OCEANOGRAPHY

An in situ digital synthesis strategy for the discovery and description of ocean life

John A. Burns[1]*, Kaitlyn P. Becker[2], David Casagrande[3], Joost Daniels[4], Paul Roberts[4], Eric Orenstein[4], Daniel M. Vogt[2], Zhi Ern Teoh[5], Ryan Wood[5], Alexander H. Yin[3], Baptiste Genot[1], David F. Gruber[6]*, Kakani Katija[4]*, Robert J. Wood[2]*, Brennan T. Phillips[3]*

Revolutionary advancements in underwater imaging, robotics, and genomic sequencing have reshaped marine exploration. We present and demonstrate an interdisciplinary approach that uses emerging quantitative imaging technologies, an innovative robotic encapsulation system with in situ RNA preservation and next-generation genomic sequencing to gain comprehensive biological, biophysical, and genomic data from deep-sea organisms. The synthesis of these data provides rich morphological and genetic information for species description, surpassing traditional passive observation methods and preserved specimens, particularly for gelatinous zooplankton. Our approach enhances our ability to study delicate mid-water animals, improving research in the world's oceans.

INTRODUCTION

The deep marine environment represents one of the lesser-known and more difficult to study ecosystems on Earth, with better maps for the surface of Mars than Earth's deep oceans (*1*). While humans have been able to access various terrestrial ecosystems for millennia, it has only been several hundred years since humans have had the capability to explore and study the ocean, beyond shallow ecosystems. Most contemporary technological advancements that facilitate interactions with deep-sea life, such as remotely operated vehicles (ROVs) equipped with titanium manipulators, have been propelled by the industrial sector and defense-related industries (*2*). However, in recent years, advances in fields such as soft robotics have focused more on the delicate manipulation of fragile marine life, developed in collaboration with and for the benefit of marine biologists (*3*–*8*). This, combined with advanced underwater imaging technology (*9*) has expanded biological inquiry that was previously only possible in controlled laboratory settings to delicate animals of the deep sea.

In this study, we combine the cross-disciplinary synergy of robotics, deep-sea specimen encapsulation, quantitative three-dimensional (3D) imaging, and molecular biology to collect rich data that can be used to identify, describe, and further understand deep-sea organisms. By combining underwater imaging and mobile robotic platforms, we unlock new mechanisms to achieve quantitative observations of deep-sea marine biota (*10*). We report a workflow and technology suite that includes structured imaging, encapsulation, in situ preservation, and genomic sequencing to provide a wealth of information about organismal systems.

This project involved two research expeditions onboard the Schmidt Ocean Institute's R/V *Falkor* with the ROV *SuBastian*, a 4500-m rated work-class ROV system. The first of these expeditions took place on 12 to 17 October 2019 off the west coast of Oahu, HI and the second on 12 to 21 August 2021 off the coast of San Diego, CA. The 2019 expedition focused on technical testing, while the 2021 expedition combined imaging and biological in situ investigation. During the integrated expedition in August 2021, a total of seven ROV dives were conducted over the course of a 10-day period off the coast of San Diego, CA (fig. S1), with details of each dive station listed in table S1. Each dive averaged ~8 hours in length, with the launch taking place in the mid-afternoon local time and recovery toward midnight to observe the upward diel migration pattern of ocean animals (*11*).

The ROV SuBastian was outfitted with three innovative midwater exploration systems consisting of the deep particle image velocimetry (DeepPIV) laser imaging system mounted in the ROV sled and extended out from the vehicle body, the EyeRIS (remote imaging system) plenoptic (light field) camera system grasped in one of the submersible's movable manipulator arms, and the rotary-actuated dodecahedron (RAD-2) for specimen encapsulation and tissue sampling grasped in the other manipulator arm (Fig. 1). These systems were operated near simultaneously during each ROV dive, with target organisms identified visually using the ROV science camera during forward-moving midwater transects.

When a biological organism of interest was encountered, we first collected in situ morphological data using ROV science camera footage in 4K, followed by quantitative imaging using the EyeRIS and/or DeepPIV (*12*, *13*) imaging systems. After imaging was complete, we captured the animal using the RAD-2, and collected animal tissues using the RAD-2 tissue sampling apparatus coupled with a McLane Labs suspended particulate rosette (SuPR) sampler (*14*) that used a high-flow pump to pull tissue samples onto a filter within a deep-sea fluid sampling rosette and then flood the filter with preservative from a reservoir mounted on the ROV.

A shipboard-mounted fisheries echosounder system (EK60) was used to identify layers of high biodensity before the start of the ROV dive and monitored throughout to guide exploratory navigation. While a total of 65 tissue and water samples were collected with RAD-2 during the course of the expedition, not every sampled specimen was considered appropriate for the imaging, and sampling systems and various technical considerations led to trade-offs and compromises with each of these emerging technologies.

[1]Bigelow Laboratory for Ocean Sciences, East Boothbay, ME 04544, USA. [2]School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA. [3]Department of Ocean Engineering, University of Rhode Island, 215 South Ferry Road, Narragansett, RI 02882, USA. [4]Monterey Bay Aquarium Research Institute, Research and Development, Moss Landing, CA 95039, USA. [5]PA Consulting, Concord, MA 01742, USA. [6]Department of Natural Sciences, Baruch College, City University of New York, New York, NY 10010, USA.
*Corresponding author: Email: jburns@bigelow.org (J.A.B.); david.gruber@baruch.cuny.edu (D.F.G.); kakani@mbari.org (K.K.); rjwood@seas.harvard.edu (R.J.W.); brennanphillips@uri.edu (B.T.P.)

Large target organisms, such as giant larvaceans and long siphonophore colonies, remain challenging to image and sample, but a wide range of sizes and morphologies can still be handled by the ROV's technical setup. Our impression is that our presented workflow is viable and efficient and can be improved upon by physical consolidation of the imaging and sampling systems and further practice in the field in less-explored midwater regions.

We present examples of rich quantitative imaging and full genome and/or transcriptome sequencing data for four representative deep-sea midwater species attained using our approach. By combining emerging state-of-the-art imaging with innovative capture and preservation techniques and genetic analyses, we create a comprehensive synthesis of an animal in its natural environment. In situ digital representation is particularly important for gelatinous zooplankton, whose delicate morphology is not usually preserved well in physical specimens (*15*–*17*).

RESULTS

An average of five selected target specimens were encountered, imaged, and sampled during each 8-hour ROV dive using the ROV SuBastian (Schmidt Ocean Institute), which is depth-rated to 4500 m. The total time of specimen visualization (close approach) to 3D imaging with DeepPIV and/or EyeRIS and to encapsulation and tissue sample preservation with the RAD-2 was often less than 20 min in total, and the time from tissue sampling to full in situ preservation was often less than 2 min per sample.

Quantitative in situ imaging of deep-sea animals captures their internal and external morphology

Imaging data were collected using DeepPIV and EyeRIS across numerous individuals in groups representing Amphipoda, Annelida, Cnidaria, Ctenophora, Echinodermata, Mollusca, Porifera, Radiozoa, Tunicata, and Vertebrata. Sizing data were collected on 61 different individuals using EyeRIS; 3D scan data were collected on 49 different individuals using DeepPIV (deep-sea) (table S2). Morphological measurements were conducted using EyeRIS data on four individuals with corresponding genetic data, two siphonophores, one salp, and one tomopterid (Fig. 2 and table S3). Siphonophore features included nectosome length, number of nectophores, and the lengths of mature and juvenile nectophores; salp features included body length, dorsal midline length, gut width, and the atrial and oral siphon widths; polychaete worm features included body length, parapodia count, and lengths of representative parapodia from anterior-most (P1) to posterior-most (P20) (table S4).

On the basis of its smooth, thick test, stolon wrapped around the gut, and two sets of muscle bands in an "X" configuration, the solitary salp specimen, RAD2-039, is identified as *Pegea* sp. (*18*). Its molecular

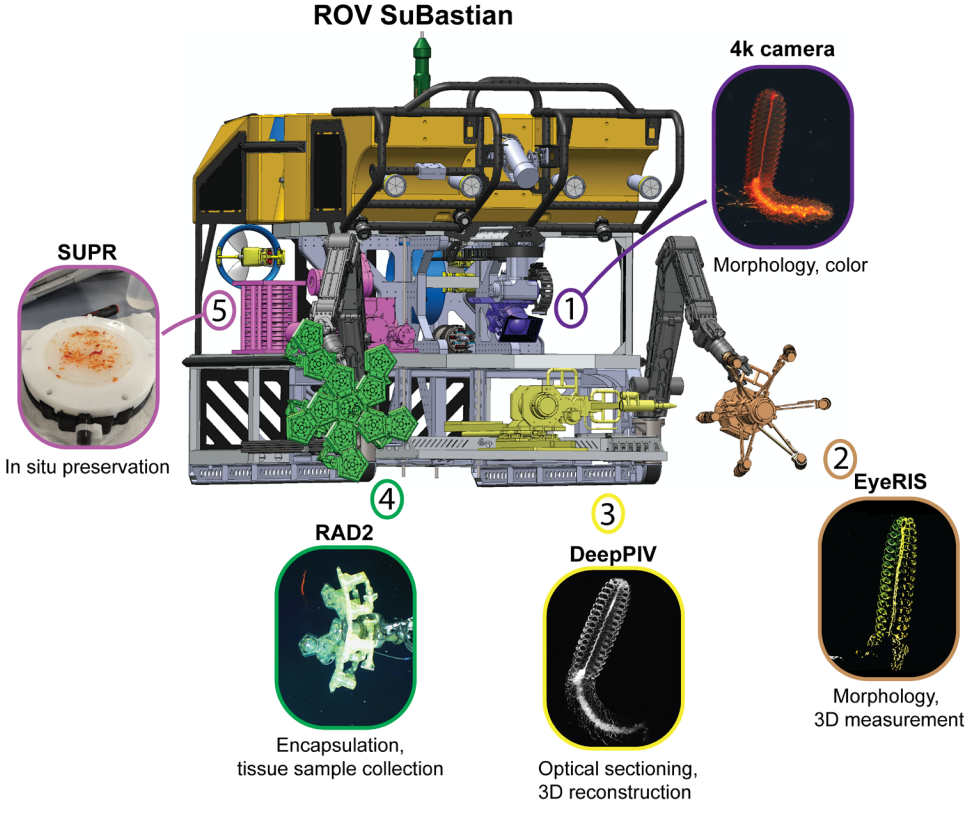

Fig. 1. An ROV-based digital synthesis strategy for in situ imaging and capture of midwater animals. ROV SuBastian's technical layout highlighting the position of our instruments in the manipulator arms and ROV sled and an example workflow for in situ digital synthesis of animal specimens for the August 2021 expedition. Imaging/sampling steps are color-coded and enumerated in the typical order of data collection with example data for one animal, *M. claudanielis*, shown for each instrument. (1) Initial specimen observations and video/still imagery using ROV 4 k scientific camera; (2) EyeRIS plenoptic light-field images; (3) DeepPIV laser imaging scans; (4) RAD-2 tissue sampling; and (5) in situ preservation. Genetic sequencing was completed on land following DNA and RNA extraction from in situ preserved tissues. ROV model by J. Williams of the Schmidt Ocean Institute.

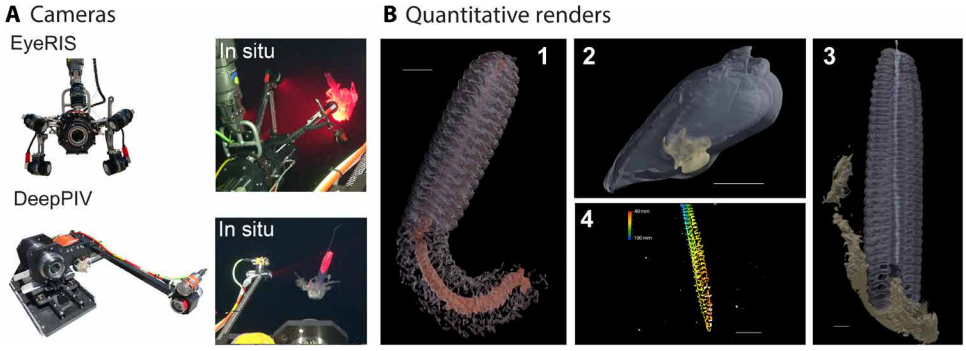

Fig. 2. Quantitative imaging using the custom imaging systems EyeRIS and DeepPIV. Quantitative imagery of highlighted specimens used for genetic analysis. (**A**) The two cameras that were mounted on ROV SuBastian and their operation in situ. EyeRIS was mounted in a manipulator arm, and DeepPIV was mounted to the center of the ROV deck. (**B**) 2D images of quantitative renders for four representative specimens. (1) Siphonophore *M. claudanielis*, (2) salp *Pegea* sp., and (3) siphonophore *Erenna* sp. reconstructed using DeepPIV laser sheet images. 4) Polychaete *Tomopteris* sp. imaged using the EyeRIS plenoptic light-field camera, with scale bar shown in top left. Video fly-arounds of each specimen are available in the data linked through table S13. In situ images in (A) captured by the 4K science camera on the ROV SuBastian, R/V *Falkor*, Schmidt Ocean Institute. 3D models in (B) generated by the Katija laboratory. Scale bars, 2 cm, do not represent depth or skew of animals or renders.

similarity to *Pegea confederata* (95% identity to 18*S*) and *Pegea bicaudata* (96% identity to 18*S*) (table S5) supports this identification to the genus level. We encourage further expert analysis of our data for more a precise description and identification of the specimen.

On the basis of identifying features including a segmented body plan, parapodia, swimming behavior, the presence of long tentacular (acicular) cirri, and near genetic identity (99.7%) to tomopterid voucher specimen *Tomopteris* sp. USNM IZ 1450332 (table S5), also captured in the Eastern Pacific, we identify RAD2-055 as *Tomopteris* sp. (*19*) and encourage additional expert examination of our data for species-level identification. We identify the siphonophore specimen RAD2-064 as *Marrus claudanielis* (*20*) based on several factors: its capture location and depth comparable to that of the holotype, a distinct deep red coloration, total nectophore count of 26 (table S4), closely packed bracts, and a 100% genetic match to the mitochondrial 16*S* sequence of *M. claudanielis* (table S5).

We identify siphonophore specimen RAD2-065 as *Errena* sp. (*21*) based on morphological measurements (table S4), flattened nectophores, distinctive pronounced arrays of tentilla, and genetic identity to 18*S* sequences of voucher specimens of *Erenna richardi* isolate T751-D3 (table S5). We refrain from identifying it to the species level due to the need for additional expert examination of its features and location.

Of the 49 individual animals scanned using DeepPIV, data were evaluated on the basis of a number of considerations (*12*), and 21 3D models (e.g., siphonophores, medusae, and tunicates) were generated. Example image system and 3D model data for the siphonophore *M. claudanielis* can be found in movie S1.

Encapsulation of planktonic animals enables tissue collection and in situ preservation

This study leverages recent advancements in deep-sea specimen encapsulation and presents an update on a self-folding polyhedron encapsulation mechanism called the RAD (*22*). The design, inspired by origami, has a single rotational input that transforms the hinged faces of a polyhedron into a ball-like structure that quickly encapsulates delicate specimens into a constrained volume (Fig. 3). Once encapsulated, in situ biological experimentation is possible inside the device while still in the open-water environment.

Expressed reference transcriptomes for four deep-sea animals provide new resources for evolutionary and ecological inference

In situ preservation of samples allowed extraction and sequencing of high-quality reference transcriptomes for all four specimens using the long-read Isoform-Sequencing (Iso-Seq) method, which captures full-length transcripts. The number of unique transcripts sequenced from each organism ranged from 47,425 (RAD2-065, siphonophore *Erenna* sp.) to 75,721 (RAD2-064, siphonophore *M. claudanielis*). Benchmarking universal single-copy ortholog (BUSCO) completeness scores for each ranged from detecting 63.0% of metazoan BUSCOs for RAD2-064 (siphonophore, *M. claudanielis*) to detecting 90.3% of metazoan BUSCOs for RAD2-039 (salp, *Pegea* sp.). The combined Iso-Seq and short-read assembly of the tomopterid RAD2-051 captures 98.9% of metazoan BUSCOs (fig. S2). Marker gene analyses link the specimens encountered in this study to voucher specimens, confirming or supporting reported identifications (table S5). Statistics for each transcriptome assembly can be found in table S6.

The salp *Pegea* sp. genome and quantitative transcriptome reveal insights into its structural composition

We report a high-quality draft assembly of the salp *Pegea* sp. genome. The genome size is 106 mega–base pairs (Mbp), assembled from 389 scaffolds with an N50 of 840 kilo–base pairs (kbp) (table S7). Structural annotation of the genome predicts 21,615 transcripts, which encode 10,464 predicted peptides. The genome of *Pegea* sp. is smaller than the only other reported genome from a salp, *Salpa thompsoni*, which has an estimated genome size of 602 Mbp and an assembled genome size of 319 Mbp (*23*). Notably, a high BUSCO score of 89.3% detected in the *Pegea* sp. genome is indicative of its functional completeness. This assessment is further supported by robust assembly statistics, including a high N50 (840 kbp) and N90 (150 kbp), as well as a low number of assembled scaffolds (389), all of which are indicative of the genome's high contiguity (Fig. 4).

A quantitative analysis of expressed transcripts enabled by in situ preservation of the specimen allowed identification of highly expressed processes in *Pegea* sp. at the time of capture. Considering the top 10% of expressed transcripts (1088 transcripts) from replicate tissue samples ($n = 2$), we see expected enrichment in protein

translation and energy metabolism (table S8). A notable highly expressed biological process from this soft-bodied organism was metabolism of hyaluronan (table S8). The genes described for hyaluronan metabolism in *Pegea* sp. all encode proteins with a hyaluronan binding domain, and homologs are associated with mediating binding between extracellular matrix proteins and hyaluronan to regulate hyaluronan synthesis, degradation, and localization. Hyaluronan is a high–molecular weight polysaccharide usually associated with the extracellular matrix of cells and animal connective tissues (*24*). Hyaluronan forms a structural hydrogel with other extracellular matrix components such as collagen, with transcripts encoding many collagens among the most highly expressed genes in *Pegea* sp. (Fig. 4).

Despite these results, *Pegea* sp. does not encode genes homologous to known animal hyaluronan synthases; however, its genome

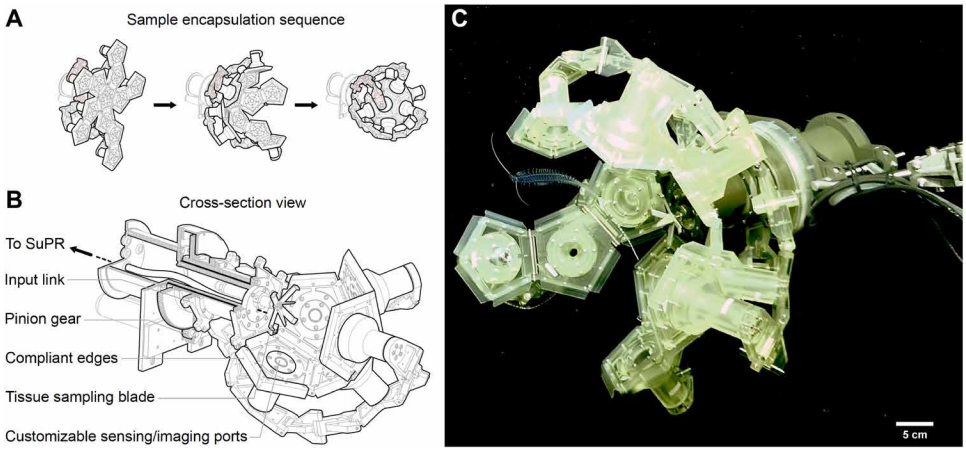

Fig. 3. An improved RAD sampler for encapsulation of pelagic animals. The RAD2 Sampler: (**A**) The RAD2 sampler geometry shown in full open, partial close, and full close positions. (**B**) A cross section view of RAD-2 showing major components integrated to achieve tissue cleaving and sample collection to the SuPR sampler. (**C**) The RAD2 system in operation, demonstrating a partial encapsulation of a tomopterid during an ROV dive. Image in (C) captured by the 4K science camera on the ROV SuBastian, R/V *Falkor*, Schmidt Ocean Institute. Scale bar, 5 cm, based on the approximate focal plane of the tomopteris and does not represent size or skew of other objects in the photograph due to their different distances to the camera.

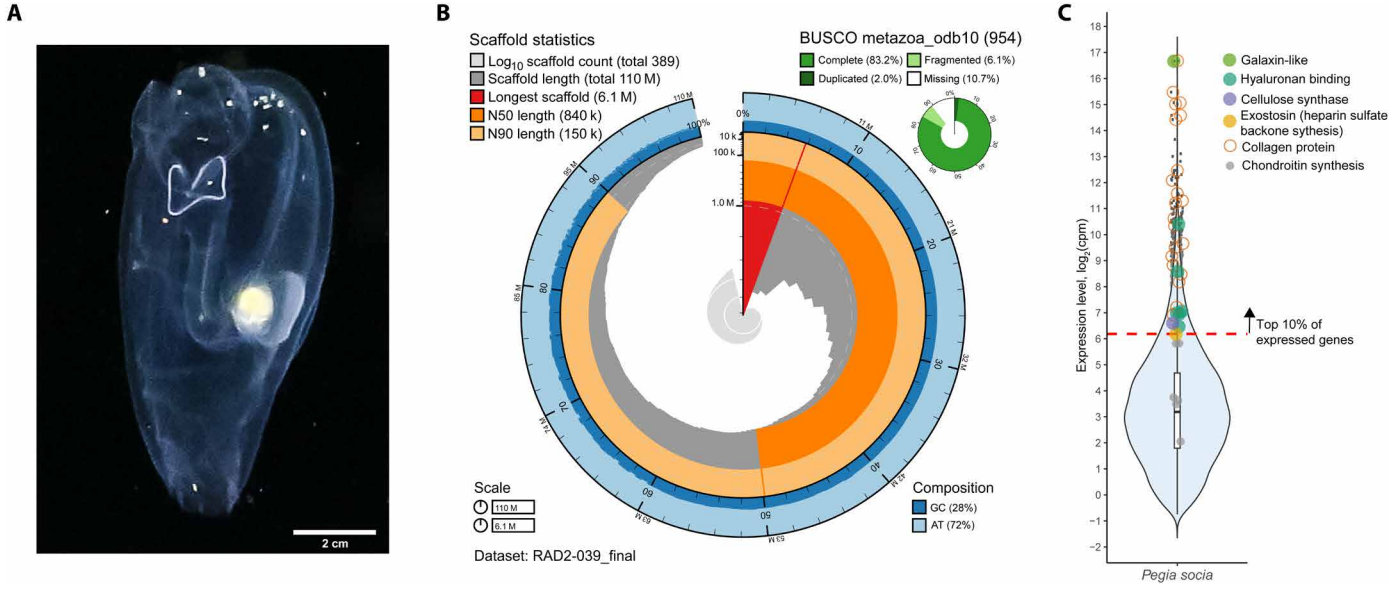

Fig. 4. Salp *Pegea* sp. genome statistics and gene expression. (**A**) *Pegea* sp. image captured by SuBastian's 4K video recorder. (**B**) Blobtools snailplot describing statistics of the *Pegea* sp. genome assembly including continuity statistics (e.g., assembly and scaffold lengths, N50, N90), GC content, and BUSCO completeness. (**C**) Violin plot describing the distribution of transcript abundances of *Pegea* sp. Genes whose functions are enriched among the top 10% most highly expressed transcripts are shown as colored dots. Gray points show relative expression levels of chondroitin synthesis genes, which are important for the extracellular matrix but have more moderate expression levels than other highly expressed extracellular matrix relevant genes such as those encoding collagen, cellulase synthase, and hyaluronan binding proteins. Image in (A) captured by the 4K science camera on the ROV SuBastian, R/V *Falkor*, Schmidt Ocean Institute. Scale bar, 2 cm, based on total length of the animal; it does not represent skew or distortion based on distance from camera.

does encode a hyaluronidase, an enzyme with biomedical applications that breaks down hyaluronan. Although hyaluronan synthases are only found in vertebrates and some bacteria (*25*), chemical analyses of mollusks have uncovered hyaluronans in that group despite a lack of hyaluronan synthase genes in molluscan genomes (*26*, *27*), indicating that there may be previously unidentified hyaluronan synthesis enzymes to be found. Another highly expressed gene in *Pegea* sp. is an animal cellulose synthase (Fig. 4), which is presumably important for building the "tunic" epidermal structure of the organism. The most highly expressed transcript encodes a galaxin (Fig. 4), a protein involved with mineralization of coral skeletons that may also interact with collagens. Given enrichment of transcripts encoding extracellular matrix proteins and glycosaminoglycans, we also considered heparan and chondroitin sulfate synthesis and see moderate expression of transcripts encoding chondroitin sulfate synthesis and heparan synthesis proteins (Fig. 4).

Acicular cirri are neurosensory organs in tomopterids

Retrieval of an intact tomopterid animal coupled with rapid in situ preservation of its tissues allowed dissection of specific tissues and quantification of expressed transcripts in those tissues (table S9). We focused on the acicular cirri of the tomopterid, which are long, modified tentacles that reach approximately half the length of the animal's body (Fig. 5). Transcripts with enhanced expression in the acicular cirri as compared to body segments highlight the role of acicular cirri as neurosensory organs (*28*). Key processes enriched in the acicular cirri include sensing with G protein–coupled receptors (GPCRs) (table S10). GPCRs are a diverse group of sensory receptors involved in both chemosensory and light detection. The GCPRs expressed in the tomopterid cirri include homologs of low-light receptors, suggesting that the organs may function in accessory light detection. In addition, transcripts encoding ammonia transporters are highly enriched in the cirri. Ammonia transporters have been confirmed as olfactory receptors in drosophila (*29*) and could function as such for tomopterids, given that their prey of fish larvae and chaetognaths (*30*, *31*) release ammonia as a waste product (*32*). Further evidence of chemosensing comes from enrichment of expression of ionotropic receptors (*33*) and two gustatory receptors that are both homologs of trehalose receptors responsible for detection of "sweet" tastes (*34*). The cirri are also enriched in expression of transcripts encoding neural-cadherin and protocadherins, cell-cell adhesion molecules important for neural cell connectivity (*35*). In comparison, the body segments, which include parapodia and a cross section of the gut, are relatively enriched in transcripts encoding cellular processes and components linked to digestion (such as nutrient transport and proteolysis), as well as muscles (such as muscle development) (table S11), while highly expressed genes common to both tissues include more general cellular processes such as the translation machinery and genes involved in respiration (table S12).

DISCUSSION

In situ digital syntheses offer the potential to enhance type material for deep-sea animals

Revolutionary strides in underwater imaging, robotics, and genomics are reshaping marine exploration and initiating new avenues for biological research. In this study, we integrate these advancements to present in situ digital specimens from the deep sea (table S13) using cybertaxonomy—the digitization of taxonomic information (*36*)—to enrich our understanding of delicate marine species.

In taxonomy, a holotype is the single, authoritative specimen upon which the original description and naming of a new species is based. This specimen serves as the primary reference for the species, establishing a standard for future comparisons and identifications. For many groups of plant and animal, holotypes are whole preserved organisms or representative portions held in a museum collection, but for gelatinous zooplankton, long-term preservation of morphology is challenging (*15*). In these cases, the type material for species description can be a drawing, image, or movie (*17*, *37*).

In zoology, there is no official mechanism to augment a holotype with new information. However, as species representation methods advance, incorporating detailed elements such as 3D computer models and linked genomic data can make digital information as or more representative than existing type material and easier to share than physical specimens (*38*). This has led researchers to propose addendums to the International Commission on Zoological Nomenclature (ICZN) rules regarding type specimens (*39*–*41*). One such proposal is the concept of "cybertypes," digitized taxonomic material that enriches existing type specimens. Our integrated image, model, and genomic data align with the guidelines for collecting and disseminating cybertype material (*39*). Alternatively, in botany, "epitypes" serve as valuable addendums to species descriptions, allowing the addition of new data types, such as in situ 3D internal and external morphology and genome-scale data, to species-type information (*42*). Although the ICZN does not officially recognize epitypes in zoology, they can be invaluable as technology enables new representations of organisms, especially among gelatinous zooplankton, where holotype specimens may be fragments and particularly information-poor (*15*, *17*). While cybertypes and epitypes have primarily been used to extend the information available for existing physical specimens (*43*), when encountering previously unknown species with our integrated methodology, the digital data could have served as fully virtual, in situ digital holotypes—potentially a crucial designation for future species identification in the oceans.

Advances in underwater imaging enable more comprehensive morphological descriptions of gelatinous zooplankton

The field of underwater imaging has seen substantial advancements in recent decades, notably the development of structured light technologies (*13*, *44*, *45*) capable of reconstructing 3D gelatinous structures of animals in their natural habitat (*12*). Coupled with newer processing approaches such as photogrammetry (*46*, *47*), these in situ imaging techniques allow complex studies of animal behavior and ecomechanics, enabling scientists to circumvent issues of structural integrity loss associated with biological fixation for long-term storage (*12*, *48*). These technologies also facilitate fine-scale measurements of morphological features and time-varying deformations, aiding research into locomotion and reproduction (*49*, *50*). Beyond generating readily shareable digital representations of marine life (*51*, *52*), the 3D models derived from these data can be harnessed for computational modeling (*53*), thus expanding our understanding of the physical limitations inherent to swimming and feeding. With these capabilities, underwater imaging is poised to fill notable gaps in species description (*17*, *37*), presenting a new frontier in ecological studies.

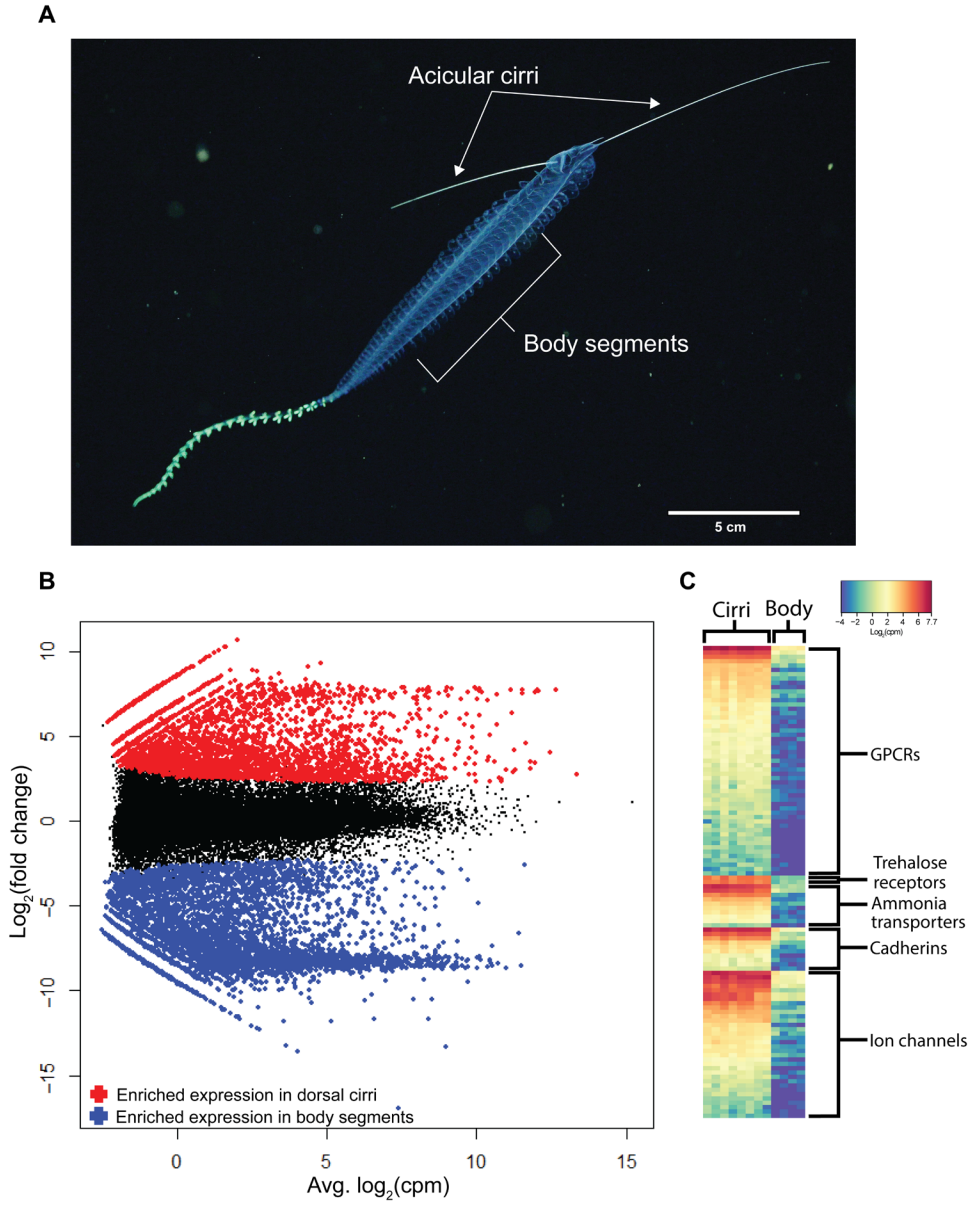

Fig. 5. Gene expression in the gossamer worm *Tomopteris* sp. reveals functions of the acicular cirri. (**A**) *Tomopteris* animal captured by SuBastian's 4K video recorder. Tissues sampled for quantitative gene expression include the acicular cirri, with $n = 4$ tissue samples from each cirrus and $n = 4$ body segments. (**B**) An MA plot [log average (A) on the horizontal axis and the log ratio (M) on the vertical axis] showing the distribution of transcripts and their relative expression and enrichment levels in acicular cirri or body segments. (**C**) Heatmap of gene expression for genes involved in sensory perception in acicular cirri across $n = 8$ tissue samples from the two cirri and $n = 4$ body segments. Image in (A) captured by the 4K science camera on the ROV SuBastian, R/V *Falkor*, Schmidt Ocean Institute. Scale bar, 2 cm, based on total length of the animal; it does not represent skew or distortion based on distance from camera. GPCRs, G protein–coupled receptors.

Innovative robotics, soft robotics, and encapsulation devices advance our ability to study ocean animals in situ

Before the RAD, only two other methods had been developed for robotic deep-sea collection of delicate midwater species (*22*). Using those methods, deep-sea animals are captured and brought to the surface accompanied with: (i) complexities in keeping the animal alive (*54*), (ii) preservatives, such as formalin or ethanol, harden and alter tissue components (*55*), and (iii) there is an overexpression of stress genes as the animal has undergone a marked change in pressure and temperature before tissue preservation (*56*). Because of those issues, when a soft-bodied animal, such as a jellyfish, is preserved with alcohol or formalin after collection, it is no longer possible to obtain detailed and nuanced information regarding gene expression from the animal at depth or of tissue volume, structure, or dynamics. Gentle encapsulation of an animal with in situ preservation of tissues with a preservative that captures the in situ transcriptional state overcomes those challenges.

Genome and transcriptome sequencing provide reference information for underwater species and offer insights into gene expression

Modern genomic-sequencing techniques have also enabled deep genomic and transcriptomic analysis of organisms based on tens of milligrams of tissue. These technologies combined with in situ preservation allows sequencing, assembly, and annotation of reference genomes and transcriptomes for an organism as well as quantitative analyses of expressed genes. For this study, sequencing was done after samples were returned to the laboratory, but, in future efforts, sequencing for initial identification could be done at sea (*57*), further accelerating the process of discovery.

The challenges of applying taxonomic names to organisms with widespread distributions underscore the need for more focused research in this area

In this study, we provide detailed imaging and sequence data for four specimens but achieve species-level identification for only one. We emphasize the taxonomic challenges posed by soft-bodied marine animals, which can contain cryptic "sibling" species, similar in morphology but distinct genetically (*58*). For example, our *M. claudanielis* specimen aligns morphologically with the eastern North Pacific holotype and shares an exact molecular match with a regional specimen. However, the original description of the species includes specimens from distinct ocean basins (*20*), which could represent unrecognized species diversity under a single name. Looking at available sequence information in more detail, while our specimen is a 100% match with the mitochondrial 16*S* sequence of an eastern North Pacific voucher specimen of *M. claudanielis* (*59*), it is also a 98% match to a cytochrome 1 oxidase sequence for a specimen named *M. claudanielis* from the western North Atlantic (*60*). These findings suggest a broad distribution for *M. claudanielis*, with morphological and genetic similarity across ocean basins, consistent with the species description (*20*); however, because of the lack of uniform molecular markers for comparison, definitive conclusions about molecular variation remain unattainable at this stage. Similarly, the classification of our other specimens necessitates careful examination of their capture locations, comparison with the locations of any corresponding holotypes, and consideration of possible cryptic diversity. Expertise in the morphology and genetics of siphonophores, salps, and polychaetes/tomopterids is essential to confidently assign species names and understand their distribution. Our study's multifaceted imaging, alongside the assembly of one complete reference genome and four reference transcriptomes, can bridge knowledge gaps if comparable data are gathered widely. We encourage the scientific community to integrate our findings into their research and to contemplate additional digital data types, such as shadowgraph imaging that may capture more traditional identifying morphological features of gelatinous zooplankton (*61*), that could further enhance taxonomic identification in future studies.

Cross-disciplinary integration opens new avenues for marine research and shapes future outlooks

While each individual technological advancement has made its mark, it is the simultaneous integration of these emerging fields that promises to revolutionize marine species discovery, description, and physiological characterization. The specimens analyzed in this study serve as prime examples of the wealth of information these combined methods can extract. In contrast to species descriptions that rely solely on imaging data (*17*, *37*), our approach weaves together diverse quantitative data types. This provides a comprehensive representation of delicate, difficult-to-preserve marine species, bridging the gap traditionally filled by physical voucher specimens.

The proposed synthesis of digital data, which compiles quantitative visual data, 3D models, and genomic data (table S13), not only outstrips physical specimens in versatility but also greatly enhances the accessibility, shareability, and comparability of the data. This development complements contemporary efforts to digitize museum collections (*52*, *62*, *63*) and aggregate imagery for species description (*64*, *65*), thus expanding the possibilities for detailed and comprehensive species information beyond the scope of conventional imaging data.

Confronted with the current, time-consuming and resource-intensive approaches to species description, our demonstrated technologies provide a solution that aligns with the rapid pace of scientific discovery. They offer a more streamlined and accessible practice for the taxonomic community. In addition, they pave the way for the incorporation of other technological innovations, setting the stage for the development of more precise and less invasive sampling techniques, such as micro-biopsies or swabs, and potentially enabling autonomous exploration (*66*).

For example, our RAD system can be upgraded with visual and force feedback mechanisms, thereby lowering costs and allowing usage on platforms other than high-end ROVs. Visual feedback could be provided by internal cameras, while force feedback could come from tactile sensors built into the soft gaskets. Considering the extraordinary longevity of some deep-sea organisms—up to 18,000 years for certain sponges (*67*)—these noninvasive techniques are increasingly vital as we strive to explore and responsibly steward our planet's vast and underexplored oceanic regions. In addition, the approach and data presented here could be useful to the scientific community working at the de-extinction interface (*68*). If a species documented with this approach was to later go extinct, these comprehensive data included would be a rich repository to draw upon, especially as current extinction rates are more than 100 times higher than background extinction rates (*69*).

It is estimated that 33 to 66% of all ocean species are not yet described (*70*), and with species description taking, on average, 21 years after the first observation (*38*), filling these gaps would require enormous effort and associated cost. The integrated strategy offered here is one approach to gain a vast amount of information on animals encountered within a relatively brief duration of observation. Given the costs of deep-sea exploration, integrative approaches that maximize information gained are necessary to fill these knowledge gaps. This is particularly urgent in the face of escalating climate change and extinction rates that is leading to species disappearance before they are ever studied or documented. The biological community has begun formal calls for the reduction and/or elimination of morphology in new species descriptions (*71*), and undescribed species are now recognized as having a higher extinction risk than described species (*72*), underscoring the need for efficient species description workflows and cross-linked species metadata (*73*).

Accelerating our understanding of these organisms is crucial for formulating effective conservation strategies to understand and protect them. This also offers a roadmap for future studies to gain maximal amount of information obtained from single specimens. Genomic information can be informative for understanding adaptation to the deep-sea environment (*74*). Quantification of motility

and rates of filtering/feeding enabled by advanced imaging can enhance our understanding of the ecological roles of deep-sea animals (*12*, *13*, *75*). In addition, artificial intelligence–enabled, structured scanning of deep-sea environments, powered by the integration of new technologies, can help quantify the distribution of animals in the water column and optimize the collection of specimens (*65*, *76*).

In a rapidly changing world, the deep sea, one of the last frontiers for scientific exploration, faces biodiversity risks even before being fully documented. The integrative approach used in this study harnesses cutting-edge technology to expedite the collection of digitized taxonomic information, optimize sample collection, and deepen our understanding of adaptations and ecological roles of gelatinous zooplankton species. Only through these collaborative efforts can we hope to document, comprehend, and ultimately protect the ocean's rich biodiversity.

MATERIALS AND METHODS

DeepPIV

DeepPIV (particle image velocimetry; https://mbari.org/technology/deeppiv/) is a ROV-deployable imaging device capable of 3D in situ visualizations of soft-bodied deep-sea organisms. DeepPIV pairs a high-resolution video camera with a laser sheet to image particles and fragile meso- and microscale gelatinous structures. The instrument is mounted in the ROV sled and extended out from the vehicle body for sampling. Pilots translate the ROV fore and aft to scan the laser sheet through the target organism while the camera is recording. DeepPIV scan data were collected on 49 different individuals during the research expedition (table S2) and only a subset of data analyzed here. Recorded video is stabilized to minimize the effect of vehicle motion, and data from an auxiliary camera, mounted to the laser projection housing, are used to compensate for vehicle acceleration/deceleration, using custom MATLAB software (*77*). Additional processing steps are identical to previous work (*48*) and consist of a segmentation using a combination of manual and automatic luminosity-based methods in 3D Slicer and additional mesh simplification in MeshLab. The output meshes represent 3D reconstructions that provide new, quantitative views of gelatinous structures.

EyeRIS

EyeRIS is an ROV-deployable lightfield (plenoptic) camera system capable of imaging particle fields and tissue surface movement in 3D (https://mbari.org/technology/eyeris/). The instrument is built around a plenoptic micro-lens array placed between the camera lens and imaging sensor, functionally subdividing the image. EyeRIS integrates a Raytrix R26 camera with a Canon 70- to 200-mm lens, resulting in a field of view of 12 cm by 12 cm by 6 cm at the widest zoom. The imaged volume is 65 cm from the viewport and is illuminated by five red (660 nm) spot lights. The light array is mounted on a hydraulic umbrella frame that expands from the housing when deployed. Real-time computer vision algorithms then use each subimage to reconstruct the data such that every object in the frame is in focus and associated with a 3D spatial coordinate. The resulting depth maps can achieve z resolution on the order of the camera pixel resolution. EyeRIS was used to collect imaging data on 61 different individuals during the research expedition (table S2), and only a subset of data are analyzed here.

Repeated quantitative measurements were made on the four individuals discussed in the present work: *Erenna* sp., *Marrus claudanielis*, *Pegia* sp., and the *Tomopteris* sp. Video sequences from EyeRIS were loaded into RxLive, Raytrix's real-time user interface, and displayed as colored depth maps (Fig 2B, no. 4; table S13). A human annotator selected 10 nonadjacent frames when the desired structure was in the sample volume, measured the feature, and recorded the results. Measurements were averaged over the 10 frames and reported with SE (table S4). Not all structures exist on each organism, and certain features were not visible between individuals. Most notably, the siphonophore *Erenna* sp. nectosome was too long to be captured in a single frame.

RAD-2 and SuPR sampling system

The RAD-2 encapsulation device was extensively modified and enhanced (Fig. 3), building upon progress presented in Teoh *et al.* (*22*). The RAD-2 encapsulation volume was scaled up 3.375 times to 0.0105 m^3. RAD-2 also has an assembly linkage layer and a folding linkage layer. Unlike the design presented in Teoh *et al.* (*22*), the input link of the assembly linkage layer rotates to cause encapsulation as opposed to the input link of the folding linkage layer. As a result, the folding linkage layer does not swirl during encapsulation. This architectural change enables the central face of the dodecahedral encapsulation volume to be instrumented by allowing the motor driving the encapsulation to be offset from the rotation axis of the input link of the assembly linkage layer. The motor is connected to the input link of the assembly linkage layer by a geared transmission. A pinion gear on the output shaft of the motor transmits torque to a ring gear attached to the input link of the assembly linkage layer. The fully encapsulated volume has 12 surfaces. Because the motor is offset from the rotation axis of the input link of the assembly linkage layer, this allows the central nonmoving surface to be instrumented with a tissue sampling device and an inlet port for pumping sampled tissue into the SuPR sampling system.

The other 11 surfaces each have mount points allowing customizable integration of imaging and sensing devices. RAD-2 was not only modified architecturally but was also enhanced by the use of materials that could withstand the forces and harshness of the marine environment. For example, the mounting structure for the input link of the assembly linkage layer and the input link of the folding linkage layer is made of aluminum 6061-T6, which was hard-anodized post machining for protection against saltwater corrosion. All other links of the assembly linkage layer and the folding linkage layer were made of Hydex 301. This material was chosen for its transparency, impact strength, and saltwater corrosion resistance. In the transmission, the pinion gear is made out of 17-4 PH stainless steel, and the ring gear is made out of Nitronic 60 stainless steel. These materials were chosen as a pair because of low wear, galling resistance, and saltwater corrosion resistance. The RAD-2 encapsulation device was constructed using a combination of anodized aluminum, stainless steel, and Hydex 301. The rotary actuator used for opening and closing RAD-2 was a 2G Engineering 2000 Series high-speed rotary actuator. The tissue cleaving apparatus was driven by a Maxon MT20 thruster motor, and the stainless steel four-blade design was derived and scaled on the basis of culinary-grade blender designs. The camera used for viewing inside RAD-2 was a DeepSea Power & Light Nano SeaCam. All control electronics for the RAD-2 and SuPR systems were enclosed in a 4500-meter rated custom titanium housing with communication with the ROV achieved via gigabit ethernet.

The in situ sample collection and preservation system was a McLane Labs SuPR Sampler, with original design presented by Breier *et al.* (*14*). This system allows for a specified amount of fluid to be sampled using a high-flow pump (8 liter/min) addressed to discrete fluid ports among 14 individual filter housings with in situ preservative injectable immediately following the sample collection. Filters were prepared onboard by layering two 160-mm support mesh discs under the plastic support frit followed by a third 160-mm support mesh disc, and lastly, a 147-mm, 100-μm mesh Nitex disc on top for tissue sample collection. The arrangement allowed collection of tissue fragments on the 100-μm mesh Nitex dish, while the assembly of the support disks and frits maintained the integrity of the space inside the filter holder. Filter holders were assembled and bolted closed and then were filled with Milli-Q water until the assembly and tubing was full. Air bubbles were removed by rotating and tapping the assembly on a table to remove as many bubbles as possible before ROV deployment.

Custom RNA preservative (described as RNALater) was produced according to the formulation by Malmstrom (*78*). The pH of the RNALater preservative was adjusted to 5.2 using concentrated sulfuric acid. Sixty liters of RNALater was prepared ahead of the cruise on site in San Diego using resources available in the Aluwihare laboratory at the Scripps Institute of Oceanography. The SuPR sampler was loaded with 10 liters of RNALater before each dive.

Animals were encased by the RAD2 sampler and then homogenized with simultaneous pumping of water through the SuPR sampler. Following capture of material, the filter was flooded with a custom formulation of RNALater (0.5 to 1 liter volume per sample) to ensure that all seawater in the tubing and filter holder with sample was replaced with RNALater. Time that blade rotation was initiated was recorded, as well as the time and volume of seawater pumped through the filter, and lastly the time and volume of RNALater pumped through the filter. The average time to complete preservation was 3:17 min with the shortest time at 1:37 min and the longest time at 13:29 mins.

Once the ROV was back on deck of the research vessel, the filter holders were immediately collected from the SUPRsampler array and brought into the onboard wet laboratory. In the laboratory, filter holders were first drained into a waste container to collect waste RNALater. When large fragments were available (e.g. Fig. 1, no. 5), tissue fragments were collected with forceps into 5-ml cryotubes that were subsequently filled with additional RNALater. When there was ample material or the material could not easily be separated from the Nitex filter, the entire filter with tissue was rolled up and placed in a 50-ml Falcon tube that was then filled with additional RNALater to 40 ml. Samples collected in this way were incubated overnight at 4°C in RNALater and were transferred to a −80°C freezer the following day. Following removal of tissues and filters from filter holders, the filter holder was rinsed with seawater and wiped with a paper towel to mechanically remove any remaining fragments. Filter holders and support frits and discs were then placed in a bath of 0.1% HCl and soaked for at least 5 min. Following soaking in HCl, filter holders and support disks and frits were transferred through subsequent baths of Milli-Q water, soaking in each for at least 5 min. All parts were dried on paper towels and then reassembled and filled with Milli-Q water before the following dive.

RNA and DNA extraction

Samples from all animals were processed the same way for RNA and DNA extraction. RNALater preserved samples were thawed, and tissue fragments were spread on a sterile petri dish. One hundred milligrams of tissue was collected from each organism using a new razor blade to cut tissue fragments and clean forceps. Tissue fragments were placed into preweighed 1.5-ml microcentrifuge tube and weighed to get the appropriate mass of tissue.

For RNA, Trizol reagent was used to extract and purify total RNA. Following tissue collection, 100 mg of tissue fragments were lysed in TRIzol reagent, and total RNA was extracted according to the manufacturer's instructions.

For DNA, QIAGEN Genomic-Tips were used to collect high–molecular weight DNA. Approximately 100 mg of tissue fragments was collected for each DNA extraction. The tissue fragments were further dissected to small pieces by manual cutting using a new razor blade. Minced tissue fragments were suspended in Qiagen lysis buffer ATL at a ratio of 220 μl of buffer ATL to 25 mg of tissue. Twenty microliters of proteinase K (20 mg/ml) per 220 μl of buffer ATL was added, and the mixture was incubated at 56°C for 1 hour with gentle mixing of the tubes by inversion every 10 min. Following proteinase K digestion, samples were centrifuged at 6000*g* for 5 min to pellet remaining tissue debris. The supernatant was recovered, and 4 μl of ribonuclease (RNase) A (100 mg/ml) per 220 μl was added to each tube. Samples were incubated at room temperature (22°C) for 10 min. Following RNase A incubation, samples were transferred to 15-ml falcon tubes containing QIAGEN buffer G2 to a final volume of 10 ml. Following dilution of the lysates into buffer G2, the samples were added to an equilibrated QIAGEN G-tip 100, and DNA cleanup and elution were completed according to the manufacturer's recommendation. The average molecular weight attainable from these samples was in the range of 25 to 50 kbp. We tested several protocols for DNA extraction including grinding tissue on dry ice and QIAGEN's MagAttract HMW DNA kit but could not improve the length profile of the extracted DNA. We suspect that RNALater preservation has the effect of shearing DNA, corroborating results from studies that note that RNALater can perform poorly in maintaining high–molecular weight DNA for some samples (*79*, *80*) and suggest future studies to test that. Forceps were cleaned between samples by soaking in 0.1% HCl for 5 min, followed by rinsing with Milli-Q water, wiping with Kimwipes, rinsing with 70% ethanol, and a final rinse with Milli-Q water.

Transcriptome sequencing

Purified total RNA integrity was checked on an Agilent TapeStation instrument using high-sensitivity RNA ScreenTape. RNA integrity values for the four specimens reported here was >7.0. Total RNA was shipped on dry ice to a sequencing facility (Azenta, Genewiz, South Plainfield, NJ). Reference transcriptomes were generated using the Pacific Biosciences Iso-Seq method that sequences full-length RNA and produces a reference transcriptome capturing all expressed isoforms without need for assembly. Functional completeness of Iso-Seq transcriptomes was analyzed using the software tool BUSCO (v5.2.2) (*81*) against the metazoan BUSCO database (metazoa_odb10).

Quantitative mRNA sequencing was performed by selecting samples for short-read sequencing on the Illumina HiSeq platform. Replicate quantitative transcriptomes were generated from two samples of salp *Pegia* sp. tissue fragments (table S9). Quantitative mRNA sequencing of the acicular sensory cirri of *Tomopteris* sp. was enabled by fortuitous collection and preservation of the entire animal without dissection. Specific tissues were dissected in the

onboard laboratory. Four 1-cm fragments from the end of each acicular cirrus were dissected by cutting each cirrus at 1-cm intervals starting from the distal tip. Comparative samples from the body were collected by dissecting the body horizontally in four segments from the middle of the animal, including left and right parapodia and the trunk section between for each tissue fragment. Total RNA was extracted using TRIzol reagent and sent on dry ice for library preparation and sequencing on the Illumina HiSeq platform at Azenta/Genewiz. We targeted 20 million 150-bp paired-end reads per sample and recovered an average of 35 million paired-end reads per sample (table S9).

To capture any tissue specific RNA, we completed a de novo transcriptome assembly for the tomopterid (RAD2-51) using pooled illumina reads from all tissues. Random sequencing errors in reads were corrected using rCorrector (v1.0.5) (*82*), and unfixable reads were removed with the python script FilterUncorrectabledPEfastq.py (*83*). Adapter and quality trimming was performed with TrimGalore (v0.6.7) (*84*). Ribosomal RNAs were removed by mapping reads against the SILVA database (SILVA_138.1) (*85*) using Bowtie2 (v2.5.0) (*86*). Transcriptome assembly was completed on the processed reads using the Trinity assembler (v2.14.0) (*87*). The resultant de novo assembly was combined with the Pacific Biosciences IsoSeq assembly using the software cd-hit-est (v4.8.1) (*88*) with an identity threshold of 95% for clustering. The assembly contained some contamination from other organisms, likely from demultiplexing errors during short-read sequencing. Identifiable contaminants were removed at this stage by blastn of the combined assembly against whole transcriptome data from *Homo sapiens*, *Danio rerio*, and *Drosophila melanogaster*. Hits with 98% or better identity more than 80% or more of the subject sequence for any of those three species were removed from the assembly at this stage. The filtered, combined assembly was used for quantitative analyses.

Genome sequencing and assembly

Whole-genome sequencing was completed for all four animals using a combination of short and long read sequencing. On the basis available information from literature on related organisms (*23*, *89*), we anticipated genome sizes approximately on the order 1 to 3 Gbp for each organism or around the size of the human genome, and we conducted our sequencing effort accordingly. We found that the salp, *Pegia* sp. (RAD2-039) had a smaller than anticipated genome with an estimated size at this stage of approximately 200 Mbp based on the inferred haploid peak in a *k*-mer plot (fig. S3), while the three other specimens had much larger than anticipated genomes, likely exceeding 23 billion bp for RAD2-051 (*Tomopteris* sp.) and RAD2-064 (siphonophore, *Marrus claudanielis*) and approximately 27 billion bp or more for RAD2-065 (siphonophore, *Erenna* sp.) (table S14). *K*-mer plots for RAD2-051 and RAD2-064 generated no peaks (fig. S3), indicating that coverage was insufficient to capture the sequence space of the genome despite having over 235 billion bases of data for each. With 33% additional read data for RAD2-065, the *k*-mer plot has a single peak at around 12× coverage (fig. S3), indicating insufficient coverage for genome assembly but giving a better lower bound estimate for the genome size of that specimen/species (siphonophore *Errena* sp.). Assembly of the larger genomes would require a sequencing effort about 10× greater than the current depth of sequencing, which was beyond the budget of the project. The apparent large genomes of soft-bodied deep-sea organisms is notable for future sequencing efforts.

Read coverage of the salp, RAD2-039 (*Pegea* sp.), genome was in excess of that needed for assembly, which can be detrimental to assembly efforts (*90*). Short-read coverage was reduced to 100× for the haploid peak by randomly downsampling the reads using the tool seqtk [v1.3 (r106)] (*91*). The 100× coverage short-read files and the complete long-read files were used in a hybrid long- and short-read assembly using the assembler MaSuRCA (v4.0.9) (*92*, *93*). Assembly quality and statistics were established using BlobTools (v1.1.1) (*94*).

Repeats in the newly sequenced and assembled genome were found using the tool RepeatModeler2 (v2.0.3) (*95*), and the genome was soft-masked using RepeatMasker (v4.1.4) (*96*). Structural annotation of the *Pegea* sp. genome was completed using the tool FINDER (v1.1.0) (*97*), which used evidence from whole transcriptome reads (the two Illumina libraries, RAD2-039-2 and RAD2-039-3) to find coding regions in the genome using the tools BRAKER2 and associated software (*98*–*106*). Resultant protein predictions were annotated for function using the tool Interproscan5 (v5.60-92.0) (*107*).

Quantitative transcriptome analyses

For the salp *Pegea* sp. (RAD2-039), the reference transcriptome was derived from the annotated genome assembly. For the tomopterid (RAD2-051), the reference transcriptome was the combined long-read IsoSeq plus short-read Illumina-Trinity assembly. Reads were mapped to each reference assembly using the tool Salmon (v1.9.0) (*108*) in mapping mode. For RAD2-039, decoy-aware mapping was done using the whole genome as decoy sequence. For RAD2-051, no genome was available, so mapping was done without decoys. Salmon quantification files were read into the R statistical environment (v4.2.2) (*109*) using the tximport tool (*110*). Quantitative and statistical analyses of transcriptome data were performed using edgeR (v3.40.1) (*111*). Functional summaries of differentially expressed genes were performed using topGO (v2.50.0) (*112*).

Supplementary Materials

This PDF file includes:

Figs. S1 to S3
Tables S1 to S14
Legend for movie S1
Legends for supplementary Excel files 1 to 4
References

Other Supplementary Material for this manuscript includes the following:

Supplementary Excel files 1 to 4
Movie S1

REFERENCES AND NOTES

1. N. Smith, Planet ocean: The last frontier. *Eng. Technol.* **14**, 58–63 (2019).
2. N. R. Council, *Undersea Vehicles and National Needs* (National Academies Press, 1996).
3. K. C. Galloway, K. P. Becker, B. Phillips, J. Kirby, S. Licht, D. Tchernov, R. J. Wood, D. F. Gruber, Soft robotic grippers for biological sampling on deep reefs. *Soft Robot.* **3**, 23–33 (2016).
4. D. F. Gruber, R. J. Wood, Advances and future outlooks in soft robotics for minimally invasive marine biology. *Sci. Robot.* **7**, eabm6807 (2022).
5. B. T. Phillips, K. P. Becker, S. Kurumaya, K. C. Galloway, G. Whittredge, D. M. Vogt, C. B. Teeple, M. H. Rosen, V. A. Pieribone, D. F. Gruber, R. J. Wood, A. Dexterous, A Dexterous, glove-based teleoperable low-power soft robotic arm for delicate deep-sea biological exploration. *Sci. Rep.* **8**, 14779 (2018).
6. B. Robison, K. Reisenbichler, R. Sherlock, The coevolution of midwater research and ROV technology at MBARI. *Oceanography* **30**, 26–37 (2017).
7. N. R. Sinatra, C. B. Teeple, D. M. Vogt, K. K. Parker, D. F. Gruber, R. J. Wood, Ultragentle manipulation of delicate structures using a soft robotic gripper. *Sci. Robot.* **4**, eaax5425 (2019).

8. D. M. Vogt, K. P. Becker, B. T. Phillips, M. A. Graule, R. D. Rotjan, T. M. Shank, E. E. Cordes, R. J. Wood, D. F. Gruber, Shipboard design and fabrication of custom 3D-printed soft robotic manipulators for the investigation of delicate deep-sea organisms. *PLOS ONE* **13**, e0200386 (2018).
9. A. T. Greer, J. C. Lehrter, B. M. Binder, A. R. Nayak, R. Barua, A. E. Rice, J. H. Cohen, M. N. McFarland, A. Hagemeyer, N. D. Stockley, K. M. Boswell, I. Shulman, S. deRada, B. Penta, High-resolution sampling of a broad marine life size spectrum reveals differing size- and composition-based associations with physical oceanographic structure. *Front. Mar. Sci.* **7**, 542701 (2020).
10. J. M. Durden, T. Schoening, F. Althaus, A. Friedman, R. Garcia, A. G. Glover, J. Greinert, N. Jacobsen Stout, D. O. B. Jones, A. Jordt, J. Kaeli, K. Koser, L. A. Kuhnz, D. Lindsay, K. J. Morris, T. W. Nattkemper, J. Osterloff, H. A. Ruhl, S. Hanumant, M. Tran, B. J. Bett, Perspectives in visual imaging for marine biology and ecology: From acquisition to understanding. *Oceanogr. Mar. Biol. An. Annu. Rev* **54**, 1–72 (2016).
11. A. R. Longhurst, W. G. Harrison, Vertical nitrogen flux from the oceanic photic zone by diel migrant zooplankton and nekton. *Deep Sea Res.* **35**, 881–889 (1988).
12. K. Katija, G. Troni, J. Daniels, K. Lance, R. E. Sherlock, A. D. Sherman, B. H. Robison, Revealing enigmatic mucus structures in the deep sea using DeepPIV. *Nature* **583**, 78–82 (2020).
13. K. Katija, R. E. Sherlock, A. D. Sherman, B. H. Robison, New technology reveals the role of giant larvaceans in oceanic carbon cycling. *Sci. Adv.* **3**, e1602374 (2017).
14. J. A. Breier, C. S. Sheik, D. Gomez-Ibanez, R. T. Sayre-McCord, R. Sanger, C. Rauch, M. Coleman, S. A. Bennett, B. R. Cron, M. Li, A large volume particulate and water multi-sampler with in situ preservation for microbial and biogeochemical studies. *Deep-Sea Res. I Oceanogr. Res. Pap.* **94**, 195–206 (2014).
15. S. H. Haddock, A golden age of gelata: Past and future research on planktonic ctenophores and cnidarians. *Hydrobiologia* **530**, 549–556 (2004).
16. D. G. Mitchell, A. Edgar, M. Q. Martindale, Improved histological fixation of gelatinous marine invertebrates. *Front. Zool.* **18**, 29 (2021).
17. D. L. Pawson, E. J. Foell, Peniagone leander new species, an abyssal benthopelagic sea cucumber (Echinodermata: Holothuroidea) from the eastern central Pacific Ocean. *Bull. Mar. Sci.* **38**, 293–299 (1986).
18. L. P. Madin, G. R. Harbison, Salps of the Genus *Pegea* Savigny 1816 (Tunicata: Thaliacea). *Bull. Mar. Sci.* **28**, 335–344 (1978).
19. K. Fauchald, *The polychaete worms. Definitions and keys to the orders, families and genera* (Natural History Museum of Los Angeles County, Science Series, 1977).
20. C. W. Dunn, P. R. Pugh, S. H. D. Haddock, *Marrus Claudanielis*, a new species of deep-sea Physonect Siphonophore (Siphonophora, Physonectae). *Bull. Mar. Sci.* **76**, 699–714 (2005).
21. P. R. Pugh, A review of the genus *Erenna* Bedot, 1904 (Siphonophora, Physonectae). *Bull.-Nat. Hist. Mus. Zool. Series* **67**, 169–182 (2001).
22. Z. E. Teoh, B. T. Phillips, K. P. Becker, G. Whittredge, J. C. Weaver, C. Hoberman, D. F. Gruber, R. J. Wood, Rotary-actuated folding polyhedrons for midwater investigation of delicate marine organisms. *Sci. Robot.* **3**, eaat5276 (2018).
23. N. K. Jue, P. G. Batta-Lona, S. Trusiak, C. Obergfell, A. Bucklin, M. J. O'Neill, R. J. O'Neill, Rapid evolutionary rates and unique genomic signatures discovered in the first reference genome for the southern ocean salp, *Salpa thompsoni* (Urochordata, Thaliacea). *Genome Biol. Evol.* **8**, 3171–3186 (2016).
24. K. T. Dicker, L. A. Gurski, S. Pradhan-Bhatt, R. L. Witt, M. C. Farach-Carson, X. Jia, Hyaluronan: A simple polysaccharide with diverse biological functions. *Acta Biomater.* **10**, 1558–1570 (2014).
25. A. B. Csoka, R. Stern, Hypotheses on the evolution of hyaluronan: A highly ironic acid. *Glycobiology* **23**, 398–411 (2013).
26. S. Kanchana, M. Arumugam, S. Giji, T. Balasubramanian, Isolation, characterization and antioxidant activity of hyaluronic acid from marine bivalve mollusc *Amussium pleuronectus* (Linnaeus, 1758). *Bioact. Carbohydr. Diet. Fibre* **2**, 1–7 (2013).
27. N. Volpi, F. Maccari, Purification and characterization of hyaluronic acid from the mollusc bivalve *Mytilus galloprovincialis*. *Biochimie* **85**, 619–625 (2003).
28. G. Purschke, C. Helm, Development and structure of the anterior nervous system and sense organs in the holopelagic annelid *Tomopteris* spp. (Phyllodocida, Errantia). *Org. Divers Evol.* **23**, 505–527 (2023).
29. A. Vulpe, H. S. Kim, S. Ballou, S.-T. Wu, V. Grabe, C. Nava Gonzales, T. Liang, S. Sachse, J. M. Jeanne, C.-Y. Su, K. Menuz, An ammonium transporter is a non-canonical olfactory receptor for ammonia. *Curr. Biol.* **31**, 3382–3390.e7 (2021).
30. K. Fauchald, P. A. Jumars, The diet of worms: A study of polychaete feeding guilds. *Oceanogr. Mar. Biol. Ann. Rev.* , (1979).
31. S. Rakusa-Suszczewski, Predation of chaetognatha by *Tomopteris Helgolandica* greff. *ICES J. Mar. Sci.* **32**, 226–231 (1968).
32. J. R. Beers, Ammonia and inorganic phosphorus excretion by the planktonic chaetognath, *Sagitta hispida* Conant. *ICES J. Mar. Sci.* **29**, 123–129 (1964).
33. A. Hussain, M. Zhang, H. K. Üçpunar, T. Svensson, E. Quillery, N. Gompel, R. Ignell, I. C. G. Kadow, Ionotropic chemosensory receptors mediate the taste and smell of polyamines. *PLoS Biol.* **14**, e1002454 (2016).
34. A. Dahanukar, K. Foster, W. M. van der Goes van Naters, J. R. Carlson, A Gr receptor is required for response to the sugar trehalose in taste neurons of *Drosophila*. *Nat. Neurosci.* **4**, 1182–1186 (2001).
35. S. Peek, K. M. Mah, J. A. Weiner, Regulation of neural circuit formation by protocadherins. *Cell. Mol. Life Sci.* **74**, 4133–4157 (2017).
36. Q. Wheeler, A. G. Valdecasas, Cybertaxonomy and ecology. *Nat. Educ. Knowledge* **1**, 6 (2010).
37. M. Ford, N. Bezio, A. Collins, *Duobrachium sparksae* (incertae sedis Ctenophora Tentaculata Cydippida): A new genus and species of benthopelagic ctenophore seen at 3,910 m depth off the coast of Puerto Rico. *Plankton Benthos Res.* **15**, 296–305 (2020).
38. B. Fontaine, A. Perrard, P. Bouchet, 21 years of shelf life between discovery and description of new species. *Curr. Biol.* **22**, R943–R944 (2012).
39. S. Faulwetter, A. Vasileiadou, M. Kouratoras, T. Dailianis, C. Arvanitidis, Micro-computed tomography: Introducing new dimensions to taxonomy. *ZooKeys* **263**, 1–45 (2013).
40. M. Schrödl, G. Haszprunar, Do we need epitypes in zoology. *Spixiana* **39**, 199–201 (2016).
41. The Code Online | International Commission on Zoological Nomenclature. https://iczn.org/the-code/the-code-online/.
42. A. N. Sennikov, The concept of epitypes in theory and practice. *Nordic J. Botany* **2022**, e03535 (2022).
43. N. Akkari, H. Enghoff, B. D. Metscher, A new dimension in documenting new species: High-detail imaging for myriapod taxonomy and first 3D cybertype of a new millipede species (Diplopoda, Julida, Julidae). *PLOS ONE* **10**, e0135243 (2015).
44. F. Caimi, F. Dalgleish, Subsea laser scanning and imaging systems, in *Subsea Optics and Imaging* (Elsevier, 2013), pp. 327–352.
45. M. Castillón, A. Palomer, J. Forest, P. Ridao, State of the art of underwater active optical 3D scanners. *Sensors(Basel)* **19**, 5161 (2019).
46. K. Calders, S. Phinn, R. Ferrari, J. Leon, J. Armston, G. P. Asner, M. Disney, 3D imaging insights into forests and coral reefs. *Trends Ecol. Evol.* **35**, 6–9 (2020).
47. T. Pulido Mantas, C. Roveta, B. Calcinai, C. G. di Camillo, C. Gambardella, C. Gregorin, M. Coppari, T. Marrocco, S. Puce, A. Riccardi, C. Cerrano, Photogrammetry, from the land to the sea and beyond: A unifying approach to study terrestrial and marine environments. *J. Mar. Sci. Eng.* **11**, 759 (2023).
48. J. Daniels, G. Sainz, K. Katija, New method for rapid 3D reconstruction of semi-transparent underwater animals and structures. *Integr. Org. Biol.* **5**, obad023 (2023).
49. F. Pereira, E. A. da Silva, G. Lafruit, *Plenoptic imaging: Representation and processing* (Academic Press Library in Signal Processing, 2018), vol. 6, pp. 75–111.
50. Z. Wang, Review of real-time three-dimensional shape measurement techniques. *Measurement* **156**, 107624 (2020).
51. T. G. Davies, I. A. Rahman, S. Lautenschlager, J. A. Cunningham, R. J. Asher, P. M. Barrett, K. T. Bates, S. Bengtson, R. B. Benson, D. M. Boyer, Open data and digital morphology. *Proc. Biol. Sci.* **284**, 20170194 (2017).
52. D. J. Irschick, F. Christiansen, N. Hammerschlag, J. Martin, P. T. Madsen, J. Wyneken, A. Brooks, A. Gleiss, S. Fossette, C. Siler, T. Gamble, F. Fish, U. Siebert, J. Patel, Z. Xu, E. Kalogerakis, J. Medina, A. Mukherji, M. Mandica, S. Zotos, J. Detwiler, B. Perot, G. Lauder, 3D visualization processes for recreating and studying organismal form. *iScience* **25**, 104867 (2022).
53. A. P. Hoover, J. Daniels, J. C. Nawroth, K. Katija, A computational model for tail undulation and fluid transport in the giant larvacean. *Fluids* **6**, 88 (2021).
54. B. Shillito, J. Ravaux, J. Sarrazin, M. Zbinden, P.-M. Sarradin, D. Barthelemy, Long-term maintenance and public exhibition of deep-sea hydrothermal fauna: The AbyssBox project. *Deep-Sea Res. II* **121**, 137–145 (2015).
55. R. Thavarajah, V. K. Mudimbaimannar, J. Elizabeth, U. K. Rao, K. Ranganathan, Chemical and physical basics of routine formaldehyde fixation. *J. Oral Maxillofac. Pathol.* **16**, 400–405 (2012).
56. G. Yan, Y. Lan, J. Sun, T. Xu, T. Wei, P.-Y. Qian, Comparative transcriptomic analysis of in situ and onboard fixed deep-sea limpets reveals sample preparation-related differences. *iScience* **25**, 104092 (2022).
57. J. J. M. Chang, Y. C. A. Ip, C. S. L. Ng, D. Huang, Takeaways from mobile DNA barcoding with BentoLab and MinION. *Genes(Basel)* **11**, 1121 (2020).
58. N. Knowlton, Sibling species in the sea. *Annu. Rev. Ecol. Syst.* **24**, 189–216 (1993).
59. C. W. Dunn, P. R. Pugh, S. H. D. Haddock, Molecular phylogenetics of the siphonophora (Cnidaria), with implications for the evolution of functional specialization. *Syst. Biol.* **54**, 916–935 (2005).
60. B. D. Ortman, A. Bucklin, F. Pagès, M. Youngbluth, DNA barcoding the medusozoa using mtCOI. *Deep-Sea Res. II Top. Stud. Oceanogr.* **57**, 2148–2156 (2010).
61. M. D. Ohman, A sea of tentacles: Optically discernible traits resolved from planktonic organisms in situ. *ICES J. Mar. Sci.* **76**, 1959–1972 (2019).
62. K. R. Johnson, I. F. P. Owens, A global approach for natural history museum collections. *Science* **379**, 1192–1194 (2023).
63. G. Nelson, S. Ellis, The history and impact of digitization and digital data mobilization on biodiversity research. *Philos. Trans. R. Soc. B Biol. Sci.* **374**, 20170391 (2018).
64. K. L. Howell, J. S. Davies, A. L. Allcock, A. Braga-Henriques, P. Buhl-Mortensen, M. Carreiro-Silva, C. Dominguez-Carrió, J. M. Durden, N. L. Foster, C. A. Game, B. Hitchin,

T. Horton, B. Hosking, D. O. B. Jones, C. Mah, C. L. Marchais, L. Menot, T. Morato, T. R. R. Pearman, N. Piechaud, R. E. Ross, H. A. Ruhl, H. Saeedi, P. V. Stefanoudis, G. H. Taranto, M. B. Thompson, J. R. Taylor, P. Tyler, J. Vad, L. Victorero, R. P. Vieira, L. C. Woodall, J. R. Xavier, D. Wagner, A framework for the development of a global standardised marine taxon reference image database (SMarTaR-ID) to support image-based analyses. *PLOS ONE* **14**, e0218904 (2019).
65. K. Katija, E. Orenstein, B. Schlining, L. Lundsten, K. Barnard, G. Sainz, O. Boulais, M. Cromwell, E. Butler, B. Woodward, K. L. C. Bell, FathomNet: A global image database for enabling artificial intelligence in the ocean. *Sci. Rep.* **12**, 15914 (2022).
66. A. Q. Byrne, Reimagining the future of natural history museums with compassionate collection. *PLoS Biol.* **21**, e3002101 (2023).
67. K. P. Jochum, J. A. Schuessler, X.-H. Wang, B. Stoll, U. Weis, W. E. G. Müller, G. H. Haug, M. O. Andreae, P. N. Froelich, Whole-ocean changes in silica and Ge/Si ratios during the Last Deglacial deduced from long-lived giant glass sponges. *Geophys. Res. Lett.* **4**, 11555–11564 (2017).
68. C. M. Schlebusch, Genomics: Testing the limits of de-extinction. *Curr. Biol.* **32**, R324–R327 (2022).
69. G. Ceballos, P. R. Ehrlich, R. Dirzo, Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proc. Natl. Acad. Sci.* **114**, E6089–E6096 (2017).
70. W. Appeltans, S. Ahyong, G. Anderson, M. Angel, T. Artois, N. Bailly, R. Bamber, A. Barber, I. Bartsch, A. Berta, M. Błażewicz-Paszkowycz, P. Bock, G. Boxshall, C. Boyko, S. Brandão, R. Bray, N. Bruce, S. Cairns, T.-Y. Chan, L. Cheng, A. Collins, T. Cribb, M. Curini-Galletti, D. Gordon, M. Guiry, F. Hernandez, B. Hoeksema, R. Hopcroft, D. Jaume, P. Kirk, N. Koedam, S. Koenemann, J. Kolb, R. Kristensen, A. Kroh, G. Lambert, D. Lazarus, R. Lamaitre, M. Longshaw, J. Lowry, E. Macpherson, L. Madin, C. Mah, G. Mapstone, P. McLaughlin, J. Mees, K. Meland, C. Messing, C. Mills, T. Molodtsova, R. Mooi, B. Nehaus, P. Ng, C. Nielsen, J. Norenburg, D. Opresko, M. Osawa, G. Paulay, W. Perrin, J. Pilger, G. Poore, P. Pugh, G. Read, J. Reimer, M. Rius, R. Roch, J. Saiz-Salinas, V. Scarabino, B. Schierwater, A. Schmidt-Rhaesa, K. Schnabel, M. Schotte, P. Schuchert, E. Schwabe, H. Segers, C. Self-Sullivan, N. Shenkar, V. Siegel, W. Sterrer, S. Stöhr, B. Swalla, M. Tasker, E. Theusen, T. Timm, M. Todaro, X. Turon, S. Tyler, P. Uetz, J. van der Land, B. Vanhoorne, L. van Ofwegen, R. van Soest, J. Vanaverbeke, G. Walker-Smith, T. Walter, A. Warren, G. Williams, S. Wilson, M. Costello, The magnitude of global marine species diversity. *Curr. Biol.* **22**, 2189–2202 (2012).
71. L. G. Cook, R. D. Edwards, M. D. Crisp, N. B. Hardy, Need morphology always be required for new species descriptions? *Invertebr. Syst.* **24**, 322–326 (2010).
72. J. Liu, F. Slik, S. Zheng, D. B. Lindenmayer, Undescribed species have higher extinction risk than known species. *Conserv. Lett.* **15**, e12876 (2022).
73. E. Pante, C. Schoelinck, N. Puillandre, From integrative taxonomy to species description: One step beyond. *Syst. Biol.* **64**, 152–160 (2015).
74. X. Wen-Jie, Z. Chenglong, G. Xueli, W. Baosheng, X. Han, H. Ming-Liang, Z. Honghui, G. Xiaoni, F. Chen-Guang, Z. Jiangmin, B. Jing, H. Lisheng, Q. Qiang, W. Wen, H. Shunping, W. Kun, Chromosome-level genome assembly of hadal snailfish reveals mechanisms of deep-sea adaptation in vertebrates. *eLife* **12**, RP87198 (2023).
75. J. Daniels, N. Aoki, J. Havassy, K. Katija, K. J. Osborn, Metachronal swimming with flexible legs: A kinematics analysis of the midwater polychaete tomopteris. *Integr. Comp. Biol.* **61**, 1658–1673 (2021).
76. D. R. Yoerger, A. F. Govindarajan, J. C. Howland, J. K. Llopiz, P. H. Wiebe, M. Curran, J. Fujii, D. Gomez-Ibanez, K. Katija, B. H. Robison, B. W. Hobson, M. Risi, S. M. Rock, A hybrid underwater robot for multidisciplinary investigation of the ocean twilight zone. *Sci Robot* **6**, eabe1901 (2021).
77. mbari / batho3dr — Bitbucket. https://bitbucket.org/mbari/batho3dr/src/master/.
78. R. Malmstrom, RNAlater Recipe. protocols.io, dx.doi.org/10.17504/protocols.io.c56y9d (2015).
79. H. A. Dahn, J. Mountcastle, J. Balacco, S. Winkler, I. Bista, A. D. Schmitt, O. V. Pettersson, G. Formenti, K. Oliver, M. Smith, W. Tan, A. Kraus, S. Mac, L. M. Komoroske, T. Lama, A. J. Crawford, R. W. Murphy, S. Brown, A. F. Scott, P. A. Morin, E. D. Jarvis, O. Fedrigo, Benchmarking ultra-high molecular weight DNA preservation methods for long-read and long-range sequencing. *Gigascience* **11**, giac068 (2022).
80. D. G. Mulcahy, K. S. M. Iii, S. G. Brady, C. Meyer, K. B. Barker, J. Coddington, Greater than X kb: A quantitative assessment of preservation conditions on genomic DNA quality, and a proposed standard for genome-quality DNA. *PeerJ* **4**, e2528 (2016).
81. M. Manni, M. R. Berkeley, M. Seppey, F. A. Simao, E. M. Zdobnov, BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. arXiv:2106.11799 [quant-ph] (22 June 2021).
82. L. Song, L. Florea, Rcorrector: Efficient and accurate error correction for Illumina RNA-seq reads. *Gigascience* **4**, s13742-015-0089-y (2015).
83. TranscriptomeAssemblyTools, Harvard Informatics, (GitHub, 2023); https://github.com/harvardinformatics/TranscriptomeAssemblyTools.
84. F. Krueger, F. James, P. Ewels, E. Afyounian, M. Weinstein, B. Schuster-Boeckler, G. Hulselmans, FelixKrueger/TrimGalore: v0.6.10, (Zenodo, 2023); https://doi.org/10.5281/zenodo.7598955.
85. C. Quast, E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, F. O. Glöckner, The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).
86. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
87. M. G. Grabherr, B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
88. W. Li, A. Godzik, Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
89. N. Ahuja, X. Cao, D. T. Schultz, N. Picciani, A. Lord, S. Shao, D. R. Burdick, S. H. Haddock, Y. Li, C. W. Dunn, Giants among Cnidaria: large nuclear genomes and rearranged mitochondrial genomes in siphonophores. bioRxiv 540511 [**Preprint**]. 14 May 2023. https://doi.org/10.1101/2023.05.12.540511.
90. A. Desai, V. S. Marwah, A. Yadav, V. Jha, K. Dhaygude, U. Bangar, V. Kulkarni, A. Jere, Identification of optimum sequencing depth especially for de novo genome assembly of small genomes using next generation sequencing data. *PLOS ONE* **8**, e60204 (2013).
91. H. Li, lh3/seqtk v1.4, (GitHub, 2023); https://github.com/lh3/seqtk.
92. A. V. Zimin, D. Puiu, M.-C. Luo, T. Zhu, S. Koren, G. Marçais, J. A. Yorke, J. Dvořák, S. L. Salzberg, Hybrid assembly of the large and highly repetitive genome of Aegilops tauschii, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. *Genome Res.* **27**, 787–792 (2017).
93. A. V. Zimin, G. Marçais, D. Puiu, M. Roberts, S. L. Salzberg, J. A. Yorke, The MaSuRCA genome assembler. *Bioinformatics* **29**, 2669–2677 (2013).
94. D. R. Laetsch, M. L. Blaxter, BlobTools: Interrogation of genome assemblies. *F1000Res.* **6**, 1287 (2017).
95. J. M. Flynn, R. Hubley, C. Goubert, J. Rosen, A. G. Clark, C. Feschotte, A. F. Smit, RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 9451–9457 (2020).
96. A. F. A. Smit, R. Hubley, P. Green, RepeatMasker Open-4.0; http://repeatmasker.org., RMDownload.html (2013).
97. S. Banerjee, P. Bhandary, M. Woodhouse, T. Z. Sen, R. P. Wise, C. M. Andorf, FINDER: An automated software package to annotate eukaryotic genes from RNA-Seq data and associated protein sequences. *BMC Bioinformatics* **22**, 205 (2021).
98. D. W. Barnett, E. K. Garrison, A. R. Quinlan, M. P. Strömberg, G. T. Marth, BamTools: A C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* **27**, 1691–1692 (2011).
99. T. Brůna, K. J. Hoff, A. Lomsadze, M. Stanke, M. Borodovsky, BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom. Bioinform.* **3**, lqaa108 (2021).
100. B. Buchfink, C. Xie, D. H. Huson, Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
101. K. J. Hoff, A. Lomsadze, M. Borodovsky, M. Stanke, Whole-genome annotation with braker. *Methods Mol. Biol.* **1962**, 65–95 (2019).
102. K. J. Hoff, S. Lange, A. Lomsadze, M. Borodovsky, M. Stanke, BRAKER1: Unsupervised RNA-seq-based genome annotation with GeneMark-ET and augustus. *Bioinformatics* **32**, 767–769 (2016).
103. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
104. A. Lomsadze, P. D. Burns, M. Borodovsky, Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Res.* **42**, e119 (2014).
105. M. Stanke, M. Diekhans, R. Baertsch, D. Haussler, Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637–644 (2008).
106. M. Stanke, O. Schöffmann, B. Morgenstern, S. Waack, Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **7**, 62 (2006).
107. P. Jones, D. Binns, H.-Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S.-Y. Yong, R. Lopez, S. Hunter, InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
108. R. Patro, G. Duggal, M. I. Love, R. A. Irizarry, C. Kingsford, Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
109. R. C. Team, R: A language and environment for statistical computing (2021).
110. C. Soneson, M. I. Love, M. D. Robinson, Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences. *F1000Res* **4**, 1521 (2015).
111. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

112. A. Alexa, J. Rahnenfuhrer, topGO: Enrichment Aanalysis for Gene Ontology, version 2.52.0, Bioconductor version: Release (3.17) (2023); https://bioconductor.org/packages/topGO.
113. W. B. F. Ryan, S. M. Carbotte, J. O. Coplan, S. O'Hara, A. Melkonian, R. Arko, R. A. Weissel, V. Ferrini, A. Goodwillie, F. Nitsche, J. Bonczkowski, R. Zemsky, Global multi-resolution topography synthesis. *Geochem. Geophys. Geosyst.* **10**, Q03014 (2009).

Acknowledgments: We thank J. Williams, N. Bingham, S. Alesandrini, J. Rodriguez, E. Campbell, D. Cobain, C. Peyres, A. Wetmore, K. Ingram, A. Miller, C. Wiener, and H. Nolan from the Schmidt Ocean Institute. We are grateful to L. Aluwihari at the Scripps Institute for Oceanography for giving us space to make the RNALater that we used on the cruise and to graduate student B. Peng for helping organize and receive shipments of chemicals needed. We thank K. Osborn of the Smithsonian Institution for helpful comments regarding tomopterid taxonomy. We additionally thank J. Williams of the Schmidt Ocean Institute for generating the ROV model indicating the placement of our instruments used in Fig. 1. **Funding:** This work was supported by the following: Schmidt Ocean Institute grant "ROV-based 3D reality capture, specimen encapsulation, and tissue voucher sampling to explore and describe midwater biodiversity in the deep sea" to B.T.P., K.K., R.J.W., and D.F.G. and its corresponding research cruise "Designing the Future 2" (https://schmidtocean.org/cruise/designing-the-future-2/); David and Lucile Packard Foundation (to K.K.); and Gordon and Betty Moore Foundation (no. 7583 to K.K.). J.A.B. was supported in part by NSF OIA-1826734. The National Geographic Society provided support (grant no. SP 12-14) to R.J.W. and D.F.G and NSF Instrument Development for Biological Research Award no.1556164 to R.J.W. and no. 1556123 to D.F.G.
Author contributions: Conceptualization: D.M.V., B.G., K.K., R.J.W., J.A.B., R.W., Z.E.T., K.P.B., D.F.G., and BTP. Investigation: D.M.V., D.C., P.R., A.H.Y., B.G., K.K., J.A.B., K.P.B., E.O., and B.T.P. Writing—review and editing: D.M.V., K.K., R.J.W., J.A.B., Z.E.T., E.O., D.F.G., and B.T.P. Methodology: D.M.V., P.R., A.H.Y., B.G., K.K., R.J.W., J.A.B., R.W., Z.E.T., K.P.B., D.F.G., B.T.P., and J.D. Resources: D.M.V., D.C., B.G., K.K., J.A.B., R.W., K.P.B., D.F.G., and B.T.P. Validation: D.M.V., D.C., A.H.Y., R.J.W., J.A.B., R.W., E.O., D.F.G., B.T.P., and J.D. Software: D.M.V., D.C., P.R., K.K., J.A.B., R.W., E.O., B.T.P., and J.D. Project administration: D.M.V., D.C., K.K., R.J.W., J.A.B., Z.E.T., D.F.G., and B.T.P. Writing—original draft: K.K., R.J.W., J.A.B., Z.E.T., D.F.G., and B.T.P. Funding acquisition: K.K., R.J.W., D.F.G., and B.T.P. Data curation: K.K., J.A.B., E.O., B.T.P., and J.D. Supervision: K.K., R.J.W., J.A.B., D.F.G., and B.T.P. Visualization: K.K., R.J.W., J.A.B., Z.E.T., E.O., D.F.G., B.T.P., and J.D. Formal analysis: J.A.B. and B.T.P. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** This Whole Genome Shotgun project (*Pegia* sp.) has been deposited at DDBJ/ENA/GenBank under the accession JAQGEQ000000000. The version described here is version JAQGEQ010000000. Reference transcriptome assemblies and associated read files for other organisms can be found under NCBI BioProject PRJNA919492 and in Zenodo repositories linked via table S13: *Pegea* sp. genome, transcriptome, and annotations: https://doi.org/10.5281/zenodo.8058080; Tomopteris sp. transcriptome assembly and annotations: https://doi.org/10.5281/zenodo.8058120; *Marrus claudanielis* transcriptome assembly and annotations: https://doi.org/10.5281/zenodo.8350091; *Erenna* sp. transcriptome assembly and annotations: https://doi.org/10.5281/zenodo.8058141. Generated models and references to the video data can be found on MBARI's Sketchfab page (https://bit.ly/46t7jWF). Code associated with reconstructing 3D models from imaging data can be found on Bitbucket (https://bitbucket.org/mbari/batho3dr). All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Submitted 29 June 2023
Accepted 19 December 2023
Published 17 January 2024
10.1126/sciadv.adj4960