

RESEARCH

Open Access



Genomic signatures of speciation and adaptation in the ctenophore *Mnemiopsis*

Remi N. Ketchum^{1,2*}, Edward G. Smith³, Leandra M. Toledo¹, Whitney B. Leach⁴, Natalia E. Padillo-Anthemides¹, Andreas D. Baxeavanis⁵, Adam M. Reitzel⁶ and Joseph F. Ryan^{1,7*}

Abstract

Background Understanding how populations diverge is one of the most compelling questions in evolutionary biology but our grasp on the genomic mechanisms underpinning divergence is limited to a handful of species. Indeed, we know even less about divergence in the pelagic zone, where barriers to gene flow are seemingly absent. The holopelagic ctenophore *Mnemiopsis leidyi* is the most widely used ctenophore in experimental biology and has become an important model system in studies ranging from developmental biology to neurobiology. In addition, its relatively small and tractable genome provides a powerful foundation for genomic and evolutionary analyses. However, we still lack a clear understanding of species boundaries, population structure, and the evolutionary forces shaping divergence within *Mnemiopsis*, limiting both evolutionary and ecological interpretations. To expand our general understanding of divergence across novel environments as well as resolve a long-standing taxonomic debate, we generated the most comprehensive genomic study to date of the holopelagic ctenophore *Mnemiopsis* across a large expanse of its native range.

Results By leveraging multiple analytical approaches and generating two near-chromosome level genomes, we identify two distinct species of *Mnemiopsis* with high levels of genome-wide divergence along the US Atlantic coast, which correspond to *M. leidyi* and *M. gardeni*. Our demographic analyses suggest that *M. leidyi* and *M. gardeni* began to diverge during the mid-to-late Pleistocene climate transitions and were later shaped by post-glacial oceanographic changes. We highlight substantial genomic rearrangements and copy number variation between species, as well as uncover key genes under selection that are likely important for environmental adaptation.

Conclusions Together, these findings provide compelling evidence that the ctenophore currently recognized as *M. leidyi* represents more than one species. Recognizing cryptic species boundaries is critical for future study designs, environmental monitoring, and developing targeted management strategies. Altogether, we connect microevolutionary processes with macroevolutionary patterns and provide new insights into how ocean dynamics drive speciation and adaptation in pelagic ecosystems.

Keywords *Mnemiopsis*, Holopelagic invertebrate, Structural variation, Speciation, Population genomics

*Correspondence:

Remi N. Ketchum
rketchum@ad.unc.edu

Joseph F. Ryan
joseph.ryan@whitney.ufl.edu

Full list of author information is available at the end of the article



© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Background

Speciation, the process through which populations diverge into groups that no longer exchange alleles [1], is a fundamental evolutionary process responsible for generating much of the biodiversity on earth. Understanding how populations diverge and eventually become reproductively isolated has been a central question in evolutionary biology for decades [2–4]. Two primary models have traditionally been invoked to explain this process. In the first, allopatric speciation, geographic barriers restrict gene flow, thereby allowing populations to diverge independently. In the second, which includes parapatric and sympatric speciation, divergence occurs despite partial or complete geographic overlap [5, 6]. Although allopatric speciation is widely regarded as the predominant mode of speciation [3, 5, 7], its prevalence in marine environments is questionable given the scarcity of physical barriers and the high dispersal potential of many organisms that can lead to patterns of extensive connectivity [8–11]. Indeed, understanding how species arise and become established is particularly challenging in marine environments, although many examples of coexisting marine sibling species suggest that speciation may often be sympatric in the ocean. This process could be driven either by rapid adaptation to different ecological niches, genomic features that limit recombination and restrict introgression at specific loci (termed barrier loci; [12–15]), and/or hidden barriers to gene flow. While the development and spread of incompatible genotypes is central to both sympatric and allopatric speciation, the exact genomic regions and architecture that mediates this incompatibility, and the chronology of these genetic changes are still widely debated [11, 16].

The genomic patterns associated with speciation depend largely on whether divergence happens in the presence or absence of gene flow either during the initial split or following secondary contact [12, 17, 18]. In allopatric speciation scenarios, divergence can proceed unfettered by the homogenizing effects of migration [18, 19]. Meanwhile, in speciation-with-gene-flow scenarios, divergence is constrained by the balance between the degree of genetic linkage and recombination versus the strength of selection or drift [17]. In this scenario, regions of low recombination (e.g., centromeric regions or structurally complex loci) can facilitate divergence by maintaining linkage among locally adapted alleles despite ongoing gene flow [20, 21]. In marine species where physical barriers to dispersal can be weak, genomic features which reduce recombination are likely strong drivers of genomic divergence. However, most studies to date have focused on single nucleotide polymorphisms (SNPs) and, while informative, this has left the contribution of structural variants to speciation comparatively unexplored

(although see [6, 20–23]). Structural variants (SVs — insertions, deletions, and other genomic changes affecting > 50 bp) can occupy more of the genome than SNPs, have the potential to dramatically change phenotypes, and are often more likely to impact organismal fitness than SNPs because they are big enough to contain entire genes and other functional elements [21, 24]. SVs have been shown to lead to reproductive isolation through several mechanisms [25], including underdominance of heterokaryotypes [26], suppression of recombination [27, 28], gene duplications causing intrinsic postzygotic isolation [29], and SVs acting as mutations of large effect [21]. Despite their potential to drive speciation, structural variants have remained understudied due to challenges in detecting and genotyping them with short-read sequencing [30]. Recent methodological advances, however, now enable the generation of chromosomal- or near-chromosomal level genome assemblies, providing new opportunities to comprehensively investigate SVs and their evolutionary significance to speciation [12].

In addition to methodological biases, the vast majority of speciation literature is based on terrestrial species with small population sizes, clear physical barriers to gene flow, and internal fertilization, all traits which are expected to facilitate the emergence of reproductive isolation [31–33]. To fully understand the mechanisms driving speciation, we must broaden the range of studied organisms to include animals with diverse life history strategies, particularly pelagic invertebrates that spend their entire life in the water column and are characterized by high fecundity [34], large population sizes [35], and high dispersal potential. These traits are expected to produce weak genetic differentiation and promote panmixia [36] and indeed, this pattern is observed in many marine species [37–40]. Holopelagic species, such as *Mnemiopsis* (A. Agassiz, 1865), remain particularly understudied in this regard.

Mnemiopsis is a genus within the phylum Ctenophora with a native range that extends continuously along more than 25,000 km of the Atlantic coast of the Americas, from Kittery, Maine to the Valdes Peninsula in Argentina [41, 42]. These comb jellies are considered to be one of the 100 most invasive species in the world [43] and their introductions into European waters have triggered community trophic cascades and collapse of fisheries [44]. The extensive geographic distribution and invasive abilities of *Mnemiopsis* are driven in part by its broad environmental tolerance [45], flexible planktivorous diet [46], extensive regeneration abilities [47], high fertility levels [48], and a capacity for self-fertilization [49]. Risk management for invasive species requires genetic information to establish management units and accurate taxonomic classification is a fundamental component of effective

control measures. Further, many invasive species have the capacity for rapid adaptation [50] which, in turn, may make them more likely to speciate following a biological invasion although relatively few studies have examined this phenomenon [51].

Despite its rich 100-year history as an evo-devo model [52] and the availability of extensive genomics resources [53–56], our understanding of species boundaries in *Mnemiopsis* is subject to debate [57–61]. The World Register of Marine Species (WoRMS) currently divides *Mnemiopsis* into two accepted species: *Mnemiopsis leidyi* (A. Agassiz, 1865; type location Massachusetts) and *Mnemiopsis gardeni* (L. Agassiz, 1860; type location South Carolina). A third species, *Mnemiopsis mccradyi* (Mayer, 1900; type location South Carolina), is currently recognized as a synonym for *Mnemiopsis leidyi* [62]. However, there is much confusion as to the ranges and differences in morphology between these two named species, which is further complicated by anthropogenic range expansion. Most modern literature considers all ctenophores along the Atlantic and Gulf coasts of the United States as a single species, specifically *Mnemiopsis leidyi* [63]. Earlier studies have raised the possibility of multiple *Mnemiopsis* species [57–61], but the molecular data available at the time were limited in either marker number or population sampling, making it difficult to resolve this question conclusively. With the compact genome of *Mnemiopsis* and the rapid democratization of next-generation sequencing (NGS) technologies, we can now examine population relationships at unprecedented resolution and address this long-standing controversy more comprehensively than was previously possible.

We conducted whole-genome short-read sequencing of 118 individuals from 13 populations within *Mnemiopsis*' native range along the United States Atlantic coast and generated two near-chromosomal length genome assemblies of a northern and southern representative of each genus. We applied a suite of genomic approaches to understand holopelagic population dynamics, demographic histories, changes in genomic architecture, and adaptation to new environments. Taken together, these findings highlight barriers to gene flow in holopelagic species, illuminate mechanisms that may underlie invasion success, and address a critical knowledge gap in marine speciation research, which has historically focused on a relatively narrow subset of species.

Results and Discussion

Whole-genome sequencing of *Mnemiopsis* populations

We collected and sequenced the genomes of 118 *Mnemiopsis* individuals from 13 populations from their native range along the US Atlantic coast (Fig. 1A, Additional file 1: Table S1). All samples were collected from < 10 m

surface waters and preserved in ethanol within an hour of collection. We generated a total of 730 Gb of Illumina reads and our samples were sequenced to an average genome depth of 18.89×. This sequence data was used to generate datasets for subsequent genomic analyses (see [Methods](#)).

We also produced two high-quality, near-chromosome level genome assemblies of one sample from Massachusetts (*M. leidyi*; northern lineage) and one sample from Florida (*M. gardeni*; southern lineage) using PacBio HiFi sequencing technology (Additional file 1: Table S2). When assembled, the *M. gardeni* genome was 206.6 Mb in length, with half of the genome assembled into contigs 3.6 Mb or larger (N50). The *M. gardeni* assembly was comprised of 174 contigs, with the 100 largest contigs making up 98% of the assembly. The *M. leidyi* genome was 215.8 Mb in length, with half of the genome assembled into contigs 2 Mb or larger. The *M. leidyi* assembly was comprised of 207 contigs, with the 100 largest contigs making up 89% of the assembly. We identified 27,570 and 26,046 protein-coding genes within the *M. gardeni* and *M. leidyi* genomes, respectively, and were able to functionally annotate 27,476 genes within the *M. gardeni* genome (Additional file 2: Table S3). Analyses using the Benchmarking Universal Single-Copy Orthologs program [64] revealed that our assemblies exhibit a similar degree of completeness to other ctenophore genomes ([53, 65], (Additional file 1: Table S2). We recovered 91% (83% complete and 8% partial) of the expected core genes for *M. leidyi* with 9% of them missing. For *M. gardeni*, we recovered 92% (84% complete and 8% partial) of the expected core genes, with 8% of them missing.

Population genomic evidence of two *Mnemiopsis* species

We performed phylogenetic analyses using nuclear and mitochondrial markers from our *Mnemiopsis* samples collected along the Atlantic Coast. Our nuclear phylogenetic tree (Fig. 1B) revealed the presence of two distinct clades that corresponded to the northern lineage (Woods Hole, MA to Roanoke Island, NC) and the southern lineage (Pivers Island, NC to Panacea, FL). We observed a similar two-clade pattern in our phylogenetic analysis of the complete mitochondrial genomes (published mitogenome is 10,326 bp [66]), driven largely by ~100 nearly fixed nucleotide differences (Fig. 1C). However, we identified six individuals from populations near the geographic interface of these two clades that exhibit mitonuclear discordance. Specifically, we observed one individual collected at Gloucester Point, Virginia (VAV10) and four individuals collected in Roanoke Island, North Carolina (NCX3–5,8) that had northern nuclear and southern mitochondrial genotypes. Likewise, we observed one individual collected from Wrightsville

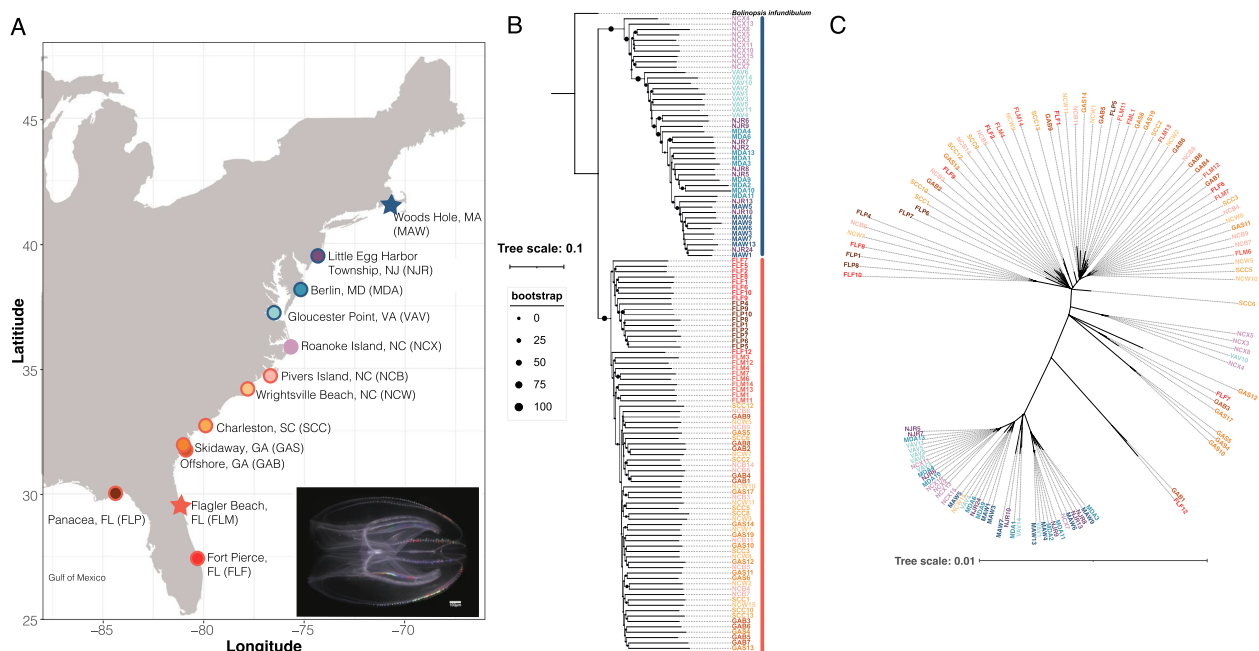


Fig. 1 Mitonuclear discordance in *Mnemiopsis*. **A** Sampling map showing all *Mnemiopsis* collection sites (site abbreviations in parentheses). The outer circle indicates the lineage assigned to each population sample (blue = northern; red = southern). Roanoke Island, NC is shown without an outer ring because it is later identified as a hybrid population. Stars denote populations also used for long-read PacBio HiFi genome sequencing and assembly. **B** Nuclear phylogeny inferred from 1,302 SNPs using ngsDist and FastME, rooted with *Bolinopsis infundibulum*. Circle size at each branch reflects bootstrap support; both the northern and southern lineages showed 100% support. Sample ID colors correspond to sampling locations (matching panel A), and the colored bars to the right of each ID denote lineage membership (blue = northern; red = southern). **C** Mitochondrial phylogeny generated from a whole-mitogenome alignment using IQ-TREE v2.3.2. Sample ID colors indicate population assignment

Beach, North Carolina (NCW15) that had a southern nuclear and northern mitochondrial genotype. This level of mitonuclear discordance could be consistent with hybridization in a secondary contact zone and early generation hybrids [61, 67].

To further investigate these admixture patterns, we evaluated the presence of 2–7 populations (by adjusting the K parameter) within our dataset using ngsADMIX [68]. Setting K=2 separates individuals into northern and southern lineages with the Roanoke Island, NC (NCX) and Gloucester Point, VA (VAV) samples showing signs of admixture (Fig. 2A). When K=3, the Roanoke Island samples form their own subpopulation and, at higher K values, more substructure within the two lineages becomes evident (Additional file 1: Fig. S1). MDS plots (Fig. 2B) support the results obtained using ngsADMIX, with the first principal component congruent with the patterns observed by setting K=2 and accounting for 39.4% of the variance in the data. The average F_{ST} between the northern and southern lineages was 0.278 when Roanoke Island, NC was included in the northern lineage, and 0.315 when it was removed from the comparison. Our pairwise F_{ST} comparisons also highlight two distinct lineages (Additional file 1: Table S4). Due to these

cumulative results, we also tested for evidence of hybridization in our data using f_3 statistics [69]. We compared the Roanoke Island, NC (NCX) population to the most northern population (Woods Hole, MA; MAW) and most southern population (Panacea, FL; FLP) and, notably, this resulted in a negative f_3 statistic and an absolute Z score ≥ 3 (threshold used for significance), highlighting a significant degree of admixture in the Roanoke Island, NC population (Additional file 1: Table S5). In line with a recent study which detected evidence of admixture in Virginia [61], we also observed signatures of admixture in Virginia (Gloucester Point, VA; VAV), as well as other northern populations (including in populations ~450 km away, namely NJR and MDA). In contrast, there is no admixture observed in any populations south of Roanoke Island, even though Pivers Island, NC (NCB) is comparatively close (~180 km south).

Our analyses reveal two divergent genetic lineages in US Atlantic *Mnemiopsis*. These results are consistent with previous studies that reported genetic differentiation between Massachusetts and Florida populations [57] and identified a population breakpoint near Cape Hatteras, NC [58]. Building on this earlier work, our high-resolution spatial and genomic sampling further

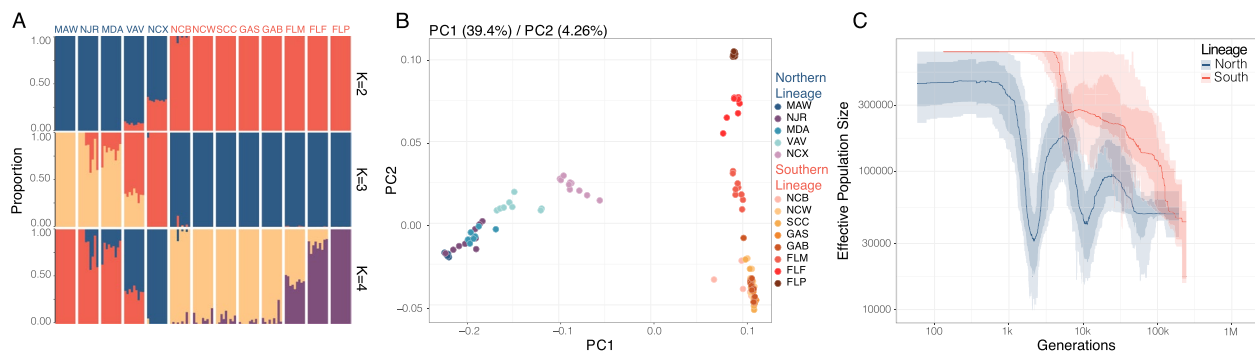


Fig. 2 *Mnemiopsis* exhibits strong population structuring along the US Atlantic Coast. **A** Admixture plots for the 118 *Mnemiopsis* samples for K=2 to K=4 where each bar represents a single individual, and the colors represent each of the K components. **B** MDS plot for PC1 and PC2 with the percent of total variance explained by each component shown above the plot in parentheses. Samples are colored according to their sampling location with *M. leidyi* samples (northern lineage) clustering on the left and *M. gardeni* samples (southern lineage) clustering on the right. **C** Stairway plot results show different demographic histories of the two lineages with a generation time of one year and a mutation rate of 6.85×10^{-8} per base pair per year

demonstrates that the genetic breakpoint between the two lineages is a narrow hybrid zone. There is cumulative evidence of admixture observed from Roanoke Island, NC up to Gloucester Point, VA (~170 km north) but little evidence of admixture observed at Pivers Island, NC, which is located only ~180 km to the south. Although f_3 -statistics suggest evidence of admixture in Little Egg Harbor Township, NJ (NJR), and Berlin, MD (MDA), this result contrasts with the other population genetic results and is therefore more likely explained by a counterintuitive property of f -statistic computations known as sister repulsion [70]. Sister repulsion can occur when there is unidirectional gene flow between study regions or when populations experience gene flow from an unsampled genetically distant source. Extending our approach to new geographic locations across the reported range of *Mnemiopsis* would make discerning between these two processes possible. However, we posit that this pattern is most likely explained by unidirectional gene flow given our demographic modeling results (see next section). Taken together, our findings indicate that the genetic breakpoint is located where the Gulf Stream leaves the coast and moves offshore into the open Atlantic Ocean, and we find evidence of a hybrid zone potentially resulting from recent secondary contact. We also see high levels of genome-wide divergence, which tends to be more typical of species that experienced geographic isolation during the early stages of speciation [71, 72]; however, demographic analyses are required to explore this further. With this in mind, we will hereafter refer to the northern lineage as *M. leidyi* and the southern lineage as *M. gardeni* (following the naming scheme employed in WoRMS).

Demographic history of *Mnemiopsis*

We employed a Stairway Plot analysis [73] to reconstruct the demographic histories of *Mnemiopsis*, assuming a generation time of 1 year (see “Methods”). The two species show strikingly different population size trajectories beginning roughly 10,000 years ago (Fig. 2C). Around this time frame, *M. gardeni* exhibits sustained population growth whereas *M. leidyi* displays a more complex history with two population bottlenecks, one around 10,000 years ago and a second between 2,000 and 3,000 years ago, followed by a recent recovery and expansion. The divergence between the effective population size curves for *M. gardeni* and *M. leidyi* that begins ~10,000 years ago coincides with major post-glacial events. Around this time, the Laurentide Glacier retreated (~8,000 years ago), and the sudden drainage of Lake Agassiz released a large influx of cold freshwater into the northern Atlantic [74]. This influx of cold freshwater likely reduced the current in the ocean conveyor belt and resulted in a decrease in the amount of warm seawater traveling northward, subsequently causing temperatures to drop in the Northern Hemisphere [74, 75]. The population expansion that followed may have been a result of a reduction in the Gulf Stream current following the Lake Agassiz flood which would have allowed *Mnemiopsis* to migrate northward. It is unclear what would have caused the population bottleneck ~2,000–3,000 years ago in *M. leidyi*. It is possible that it was caused by the Little Ice Age — a cooling period in the northern hemisphere, although this occurred roughly between 1300 and 1850 [76, 77] and therefore does not line up perfectly with our estimated bottleneck.

Our results differ from the Pairwise Sequentially Markovian Coalescent (PSMC) results reported in Jaspers et al.

2021 [57], which were based on a pairwise comparison of populations from Massachusetts and Florida. While both studies identify population expansions about ~50,000 generations ago and a population bottleneck ~10,000 generations ago, their *M. leidy* samples did not show oscillating population size changes. Additionally, their *M. gardeni* samples showed a steady increase in population size from ~500,000 to 50,000 generations ago and then a decrease in population size from then onwards. In contrast, ours has shown a steady increase in population size from 100,000 generations ago to present day. This difference in findings can likely be explained by the increased geospatial resolution in our study. In addition, while the PSMC method [78] performs well for inferring ancient histories, the Stairway Plot method is more accurate for inference of more recent population histories and we see the greatest discordance in our results when looking at more recent population histories [73, 79].

Our demographic analysis using *Moments* further refined our understanding [80]. The best-fit demographic model (sc3ielsm1, Additional file 1: Fig. S2-4) included secondary contact and three epochs in each species following their initial divergence. The split between *M. leidy* and *M. gardeni* occurred approximately 223 Kya, followed by ~107 Kya of low-to-moderate bidirectional gene flow. The second epoch began around 116 Kya and was characterized by genetic isolation lasting roughly 89 Kya. The most recent epoch started approximately 27 Kya, during which asymmetrical migration occurred predominantly from *M. gardeni* to *M. leidy*. Across these epochs, *M. leidy* experienced an initial population expansion, followed by a sustained decline, whereas *M. gardeni* displayed the opposite pattern, an initial contraction followed by consistent population growth.

Together, Stairway Plot and *Moments* analyses reveal broadly consistent demographic patterns. *Moments* infers a reasonably ancient divergence time, perhaps because incorporating migration into the model allows some population differences to be explained by a more ancient split rather than recent divergence. Both approaches recover evidence of pronounced population size reductions in *M. leidy*. The magnitude and number of contractions differ between methods, likely because *Moments* is limited to modeling a maximum of three population size changes, whereas Stairway Plot can recover more complex histories. These methods also concur that *M. gardeni* has undergone a recent expansion, though they differ on whether *M. leidy* has rebounded or continues to experience a population contraction. Overall, the demographic models suggest that divergence likely began in the mid-to-late Pleistocene, potentially caused by distant colonization events, geological change, sea level changes or changes in prevailing currents [81]. However, it remains

unclear whether genetic isolation was initiated by glacially-driven flooding and subsequent re-strengthening of the Gulf Stream, or whether divergence predates these events, with post-glacial oceanographic changes instead facilitating secondary contact and asymmetric migration from *M. gardeni* to *M. leidy*.

Structural genomic evidence of speciation

Chromosome-scale linkages of orthologous genes tend to be conserved even over very long time frames [82–86]. As such, chromosomal rearrangements are major evolutionary events that can either cause speciation events and/or provide evidence of long periods of isolation between lineages. To investigate structural differences between *M. leidy* and *M. gardeni* genomes, we produced two high-quality, near-chromosome level genome assemblies of one *M. leidy* sample from Massachusetts and one *M. gardeni* sample from Florida. We conducted macrosynteny analyses to check for evidence of chromosomal rearrangements between the two species' genome assemblies. In addition, we compared both genomes to the recently published chromosome-level genome assembly of *Bolinopsis microptera* [85], a lineage that is closely related to *Mnemiopsis* [87], in order to infer ancestral chromosomal states. We identified at least one chromosomal rearrangement between the *M. gardeni* and *M. leidy* genomes (Fig. 3A – C). The *M. leidy* scaffold L006 maps to two *M. gardeni* scaffolds: G002, which maps completely to chromosome one of *B. microptera*, and G004, which maps completely to chromosome eight of *B. microptera*. The fact that this linkage group is present in both *M. gardeni* and *B. microptera* suggests that the genomic rearrangement occurred in *M. leidy*. There are potentially other rearrangements involving the *M. leidy* scaffolds L012 and L018, but a chromosome-scale assembly is needed to confirm or refute those possibilities.

As copy number variants (CNVs) can drive genome evolution [88] and contribute to both speciation events [89] and environmental adaptation [90], we examined CNV patterns between the two species. Genome-wide patterns of CNV across samples largely mirror that of SNP variation, with separation of samples into clusters associated with *M. leidy*, samples south of the Outer Banks, and the Floridian samples (Additional file 1: Fig. S5). We assessed the 67 genes in the top 1% of fixation index values (V_{ST} ; Additional file 3: Table S6). Of these 67 genes, only 21 (31%) were assigned a gene ontology (GO) category [91] based on sequence similarity (Additional file 1: Table S7). We ran a Monte Carlo analysis and showed that this is significantly less ($P=0.0006$) than would be expected by chance. On average, a random set of 67 genes would produce 32 (48%) genes within an associated GO category. These data suggest that the CNV

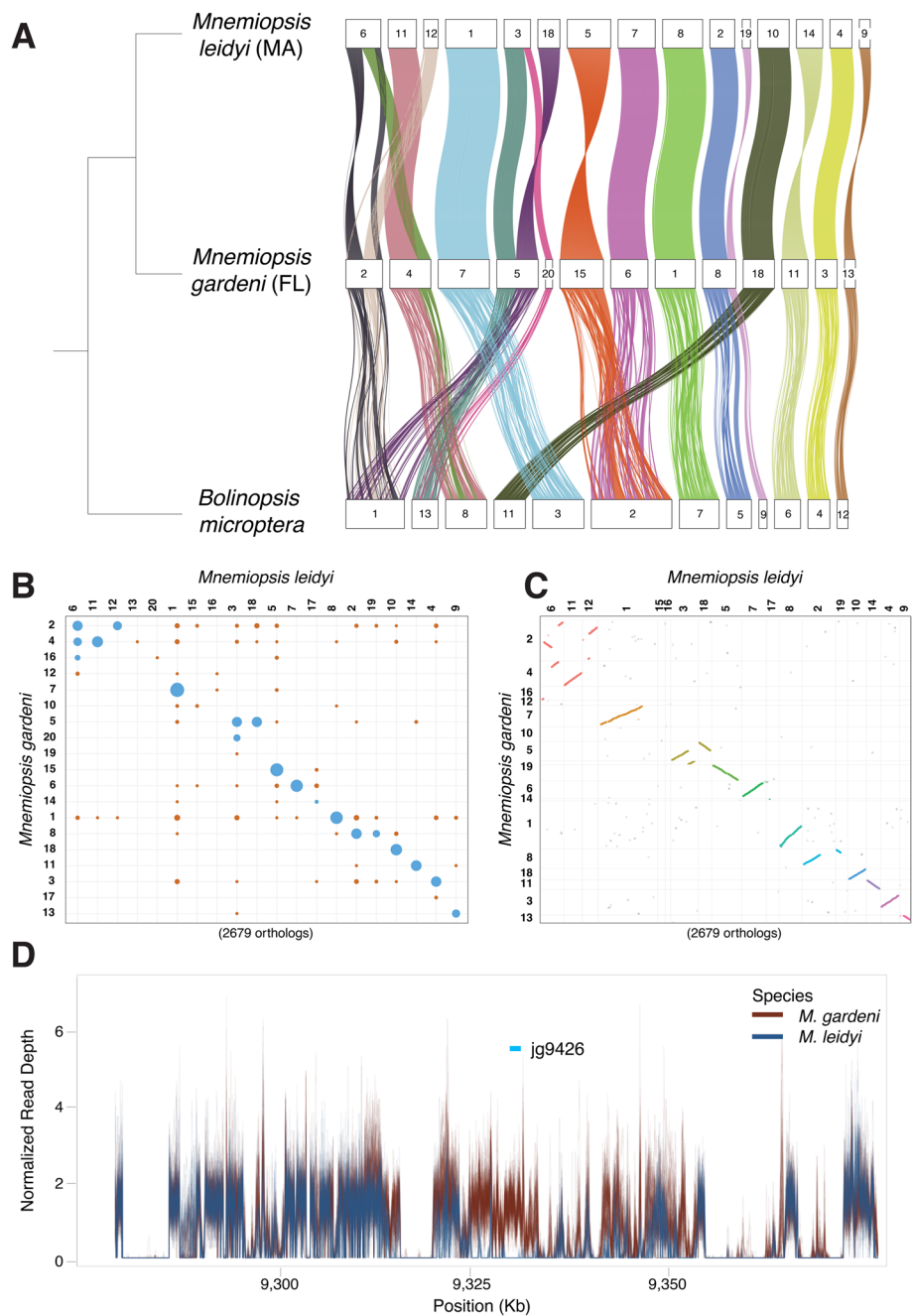


Fig. 3 Large-scale genomic rearrangements between species. **A** Chord diagram of macrosynteny on a near-chromosome scale between *Mnemiopsis leidyi* (top), *M. gardeni* (middle), and *Bolinopsis microptera* (bottom). The coordinates of 12,134 orthologs (based on reciprocal best BLAST hits) are plotted. **B** Plot of macrosynteny between *M. leidyi* and *M. gardeni*. **C** Oxford dot plot between *M. leidyi* and *M. gardeni*. Each dot represents an ortholog. **D** The genomic landscape neighboring the top copy number variation outlier gene based on fixation index values (V_{ST}). Normalized read depth plot for short-read sequences from *M. leidyi* (blue) and *M. gardeni* (red) for *M. gardeni* scaffold 001 from position 9,278,170–9,378,170

outliers are enriched in so-called orphan genes. This is not unexpected, as “orphan” genes are often associated with effector functions (e.g., [92]) rather than highly conserved functions that are often critical to survival.

The top V_{ST} CNV gene (jg9426 in the *M. gardeni* genome) is a putative G-protein-coupled receptor (GPCR), a class of cell surface receptors that detect molecules (e.g., light energy, lipids, proteins) and

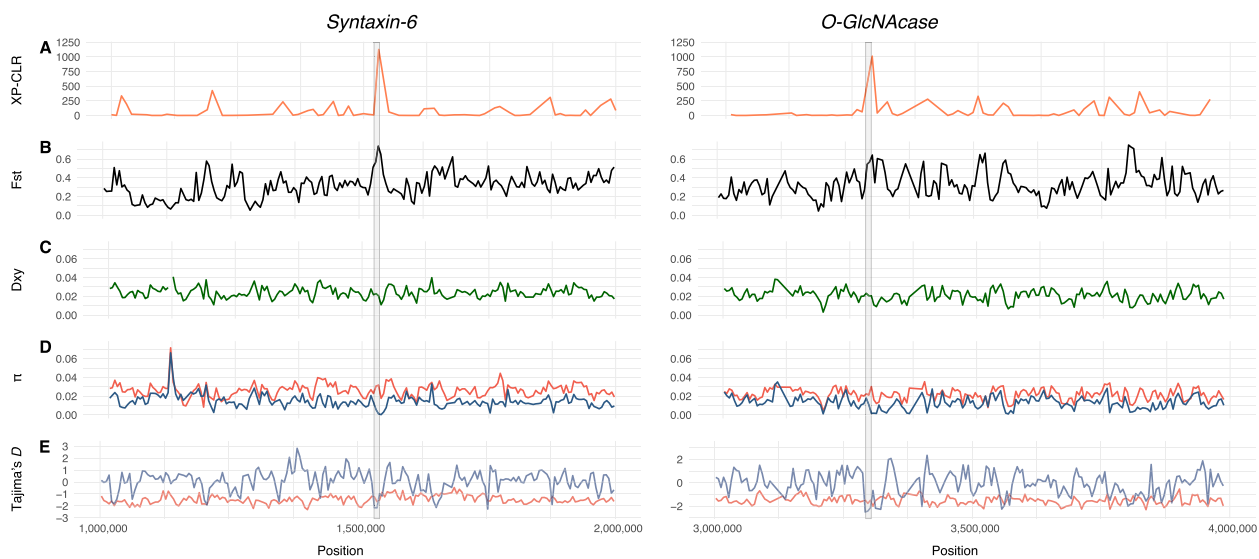


Fig. 4 Genes involved in autophagic flux are under strong positive selection between species. The two genes: *syntaxin-6* (left) and *O-GlcNAcase* (right) that are under putative positive selection. The **A** XP-CLR, **B** F_{ST} , **C** D_{XY} , **D** nucleotide diversity (π), and **E** Tajima's D values were calculated with nonoverlapping windows and step sizes. The locations of the two sweeping windows are shown with gray boxes. For panels **D** and **E**, blue lines indicate *Mnemiopsis leidyi* and red lines indicate *M. gardeni*

activate cellular responses. This GPCR gene copy is missing from *M. leidyi* and appears to be the result of either a species-specific duplication in *M. gardeni* or from a deletion of an ancestrally duplicated gene in *M. leidyi* (*M. gardeni* gene IDs: jg9426 and jg9432, *M. leidyi* gene ID: ML07987a; Fig. 3D, Additional file 1: Fig. S6). It is one of four predicted GPCRs in these 67 genes (Additional file 1: Table S8). These GPCRs are all part of GPCR clades that have independently radiated in ctenophores and have a many-to-many relationship with non-ctenophore GPCRs, making it difficult to infer function. Loss or duplications in these genes would likely lead to changes in sensory capabilities or cellular responses downstream of sensory inputs, changes that are widely hypothesized to facilitate ecological speciation (speciation caused by divergent selection; [93, 94]). These microevolutionary changes are consistent with the macroevolutionary pattern of loss and expansion observed in GPCRs, offering insight into mechanisms that underlie the independent radiation commonly seen in phylogenetic analyses of channel and receptor genes from multiple species. These observations are consistent with the “accordion model” of gene evolution, where rounds of extensive loss are followed by rounds of extensive duplication [95]. Overall, patterns in the CNV data mirror those seen at the single nucleotide level and, although gene loss in *M. leidyi* may be a result of severe bottlenecks or selective pressures, it may also stem from reference

bias, highlighting the need for pangenomic approaches [96, 97].

Genomic signatures of selection in genes mediating environmental sensing

To investigate signatures of selection and local adaptation, we used the cross-population composite-likelihood ratio (XP-CLR) test [98]. We observed key sensory related and neural genes found in regions within the top 10% of XP-CLR values (Additional file 4: Table S9). These included 79 GPCRs, four TRP channels, five glutamate receptors, the single *Mnemiopsis* voltage-gated calcium channel, one of the two *Mnemiopsis* voltage-gated sodium channels, 10 voltage-gated potassium channels, three acid-sensing ion channels (ASIC), two calcium-activated potassium channels, a voltage-gated chloride channel, and a DEG/ENaC channel (Additional file 1: Tables S10–11). Our outlier dataset also contains many other genes that are relevant for nutrient sensing and metabolism; these include the mechanistic target of rapamycin (mTOR), acetyl coenzyme A synthase, malate dehydrogenase, and cytochrome c oxidase subunit 2 genes.

The top XP-CLR outlier region contains the *syntaxin-6* gene (jg7623), which promotes the movement of transport vesicles to target membranes. We also observe the *O-GlcNAcase* (jg8681) gene in the outlier dataset, which codes for a post-translational modifier of serine and threonine residues on a wide range of proteins [99]. We highlight these two genes in Fig. 4 as they have both been

implicated in modulating autophagic flux in response to nutrient status in *Caenorhabditis elegans* [100, 101]. Both sweep regions show multiple signatures of selection, including reduced nucleotide diversity in *M. leidy*, elevated F_{ST} values, and an excess of rare alleles (as evidenced by the more negative Tajima's D values for *M. leidy*). We hypothesize that overwintering in northern populations may be a driver of these selective sweeps, as organisms are likely nutrient-deficient during this time. Interestingly, *M. gardeni* populations have significantly lower Tajima's D values than *M. leidy* populations across the entire genome ($W=1.16 \times 10^9$; $P<0.01$). However, this pattern breaks down at the target of selective sweeps. This may be consistent with overall genomic patterns highlighting a population expansion in *M. gardeni* and population contractions in *M. leidy* that is disrupted at the site of selective sweeps.

To highlight functional categories of these genes, we conducted gene ontology (GO) analysis [91] of the 2,093 genes that overlap with the genomic windows ranked in the top 10% of XP-CLR values (Additional file 1: Table S12). These genes were significantly enriched ($P \leq 0.05$) for GO molecular function categories, including those associated with sensory processing, neural activity, or diet, such as GPCR signaling pathways ($P=0.0136$), GPCR activity ($P=0.0152$), Golgi vesicle transport ($P=0.0399$), neurotransmitter transport ($P=0.0424$), and chitin metabolic processes ($P=0.0056$). The chitin metabolic processes category includes two chitinase genes, which are critical for digesting prey. These findings suggest that recent selection has acted on genes involved in environmental sensing and resilience to nutrient deficiency and provide important candidate genes for future functional verification.

Conclusions

Both the population genomic and structural genomic evidence are consistent with there being two species of *Mnemiopsis* on the US Atlantic coast. We report high levels of genome-wide divergence, as well as mitonuclear discordance. Demographic modeling suggests that these species diverged in the mid-to-late Pleistocene, with long-term genetic isolation followed by secondary contact and asymmetric migration from *M. gardeni* into *M. leidy*, a scenario supported by the hybrid zone we identify along the North Carolina coast. We highlight large-scale genomic rearrangements between species that are consistent with a history of population contractions or small founder populations. We find evidence of selection on sensory and nervous system genes, with similar trends being reported in other species following speciation events [87, 102–105]. Further research is needed to investigate morphological, reproductive, and life-history

traits that may help provide further evidence for multiple species, as well as to examine the possibility that more *Mnemiopsis* species exist in Central and South America, as the mitochondrial results of Bayha et al. suggest [58]. More broadly, our knowledge of evolution is heavily biased across the tree of life, and this study reinforces the importance of expanding taxon sampling as a means towards gaining a much richer understanding of how evolutionary forces shape the genomic variation between populations and species.

Methods

Sampling

One individual was collected from each of two locations for genome assembly: Woods Hole, MA and Flagler Beach, FL. These samples were collected in August 2023 from surface waters (<2 m) using a cteno-dipper [48]. Samples were preserved in ethanol and placed in a -20 °C freezer until further processing.

A total of 118 *Mnemiopsis* samples were collected from thirteen locations along the Atlantic and Gulf coast of the United States. These locations included: Woods Hole, MA, Little Egg Harbor Township, NJ, Berlin, MD, Gloucester Point, VA, Roanoke Island, NC, Pivers Island, NC, Wrightsville Beach, NC, Charleston, SC, Skidaway Island, GA, Skidaway Island offshore, GA, Flagler Beach, FL, Fort Pierce, FL, and Panacea, FL (Fig. 1). Samples were collected between August 2021 and November 2022 from surface waters (<10 m) using either a cteno-dipper or a tow net. Samples were immediately preserved in ethanol and placed in a -20 °C freezer until further processing.

DNA extraction and sequencing

High molecular weight DNA was extracted according to the CTAB/chloroform protocol detailed in Erlich et al. 2017 [106]. DNA extractions were visualized on a 1% agarose gel for quality control and quantified using a Qubit dsDNA High Sensitivity Assay Kit on a Qubit 2.0 Fluorometer. For the genome assemblies, sequencing was performed on a Pacific Biosciences Sequel IIe platform in the Circular Consensus sequencing (CCS) mode. A total of 17.7–20.2 Gb of HiFi reads were generated using CCS for the two individuals. PacBio library preparation and sequencing were performed at NIH Intramural Sequence Center (NISC) in Rockville, MD.

For short-read whole-genome sequencing, all DNA extractions were visualized on a 1% agarose gel for quality control. One hundred and eighteen individually indexed, paired-end libraries with an approximate insert size of 550 bp were constructed using the NEBNext Ultra II FS DNA Library Preparation kit. Whole-genome

sequencing was performed at the University of Florida's ICBR Bioinformatics Core Facility on a single flowcell lane of a NovaSeq6000.

De novo genome assembly

Hifiasm v0.19.8 [107] was used to assemble the two genomes. For the Flagler Beach, FL individual, the `--hom-cov` flag was set to 90. To reduce the amount of duplication, PurgeHaplotigs v1.1.2 [108] was run with the “contigcov” option for low coverage, with the low point between the peaks, and high coverage cutoffs set to 10, 70, and 175, respectively. The percent cutoff for identifying a contig as a haplotig (`-a` flag) was set to 50% and this assembly was subsequently run through the PurgeHaplotigs “clip” option that identifies and trims overlapping contig ends. For the Woods Hole, MA individual, the `--hom-cov` flag was set to 101, the “contigcov” options were set to 25, 70, and 155, and the `-j` and `-s` flags were both set to 55. The percent cutoff `-a` flag was set to 60% and the final assembly was also clipped to remove overlapping contig ends.

Gene prediction and annotation

Genomic repetitive elements were identified with RepeatModeler v1.0.11 [109] to generate a *Mnemiopsis*-specific repeat element library. Repetitive regions were soft masked prior to gene prediction and annotation using RepeatMasker v4.0.8. We aligned previously published bulk RNA-Seq reads from tissues (PRJEB28334; [110]) to the above genome assembly using Bowtie2 [111]. Aligned RNA-seq data and gene models for *M. leidyi* (ML2.2.aa) were input into Braker v2.1.5 [112]. Assembled gene/protein models were functionally annotated using BLASTP v2.9.0+ with three protein databases: UniProt Knowledgebase Swiss-Prot protein models v2021-03, RefSeq invertebrate protein models, and the *M. leidyi* protein models [53]. Finally, BUSCO v5 [64] was used to measure the completeness of the genome assembly and protein sequences using the Eukaryota database.

Macrosynteny analysis

We analyzed conserved chromosome-scale synteny (macrosynteny) between these new Florida and Massachusetts *Mnemiopsis* assemblies, as well as the published *Bolinopsis microptera* genome assembly. To identify conserved genes, we performed diamond BLAST [113] to compute the reciprocal best BLAST hits between the 2013 ML2.2 *Mnemiopsis* gene models and the gene models of each of the new *Mnemiopsis* assemblies and the published *B. microptera* gene models. These reciprocal best BLAST hits were used to identify significantly conserved macrosyntentic blocks in MacrosyntR (version 0.3.3; [114]).

Copy number analysis

Copy number variation between the two species was assessed using the GATK germline CNV (gCNV) pipeline in cohort mode [115]. The pipeline was run using the exome approach to focus specifically on genic regions, and gene intervals were padded by 250 bp. Intervals with fewer than 50 reads in greater than 50% of individuals were excluded, and samples from Roanoke Island in North Carolina were excluded due to their admixture. Denoised copy number ratios were processed in R to remove copy number invariant loci, perform principal component analysis, and to calculate the V_{ST} statistic. Loci within the top 1% of V_{ST} values were considered outliers. Microsynteny plots for CNV loci were generated by performing contig alignments in MUMmer [116]. SYRI was subsequently used to identify synteny and structural rearrangements, with plots generated using Plotsr [117, 118], and read depth was calculated using Samtools v.1.9 [119].

Data processing and filtering of short-read data

Raw sequence data was trimmed using Trimmomatic v0.39 [120] with a minimum phred score of 33. Once the data was filtered and adaptors were removed, the paired sequences were aligned to the Flagler Beach, FL genome using BWA-mem v0.7.12 [78]. Sequence alignments were filtered using Samtools for a minimum quality score of 30, Sambamba v0.6.6 [121] was used to filter duplicate reads with default parameters, and bamutil v1.0.15 [122] was used to clip overlapping read pairs.

We generated three datasets for our different analyses: LD-Filtered (89,750 SNPs), Lineage-Specific Filtered (3.8 million SNPs), and the All-Sites Filtered (81 million positions). For each dataset, we used the Flagler Beach, FL genome as our reference and ancestral genome as it was the most contiguous assembly.

LD-filtered dataset

This dataset was produced by using ANGSD v0.921 [123] to call SNPs and estimate genotype likelihoods using the Samtools model (`-GL 1`). We filtered the SNP data for (1) positions that were present in at least 94 out of 118 samples, (2) a maximum coverage of $2,360 \times$ per site across all individuals to avoid calling sites in highly repetitive regions, (3) a minimum of $354 \times$ coverage per site across all individuals, (4) a minimum base quality score > 20 , (5) a P -value cut off of 10^{-6} for calling polymorphic loci and (6) retained only SNPs with a global minor allele frequency $> 5\%$. The ANGSD output was then subsampled for one SNP out of every 50 and pairwise linkage disequilibrium (LD) was estimated with ngsLD v1.1.1 [124] on this subsampled dataset. Linked sites were pruned with the script “prune_graph.pl” with

a maximum distance of 10 kb between SNPs and a minimum weight of 0.5 [125]. The final LD-filtered dataset was used to generate phylogenetic trees, ngsADMIX and MDS plots, and to calculate population pairwise F_{ST} statistics.

Lineage-specific filtered dataset

The second dataset was filtered using the same ANGSD flags as the first dataset (excluding the LD filtering steps) but then underwent a second round of lineage-level filtering. Specifically, sampling locations from Woods Hole, MA to Roanoke Island, NC were included in the “northern lineage” group and sampling locations from Pivers Island, NC to Panacea, FL were included in the “southern lineage” group. These two groups were filtered separately and then later recombined. For the northern lineage, we filtered the dataset for (1) positions that were present in at least 36 out of 45 individuals, (2) a maximum coverage of $900 \times$ coverage per site across all individuals, (3) a minimum of $108 \times$ coverage per site across all individuals. For the southern lineage, we filtered the dataset for (1) positions that were present in at least 58 out of 73 individuals, (2) a maximum coverage of $1,460 \times$ coverage per site across all individuals, (3) a minimum of $175 \times$ coverage per site across all individuals. We restricted the secondary lineage-specific filtering step to the subset of positions that were retained in the first round of filtering using the sites option in ANGSD. Filtering in this manner guarantees that sites with an allele that is fixed in one lineage will be retained. The final lineage-specific filtered dataset was used to generate sliding window F_{ST} values, XP-CLR values, and to calculate f_3 statistics.

All-sites filtered dataset

The third dataset was filtered with the same method as the Lineage-specific filtered dataset, but we included all non-polymorphic sites. This dataset was used as input to our demographic analyses and to calculate nucleotide diversity [126], Tajima's D [127] and D_{XY} values.

Population structure analysis

We used the ngsDist option in ngsTools [128] on the LD-filtered dataset to compute genetic distances between individuals and perform multiple dimensional scaling (MDS). We then looked for signatures of admixture using ngsADMIX [68] with runs ranging from $K=2$ to $K=7$. Weighted pairwise F_{ST} values were estimated between the individuals from the North (Woods Hole, MA to Roanoke Island, NC) and the individuals from the South (Pivers Island, NC to Panacea, FL) using ANGSD with the `-doSaf 1` flag. We then used realSFS to estimate the 2D site frequency spectrum for each population and calculated the average pairwise weighted F_{ST} using realSFS

fst. Weighted pairwise F_{ST} estimates were also calculated between each sampling location using ANGSD with the same parameters.

NgsDist was used to compute genetic distances from genotype probabilities with *Bolinopsis infundibulum* as the outgroup (PRJNA818620; [85]). Raw *Bolinopsis* sequences were filtered according to the same protocol as was applied to *Mnemiopsis* and then we ran ANGSD with *Bolinopsis* included, a minor allele frequency filter of 0.05 and a SNP P -value cut off of 1×10^{-6} . This set of filters resulted in 1,302 SNPs that were used to generate the phylogenetic tree based on the nuclear genome. NgsDist was run with 100 bootstrap replicates and a boot block size of 20. Finally, FastME v2.1.6.1 [129] was used to visualize the pairwise genetic distances and RAxML-NG v0.9.0 [130] was used to place bootstrap supports onto the tree. Final figures were generated using the Interactive Tree of Life tool v6.9.1 [131].

Novoplasty v4.3.1 [132] was used to extract the mitochondrial genome from each trimmed sample using the published *Mnemiopsis leidyi* COI sequence (KF435121.1 [133]) as the seed. When this was unsuccessful, we instead used the published *Bolinopsis microptera* (PRJNA716277 [87]) whole-genome sequences and ran Novoplasty with the published *B. microptera* COI seed (MW735734.1 [134]) to extract the complete mitochondrial genome. The complete *B. microptera* mitochondrial genome was then used as a seed for the remaining samples that initially failed Novoplasty. Using this approach, we were able to successfully assemble 112 out of 118 complete *Mnemiopsis* mitochondrial genomes. Next, we used MARS [135] to ensure that each mitochondrial genome began at the same location using the branch and bound method and the block length set to 100. We used MAFFT v7.480 [136] to generate an alignment and IQ-TREE v2.3.2 [137] to generate a phylogenetic tree.

F_3 statistics

We used f_3 statistics, included as part of the ADMIXTOOLS package v.7.0.2 [69], as a test for detecting hybridization. The f_3 statistic is a tool to test whether a population is admixed between two source populations or to measure shared genetic drift between two test populations compared to an outgroup population. If the f_3 statistic is negative, this suggests that admixture has likely occurred. For this analysis, we tested the most northern population sample (MAW) and the most southern population sample (FLP) against Roanoke Island, NC and Gloucester Point, VA.

Demographic analysis

To infer the demographic history of these species, we used two approaches based on allele frequency spectra.

First, StairwayPlot v2 [73] was used to reconstruct effective population size estimates based on the folded SFS generated from the All-Sites Filtered dataset. Stairway Plot is an unsupervised analysis method that does not require prespecified demographic models. We used a mutation rate of 6.85×10^{-8} per base pair per year [57] and a generation time of 1 year [138]. The hermaphroditic *Mnemiopsis* can reproduce within the first month of its lifespan [139], its lifespan is approximately 2 years [138], and there is no information indicating that they ever lose the ability to reproduce. As such, the mean age of reproduction was set to 1 year but due to the surrounding uncertainty, we plotted the results in generations.

Second, we used the *Moments* Python library [80] based on the two-dimensional AFS to estimate population size changes and the timing and magnitude of introgression between each pair of genetic lineages. We subsampled our data so that we only included four populations, two *M. leidyi* populations (Woods Hole, MA and Little Egg Harbor, NJ) and two *M. gardeni* populations (Fort Pierce, FL and Panacea, FL), to conservatively remove putative hybrids and for computational efficiency. This dataset returned approximately 21 million sites in total. Unlike the Stairway Plot analysis, the *Moments* workflow requires demographic models to be specified at the start of the analysis. Here, we explicitly tested 102 demographic models (described in detail at <https://github.com/zoon/AFS-analysis-with-moments/>) that represent a wide range of possible demographic histories. Following the *Moments* pipeline described in Rippe et al. (2021), we first generated 100 bootstrapped two-population AFS datasets [140]. For a subset of 10 of these bootstrapped AFS, each of the 102 models was run six times using randomized starting parameters. Model log likelihoods were converted to Akaike's Information Criterion (AIC) which were then used to find the best-fit run out of the six runs for each bootstrapped replicate. From the resulting set of AIC values, the demographic model with the lowest median AIC value across all 10 bootstrapped AFS replicates was chosen. Finally, the "winning" demographic model was fit to all 100 bootstrapped AFS's to evaluate parameter uncertainties. As with the Stairway plot analysis, we assumed a mutation rate of 6.85×10^{-8} per base pair per year and a generation time of 1 year.

Genomic signatures of selection

To detect regions of the genome under putative selection between the two lineages, we used XP-CLR [98] which uses the AFS to calculate local deviations between populations. We used a sliding window and step size of 10 kb with the `maxSNPs` flag set to 200. The genomic regions in the top 10% of the XP-CLR values were considered to represent selective sweeps and were associated with

overlapping genes using Bedtools v.2.29.0 [141]. We used InterProScan (version 5.25–64) to associate gene ontology categories with the *M. gardeni* gene models. We conducted a Monte Carlo analysis to identify overrepresented gene ontology terms as was conducted in Babonis et al. 2018 [110].

To further characterize genomic patterns of regions under selection, we generated sliding window scans for F_{ST} , D_{XY} , nucleotide diversity (π), and Tajima's D values, with the hybrid Roanoke Island, NC population removed. The lineage-specific filtered dataset was used to generate the F_{ST} values. We used `realSFS` F_{ST} index in ANGSD to compute per-site F_{ST} indexes and `realSFS` F_{ST} stats2 to perform a sliding window analysis with the window and step size set to 5 kb to ensure nonoverlapping windows. The all-sites filtered dataset was used to generate nucleotide diversity, Tajima's D , and D_{XY} values. For nucleotide diversity and Tajima's D calculations, we used `realSFS` `saf2theta` in ANGSD to generate folded SFS's and then calculated these values along a sliding window (window size 5 kb, step size 5 kb) with `thetaStat` `do_stat`. For D_{XY} calculations, we used the `CalcDxy.R` code by Joshua Penalba in `ngsTools` [128] and then averaged the resulting values over the same 5 kb windows as the F_{ST} and nucleotide calculations. By combining several tests and summary statistics, we can make inferences about evolutionary forces acting on regions of interest with greater certainty.

Abbreviations

SNP	Single nucleotide polymorphism
SV	Structural variant
MDS	Multidimensional scaling
PSMC	Pairwise Sequentially Markovian Coalescent
AIC	Akaike's Information Criterion
GO	Gene ontology
COI	Mitochondrial Cytochrome C Oxidase subunit I
LD	Linkage disequilibrium
CNV	Copy number variation

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-026-02578-8>.

Additional File 1: Tables 1, 2, 4, 5, 7, 8, 10, 11, 12 and Figs. 1–6. Table S1 – Collection information for all samples collected along the US Atlantic coast including date of collection, number of samples collected, coordinates, and collection method. Table S2 – Comparison of genome assembly benchmarks between the two published *Mnemiopsis leidyi* genomes and the two new genome assemblies. Table S4 – F_{ST} values across all populations of *Mnemiopsis* collected based on the LD-Filtered SNP dataset. Table S5 – f_3 results where each row represents a scenario where the FLP and MAW populations are the source for admixture and the third population is the target of hybridization. A f_3 statistic, the standard error (SE) and a Z-score value calculated with jackknife is provided for each scenario. Table S7 – Gene Ontology (GO) analysis of genes in the top 1% regions of copy number variation outliers reveals enrichment in molecular functions (MF), cellular components (CC), and biological processes (BP). Table S8 – Putative G protein-coupled receptors in the top 1% of copy number variation outliers. Table S10 – Putative G-coupled protein

receptors in the top 10% of XP-CLR outliers with corresponding *M. gardeni* gene ID. Table S11 – Channels and receptors in the top 10% of XP-CLR outliers with corresponding *M. gardeni* gene ID. Table S12 – Gene Ontology (GO) analysis of genes in the top 10% of XP-CLR outlier regions reveals enrichment in molecular functions, cellular components, and biological processes. Figure S1 – Admixture plots for the 118 *Mnemiopsis* samples collected for $K = 5$ to $K = 7$ where each bar represents a single individual, and the colors represent each of the K components. Figure S2 – Boxplots summarizing the best AIC scores for each model for all bootstrapped replicates for each of 102 demographic models tested. The inset plot shows the boxplot for the median AIC for the top ten best-fitting models in the set. Figure S3 – Plots showing the goodness-of-fit of the best fitting model (sc3ielsm1). The top two panels show the observed AFS (top left) and the modeled AFS (top right). The lower two plots show model errors plotted as residuals between the best-fit model and the data (bottom left) and the frequency histogram of residuals (bottom right). Figure S4 – The best-fit demographic model (sc3ielsm1) inferred using *Moments* for the demographic history between *M. leidyi* and *M. gardeni*. This best-fit model includes three epochs in each species with symmetric migration in the first and asymmetric migration in the last. Figure S5 – PCA plot based on copy number variation among samples. Copy number estimates were generated as denoised copy number ratios in GATK. Individuals are colored by their sampling location. Figure S6 – The top V_{ST} outlier gene occurs in an *M. gardeni*-specific duplicated region. This dot plot shows BLAST hits (e -value $< 1e-10$; length ≥ 500 bp) between corresponding regions in the *M. leidyi* and *M. gardeni* assemblies. The V_{ST} outlier gene (jg9426 – Mgar.v1.001:9329910–9331046) occurs as part of a larger duplicated region in *M. gardeni*. In *M. leidyi*, this region is single-copy. Arrows highlight the location of the *M. gardeni* duplicated genes.

Additional File 2: Table S3. Table S3 – Assembled gene/protein models that were annotated using three protein databases: UniProt Knowledgebase Swiss-Prot protein models v2021-03, RefSeq invertebrate protein models, and the *M. leidyi* protein models.

Additional file 3: Table S6. Table S6 – The top 1% of fixation index values (V_{ST}) for the copy number variation analyses with their associated annotations taken from Additional File 2: Table S3.

Additional file 4: Table S9. Table S9 – The top 10% of XP-CLR values calculated between the two lineages using a sliding window approach, a step-size of 10 kb and the `maxSNPs` flag set to 200. We have appended their associated annotations from Additional File 2: Table S3.

Acknowledgements

We would like to thank Dr. Ann Tarrant, Roland Hagan, Robert Aguilar, Dr. Deborah Steinberg, Tor Mowatt-Larssen, Eli at Outer Banks Kayaking Adventures, Ken Song, Carol Smith, Dr. Joshua Osterberg, Dr. Jacob Warner, Dr. Daniel Sasson, Dr. Marc Frischer, Dr. Adam Greer, Kyle Aaron, Dr. Holly Sweat, and Dr. Yuriy Bobkov for help collecting *Mnemiopsis* samples along the US Atlantic coast. We also thank Dr. Dan Schrider and Austin Daigle at UNC Chapel Hill for help analyzing and interpreting demographic modeling results. We would like to acknowledge the ICBB for their excellent sequencing and service (RRID:SCR_019152). We would like to thank Dr. James Thomas at the NIH Intramural Sequencing Center (NISC) for his help with the PacBio sequencing. We would also like to thank Dr. Gustav Paulay and Dr. Claudia Mills for their help navigating the complex taxonomy of *Mnemiopsis*. Finally, we want to thank Dr. Mark Ravinet for his help brainstorming ideas pertaining to the hybridization of these two species.

Authors' contributions

Conceptualization: R.N.K and J.F.R. Investigation (fieldwork): R.N.K, W.B.L, E.G.S, J.F.R. Investigation (analysis): R.N.K, E.G.S, J.F.R, L.M.T, N.E.P. Resources: J.F.R, A.D.B, A.M.R. All authors read and approved the final manuscript.

Funding

This work was supported in part with a Postdoctoral Research Fellowship in Biology (PRFB) from the National Science Foundation (award number: 2109712) to R.N.K, an Allen Distinguished Investigator Award from the Paul G. Allen Frontiers Group of the Paul G. Allen Family Foundation to J.F.R. This research was also supported in part by the Intramural Research Program of

the National Human Genome Research Institute, National Institutes of Health (ZIA HG000140 to A.D.B).

Data availability

The datasets supporting the conclusions of this article are available at the NCBI under BioProject accession number PRJEB87530. The genome assemblies for each *Mnemiopsis* species are also available under the above BioProject accession codes. All scripts are publicly available at https://github.com/remiketchum/Mnemiopsis_2026.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

The authors have all been consulted on draft and final versions of the manuscript and have consented to publication.

Competing interests

Authors declare no competing interests.

Author details

¹Whitney Laboratory for Marine Bioscience, University of Florida, St. Augustine, Florida, USA. ²Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. ³School of Life Sciences, The University of Warwick, Coventry, UK. ⁴Department of Biology, DeSales University, Center Valley, PA, USA. ⁵Center for Genomics and Data Science Research, Division of Intramural Research, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA. ⁶Department of Biological Sciences, University of North Carolina at Charlotte, Charlotte, NC, USA. ⁷Department of Biology, University of Florida, Gainesville, FL, USA.

Received: 23 July 2025 Accepted: 17 March 2026

Published online: 21 April 2026

References

- Matute DR, Cooper BS. Comparative studies on speciation: 30 years since Coyne and Orr. *Evolution*. 2021;75:764–78. <https://doi.org/10.1111/evo.14181>.
- Mayr E. *Animal species and evolution*. Harvard University Press; 1963.
- Coyne JA, Orr HA. *Speciation: a catalogue and critique of species concepts*. *Philos Biol Anthol*. 2004;272–92.
- Frazer ME, Robles NV, Rodríguez-Barrera MJ, Coughlan JM, Schumer M. The molecular evolutionary basis of species formation revisited. *Trends Genet*. 2025;41:1068–95. <https://doi.org/10.1016/j.tig.2025.07.003>.
- Bolnick DI, Fitzpatrick BM. Sympatric speciation: models and empirical evidence. *Annu Rev Ecol Syst*. 2007;38:459–87.
- Paudel Y, Madsen O, Megens H-J, Frantz LAF, Bosse M, Crooijmans RPMA, et al. Copy number variation in the speciation of pigs: a possible prominent role for olfactory receptors. *BMC Genomics*. 2015;16:330. <https://doi.org/10.1186/s12864-015-1449-9>.
- Anderson SA, Weir JT. The role of divergent ecological adaptation during allopatric speciation in vertebrates. *Science*. 2022;378:1214–8.
- De Jode A, Le Moan A, Johannesson K, Faria R, Stankowski S, Westram AM, et al. Ten years of demographic modelling of divergence and speciation in the sea. *Evol Appl*. 2023;16:542–59.
- Palumbi SR. Genetic divergence, reproductive isolation, and marine speciation. *Annu Rev Ecol Syst*. 1994;547–72.
- Pogson GH. Studying the genetic basis of speciation in high gene flow marine invertebrates. *Curr Zool*. 2016;62:643–53.
- Filatov DA. How does speciation in marine plankton work? *Trends Microbiol*. 2023;31:989–91. <https://doi.org/10.1016/j.tim.2023.07.005>.
- Ravinet M, Faria R, Butlin RK, Galindo J, Bierre N, Rafajlović M, et al. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *J Evol Biol*. 2017;30:1450–77.
- Poikela N, Laetsch DR, Kankare M, Hoikkala A, Lohse K. Experimental introgression in *Drosophila*: asymmetric postzygotic isolation

- associated with chromosomal inversions and an incompatibility locus on the X chromosome. *Mol Ecol.* 2023;32:854–66.
14. Dettman JR, Anderson JB, Kohn LM. Divergent adaptation promotes reproductive isolation among experimental populations of the filamentous fungus *Neurospora*. *BMC Evol Biol.* 2008;8:35. <https://doi.org/10.1186/1471-2148-8-35>.
 15. Dobzhansky T. Genetics and the origin of species. Columbia university press; 1982.
 16. Foote AD. Sympatric speciation in the genomic era. *Trends Ecol Evol.* 2018;33:85–95.
 17. White NJ, Snook RR, Eyres I. The past and future of experimental speciation. *Trends Ecol Evol.* 2020;35:10–21. <https://doi.org/10.1016/j.tree.2019.08.009>.
 18. Feder JL, Gejji R, Yeaman S, Nosil P. Establishment of new mutations under divergence and genome hitchhiking. *Philos Trans R Soc Lond B Biol Sci.* 2012;367:461–74. <https://doi.org/10.1098/rstb.2011.0256>.
 19. Gavrilets S. Fitness landscapes and the origin of species. Princeton university press; 2004.
 20. Michel AP, Sim S, Powell TH, Taylor MS, Nosil P, Feder JL. Widespread genomic divergence during sympatric speciation. *Proc Natl Acad Sci USA.* 2010;107:9724–9.
 21. Weissensteiner MH, Bunikis I, Catalán A, Francoijs K-J, Knief U, Heim W, et al. Discovery and population genomics of structural variation in a songbird genus. *Nat Commun.* 2020;11:3403.
 22. Davey JW, Chouteau M, Barker SL, Maroja L, Baxter SW, Simpson F, et al. Major improvements to the *Heliconius melpomene* genome assembly used to confirm 10 chromosome fusion events in 6 million years of butterfly evolution. *G3 Genes[Genomes]Genetics.* 2016;6(3):695–708.
 23. Giménez MD, Förster DW, Jones EP, Jóhannesdóttir F, Gabriel SI, Panithanarak T, et al. A half-century of studies on a chromosomal hybrid zone of the house mouse. *J Hered.* 2017;108:25–35.
 24. Dierckxsens N, Li T, Vermeesch JR, Xie Z. A benchmark of structural variation detection by long reads through a realistic simulated model. *Genome Biol.* 2021;22:1–16.
 25. Zhang L, Reifová R, Halenková Z, Gompert Z. How important are structural variants for speciation? *Genes.* 2021;12:1084.
 26. Ferree PM, Barbash DA. Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in *Drosophila*. *PLoS Biol.* 2009;7:e1000234.
 27. Ayala D, Guerrero RF, Kirkpatrick M. Reproductive isolation and local adaptation quantified for a chromosome inversion in a malaria mosquito. *Evolution.* 2013;67:946–58.
 28. Hager ER, Harringmeyer OS, Wooldridge TB, Theingi S, Gable JT, McFadden S, et al. A chromosomal inversion contributes to divergence in multiple traits between deer mouse ecotypes. *Science.* 2022;377:399–405.
 29. Zuellig MP, Sweigart AL. Gene duplicates cause hybrid lethality between sympatric species of *Mimulus*. *PLoS Genet.* 2018;14:e1007130.
 30. Ho SS, Urban AE, Mills RE. Structural variation in the sequencing era. *Nat Rev Genet.* 2020;21:171–89.
 31. Choo LQ, Spagliardi G, Malinsky M, Choquet M, Goetze E, Hoarau G, et al. Genome-wide phylogeography reveals cryptic speciation in the circumglobal planktonic calcifier *Limacina bulimoides*. *Mol Ecol.* 2023;32:3200–19.
 32. Breusing C, Johnson SB, Tunnicliffe V, Clague DA, Vrijenhoek RC, Beinart RA. Allopatric and sympatric drivers of speciation in *Alviniconcha* hydrothermal vent snails. *Mol Biol Evol.* 2020;37:3469–84. <https://doi.org/10.1093/molbev/msaa177>.
 33. Schluter D, Howard DJ, Berlocher SH. Endless forms: species and speciation. Oxford Univ. Press; 1998.
 34. Jaspers C, Costello JH, Colin SP. Carbon content of *Mnemiopsis leidyi* eggs and specific egg production rates in northern Europe. *J Plankton Res.* 2015;37:11–5.
 35. Shiganova T. Changes in appendicularian Oikopleura dioica abundance caused by invasion of alien ctenophores in the Black Sea. *J Mar Biol Assoc U K.* 2005;85:477–94.
 36. Ward RD, Woodward M, Skibinski DOF. A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. *J Fish Biol.* 1994;44:213–32.
 37. Ketchum RN, Smith EG, DeBiasse MB, Vaughan GO, McParland D, Leach WB, et al. Population genomic analyses of the sea urchin *Echinometra* sp. EZ across an extreme environmental gradient. *Genome Biol Evol.* 2020;12:1819–29.
 38. Knutsen H, Jorde PE, André C, Stenseth NC. Fine-scaled geographical population structuring in a highly mobile marine species: the Atlantic cod. *Mol Ecol.* 2003;12:385–94.
 39. Yan R-J, Schnabel KE, Rowden AA, Guo X-Z, Gardner JP. Population structure and genetic connectivity of squat lobsters (*Munida* Leach, 1820) associated with vulnerable marine ecosystems in the southwest Pacific Ocean. *Front Mar Sci.* 2020;6:791.
 40. Kyle CJ, Boulding EG. Comparative population genetic structure of marine gastropods (*Littorina* spp.) with and without pelagic larval dispersal. *Mar Biol.* 2000;137:835–45.
 41. Mianzan HW, Martos P, Costello JH, Guerrero RA. Avoidance of hydrodynamically mixed environments by *Mnemiopsis leidyi* (Ctenophora: Lobata) in open-sea populations from Patagonia, Argentina. *Jellyfish Blooms New Probl Solut.* 2010;645:113–24.
 42. Harbison GR, Volovik SP. The ctenophore, *Mnemiopsis leidyi*, in the Black Sea: a holoplanktonic organism transported. In: Nonindigenous Estuarine & Marine Organisms (NEMO): Proceedings of the Conference and Workshop, Seattle, Washington, April 1993. US Department of Commerce, National Oceanic and Atmospheric Administration; 1994. p. 25.
 43. Lowe S, Browne M, Boudjelas S, De Poorter M. 100 of the world's worst invasive alien species: a selection from the global invasive species database. Auckland: Invasive Species Specialist Group; 2000.
 44. Doyle L. *Mnemiopsis leidyi*. AN INTRODUCTION TO MARINE INVASIVE SPECIES. 1984;52(560):561.
 45. Shiganova TA. Adaptive strategies of *Mnemiopsis leidyi* A. Agassiz 1865 in different environments of the Eurasian seas. *Mar Pollut Bull.* 2020;161:111737.
 46. Jaspers C, Möller LF, Kiørboe T. Reproduction rates under variable food conditions and starvation in *Mnemiopsis leidyi*: significance for the invasion success of a ctenophore. *J Plankton Res.* 2015;37:1011–8.
 47. Edgar A, Mitchell DG, Martindale MQ. Whole-body regeneration in the lobate ctenophore *Mnemiopsis leidyi*. *Genes.* 2021;12:867.
 48. Sasson DA, Ryan JF. The sex lives of ctenophores: the influence of light, body size, and self-fertilization on the reproductive output of the sea walnut, *Mnemiopsis leidyi*. *PeerJ.* 2016;4:e1846.
 49. Sasson DA, Jacquez AA, Ryan JF. The ctenophore *Mnemiopsis leidyi* regulates egg production via conspecific communication. *BMC Ecol.* 2018;18:1–10.
 50. Prentis PJ, Wilson JR, Dormontt EE, Richardson DM, Lowe AJ. Adaptive evolution in invasive species. *Trends Plant Sci.* 2008;13:288–94.
 51. Lee CE. Evolutionary genetics of invasive species. *Trends Ecol Evol.* 2002;17:386–91.
 52. Child CM. The gradient in susceptibility to cyanides in the meridional conducting path of the ctenophore *Mnemiopsis*. *Am J Physiol-Leg Content.* 1917;43:87–112.
 53. Ryan JF, Pang K, Schnitzler CE, Nguyen A-D, Moreland RT, Simmons DK, et al. The genome of the ctenophore *Mnemiopsis leidyi* and its implications for cell type evolution. *Science.* 2013;342:1242592.
 54. Moreland RT, Nguyen A-D, Ryan JF, Schnitzler CE, Koch BJ, Siewert K, et al. A customized web portal for the genome of the ctenophore *Mnemiopsis leidyi*. *BMC Genomics.* 2014;15:1–13.
 55. Moreland RT, Nguyen A-D, Ryan JF, Baxevanis AD. The *Mnemiopsis* genome project portal: integrating new gene expression resources and improving data visualization. *Database.* 2020;2020:baaa029.
 56. Moreland RT, Zhang S, Barreira SN, Ryan JF, Baxevanis AD. An AI-generated proteome-scale dataset of predicted protein structures for the ctenophore *Mnemiopsis leidyi*. *Proteomics.* 2024;24:2300397.
 57. Jaspers C, Ehrlich M, Pujolar JM, Künzel S, Bayer T, Limborg MT, et al. Invasion genomics uncover contrasting scenarios of genetic diversity in a widespread marine invader. *Proc Natl Acad Sci.* 2021;118:e2116211118.
 58. Bayha KM, Chang MH, Mariani CL, Richardson JL, Edwards DL, DeBoer TS, et al. Worldwide phylogeography of the invasive ctenophore *Mnemiopsis leidyi* (Ctenophora) based on nuclear and mitochondrial DNA data. *Biol Invasions.* 2015;17:827–50.
 59. Verwimp C, Vansteenbrugge L, Derycke S, Kerkhove T, Muylle H, Honnay O, et al. Population genomic structure of the gelatinous

- zooplankton species *Mnemiopsis leidyi* in its nonindigenous range in the North Sea. *Ecol Evol.* 2020;10:11–25.
60. Reusch TB, Bolte S, Sparwel M, Moss AG, Javidpour J. Microsatellites reveal origin and genetic diversity of Eurasian invasions by one of the world's most notorious marine invader, *Mnemiopsis leidyi* (Ctenophora). *Mol Ecol.* 2010;19:2690–9.
 61. Pujolar JM, Breitburg D, Lee J, Decker MB, Jaspers C. Hybridization and adaptive introgression in a marine invasive species in native habitats. *iScience.* 2023;26.
 62. Ah Yong S, Boyko CB, Bernot J, Brandão SN, Daly M, De Grave S, et al. World Register of Marine Species (WoRMS). 2024.
 63. Seravin LN. The systematic revision of the genus *Mnemiopsis* (Ctenophora, Lobata). *Zool Ž.* 1994;73:9–18.
 64. Manni M, Berkeley MR, Seppey M, Zdobnov EM. BUSCO: assessing genomic data quality and beyond. *Curr Protoc.* 2021;1:e323.
 65. Koutsouveli V, Torres-Oliva M, Bayer T, Fuß J, Grossschmidt N, Marulanda-Gomez AM, et al. The chromosome-level genome of the ctenophore *Mnemiopsis leidyi* A. Agassiz, 1865 reveals a unique immune gene repertoire. *Genome Biol Evol.* 2025;17.
 66. Pett W, Ryan JF, Pang K, Mullikin JC, Martindale MQ, Baxevanis AD, et al. Extreme mitochondrial evolution in the ctenophore *Mnemiopsis leidyi*: insight from mtDNA and the nuclear genome. *Mitochondrial DNA.* 2011;22:130–42. <https://doi.org/10.3109/19401736.2011.624611>.
 67. Toews DP, Brelsford A. The biogeography of mitochondrial and nuclear discordance in animals. *Mol Ecol.* 2012;21:3907–30.
 68. Skotte L, Korneliusen TS, Albrechtsen A. Estimating individual admixture proportions from next generation sequencing data. *Genetics.* 2013;195:693–702.
 69. Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, et al. Ancient admixture in human history. *Genetics.* 2012;192:1065–93.
 70. Ataç G, Waldman S, Carmi S, Somel M. An explanation for the sister repulsion phenomenon in Patterson's f-statistics. *Genetics.* 2024;228:iyae144.
 71. Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, Simpson F, et al. Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Res.* 2013;23:1817–28.
 72. Li M-R, Wang H-Y, Ding N, Lu T, Huang Y-C, Xiao H-X, et al. Rapid divergence followed by adaptation to contrasting ecological niches of two closely related columbine species *Aquilegia japonica* and *A. oxysepala*. *Genome Biol Evol.* 2019;11:919–30.
 73. Liu X, Fu Y-X. Stairway plot 2: demographic history inference with folded SNP frequency spectra. *Genome Biol.* 2020;21:280.
 74. Kleiven H (Kikki)F, Kissel C, Laj C, Ninnemann US, Richter TO, Cortijo E. Reduced North Atlantic deep water coeval with the glacial Lake Agassiz freshwater outburst. *Science.* 2008;319:60–4.
 75. Teller JT, Leverington DW, Mann JD. Freshwater outbursts to the oceans from glacial Lake Agassiz and their role in climate change during the last deglaciation. *Quat Sci Rev.* 2002;21:879–87. [https://doi.org/10.1016/S0277-3791\(01\)00145-7](https://doi.org/10.1016/S0277-3791(01)00145-7).
 76. Martínez-García L, Ferrari G, Oosting T, Ballantyne R, Van der Jagt I, Ystgaard I, et al. Historical demographic processes dominate genetic variation in ancient Atlantic cod mitogenomes. *Front Ecol Evol.* 2021;9:671281.
 77. Atmore LM, Martínez-García L, Makowiecki D, André C, Lôugas L, Barrett JH, et al. Population dynamics of Baltic herring since the Viking Age revealed by ancient DNA and genomics. *Proc Natl Acad Sci.* 2022;119:e2208703119.
 78. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. *Nature.* 2011;475:493–6.
 79. Liu X, Fu Y-X. Exploring population size changes using SNP frequency spectra. *Nat Genet.* 2015;47:555–9.
 80. Jouganous J, Long W, Ragsdale AP, Gravel S. Inferring the joint demographic history of multiple populations: beyond the diffusion approximation. *Genetics.* 2017;206:1549–67.
 81. Prada C, Hellberg ME. Speciation-by-depth on coral reefs: sympatric divergence with gene flow or cryptic transient isolation? *J Evol Biol.* 2021;34:128–37.
 82. Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, et al. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science.* 2007;317:86–94.
 83. Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, Kawashima T, et al. The amphioxus genome and the evolution of the chordate karyotype. *Nature.* 2008;453:1064–71.
 84. Simakov O, Marletaz F, Cho S-J, Edsinger-Gonzales E, Havlak P, Hellsten U, et al. Insights into bilaterian evolution from three spiralian genomes. *Nature.* 2013;493:526–31.
 85. Schultz DT, Haddock SH, Bredeson JV, Green RE, Simakov O, Rokhsar DS. Ancient gene linkages support ctenophores as sister to other animals. *Nature.* 2023;618:110–7.
 86. Lewin TD, Liao I-Y, Luo Y-J. Annelid comparative genomics and the evolution of massive lineage-specific genome rearrangement in bilaterians. *Mol Biol Evol.* 2024;41:msae172.
 87. Johnson SB, Winnikoff JR, Schultz DT, Christianson LM, Patry WL, Mills CE, et al. Speciation of pelagic zooplankton: invisible boundaries can drive isolation of oceanic ctenophores. *Front Genet.* 2022;13:970314.
 88. Mérot C, Oomen RA, Tigano A, Wellenreuther M. A roadmap for understanding the evolutionary significance of structural genomic variation. *Trends Ecol Evol.* 2020;35:561–72.
 89. O'Neill MJ, O'Neill RJ. Sex chromosome repeats tip the balance towards speciation. *Mol Ecol.* 2018;27:3783–98.
 90. Hull RM, Cruz C, Jack CV, Houseley J. Environmental change drives accelerated adaptation through stimulated copy number variation. *PLoS Biol.* 2017;15:e2001333.
 91. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. *Nat Genet.* 2000;25:25–9.
 92. Babonis LS, Martindale MQ, Ryan JF. Do novel genes drive morphological novelty? An investigation of the nematosomes in the sea anemone *Nematostella vectensis*. *BMC Evol Biol.* 2016;16:114. <https://doi.org/10.1186/s12862-016-0683-3>.
 93. Schluter D. Evidence for ecological speciation and its alternative. *Science.* 2009;323:737–41.
 94. Carruthers M, Carleton KL, Linderot T, Bridle J, Hudson AG, Malinsky M, et al. Rapid divergence of visual systems and signaling traits to contrasting light regimes during early speciation of African crater lake cichlid fish. *Mol Biol Evol.* 2025;42:msaf204.
 95. Klein J, Ono H, Klein D, O'hUigin C. The accordion model of MHC evolution. In: *Progress in Immunology Vol. VIII.* Springer; 1993. p. 137–43.
 96. Secomandi S, Gallo GR, Rossi R, Rodríguez Fernandes C, Jarvis ED, Bonisoli-Alquati A, et al. Pangenome graphs and their applications in biodiversity genomics. *Nat Genet.* 2025;57:13–26. <https://doi.org/10.1038/s41588-024-02029-6>.
 97. Sirén J, Monlong J, Chang X, Novak AM, Eizenga JM, Markello C, et al. Pangenomics enables genotyping of known structural variants in 5202 diverse genomes. *Science.* 2021;374:abg8871.
 98. Chen H, Patterson N, Reich D. Population differentiation as a test for selective sweeps. *Genome Res.* 2010;20:393–402.
 99. Torres C-R, Hart GW. Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc. *J Biol Chem.* 1984;259:3308–17.
 100. Guo B, Liang Q, Li L, Hu Z, Wu F, Zhang P, et al. O-GlcNAc-modification of SNAP-29 regulates autophagosome maturation. *Nat Cell Biol.* 2014;16:1215–26. <https://doi.org/10.1038/ncb3066>.
 101. Mizushima N. Sugar modification inhibits autophagosome-lysosome fusion. *Nat Cell Biol.* 2014;16:1132–3. <https://doi.org/10.1038/ncb3078>.
 102. Morinaga G, Balcazar D, Badolo A, Iyaloo D, Tantely ML, Mouillaud T, Sharakhova M, Geib SM, Paupy C, Ayala D, Powell JR. From macro to micro: De novo genomes of *Aedes* mosquitoes enable comparative genomics among close and distant relatives. *Genome Biology and Evolution.* 2025;17(8):evaf142.
 103. MacGuigan DJ, Krabbenhoft TJ, Harrington RC, Wainwright DK, Backenstose NJC, Near TJ. Lacustrine speciation associated with chromosomal inversion in a lineage of riverine fishes. *Evolution.* 2023;77:1505–21.
 104. Wu N, Evans E, van Schooten B, Meléndez-Rosa J, Ortiz Y, Planas Soto-Navarro SM, et al. Widespread gene expression divergence in butterfly sensory tissues plays a fundamental role during reproductive isolation and speciation. *Mol Biol Evol.* 2022;39.
 105. Yoshida K, Rödelberger C, Röseler W, Riebesell M, Sun S, Kikuchi T, et al. Chromosome fusions repatterned recombination rate and facilitated reproductive isolation during *Pristionchus* nematode speciation. *Nat Ecol Evol.* 2023;7:424–39.

106. Ehrlich M. Invasion Genomics: population structure and diversity patterns in the invasive ctenophore *Mnemiopsis leidyi* based on whole-genome re-sequencing. Christian-Albrechts-Universität Kiel; 2017.
107. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods*. 2021;18:170–5.
108. Roach MJ, Schmidt SA, Borneman AR. Purge haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics*. 2018;19:1–10.
109. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci*. 2020;117:9451–7.
110. Babonis LS, DeBiasse MB, Francis WR, Christianson LM, Moss AG, Haddock SH, et al. Integrating embryonic development and evolutionary history to characterize tentacle-specific cell types in a ctenophore. *Mol Biol Evol*. 2018;35:2940–56.
111. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.
112. Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics Bioinformatics*. 2021;3:lqaa108.
113. Buchfink B, Reuter K, Drost H-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods*. 2021;18:366–8.
114. El Hilali S, Copley RR. macrosyntR: Drawing automatically ordered Oxford Grids from standard genomic files in R. *BioRxiv*. 2023:2023-01.
115. Babadi M, Fu JM, Lee SK, Smirnov AN, Gauthier LD, Walker M, et al. GATK-gCNV enables the discovery of rare copy number variants from exome sequencing data. *Nat Genet*. 2023;55:1589–97.
116. Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: a fast and versatile genome alignment system. *PLoS Comput Biol*. 2018;14:e1005944.
117. Goel M, Sun H, Jiao W-B, Schneeberger K. SyRl: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol*. 2019;20:1–13.
118. Goel M, Schneeberger K. Plotsr: visualizing structural similarities and rearrangements between multiple genomes. *Bioinformatics*. 2022;38:2922–6.
119. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *bioinformatics*. 2009;25:2078–9.
120. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20.
121. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics*. 2015;31:2032–4.
122. Jun G, Wing MK, Abecasis GR, Kang HM. An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. *Genome Res*. 2015;25:918–25.
123. Korneliusen TS, Albrechtsen A, Nielsen R. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics*. 2014;15:1–13.
124. Fox EA, Wright AE, Fumagalli M, Vieira FG. ngsLD: evaluating linkage disequilibrium using genotype likelihoods. *Bioinformatics*. 2019;35:3855–6.
125. Lou RN, Jacobs A, Wilder AP, Therkildsen NO. A beginner's guide to low-coverage whole genome sequencing for population genomics. *Mol Ecol*. 2021;30:5966–93.
126. Nei M, Li W-H. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci*. 1979;76:5269–73.
127. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 1989;123:585–95.
128. Fumagalli M, Vieira FG, Linderot T, Nielsen R. ngsTools: methods for population genetics analyses from next-generation sequencing data. *Bioinformatics*. 2014;30:1486–7.
129. Lefort V, Desper R, Gascuel O. FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program. *Mol Biol Evol*. 2015;32:2798–800.
130. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*. 2019;35:4453–5.
131. Letunic I, Bork P. Interactive tree of life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res*. 2024;gkae268.
132. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res*. 2017;45:e18–e18.
133. Ghabooli S, Shiganova TA, Briski E, Piraino S, Fuentes V, Thibault-Botha D, et al. Invasion pathway of the ctenophore *Mnemiopsis leidyi* in the Mediterranean Sea. *PLoS ONE*. 2013;8:e81067.
134. Christianson LM, Johnson SB, Schultz DT, Haddock SH. Hidden diversity of Ctenophora revealed by new mitochondrial COI primers and sequences. *Mol Ecol Resour*. 2022;22:283–94.
135. Ayad LA, Pissis SP. MARS: improving multiple circular sequence alignment using refined sequences. *BMC Genomics*. 2017;18:1–10.
136. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30:772–80.
137. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 2020;37:1530–4.
138. Presnell JS, Bubel M, Knowles T, Patry W, Browne WE. Multigenerational laboratory culture of pelagic ctenophores and CRISPR–Cas9 genome editing in the lobate *Mnemiopsis leidyi*. *Nat Protoc*. 2022;17:1868–900. <https://doi.org/10.1038/s41596-022-00702-w>.
139. Edgar A, Ponciano JM, Martindale MQ. Ctenophores are direct developers that reproduce continuously beginning very early after hatching. *Proc Natl Acad Sci USA*. 2022;119:e2122052119.
140. Rippe JP, Dixon G, Fuller ZL, Liao Y, Matz M. Environmental specialization and cryptic genetic divergence in two massive coral species from the Florida Keys Reef Tract. *Mol Ecol*. 2021;30:3468–84.
141. Quinlan AR, Hall IM. Bedtools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26:841–2.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.