	Fernando (1992)	*6.82*	*79.87*	Mount Lavinia, Colombo, LK
Chiropsoides buitendijki	Fernando (1992)	*6.86*	*79.86*	Mount Lavinia, Colombo, LK
Chiropsoides buitendijki	Menon (1936)	*9.23*	*79.22*	Krusadai Island, Tamil Nadu, IN
Chiropsoides buitendijki	NCL 106964	13.09	80.33	Chennai, Tamil Nadu, IN
Chiropsides quadrigatus	Haeckel (1880)	*16.37*	*96.36*	Rangoon, Burma
Chiropsoides buitendijki	Tahera & Kazmi (2006)	24.84	66.92	Sandspit, Karachi, PK

Table 3 Parameters used for model building, including the loadings of the first principal components (PC) and their importance expressed in cumulative proportions of variance; loadings < 0.1 not shown. All data were continuous. The spatial resolution of topography data was 2'; that of all other parameters was 2.52'. Topography data were acquired from http://topex.ucsd.edu/ (accessed August 2002); all other data were from http://oceancolor.gsfc.nasa.gov/ (accessed December 2007). nlw: normalized water-leaving radiance

Parameter	PC 1	PC 2	PC 3	PC 4	PC 5
Seafloor Topography	-0.224	0.116	-0.607	0.682	0.329
Sea Surface Temperature	0.211	0.115	-0.744	-0.419	-0.408
Chlorophyll	0.423			-0.140	
Calcite	-0.365	0.278	0.193		-0.187
Diffuse Attenuation	-0.415			-0.193	
nlw at 412 nm	0.386	0.244			0.126
nlw at 443 nm	0.365	0.322			
nlw at 488 nm	0.529	0.529	0.156		
nlw at 531 nm	-0.243	0.485		0.114	-0.505
nlw at 551 nm	-0.183	0.465		-0.491	0.637
Cumulative Proportion of Variance	0.54	0.79	0.89	0.95	0.99

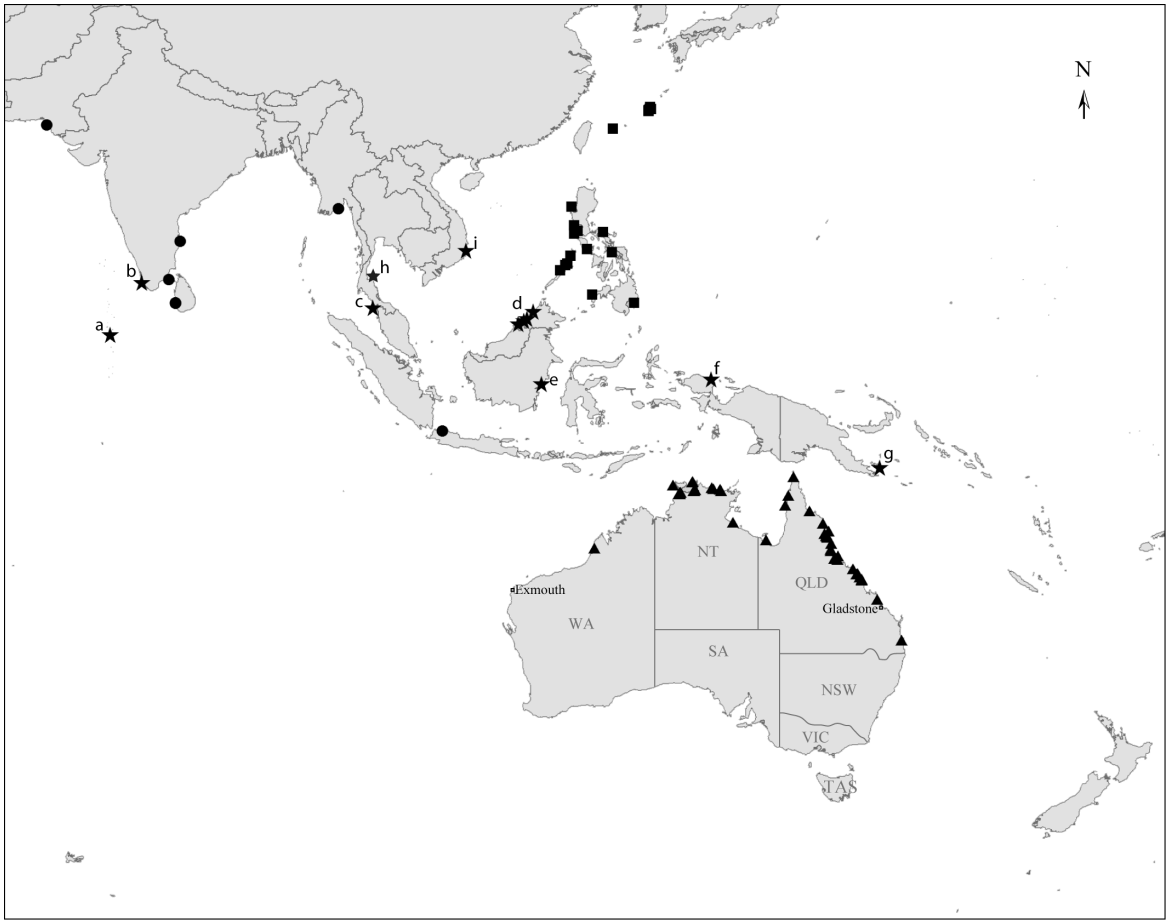


Fig. 13 Occurrence points used in ecological niche modeling, as well as locations from which undetermined chirodropid box-jellyfishes have been reported (the latter were not used in model-building). Reports from (a) the Maldives in Stiasny (1937), (b) India in Nair (1951), (c–g) Thailand (Andaman Sea), Malaysia, Brunei, Sulawesi, Irian Jaya, and Papua New Guinea, respectively, in Fenner & Williamson (1996), (h) Thailand (Gulf of Siam) in Suntrarachun *et al.* (2001), (i) Vietnam in Dawydoff (1936) and Ranson (1945). Triangle: *Chironex fleckeri*; dots: *Chiropsoides buitendijki*; squares: *Chironex yamaguchii*; stars: approximate locations from which indeterminate chirodropids have been reported. ACT: Australian Capital Territory, NSW: New South Wales, NT: Northern Territory, QLD: Queensland, TAS: Tasmania, VIC: Victoria, WA: Western Australia

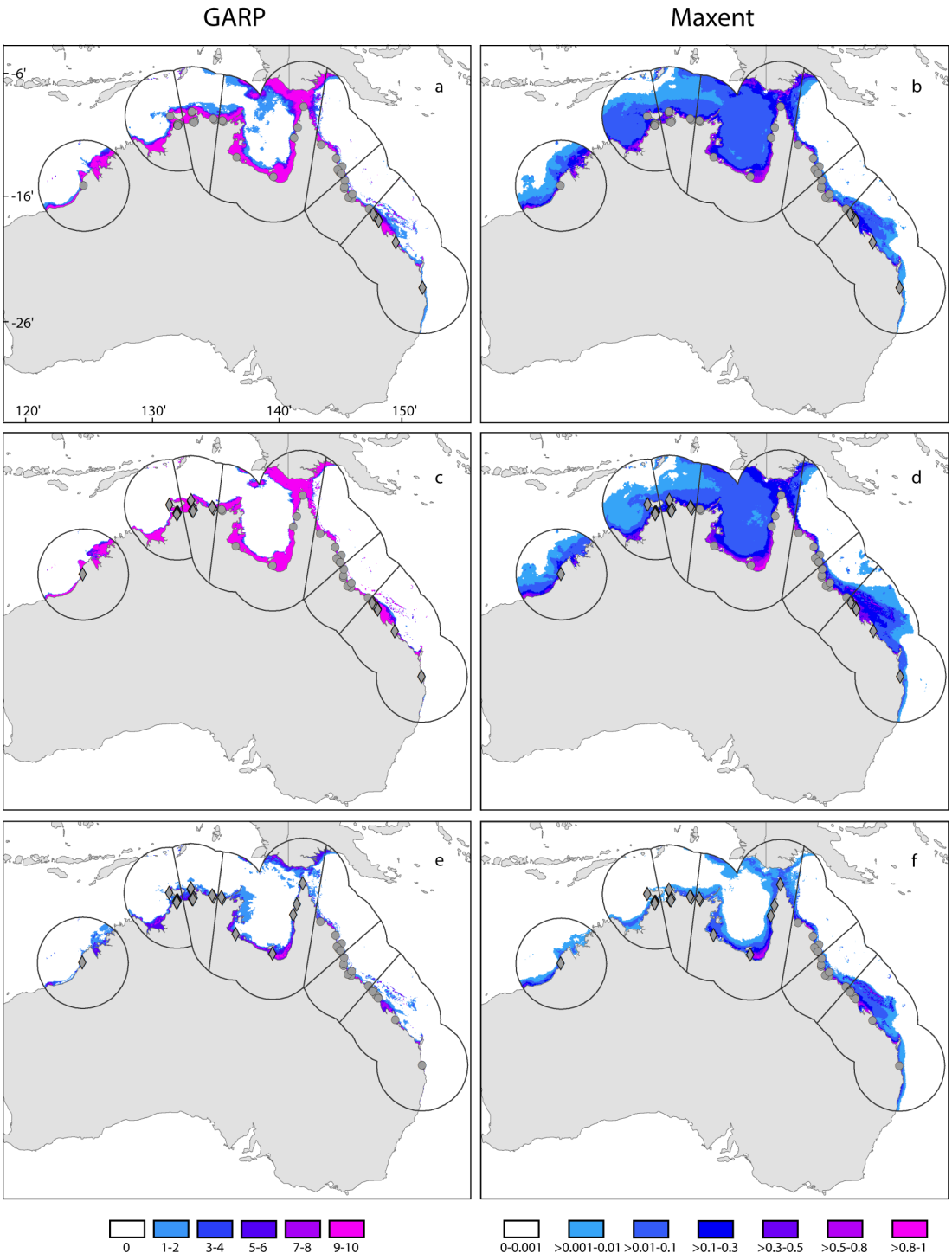


Fig. 14 Representative results from model-validation experiments with Australian *Chironex fleckeri*. Niche models were generated by GARP and Maxent within 5 or 3 of the 6 sectors

(sectors are delineated by straight lines) of the buffer zone that encompasses an area of 250 nautical miles (= 463 km) around each occurrence point. Strength of prediction is indicated according to the color gradient (see key). The GARP scores from 1 to 10 represent congruence among the 10 best models; Maxent output represents a probability distribution of occurrence. Circles: occurrence records used for model building; diamonds: occurrence points used for model evaluation

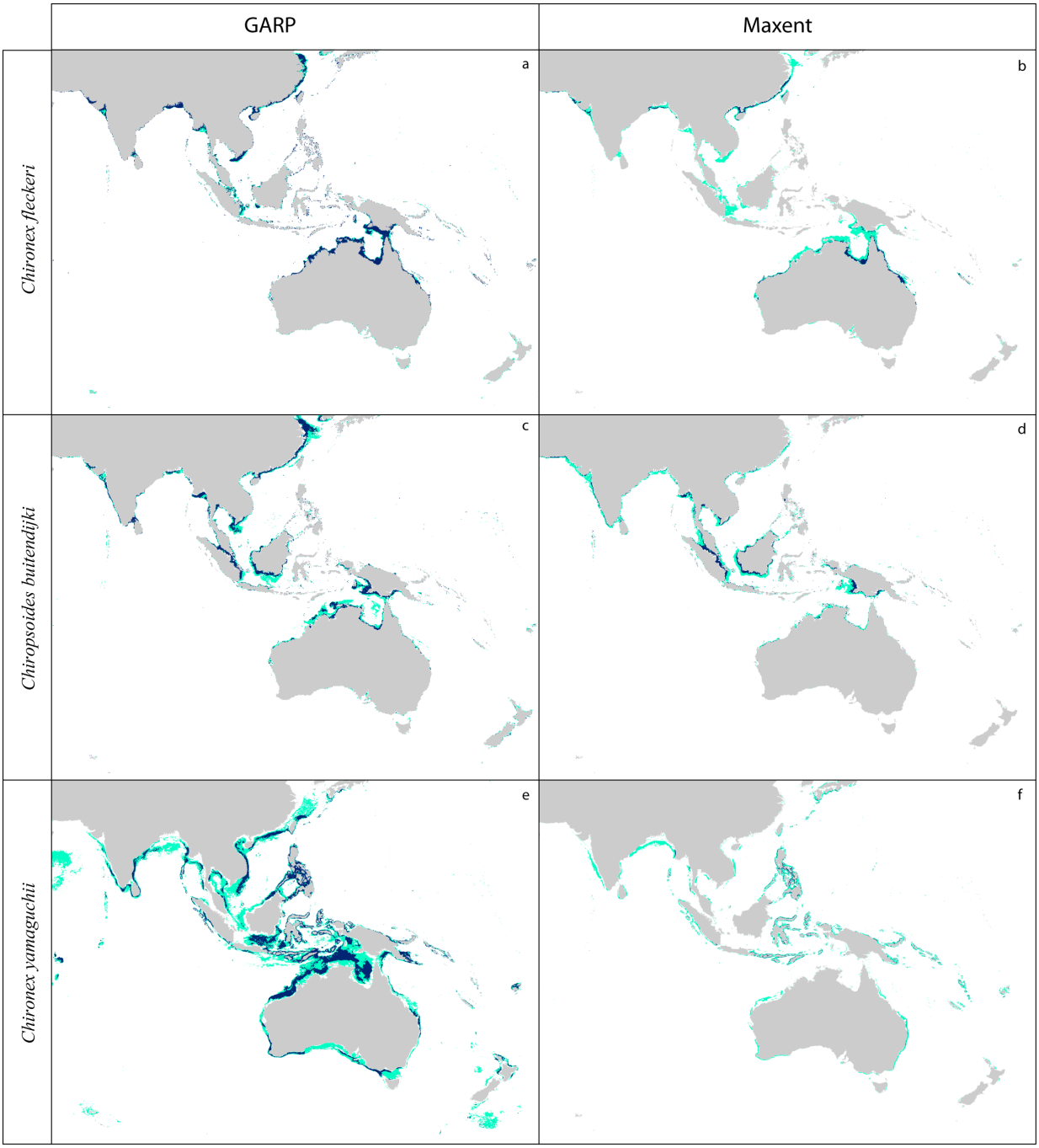


Fig. 15 Geographic distribution predicted for the 3 chirodropid species throughout Oceania and Asia, as inferred by GARP and Maxent. Color indicates the threshold applied to the broad-scale predictions. Green: lenient threshold; blue: stringent threshold

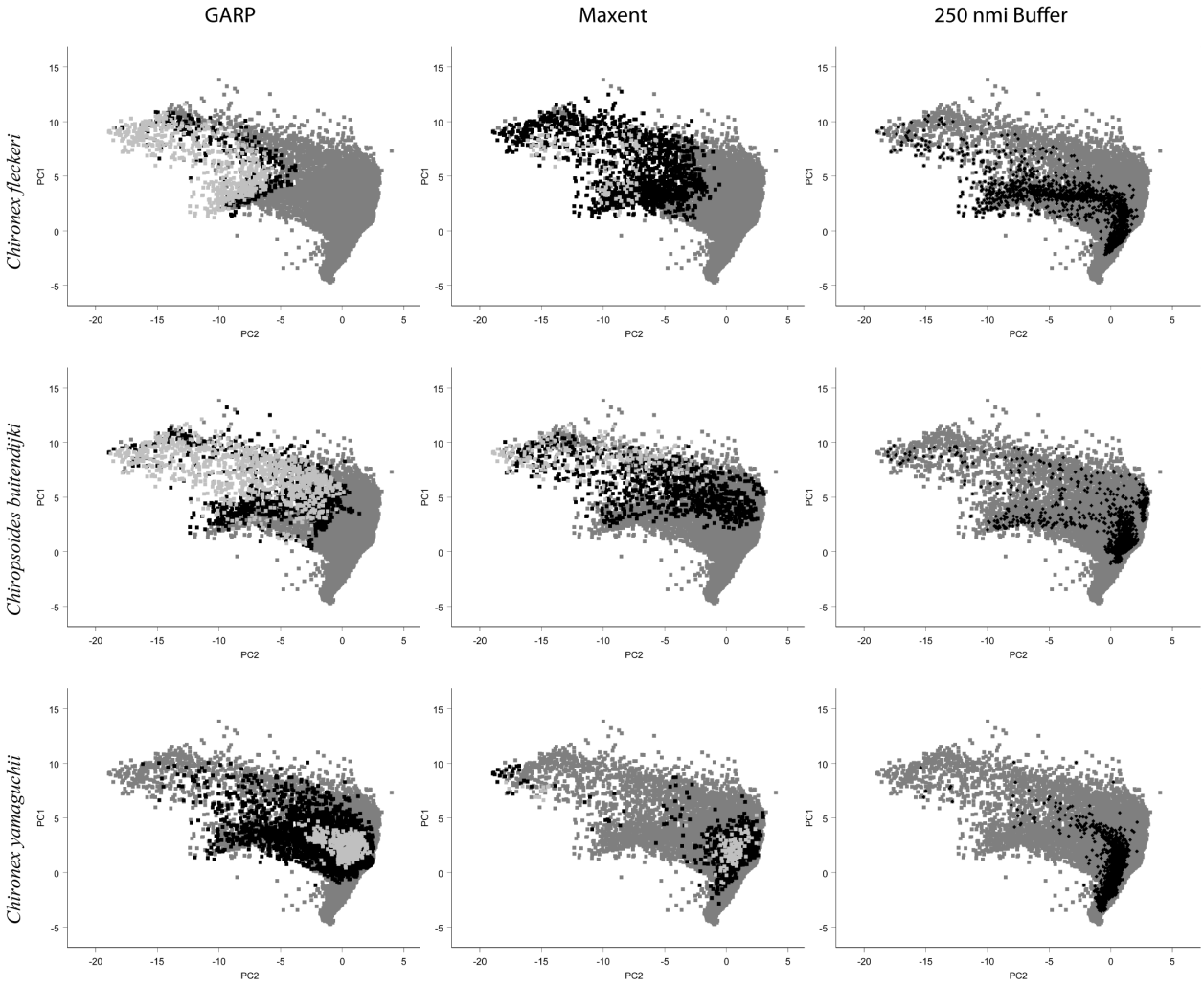


Fig. 16 Distribution of the first 2 principal components (PC) of each final niche model in the entire environmental space of the study area and proportion of the entire environmental space 'accessible' to each species. Dark gray squares: environmental space available in the study region; black squares (left and middle column): area occupied by models under lenient threshold; light gray squares: area occupied by models under stringent threshold; black diamonds (right column): environmental conditions present in a 250 nautical mile (nmi) (= 463 km) buffer around occurrence data points.

CHAPTER 4:

Plumbing the depths: extending ecological niche modeling and species distribution modeling in three-dimensional space

Abstract

Ecological niche modeling (ENM) and species distribution modeling (SDM) have been used extensively to study biogeographic and macroecological patterns of terrestrial fauna and flora. Few studies to date have applied ENM and SDM to marine ecosystems, and those that did treated the marine environment as a two-dimensional space owing to limitations of the implementations of current ENM/SDM tools. For many marine organisms, however, ENM/SDM would ideally be performed in three-dimensional space, taking into account latitude, longitude, and depth. Here, we present a case study using a modified ENM/SDM approach that can be used for three-dimensional ENM/SDM, using existing implementations of tools in geographic information systems. We decompose the three-dimensional structure of marine environmental data and species occurrence data into two-dimensional spaces using an easy to implement transformation, after which existing ENM and SDM tools can be used to analyze the data. We demonstrate our approach with the example of the potential distribution of a deep-sea-dwelling jellyfish (*Periphylla periphylla*). We employed two commonly-used algorithms, and assessed robustness of our models to missing data using a randomization approach. Transferability of models was assessed in a set of cross-prediction experiments. We were highly successful in building models that provided a predictive understanding of the three-dimensional distribution of the species in question. The approach presented herein can be applied to marine ecosystems

using existing tools and datasets, although similar to terrestrial ENM and SDM experiments, care needs to be taken in experimental design and model interpretation. In particular, interpolation among occurrence data-points and extrapolation into unsampled habitats present distinct challenges that may require different approaches and interpretations. Our modified ENM/SDM approach is straight forward to implement, and can be used to model habitats that have heretofore been beyond the reach of ENM/SDM applications. Potential applications of this new approach include estimating the geographic distributions and ecological niches of organisms inhabiting the water column in marine and freshwater environments, as well as the lower parts of the atmosphere (e.g., insetcs, birds, or bats).

Introduction

Studying the biogeography of organisms that inhabit the open oceans remains challenging owing to the difficulties of sampling and characterizing these vast and often inaccessible environments. As a result, knowledge of the distributions and ecology of many pelagic organisms remains limited. Open-ocean biogeographic patterns are best described and understood for the top-most layers of the water column (UNESCO 2009), and distributions of organisms inhabiting shallow waters are strongly correlated with environmental conditions at the sea surface (e.g., sea-surface temperatures and chlorophyll concentrations). Several of these sea surface parameters can readily be obtained from satellite imagery, which has led to the division of the oceans into well-defined biomes that describe large-scale distributions of marine organisms rather well (Longhurst 2007; UNESCO 2009).

In addition to defining large-scale biomes, sea-surface environmental parameters have been used in modeling studies that detect associations between species' occurrence records and observed environmental parameters to estimate distributions of individual species (this approach is usually referred to as ecological niche modeling [ENM] or species distribution modeling [SDM]). To a limited degree, this approach has been applied to study distributions of deep-sea organisms (e.g., Tittensor *et al.* 2010; Owens *et al.* 2012). Still, comparatively few studies have used ENM or SDM to study marine organisms (e.g., Robinson *et al.* 2011). Considering that environmental data for the global sea surface have recently become available in user-friendly formats (Tybergheim *et al.* 2010), and given the many opportunities that ENM and SDM open for studying marine biogeography and macroecology (e.g., Robinson *et al.* 2011), the field of marine ENM and SDM may expand greatly in the near future. However, current ENM and SDM

implementations treat occurrences and environmental data as two-dimensional which may impede open ocean ENM and SDM. We address this issue here by expanding ENM/SDM to three-dimensional spaces.

ENM and SDM in three-dimensional environments. ENM and SDM address the problem of limited occurrence data that are suitable to describe distributions of organisms by providing predictions about a species' range in ecological and geographic dimensions. ENM and SDM represent correlative approaches, in which occurrence records for an organism are related to raster environmental data layers using diverse algorithms to reconstruct limits in environmental dimensions. This ecological niche (cf., Soberón 2007; Soberón & Nakamura 2009; Godsoe 2010; Peterson *et al.* 2011) can then be projected across broader landscapes than the original sampling area to identify the potential geographic distribution of the species. In addition to geographically explicit predictions about distributions, the ecological niche is estimated, at least in a relatively few coarse-resolution dimensions, allowing inferences regarding ecological properties of the species in question (Soberón & Nakamura 2009; Godsoe 2010; Peterson *et al.* 2011).

The approach outlined above has been applied most widely in terrestrial systems, as the tools developed for ENM and SDM were implemented for two-dimensional geographic spaces (i.e., grid-cells are addressed by latitude and longitude in a geographic information system [GIS]). This limited dimensionality does not necessarily represent a problem in marine environments if the species of interest is restricted to shallow water that is well-described by sea-surface environmental parameters, or if it inhabits the benthos and environmental data layers for the ocean-/sea-floor are available. Pelagic species, inhabiting the middle layers of the water

column (the mid-water), however, present challenges to current ENM and SDM implementations because they inhabit a genuinely three-dimensional environment in which locations can be described completely only by longitude, latitude, and depth. Such organisms may move, actively or passively transported by currents, in all three of these dimensions. In addition to holopelagic mid-water species, some species are meroplanktonic, inhabiting the benthos at some stage of their life cycle, while entering the water column during another life cycle stage. In summary, situations exist in which using two-dimensional models is inadequate in an environment that is actually three-dimensional and mostly lacks fast physical boundaries to dispersal in horizontal and vertical directions. Nonetheless, ENM/SDM offers a means to address challenges posed by sparse occurrence data for studying mid-water biogeography, allowing characterization of distributions of species throughout the oceans in both geographic and ecological spaces. In this contribution, we present and evaluate a strategy that allows existing implementations of ENM/SDM algorithms to be applied to problems in three-dimensional environments, in the hopes of stimulating further ENM/SDM applications to ocean biogeography and macroecology. To demonstrate our approach, we model the potential distribution of the open-ocean-inhabiting scyphozoan (phylum Cnidaria) jellyfish *Periphylla periphylla* on a global scale.

Materials and methods

Conceptual framework. Much debate has centered on what it is that is modeled when one estimates correlations between environmental variables and observed presences of a species (cf., Araújo & Peterson, in press). To avoid adding to the confusion, we outline the conceptual framework we have employed (cf., Peterson *et al.* 2011). In essence, the distributional limits of a

species are determined by the scenopoetic conditions (e.g., climate) **A**, as well as the bionomic conditions (e.g., biological interactions among species, density-dependent resource consumption) **B** that favor a positive growth rate of the populations of a species. This area can be referred to as the potential distribution $\mathbf{G}_P = \mathbf{G}_O \cup \mathbf{G}_I$, where $\mathbf{G}_O$ is the occupied distributional area and $\mathbf{G}_I$ is the distributional area that allows for a positive growth rate of the species' populations but has not been colonized by the species. The geographic area **M** (a subset of the study area **G**) determines which part of $\mathbf{G}_P$ was accessible to the species (i.e., that the species has been able to reach via dispersal) and is occupied by the species ($\mathbf{G}_O$).

Scenopoetic environmental data are readily available at various spatial extents, and at various spatial and temporal resolutions; few datasets describing strictly bionomic conditions **B** are available. At large spatial extents and coarse resolutions however, **B** may be less important (at such scales one may assume that the relevant area $\mathbf{A} \cup \mathbf{B}$ is small with respect to **B**) than at smaller, more local scales, at which local population extinction events due to biological interactions are prevalent (cf., Pearson & Dawson 2003; Soberón 2010). Despite generally omitting **B** from the model fitting process, ENMs and SDMs appear to capture a significant amount of ecological signature (Pearson & Dawson 2003; Elith & Leathwick 2009), and have repeatedly resulted highly informative about complex biological phenomena. Lastly, **M** plays a crucial role in both model fitting and interpretation in ENM/SDM exercises (Anderson & Raza 2010; Soberón 2010; Barve *et al.* 2011). Estimation of **M** can be complicated and complex (Barve *et al.* 2010), as it involves knowledge of the history of the species and its dispersal capabilities through space and time. In some cases, good *a priori* assumptions of **M** can be made, for example by considering the history of the study region **G** (Soberón 2010). Depending on the

algorithm employed, the geographic setting, and the nature of the data (e.g., presence-only versus presence-absence), ENM/SDM algorithms may recover different parts of $\mathbf{A} \cup \mathbf{B} \cup \mathbf{M}$.

Data transformation: 2D to 3D. Current ENM and SDM algorithms use environmental variables and occurrence data for model fitting on a cell-by-cell basis on GIS grids without considering topological effects among cells. As a consequence, the spatial arrangement of cells in the grids does not affect the outcome of ENM and SDM analyses, and we can apply a straight-forward transformation of the geo-referenced environmental and species occurrence datasets to permit representation of the three dimensions of aquatic environments. Environmental grids summarizing environmental variables used for model fitting in three-dimensional space are associated with particular water depths. These grids can be thought of as being part of an ordered sequence of size **n**, where **n** equals the number of grids in the sequence and $\mathbf{d}_i$ refers to the *i*-th depth layer. Each grid corresponds to an element of the sequence of depths ($\mathbf{d}_0$,...,$\mathbf{d}_{n-1}$), which ranges from the shallowest layer ($\mathbf{d}_0$) to the deepest layer ($\mathbf{d}_{n-1}$). Using this framework, we can transform the latitudes of each layer to create an array of depth layers (Fig. 1). That is, we create a strip of equirectangular grids in which adjacent grids correspond to consecutive depth levels (Fig. 1). If **r** is the range of the set of original latitudes (e.g., if the geographic extent of the grids range from 90ºN to 90ºS and the grid-cell-size = 1º then **r** = 180) we can transform all cells of the environmental grids and the coordinates of the species occurrences as $\text{Latitude}' = \text{Latitude} + \mathbf{r} \cdot i$. In practice, environmental layers can be transformed by concatenating ASCII grid files; shell scripts for this conversion can be obtained from http://purl.org/3D-ENM-and-SDM-conversion-tools.

Example analysis and performance evaluation. *Species occurrence records.* We

present an example analysis of a species of scyphozoan jellyfish (Cnidaria: Medusozoa), *Periphylla periphylla*. *P. periphylla* makes a good example to test our strategy because it is the sole member of the genus *Periphylla*, and is readily identifiable even by non-specialists, making occurrence data reliable. In addition, its global distribution is relatively well understood. *P. periphylla* generally inhabits the meso- (200-1000 m depth) and bathypelagic (1000-4000 m depth) and is cosmopolitan in distribution. *P. periphylla* is holopelagic, and all of its life cycle stages co-occur in the water column (Jarms *et al.* 1999; Jarms *et al.* 2002), suggesting that no ontogenetic shifts in ecological niche characteristics are to be expected. Considering the areas from which *P. periphylla* has bee sampled, we expect that effective models for *P. periphylla* will predict large, cohesive areas of the deep oceans, plus shallower areas at higher and colder latitudes and in upwelling regions. We obtained occurrence records from the catalog of the National Museum of Natural History (Smithsonian Institution, Washington, DC, USA), the Automated Video Annotation and Reference System (http://www.mbari.org/vars/) of the Monterey Bay Aquarium Research Institute (Moss Landing, CA, USA), and the Ocean Biogeographic Information System's online database (http://www.iobis.org/). After removing records that were not spatially unique in all three dimensions (i.e., latitude, longitude, and depth), our dataset consisted of 100 records (Fig. 2).

Environmental data. We used the global 1° grid dataset of objectively analyzed climatological fields of the 2005 World Ocean Atlas (WOA05; http://www.nodc.noaa.gov/OC5/WOA05/pr_woa05.html), which represent interpolated mean fields for oceanographic variables at standard depths. The depth resolution of the WOA05 was coarsened for the first 12 standard depths, as the high vertical resolution of the WOA05 at

shallow depths exceeded the accuracy of depth data associated with the species occurrence data in several cases. For this reason, oceanographic variables for standard depths 1 through 7, 8 through 10, and 11 and 12 were averaged across depth layers, resulting in environmental grids for 24 depth layers (Table 1).

We calculated minimum and maximum values for temperature, salinity, O_2 concentration, and apparent oxygen utilization from WOA05 seasonal climatologies (Antonov *et al.* 2006; Garcia *et al.* 2006a; Locarini *et al.* 2006). We used averaged annual climatologies for nitrate (N) and phosphate (P) concentrations (Garcia *et al.* 2006b), as the depth resolution of seasonal climatologies for N and P did not allow calculation of extreme values for N and P. Furthermore, we calculated grids of "potential productivity" according to a stoichiometric nutrient ratio of N:P of 16:1 (i.e., the Redfield ratio; cf., e.g., Lenton & Watson 2000). Since strong correlations among environmental grids were apparent, we used principal component analysis in ArcMap ver. 9.2 (Esri, Redlands, CA, USA) to transform the data. Of the 11 components, we retained the first 3 which explained 99.7% of the total variation in the data; these three principal components were the basis for all further analyses.

Modeling approach. We employed two commonly used algorithms, the maximum entropy algorithm implemented in Maxent (Phillips *et al.* 2006) and the Genetic Algorithm for Rule-set Prediction (GARP; Stockwell & Peters 1999). Maxent runs were performed using the default settings (autofeatures = yes; random test percentage = 0%; regularization multiplier = 1; maximum iterations = 500; convergence threshold = 0.00001; maximum number of background points = 10000) in Maxent ver. 3.1.0. The logistic output of Maxent was used in all modeling experiments and model evaluations. GARP runs were performed using the multi-threaded

openModeller (ver. 1.1) "best subsets" implementation of GARP (Muñoz *et al.* 2009) using the following settings: training proportion = 50%; total runs = 100; soft omission threshold; models under omission threshold = 20; commission threshold = 50%; commission sample size = 999; maximum number of generations = 400; convergence limit = 0.01; population size = 50; resamples = 2500; number of threads ranged from 2 to 16.

Models were fit using the whole world as training area and projected across the same region. The lowest prediction score of an occurrence record (the lowest presence threshold) was used as a threshold to convert continuous or categorial predictions into binary predictions. In cases where occurrence records used for training fell into cells with zero prediction scores, we used the next lowest prediction score as a threshold. Models for cross-predictions (see below) were fit on subsets of the grids (e.g., the Northern Hemisphere).

Note that our modeling approach assumes that *P. periphylla* was able to disperse to all suitable habitats (**M = G**). Considering that *P. periphylla* can be found in large abundances in all oceans, belongs to a lineage dating back to the Cambrian (Cartrwight *et al.* 2007), and that the oceans most likely lack hard barriers to dispersal of pelagic organisms over this time-frame (i.e., ~500-600 million years), *P. periphylla* probably had the potential to disperse to all suitable habitats ($\mathbf{G}_P$).

R*obustness and transferabilit*y. We assessed sensitivity of models to missing occurrence data (model robustness) using the following resampling approach (Fig. 3). We generated 100 sets of occurrence data, each consisting of 10 pseudo-replicates. Pseudo-replicates were generated by drawing 50% of the original occurrences at random without replacement. Maxent and GARP models were fit and projected as outlined above, but not transformed into binary predictions.

Instead, model predictions were rescaled by dividing each prediction score in a model prediction grid by the sum of all scores of the grid (normalization *sensu* Warren *et al.* 2008). We paired normalized pseudo-replicate and normalized full-data (i.e., all occurrences were used during model fitting) predictions, and calculated overlap among pseudo-replicate and full-data niche models using Hellinger Distance (*I*) and Bray Curtis Distance (*BC*). This resulted in 10 *I* and *BC* values for each set of pseudo-replicates. Averages, $\bar{I}$ and $\bar{BC}$, were calculated for each pseudo-replicate dataset, which led to a distribution of 100 averages from which we calculated the average ("average of averages"), and 2.5% and 97.5% percentiles as a confidence interval (cf., bootstrap procedure for estimating a confidence interval in Zoubir & Boashash 1998). This analysis was performed using the Bash command language (http://www.gnu.org/software/bash/) and the functions provided in the NichePy library (ver. 1.1; Bentlage & Shcheglovitova 2012).

Model transferability was assessed by performing two different cross-prediction experiments. We subset species' occurrence data spatially and divided the environmental data into pairs as follows: Northern Hemisphere (68 occurrence records) versus Southern Hemisphere (32 occurrence records), "shallow" (≤ 650 m depth; 38 occurrence records) versus "deep" (> 650m depth; 62 occurrence records) water, Atlantic (53 occurrence records) versus combined Pacific and Indian Oceans (47 occurrence records).

Using binary predictions, we evaluated whether or not occurrence data omitted from the model fitting process fell into grid cells of predicted presence more often than would be expected by chance (Anderson *et al.* 2002; Peterson *et al.* 2011). For this purpose, we fit models across the spatial subset, and subsequently projected to the spatial subset omitted from the model fitting process. Models were transformed into binary predictions by applying a threshold as described

above. Significance was assessed by calculating cumulative, 1-tailed binomial probabilities of observed proportions of occurrence data falling into grid cells of predicted presence versus absence.

Whether or not two models were significantly different was assessed using Warren *et al.*'s (2008) niche identity test as implemented in NichePy (ver. 1.1; Bentlage & Shcheglovitova 2012). For this test, occurrences from each subset pair were pooled and pseudo-replicates corresponding to the size of each subset occurrence dataset were drawn from the pooled dataset at random without replacement (cf., Bentlage & Shcheglovitova 2012); this process was repeated 100 times. For each pseudo-replicate, models were fit using the whole world as training and projection area. *I* and *BC* were calculated for all pairwise pseudo-replicate comparisons after normalizing models, thus creating a distribution of *I* and *BC* values. Models for the subset occurrence data were generated by fitting models in spatial subsets (e.g., the Northern and Southern Hemispheres) and then projected across the whole world; *I* and *BC* were calculated after normalization of models to compare subset models. Subset *I* and *BC* values were compared to the distribution of pseudo-replicate *I* and *BC* values using a 1-tailed test to assess whether or not the subset models were significantly less similar than pseudo-replicate models.

Results

Maxent assigned high predictive values to large geographic areas in the deeper ocean layers. In addition, shallow waters in higher latitudes as well as upwelling regions (e.g., the western coasts of the Americas and Africa) received higher predictive scores than other surface water areas (Fig. 4). We observed strong disagreement between initial Maxent and GARP predictions when using

all occurrence data to fit the GARP model (not shown). In particular, GARP assigned high predictive values to sea-surface habitats throughout the world, including the sub-tropics and tropics. This prediction was in disagreement with existing knowledge of the distribution of *P. periphylla* and was a likely artifact of bias in the density of distributional records (68 for the Northern Hemisphere versus 32 for the Southern Hemisphere). In this case, thresholding of model predictions proved ineffective to dampen the effects of this bias but exaggerated it. However, subsampling of the occurrence data from the Northern Hemisphere to the density of the Southern Hemisphere (drawing 32 occurrences at random from the 68 Northern Hemisphere occurrences) led to GARP predictions that were more congruent but not identical with Maxent predictions (Fig. 4). GARP generally assigned higher predictive values to the northern and southern areas of ocean basins compared to the central areas. In addition, GARP assigned high predictive values to surface waters on western coasts that far exceeded the range of upwelling regions. Maxent assigned high predictive scores across all oceans and depths, in contrast to GARP (last panel Fig. 4).

Model predictions appeared relatively robust to missing data. Models built with 50% of the occurrences were more similar to than different from models generated using all occurrences (averages of $\bar{I}$ and $\bar{BC} > 0.5$ on a scale where 0 = no niche overlap and 1 = total niche overlap; Table 2). *BC* values were generally smaller than *I* values, which is expected considering that *BC* represents a stricter measure of niche overlap (Rödder & Engler 2011; Bentlage & Shcheglovitova 2012). When challenged with predicting between geographic areas (i.e., extrapolation challenge), GARP appeared to perform better than Maxent for our data. GARP was successful in predicting occurrence data omitted from the model fitting process better than

random predictions in 5 of the 6 cross-prediction experiments; GARP failed to predict occurrences in the Northern Hemisphere using models fit in the Southern Hemisphere ($p > 0.05$). By contrast, Maxent failed to predict occurrence data omitted during model fitting in half of the experiments. Maxent was able to cross-predict occurrences from "shallow" and "deep" waters better than random expectations ($p < 0.01$), and was able to predict occurrences in the Atlantic using models fit in the Indian and Pacific Oceans ($p < 0.01$), but failed to predict occurrences in the Indian and Pacific Oceans when fit with data from the Atlantic ($p > 0.05$). Maxent also failed to cross-predict occurrences across hemispheres ($p > 0.05$). Analogous to these results, Maxent and GARP models fit in the Northern and Southern Hemispheres, using occurrences that were sampled from largely distinct environments, were significantly different (uppermost panel, Fig. 6). Even though models fit in different basins (Atlantic Ocean versus Indian and Pacific Oceans) were significantly different for both GARP and Maxent, GARP models were highly similar in this case while Maxent models were quite distinct (middle panel, Fig. 6). Here, occurrences from the Atlantic and combined Indian and Pacific Oceans displayed more overlap in environmental dimensions compared to predictions across hemispheres. Lastly, models fit at different water depths where highly similar for both GARP and Maxent, with GARP models not displaying any significant differences (lowest panel, Fig. 6). In this cross-prediction experiment, occurrences from "shallow" and "deep" waters were sampled from similar and overlapping environments.

Discussion

The approach presented here expands the ability to study ocean biodiversity using ENM and

SDM to gain insight into distributional patterns in three-dimensional habitats. Although our demonstration used presence-only data, the same strategy can be used in presence-absence modeling (but see Bentlage *et al.* 2009 for a discussion of the difficulty of obtaining absence data for marine invertebrates). It is important to note that several challenges and conceptual issues pertinent to terrestrial ENM and SDM are equally important to consider when using ENM and SDM to make predictions about distributions of marine organisms in geographic and environmental space. Below, we discuss these issues in light of the exemplar analysis presented in this contribution.

What did we model? Considering *P. periphylla*'s evolutionary history and documented geographic distribution, it seems reasonable to assume that the species is in distributional equilibrium on a global scale. In other words, it seems probable that the entire study region **G** has been accessible to *P. periphylla* over relevant time-scales (i.e., **M** = **G**). That is, for *P. periphylla* the potential distribution ($\mathbf{G}_P$) is equal to $\mathbf{G}_0$ and $\mathbf{G}_I$ is null. It is worth noting that the environmental variables employed for modeling contain both strictly scenopoetic variables (e.g., temperature and salinity) and variables that could be part of **B** depending on the spatio-temporal scale of the analysis (e.g., community effects on dissolved nutrients like phosphorous, nitrogen, etc. may change depending on the spatio-temporal scale). At the large spatial scales and coarse resolutions of the environmental variables employed herein, this distinction may not be critical (see "conceptual framework"). In summary, if we assume that *P. periphylla* was able to reach all suitable habitats, global niche models (Fig. 4) may be interchangeably interpreted as suitable habitat models or models representing geographic distributions.

***Periphylla periphylla*: modeled versus documented distributio**n. The distribution of *P.*

periphylla is relatively well-known and includes deep waters worldwide except for the Arctic Ocean (e.g., Russel 1970). In addition, museum records (some of which we used for model fitting) show that *P. periphylla* occurs in surface waters in several places (e.g., upwelling regions in which cold, nutrient-rich waters are transported from the deep sea to the surface). The Maxent model (Fig. 4) generally fit the documented distribution of *P. periphylla* well and predicted large areas of the deep-sea as part of its potential distribution $\mathbf{G}_{P}$. In addition, the surface waters in upwelling regions like the west coasts of the Americas and Africa were part of Maxent's prediction of $\mathbf{G}_{P}$. Initial GARP predictions strongly contradicted Maxent predictions, which appeared to be an artifact of bias in the density of distributional records. Random sub-sampling of occurrences from the Northern Hemisphere to the density of the Southern Hemisphere reduced this artifact somewhat and led to GARP predictions that were more congruent with *P. periphylla*'s documented distribution and Maxent predictions (Fig. 4). This GARP model seemed to generally favor higher latitudes while omitting lower latitudes from its prediction of $\mathbf{G}_{P}$. In addition, parts of the Arctic Ocean and large areas of the eastern Atlantic and Pacific oceans were part of GARP's prediction of $\mathbf{G}_{P}$; the latter areas extended far beyond upwelling regions. These parts of the prediction likely represent overprediction in both geographic and environmental dimensions. Even though potential overprediction may be undesirable in some circumstances, it can be of much use in cases where one is interested in making predictions about environments that were not used during model fitting (i.e., extrapolation).

Model robustness and transferability. At the extreme ends of the spectrum, SDM and ENM are confronted with two main challenges: interpolation and extrapolation. In the case of interpolation, a dataset of species occurrence data exists that describes the distribution of the

species in environmental dimensions well. The challenge, then, is to fill in the gaps among known occurrence data-points, given the environmental information contained in the known occurrences. In the case of extrapolation, on the other hand, one wishes to make predictions about areas in geographic and/or environmental space using occurrence records sampled from elsewhere (e.g., risk assessment of species' invasive potential, responses of biota to climate change, or simply sampling bias). While interpolation challenges are the realm of SDM, extrapolation is the realm of ENM.

With our resampling and cross-prediction experiments, we attempted to simulate these two challenges, to visualize and evaluate their effects on modeling results. Different algorithms may perform well in interpolation exercises while failing at extrapolation or *vice versa*. At first glance, Maxent seemed more robust to missing data than GARP for our data (Fig. 5). However, direct comparisons of the effect of missing data should not be drawn between these two algorithms here since differences in Maxent and GARP outputs may lead to systematic differences in the calculation of *I* and *BC*. In particular, Maxent predictions are continuous while GARP predictions are categorical, which may amplify differences among GARP predictions compared to comparisons among Maxent predictions. This issue is analogous to systematic biases affecting other metrics for model comparisons (e.g., Peterson *et al.* 2008). Regardless, Maxent produced predictions that were extremely robust to missing occurrences while GARP pseudo-replicate models were more similar to than different from models fit using all occurrences (Fig. 5; Table 2).

When challenged with extrapolation, GARP was highly successful in predicting unsampled occurrences and geographic areas, as exemplified by its ability to predict missing data

and produce similar models across a depth gradient and ocean basins (Fig. 6). At the same time GARP, failed in a cross-prediction experiments. For our data, Maxent succeeded in fewer cross-prediction experiments than GARP; comprehensive evaluation of this topic is beyond the scope of this contribution, and this issue has been discussed elsewhere (e.g., Peterson *et al.* 2007; Phillips 2008). However, when designing ENM/SDM experiments to study marine organisms using our modified approach, the differences among interpolation and extrapolation should be considered during experimental design and model interpretation: for example, sampling that fails to capture the full environmental range of the species can lead to overfitting with respect to the true distribution of the species in environmental dimensions (e.g., Thuiller *et al.* 2004).

Concluding remarks. It is our hope that three-dimensional ENM and SDM applications will help to further the understanding of the biogeography and macroecology of marine species, in particular for species that inhabit areas of the oceans that are difficult to sample. The paucity of occurrence data from mid- and deep-water habitats has impeded the study of marine biogeography and macroecology by preventing delineation of species' distributions in geographic and environmental dimensions (cf., UNESCO 2009). ENM and SDM tools are capable of addressing this issue by deriving predictions for mid-water and deep-sea species' distributions based on limited occurrence data. In our example analysis, the potential distribution and realized geographic distribution of *P. periphylla* were expected to be identical, but for other species this may not be the case. ENM and SDM derive a model of part of the species' niche which may predict geographic areas that are suitable for a positive population growth rate but that have not been colonized by the species. In such cases, post-processing of ENMs/SDMs, for example with the aid of ocean circulation models, may prove necessary for development of accurate

descriptions of species' distributions in geographic space. In addition to marine ENM/SDM, the strategy presented here can be applied to species' distributions in other three-dimensional habitats, like ponds and lakes, or potentially the atmosphere for volant animals or pollen.

Table 4 Depth resolution of the 2005 World Ocean Atlas (WOA05) grids employed in this contribution. A key to the standard depths can be found in the WOA05 documentation (ftp://ftp.nodc.noaa.gov/pub/WOA05/DOC/woa05documentation.pdf). **d** represents the factor used to transform latitudes. Each value of **d** corresponds to a depth interval to which species occurrence data were mapped. [a, b]: a and b included in interval; (a, b]: a excluded from interval and b included in interval.

Standard Depth	d	Depth Interval in Meters
1-7	0	[0, 112.5]
8-10	1	(112.5, 225]
11-12	2	(225, 350]
13	3	(350, 450]
14	4	(450, 550]
15	5	(550, 650]
16	6	(650, 750]
17	7	(750, 850]
18	8	(850, 950]
19	9	(950, 1050]
20	10	(1050, 1100]
21	11	(1150, 1250]
22	12	(1250, 1350]
23	13	(1350, 1450]
24	14	(1450, 1625]
25	15	(1625, 1875]
26	16	(1875, 2250]
27	17	(2250, 2750]
28	18	(2750, 3250]
29	19	(3250, 3750]
30	20	(3750, 4250]
31	21	(4250, 4750]
32	22	(4750, 5250]
33	23	(5250, 5750]

Table 5 Average of *I* and *BC* averages from pseudo-replicate/full-occurrence-dataset comparisons; see Fig. 4 for details on the resampling procedure. The 2.5% and 97.5% bounds are the percentiles of the resampling averages of *I* and *BC*; both *I* and *BC* range from 0 (no niche overlap) to 1 (niches identical).

	$\bar{I}$			$\bar{BC}$		
	Average	*2.5%*	*97.5%*	*Average*	*2.5%*	*97.5%*
Maxent	0.987	0.985	0.989	0.879	0.871	0.889
GARP	0.788	0.744	0.821	0.618	0.572	0.661

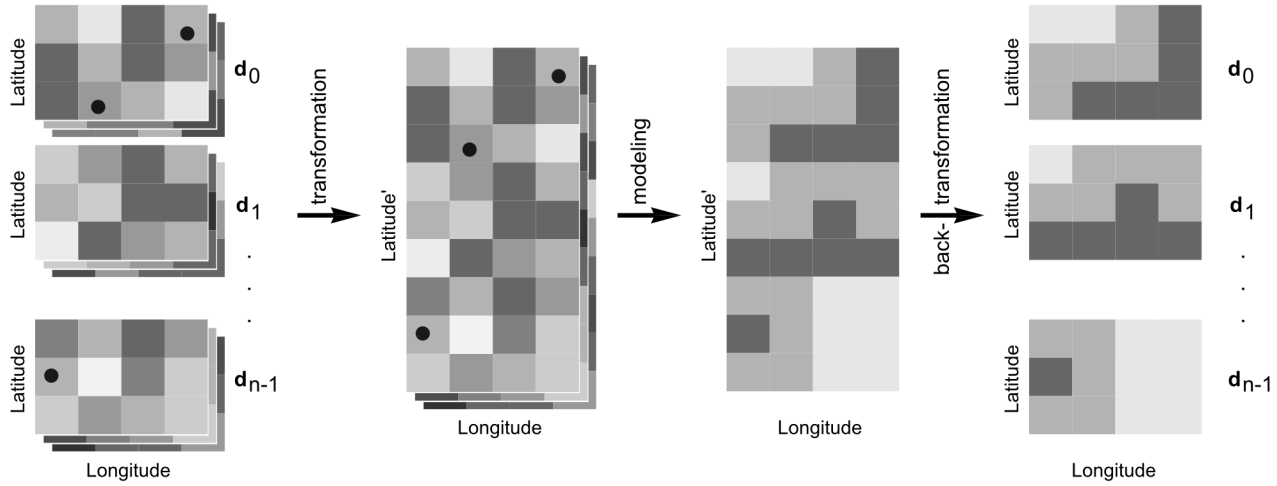


Fig. 17 Transformation of GIS data for three-dimensional ENM/SDM. Species occurrence data (dots) and environmental grids are combined by transforming latitudes based on their association with a depth layer (**d**). The transformed species occurrence data and environmental grids can then be used for ENM/SDM. Post-processing of ENMs/SDMs involves the back-transformation of latitudes.

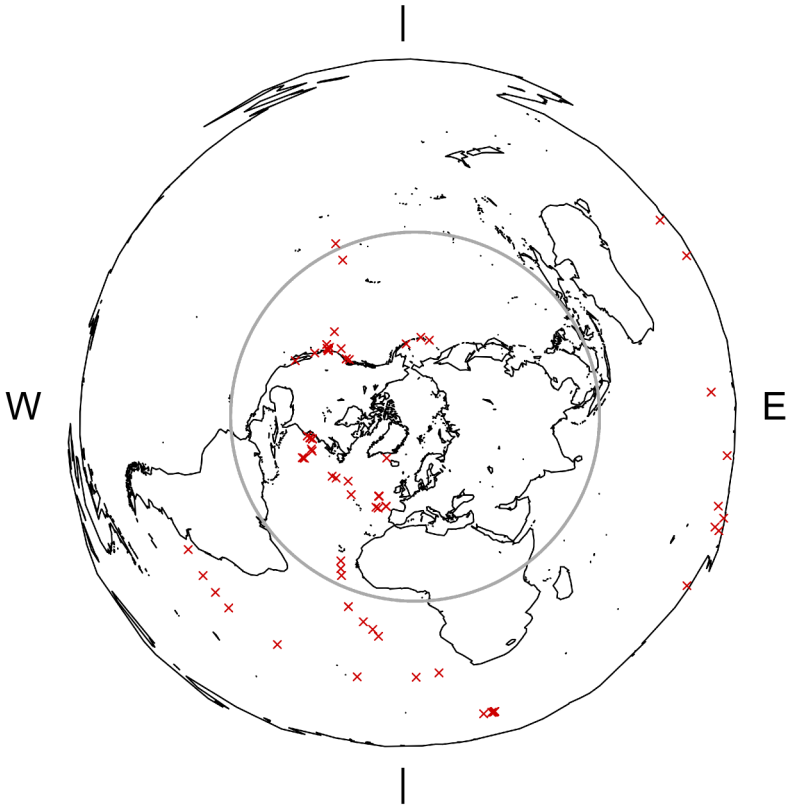


Fig. 18 Azimuthal equidistant projection depicting the location of occurrence data (red crosses). Gray line: equator; E: Eastern Hemisphere; W: Western Hemisphere; tick-marks (|): Prime Meridian and International Date Line.

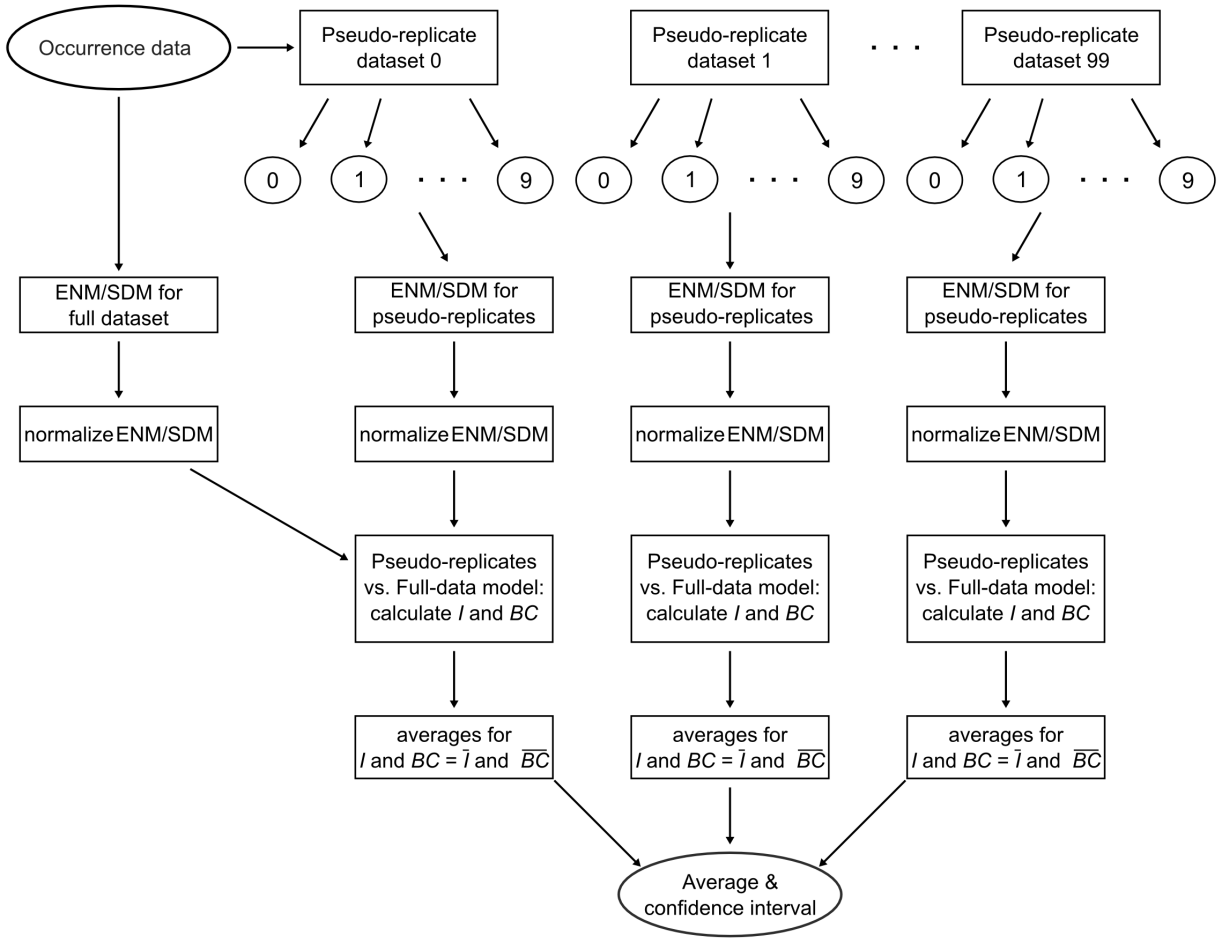


Fig. 19 Flow-chart outlining the occurrence data resampling procedure employed to test model robustness to missing data. Pseudo-replicates are generated by drawing 50% of occurrences from full occurrence dataset without replacement.

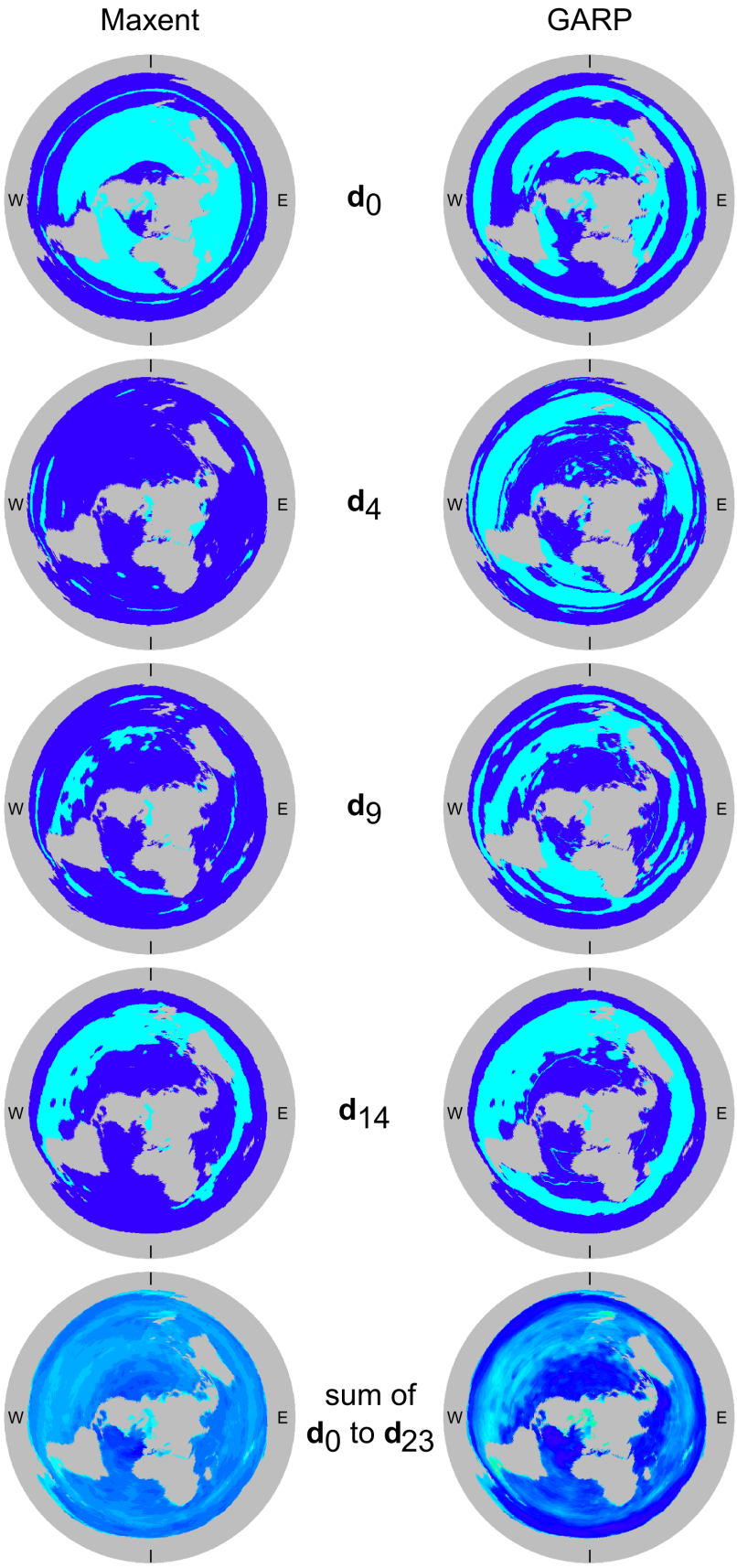

Maxent
GARP
W
E
$\mathbf{d}_0$
$\mathbf{d}_4$
$\mathbf{d}_9$
$\mathbf{d}_{14}$
sum of
$\mathbf{d}_0$ to $\mathbf{d}_{23}$

Fig. 20 (previous page) Azimuthal equidistant projections of Maxent and GARP models obtained using all species occurrence records. $\mathbf{d}_0$, $\mathbf{d}_4$, $\mathbf{d}_9$, and $\mathbf{d}_{14}$: binary predictions for depth level where dark blue = 1 (predicted presence) and light blue = 0 (predicted absence). Sum of $\mathbf{d}_0$ to $\mathbf{d}_{23}$: sum of binary predictions across all 24 depths; darker shades of blue indicate higher level of model congruence across depths while lighter shades indicate lower levels. Refer to Table 1 to obtain depth intervals in meters associated with depth levels. Gray area: landmass and seabed; E: Eastern Hemisphere; W: Western Hemisphere; tick-marks (|): Prime Meridian and International Date Line.

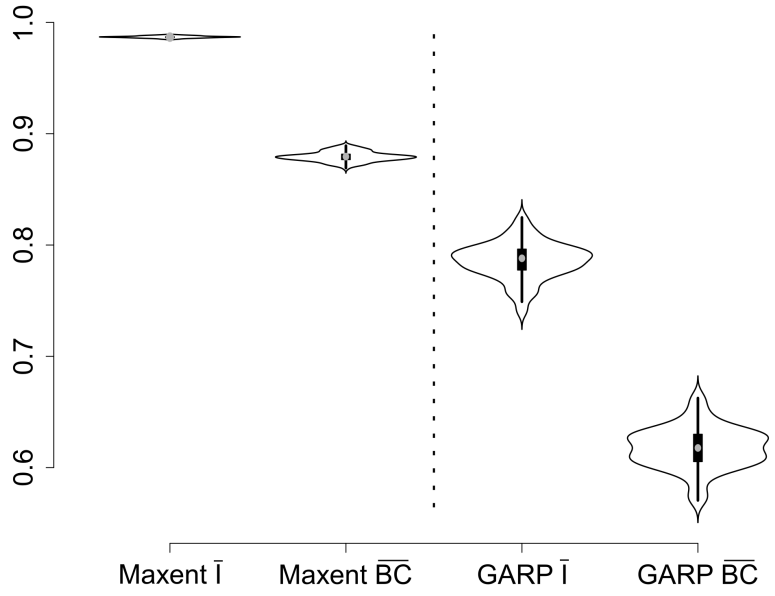


Fig. 21 Violin plots depicting results of occurrence data resampling experiments. The spread of pseudo-replicate versus full-dataset averages of *I* and *BC* values (black box-plots) is shown as well as the smoothed frequency of the distribution of averages (distributions on either side of box-plots) and median of the distribution of averages (gray dots). Both *I* and *BC* range from 0 (no niche overlap) to 1 (niches identical).

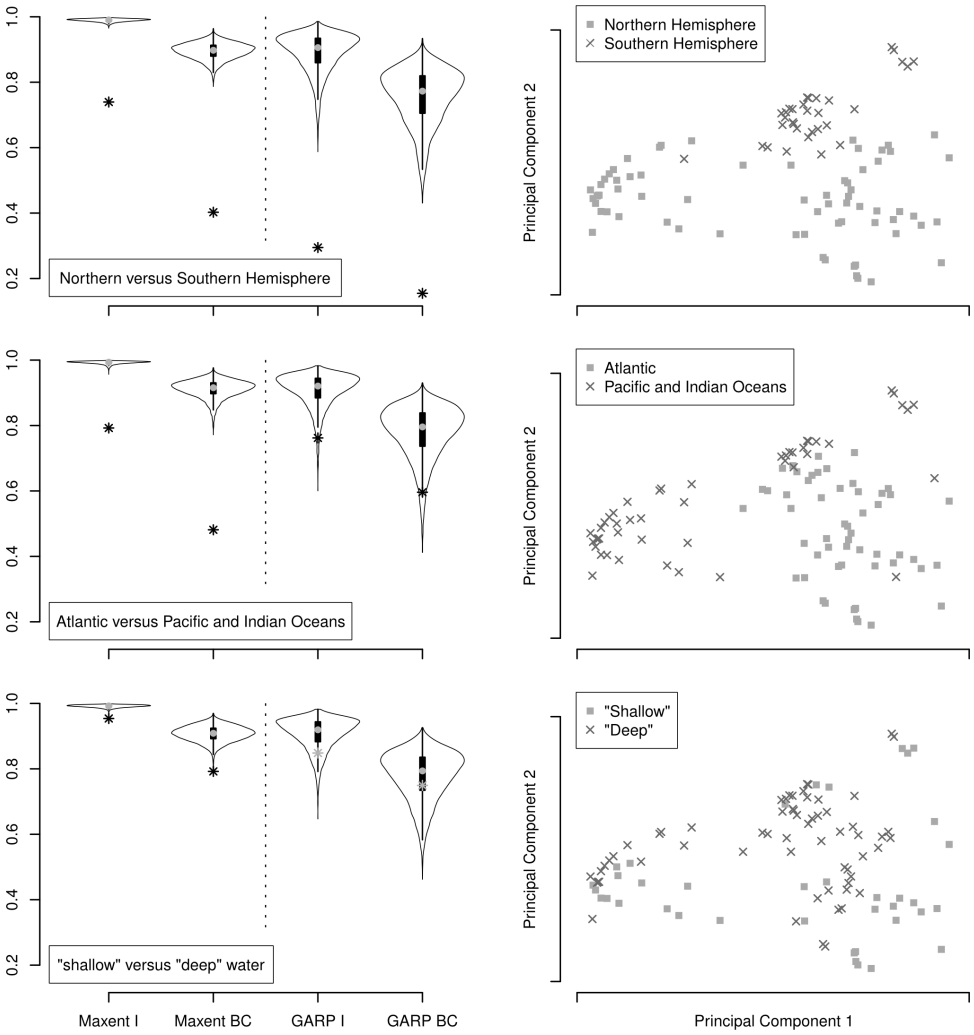


Fig. 22 Violin plots depicting the results of the niche identity test used to evaluate the similarity of model predictions when extrapolating from one geographic area into another using Maxent and GARP. The distribution (black box-plots) and smoothed frequency (distributions on either side of box-plots) of *I* and *BC* values, as well as the median (gray dots) of pseudo-replicate models are shown. Asterisks show *I* and *BC* values calculated for the subset models (e.g. Atlantic Ocean versus Pacific and Indian Oceans). White asterisks indicate $p < 0.05$; the null hypothesis tested was that the subset models are not identical based on a one-tailed test (cf., Bentlage & Shcheglovitova 2012). Both *I* and *BC* range from 0 (no niche overlap) to 1 (niches identical).

CHAPTER 5:

Marine biogeography: global panmixia in the open oceans

Abstract

Population genetic and phylogeographic studies have uncovered strong population structuring and previously unrecognized amounts of cryptic sibling species in many marine habitats. In the open oceans, most studies have investigated patterns of population structuring and species diversity for the uppermost parts of the water column. These patterns can be explained using present-day environmental discontinuities of water masses. The deeper parts of the open ocean water column (the mid-water) have been largely neglected due to the difficulties of sampling this environment. Here, we present a case-study integrating three-dimensional correlative ecological niche modeling with population genomics to investigate population genetics of a holopelagic hydrozoan jellyfish species, *Halicreas minimum* (Cnidaria: Hydrozoa: Trachylina). Our ecological niche models predicts broad ranges of suitable habitat across all oceans in the mid-water but not in shallow waters. Our population genomic results are consistent with these geographically explicit predictions and suggest that this species of mid-water jellyfish has a globally panmictic distribution. In particular, its distribution appears to be continuous across large areas of the mid-water and a large number of effective migrants per generation are exchanged between geographically distant demes. These findings suggest that the evolutionary dynamics of mid-water habitats differ from those of shallow open ocean waters.

Introduction

Many marine species have a high potential for long-range dispersal through larval stages that may drift in the water column for hundreds or thousands of kilometers before metamorphosing into adults (cf., Palumbi 1992). Despite this potential for wide-ranging dispersal, population genetic and phylogeographic studies have uncovered strong population genetic structuring and previously unappreciated amounts of richness in terms of unrecognized species or "evolutionary significant units" in marine environments (reviewed in Knowlton 1993; Heads 2005). These patterns have been widely documented for organisms inhabiting the continental shelves (the neritic zone) and are the likely the result of sea-level fluctuations, continental drift, the emergence of remote atolls and islands, and changing patterns of sea surface circulation have the potential to create deep, persisting population structuring which may eventually lead to speciation (e.g., Barber *et al.* 2000; Vogler *et al.* 2008; Abe & Lieberman 2009; Kool *et al.* 2011).

In open ocean communities by contrast, there appear to be few clear and persisting physical boundaries to dispersal that may limit gene flow (Angel 1993; Norris 2000). In addition, species inhabiting the open oceans are mostly adapted to a holopelagic life style. That is, the entire life cycle occurs in the water column, likely removing ontogenetic constraints on ranges of dispersal since both adults and juvenile stages may co-disperse. This suggests that the apparent paucity of physical barriers to dispersal, and lack of ontogenetic constraints, may be limiting population sub-structuring and ultimately diversification through speciation in the open oceans, compared to other marine habitats. In Hydrozoa (phylum Cnidaria), for example, the number of

species with a benthic stage in their life cycle outnumber holopelagic species by a factor of ten (Daly *et al.* 2007). In addition to reduced species richness, holopelagic hydrozoans appear to occupy larger distributional areas (Gibbons *et al.* 2010). Population genetic analyses of holopelagic hydrozoans have not been undertaken on a large scale to date (Collins *et al.* 2008; Gibbons *et al.* 2010). However, some open ocean inhabiting Foraminifera bear signatures of large-scale population cohesion consistent with the above notions, although paths of migration and timescales at which gene flow is or was occurring remain unclear and subject to speculation (e.g., Darling *et al.* 2000).

However, environmental discontinuities in open ocean water masses may lead to temporary or persistent population isolation and possibly allopatric speciation. Such patterns have, for example, been observed in copepods (e.g., Bucklin *et al.* 1996; Bucklin *et al.* 2000; Goetze 2003; Goetze 2005; Goetze & Ohman 2010) and foraminifers (e.g., de Vargas *et al.* 1999; Norris & de Vargas 2000; Darling *et al.* 2000; Darling & Wade 2008). In general, the distribution of plankton inhabiting the upper several hundred meters of the water column in the open oceans (the epipelagic and upper mesopelagic) are strongly correlated with large-scale environmental structuring of surface water masses (e.g., Fernández-Álamo & Färber-Lorda 2006; Longhurst 2007). Furthermore, many species of plankton have been shown to consist of several genetically distinct lineages or cryptic species (Bucklin *et al.* 2010), possibly as a result of environmental heterogeneity of surface water masses (cf., Goetze 2005).

Structuring of sea surface waters can be described using satellite imagery or well-defined sea surface circulation models. Using these data and information, hypotheses about plankton distributions and population structure can be generated and subsequently tested using molecular

genetic methods. For example, ocean gyres are often assumed to affect gene-flow among populations (e.g., Bucklin *et al.* 2000). Evaluating the distributions of organisms inhabiting the deep sea and deep parts of the water column (the mid-water) has proven more challenging (UNESCO 2009; Bucklin *et al.* 2010). Technical hurdles associated with sampling deep water habitats are being overcome through continuous technological advances, but explorations of deep-sea habitats are comparatively rare and geographically restricted (cf., Bucklin *et al.* 2010). Consequently, our knowledge of the mid-water remains patchy (UNESCO 2009; Bucklin *et al.* 2010). To address the issue of limited knowledge on the distributions of species, one may employ correlative ecological niche models that relate observations of species occurrence to environmental data to derive spatially explicit models of the species' niche (cf., Chapter 4). Niche models can be used to visualize deep sea habitats that are suitable for a positive growth rate of the populations of the species of interest (e.g., Tittensor *et al.* 2010; Owens *et al.* 2012; Chapter 4) and have the potential to allow for making explicit *a priori* hypotheses about population genetic or phylogeographic patterns of mid-water organisms based on the distribution of suitable habitat (i.e., based on the organism's potential distribution).

To evaluate this approach, and investigate the distributional patterns of the mid-water dwelling, holopelagic jellyfish *Halicreas minimum* (Cnidaria: Hydrozoa: Trachylina). *Halicreas minimum* is a holopelagic species that develops directly from larva to adult in the water column and has been described to occur in the Atlantic, Pacific, and Indian Oceans (Kramp 1959, Kramp 1965, Kramp 1968), suggesting that it may have the capability for long-distance dispersal. In order to extrapolate beyond the documented observations of occurrence of *Halicreas minimum*, we used a potential distribution model to make predictions about the species' expected, large-

scale geographic distribution. This distribution model was then used to make predictions about population structure. In order to assess these predictions, individuals of *H. minimum* were sampled from disparate geographic regions. Next-generation sequencing technologies were used for whole genome shotgun genotyping and subsequent population genomic analysis.

Materials and Methods

Ecological Niche Modeling. The potential distribution of *H. minimum* was inferred in three-dimensional space throughout the water column on a global scale (cf., Chapter 4) using two niche modeling algorithms, the Genetic Algorithm for Rule-Set Prediction (GARP; Stockwell & Peters 1999) and the maximum entropy machine learning algorithm implemented in Maxent (Phillips *et al.* 2004). Models were fitted using Desktop GARP (version 1.16) and Maxent (version 3.3.3) with default settings. 1° grids of objectively analyzed annual climatologies were obtained from the 2005 World Ocean Atlas (WOA05; Antonov *et al.* 2006; Garcia *et al.* 2006a; Garcia *et al.* 2006b; Locarini *et al.* 2006). Interpolated fields for temperature, salinity, oxygen concentration, and apparent oxygen utilization were used in addition to a potential productivity grid that was calculated as outlined in Chapter 4; all 33 depth layers included in the WOA05 were used. 101 spatially unique occurrences for *H. minimum* were obtained from the catalog of the National Museum of Natural History (Smithsonian Institution, Washington, DC, USA), the Automated Video Annotation and Reference System (http://www.mbari.org/vars/) of the Monterey Bay Aquarium Research Institute (Moss Landing, CA, USA), and the Ocean Biogeographic Information System's online database (http://www.iobis.org/). GARP and Maxent models were fitted using the whole world as training and projection area, for the same reasons

outlined for another mid-water jellyfish in chapter 4. That is, *Halicreas minimum* was likely to have reached all suitable habitats in relevant time-frames and is likely in distributional equilibrium on a global scale. To facilitate interpretation of 3D models, best subset GARP model and the Maxent model were transformed into binary predictions using a threshold that allows for 5% of the training data with the lowest prediction scores to be omitted from the binary predictions. Only those pixels that were predicted by both GARP and Maxent were retained for the final prediction, a "consensus model" of GARP and Maxent predictions.

Shotgun genotyping. Tissue samples from *H. minimum* specimens were collected in eastern Antarctica off Dumont D'Urville Station (2 specimens), the Charlie-Gibbs Fracture Zone in the North Atlantic (5 specimens), the Monterey Canyon off the coast of California (4 specimens), and off the coasts of Japan (5 specimens) during the years 2006-2010. DNA was extracted using phenol-chloroform extractions as described in Cartwright *et al.* (2008) or using the automated DNA isolation system AutoGenPrep 965 (AutoGen Inc., Holliston, MA, USA) following the manufacturer's protocol. Extracted genomic DNA was amplified using the illustra GenomiPhi chemistry (GE Healthcare, Little Chalfont, United Kingdom) following the manufacturer's protocol. Amplified DNA was purified using Agencourt AMPure beads (Beckman Coulter Genomics Inc., Brea, CA, USA). Genomic DNA libraries for whole genome shotgun sequencing were prepared with the TruSeq DNA sample prep kit rev. A following the manufacturers protocol (Illumina, San Diego, CA, USA); all 6 supplied barcodes were used to index libraries for multiplexing. Libraries were pooled up to a depth of 6 samples per well (multiplexing ranged from 4-6 samples per lane) and sequenced using Illumina's HiSeq2000 sequencing technology at the University of Kansas Medical Center Genome Sequencing Facility

(http://www.kumc.edu/genome-sequencing-facility.html) to obtain 100 bp long paired-end reads. After trimming bases with a Phred quality score < 20 using custom Python scripts, the reads of 3 specimens (1 from Antarctica, 1 from California, and 1 from Japan) were individually assembled *de novo* with the short read assembler Velvet (Zerbino & Birney 2008) using an Amazon EC2 instance with 68 GB of physical RAM and ~40 GB of swap distributed across multiple hard drives with equal swapping priorities. Assemblies were performed using using varying k-mer lengths. The resulting 3 sets of contigs derived from the raw reads were then assembled into longer scaffolds using CAP3 (Huang & Madan 1999); this assembly was used as a reference to which raw reads were mapped (see below).

In order to genotype each sequenced specimen the following pipeline was implemented using Python (http://www.python.org/) and Bash (http://www.gnu.org/software/ bash/) scripts in conjunction with the Genome Analysis Toolkit (GATK release 1.4; McKenna *et al.* 2010), SAMtools (version 1.18; Li *et al.* 2009), and the Picard command-line tools (version 1.63; http://sourceforge.net/ projects/picard/) (Fig. 23). Raw reads were mapped against the *de novo* assembled reference scaffolds using Bowtie2 (version 2 beta5; Langmead & Salzberg 2012). Mapped reads were locally re-aligned using the Smith-Waterman algorithm (Smith & Waterman 1981) implemented in GATK. The following series of steps were performed to avoid artifacts and low-quality variants in the final SNP datasets. The re-aligned mapped reads were used to identify potential single nucleotide polymorphisms (SNPs) and indels. Potential SNPs adjacent to these indels were masked since such SNPs tend to contain a high number of false positives according to the GATK documentation. The remaining potential SNPs were further filtered. All SNPs called in regions with more than 10% of reads having a mapping quality score of 0 were

excluded from further analysis. A sliding window of 20 bp size was used to mask clusters of 3 or more SNPs; false positive SNPs may be expected to form clusters due to systematic sequencing errors, and thus removed from further analysis. Further, all SNPs called from mapped regions with strand bias were excluded (strand bias > 0). Lastly, only those SNPs were retained for downstream analysis that could be ascertained for every specimen. In this final dataset several SNPs were located on the same reference scaffold. To reduce potential linkage effects on downstream analyses only one SNP per scaffold was selected at random using custom Bash scripts. Two datasets, A and B, were created in this manner, so that replicate analyses could be performed to evaluate the effect different data may have on the analysis .

Population genetic analysis: Tajima's *D* and *F*-statistics. The frequency of polymorphisms in all genotyped loci was evaluated using Tajima's *D*. *D* was calculated using a 200 bp sliding window in VCFtools (version 0.1.8a; Danecek *et al.* 2011). Departures from Hardy-Weinberg expectations (HWE) within populations were evaluated by calculating the inbreeding coefficient F_{is} using the observed frequency of heterozygotes (H_o) and frequency of expected heterozygotes (H_e) under HWE for both datasets A and B; $F_{is} = 1 - H_o/H_e$. H_o and H_e were calculated using the R (R Development Core Team 2010) package HIERFSTAT (version 0.04-6; Goudet 2005). 95% confidence intervals for F_{is} were estimated using the bootstrap procedure implemented in HIERFSTAT with 1000 replicates. Genetic differentiation among populations based on Weir & Cockerham's (1984) unbiased estimator of F_{st}, theta (Θ), as well as the 95% bootstrap confidence intervals (1000 replicates) for pairwise F_{st}s were calculated using HIERFSTAT.

Population genetic analysis: Model-based clustering. Population structure was inferred

using the model-based clustering methods implemented in Structure (version 2.3.3; Pritchards *et al.* 2000; Falush *et al.* 2003; Falush *et al.* 2007). In particular, *K*, the prior for the number of potential populations, was set to 4, equaling the number of large geographic sampling locations or demes (Antarctica, North Atlantic, California, and Japan). The admixture model was used for clustering, allowing each individual to have admixed ancestry in any of the *K* populations. The admixture parameter α was estimated for each population individually. Similarly, allele frequency λ was estimated individually for each *K*. Note that individuals were not assigned to populations for the clustering analysis; individuals were assigned to any of the *K* populations based on genotype information alone. Several runs were performed in Structure to assess when the Markov chain of the implemented MCMC algorithm had stabilized. Final analyses were performed with 50,000 samples that were discarded as burn in and 50,000 samples that were recorded.

Simulation. In order to assess the amount of migration among populations necessary to explain the observed overall estimator of F_{st}, theta (Θ) and clustering of individuals, simulation analyses were performed. Datasets for 29 effective migration rates ($4N_0m$), ranging from 0.5 to 14.5, were generated under a neutral model of evolution using the computer program ms (Hudson 2002). Simulated datasets had the same number of segregating sites and individuals per deme as the empirical datasets A and B. Overall theta was calculated for 1000 datasets for each migration rate. The arithmetic means and range containing 95% of theta for the simulated data for each migration rate were compared to theta calculated from datasets A and B.

For clustering analyses, 100 datasets were simulated for each migration rate. Simulated datasets were analyzed using the no-admixture model in Structure, which assigns each individual

to one of *K* populations (here *K* equaled 4) without allowing for admixed ancestry of individuals. Analogous to the simulated datasets, datasets A and B were analyzed under the no-admixture models using 100 independent runs in Structure for each dataset. The no-admixture model analyses were performed with 10,000 burn in samples that were discarded, followed by 10,000 samples that were recorded. Replicate analyses using clustering approaches may lead to major differences among results due to "label switching" of clusters across replicates or multi-modality of the data that may lead to truly different clustering solutions (Jacobsen & Rosenberg 2007). To account for these issues, clusters were matched for each series of replicate clustering analyses using the cluster matching program CLUMPP (version 1.1.2; Jacobsen & Rosen 2007); resulting clusters were plotted for visual comparisons.

Results

Potential distribution. The modeled potential distribution of *H. minimum* shows that the suitable range of habitats encompasses large, cohesive areas of the deep mid-water (Fig. 24). In particular, the potential distribution spanned the global oceans and seas in almost entirety in mid- and deep water between depths of 250m and 3000m, omitting Arctic waters and smaller areas close to Antarctica. In addition, the Mediterranean was not predicted. This is in contrast to the sea surface, where the potential distribution is predicted to have very narrow ranges in higher latitudes. In particular, in surface waters, parts of the Kuroshio and North Pacific Currents, and parts of the North Atlantic Current are contained in the prediction in the Northern Hemisphere. In the Southern Hemisphere, large areas of the subtropical front, the Benguela Current, and parts of the Humboldt Current are contained in the prediction of the Southern Hemisphere. In summary,

the potential distribution of *H. minimum* spans cohesive areas across the globe, with the largest areas of potential distribution being contained in the deep mid-water.

Population genetics. The final dataset contained 1263 loci that were genotyped for each individual. Tajima's *D* for 82% of these loci equaled 0, while ~1% of all SNPs discovered in *H. minimum*'s genome had a value of $2 < D < 2.35$ and ~1.7% of SNPs had a value of $-1.89 < D < -1.8$ (Fig .3). The final datasets (A and B) for analysis contained 1092 genotypes. Observed and expected heterozygosities roughly equaled each other in Antarctica, while all other demes appeared to harbor an excess in heterozygous individuals when compared to HWE (Fig. 26). Pairwse F_{st}s indicated little differentiation among demes (Table 6), with the highest differentiation observed between Pacific demes and the North Atlantic (Table 6). Inference of population structure using a Bayesian clustering approach demonstrated a lack of separation of individuals into distinct populations (Fig. 27). All individuals had varying amounts of mixed ancestry in the prior $K = 4$ populations, with most individuals having the majority of their genotypes likely originating in the same ancestral population.

Simulation. Simulation results suggested that migration rates much larger than 1 effective migrant per generation are needed to explain the observed genotype diversity among individuals (Fig. 28). Likewise, the lack of population structure observed for datasets A and B seems to require migration rates among sampling sites that are much larger than 1 (Fig. 29). At a migration rate of 0.5 all individuals are correctly assigned to each of the 4 populations. At a migration rate of 1.0 all but one individual are correctly assigned, but at migration rates larger than 1 and smaller than 2, the smallest population containing 2 individuals cannot be resolved correctly anymore. With increasing migration rates population assignment of individuals to

populations became increasingly spurious, until the majority or all individuals were more or less consistently assigned to only a single population.

Discussion

Similar to other mid-water species, *H. minimum* has a documented global distribution including all three oceans, but is absent from both the Mediterranean and the Arctic (Kramp 1959, Kramp 1965, Kramp 1968, Gili *et al.* 1992), consitent with our potential distribution model (Fig. 24). *H. minimum*'s depth distribution appears to focus on the mesopelagic and bathypelagic zones (Kramp 1959, Kramp 1965, Kramp 1968; Larson *et al.* 1991; Pugh *et al.* 1997; Vinogradov & Shushkin 2002), even though individuals have been collected from shallower, epipelagic (including the sea surface) and deeper, abyssopelagic waters (e.g., Kramp 1965; cataloged specimens, National Museum of Natural History, Smithsonian Institution, Washington, DC). Ecological niche modeling enabled us to make predictions beyond the snapshots gained from sampling expeditions. In this way we visualized the potential distribution (cf., Chapter 4) of *H. minimum* on a global scale across different horizons of the water column. In agreement with *in situ* observations, the potential distribution of *H. minimum* is expansive and cohesive in mid-water habitats throughout all oceans. The environmental discontinuities that may present barriers to the dispersal in shallow waters are apparently lacking in the mid- and deep-water. These predictions suggest that *H. minimum* may well be capable of maintaining population cohesion on a global scale.

Consistent with our geographic models, we were unable to find any signatures of population structure among four distant geographic areas in the Atlantic and Pacific Oceans

using more than one thousand loci from each individual's genome. Observed heterozygotes were either in agreement with HWE, suggesting neither in- nor outbreeding, or in excess, indicating the possibility of outbreeding. In agreement with this finding, pairwise F_{st}s indicated little differentiation among demes despite large geographic distances between them. Analogous to the results obtained using *F*-statistics, Bayesian clustering approaches did not uncover any appreciable amount of population structure. The lack of clustering of individuals into genetically distinguishable populations (Fig. 27) suggests that none of the sampled demes are acting as distinct populations or "evolutionary significant units", that would suggest the potential for cryptic, unrecognized speciation as observed in other zooplankton (e.g., Goetze 2003). In contrast, the lack of structure observed suggests panmixia among demes. Simulations further supported this view (Figs. 6 & 7). In particular, migration rates far greater than 1 effective migrant per generation are needed to explain the observed lack of population structure.

The geographic scale at which panmixia appears to be operating in the mid-water holopelagic species investigated here may be surprising. The observed lack of population sub-structuring is based on relatively few samples due to the difficulties in sampling the remote depths that *H. minimum* inhabits in the open oceans. The apparent lack of individuals for population genetic analysis was addressed, however, by the large amount of genotypic data obtained from each individual's genome. It seems unlikely that our findings are largely the result of selection; the limited number of individuals used here may have inflated or deflated *D* artificially, leading to the few extreme values observed (Fig. 25). Importantly, our population genomic results are consistent with the results of geographically explicit models of *H. minimum*'s niche. In combination, these two lines of evidence suggest that *H. minimum* has the ability to

maintain population cohesion and global panmixia by inhabiting a large area of contiguous suitable habitat which facilitates gene flow among the demes inhabited by this holopelagic species.

Mid-water connectivity through space and time. The present-day thermohaline circulation, driving the Ocean Conveyor Belt, offers a means for the dispersal of mid-water organisms, and thus *H. minimum*, among ocean basins. In particular, the Antarctic Circumpolar Current allows for a free exchange of waters among oceans at all depths (e.g., Comiso 2010). The narrow Farm and Berring Straits connecting the Arctic to the Atlantic and Pacific Oceans, on the other hand, appear to represent effective barriers to present-day dispersal of plankton in the Northern Hemisphere (e.g., Darling *et al.* 2007). Many mid-water inhabiting jellyfish can indeed be found in all three oceans (e.g., Gili *et al.* 1998) and it seems likely that the distributional pattern of panmixia discovered for *H. minimum* is applicable to other mid-water jellyfish. For example, Collins *et al.* (2008) found no genetic differentiation among two bathypelagic species sampled from opposite ends of the Pacifc basin, while two epi- or upper meso-pelagic species sampled from distant geographic regions may likely harbor previously unrecognized cryptic species.

Connectivity among distant demes may be explained using patterns of present-day ocean circulation, but in the past, changes in global climate periodically slowed or stopped the thermohaline circulation, leading to reorganizations of large-scale circulation patterns (e.g., Broecker 1997). Similar slow-downs in the thermohaline circulation may occur in the future or may have already begun due to anthropogenic climate change (e.g., Bryden *et al.* 2005; Schiermeier 2006), potentially affecting future habitat connectivity in the mid-water. Further

population genetic studies need to critically examine how wide-spread the pattern of global panmixia is among mid-water species. The results presented herein, the preliminary data based on limited sampling in Collins *et al.* (2008), and distribution modeling of another mid-water jellyfish (Chapter 4), in conjunction with present-day ocean circulation patterns, suggest that mid-water jellyfish, and other zooplamkton, have the potential for present-day panmixia on a global scale.

Past reorganizations in circulation patterns may have had the potential to isolate populations in the mid-water, leading to speciation. This may have been followed by dispersal when circulation patterns changed again, analogous to geodispersal in shelf habitats (cf., Lieberman & Eldredge 1996). Such a scenario could explain the probable lack of present-day barriers to gene-flow in conjunction with the observed large-scale co-distributions of species in the mid-water (e.g., open ocean hydrozoans appear largely co-distributed: Kramp 1959; Kramp 1965; Kramp 1968); diversification through the rapid evolution of mate recognition systems may represent an alternative scenario capable of explaining the radiation of mid-water plankton (cf., Palumbi 2009). Lastly, multiple invasions of the mid-water from shallow open ocean habitats (e.g., the epipelagic) may have led to the colonization of the mid-water rather than *in situ* diversification, similar to patterns described for organisms inhabiting the benthos (Jablonski *et al.* 1983; Jablonski & Bottjer 1990; Barbeitos *et al.* 2010; but see Lindner *et al.* 2008). Whatever the process, the pattern of global panmixia observed in the case study presented here is in stark contrast with the patterns of strong population sub-structuring and rampant discovery of cryptic sibling species in shallow open ocean and shelf habitats.

Table 6 Pairwise F_{st}s calclated using Weir & Cockerham's (1984) estimator. Middle panel contains observed F_{st}s (obs. F_{st}). The upper and lower panels contain the lower (2.5 percentile) and upper (97.5 percentile) boundaries of the 95% confidence interval estimated using 1000 bootstrap replicates as implemented in HIERFSTAT. Results for dataset A are shown in black; results for dataset B are shown in grey. Negative estimates were set to zero.

2.5%	Antarctica	N Atlantic	California	Japan
Antarctica		0.049	0	0
N Atlantic	0.046		0.022	0.019
California	0	0.023		0
Japan	0	0.002	0	
obs. *F*st	Antarctica	N Atlantic	California	Japan
Antarctica		0.061	0.012	0.012
N Atlantic	0.059		0.032	0.027
California	0.007	0.032		0.005
Japan	0.006	0.027	0.003	
97.5%	Antarctica	N Atlantic	California	Japan
Antarctica		0.075	0.029	0.026
N Atlantic	0.071		0.042	0.036
California	0.024	0.041		0.014
Japan	0.019	0.035	0.012	

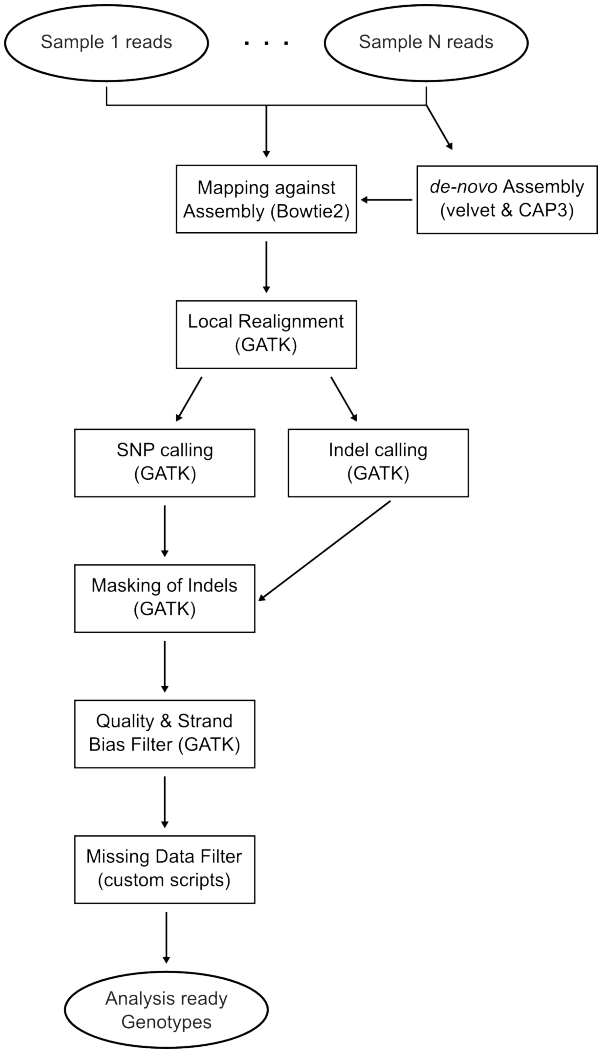


Fig. 23 Strategy employed for shotgun genotyping. Genomic DNA from samples were sequenced using the Illumina (San Diego, CA) HiSeq 2000 sequencing system. A subset of 3 samples was used to assemble reads *de novo*. All sample reads were mapped against the assembly. Mapped reads were re-aligned using the Smith-Waterman algorithm (Smith & Waterman 1981). Both single nucleotide polymorphisms (SNPs) and indels were called from the re-aligned, mapped reads. Indels were used to mask SNPs that occurred adjacent to indels. Subsequently, low quality SNPs and SNPs called with strand bias were filtered out. Lastly, SNPs that could not be ascertained for every sample were pruned from the final dataset.

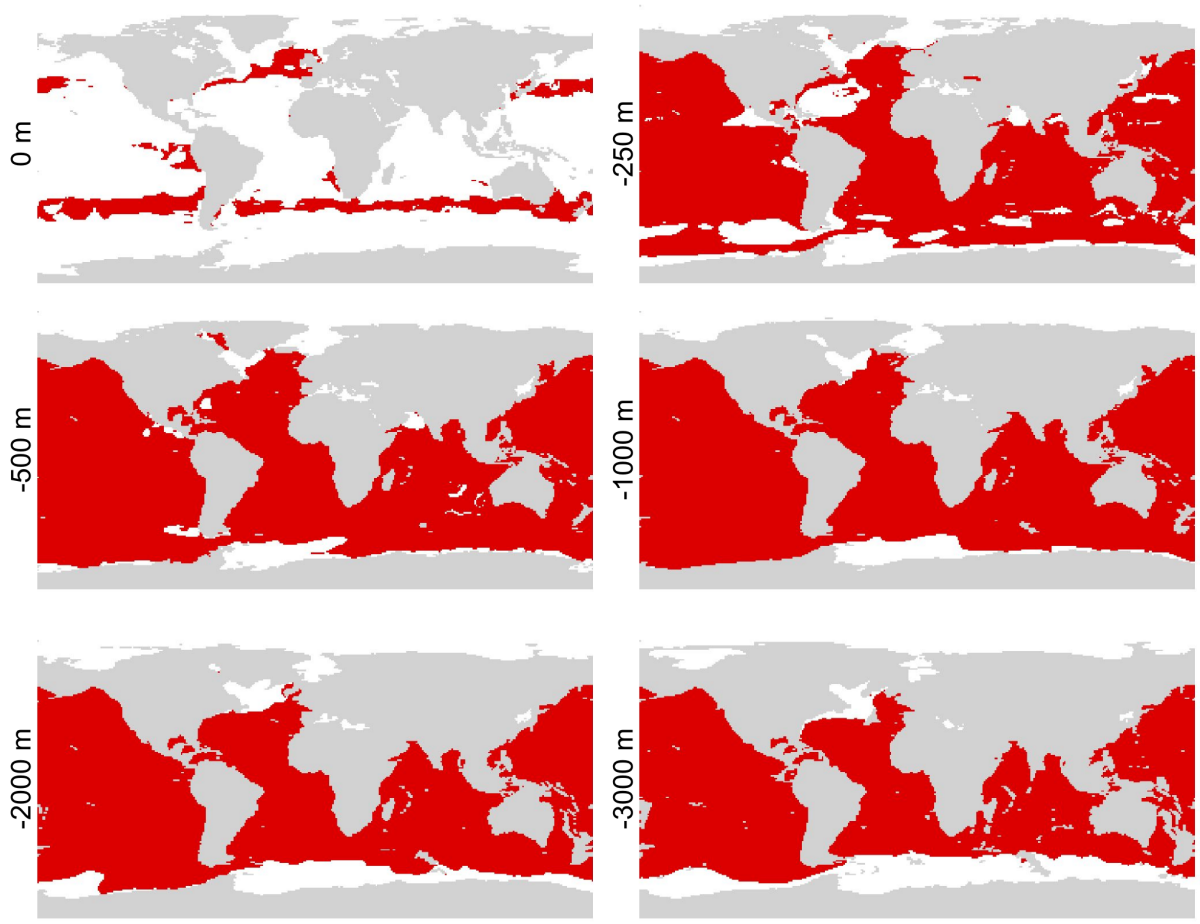


Fig. 24 Potential distribution of *Halicreas minimum*, inferred using ecological niche modeling. Pixels in red represent pixels predicted present by both GARP and Maxent. Gray areas correspond to land masses and the ocean floor. Red corresponds to areas of predicted suitable habitat whereas white are areas are not predicted as suitable environments for *Halicreas minimum*.

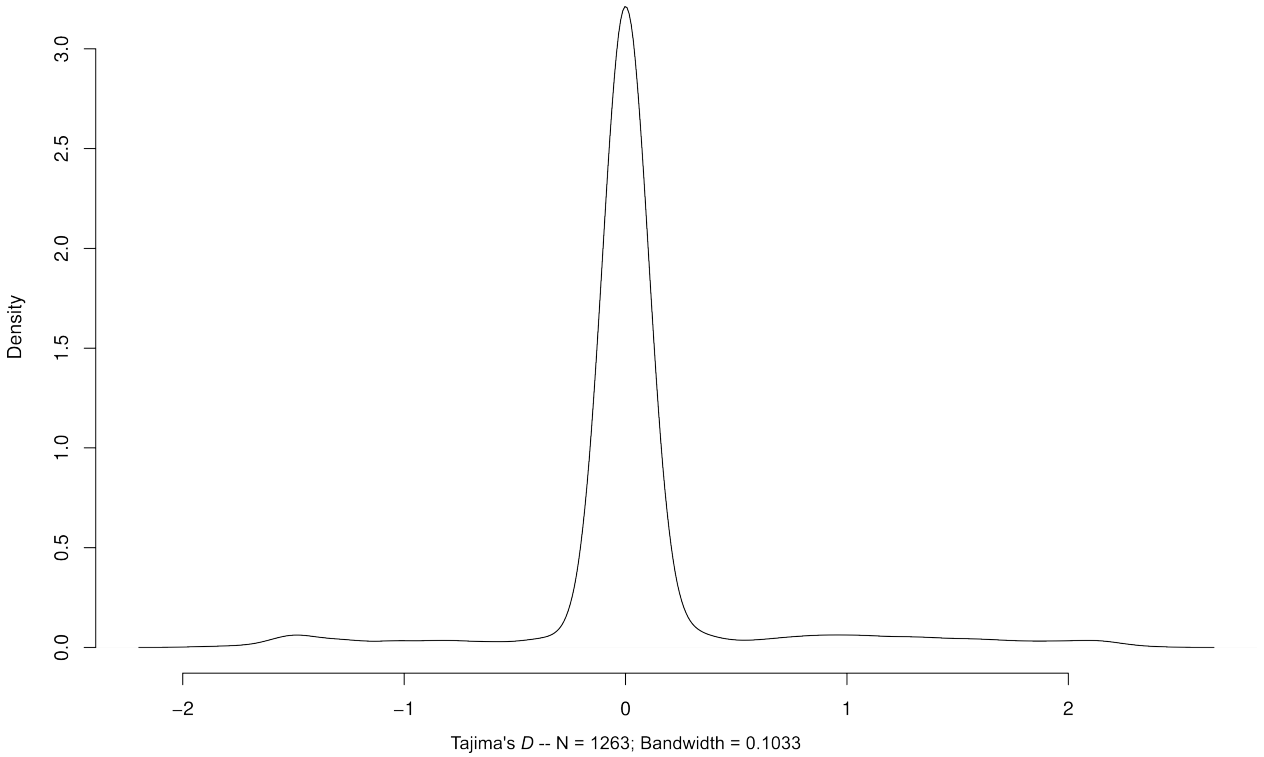


Fig. 25 Density distribution of Tajima's *D* for 1263 loci identified from the *Halicreas minimum* genome.

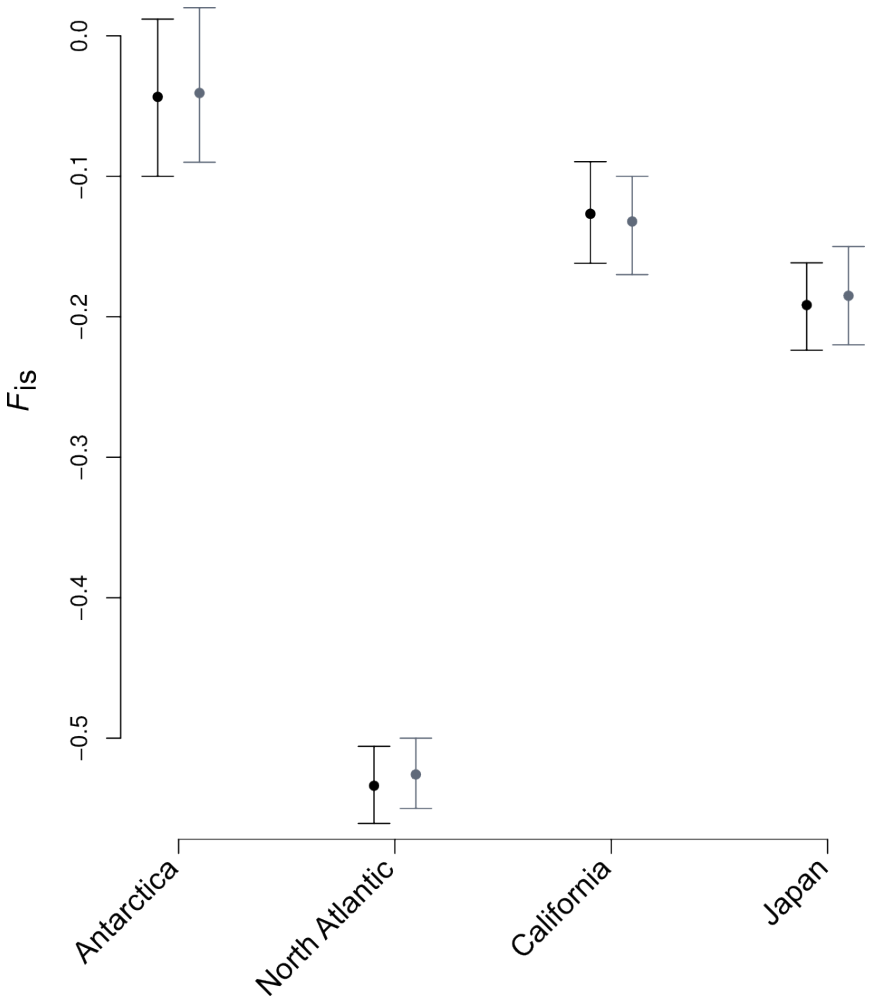


Fig. 26 F_{is} for each sampled population (dots) and the 95% confidence intervals obtained from 1000 bootstrap replicates in HIERFSTAT (Goudet 2005). Dataset A in black; dataset B in grey.

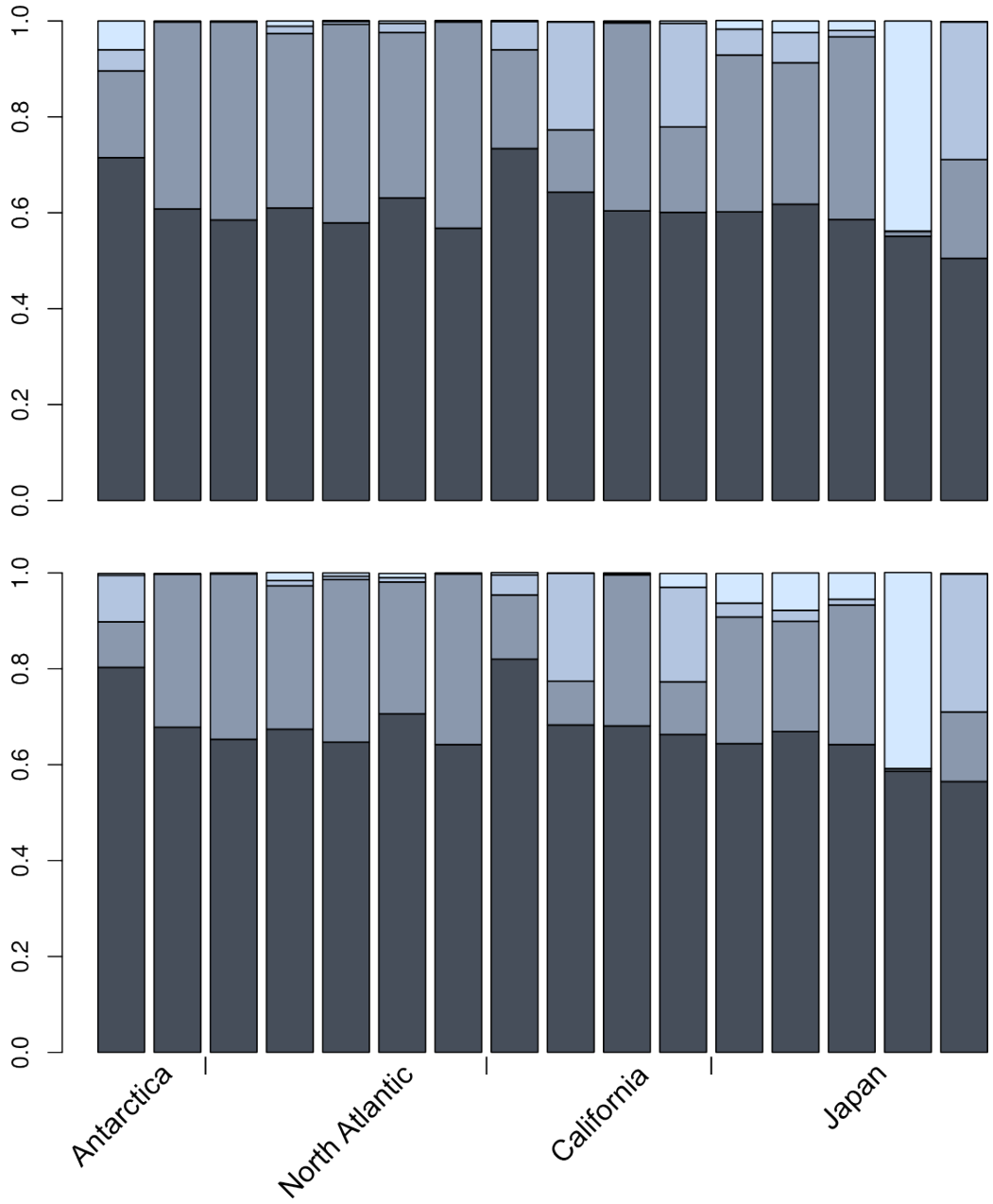


Fig. 27 Inference of population structure under an admixture model that allows individuals to have mixed ancestries (Pritchard *et al.* 2000; Falush *et al.* 2003). In this analysis, each individual's genotypes were allowed to be drawn from *K*=4 populations. Barplots show the percentage of each of 16 individual's genotypes inferred to belong to each of the *K* populations.

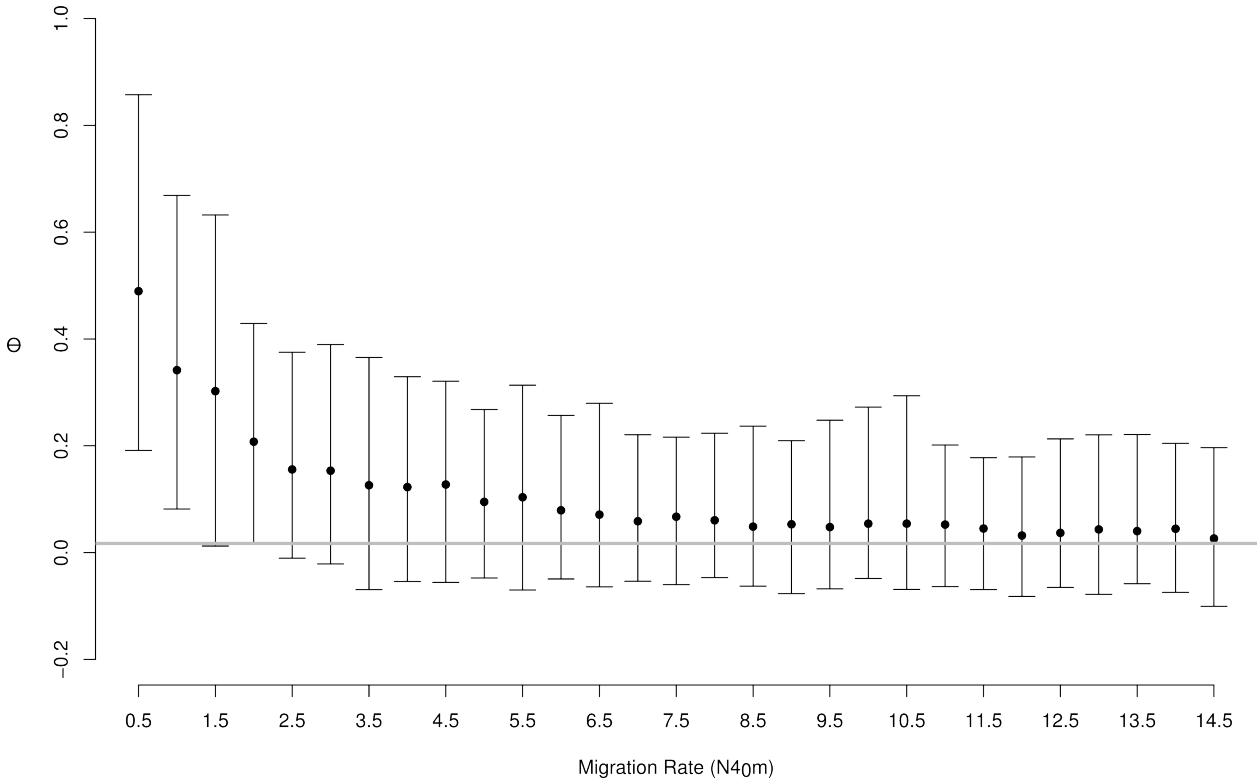


Fig. 28 Mean theta (dots; intervals show the range containg 95% of theta estimates from simulated data) versus migration rates. Theta was estimated from simulated datasets for a range of migration rates. Gray horizontal line indicates the range of theta observed from datasets A and B.

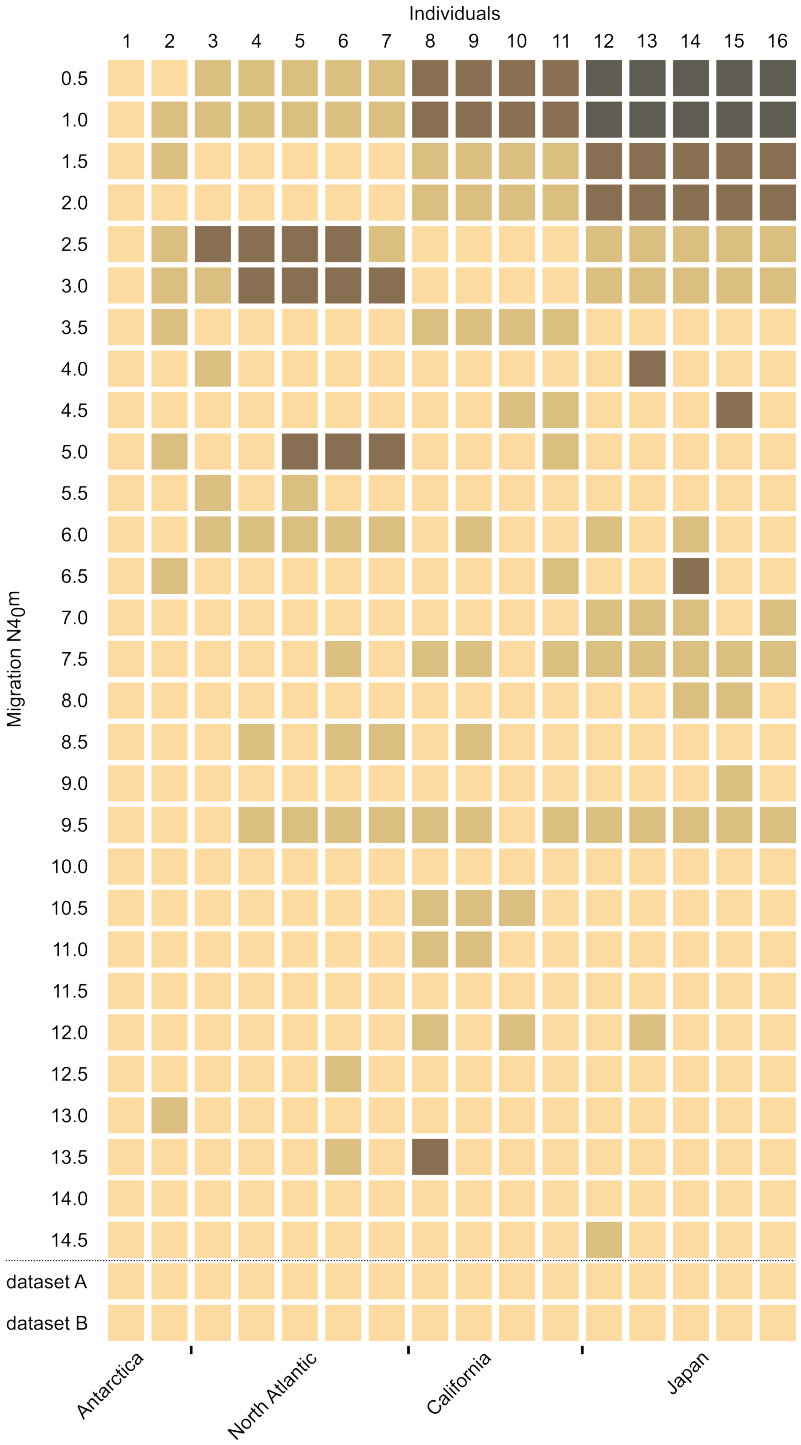


Fig. 29 Assignment of individuals to one of *K*=4 populations (color-coded squares). Datasets were simulated under different migration rates. In addition, clustering results for datasets A and B are shown for comparison with simulated data.

REFERENCES

Abe FR & Lieberman BS (2009) The nature of evolutionary radiations: A case study involving devonian trilobites. *Evolutionary Biology* 36: 225-234.

Anderson RP, Gómez-Laverde M & Peterson AT (2002) Geographical distributions of spiny pocket mice in South America: insights from predictive models. *Global Ecology Biogeography* 11:131-141.

Anderson RP, Lew D & Peterson AT (2003) Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecological Modelling* 162:211-232.

Anderson RP & Raza A (2010) The effect of extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: preliminary tests with montane rodents (genus *Nephelomys*) in Venezuela. *Journal of Biogeography* 37: 1378-1393.

Angel MV (1993) Biodiversity of the pelagic ocean. *Conservation Biology* 7: 760-772.

Antonov JI, Locarnini RA, Boyer TP, Mishonov AV & Garcia HE (2006) World Ocean Atlas 2005, Volume 2: Salinity. In: *NOAA Atlas NESDIS 62* (ed. S Levitus), pp. 182. US Government Printing Office, Washington, DC.

Araújo MB & Peterson AT (in press) Uses and misuses of bioclimatic envelope modelling. *Ecology*. DOI: dx.doi.org/10.1890/11-1930.1

Arneson, AC (1976) Life history of *Carybdea alata* Reynaud, 1830 (Cubomedusae). M.S. thesis, University of Puerto Rico, Mayaguez, Commonwealth of Puerto Rico, USA.

Arneson AC & Cutress CE (1976) Life history of *Carybdea alata* Reynaud, 1830 (Cubomedusae). In: *Coelenterate ecology and behavior* (ed. GO Mackie), pp. 227-236.

New York, NY: Plenum Press.

Bahn V & McGill BJ (2007) Can niche-based distribution models outperform spatial interpolation? Global Ecology and Biogeography 16:733-742.

Bailey PM (2003) Fatal envenomation by jellyfish causing Irukandji syndrome. *The Medical Journal of Australia* 178: 139-140.

Bailey PM, Little M, Jelinek GA & Wilce JA (2003) Jellyfish envenoming syndromes: unknown toxic mechanisms and unproven therapies. *The Medical Journal of Australia* 178: 34-37.

Bailey PM, Bakker AJ, Seymour JE & Wilee JA (2005) A functional comparison of the venom of three Australian jellyfish – *Chironex fleckeri*, *Chiropsalmus* sp., and *Carybdea xaymacana* – on cytosolic Ca2þ, haemolysis and *Artemia* sp. lethality. *Toxicon* 45:233-242.

Barber PH, Palumbi SR, Erdmann MV & Moosa MK (2000) Biogeography: A marine Wallace's line? *Nature* 406: 692-693.

Barbeitos M, Romano SL & Lasker HR (2010) Repeated loss of coloniality and symbiosis in scleractinian corals. *Proceedings of the National Academy of the USA* 107: 11877-11882.

Barker FK & Lutzoni FM (2002) The utility of the incongruence length difference test. Systematic Biology 51: 625-637.

Barnes JH (1964) Cause and effect in Irukandji stingings. *The Medical Journal of Australia*. 178: 34-37.

Barve N, Barve V, Jiménez-Valverde A, Lira-Noriega A, Maher SP, Peterson AT, Soberón J & Villalobos F (2011) The crucial role of the accessible area in ecological niche modeling

and species distribution modeling. *Ecological Modelling* 222: 1810-1819.

Bengtson K, Nichols MM, Schnadig V & Ellis MD (1991) Sudden death in a child following jellyfish envenomation by *Chiropsalmus quadrumanus*. Case report and autopsy findings. The Journal of the American Medical Association 266: 1404-1406.

Bentlage B (2010) *Carybdea alata* auct. (Cubozoa): rediscovery of the *Alatina grandis* type. *Zootaxa* 2713: 52-54.

Bentlage B, Cartwright P, Yanagihara AA, Lewis C, Richards GS & Collins AG (2010) Evolution of box jellyfish (Cnidaria: Cubozoa), a group of highly toxic invertebrates. *Proceedings of the Royal Society B* 277: 493-501.

Bentlage B, Peterson AT & Cartwright P (2009) Inferring distributions of chirodropid box-jellyfishes (Cnidaria: Cubozoa) in geographic and ecological space using ecological niche modeling. Marine Ecology Progress Series 384: 121-133.

Bentlage B & Shcheglovitova M (2012) NichePy: modular tools for estimating the similarity of ecological niche and species distribution models. *Methods in Ecology and Evolution* online early. DOI: 10.1111/j.2041-210X.2011.00184.x

Boecker WS (1997) Thermohaline circulation, the Achilles Heel of our climate system: Will man-made CO_2 upset the current balance? *Science* 278: 1582-1588.

Brinkman DL & Burnell JN (2009) Biochemical and molecular characterisation of cubozoan protein toxins. *Toxicon* 54: 1162-1173.

Bryan TL & Metaxas A (2007) Predicting suitable habitat for deep-water gorgonian corals on the Atlantic and Pacific continental margins of North America. *Marine Ecology Progress Series* 330: 113-126.

Bryden HL, Longworth HR & Cunningham SA (2005) Slowing of the Atlantic meridional overturning circulation at 25°N. *Nature* 438: 655-657.

Bucklin A, LaJeunesse TC, Curry E, Wallinga J & Garrison K (1996) Molecular genetic diversity of the copepod, *Nannocalanus minor*: genetic evidence of species and population structure in the North Atlantic Ocean. *Journal of Marine Research* 54: 285-310.

Bucklin A, Astthorsson, Gislason A, Allen LD, Smolenack SB & Wiebe PH (2000) Population genetic variation of *Calanus finmarchius* in Icelandic waters: preliminary evidence of genetic differences between Atlantic and Arctic populations. *ICES Journal of Marine Science* 57: 1592-1604.

Bucklin A, Nishida S, Schnak-Schiel S, Wiebe PH, Lindsay D, Machida RJ & Copley NJ (2010) A census of zooplankton of the global open. In: *Life in the World's Oceans: Diversity, Distribution, and Abundance* (ed Alstair McIntyre) pp. 247-265. Hoboken, NJ, USA: Wiley-Blackwell.

Carson HL & Clague DA (1995) Geology and biogeography of the Hawaiian Islands. In: *Hawaiian biogeography – evolution on a hot spot archipelago* (eds WL Warren & VA Funk), pp. 14-29. Washington, DC: Smithsonian Institution Press.

Cartwright P & Collins AG (2007) Fossils and phylogenies: integrating multiple lines of evidence to investigate the origin of early major metazoan lineages. Integrative and Comparative Biology 47: 744-751.

Cartwright P, Halgedahl SL, Hendricks JR, Jarrard RD, Marques AC, Collins AG & Lieberman BS (2007) Exceptionally preserved jellyfishes from the middle Cambrian. *PLoS ONE* 2: e1121.

Cartwright P, Evans NM, Dunn CW, Marques AC, Miglietta MP, Schuchert P & Collins AG (2008) Phylogenetics of Hydroidolina (Cnidaria: Hydrozoa). *Journal of the Marine Biological Association of the UK* 88: 1663-1672.

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolutio*n 17: 540-552.

Chung JL, Ratnapala LA, Cooke IM & Yanagihara AA (2001) Partial purification and characterization of a hemolysin (CAH1) from Hawaiian box jellyfish (*Carybdea alata*) venom. *Toxicon* 39: 981-990.

Cleland JB & Southcott RV (1965) Injuries to man from marine invertebrates in the Australian region. *National Health and Research Council, Special Report Series 12*, 282 p., Canberra, Australia.

Clement M, Posada D & Crandall K (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology* 9: 1657-1660.

Coates MM (2003) Visual ecology and functional morphology of Cubozoa (Cnidaria). *Integrative and Comparative Biology* 43: 542-548.

Coles SL, DeFelice RC, Eldredge LG & Carlton JT (1999) Historical and recent introductions of nonindigenous marine species into Pearl Harbor, Oahu, Hawaiian Islands. *Marine Biology* 135: 147-158.

Collins AG (2002) Phylogeny of Medusozoa and the evolution of cnidarian life cycles. *Journal of Evolutionary Biology* 15: 418-432.

Collins AG, Schuchert P, Marques AC, Jankowski T, Medina M & Schierwater B (2006) Medusozoan phylogeny and character evolution clarified by new large and small subunit

rDNA data and an assessment of the utility of phylogenetic mixture models. *Systematic Biology* 55: 97-115.

Collins AG, Bentlage B, Lindner A, Lindsay D, Haddock SHD, Jarms G, Norenburg JL, Jankowski T & Cartwright P (2008) Phylogenetics of Trachylina (Cnidaria: Hydrozoa) with new insights on the evolution of some problematical taxa. *Journal of the Marine Biological Association of the UK* 88: 1673-1685.

Collins AG (2009) Recent insights into cnidarian phylogeny. *Smithsonian Contributions to the Marine Sciences* 38: 140-149.

Collins AG, Bentlage B, Gillan W, Lynn TH, Morandini AC & Marques AC (2011) Naming the Bonaire banded box jelly, *Tamoya ohboya*, n. sp. (Cnidaria: Cubozoa: Carybdeida: Tamoyidae). Zootaxa 2753: 53-68.

Comiso JC (2008) *Polar Oceans from Space.* 507pp. Springer, NY, USA.

Conant FS (1898) The Cubomedusae. *Memoirs from the Biological Laboratory of the Johns Hopkins University* 4: 1-61.

Cunningham, C. W. 1997 Can three incongruence tests predict when data should be combined? *Molecular Biology and Evolution* 14: 733-740.

Currie BJ (2003) Marine antivenoms. *Jo*urnal of *Toxicol*ogy - *Clinic*al *Toxicol*ogy 41: 301-308.

Daly M, Brugler MR, Cartwright P, Collins AG, Dawson MN, France CS, McFadden CS, Opresko DM, Rodriguez E, Romano S & Stake J (2007) The phylum Cnidaria: a review of phylogenetic patterns and diversity three hundred years after Linnaeus. *Zootaxa* 1668: 127-182.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G,

Marth GT, Sherry ST, McVean G, Durbin R & 1000 Genomes Project Analysis Group (2011) The variant call format and VCFtools. *Bioinformatics* 27: 2156-2158.

Darling KF, Wade CM, Stewart IA, Kroon D, Dingle R & Brown AJL (2000) Molecular evidence for mixing of Arctic and Antarctic subpolar populations of planktonic foraminifers. *Nature* 405: 43-47.

Darling KF, Kucera M & Wade CM (2007) Global molecular phylogeography reveals persistent Arctic circumpolar isolation in a marine planktonic protist. *Proceedings of the National Academy of the USA* 104: 5002-5007.

Darling KF & Wade (2008) The genetic diversity of planktic foraminifera and the global distribution of ribosomal RNA genotypes. *Marine Micropaleotology* 67: 216-238.

Dawson MN, Raskoff KA & Jakobs DA (1998) Field preservation of marine invertebrate tissue for DNA analyses. *Molecular Marine Biology and Biotechnology* 7: 145-152.

Dawydoff C (1936) Observations sur la faune pelagique des eaux indochinoises de la mer de Chine meridionale (note preliminaire). *Bulletin de la Société Zoologique de France* 61: 461-484.

de Vargas C, Norris RD, Zaninetti L, Gibb SW, Pawlowski J (1999) Molecular evidence of cryptic specieation in planktonic formainifers and their relation to oceanic provinces. *Proceeding of the National Academey of the USA* 96: 2864-2868.

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

Elith J, Graham CH, Anderson RP, Dudík M et al. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129-151

Elith J & Leathwick J (2009) Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics* 40: 677-697.

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

Falush D, Stephens D & Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes*.

Farris JS, Kallersjö M, Kluge AG & Bult C (1995a) Constructing a significance test for incongruence. *Systematic Biology* 44: 570-572.

Farris JS, Kallersjö M, Kluge AG & Bult C (1995b) Testing significance of incongruence. *Cladistics* 10: 315-319.

Fenner PJ (1997) The global problem of cnidarian (jellyfish) stinging. M.D. thesis, University of London, London, UK.

Fenner PJ (1998) Dangers in the ocean: the traveller and marine envenomation. Journal of Travel Medicine 5: 135-141.

Fenner PJ (2006) Jellyfish responsible for Irukandji syndrome. *QJM: An International Journal of Medicine* 99: 802-803.

Fenner PJ & Hadok JC (2002) Fatal envenomation by jellyfish causing Irukandji syndrome. *The Medical Journal of Australia* 177: 362-363.

Fenner PJ & Williamson JA (1996) Worldwide deaths and severe envenomation from jellyfish stings. *The Medical Journal of Australia* 165: 658.

Fenner PJ, Fitzpatrick PF, Hartwick RJ & Skinner R (1985) 'Morbakka', another

cubomedusan. *The Medical Journal of Australia* 143: 550-555.

Fenner PJ, Williamson JA & Burnett JW (1996) Clinical aspects of envenomation by marine animals (Abstract). *Toxicon* 34: 145 (full article available from www.marine-medic.com.au/pages/articles/pdf/parisArticle.pdf).

Fernández-Álamo MA & Färber-Lorda J (2006) Zooplankton and the oceanography of the eastern tropical Pacific: A review. *Progress in Oceanography* 69: 318-359.

Fernando M (1992) Some hazards of diving. *Ceylon Medical Journal* 37: 72-80

Fischer AB & Hofmann DK (2004) Budding, bud morphogenesis, and regeneration in *Carybdea marsupialis* Linnaeus, 1758 (Cnidaria: Cubozoa). In: *Developments in hydrobiology – coelenterate biology 2003* (eds DG Fautin, JA Westfall, P Cartwright, M. Daly & CR Wyttenbach), pp. 331-337. Dordrecht, The Netherlands: Springer.

Gadagkar SR, Rosenberg MS & Kumar S (2005) Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. Journal of Experimental Zoology 304B: 64-74.

Garcia HE, Locarnini RA, Boyer TP & Antonov JI (2006a) World Ocean Atlas 2005, Volume 3: Dissolved Oxygen, Apparent Oxygen Utilization, and Oxygen Saturation. *NOAA Atlas NESDIS 63* (ed S Levitus), pp. 342. US Government Printing Office, Washington, DC.

Garcia, H.E., Locarnini, R.A., Boyer, T.P. & Antonov, J.I. (2006b) World Ocean Atlas 2005, Volume 4: Nutrients (phosphate, nitrate, silicate). *NOAA Atlas NESDIS 64* (ed S Levitus), pp. 396. US Government Printing Office, Washington, DC.

Gili JM, Bouillion J, Pagès F, Palanques A, Puig P & Heussner S (1998) Origin and biogeography of the deep-water Mediterranean Hydromedusae including the description

of two new species collected in submarine canyons of Northwestern Mediterranean. *Scientia Marina* 62: 113-134.

Gershwin L (2005a) *Carybdea alata* auct. and *Manokia stiasnyi*, reclassification to a new family with description of a new genus and two new species. *Memoirs of the Queensland Museum – Nature* 52: 501-523.

Gershwin L (2005b) Two new species of jellyfishes (Cnidaria: Cubozoa: Carybdeida) from tropical Western Australia, presumed to cause Irukandji Syndrome. *Zootaxa* 1084: 1-30.

Gershwin L (2006a) Jellyfish responsible for Irukandji syndrome. *QJM: An Internatinal Journal of Medicine* 99: 801-802.

Gershwin L (2006b) Nematocysts of the Cubozoa. *Zootaxa* 1232: 1-57.

Gershwin L (2006c) Comments on *Chiropsalmus* (Cnidaria: Cubozoa: Chirodropida): a preliminary revision of the Chiropsalmidae, with descriptions of two new genera and two new species. *Zootaxa* 1231: 1-42.

Gershwin L (2007) *Malo kingi*: a new species of Irukandji jellyfish (Cnidaria: Cubozoa: Carybdeida), possibly lethal to humans, from Queensland, Australia. *Zootaxa* 1659: 55-68.

Gershwin L (2008) *Morbakka fenneri*, a new genus and species of Irukandji jellyfish. *Memoirs of the Queensland Museum – Nature* 54: 23-33.

Gershwin L & Alderslade P (2005) A new genus and species of box jellyfish (Cubozoa: Carybdeidae [*sic*]) from tropical Australian waters. *The Beagle, Records of the Museum and Art Galleries of the Northern Territory* 21: 27-36.

Gibbons MJ, Janson LA, Ismail A & Samaii T (2010) Life cycle strategy, species richnes and

dsitribution in marine Hydrozoa (Cnidaria: Medusozoa). *Journal of Biogeography* 37: 441-448.

Godsoe W (2010) I can't define the niche but I know it when I see it: a formal link between statistical theory and the ecological niche. *Oikos* 119: 53-60.

Goetze, E. (2003) Cryptic speciation on the high seas; global phylogenetics of the copepod family Eucalanidae. Proceedings of the Royal Society of London B 270: 2321-2331.

Goetze E (2005) Global population genetic structure and biogeography of the oceanic copepods *Eucalanus hyalinus* and *E. spinifer*. *Evolution* 59: 2378-2398.

Goetze E & Ohman MD (2010) Integrated molecular and morphological biogeography of the calanoid copepod family Eucalanidae. *Deep-Sea research II* 57: 2110-2129.

Goudet F (2005) HIERFSTAT, a package for R to compute and test hierarchical *F*-statistics. *Molecular Ecology Notes* 5: 184-186.

Graham CH, Ron SR, Santos JC, Schneider CJ & Moritz C (2004) Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. *Evolution* 58: 1781-1793

Guinotte JM, Bartley JD, Iqbal A, Fautin DG & Buddemeier RW (2006) Modeling habitat distribution from organism occurrences and environmental data: case study using anemonefishes and their sea anemone hosts. *Marine Ecology Progress Series* 316: 269-283

Guisan A, Broennimann O, Engler R, Vust M, Yoccoz NG, Lehmann A & Zimmermann NE (2006) Using niche-based models to improve the sampling of rare species. *Conservation Biology* 20: 501-511

Haeckel E (1880) System der Acraspeden – Zweite Hälfte des Systems der Medusen. *Denkschriften der Medizinisch-Naturwissenschaftlichen Gesellschaft zu Jena*, Jena.

Hartwick RF (1991a) Observations on the anatomy, behaviour, reproduction and life cycle of the cubozoan *Carybdea sivickisi*. *Hydrobiologia* 216/217: 171-179.

Hartwick RF (1991b) Distributional ecology and behavior of the early life stages of the box-jellyfish *Chironex fleckeri*. *Hydrobiologia* 216/217: 181-188.

Holt RD & Gaines MS (1992) Analysis of adaptation in heterogenous landscapes: implications for the evolution of fundamental niches. *Evolutionary Ecology* 6: 433-447.

Hooper JNA (2000) Box jellyfish. In: *Wildlife of tropical north Queensland* (eds M Ryan & C Burwell). Queensland Museum, Brisbane, Australia.

International Commission on Zoological Nomenclature (1999) International Code of Zoological Nomenclature, 4th edition. The Natural History Museum, London.

Heads M (2005) Towards a panbiogeography of the seas. *The biological Journal of the Linnean Society* 84: 675-723.

Huang X & Madan A (1999) CAP3: A DNA sequence assembly program. *Genome Research* 9: 868-877.

Hudson RR (2002) Generating samples under a Wright-Fisher neutral model. *Bioinformatics* 18: 337-338.

Iwama Y (2001) Kurage sono maryoku to kaikata. Kyodo Publishers Ltd., Tokyo. (in Japanese)

Jablonski D, Sepkoski Jr. JJ, Bottjer DJ & Sheehan PM (1983) Onshore-offshore patterns in the evolution of Phanerozoic shelf communities. *Science* 222: 1123-1125.

Jablonski D & Bottjer DJ (1990) Onshore-offshore trends in marine invertebrate evolution. In: *Causes of Evolution: A paleontological perspective* (eds Ross RM & Allmon WD), pp. 21-75, University of Chicago press, Chicago, IL, USA.

Jacobsen M & Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23: 1801-1806.

Jarms G, Båmstedt U, Tiemann H, Martinussen MB & Fosså JH (1999) The holopelagic life cycle of the deep-sea medusa *Periphylla periphylla* (Scyphozoa, Coronatae). *Sarsia* 84: 55-65.

Jarms G, Tiemann H & Båmstedt U (2002) Development and biology of *Periphylla periphylla* (Scyphozoa: Coronatae) in a Norwegian fjord. *Marine Biology* 141: 647-657.

Jaynes ET (1957) Information theory and statistical mechanics. *Physical Review* 106: 620-630.

Kambhampati S, Peterson AT (2007) Ecological niche conservation and differentiation in the wood-feeding cockroaches, Cryptocercus, in the United States. *Biological Journal of the Linnean Society* 90: 457-466.

Kishinouye K (1891a) Zwei neue Medusen von *Charybdea* (*Ch. brevipedalia* n. sp., *Ch. latigenitalia* n. sp.). *Dobutsugaku Zasshi* 3: 437-440. (in Japanese with German diagnoses)

Kishinouye K (1891b) Hikurage. *Dobutsugaku Zasshi* 3: 508-509. (in Japanese)

Kishinouye K (1910) Some medusae of Japanese waters. *The Journal of the College of Science, Imperial University of Tokyo, Japan* 27: article 9.

Kluza DA & McNyset M (2005) Ecological niche modeling of aquatic invasive species. *Aquatic*

Invaders 16: 1-7.

Knowlton N (1993) Sibling species in the sea. *Annual Review of Ecology, Evolution, and Systematics* 24: 189-216.

Kool JT, Paris CB, Barber PH & Cowen RK (2011) Connectivity and the development of population genetic structure in Indo-West Pacific coral reef communities. *Global Ecology and Biogeography* 20: 695-706.

Kramp (1959) The hydromedusae of the Atlantic Ocean and adjacent waters. *Dana Report* 46: 1-279.

Kramp PL (1961) Synopsis of the medusae of the world. *Journal of the Marine Biological Association of the United Kingdom* 40: 1-469.

Kramp (1965) The hydromedusae of the Pacific and Indian Oceans. *Dana Report* 63: 1-161.

Kramp (1968) The hydromedusae of the Pacific and Indian Oceans – Sections II and III. *Dana Report* 72: 1-200.

Kubota S (1998) Faunal list of Coelenterata collected from Tanabe Bay and its vicinities. Order Cubomedusae (Phylum Cnidaria, Class Cubozoa). *Annual Report of the Seto Marine Biological Laboratory* 11: 33-34.

Langmead B & Salzberg S (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9: 357-359.

Larson RI, Mills CE & Harbison GR (1991) Western Atlantic midwater hydrozoan and scyphozoan medusae: *in situ* studies using manned submersibles. *Hydrobiologia* 216/217: 311-317

Lenton TM & Watson AJ (2000) Redfield revisited – 1. Regulation of nitrate, phosphate, and

oxygen in the ocean. *Global Biogeochemical Cycles* 14: 225-248.

Lewis C & Bentlage B (2009) Clarifying the identity of the Japanese Habu-kurage, *Chironex yamaguchii*, sp. nov. (Cnidaria: Cubozoa: Chirodropida). *Zootaxa* 2030: 59-65.

Lewis C & Long TAF. 2005. Courtship and reproduction in *Carybdea sivickisi* (Cnidaria: Cubozoa). *Marine Biology* 147: 477-483.

Lewis C, Kubota S, Migotto AE & Collins AG. 2008. Sexually dimorphic cubomedusa *Carybdea sivickisi* (Cnidaria: Cubozoa) in Seto, Wakayama, Japan. *Publications of the Seto Marine Biology Laboratory* 40: 1-8.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis, Durbin & 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25: 2078-2079.

Lieberman BS & Eldredge N (1996) Trilobite biogeography in the Middle Devonian: Geological processes and analytical methods. *Paleobiology* 22: 66-79.

Lindner A, Cairns SD & Cunningham CW (2008) From offshore to onshore: multiple origins of shallow-water corals from deep-sea ancestors. *PloS ONE* 3: e2429.

Little M, Pereira P, Carrette T & Seymour J (2006) Jellyfish responsible for Irukandji syndrome. QJM: An International Journal of Medicine 99: 425-427.

Locarnini RA, Mishonov AV, Antonov JI, Boyer TP & Garcia HE (2006) World Ocean Atlas 2005, Volume 1: Temperature. *NOAA Atlas NESDIS 61* (ed S Levitus), pp. 182. US Government Printing Office, Washington, DC.

Longhurst A (2007) *Ecological Geography of the Sea*, 2nd edn. Academic Press. Waltham MA, USA.

Martínez-Meyer E & Peterson AT (2006) Conservatism of ecological niche characteristics in North American plant species over the Pleistocene-to-Recent transition. *Journal of Biogeography* 33:1779-1789.

Martínez-Meyer E, Peterson AT & Navarro-Sigüenza AG (2003) Evolution of seasonal ecological niches in Passerina buntings (Aves: Cardinalidae). *Proceedings of the Royal Society B* 271: 1151-1157

Marques AC, Morandini AC & Pinto MM. 1997. Cnidome of *Chiropsalmus quadrumanus* (Cnidaria, Cubozoa) from Brazil. In: *'Congresso Latino-americano sobre Ciências do Mar, VII,Santos, 1997. Resumos Expandidos vol. II, Santos, IO-USP, ALICMAR'*. pp.136-138.

Mayer AG (1906) Medusae collected in the Hawaiian Islands by the steamer Albatross in 1902. Bull. *United States Fish Commission 1903* 23: 1131-1143.

Mayer AG (1910) Medusae of the World, Vol III, The Scyphomedusae. Carnegie Institution of Washington, Washington DC, USA.

Mayer AG (1915) Medusae of the Philippines and Torres Straits. Being a report on the Scyphomedsae collected by the U.S. Fisheries Bureau steamer “Albatross” in the Philippine Islands and Malay Archipelago, 1907–1910, and upon the medusae collected by the expedition of the Carnergie Institution of Washington to Torres Straits, in 1913. *Papers from the Tortugas Laboratory of the Carnegie Institution of Washington* 8: 157-202.

Mayer AG (1917). Report upon the Scyphomedusae collected by the U.S. Bureau of Fisheries Steamer "Albatross" in the Philippine and Malay Archipelago. *Bulletin of the United States National Museum* 100: 175-233.

Mc Kenna A, Hanna M, Banks E, Sivachenko A, Cibuliskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M & DePristo MA (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20: 1297-1303.

Menon MGK (1936) Scyphomedusae of Krusadai Island. *Bulletin of the Madras Government Museum – Natural Science and Natural History Section* 1: 1-9

Migotto AE, Marques AC, Morandini AC & da Silveira FL (2002) Checklist of the Cnidaria Medusozoa of Brazil. *Biota Neotropica* 2: 1-31.

Mishra DR, Narumalani S, Rundquist D & Lawson M (2005) High-resolution ocean color remote sensing of benthic habitats: a case study at the Roatan Island, Honduras. *IEEE Transactions on Geoscience and Remote Sensing* 43: 1592-1604.

Moore S (1988) A new species of cubomedusan (Cubozoa: Cnidaria) from Northern Australia. *The Beagle, Records of the Museum and Art Galleries of the Northern Territory* 5: 1-4.

Morandini AC (2003) Deep-Sea medusae (Cnidaria: Cubozoa, Hydrozoa and Scyphozoa) from the coast of Bahia (western south Atlantic, Brazil). *Mitteilungen Hamburgisches Zoologisches Museum und Institut* 100: 3-25.

Morandini AC & Marques AC (1997) 'Morbakka' syndrome: first report of envenomation by Cubozoa (Cnidaria) in Brazil. In: *VII Congreso Latino-Americano sobre Ciencias do Mar, Resumos Expandidos*, vol. 2, pp. 188 –189, Oceanographic Institute of the

University of Sao Paulo (IO-USP), Santos, Brazil.

Muñoz MES, Giovanni R, Siqueira MF, Sutton T, Brewer P, Pereira RS, Canhos DAL & Canhos VP (2009) openModeller: a generic approach to species' potential distribution modelling. *GeoInformatica* 15: 111-135.

Nagai H (2003) Recent progress in jellyfish toxin study. *Journal of Health Science* 49: 337-340.

Nagai H, Takuwa-Kuroda K, Nakao M, Sakamoto B, Crow GL & Nakajima T (2000a) Isolation and characterization of a novel protein toxin from the Hawaiian box jellyfish (sea wasp) *Carybdea alata*. *Biochemical and Biophysical Research Communications* 275: 589-594.

Nagai H, Takuwa-Kuroda K, Nakao M, Ito E, Miyake M, Noda M & Nakajima T (2000b) Novel proteinaceous toxins from the box jellyfish (sea wasp) *Carybdea rastoni*. *Biochemical and Biophysical Research Communications* 275: 582-588.

Nagai H, Takuwa-Kuroda K, Nakao M, Oshiro N, Iwanga S & Nakajima T (2002) A novel protein toxin from the deadly box jellyfish (sea wasp, Habu-kurage) *Chiropsalmus quadrigatus*. *Biochemical and Biophysical Research Communications* 66: 97-102.

Nair KK (1951) Medusae of the Trivandrum coast. Part 1. Systematics. *Bulletin of the Central Research Institute, University of Travancore, Series C, Natural Sciences* 2: 47-75

Nilsson D, Gislen L, Coates MM, Skogh C & Garm A (2005) Advanced optics in a jellyfish eye. *Nature* 435: 201-205.

Norris RD (2000) Pelagic species diversity, biogeography, and evolution. *Paleobiology* 26: 236-258.

Norris RD & de Vargas C (2000) Evolution all at sea. *Nature* 405: 23-24.

Nylander JAA (2004) MRMODELTEST v. 2. Uppsala University. Program distributed by the

author.

Orellana ER & Collins AG (2011) First report of the box jellyfish *Tripedalia cystophora* (Cubozoa: Tripedaliidae) in the continental USA, from Lake Wyman, Boca Raton, Florida. *Marine Biodiversity Records* 4: e54.

Owens H, Bentley AC & Peterson AT (2012) Predicting suitable environments and potential occurrences for coelocanth (*Latimeria* spp.). *Biodiversity and Conservation* 21: 577-587.

Palumbi SR (1992) Marine speciation on a small planet. *Trends in Ecology & Evolution* 7: 114-118.

Palumbi (2009) Speciation and the evolution of gamete recognition genes: pattern and process. *Heredity* 102: 66-76

Pearson AG. & Dawson TP (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography* 12: 361-371.

Pearson RG, Raxworthy CJ, Nakamura M & Peterson AT (2007) Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography* 34: 102-117.

Peterson AT (2003) Predicting the geography of species' invasions via ecological niche modeling. *The Quarterly Review of Biology* 78: 419-433

Peterson AT (2005) Predicting potential geographic ranges of invading species. *Current Science* 89: 9.

Peterson AT (2006) Uses and requirements of ecological niche models and related distributional models. *Biodiversity Informatics* 3: 59-72.

Peterson AT & Nyári AS (2008) Ecological niche conservatism and Pleistocene refugia in the Thrush-like Mourner,*Schiffornis* sp., in the neotropics. *Evolution* 62: 173-183

Peterson AT, Papes M & Eaton M (2007) Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. *Ecography* 4: 550-560

Peterson AT, Papeş M & Soberón J (2008) Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling* 213: 63-72.

Peterson AT, Soberón J & Sánchez-Cordero V (1999) Conservatism of ecological niches in evolutionary time. *Science* 285: 1265-1267

Peterson AT, Soberón J, Pearson RG, Anderson RP, Martínez-Meyer E, Nakamura M & Araújo MB (2011) Ecological niches and geographic distributions. *Monographs in Population Biology 49* (eds SA Levin & HS Horn), 314 pp. Princeton University Press, Princeton.

Phillips SJ (2008) Transferability, sample selection bias and background data in presence-only modelling: a response to Peterson et al. (2007). *Ecography* 31 272-278.

Phillips SJ & Dudík M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31: 161-175

Phillips SJ, Dudík M & Schapire RE (2004) A maximum entropy approach to species distribution modeling. *Proceedings of the 21st International Conference on Machine Learning (ACM, Banff, AB)* p. 83.

Phillips SJ, Anderson RP & Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231-259.

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

Pugh PR, Pagès F & Boorman B (1997) Vertical distribution and abundance of pelagic cnidarians in the eastern Weddell Sea, Antarctica. *Journal of the Marine Biological Association of the UK* 77: 341-360.

R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Ranson G (1945) Les scyphomeduses de la collection du Museum National d'Histoire Paris. II. Catalogs raisonne; origine des recoltes. *Bulletin du Museum d'Histoire Naturelle, Paris* 2: 312-320.

Raxworthy CJ, Martínez-Meyer E, Horning N, Nussbaum RA, Schneider GE, Ortega-Huerta MA & Peterson AT (2003) Predicting distributions of known and unknown reptile species in Madagascar. *Nature* 426: 837-841.

Reynaud M (1830) *Carybdea alata* n. sp. In: *Centurie Zoologique*. (ed RP Lesson) p. 95, pl. 33. Levrault, Paris, France.

Rice NH, Martínez-Meyer E & Peterson AT (2003) Ecological niche differentiation in the Aphelocoma jays: a phylogenetic perspective. Biological Journal of the Linnean Society 80: 369-383.

Robinson LM, Elith J, Hobday AJ, Pearson RG, Kendall BE, Possingham HP & Richardson AJ (2011) Pushing the limits in marine species distribution modelling: lessons from the land present challenges and opportunities. *Global Ecology and Biogeography* 20: 789-802.

Rödder D & Engler JO (2011) Quantitative metrics of overlaps in Grinnellian niches: advances and possible drawbacks. *Global Ecology and Biogeography* 20: 915-927.

Rottini D, Gusmani L, Parovel E, Avian M & Patriarca P (1995) Purification and properties of a cytolytic toxin in venom of the jellyfish *Carybdea marsupialis*. *Toxicon* 33: 315-326.

Russel FS (1970) *The Medusae of the British Isles*, vol 2. Cambridge University Press, Cambridge, UK.

Sanchez-Rodriguez J, Torrense E & Segura-Puertas L (2006) Partial purification and characterization of a novel neurotoxin and three cytolysins from box jellyfish (*Carybdea marsupialis*) nematocyst venom. Archives of Toxicology 80: 163-168.

Savtchenko A, Ouzounov D, Ahmad S, Acker J, Leptoukh G, Koziana J & Nickless D (2004) Terra and Aqua MODIS products available from NASA GES DAAC. Advances in Space Research 34: 710-714.

Schiermeier Q (2006) Climate change: A sea change. *Nature* 439: 256-260.

Smith WHF & Sandwell DT (1997) Global seafloor topography from satellite altimetry and ship depth soundings. *Science* 277: 1956-1962.

Smith TF & Waterman (1981) Identification of common molecular subsequences. *Journal of Molecular Biology* 147: 195-197.

Soberón J (2007) Grinellian and Eltonian niches and geographic distributions of species. *Ecology Letters* 10: 1115-1123.

Soberón J (2010) Niche and area of distribution modeling: a population ecology perspective. *Ecography* 33: 159-167.

Soberón J & Peterson AT (2004) Biodiversity informatics: managing and applying primary biodiversity data. Philosophical Transactions of the Royal Society B 359: 689-698.

Soberón J & Peterson AT (2005) Interpretation of models of fundamental ecological niches and

species' distributional areas. Biodiversity Informatics 2: 1-10.

Southcott RV (1967) Revision of some Carybdeidae (Scyphozoa: Cubomedusae), including a description of the jellyfish responsible for the "Irukandji syndrome". *Australian Journal of Zoology* 15: 651-671.

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

Stangl K, Salvini-Plawen LV & Holstein TW (2002) Staging and induction of medusa metamorphosis in *Carybdea marsupiali*s (Cnidaria, Cubozoa). *Vie et Milieu* 52: 131-140.

Stiasny G (1937) Scyphomedusae. John Murray Expedition 1933–1934. Sci Rep 4:203–242

Stockwell D, Peters D (1999) The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographic Information Science* 13: 143-158

Stockwell D & Peters D (1999) The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographical Information Scienc*e 13:143-158.

Stockwell D & Peterson AT (2002) Controlling bias in biodiversity data. In: *Predicting species occurrences: issues of scale and accuracy*. (eds JM Scott, PJ Heglund & ML Morrison) pp. 537-546. Island Press, Washington DC, USA.

Straehler-Pohl I & Jarms G (2005) Life cycle of *Carybdea marsupialis* appears to be a modified strobila. *Marine Biology* 147: 1271-1277.

Straehler-Pohl I & Jarms G (2011) Morphology and life cycle of *Carybdea morandinii*, sp. nov. (Cnidaria), a cubozoan with zooxanthellae and peculiar polyp anatomy. *Zootaxa* 2755:

36-56.

Studebaker JP (1972) Development of the cubomedusa, *Carybdea marsupialis*. M.S. thesis, University of Puerto Rico, Mayaguez, Commonwealth of Puerto Rico, USA.

Suntrarachun S, Roselieb M, Wilde H & Sitprija V (2001) A fatal jellyfish encounter in the Gulf of Siam. *Journal of Travel Medicine* 8: 150-151.

Swenson NG (2008) The past and future influence of geographic information systems on hybrid zone, phylogeographic and speciation research. Journal of Evolutionary Biology 21: 421-434.

Swofford DL (2003) PAUP*. Phylogenetic analysis using parsimony (*and other methods), version 4. Sunderland, MA: Sinauer Associates.

Tahera Q & Kazmi QB (2006) New records of two jellyfish medusae (Cnidaria: Scyphozoa: Catostylidae: Cubozoa: Chirodropidae) from Pakistani waters. *Journal of the Marine Biological Association of the UK* 86: 1482.

Therriault TW & Herborg LM (2008) Predicting the potential distribution of the vase tunicate Ciona intestinalis in Canadian waters: informing a risk assesment. *ICES Journal of Marine Science* 65: 788-794.

Thiel ME (1928) Die Scyphomedusen des Zoologischen Staatsinstituts und Zoologischen Museums in Hamburg: I, Cubomedusae, Stauromedusae und Coronatae. *Mitteilungen Hamburgisches Zoologisches Museum und Institut* 43: 1-34.

Thomas CS, Scott SA, Galanis DJ & Goto RS (2001) Box jellyfish (*Carybdea alata*) in Waikiki: their influx cycle plus the analgesic effect of hot and cold packs on their stings to swimmers at the beach: a randomized, placebo-controlled, clinical trial. Hawaii Medical

Journal 60: 278.

Thuiller W, Brotons L, Araújo MB & Lavorel S (2004) Effects of restricting environmental range of data to project current and future species distributions. *Ecography* 27: 165-172.

Tittensor DP, Baco AR, Hall-Spencer JM Orr JC & Rogers AD (2010) Seamounts as refugia from ocean acidification for cold-water stony corals. *Marine Ecology: An Evolutionary Perspective* 31: 212-225.

Tybergheim, L., Verbruggen, H., Pauly, K., Troupin, C., Mineur, F. & De Clerk, O. (2011) Bio-ORACLE: a global environmental dataset for marine species distribution modelling. *Global Ecology and Biogeography* 21: 272-281.

van der Horst R (1907) On a new cubomedusa from the JavaSea: *Chiropsalmus buitendijki*. *Notes from the Leyden Museum* 29: 101-106.

Uchida T (1929) Studies on the Stauromedusae and Cubomedusae, with special reference to their metamorphosis. *Japanese Journal of Zoology* 2: 103-193.

Uchida T (1947) Some medusae from the central Pac*ific. Journal of the Faculty of Science, Hokkaido University Series 6, Zoology* 9: 297-319.

Uchida T (1954) Distribution of Scyphomedusae in Japanese and its adjacent waters. *Journal of the Faculty of Science, Hokkaido University Series 6, Zoology* 12: 209-219.

Uchida T (1970) Revision of Japanese Cubomedusae. *Publications of the Seto Marine Biology Laboratory* 17: 289-297.

UNESCO (2009) Global Open Oceans and Deep Seabed (GOODS) – Biogeographic Classification. *IOC Technical Series 84* (eds M Vierros, I Cesswell, E Escobar Briones, J Rice & J Ardron), 89 pp. UNESCO-IOC, Paris, France.

Vinogradov ME & Shushkina EA (2002) Vertical distribution of gelatinous macroplankton in the North Pacific observed by manned submersibles *Mir-1* and *Mir-2*. *Journal of Oceanography* 58: 295-303.

Vogler C, Benzie J, Lessios H, Barber P & Wörheide G (2008) A threat to coral reefs multiplied? Four species of crown-of-thorns starfish. *Biology Letters* 4: 696-699.

Warren DL, Glor RE & Turelli M (2008) Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* 62: 2868-2883.

Weir B and Cockerham C (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution* 38: 1358-1370.

Werner B (1973) Spermatozeugmen und Paarungsverhalten bei *Tripedalia cystophora* (Cubomedusae). *Marine Biology* 18: 212-217.

Werner B, Cutress CE & Studebaker JP (1971) Life cycle of *Tripedalia cystophora* Conant (Cubomedusae). *Nature* 232: 582-583.

Wiley EO, McNyset KM, Peterson AT, Robins CR & Stewart AM (2003) Niche modeling and geographic range predictions in the marine environment using a machine learning algorithm. *Oceanography (Washington DC)* 16: 120-127.

Williamson J, Fenner P, Burnett J & Rifkin J (1996) Venomous and poisonous marine animals: a medical and biological handbook. NSW University Press, Sydeney, Australia.

Winkel KD, Hawdon GM, Fenner PJ, Gershwin L, Collins AG & Tibbals J (2003) Jellyfish antivenoms: past, present, and future. *Journal of Toxicology* 22: 115-127.

Wiltshire CJ, Sutherland SK, Fenner PJ & Young AR (2000) Optimization and preliminary characterization of venom isolated from 3 medically important jellyfish: the box

(*Chironex fleckeri*), Irukandji (*Carukia barnesi*), and blubber (*Catostylus mosaicus*) jellyfish. *Wilderness and Environmental Medicine* 11: 241-250.

Yamaguchi M & Hartwick R (1980) Early life history of the sea wasp, *Chironex fleckeri* (class Cubozoa). In: *Development and cellular biology of coelenterates* (eds P Tardent & R Tardent), pp. 11–16. Elsevier/North-Holland Biomedical Press, Amsterdam, The Netherlands.

Yamaguchi M (1982) Cubozoans and their life histories. *Kaiyou to Seibutsu* 4: 248-254.

Yamasu T & Yoshida M (1976) Fine structure of complex ocelli of a cubomedusan, *Tamoya bursaria* Haeckel. *Cell and Tissue Research* 70: 325-339.

Yanagihara AA, Kuroiwa J, Oliver L, Chung J & Kunkel D (2002) Ultrastructure of a novel eurytele nematocyst of *Carybdea alata* Reynaud (Cubozoa, Cnidaria). Cell and Tissue Research 308: 307-318.

Yoshimoto CM & Yanagihara AA (2002) Cnidarian (coelenterate) envenomations in Hawai'i improve following heat application. Transactions of the Royal Society of Tropical Medicine and Hygiene 96: 300-303.

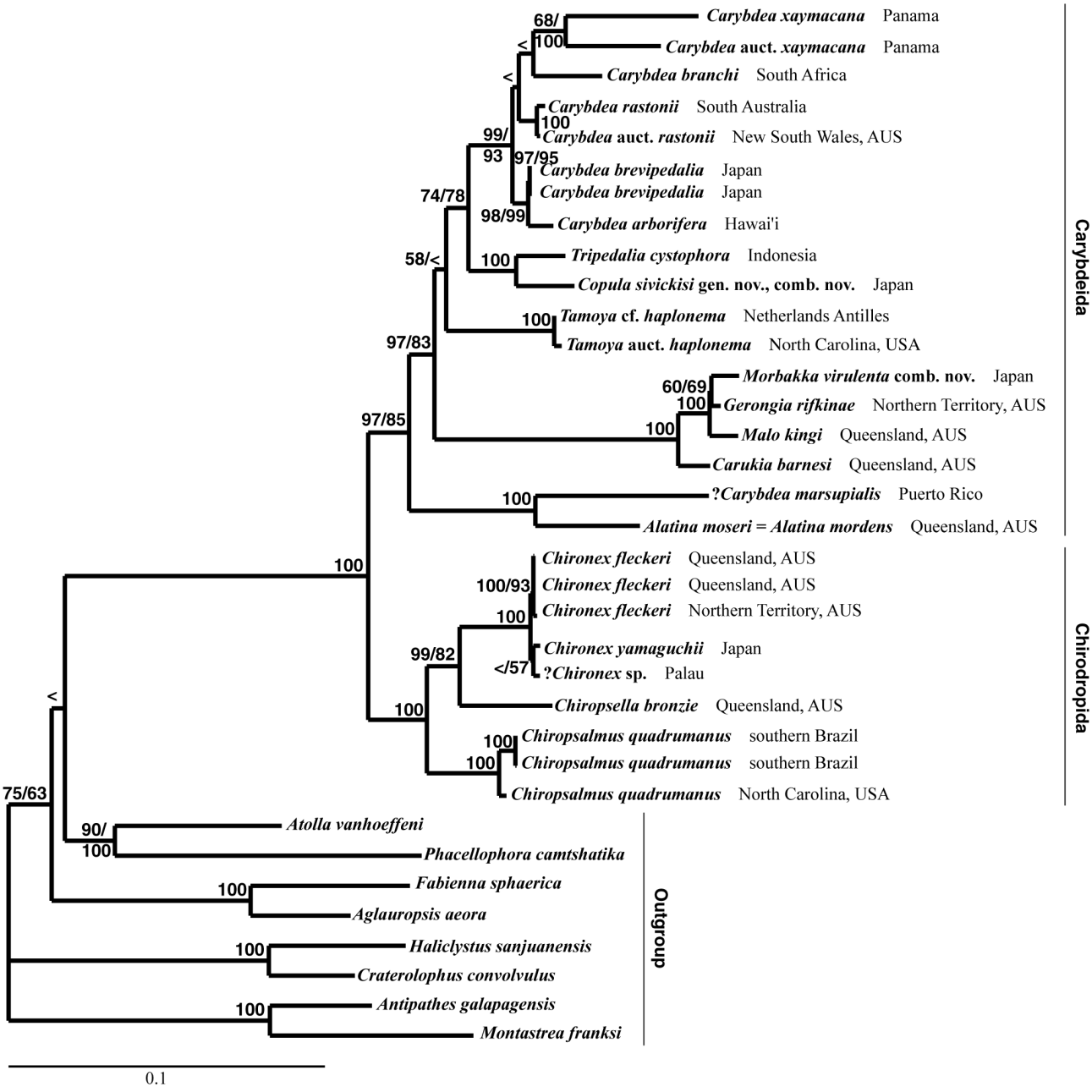
Zerbino DR & Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* 18: 821-829.

Zoubir A & Boashash B (1998) The bootstrap and its application in signal processing. *IEEE Signal Processing Magazine* 15: 56-76.

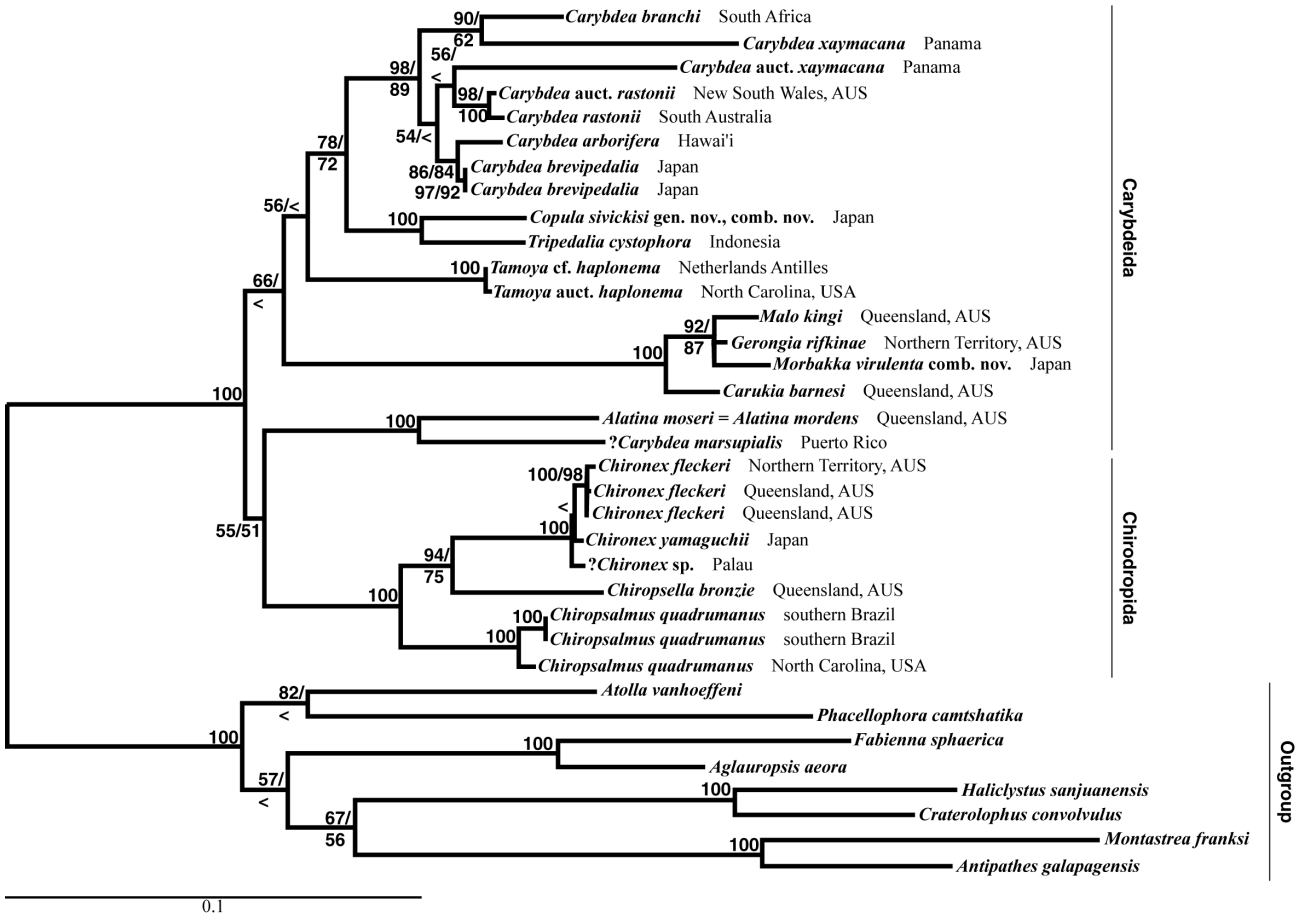
APPENDIX

Appendix I Samples included in phylogenetic analyses, sample origin, Genbank accession numbers (accession numbers in italics had been generated previously) as well as museum catalogue numbers for voucher specimens (where applicable) are provided. Specimens were deposited at the National Museum of Natural History, Smithsonian Institution [USNM (Washington, DC, USA)], the University of Kansas Natural History Museum [KUNHM (Lawrence, KS, USA)], Australian Museum [AMS (Sydney, NSW, Australia)], the Queensland Museum [QM (Brisbane, QLD, Australia)], the South Australian Museum [SAM (Adelaide, South Australia)] and the collection of the Departmento de Zoologia, Universidade Federal do Rio de Janeiro [DZ-UFRJ (Rio de Janeiro, Brasil)]. Paravouchers: 1) Specimens collected in Queensland (QM G327938 - G327941 and USNM 112754 & 2048431) match the published *Chiropsella bronzie* SSU and LSU sequences, but we decided to use the published sequence data in the phylogenetic analyses presented here; 2) DNA was extracted from specimens USNM 1125365 - 1125367, but it is not clear which one of the DNA samples corresponds to which one of the formalin-preserved specimens.

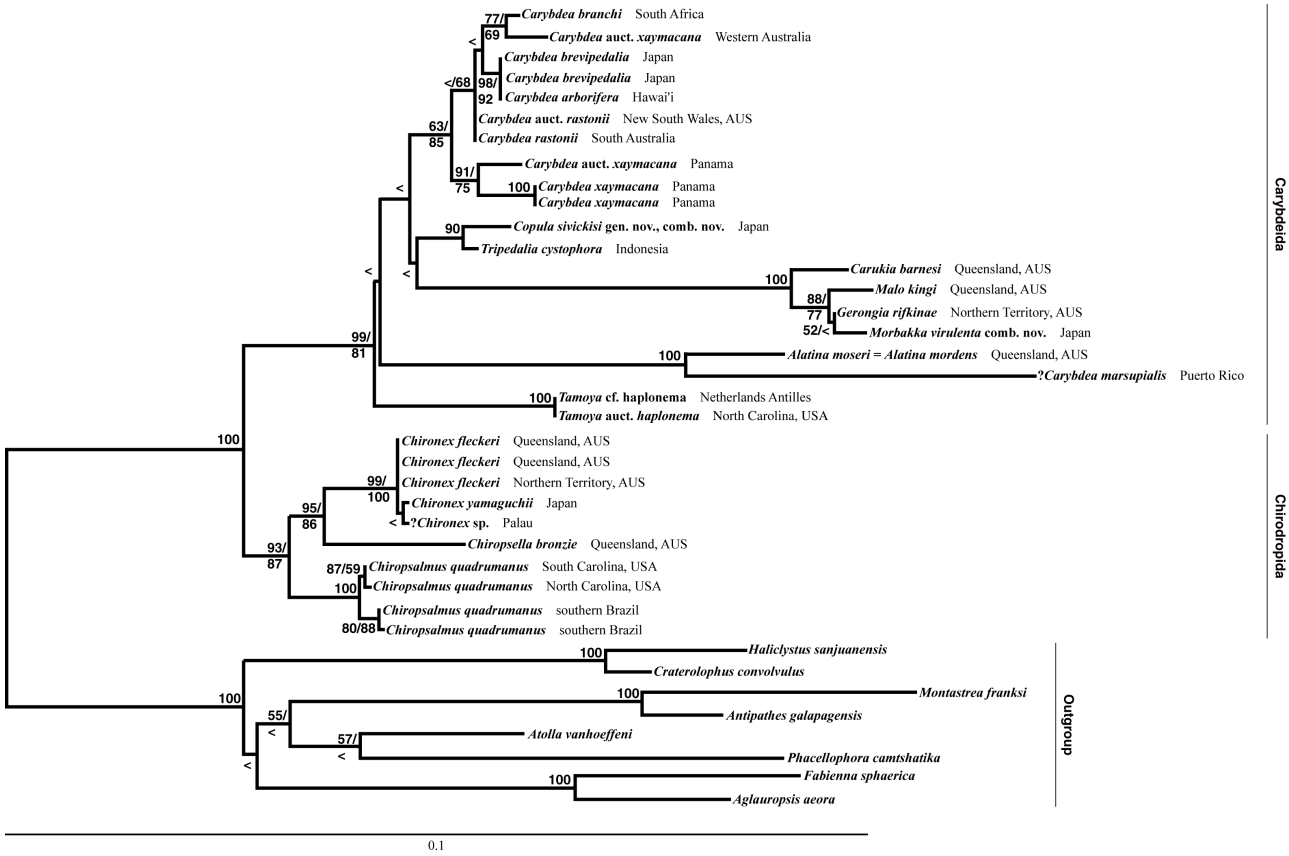
	16S	SSU	LSU	Voucher	Origin
Chirodropida					
Chirodropidae					
Chironex fleckeri	GQ849101	GQ849073	*AY920785*		Darwin, NT, AUS
Chironex fleckeri	GQ849102	GQ849074	GQ849051		Weipa, QLD, AUS
Chironex fleckeri	GQ849103				Weipa, QLD, AUS
Chironex fleckeri		GQ849075	GQ849052		Townsville, QLD, AUS
Chironex yamaguchii		GQ849076	GQ849053	USNM 1121554	Okinawa, Japan
?*Chironex* sp.	GQ849104	GQ849077	GQ849054		Palau
Chiropsalmidae					
Chiropsaella bronzie		*AF358103*	*AY920786*	paravouchers[1]	Queensland, AUS
Chiropsella bronzie	GQ849099			USNM 1129754	Port Douglas, QLD, AUS
Chiropsalmus quadrumanus	GQ849111	GQ849079	GQ849056	DZ-UFRJ 1-103	Southern Brazil
Chiropsalmus quadrumanus	GQ849110	GQ849078	GQ849055	DZ-UFRJ 1-104	Southern Brazil
Chiropsalmus quadrumanus		GQ849080			South Carolina, USA
Chiropsalmus quadrumanus	GQ849109	GQ849081	GQ849057	USNM 1124254	North Carolina, USA
Carybdeida					
Alatinidae					
Alatina mordens		GQ849082	GQ849058	USNM 1124421	Cairns, QLD, AUS
Alatina mordens	GQ506980			USNM 1124410	Cairns, QLD, AUS
Carukiidae					
Carukia barnesi			GQ849059	USNM 1124560	Cairns, QLD, AUS
Carukia barnesi	GQ849097			USNM 1124558	Cairns, QLD, AUS
Carukia barnesi		*AF358107*		USNM 1124559	Cairns, QLD, AUS
Gerongia rifkinae	GQ849119	*AF358105*	*AY920788*		Northern Territory, AUS
Morbakka virulenta	GQ849121	GQ849083	GQ849060	paravouchers[2]	Hiroshima, Japan
Malo kingi		GQ849084	GQ849061	SAM H1062	Queensland, AUS
Tamoyidae					
Tamoya auct. *haplonema*	GQ849122	GQ849085	GQ849062		North Carolina, USA
Tamoya cf. *haplonema*	GQ849095	GQ849086	GQ849063	USNM 1124022	Bonaire, Netherlands Antilles
Tripedaliidae					
Copula sivickisi		GQ849087	GQ849064	KUNHM 002822	Wakayama, Japan
Copula sivickisi	GQ849113			USNM 1124561	Cairns, QLD, AUS
Tripedalia cystophora		GQ849088	GQ849065	USNM 1124454	Kalimantan, Indonesia
Tripedalia cystophora	GQ849123			USNM 1124457	Kalimantan, Indonesia
Carybdeidae					
Carybdea arborifera	GQ849096	GQ849091	GQ849068	USNM 1124452	Honolulu, O'ahu, Hawai'i
Carybdea branchi		GQ849089	GQ849066	USNM 1124259	Cape Town, South Africa
Carybdea brevipedalia	GQ849106	GQ849092	GQ849069	KUNHM 002614	Wakayama, Japan
Carybdea brevipedalia		GQ849093	GQ849070	USNM 1124250	Hiroshima Bay, Japan
Carybdea rastonii	GQ849112	*AF358108*	*AY920787*		South Australia, AUS
Carybdea auct. *rastonii*	GQ849116	GQ849094	GQ849071	AMS G17493	New South Wales, AUS
Carybdea xaymacana	GQ849114	*AY920775*	GQ860995		Panama (Caribbean)
Carybdea xaymacana		GQ849090		KUNHM 002755	Panama (Caribbean)
Carybdea auct *xaymacna*	GQ849118		GQ849067	USNM 1073334	Panama (Caribbean)
Carybdea auct *xaymacana*	GQ849115	*AF358109*		USNM 100349	Busselton, WA, AUS
?*Carybdea marsupialis*	*AF360118*	*AF358106*	GQ849072		Puerto Rico



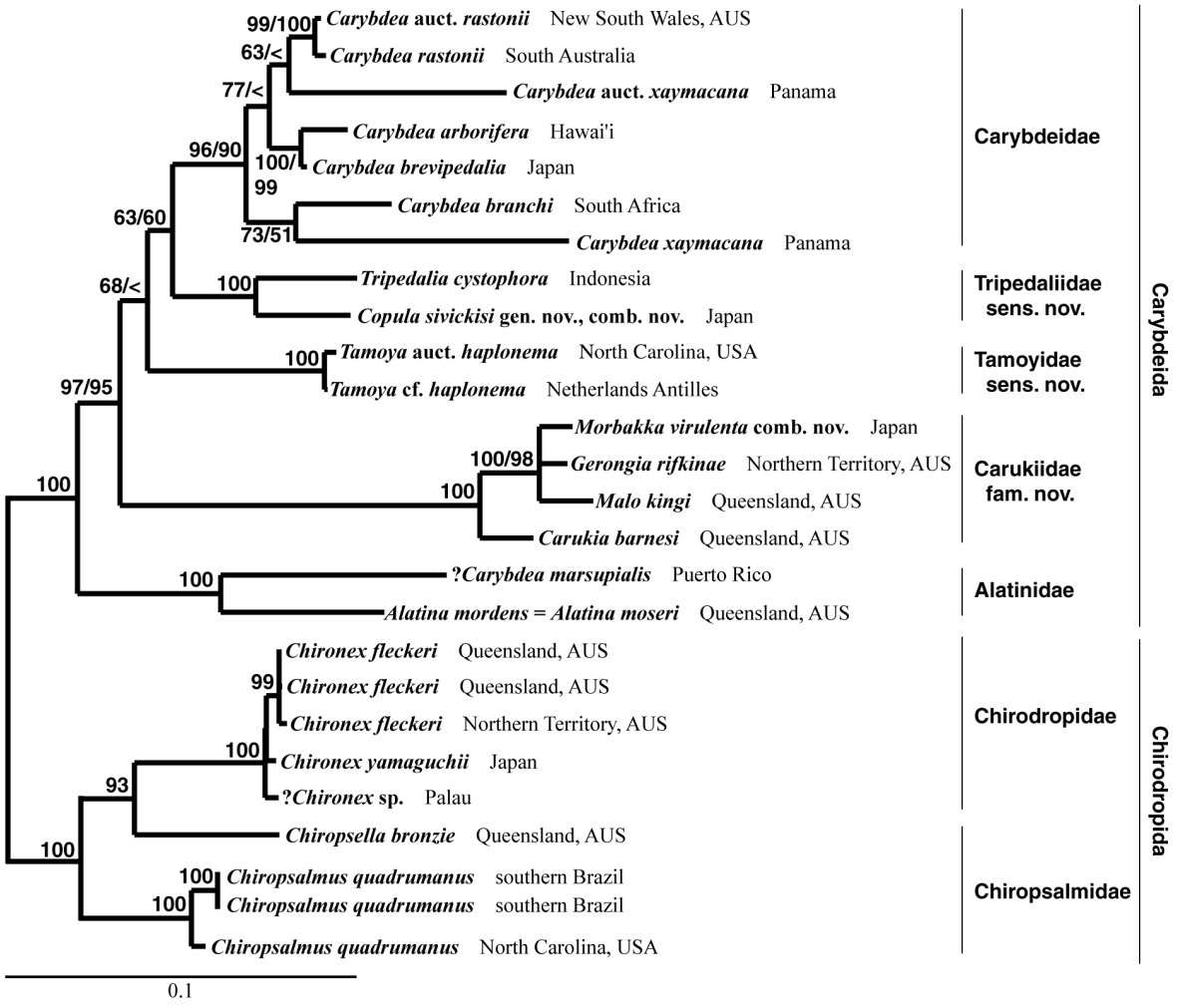
Appendix II Combined SSU and LSU likelihood analysis (GTR+I+G) including a cnidarian outgroup. The alignment contains 4956 characters; 1057 are parsimony informative and 3609 are invariant. The LSU partition contains 3172 characters and the SSU partition contains 1784 characters. Values on nodes represent 1000 parametric ML/MP bootstrap replicates; if only one value is provided it applies to both ML and MP bootstrap analysis. < = bootstrap value below 50.



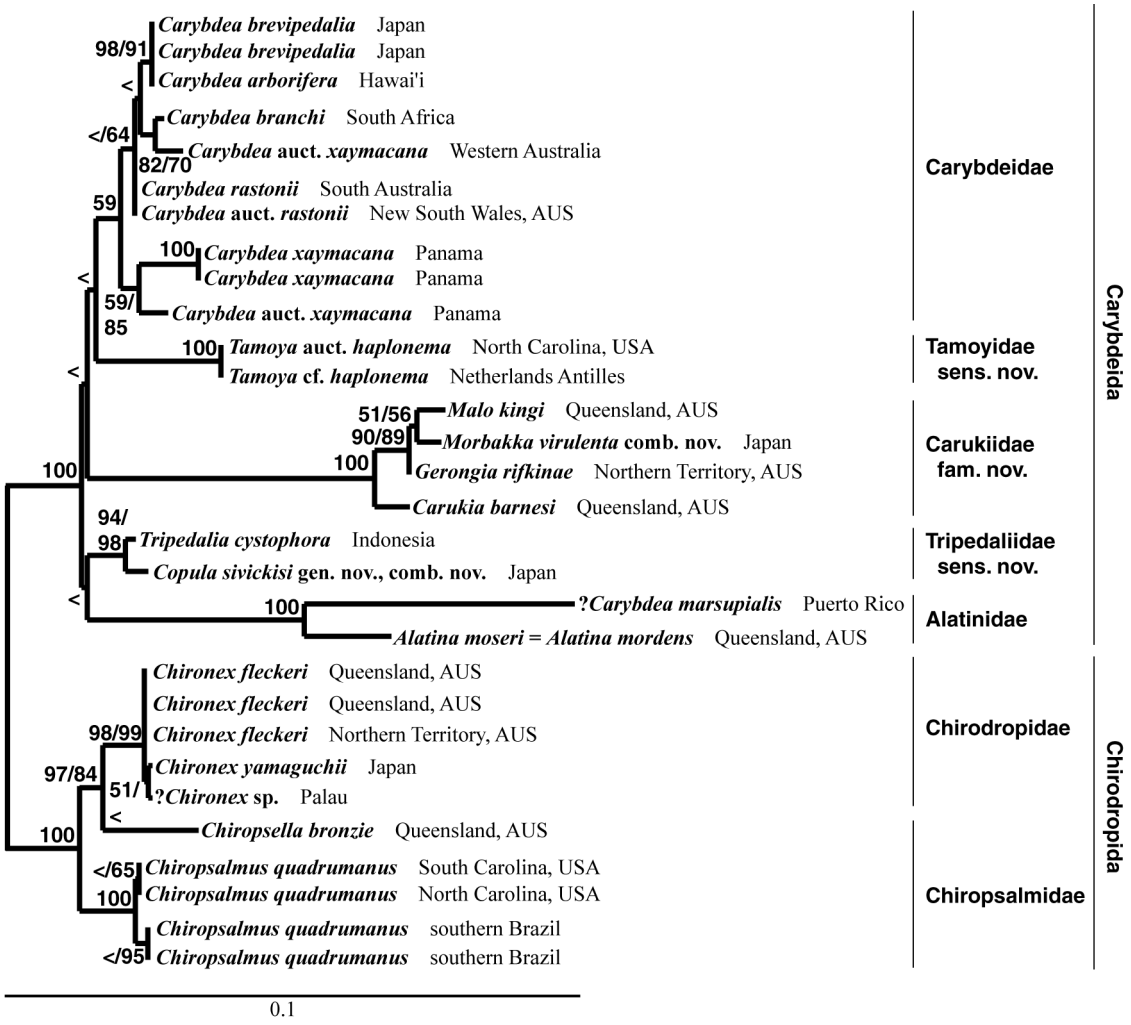
Appendix III LSU likelihood analysis (GTR+I+G) including a cnidarian outgroup. The alignment contains 3129 characters; 716 are parsimony informative and 2197 are invariant. Values on nodes represent 1000 parametric ML/MP bootstrap replicates; if only one value is provided it applies to both ML and MP bootstrap analysis. < = bootstrap value below 50.



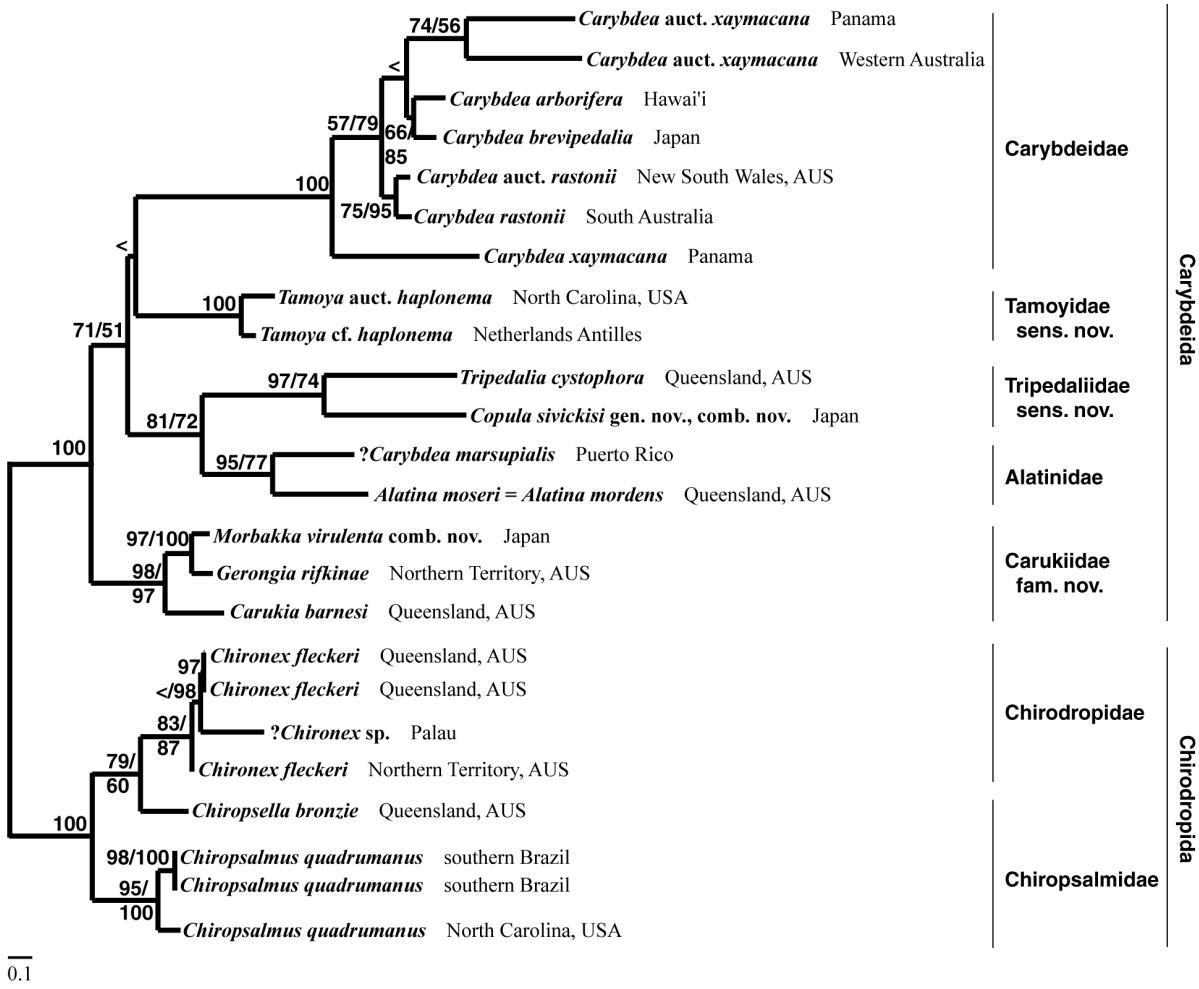
Appendix IV SSU likelihood analysis (GTR+I+G) including a cnidarian outgroup. The alignment contains 1742 characters; 298 are parsimony informative and 1378 are invariant. Values on nodes represent 1000 parametric ML/MP bootstrap replicates; if only one value is provided it applies to both ML and MP bootstrap analysis. < = bootstrap value below 50.



Appendix V LSU likelihood topology (GTR+I+G). The alignment contains 3292 characters; 523 are parsimony informative and 2605 are invariant. Values on nodes represent 1000 parametric ML/MP bootstrap replicates; if only one value is provided it applies to both ML and MP bootstrap analysis. < = bootstrap value below 50.



Appendix VI SSU likelihood topology (GTR+I+G). The alignment contains 1777 characters; 175 are parsimony informative and 1554 are invariant. Bootstrap values on nodes represent 1000 parametric ML/MP bootstrap replicates; if only one value is provided it applies to both ML and MP bootstrap analysis. < = bootstrap value below 50.



Appendix VII 16S likelihood topology (GTR+I+G). The alignment contains 486 characters; 248 are parsimony informative and 203 are invariant. Bootstrap values on nodes represent 1000 parametric ML/MP bootstrap replicates; if only one value is provided it applies to both ML and MP bootstrap analysis. < = bootstrap value below 50**.**