Global genomics of man-o'-war (*Physalia*) reveal biodiversity at the ocean surface

Samuel H. Church[1,*], River B. Abedon[1], Namrata Ahuja[1], Colin J. Anthony[2], Diego A. Ramirez[1], Lourdes M. Rojas[3], Maria E. Albinsson[4], Itziar Álvarez Trasobares[5], Reza E. Bergemann[1], Ozren Bogdanovic[6], David R. Burdick[2], Tauana J. Cunha[7], Alejandro Damian-Serrano[8], Guillermo D'Elía[9], Kirstin B. Dion[1], Thomas K. Doyle[10], João M. Gonçalves[11], Alvaro Gonzalez Rajal[12], Rebecca R. Helm[13], Diane D. G. Le Gouvello[14], Steven H. D. Haddock[15,16], Zachary R. Lewis[17], Javier Sellanes[18], Bruno I. M. M. Magalhães[11], Maciej K. Mańko[19], Alex de Mendoza[20], Carlos J. Moura[11], Ronel Nel[14], Jessica N. Perelman[21,22], Laura Prieto[5], Catriona Munro[23], Kohei Oguchi[24], Kylie A. Pitt[25,26,27], Amandine Schaeffer[28], Andrea L. Schmidt[21,22], Nerida G. Wilson[29,30], Gaku Yamamoto[31], Eric A. Lazo-Wasem[3], Chris Simon[32], Mary Beth Decker[1], Jenn M. Coughlan[1], Casey W. Dunn[1,3]

* corresponding author: samuelhchurch@gmail.com

1. Yale University, Department of Ecology and Evolutionary Biology, New Haven, Connecticut, USA
2. University of Guam, Marine Laboratory, Guam
3. Yale University, Yale Peabody Museum, New Haven, Connecticut, USA
4. Tassal, Animal Welfare and Genetics, Hobart, Australia
5. Consejo Superior de Investigaciones Científicas (CSIC), Instituto de Ciencias Marinas de Andalucía (ICMAN), Cádiz, Spain
6. CSIC-Universidad Pablo de Olavide-Junta de Andalucía, Centro Andaluz de Biología del Desarrollo, Seville, Spain
7. Field Museum of Natural History, Chicago, Illinois, USA
8. University of Oregon, Institute of Ecology and Evolution, Department of Biology, Oregon, USA
9. Universidad Austral de Chile, Instituto de Ciencias Ambientales y Evolutivas, Valdivia, Chile
10. University College Cork, School of Biological, Earth and Environmental Sciences, Cork, Ireland
11. Universidade dos Açores, Okeanos-UAc – Instituto de Investigação em Ciências do Mar, Horta, Açores, Portugal

12. University of New South Wales Sydney, School of Biotechnology and Biomolecular Sciences (BABS), Faculty of Science, Sydney, Australia
13. Georgetown University, Earth Commons, Washington D. C., USA
14. Nelson Mandela University, Department of Zoology, Port Elizabeth, South Africa
15. Monterey Bay Aquarium Research Institute, Moss Landing, California, USA
16. University of California Santa Cruz, Department of Ecology and Evolutionary Biology, Santa Cruz, California, USA
17. Allen Institute for Brain Science, Molecular Genetics, Seattle, Washington, USA
18. Universidad Católica del Norte, Departamento de Biologia Marina, El centro de Ecología y Manejo Sustentable de Islas Oceánicas (ESMOI), Coquimbo, Chile
19. University of Gdańsk, Department of Marine Biology and Biotechnology, Gdańsk, Poland
20. Queen Mary University of London, School of Biological and Behavioural Sciences, London, England
21. Cooperative Institute for Marine and Atmospheric Research (CIMAR), University of Hawaii, Honolulu, Hawaii, USA
22. NOAA Pacific Islands Fisheries Science Center (PIFSC), National Marine Fisheries Service, Honolulu, Hawaii, USA
23. Sorbonne Université, Centre national de la recherche scientifique (CRNS), Laboratoire de Biologie du Développement de Villefranche-sur-Mer (LBDV), Villefranche-sur-Mer, France
24. University of Tokyo, Misaki Marine Biological Station, Miura, Japan
25. Griffith University, Australian Rivers Institute, Queensland, Australia
26. Griffith University, Coastal and Marine Research Centre, Queensland, Australia
27. Griffith University, Griffith School of Environment and Science, Queensland, Australia
28. University of New South Wales Sydney, Mathematics and Statistics, Sydney, Australia
29. University of Western Australia, School of Biological Sciences, Western Australia, Australia
30. Western Australian Museum, Collections and Research, Western Australia, Australia
31. Enoshima aquarium, Exhibitions and breeding, Fujisawa, Japan
32. University of Connecticut, Department of Ecology and Evolutionary Biology, Storrs, Connecticut, USA

Abstract

The open ocean is a vast, highly connected environment, and the organisms found there have been hypothesized to represent massive, well-mixed populations. Of these, the Portuguese man-o'-war (*Physalia*) is uniquely suited to dispersal, sailing the ocean surface with a muscular crest. We tested the hypothesis of a single, panmictic *Physalia* population by sequencing 133 genomes, and found five distinct lineages, with multiple lines of evidence showing strong reproductive isolation despite range overlap. We then scored thousands of citizen-science photos and identified four recognizable morphologies linked to these lineages. Within lineages, we detected regionally endemic subpopulations, connected by winds and currents, and identified individual long-distance dispersal events. We find that, even in these sailing species, genetic variation is highly partitioned geographically across the open ocean.

Summary

The open ocean is a vast and highly connected environment. The organisms that live there have a significant capacity for dispersal and few geographic boundaries to separate populations. Of these, the Portuguese man-o'-war or bluebottle (genus *Physalia*) is uniquely suited to long-distance travel, using its gas-filled float and muscular crest to catch the wind and sail the sea surface. *Physalia* are distributed across the globe, and like many pelagic organisms, have been hypothesized to represent a massive, well-mixed population that extends across ocean basins. We tested this hypothesis by sequencing whole genomes of 133 samples collected from waters of over a dozen countries around the globe. Our results revealed five distinct lineages, with multiple lines of evidence indicating strong reproductive isolation, despite regions of range overlap. We combined these data with an independent dataset of thousands of images of *Physalia* uploaded to the citizen-science website `inaturalist.org`, which we scored for morphological characters including sail size, tentacle arrangement, and color. From these images, we identified four recognizable morphologies, described their geographical distribution, and linked them to four of the lineages identified with genomic data. We propose there are at least four species, three of which correspond to species hypothesized by scientists in the 18th and 19th centuries: *P. physalis*, *P utriculus*, and *P. megalista*, along with one as yet unnamed species *Physalia sp.* from the Tasman Sea. Within each species, we observe significant population structure, with evidence of persistent subpopulations at a regional scale, as well as evidence for individual long-distance dispersal events. Our findings indicate that, instead of one well-mixed, cosmopolitan species, there are in fact multiple *Physalia* species with distinct but overlapping ranges, each made up of regionally endemic subpopulations that are connected by major ocean currents and wind patterns.

Main text

Introduction

The open ocean has few geographic barriers that might limit connectivity (*1*). The organisms that live there often have strong dispersal potential (*2*) and massive effective population sizes (*3*), contributing to the assumption that populations are predominantly well-mixed, even at a global scale. However, a series of recent studies have found evidence for population structure in the open ocean, despite the absence of geographic barriers (*4–6*). The studies challenge expectations of uninterrupted gene flow and bolster claims that open-ocean diversity has routinely been underestimated (*7*).

Studies of oceanic population structure have largely focused on benthic and planktonic species (either holoplanktonic or planktonic in the larval stage), meaning far less is known about populations that live at or near the ocean surface (*8*), collectively termed neuston (*9*). The surface ecosystem represents a massive and biologically rich environment, and the physical processes at play at the air-water interface (e.g. winds, surface currents) have distinct potential to mediate dispersal. At the same time, the ocean surface ecosystem is imperiled by plastics and pollutants that aggregate there, as well as efforts to clean pollutants at a large scale (*10*). A common but unproven justification for potentially destructive clean-up efforts is that there is relatively little diversity at the ocean surface, and the organisms present there have robust population sizes (*11*). It is urgent that we evaluate this claim by examining genetic diversity at the surface to build informed strategies moving forward (*12*).

Bluebottles or Portuguese man o' war, cnidarians in the genus *Physalia*, present a compelling test case for exploring open-ocean population structure. They are among the few invertebrates to utilize wind-powered movement, sailing the ocean surface with a muscular crest, and they are the largest to do so, making them particularly capable of long-distance dispersal. There is only one species of *Physalia* currently recognized, with a hypothesized global population that extends across the Atlantic, Indian, Pacific, and Southern Oceans (*13–15*). However, a recent study, analyzing marker genes from samples around New Zealand, found preliminary evidence of substantial genetic variation, even within a relatively small geographic area (*16*). A global analysis of genomic variation in *Physalia* therefore represents a prime test for the existence of a globally panmictic population, targeting a widespread taxon with a significant capacity for long-distance dispersal (*17*).

Physalia populations are potentially influenced by the dynamics both at and below the ocean surface. Reproduction occurs below the surface, as reproductive structures (gonodendra) separate from the main body, sink, and release gametes into the water column (*18*). Following fertilization, juvenile *Physalia* return to the surface using specialized gas-producing tissues to inflate their nascent float (*19*). Growth occurs through the addition of asexually-budded, clonal bodies that remain integrated to one another through shared nervous and gastric systems, similar to a colony of coral, but in *Physalia* these bodies perform specialized functions (e.g., reproduction, prey-capture, digestion) (*13*). Mature *Physalia* colonies are key predators

within the neuston assemblage (*20*), extending their tentacles up to tens of meters into the water column to kill and retrieve fish prey (*21*). Onshore winds can blow these colonies onto beaches, often in large numbers (*22*), where, given their potent sting, they present a medical risk to humans and affect tourism via beach closures. These impacts create an additional need to understand the factors influencing their dispersal and distribution (*23–25*).

Variation in colony size is associated with ocean basins (e.g., Atlantic specimens are typically the largest). Two alternative hypotheses can explain this pattern (*13*, *15*): [1] the large *Physalia* in certain parts of the world represent the oldest colonies, ones that sailed in from elsewhere, or [2] there are distinct populations in different regions that mature at different sizes. New technologies make it possible to distinguish between these hypotheses, including increased efficiency of next-generation sequencing that has made it feasible to collect genomic data despite their large genome size (estimated at 2-3Gb). In addition participatory science on the internet has generated thousands of images of *Physalia* from beaches and waters around the world (Fig. 1B). In this study, we evaluate the population structure and diversity of *Physalia* by evaluating two independent datasets of *Physalia* diversity: [1] whole genome sequencing of 133 specimens, and [2] morphological data from more than 4,000 images submitted via participatory science to the natural history website `inaturalist.org`. We test for evidence of multiple species associated with distinct morphs, describe their ranges and distributions, and analyze the spatiotemporal dynamics within each lineage.

Results

Reference genome

We generated a new genome assembly for *Physalia* from a specimen collected in Texas, USA in 2017. This assembly, along with its alternate haplotype counterpart, has high contiguity (N50 of 10.4 and 4.6 megabases, see Table S1) and high `BUSCO` completeness scores (89.7% and 86.9%, Table S2). The length of the primary and alternate assemblies are 3.33 and 2.69 gigabases (Gb) respectively, and like other siphonophores (*26*), the *Physalia* genome is characterized by a substantial fraction of repeat sequences (~65%, Fig. S1). Given the potential for these repeats to affect assembly size, we used a _k_mer analysis to estimate the genome size of ten specimens that were sequenced to a sufficient depth, and found that genome sizes vary between 1.5 and 2.0 gigabases, Gb, indicating an artificially inflated number of repeat sequences in the assembly (note however that these estimates are significantly smaller than previous estimates based on flow cytometry that estimated the size as 3.2 Gb (*27*)). To account for potential repeat inflation, for all downstream analyses we used only reads mapped to non-repeat regions of the primary assembly.

In addition to the genome assembly, we also generated a new transcriptome using full-length cDNA generated with PacBio Iso-Seq on an additional specimen of *P. physalis*, collected in Florida in 2023. To test the robustness of our results to reference assembly, analyses were repeated over both the genome and transcriptome assemblies.

Distinct lineages

To test hypotheses about *Physalia* diversity and population structure, our global team collected >350 specimens, the majority of which were accessioned at the Yale Peabody Museum (Fig 1, and see supplementary text). We sequenced the genomes of 123 samples and performed a principal component analysis (PCA). The results show samples are divided into five clusters along the first two principal components of genomic variation (Fig. 1C, S2). These five clusters are labeled as A, B1, B2, C1, and C2, given the adjacency of the latter pairs to one another along principle components. We repeated the PCA including an additional ten sequenced samples found to be of moderate (rather than high) quality, and observed the same five clusters (Fig. S3). We also repeated the analysis mapping reads to the Iso-Seq transcriptome reference, and observed the same results (Fig. S4), indicating that population genomic studies similar to those presented here may not require reference genomes.

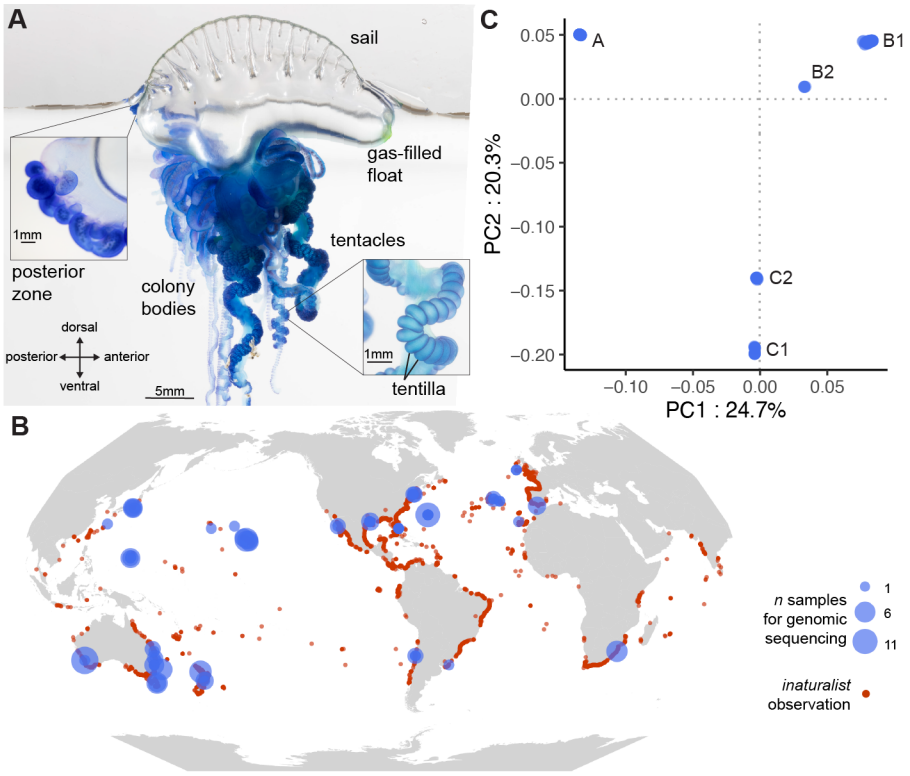

Figure 1: Biology, distribution, and genomic variation of *Physalia*. A, *Physalia* colonies comprise a muscular sail attached to a gas-filled float which maintains the mature animal at the surface of the water. Colony bodies (zooids), including those specialized for feeding (gastrozooids), prey-capture (palpons with tentacles), and reproduction (gonozooids) are added to the float via asexual reproduction at growth zones. Tentacles drape below the float to trap, sting, and retrieve fish using batteries of stinging cells contained in tentilla. Photos of *Physalia sp.* C2, specimens YPM-IZ-111236 (main), YPM-IZ-111237 (growth zone), and YPM-IZ-111240 (tentacle). B, *Physalia* are observed throughout the world, as shown by observations posted to `inaturalist.org` (red). Samples for genomic analysis (blue) were collected by a global collaboration of scientists. C, The first three principal components of variation reveal five lineages, labeled as clusters A, B1, B2, C1, and C2.

Genomic differentiation

The geographic distribution of the five lineages shows that at least two clusters were observed across multiple ocean basins (Fig. 2A): cluster B1 was found in the S. Atlantic, S. Indian, and S. and N. Pacific; cluster C1 was found on both sides of the S. Indian and S. Pacific oceans. By contrast, cluster A was observed only in the Atlantic, B2 on the northernmost sampling locations on both sides of the N. Pacific, and C2 only in New Zealand and Tasmania. We

evaluated genomic differentiation by calculating the reciprocal fixation index (*Fst*), averaged across non-repeat windows. Average *Fst* values range from 0.29 between B1 and B2, to 0.64 between A and C1, suggesting little genetic exchange between any pair of lineages (Fig. 2B, see Fig. S5 for range across genomic windows). Estimates of nucleotide diversity, *pi*, indicate that cluster A has the lowest overall diversity and clusters B1 and C2 have the highest (Fig. S6), consistent with estimates of individual heterozygosity (Fig. S1B).

We tested the monophyly and phylogenetic relationships of these lineages using two approaches. First, we assembled mitochondrial genomes for each sample and inferred a mitochondrial tree. For this analysis, we combined the mitogenomes generated in this study with all publicly available *Physalia* mitogenomes, as well as mitogenomes from *Rhizophysa* as an outgroup. The most likely tree shows clusters are monophyletic, with relatively little sequence variation within clusters (Fig. 2C). The clusters B1 and B2 were found to be sister to one another with high bootstrap support, and the clade of B1+B2 sister to cluster A. Support values were lower (bootstrap of 82) for the relationships at the base of the *Physalia* phylogeny.

Second, we estimated the phylogeny from a dataset of 800k high-quality SNPs, using the coalescent-based software `SVDQuartets`. A phylogeny of all specimens confirmed the reciprocal monophyly of the five lineages (Fig. S7). We examined the relationships between lineages by estimating a tree with individuals assigned to their respective clusters. Our results again indicated a split between the clade (C1, C2), and the clade of (A and B1+B2) (Fig. S7C). Support values for both partitions in this unrooted tree showed unanimous support (bootstrap of 100).

We used an admixture analysis to understand how genetic ancestries are partitioned across these lineages. The results favored five ancestry groups, corresponding to the five clusters above, and showed little evidence of mixture between groups (Fig. 2D). Repeating this analysis including the ten samples of moderate quality returned the same general results (Fig. S3), with the exception of three moderate-quality specimens of C2 that showed a minor proportion of admixture with C1. Repeating analyses using the reference transcriptome returned the same results (Fig. S4).

Several studies have generated data on individual genetic markers from *Physalia* (*16*, *28–31*). In order to place those data in the context of our findings, we inferred individual gene trees for four genes: mitochondrial CO1 and 16S, and nuclear ITS and 18S (Fig. S8-S11). We combined publicly available sequences from the National Center for Biotechnology Information (NCBI) with assembled marker sequences from our specimens, inferred using *in silico* PCR as implemented in our custom software `sharkmer`. This tool uses PCR primer sequences to seed a de Bruijn graph assembly of raw sequencing reads. These results furthered our understanding of *Physalia* diversity in the following ways: [1] a specimen reported from the Sargasso Sea (N. Atlantic) extended the predicted range of B1; [2] a specimen reported from Pakistan (N. Indian) extended the predicted range of B2; [3] using the internally transcribed spacer gene ITS, we were able to assign three clans, described in New Zealand (*16*), to clusters we describe here: clan 1 = cluster C2, clan 2 = cluster B1, clan 3 = cluster C1; however, using COI we found an incongruent result for the identity of clan 3. Without further information we cannot

determine whether this result is due may be due to a potential exchange of mitochondrial sequences between clusters C1 and C2 in New Zealand.

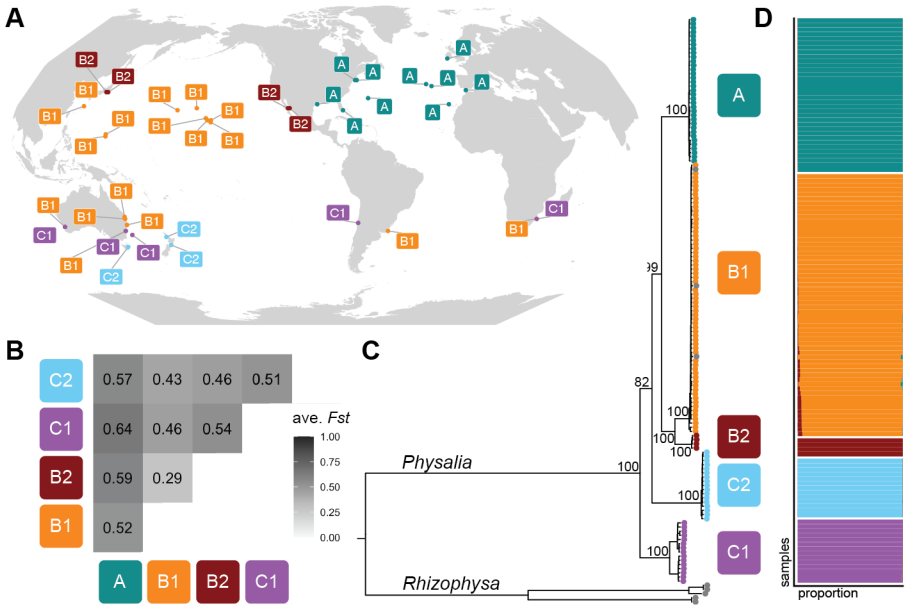

Figure 2: Multiple lines of evidence indicate reproductive isolation between lineages. A, The distribution of the five clusters from Fig. 1 shows some lineages span ocean basins (e.g., B1, C1) and others are restricted to smaller areas (e.g., C2 observed in New Zealand and Tasmania). B, Reciprocal fixation index (*Fst*) averaged across non-repeat genomic windows indicate high levels of reproductive isolation between all lineages, with the weakest between B1 and B2. C, Phylogenetic analysis of 141 mitochondrial genomes shows reciprocal monophyly of lineages. Bootstrap values are shown at internal *Physalia* nodes. D, Admixture analysis of 123 samples recovers five lineages with little evidence of mixture.

Morphology-based analysis

We tested for the evidence of distinct morphologies of *Physalia* by analyzing a dataset of images of *Physalia* uploaded to the citizen-science website `inaturalist.org`. While most of these images are of beached specimens, many aspects of the gross morphology are often preserved and identifiable. We scored the following characters (Fig. 3A): the height and length of the sail relative to the float; the color of the sail apex and the colony bodies (primarily gastrozooids); the arrangement of principle tentacles (defined as those with dense aggregations of tentilla); and the visible presence of a gap between the posterior and main zone of the colony. To ensure reproducibility of scoring, we had three independent observers score the same set of 100 images. We scored characters on a dataset of 4,047 images, selected to include multiple images from all

represented countries and time zones, with additional images scored for locations hypothesized to have increased diversity (e.g., New Zealand (*16*))

From these images, we identified four distinct morphologies (Fig. 3A-C). These were defined by describing a series of rules for positive identification based on suites of characters (Fig. S12), excluding images of poor quality or of specimens scored as having juvenile characteristics (e.g., globular float, few zooids). These rules constitute a strict definition for a high-confidence observation of each type; for example, images were positively identified as the *P. physalis* morphology if they had reddish feeding bodies, multiple major tentacles, and a sail that is as tall as the float and extends nearly to the anterior end. While other specimens of *P. physalis* may deviate from these characters (e.g., if the sail is not raised), the rules were designed to minimize overlap between morphologies and allow for high-confidence identifications.

Three of the morphologies we identified correspond to species hypothesized by scientists centuries ago (*15*). *P. physalis* was named by Linneaus in 1758 based on specimens from the Atlantic that had large sails and multiple major tentacles (Fig. 3A). *P. utriculus* was named by Gmelin (1788) (*32*), based on illustrations by La Martinière (1787) (*33*) of a Pacific specimen collected on the Lapérouse expedition that had a single major tentacle, yellow-tipped gastrozooids, and a flared posterior growth zone. *P. megalista* was named and illustrated by Lesueuer and Petit (1807) (*34*)] from specimens from the Southern Ocean that had an incomplete sail and a sinuously postured float. Each of these species names was synonymized with *P. physalis* in later centuries (*13*, *15*, *35*, *36*); our results indicate these synonymizations to have been incorrect.

We linked these morphotypes to clusters identified through genome sequencing by analyzing the morphology of specimens we had analyzed genetically, using images taken upon collection, when available, as well as the morphology of fixed specimens (Fig. 3D, see supplementary text for images). Our results confirm that cluster A corresponds to *P. physalis*, B1 to *P. utriculus*, C1 to *P. megalista*, and C2 to *P. sp.*. Cluster B2 could not be assigned given that no images of specimens were taken upon collection; analysis of the morphology of the single available fixed specimen suggested a general similarity to specimens of B1, *P. utriculus*.

Based on the assignment of morphotypes to clusters, we re-examined the distribution of the lineages using positively identified images (Fig. 3C). We found that cluster A, *P. physalis* was observed in the N and SW Atlantic; B1, *P. utriculus* was found throughout the Pacific and Indian and extending into the SW Atlantic and Gulf of Mexico; C1, *P. megalista* was found in the Southern edges of the Atlantic, Indian, and Pacific; and C2, *P. sp* was found in New Zealand, Tasmania, and E Australia.

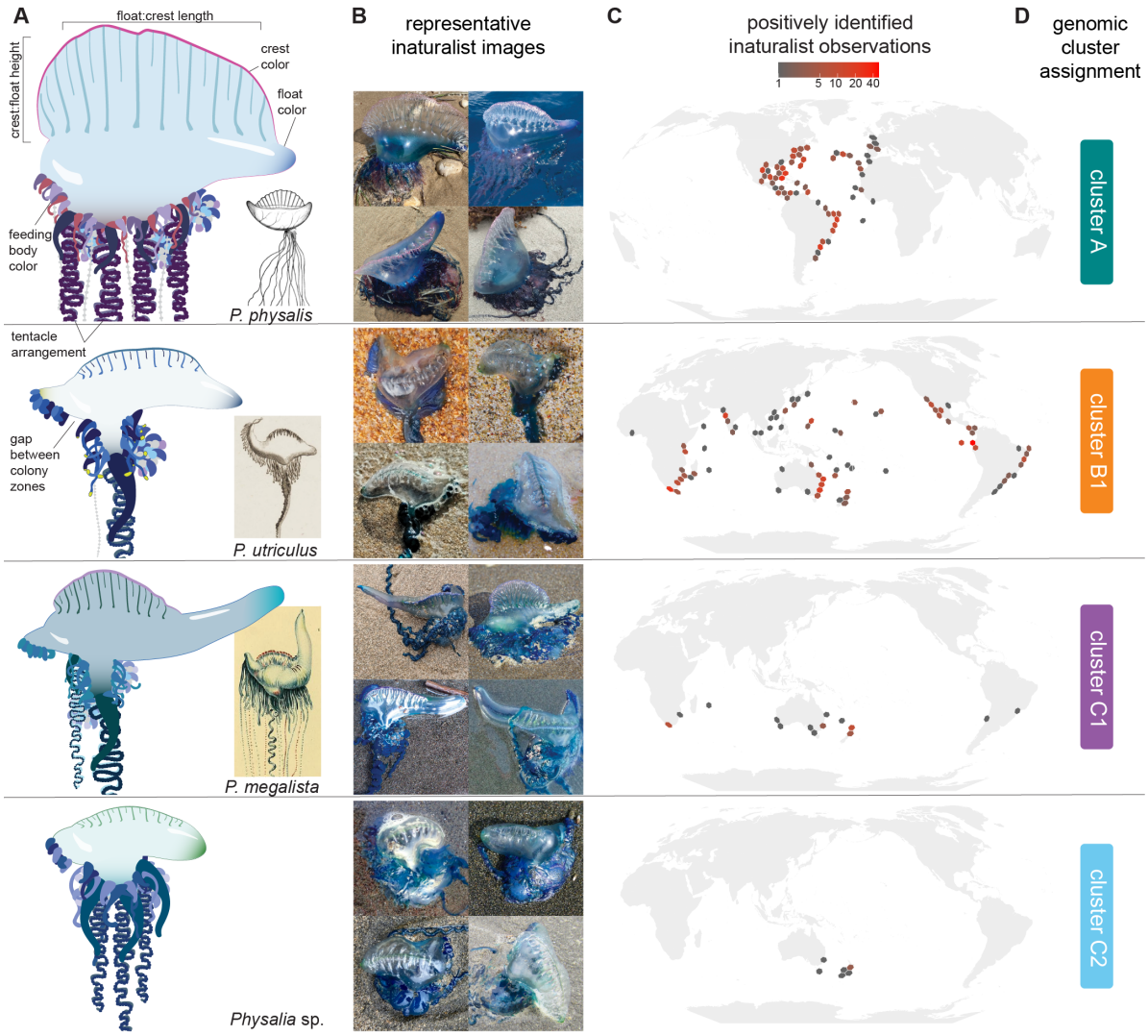

Figure 3: Distinct morphologies are detectable in citizen science images. A, Morphological traits such as aspects of size, color, and tentacle number were scored for thousands of images on `inaturalist.org`. From these, four morphologies were identified, three of which correspond to historically hypothesized species. B, Representative photos of each morphology from `iNaturalist`, image credits listed below. C, Ranges of positively identified `iNaturalist` records for each morphology, using a rule-based analysis of morphological traits. D, Morphologies were assigned to a genomic cluster by scoring the same traits of genomic specimens. Cluster B2 could not be definitively assigned due to lack of images and fixed material.

Geographic population structure

Given that these animals can move with wind and currents, combined with the evidence of distributions extended across ocean basins, we tested for evidence of long-distance dispersal and subpopulation structure by performing PCA within each of the four species: *P. physalis*, *P. utriculus*, *P. megalista*, and *P. sp* C2. For *P. utriculus* we repeated this analysis both including and excluding cluster B2 (Fig. S13). Within species, samples are largely grouped by geographic region (Fig. 4), and not by date of collection (Fig. S14). The observation of a strong geographic signature, persistent even at sites with collection events over the span of multiple years, suggests that *Physalia* subpopulations largely stay in the same region over time. The extent of these geographic regions appears linked to major ocean currents; for example *P. physalis* samples from Florida, Bermuda, and New England are highly similar, without substructure corresponding to collection sites, indicating these samples are part of one large subpopulation aligned with the Gulf Stream current (Fig. 4A-B) and wind patterns.

We observed several individual exceptions to the pattern of persistent regional subpopulations within the dataset, indicating individuals can disperse over long distances. In *P. physalis*, two of the samples collected in Bermuda showed an Eastern Atlantic subpopulation signature, and one sample in the Azores had a Western Atlantic signature, suggesting dispersal events in both directions across the N. Atlantic. In *P. utriculus*, one sample collected in Hawai'i showed a genomic signature associated with samples collected in Guam and Japan, suggesting an individual eastward dispersal event across the N. Pacific. In *P. megalista*, one sample collected in Eastern Australia had an Indian Ocean signature, suggesting dispersal across ocean basins.

We also tested the strength of differentiation between subpopulations within species. Using a k-means clustering analysis, we identified two subpopulations within *P. physalis*, three in *P. utriculus*, and two each in *P. megalista* and *P. sp* C2 (Fig. S15). Lineage B2 was treated as a single cluster, given limited sampling. Genomic differentiation between subpopulations (calculated as average *Fst*) was small (<0.05), with the exception of the division within *P. megalista* (average *Fst* of 0.12, Fig. S16), a division also reflected in the mitochondrial and nuclear phylogenetic results (Fig. 2C, S7A), suggestive of a barrier to gene flow within *P. megalista*. We also examined genomic differentiation between specimens that shared the same sampling region (Fig. S17). This comparison confirmed that genomic differentiation is not lower for groups of specimens collected in the same region (e.g., *utriculus* and *megalista* that co-occur in the SW Pacific, *Fst* value of 0.43).

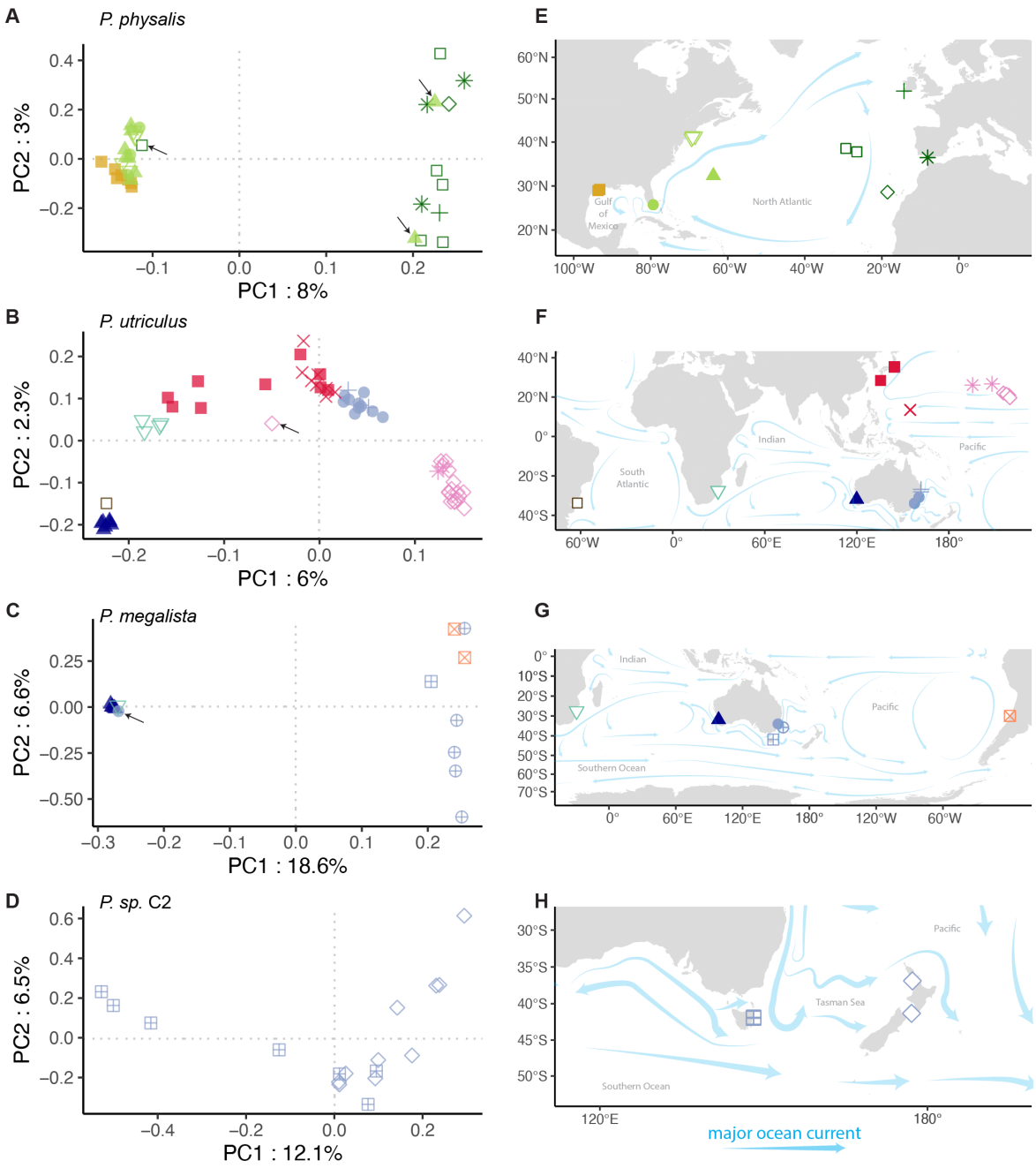

Figure 4: Principal component analyses within species show that subpopulations are largely defined by region. Exceptions to this pattern are marked with black arrows; these individuals suggest long-distance dispersal events across regions. A, *P. physalis*, B, *P. utriculus*, C, *P. megalista*, and D, *P. sp.* C2. Colors indicate regions of the ocean (e.g., Northwest Atlantic), and shapes indicate sampling location (e.g., Florida). E-H, Corresponding sampling locations and major ocean currents are shown, currents from National Oceanic and Atmospheric Administration (*37*)).

Discussion

This study, targeting an organism capable of long distance dispersal via ocean currents and winds, suggests that panmictic, cosmopolitan populations are the exception, and not the rule, for marine invertebrates. Our results show multiple lines of evidence that there are at least four species of *Physalia*, each composed of regionally endemic subpopulations. These lines of evidence include high genomic differentiation (measured as average reciprocal *Fst*), reciprocal monophyly of mitochondrial and nuclear phylogenies, clear morphological differentiation, and consistent mapping between genomic and morphological groupings. The four species we identify have distinct but overlapping ranges: *Physalia physalis* from the Atlantic; *P. utriculus*, present in the Pacific, Indian, and Atlantic; *P. megalista*, present in the Southern Ocean, and *P. sp.* C2, present in the Tasman Sea. We also find evidence to suggest a potential fifth species (cluster B2), but the absence of morphological data for genome sequenced species, combined with the relatively lower *Fst* value and phylogenetic proximity to *P. utriculus*, precludes its designation at this time.

Within species we observe genomic signatures endemic to specific regions, which are persistent over multiple years of sampling. In the case of *P. utriculus* in Hawai'i, we collected reproductively mature adults (e.g., YPM-IZ-110777), juveniles collected before surfacing (e.g. YPM-IZ-110881, YPM-IZ-110882, YPM-IZ-110883, collected at ~6 meters depth), and a range of sizes in between (see supplementary text), and observed that they shared the same subpopulation signature, suggesting reproduction happens *in situ*. Regional genomic signatures are robust despite our observation of five individual specimens with incongruent genomic and geographical signatures that suggest cross-regional dispersal events. Subpopulation boundaries appear to be defined by patterns of winds and currents, as demonstrated by the close genetic affinity of samples collected at sites adjacent to the Gulf Stream (Texas, Florida, Bermuda, and New England).

The southern hemisphere, and in particular the southwest regions of ocean basins, consistently represent hotspots of *Physalia* diversity: three species are found in the SW Pacific (*P. utriculus*, *P. megalista*, and *P. sp.* C2), two species in the SW Indian Ocean (*P. utriculus* and *P. megalista*), and three species in the SW Atlantic (*P. physalis*, *P. utriculus*, and *P. megalista*). Furthermore, the phylogenetic relationships between species suggest that diversification may have originated from the Pacific Ocean. The most recent species division is between *P. physalis* and *P. utriculus*, potentially as the ancestral population of these lineages came to occupy the Atlantic Ocean. In addition, we observe moderate genomic differentiation between clusters B1 and B2 in the N Pacific (ave. *Fst* of 0.29, Fig. 2), and between subpopulations of *P. megalista* in the Pacific and Indian Oceans (ave. *Fst* of 0.12, Fig. S16), suggestive of nascent reproductive isolation events.

This study builds on the work and observations of sailors, swimmers, and scientists over the course of centuries. As early as the 18th century, hypotheses about multiple species emerged, based on reports from global voyages (*15*). Among these are three of the species we observed, *P. physalis*, *utriculus*, and *megalista*. These species were not "cryptic"; they

were hypothesized, debated, and ultimately rejected over the course of 250 hundred years. Our results vindicate the original descriptions, showing clear and strong support for distinct species matching the original illustrations. The central challenge faced by taxonomists in past centuries was that there was no way to simultaneously observe live or recently beached *Physalia* across its huge range, and key characteristics like posture, color, and behavior are lost during fixation. These results underscore the power of participatory science and social media to provide an unprecedented lens on biodiversity.

Conflicting expectations and observations of the number of planktonic species have spawned multiple discourses in the literature (e.g., "the paradox of the plankton" (*38*) and its companion, "the inverted paradox of the plankton" (*39*)). Here we demonstrate that, in the case *Physalia*, there is more diversity than previously assumed (four species instead of one), and that the open ocean ecosystem might indeed have high evolutionary potential (*4*). Across the open ocean we observe substantial geographic partitioning of genetic variation, evidence for reproductive isolation events that have resulted in strong barriers to reproduction, as well as events that suggest isolation may be currently underway (e.g. clusters B1 and B2), and in the case of *P. sp* C2, we report a previously undescribed species that represents a single-sea endemic. Future research into the physical, environmental, and biological processes that generate and maintain this genetic variation will be crucial in recalibrating our expectations towards open-ocean biodiversity.

Acknowledgments

We thank Adachi Aya, Stephen Chiswell, Lynne Christianson, Andy Coetzee, Leanne Elder, Steve Gerber, Rosemary Hill, Kazuhisa Hori, Michiyo Kawabata, Mitsugu Kitada, Izumi Komori, Hiromi Kurihara, Yukio Kurihara, Juan Moles, Natasha Picciani, Nicola Puharich, Fae Spafford, John Stimson, Kosuke Uchiwa, the Miura Fisher Cooperative Association, and the East Matunuck State Beach lifeguards for additional support in collecting specimens. We thank Lawrence Gall, Giselle Caccone, and the Caccone laboratory for early feedback. We thank the Western Australia Museum, Tasmanian Museum and Art Gallery, and the Field Museum of Natural History for loaning specimens. We also thank Yale Center for Genome Analysis and Keck Microarray Shared Resource at Yale University for providing the necessary sequencing services.

Image credits

Images reproduced in Figure 3B were retrieved from https://inaturalist.org and are all made available under Creative Commons licence CC-0 or CC-BY. They are credited as follows, left to right and top to bottom:

1. Eileen Mattei, Texas, USA. CC-0. https://inaturalist-open-data.s3.amazonaws.com/photos/271981781/original.jpg
2. Seth Wollney, Atlantic Ocean. CC-BY. https://inaturalist-open-data.s3.amazonaws.com/photos/314857468/original.jpg
3. Amanda Chase, Texas, USA. CC-BY. https://inaturalist-open-data.s3.amazonaws.com/photos/4323195/original.jpeg
4. Lucian Vazquez, Matanzas, Cuba. CC-BY. https://inaturalist-open-data.s3.amazonaws.com/photos/249736410/original.jpg
5. Tim (user `twan3253`), New South Wales, Australia. CC-BY. https://inaturalist-open-data.s3.amazonaws.com/photos/116503028/original.jpeg
6. William Stephens. San Cristobal, Ecuador. CC-BY. https://inaturalist-open-data.s3.amazonaws.com/photos/64365445/original.jpg
7. Andrew Gillespie, Eastern Cape, South Africa. CC-BY. https://inaturalist-open-data.s3.amazonaws.com/photos/349817115/original.jpeg
8. Manuel R. Popp, Eastern Cape, South Africa. CC-BY. https://inaturalist-open-data.s3.amazonaws.com/photos/180673664/original.jpeg
9. Hazel Valerie, Northland, New Zealand. CC-BY. https://inaturalist-open-data.s3.amazonaws.com/photos/145998383/original.jpeg
10. Jacqui Geux, Auckland, New Zealand. CC-BY. https://inaturalist-open-data.s3.amazonaws.com/photos/3560793/original.JPG
11. Emily Roberts, Taranaki, New Zealand. CC-BY. https://inaturalist-open-data.s3.amazonaws.com/photos/205384503/original.jpeg
12. Arnim LIttek, Manawatu-Wanganui, New Zealand. CC-BY. https://inaturalist-open-data.s3.amazonaws.com/photos/227840858/original.jpeg
13. Arnim LIttek, Manawatu-Wanganui, New Zealand. CC-BY. https://inaturalist-open-data.s3.amazonaws.com/photos/72081497/original.jpeg
14. Jacqui Geux, Auckland, New Zealand. CC-BY. https://inaturalist-open-data.s3.amazonaws.com/photos/65125848/original.jpeg
15. Roderic Page, Auckland, New Zealand. CC-BY. https://inaturalist-open-data.s3.amazonaws.com/photos/340905787/original.jpg
16. Arnim LIttek, Manawatu-Wanganui, New Zealand. CC-BY. https://inaturalist-open-data.s3.amazonaws.com/photos/36406644/original.jpeg

Funding

1. SHC was funded by the National Science Foundations (NSF) Fellowship 2109502 and the Yale Institute of Biospheric Sciences (YIBS) Grant PJ111850.
2. NA was funded by the YIBS grant PJ021802 and the Tal Waterman gift GE100374.
3. DRB was funded by NSF grant OIA-1946352.
4. TJC was funded by the University of Chicago Women's Board Postdoctoral Fellowship.
5. JMG, BIMMM, and CJM were funded by the Açores FEDER (85%) and regional funds

(15%) under Programa Operacional Açores 2020 project Águas-VivAz ACORES-01-0145-FEDER-000119.
6. BIMMM was additionally funded by the Foundation for Science and Technology Portugal PhD Grant SFRH/BD/129917/2017.
7. SHDH was funded by the David and Lucile Packard Foundation.
8. DDGLG and RN were funded by Nelson Mandela University Council Funding.
9. KAP was funding the by Australian Research Council (ARC) Linkage LP210200689.
10. KO was funded by a Grant-in-Aid for Research Activity Start-up 22K20662 from the Ministry of Education, Culture, Sports, Science and Technology of Japan, and a grant from the Research Institute of Marine Invertebrates.
11. AS was funded by ARC linkage LP210200689.
12. JS was funded by ANID ATE 220044.
13. The Yale Center for Genome Analysis and Keck Microarray Shared Resource are in part funded by the National Institutes of Health instrument grant 1S10OD028669-01 and HPC grant (1S10OD030363-01A1).

Competing interests

The authors declare no competing interests

Data and materials availability

Sequence data for genome assembly (PacBio reads and 10X Genomics Chromium linked-reads) are available at the National Center for Biotechnology Information (NCBI) Bioproject PRJNA735958, and the principal and haplotype assembly sequences are available at NCBI Bioproject PRJNA1040906. Full-length, non-chimeric PacBio Iso-Seq RNA data are available at BioProject PRJNA1126252. Illumina sequence data for specimens intended for population genomic analysis are available at Bioproject PRJNA1092115. Code used to analyze data and reproduce the figures shown here is available at https://github.com/shchurch/Physalia_population_genomics, commit `648e532`. Specimens collected for this study are deposited at the Yale Peabody Museum, with the exception of specimens loaned by the Western Australian Museum, Tasmanian Museum and Art Gallery, and the Field Museum of Natural History. Specimen accession numbers listed in supplementary text: specimen collection information.

References

1. S. Van der Spoel, The basis for boundaries in pelagic biogeography. *Progress in Oceanography* **34**, 121–133 (1994).

2. M. V. Angel, Biodiversity of the pelagic ocean. *Conservation Biology* **7**, 760–772 (1993).

3. R. D. Norris, Pelagic species diversity, biogeography, and evolution. *Paleobiology* **26**, 236–258 (2000).

4. K. T. Peijnenburg, E. Goetze, High evolutionary potential of marine zooplankton. *Ecology and Evolution* **3**, 2765–2781 (2013).

5. B. W. Bowen, M. R. Gaither, J. D. DiBattista, M. Iacchei, K. R. Andrews, W. S. Grant, R. J. Toonen, J. C. Briggs, Comparative phylogeography of the ocean planet. *Proceedings of the National Academy of Sciences* **113**, 7962–7969 (2016).

6. S. B. Johnson, J. R. Winnikoff, D. T. Schultz, L. M. Christianson, W. L. Patry, C. E. Mills, S. H. Haddock, Speciation of pelagic zooplankton: Invisible boundaries can drive isolation of oceanic ctenophores. *Frontiers in Genetics* **13**, 970314 (2022).

7. S. H. Haddock, A golden age of gelata: Past and future research on planktonic ctenophores and cnidarians. *Hydrobiologia* **530**, 549–556 (2004).

8. J. B. Pfaller, A. C. Payton, K. A. Bjorndal, A. B. Bolten, S. F. McDaniel, Hitchhiking the high seas: Global genomics of rafting crabs. *Ecology and Evolution* **9**, 957–974 (2019).

9. R. R. Helm, The mysterious ecosystem at the ocean's surface. *PLoS Biology* **19**, e3001046 (2021).

10. F. Chong, M. Spencer, N. Maximenko, J. Hafner, A. C. McWhirter, R. R. Helm, High concentrations of floating neustonic life in the plastic-rich north pacific garbage patch. *Plos Biology* **21**, e3001646 (2023).

11. Z. Long, Begin ocean garbage cleanup immediately. *Science* **381**, 612–613 (2023).

12. I. Tessnow-von Wysocki, M. Wang, C. Morales-Caselles, L. C. Woodall, K. Syberg, B. Carney Almroth, M. Fernandez, L. Monclús, S. P. Wilson, M. Warren, others, Plastics treaty text must center ecosystems. *Science* **382**, 525–526 (2023).

13. A. K. Totton, Studies on *Physalia physalis* (L.). *Natural history and morphology. Discovery Reports* **30**, 301–368 (1960).

14. J. Bardi, A. C. Marques, Taxonomic redescription of the Portuguese man-of-war, *Physalia physalis* (Cnidaria, Hydrozoa, Siphonophorae, Cystonectae) from Brazil. *Iheringia. Série Zoologia* **97**, 425–433 (2007).

15. P. Pugh, A history of the sub-order Cystonectae (Hydrozoa: Siphonophorae). *Zootaxa* **4669**, 1–91 (2019).

16. D. Pontin, R. Cruickshank, Molecular phylogenetics of the genus *Physalia* (Cnidaria: Siphonophora) in New Zealand coastal waters reveals cryptic diversity. *Hydrobiologia* **686**, 91–105 (2012).

17. M. J. Costello, P. Tsai, P. S. Wong, A. K. L. Cheung, Z. Basher, C. Chaudhary, Marine biogeographic realms and species endemicity. *Nature Communications* **8**, 1057 (2017).

18. C. Munro, Z. Vue, R. R. Behringer, C. W. Dunn, Morphology and development of the Portuguese man of war, *Physalia physalis*. *Scientific Reports* **9**, 15522 (2019).

19. Y. K. Okada, Développement post-embryonnaire de la Physalie Pacifique. *Memoirs of the College of Science, Kyoto Imperial University. Series B* **8**, 1–26 (1932).

20. J. E. Purcell, Predation on fish larvae by *Physalia physalis*, the Portuguese man of war. *Marine Ecology Progress Series. Oldendorf* **19**, 189–191 (1984).

21. A. Damian-Serrano, S. H. Haddock, C. W. Dunn, The evolution of siphonophore tentilla for specialized prey capture in the open ocean. *Proceedings of the National Academy of Sciences* **118**, e2005063118 (2021).

22. L. Prieto, D. Macías, A. Peliz, J. Ruiz, Portuguese Man-of-War (*Physalia physalis*) in the Mediterranean: A permanent invasion or a casual appearance? *Scientific Reports* **5**, 11545 (2015).

23. D. R. Pontin, M. J. Watts, S. P. Worner, "Using multi-layer perceptrons to predict the presence of jellyfish of the genus *Physalia* at New Zealand beaches" in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)* (IEEE, 2008), pp. 1170–1175.

24. A. Canepa, J. E. Purcell, P. Córdova, M. Fernández, S. Palma, Massive strandings of pleustonic Portuguese Man-of-War (*Physalia physalis*) related to ENSO events along the southeastern Pacific Ocean. *Latin American Journal of Aquatic Research* **48**, 806–817 (2020).

25. N. Bourg, A. Schaeffer, P. Cetina-Heredia, J. C. Lawes, D. Lee, Driving the blue fleet: Temporal variability and drivers behind bluebottle (*Physalia physalis*) beachings off Sydney, Australia. *Plos One* **17**, e0265593 (2022).

26. N. Ahuja, X. Cao, D. T. Schultz, N. Picciani, A. Lord, S. Shao, K. Jia, D. R. Burdick, S. H. Haddock, Y. Li, others, Giants among cnidaria: Large nuclear genomes and rearranged mitochondrial genomes in siphonophores. *Genome Biology and Evolution* **16**, evae048 (2024).

27. K. Adachi, H. Miyake, T. Kuramochi, K. Mizusawa, S. Okumura, Genome size distribution in phylum cnidaria. *Fisheries Science* **83**, 107–112 (2017).

28. A. Collins, Phylogeny of medusozoa and the evolution of cnidarian life cycles. *Journal of Evolutionary Biology* **15**, 418–432 (2002).

29. A. G. Collins, S. Winkelmann, H. Hadrys, B. Schierwater, Phylogeny of Capitata and Corynidae (Cnidaria, Hydrozoa) in light of mitochondrial 16S rDNA data. *Zoologica Scripta* **34**, 91–99 (2005).

30. C. W. Dunn, P. R. Pugh, S. H. Haddock, Molecular phylogenetics of the Siphonophora (Cnidaria), with implications for the evolution of functional specialization. *Systematic Biology* **54**, 916–935 (2005).

31. B. D. Ortman, A. Bucklin, F. Pages, M. Youngbluth, DNA barcoding the medusozoa using mtCOI. *Deep Sea Research Part II: Topical Studies in Oceanography* **57**, 2148–2156 (2010).

32. Linné Carl von, Gmelin Johann Friedrich, Bernuset Pierre, Delamollière Jean Baptiste, *Caroli a Linné, Systema Naturæ Per Regna Tria Naturæ, Secundum Classes, Ordines, Genera, Species; Cum Characteribus, Differentiis, Synonymis, Locis* (1788; https://www.biodiversitylibrary.org/item/218817)vols. I, Part VI.

33. La Martinière, Sur quelques insects. *Observations et Mémoires sur la Physique, sur l'Histoire Naturelle et sur les Arts et Métiers* **tome 31**, Pl. II (1787).

34. Péron François, Freycinet Louis Claude Desaulses de, Lesueur Charles Alexandre, Petit Nicolas-Martin, Péron François, Baudin Nicolas, *Voyage de Découvertes Aux Terres Australes* (Paris, De l'Imprimerie impériale, 1807; https://www.biodiversitylibrary.org/item/95867)vols. Historique, t. 1.

35. C. Chun, Die Siphonophoren der Plankton-Expedition. *Ergebnisse der in dem Atlantischen Ocean von Mitte Juli bis Anfang November* **Bd. 2.K.b**, 1–126 (1889).

36. K. C. Schneider, Mittheilungen über Siphonophoren. III. Systematische und andere Bemerkungen. *Zoologischer Anzeiger* **Bd. 21**, 51–53, 73–93, 114–133, 153–173, 185–200 (1898).

37. National Oceanographic and Atmospheric Administration. National Weather Service, United States Army. arcGIS user: dhwilso1_NASA., Major ocean currents map. https://www.arcgis.com/home/item.html?id=5f4a73d0940f48859ee1fa9d6ff4a00f.

38. G. E. Hutchinson, The paradox of the plankton. *The American Naturalist* **95**, 137–145 (1961).

39. M. J. Behrenfeld, R. O'Malley, E. Boss, L. Karp-Boss, C. Mundt, Phytoplankton biodiversity and the inverted paradox. *ISME Communications* **1**, 52 (2021).