

# Hidden diversity of Ctenophora revealed by new mitochondrial COI primers and sequences

Lynne Christianson<sup>1</sup>, Shannon Johnson<sup>1</sup>, Darrin Schultz<sup>2</sup>, and Steven Haddock<sup>1</sup>

<sup>1</sup>MBARI

<sup>2</sup>University of California, Santa Cruz

November 30, 2021

## Abstract

**[Definitive version of this article may be found [here](#)]** The mitochondrial gene cytochrome-c-oxidase subunit 1 (COI) is useful in many taxa for phylogenetics, population genetics, metabarcoding, and rapid species identifications. However, the phylum Ctenophora (comb jellies) has historically been difficult to study due to divergent mitochondrial sequences and the corresponding inability to amplify COI with degenerate and standard COI ‘barcoding’ primers. As a result, there are very few COI sequences available for ctenophores, despite over 200 described species in the phylum. Here, we designed new primers and amplified the COI fragment from members of all major groups of ctenophores, including many undescribed species. Phylogenetic analyses of the resulting COI sequences revealed high diversity within many groups that was not evident from more conserved 18S rDNA sequences, in particular among the Lobata. The COI phylogenetic results also revealed unexpected community structure within the genus *Bolinopsis*, suggested new species within the genus *Bathocyroe*, and supported the ecological and morphological differences of some species such as *Lampocteis cruentiventer* and similar lobates (*Lampocteis* sp. ‘V’ stratified by depth, and ‘A’ differentiated by color). The newly described primers reported herein provide important tools to enable researchers to illuminate the diversity of ctenophores worldwide via quick molecular identifications, improve the ability to analyze environmental DNA by improving reference libraries and amplifications, and enable a new breadth of population genetic studies.

Hidden diversity of Ctenophora revealed by new mitochondrial *COI* primers and sequences

Lynne M. Christianson<sup>1\*</sup>, Shannon B. Johnson<sup>1\*</sup>, Darrin T. Schultz<sup>1,2</sup>, Steven H.D. Haddock<sup>1\*</sup>

<sup>1</sup>Monterey Bay Aquarium Research Institute, Moss Landing, CA 95039, US

<sup>2</sup>Department of Biomolecular Engineering and Bioinformatics, University of California, Santa Cruz, CA 95064, US

\* Correspondence: lynne@mbari.org, sjohnson@mbari.org, haddock@mbari.org

## Abstract

**[The final version of this article may be found [here](#)]**

The mitochondrial gene cytochrome-c-oxidase subunit 1 (*COI*) is useful in many taxa for phylogenetics, population genetics, metabarcoding, and rapid species identifications. However, the phylum Ctenophora (comb jellies) has historically been difficult to study due to divergent mitochondrial sequences and the corresponding inability to amplify *COI* with degenerate and standard *COI* ‘barcoding’ primers. As a result, there are very few *COI* sequences available for ctenophores, despite over 200 described species in the phylum. Here,

we designed new primers and amplified the *COI* fragment from members of all major groups of ctenophores, including many undescribed species. Phylogenetic analyses of the resulting *COI* sequences revealed high diversity within many groups that was not evident from more conserved *18S* rDNA sequences, in particular among the Lobata. The *COI* phylogenetic results also revealed unexpected community structure within the genus *Bolinopsis*, suggested new species within the genus *Bathocyroe*, and supported the ecological and morphological differences of some species such as *Lampocteis cruentiventer* and similar undescribed lobates (*Lampocteis* sp. ‘V’ stratified by depth, and ‘A’ differentiated by color). The newly designed primers reported herein provide important tools to enable researchers to illuminate the diversity of ctenophores worldwide via quick molecular identifications, improve the ability to analyze environmental DNA by improving reference libraries and amplifications, and enable a new breadth of population genetic studies.

## KEYWORDS

barcoding, biodiversity, cytochrome-c oxidase, Ctenophora, eDNA, metabarcoding, primers

## 1 | Introduction

Ctenophora is a phylum of gelatinous marine zooplankton with nearly 200 named species (Mills 2017). Their unique body plan separates them from other gelatinous plankton and from Cnidarians with which they are sometimes associated (Dunn *et al.* 2015). They are found throughout the world ocean, from both poles to the equator and from the surface to the deep-sea (Harbison *et al.* 1978), with the deepest ctenophore observed at over 7,000 meters depth (Lindsay & Miyake 2007). Many species are common and well-studied, particularly coastal species like *Mnemiopsis leidyi*, which is noted for having been introduced to habitats around the world. However, most deep-living ctenophores remain undescribed because specimens are delicate, difficult to access, and often damaged during collection (Haddock 2004). The use of remotely operated vehicles (ROVs) and specialized sampling equipment have expanded our ability to observe and collect ctenophores in the deep-sea during the last 30 years of research (Haddock *et al.* 2017). Morphological investigations of specimens collected from the deep sea suggest that we have only begun to reveal the remarkable diversity within this phylum. However, there are few taxonomic experts who work on ctenophores, and morphological identifications often are stymied by damaged specimens, cryptic morphology, and poor preservation in all fixatives. Molecular identifications can provide relatively quick identifications, especially for taxa like ctenophores that have few taxonomic experts, although polymorphic loci and good reference libraries are critical to achieve this goal.

Sequence data from the nuclear *18S* ribosomal gene provide a molecular phylogenetic framework for the broad relationships within ctenophores (Podar *et al.* 2001; Simion *et al.* 2015), and transcriptomes from a handful of species allow for more in-depth studies of some representative diversity (Simion *et al.* 2017). However, the *18S* rDNA gene fragment is highly conserved, and the phylogenies often do not effectively discriminate between many species and closely related genera particularly in groups such as Lobata. For example, some genera, such as *Bathocyroe*, *Eurhamphaea*, *Deiopea* and *Kiyohimea*, have nearly identical *18S* rDNA sequences, showing the limitations of the utility of the *18S* fragment with respect to species delineation (Haddock *et al.* in prep). The “barcoding” mitochondrial cytochrome-c-oxidase subunit-I (*COI*) sequence fragment is typically useful for species identification and delimitation (DeSalle & Goldstein 2019). Many degenerate PCR primer sets are available to amplify a broad swath of taxa, from bacteria to humans (Folmer *et al.* 1994; Gelleret *et al.* 2013; Leray *et al.* 2013; Siddall *et al.* 2009). While the barcode locus is successfully amplified for many taxa, amplification and/or its utility is problematic for many others, often for non-model organisms (Vrijenhoek & Waples 2012). One such example is Ctenophora, since many species have extremely high rates of mitochondrial evolution, are rich in adenine (A) and thymine (T) residues, and have variable gene order within the mitochondrial genome, even within a genus (Arafat *et al.* 2018; Kohn *et al.* 2012; Pett *et al.* 2011; Schultz *et al.* 2020). A consequence of high mitochondrial variability is that common primers are often poorly suited to amplify the *COI* fragment. Unsuccessful amplification of ctenophores by commonly used barcoding primers has a number of important ramifications including: a lack of quick and easy molecular identifications that results in difficulties for revealing diversity, few ctenophore sequences in public databases, and a deficiency of easily amplified phylogenetic and population-genetic markers.

The lack of robust *COI* primers and the resulting paucity of ctenophore sequences available in public repositories also hampers our understanding of the role of ctenophores in ocean ecology. Although they are delicate animals, ctenophores are carnivorous and play a critical role as predators in food webs. Despite their seemingly low nutrient content, they are also prey for a large range of animals (Choy *et al.* 2017; Thiebot & McInnes 2020; Yeh *et al.*

2020). Metabarcoding and eDNA studies are powerful tools to assess community diversity, ecosystem monitoring, and function (Eble *et al.* 2020). Many manuscripts have highlighted the presence and abundance of ctenophores in zooplankton metabarcoding analyses based on *18S*rDNA fragments (Günther *et al.* 2018; López-Escardó *et al.* 2018; Preston *et al.* 2020; Schroeder *et al.* 2020; Yeh *et al.* 2020); However, since the fragment of *18S* is often used in metabarcoding studies but lacks resolution to discriminate between most species of ctenophores, often all members of the entire phylum are lumped together (Günther *et al.* 2018; López-Escardó *et al.* 2018; Preston *et al.* 2020; Schroeder *et al.* 2020; Yeh *et al.* 2020). In manuscripts that used multiple loci in a metabarcoding framework, a significant proportion of *18S* sequences were from ctenophores, yet *COI* often failed to detect any (Djurhuus *et al.* 2018; Günther *et al.* 2018), or the ctenophore sequences had poor taxonomic assignments so the results were not addressed (Pitz *et al.* 2020). For other environmental studies of metabarcoding and eDNA, the entire phylum of Ctenophora is often missing from diversity estimates (Lacoursière-Roussel *et al.* 2018) or they are lumped in an ‘unassigned taxa’ category (Leray & Knowlton 2015).

For this study, we designed multiple primers to amplify *COI* from all major clades of ctenophores, including many deep-living, undescribed species. We applied those primers across the phylum and tested taxonomic assignments for various groups. Finally, as a case study, we used our newly generated sequences as a library for an eDNA study along the eastern Pacific coast, and provided species-level resolution for taxonomic assignments of ctenophores.

## 2 | Materials and Methods

### 2.1 | Sample Collection

Ctenophores were primarily collected by remotely operated vehicles (ROVs), blue-water SCUBA diving, and midwater trawls, along with contributions from collaborators around the globe (Table S1). Tissue samples for genomic DNA or RNA were flash frozen in liquid nitrogen, then stored at -80°C. Genomic DNA was isolated from frozen samples using the DNeasy DNA Blood and Tissue kit (Qiagen, Germantown, MD) or the Monarch Genomic DNA Purification Kit (New England Biolabs, Ipswich, MA) according to the manufacturers’ directions. When possible, multiple individuals within a species from different localities and depths were separately amplified and sequenced.

### 2.2 | Primer design

Initial ctenophore *COI* primers were designed based on *COI* sequences from our unpublished ctenophore transcriptome data, published mitochondrial sequences of *Mnemiopsis leidyi* (GenBank accession #: NC016117, JF760210), *Pleurobrachia pileus* (GenBank accession #: JF760211), and the ctenophores which we could successfully amplify with ‘Folmer primers’ (Folmer *et al.* 1994) (Figure 1). RNA extractions, transcriptome sequencing, and analyses were conducted as previously described (Francis *et al.* 2013). We designed primers for the mitochondrial *COI* gene with Primer3 (Untergasser *et al.* 2012) within Geneious Prime (v. 2020.2.3, www.geneious.com), using PrimerQuest (www.idtdna.com), or by eye based on nucleotide alignments. Primer design was an iterative process, with multiplexed PCR assays suggesting which variants and combinations of forward and reverse primers worked best. Primer positions were numbered relative to the sites as they occur in *M. leidyi* (JF760210; Pett *et al.* 2011), as indicated in Figure 1.

### 2.3 | Amplification and Sequencing

Fragments of *18S* rDNA were amplified using primers MitchA and MitchB (Medlin *et al.* 1988). Both *18S* and *COI* fragments were amplified using Phusion High-Fidelity PCR Master Mix with HF buffer (New England BioLabs, Ipswich, MA) in a Veriti PCR thermal cycler (Life Technologies, Carlsbad, CA). Annealing

temperatures were most successful at 50°C for most species. Gene fragments were sequenced bi-directionally with PCR primers and the BigDyeTerminator v3.1 sequencing kit according to the manufacturer’s protocol and analyzed on a 3500xL Genetic Analyzer (Life Technologies, Carlsbad, CA).

## 2.4 | Statistical Methods

Sequence fragments were assembled, edited, and aligned with MUSCLE (Edgar 2004) within Geneious Prime. The *COI* alignment was translated with the Mold, Protozoan, and Coelenterate translation mitochondrial code table to ensure that the sequences were aligned in the correct reading frame and that no stop codons were present.

We used jModelTest (Posada & Crandall 1998) within Geneious to select the best evolutionary model. We estimated phylogenies with MrBayes (v3.2, Ronquist *et al.* 2012) and IQ-Tree 2 with 1000 bootstrap replicates (Minh *et al.* 2020) for all ctenophores available based on the most appropriate models selected by the AIC (Akaike 1974). Bayesian phylogenies estimated with MrBayes included multiple runs of  $5 \times 10^6$  generations with a 10% burn-in, with six chains, that were sampled and printed every 1000 generations. Convergence was determined with TRACER v1.7 (Rambaut *et al.* 2018) and by comparing topologies of multiple runs.

Phylogenies were visualized with FigTree (v1.4.4, tree.bio.ed.ac.uk/software/figtree/). To illuminate within-order diversity, alignments of lobate species for both *COI* and *18S* fragments were analyzed separately with the same parameters that had been used for the full alignment, and both Bayesian and likelihood support values were reported on the phylogeny. The lobate phylogenies were rooted with the cydippid *Pleurobrachia bachei* as the outgroup.

To assess diversity within sequenced ctenophores for *COI* we calculated percent general time reversible (GTR) distance within and between molecular operational taxonomic units (MOTUs). Base composition was calculated with MegaX (v10.0.5, Kumar *et al.* 2018) and pairwise distances were calculated in Geneious Prime. We calculated the number of parsimony-informative sites within the Lobata for both *COI* and *18S* with DnaSP (v6.12.01, Rozas *et al.* 2017). We tested for saturation of observed proportions of transitions and transversions versus GTR distance among all MOTUs with the software DAMBE7 (v7, Xia 2018). We also tested five other mitochondrial loci for saturation from the published mitochondrial genomes of eight species of ctenophores with DAMBE7.

To illustrate how a more complete reference library can affect metabarcoding sequence assignments, we used data for the *COI* fragment for ctenophores from Pitz *et al.* (2020) and queried our ctenophore-specific library with the same methods as the authors of the study. Data for stations where no ctenophore sequences were recovered were not reported. We plotted species assignments based on the presence or absence of reads in each sample in Rstudio (Team 2015) with ggplot2 (Wickham 2016).

## 3 | Results

We successfully amplified and sequenced 174 individuals of 67 distinct species. Phylogenetic analyses of the resulting *COI* sequences revealed high diversity and a number of cryptic species within all groups, but especially within Lobata, which had not been evident from previous *18S* rDNA sequence analyses. *COI* and *18S* sequence fragments were deposited in GenBank with accession numbers (xxx-xxx) (Table S1). Taxon-optimized primer sequences are indicated in Figure 1.

Analysis of multiple isolates of *Bolinopsis* suggested that the genus is polyphyletic. In addition, we confirmed morphological differences of distinct genera with *COI* sequence fragments, such as *Deiopea*, *Kiyohimea*, and *Eurhamphaea*, that were unresolved with the *18S* fragment, and found support for subtle morphological differences between undescribed species within the genera *Lampocteis* and *Bathocyroe*.

### 3.1 | Primer Design and Combinations

Successful primer pairs varied by taxon (Table S1, Figure 1, and branches of Figure 2). Primer combinations amplified a range of fragment lengths (Figure 1), so we trimmed the ends of the alignments used to generate phylogenies to exclude missing data. A few sets of primers such as F259/1060R worked well for many species,

and reverse primer R866 was successful for many species of cydippids (Figure 2, starred primers). However, amplification of many genera such as *Euplokamis*, *Ocyropsis*, *Beroe*, and *Lampea* required customized primers. Some primers designed for one species also worked for other closely related taxa, such as Bfor259F, which was specific to *B. forskalii* but also amplified most Beroids. For several species, multiple combinations of primer pairs successfully amplified the *COI* fragment (Figure 1, 2).

### 3.2 | Phylogenetics and Species Delimitation

The *COI* alignment included 174 sequences representing 67 MOTUs (molecular operational taxonomic units). It was trimmed to a 765 base-pair (bp) fragment and we used a GTR+i+ $\Gamma$  model for phylogenetic analyses based on jModelTest results (Figure 2). Mitochondrial loci in ctenophores are generally rich in the bases adenine (A) and thymine (T) (Pett *et al.* 2011). Average base composition for all ctenophores sequenced were  $\sim$  50% A, 21% T, 16% C, and 13% G, and the vast majority of changes were at the third codon position.

In order to examine saturation between and within taxa, we plotted the proportion of transitions and transversions versus GTR distance among all sequences (Figure 3a). For ctenophores sequenced, we found that the proportion of transitions was greater than that of transversions, which is in contrast to the general rule (for mammal mitochondrial DNA), where transversions are often two-fold more common than transitions (Xia *et al.* 1996). Transitions were saturated for species that were more than  $\sim$ 25% distinct and transversions were saturated for species that were more than  $\sim$ 30% distinct. We saw the same pattern for the other mitochondrial loci that we could align without the presence of stop codons including; *COII*, *COIII*, *CytB*, *ND4* and *ND5* (Figure S1). In these genes as well, transversions were more common than transitions and there were high levels of saturation.

The high levels of saturation between distantly related species (Figure 3a) were reflected in the poor resolution of deeper level relationships amongst taxa in the Bayesian and maximum likelihood phylogeny trees. Phylogenetic relationships for *COI* among all the ctenophores were therefore portrayed in an unrooted tree without support values (Figure 2). In order to illustrate levels of diversity within the phylum, we also plotted within and between pairwise GTR distances for all individuals sequenced (Figure 3b). In comparison to other phyla, distances between MOTUs were relatively large, ranging between from 0–43%, and showed four peaks. The first peak represented within-species variability, around 0–4% GTR, although we had few within-species samples for this estimate. The second small peak centered at 10% GTR distance, was represented by closely-related species complexes. The two largest peaks were represented by the majority of our sequencing efforts and occurred at  $\sim$ 17% (within group; pink, blue, or green) and 30% GTR distance among different orders (gray) of ctenophores (Figure 3b).

The *18S* alignment of the Lobata included 82 sequences for 29 MOTUs. This was trimmed to 1780 bp and had a TrN+i+ $\Gamma$  model determined with jModelTest for phylogenetic analyses. The *18S* fragment was very conserved and included only 47 parsimony-informative sites within all sequence data for the Lobata and phylogenetic analyses showed little differentiation within and even between many genera (Fig 4a). Posterior probabilities and bootstrap values were relatively poor and unresolved for most relationships among taxa due to low resolution among species and genera (Figure 4a). Distinguishing taxa using the *COI* fragment limited to the Lobata was more successful since the marker is more variable than *18S*. The *COI* fragment had 199 parsimony informative sites between 27 MOTUs (Figure 4b). Although deeper level relationships amongst more distantly related taxa were still not well supported and some saturation was evident (Figure 3a), analyses revealed high support for several cryptic species complexes and new MOTUs. Conversely, other species were revealed to be truly cosmopolitan with worldwide or at least ocean-basin wide distributions (Figure 4b).

### 3.3 | Metabarcoding ctenophores

In 2012 on a transit from Monterey Bay, CA, USA to the mouth of the Gulf of California, MX Pitz *et al.* (2020) sampled 15 offshore stations with 100 meter-depth vertical net tows which they sequenced for *18S* and *COI* metabarcoding fragments. The authors found 9105 *COI* sequences from thirteen of the stations that were assigned only to the family level (Pitz, pers. com). We used our newly assembled ctenophore sequence

library and assigned their fragments more precisely to ten different species rather than three families. Here, species identifications were reduced to a presence/absence matrix (Figure 6).

## 4 | Discussion

### 4.1 | Phylogenetic analyses

Newly-designed primers combined with a few existing primer pairs enabled us to amplify ~700-1000 base pairs of the mitochondrial *COI* fragment for 66 MOTUs from all groups of ctenophores (Figure 2). The *COI* phylogeny illuminated a great deal of species-level diversity that was not evident for the *18S* fragment, within and between groups. However, the levels of divergence and saturation for mitochondrial genes among ctenophores are some of the highest in the Metazoa (Lavrov & Pett 2016). Despite high levels of mitochondrial variation, within-species diversity of ctenophores rarely exceeded 4% (Figure 3b). Saturation within the phylum, especially at third codon positions, obscured many higher-level relationships and resulted in basal polytomies. Curiously, we found a higher proportion of transitions than transversions for the *COI* fragment among ctenophores (Figure 3a). For protein-coding loci, transitions are often more common than transversions because they usually result in synonymous mutations and involve the exchange of bases of similar shapes (Xia et al. 1996). It is plausible that the proportion of transitions within our data were overly saturated and additional mutations were obscured. The mitochondrial genomes of ctenophores are enriched with A/T residues, which also could contribute to a greater incidence of transversions (purines=pyrimidines) than transitions. Plots of other mitochondrial loci also revealed the same pattern, where transitions saturated quickly, often at ~20% GTR distance, and there were higher proportions of transversions (Figure S1). Despite high levels of saturation, closely related species and groups were well supported, especially within the Lobata (Figure 4b). Although saturation among divergent taxa resulted in poor phylogenetic resolution, and it was difficult to make strong conclusions based on the *COI* fragment alone, *COI* sequences were easily attributable to MOTUs, and proved useful for species delimitation.

### 4.2 | The Lobata

When the *COI* phylogeny was limited to the lobate ctenophores, which form a monophyletic clade (including *Thalassocalyce* and Cestids; Podar, 2001), phylogenetic resolution was improved, and the *COI* fragment revealed that many genera were polyphyletic (Figure 4b). One example; tropical species of *Bolinopsis mikado* (Japan), *B. vitrea* (Bahamas), *B. aff. vitrea* (Hawaii) and *B. ashleyi* (Australia) all formed a well-supported clade with *Mnemiopsis leidyi*. The temperate species *Bolinopsis infundibulum* (Atlantic Ocean) and a new *Bolinopsis* MOTU from the eastern Pacific were in a distinct and well-supported clade (Figure 4b; and Johnson *et al.* in prep). This was in contrast to the *18S* fragment where species within *Bolinopsis* genus were undifferentiated, with the exception of a few individuals of *B. aff. vitrea* collected near Hawaii and Moorea (Figure 4a). Clearly, more data from more localities, along with transcriptomes and genomes, are critical in understanding relationships among ctenophores, but the *COI* gene provides a useful starting point.

### 4.3 | Morphological and molecular concordance

Mitochondrial sequencing reinforced the designation of several undescribed species that returned one *18S* MOTU, but differed morphologically or ecologically (Figure 4). *Lampocteis cruentiventer* (Figure 5a) and *Lampocteis* ‘sp. V’ (Figure 5b), differed by ~10% GTR distance and were segregated by depth. A third suspected species within the genus was *Lampocteis* ‘sp. A,’ (Figure 5c) which are typically amber rather than a brilliant red color, although color is often a misleading trait. However, *COI* sequences for *Lampocteis* ‘sp. A,’ differed from the other two MOTUs by ~17% GTR distance, which gave strong support for a distinct species.

The *COI* fragment revealed three distinct *Bathocyroe* MOTUs collected from California, and one from the Gulf of California, Mexico, in contrast to the single MOTU that was returned by the *18S* fragment (Figs. 4a, b, 5d-f). There are three described species within the genus including *B. fosteri* (Madin & Harbison 1978) from the Gulf of Mexico, *B. paragaster* (Ralph & Kaberry 1950) from the South Western Pacific, and

*B. longigula* (Horita et al. 2011) from shallow waters in Japan. *Bathocyroe* aff. *fosteri* ‘A’ collected from California most closely matched the description of *B. fosteri* (Figure 5d). *Bathocyroe* aff. *longigula* (Figure 5f) from California resembled *B. longigula* somewhat, in having an elongated stomodaeum, but it lacked any colored spots along the meridional canals and the gut appeared wider than that of *B. longigula*. In addition, the *B. aff. longigula* from California was mainly found from ~2000–3000 meters depth and *B. longigula* in Japan was collected at the surface (although deep species are known to be occasionally upwelled to the surface in those waters.). In California, *B. aff. longigula* and *B. aff. fosteri* ‘A’ had overlapping depth ranges, distributions, but had small distinctions in gut shape (Figure 5d, f) and were the most divergent from one another molecularly (15% GTR). The third MOTU from California, *Bathocyroe* aff. *fosteri* ‘B’, had a distinct gut shape and color from all the other MOTUs, was relatively rare, and was only collected below 3000 meters depth (Figure 5e). Unfortunately, specimens from type localities were unavailable for sequencing so it is unclear if any of the MOTUs we sequenced match those already described or new to science. However, with our new primers researchers world-wide can now amplify species of *Bathocyroe* and help to confirm species identifications.

The *COI* fragment sequences also resolved the relationships between *Deiopea kaloktenota* (Chun 1880) and *Kiyohimea usagi* (Matsumoto & Robison 1992). It was first suspected that despite morphological distinctions, *Deiopea* might be a juvenile form of *Kiyohimea* (Matsumoto & Robison 1992). By sequencing the *COI* fragment we found two distinct lineages of *Deiopea* from California in addition to that of *K. usagi*. A single specimen collected offshore of the Hawaiian Islands also was the same MOTU as one of the *Deiopea* from California, showing this species has high dispersal capabilities and a broad oceanic distribution. Specimens of *D. kaloktenota*, which were described from the N. Atlantic, and *K. aurita* (Komai & Tokioka 1940) from Japan were unavailable for sequencing, so it is unclear whether one or both species of *Deiopea* from Monterey Bay are undescribed.

#### 4.4 | Population subdivision of *Ocyropsis*

The high levels of within-species diversity of Ctenophores make the *COI* fragment, and possibly mitochondrial sequencing in general, a good marker for population genetics. Mitochondrial sequencing revealed many morphologically cryptic but genetically distinct species. Within Lobata, phylogenetic relationships were better supported and many cryptic species complexes were revealed for those with both sympatric distributions and allopatric isolation (Figure 4b).

The genus *Ocyropsis* was a good example. Currently, *Ocyropsis* contains two robustly described species, each with two subspecies based on coloration, including *O. crystallina crystallina* and *O. maculata maculata*. These have sympatric distributions and were both described from the Atlantic Ocean (near Europe), but are thought to be distributed worldwide. Two other subspecies were described from the Gulf of Mexico, including *O. crystallina guttata* which we sequenced one specimen from Florida. The *COI* fragment for specimens of *O. maculata maculata* sequenced from Florida, Australia, and Tahiti were closely related and all represented one MOTU. In contrast, specimens of *O. crystallina crystallina* sequenced from Florida, the Gulf of California, and Hawaii, represented three distinct lineages. Unfortunately, specimens from other regions were not available for sequencing. Regardless, the *Ocyropsis* genus represents just one interesting dichotomy where one subspecies such as *O. crystallina crystallina* may have many genetic and morphological distinctions that lead to many named subspecies, yet *O. maculata maculata*, a closely related species within the genus has a nearly worldwide distribution (Johnson et al., in prep).

#### 4.5 | eDNA and metabarcoding

In order to explore biodiversity, ecosystem function, and population genetics, a high-resolution marker with a good reference library is critical. As of the writing of this manuscript there was *COI* data (including mitochondrial genomes) for ~eight species of ctenophores on GenBank and BOLD databases. The paucity of mitochondrial sequencing data is mostly a result of the failure of commonly used ‘barcoding’ primers to amplify the highly divergent phylum. In a study that took 100-meter depth vertical net tows at stations from Monterey Bay, CA to the Gulf of California, Mexico, Pitz *et al.*, (2020) successfully sequenced at least

some ctenophores for *COI* and *18S*, but was only able to assign taxonomy mostly to the family level. When net tow data were queried against sequences generated by the new ctenophore primers, we were able to assign thousands of sequences to ten distinct species of ctenophores, illuminating how diversity changed over two important biogeographic barriers, Point Conception and Punta Eugenia (Figure 6). Common primers only amplify a few groups of ctenophores and although we had successful taxonomic assignments, there were certainly more than ten species in the net-tow samples. With the addition of new primers and sequences, researchers are now able to redesign common primers to include the full spectrum of diversity of ctenophores in the world ocean.

In conclusion, the primers designed herein will enable researchers worldwide to amplify and sequence ctenophores for the mitochondrial *COI* locus for quick identification, as a population genetic marker, and for metabarcoding studies. With the publication of our sequences amplified from our new primers, researchers can increase our understanding of processes large and small in the world ocean. As we continue to sequence a broader diversity of ctenophores, new sequence information will help with species identification and descriptions, and will also provide a better understanding of relationships within Ctenophora.

### Author Contributions

Lynne M. Christianson designed the study, primers, collected, amplified and sequenced most of the ctenophores, and helped draft the manuscript. Shannon B. Johnson helped to amplify and sequence some of the ctenophores, performed statistical analyses, and drafted the manuscript. Darrin T. Schultz provided valuable specimens, transcriptomic data, mitochondrial sequences, and edited the manuscript. Steven H.D. Haddock helped conceive the study, provided specimens, funding, transcriptomic data, performed statistical analyses and helped to draft the manuscript.

### Data Accessibility

*COI* and *18S* sequence fragments were deposited in GenBank with accession numbers (xxx-xxx) (Table S1). Primer sequences are indicated in Figure 1.

### Acknowledgements

We gratefully acknowledge the crew and pilots of the *R/V Western Flyer* and *ROV Doc Ricketts*, *R/V Rachel Carson* and *ROV Ventana* for their help with collecting these delicate specimens and assistance with blue-water diving. We would like to thank Warren Francis for generating many of our transcriptome data and for providing helpful comments on the manuscript. April Woods and Maile Prickett were instrumental in PCR and sequencing help, and we thank Kyra Schlining and Susan Von Thun for support at sea. We would also like to thank Jacob Winnikoff who created some of the silhouettes for figure 2, helped collect specimens, and for his constructive comments on the manuscript. Numerous specimens were donated by Bruce Robison, Rob Sherlock, Kristine Walz, David Clague, Lonny Lundsten, William Browne, Stacy Kim, Keith Bayha, Larry Madin, Wyatt Patry (Monterey Bay Aquarium), California Academy of the Sciences, and Ocean Genome Legacy Center to increase our sampling breadth. We also graciously acknowledge Katie Pitz, who provided ctenophore sequences from her metabarcoding library. This work was funded by the David and Lucile Packard Foundation, the National Science Foundation (NSF-DEB 1542679 to SHDH) and the National Institutes of Health (R01-GM087198).

### References

Akaike H (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**, 716-723. Arafat H, Alamaru A, Gissi C, Huchon D (2018) Extensive mitochondrial gene rearrangements in the Ctenophora: insights from benthic Platyctenida. *BMC Evolutionary Biology* **18**, 11. Choy CA, Haddock SHD, Robison BH (2017) Deep pelagic food web structure as revealed by *in situ* feeding observations. *Proceedings of the Royal Society B: Biological Sciences* **284**, 20172116. Chun C (1880) Die Ctenophoren des Golfes von Neapel und der angrenzenden Meeres-Abschnitte. *Flora und Fauna des Golfes von Neapel*, **1** Engelmann, Leipzig. 1-311 DeSalle R, Goldstein P (2019) Review and interpretation of trends in DNA



barcoding. *Frontiers in Ecology and Evolution* **7**, 302. Djurhuus A, Pitz K, Sawaya NA, *et al.* (2018) Evaluation of marine zooplankton community structure through environmental DNA metabarcoding. *Limnology and Oceanography: Methods* **16**, 209-221. Dunn CW, Leys SP, Haddock SHD (2015) The hidden biology of sponges and ctenophores. *Trends in Ecology and Evolution* **30**, 282-291. Eble JA, Daly-Engel TS, DiBattista JD, Koziol A, Gaither MR (2020) Marine Environmental DNA: Approaches, applications, and opportunities. In: *Advances in Marine Biology* (ed. Sheppard C), pp. 141-169. Elsevier Ltd., London, United Kingdom. Edgar R (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792-1797. Folmer O, Black MB, Hoeh WR, Lutz RA, Vrijenhoek RC (1994) DNA primers for amplification of mitochondrial cytochrome C oxidase subunit I from metazoan invertebrates. *Molecular Marine Biology and Biotechnology* **3**, 294-299. Francis WR, Christianson LC, Kiko R, *et al.* (2013) A comparison across non-model animals suggests an optimal sequencing depth for *de novo* transcriptome assembly. *BMC Genomics* **14**, 167. Geller J, Meyer C, Parker M, Hawk H (2013) Redesign of PCR primers for mitochondrial cytochrome *c* subunit I for marine invertebrates and application in all-taxa biotic surveys. *Molecular Ecology Resources* **13**, 851-861. Günther B, Kneibelsberger T, Neumann H, Laakmann S, Arbizu PM (2018) Metabarcoding of marine environmental DNA based on mitochondrial and nuclear genes. *Scientific Reports* **8**, 14822. Haddock SHD (2004) A golden age of gelata: past and future research on planktonic ctenophores and cnidarians. *Hydrobiologica* **530/531**, 549-556. Haddock SHD, Christianson LM, Francis WR, *et al.* (2017) Insights into the biodiversity, behavior, and bioluminescence of deep-sea organisms using molecular and maritime technology. *Oceanography* **30**, 39-47. Harbison GR, Madin LP, Swanberg NR (1978) On the natural history and distribution of oceanic ctenophores. *Deep-Sea Research* **25**, 233-256. Hoang D, Chernomor O, von Haeseler A, Minh B, Vinh L (2018) UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* **35**, 518-522. Johnson SB, Winnikoff JR, Schultz DT, *et al.* (in prep) Speciation in the midwater; a new lobate ctenophore in the eastern Pacific. Kohn AB, Citarella MR, Kocot KM, *et al.* (2012) Rapid evolution of the compact and unusual mitochondrial genome in the ctenophore, *Pleurobrachia bachei*. *Molecular Phylogenetics and Evolution* **63**, 203-207.

Komai T, Tokioka T (1940) *Kiyohimea aurita*, n. gen., n. sp., type of a new family of lobate Ctenophora. *Annot. Zool. Japan* **19**, 43-46.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution* **35**, 1547-1549. Lacoursière-Roussel A, Howland K, Normandeau E, *et al.* (2018) eDNA metabarcoding as a new surveillance approach for coastal Arctic biodiversity. *Ecology and Evolution* **8**, 7763-7777. Lavrov DV, Pett W (2016) Animal mitochondrial DNA as we do not know it: mt-genome organization and evolution in nonbilaterian lineages. *Genome Biology and Evolution* **8**, 2896-2913. Leray M, Knowlton N (2015) DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proceedings of the National Academy of Sciences* **112**, 2076-2081. Leray M, Yang JY, Meyer CP, *et al.* (2013) A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Frontiers in Zoology* **10**, 34. Lindsay DJ, Miyake H (2007) A novel benthopelagic ctenophore from 7,217 m depth in the Ryukyu Trench, Japan, with notes on the taxonomy of deep-sea cydippids. *Plankton and Benthos Research* **2**, 98-102. López-Escardó D, Paps J, C. dV, *et al.* (2018) Metabarcoding analysis on European coastal samples reveals new molecular metazoan diversity. *Scientific Reports* **8**, 9106.

Madin LP, Harbison GR (1978) *Bathocyroe fosteri* gen.nov., sp.nov.: a mesopelagic ctenophore observed and collected from a submersible. *Journal of the Marine Biological Association of the United Kingdom* **58**, 559-564.

Matsumoto GI, Robison BH (1992) *Kiyohimea usagi*, a new species of lobate ctenophore from the Monterey submarine canyon. *Bulletin of Marine Science* **51**, 19-29. Medlin LK, Elwood HJ, Stickel S, Sogin ML (1988) The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. *Genetica* **71**, 491-499. Mills CE (2017) *Phylum ctenophora: list of all valid species names*. <https://faculty.washington.edu/cemills/Ctenolist.html> Minh B, Schmidt H, Chernomor O, *et al.* (2020) IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology*

*and Evolution* **37**, 1530-1534. Nguyen L, Schmidt H, von Haeseler A, Minh B (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Molecular Biology and Evolution* **32**, 268-274. Pett W, Ryan JR, Pang K, *et al.* (2011) Extreme mitochondrial evolution in the ctenophore *Mnemiopsis leidyi*: insights from mtDNA and the nuclear genome. *Mitochondrial DNA* **22**, 130-142. Pitz K, Guo J, Johnson SB, *et al.* (2020) Zooplankton biogeographic boundaries in the California Current System as determined from metabarcoding. *Plos One* **15**, e0231519-0231520. Podar M, Haddock SHD, Sogin ML, Harbison GR (2001) A molecular phylogenetic framework for the phylum Ctenophora using 18S rRNA genes. *Molecular Phylogenetics and Evolution* **21**, 218-230. Posada D, Crandall KA (1998) Modeltest: Testing the model of DNA substitution. *Bioinformatics* **14**, 817-818. Preston CM, Durkin CA, Yamahara KM (2020) DNA metabarcoding reveals organisms contributing to particulate matter flux to abyssal depths in the North East Pacific ocean. *Deep-Sea Research Part II* **173**, 104708. Rambaut A, Drummond A, Xie D, Baele G, Suchard MA (2018) Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology* **syy032**. Ronquist F, Teslenko M, van der Mark P, *et al.* (2012) MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61**, 539-542. Rozas J, Ferrer-Mata A, Sánchez D, JC, *et al.* (2017) DnaSP 6: DNA sequence polymorphism analysis of large datasets. *Molecular Biology and Evolution* **34**, 3299-3302. Schroeder A, Stanković D, Pallavicini A, *et al.* (2020) DNA metabarcoding and morphological analysis - Assessment of zooplankton biodiversity in transitional waters. *Marine Environmental Research* **160**, 104946. Schultz DT, Eizenga JM, Corbett-Detig RB, *et al.* (2020) Conserved novel ORFs in the mitochondrial genome of the ctenophore *Beroe forskalii*. *PeerJ* **8**. Siddall ME, Fontanella FM, Watson SC, Kvist S, Erséus C (2009) Barcoding bamboozled by bacteria: convergence to metazoan mitochondrial primer targets by marine microbes. *Systematic Biology* **4**, 445-451. Simion P, Bekkouche N, Jager M, Quéinnec E, Manuel M (2015) Exploring the potential of small RNA subunit and ITS sequences for resolving phylogenetic relationships within phylum Ctenophora. *Zoology* **118**, 102-114.

Ralph PM, Kaberry C (1950) New Zealand Coelenterates. Ctenophores from Cook Strait. *Zoology Publications from Victoria University College* **3**, 1-11.

R Team (2015) RStudio: Integrated Development for R. RStudio Inc, Boston, MA. Thiebot J-B, McInnes JC (2020) Why do marine endotherms eat gelatinous prey? *ICES Journal of Marine Science* **77**, 58-71. Trifinopoulos J, Nguyen L, von Haeseler A, Minh B (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analyses. *Nucleic Acids Research* **44**, W232-W235. Untergasser A, Cutcutache I, Koressaar T, *et al.* (2012) Primer3—new capabilities and interfaces. *Nucleic Acids Research* **40**, e115. Vrijenhoek RC, Waples RS (2012) Popular misconceptions. *Molecular Ecology* **21**, 4155-4156. Wickham H (2016) *ggplot2: Elegant graphics for data analysis* Springer-Verlag, New York. Xia X (2018) DAMBE7: New and improved tools for data analysis in molecular biology and evolution. *Molecular Biology and Evolution* **35**, 1550-1552.

Xia X, Hafner MS, Sudman PD (1996) On transition bias in mitochondrial genes of pocket gophers. *Journal of Molecular Evolution* **43**, 32-40.

Yeh HD, Questel JM, Maas KR, Bucklin A (2020) Metabarcoding analysis of regional variation in gut contents of the copepod *Calanus finmarchicus* in the North Atlantic Ocean. *Deep-Sea Research Part II*.

## Figure Legends

FIGURE 1(a) Primer binding position, taxa amplified, and PCR reaction conditions for each primer set relative to published *Mnemiopsis leidyi* sequence, including the ‘Folmer’ region (b) primer sequences and position used to amplify and sequence all groups of ctenophores including LCO1490/HCO2198 (Folmer *et al.* 1994) and meta-COI-F2/meta-COI-R2 (Pett *et al.* 2011). Differences from *M. leidyi* are indicated in bold and colored according to base. Primers marked with \* are more universal.

FIGURE 2 Bayesian unrooted phylogeny for 765 bp *COI* fragment for all ctenophores sequenced with successful primer pairs indicated by colored triangles. Closed triangles indicate forward and open indicate reverse direction primers. Multiple triangles indicate multiple successful primer pairs. Branches colored by

group including; Lobates (pink), Beroids (blue), and Platyctenes (green).

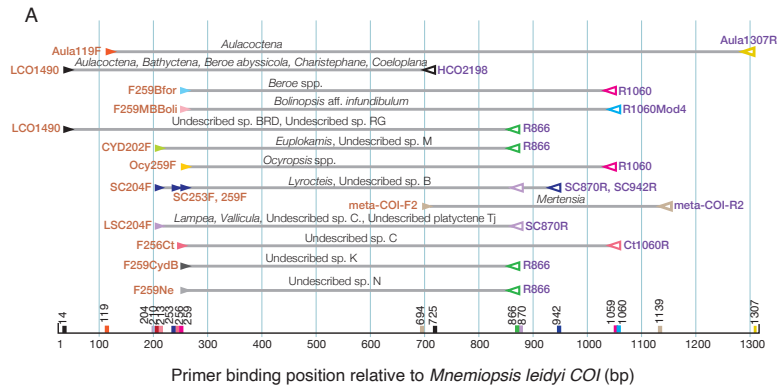
FIGURE 3(a) Transitions and transversions versus GTR sequence divergence and (b) Pairwise percent sequence distance (GTR) within and amongst all ctenophore species sequenced for the *COI* fragment.

FIGURE 4 Bayesian and maximum likelihood estimated phylogenies for fragments of (a) *18S* (1780 bp, TrN+i+ $\Gamma$ ) and (b) *COI* (765 bp, GTR+i+ $\Gamma$ ) limited to Lobate ctenophores. *Bolinopsis* species are bolded, showing the scattered distribution with *18S*. Colored backgrounds indicate the species of *Lampocteis* which are identical with *18S* but form distinct clades with *COI*. Both likelihood and Bayesian support values are indicated by triangles on the nodes, with unlabeled nodes having low support. Locations (Table S1) are indicated in the blue labels including: (AU) Australia, (BA) Bahamas, (CA) California, (FL) Florida, (GC) Gulf of California, (HI) Hawaii, US, (JP) Japan, (NO) Norway, (TA) Tahiti, (WA) Washington, (WH) Woods Hole

FIGURE 5 Images and known depths in meters of specimens of (a) *Lampocteis cruentiventer* (~250–1500 m) (b) *Lampocteis* sp. ‘V’ (~1500–3000 m), (c) *Lampocteis* sp. ‘A’, (~500–2800 m), (d) *Bathocyroe* aff. *fosteri* ‘A’ (560–3572 m, small red gut), (e) *Bathocyroe* aff. *fosteri* ‘B’ (> 3000 m, peach gut), (f) *Bathocyroe* aff. *longigula* (1900–2858 m, big gut)

FIGURE 6 Presence and absence matrix of assigned ctenophore species for a 311 bp fragment of *COI* from 100 meter-depth net tows collected in 2012 along a transect that spanned from Central California, USA, to Southern Baja California, Mexico from Pitz *et al.* (2020).

SUPPLEMENTARY FIGURE S1 Transitions and transversions versus GTR sequence divergence for *Beroe cucumis* (MK361035), *Beroe forskalii* (MG655623), *Coeloplana loyai* (LN898113), *Coeloplana yulianicorum* (LN898115), *Hormiphora californensis* (MN544300), and *Mnemiopsis leidyi* (JF760210) for *COI* (1311 bp), *COII* (536 bp), *COIII* (601), *CytB* (570 bp), *ND4* (429 bp), and *ND5* (506 bp).



**B**

5' FORWARD 3'		5' REVERSE 3'	
<i>M. leidyi</i>	14 TTTCTACTAATCACAAAGATATAGC 38	<i>M. leidyi</i>	725 TAAACTTCTGGATGGCCAAAAATCA 700
*LCO1490	▶ GGTCAACAATAATCAATAAGATATGG	*HCO2198	◀ TAAACTTCAAGGTGCCAAAAATCA
<i>M. leidyi</i>	102 GCGTTTAGCTTATAGTTGACCTTTTGC 119	<i>M. leidyi</i>	870 CATATGATGGGCTCAAACATAACACCT 843
Aula119F	▶ GCGTCTTCAATTAGTATGACCTTATTC	*R866	◀ CATRTGRTGNGCYCANACNANAC
<i>M. leidyi</i>	202 ATGCCTTTTACTATAGGTGGTTTATC 227	SC870R	◀ GATGAGCTCAAACAATACAACCT
Cyd202F	▶ ATGCCTTTTWSHWTWGGNGGNTT	<i>M. leidyi</i>	942 AGTAGGAATAGCTATGATTAAGGTAGC 913
SC204F	▶ GCCTTTTCTATGGTGGTTTAGG	SC942R	◀ GGTAGGAACCTGCAATAACTAAAGTAGC
LSC204F	▶ GCCWTTTTCATGCGNGWTRGG	<i>M. leidyi</i>	1060 TACCGGATAGACCTCCAAATGT 1039
<i>M. leidyi</i>	253 CATTGGCTGATATGTTTACCTCGTAT 281	*R1060	◀ ACCRGGAYARGCCACAAAAGT
*F259	▶ GCWGATATGTTTACCTYMG	*R1060Mod4	◀ TWCCAGAYARRCCWCCAAAAGT
F259Bfor	▶ GCGATATGTCYVINCIMMG	Ct1060R	◀ ACCAGTTACACCACCAACTGT
F259MBBoli	▶ GCGGATATGCTTCCACCG	<i>M. leidyi</i>	1139 GATAAAACGTAATGAAAATGAGC 1117
F256C1	▶ GCTCTATGATATGGTTTACCTAG	*meta-COI-R2	◀ GAKAAKACGTAGTRAAARTGNCG
F259CydB	▶ CCTGATATGTTTACCTAG	<i>M. leidyi</i>	1307 CGGGGTAACCTCATAAAC 1289
F259Ne	▶ CTTGATATGTTTCCCTAG	Aula1307R	◀ CGAGGTAACCTCATAAAC
F259Ocy	▶ CTTGATATGTTTCCCTAG		
SC259F	▶ GCTGACATGGTCTTCCACCGAGT		
SC253F	▶ GGTGCTGCTGACATGGTCTTCC		
<i>M. leidyi</i>	694 TTATTTTGATTTTTGGCCATCC 716		
*meta-COI-F2	▶ TTKTTYTGRRTTYTYGGNCAATCC		

