Check for updates

Received: 24 June 2022 | Revised: 29 July 2023 | Accepted: 9 August 2023

DOI: 10.1111/mec.17118

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Evidence of introgression, ecological divergence and adaptation in *Asterias* sea stars

Melina Giakoumis[1,2,3] | Gonzalo E. Pinilla-Buitrago[1,2] | Lukas J. Musher[4] | John P. Wares[5] | Stuart J. E. Baird[6] | Michael J. Hickerson[1,2,3]

[1]The Graduate Center, The City University of New York, New York, New York City, USA

[2]The City College of New York, New York, New York City, USA

[3]The American Museum of Natural History, New York, New York City, USA

[4]The Academy of Natural Sciences of Drexel University, Pennsylvania, Philadelphia, USA

[5]Odum School of Ecology and Department of Genetics, University of Georgia, Georgia, Athens, USA

[6]Institute of Vertebrate Biology, Czech Academy of Sciences, Brno, Czechia

Correspondence
Melina Giakoumis, The American Museum of Natural History, New York City, NY, USA.
Email: mgiakoumis@amnh.org

Funding information
National Science Foundation, Grant/Award Number: DEB-1253710

Handling Editor: Michael M. Hansen

Abstract

Hybrid zones are important windows into the evolutionary dynamics of populations, revealing how processes like introgression and adaptation structure population genomic variation. Importantly, they are useful for understanding speciation and how species respond to their environments. Here, we investigate two closely related sea star species, *Asterias rubens* and *A. forbesi*, distributed along rocky European and North American coastlines of the North Atlantic, and use genome-wide molecular markers to infer the distribution of genomic variation within and between species in this group. Using genomic data and environmental niche modelling, we document hybridization occurring between northern New England and the southern Canadian Maritimes. We investigate the factors that maintain this hybrid zone, as well as the environmental variables that putatively drive selection within and between species. We find that the two species differ in their environmental niche breadth; *Asterias forbesi* displays a relatively narrow environmental niche while conversely, *A. rubens* has a wider niche breadth. Species distribution models accurately predict hybrids to occur within environmental niche overlap, thereby suggesting environmental selection plays an important role in the maintenance of the hybrid zone. Our results imply that the distribution of genomic variation in North Atlantic sea stars is influenced by the environment, which will be crucial to consider as the climate changes.

KEYWORDS
echinoderms, ecological genetics, environmental niche modelling, hybridization, marine invertebrates, molecular evolution

1 | INTRODUCTION

Adaptive ecological divergence between populations due to environmental variation is a key factor in the generation and maintenance of species-level biodiversity (Kawecki & Ebert, 2004; Nosil, 2012; Palumbi, 1994). This is exemplified in the nearshore marine environment, which often has steep gradients in temperature, salinity and other abiotic variables (Harley & Helmuth, 2003), which might promote local adaptation in species that occur there (Zardi et al., 2011). Although growing evidence in marine species is now challenging the long-held assumption that high dispersal ability can limit the potential for adaptive divergence (Sanford & Kelly, 2011; Bernatchez, 2016), examining the population genomics and environment of hybrid zones between closely related coastal marine species is particularly informative about how adaptation, gene flow and intrinsic reproductive barriers can influence the speciation process (Abbott et al., 2013; Barton & Hewitt, 1985, 1989; Chown et al., 2015).

Hybrid zones are typically studied in the context of cline theory, which describes spatial variation in traits or allele frequencies

 © 2023 John Wiley & Sons Ltd.

as spatial gradients that change across the region where hybridizing species meet (Barton et al., 1981; Harrison & Larson, 2014). There are many factors that are known to affect the shape, geographic extent and maintenance of hybrid zones; in marine systems, hybrid zones are likely to be dually shaped by physical seascape features (i.e. soft barriers, habitat availability and ocean currents) and selective processes whereby hybrid genotypes have higher or lower fitness than parental forms (El Ayari et al., 2019). Although research on marine hybrid zones often focuses on the effects of paleoenvironmental or physiographic barriers, such as ocean currents (Fenberg et al., 2014; Hare et al., 2005), many marine species exhibit transitional hybrid clines without the existence of an obvious barrier (Kelly & Palumbi, 2010; Wares, 2001). Endogenous selection against hybrids can arise via genomic conflict that leads to post-zygotic reproductive incompatibilities (Bronson et al., 2003; Dobzhansky, 1934; Svedin et al., 2008). This can lead to a tension hybrid zone where the selection against hybrids occurs within a narrow region of contact and a steep cline of hybridization (Barton & Hewitt, 1985). Such clines formed by endogenous selection can often coincide with biogeographic barriers or sharp environmental gradients that do not directly drive selection (i.e. exogenous selection), as would happen when environmental filtering prevents recruitment of hybrids within parental ranges that are climatically divergent from one another. In contrast to tension zones, cases of exogenous selection driven by environmentally divergent factors can lead to a mosaic hybrid zone (Howard et al., 1993). Often, mosaic hybrid zones have broad and varying areas of overlap in habitats that are intermediate between those of the parental forms, thereby leading to a patchwork of hybrid and parental forms often associated with gradual environmental or ecological transitions (Bierne et al., 2003). The latter may also occur when hybrid individuals show increased fitness outside of the parental niches, a condition referred to as transgressive segregation (Rieseberg et al., 1999). These two types of forces (i.e. endogenous and exogenous selection) underlying tension and mosaic hybrid zones can be difficult to disentangle and likely do not operate exclusively (Bierne et al., 2011).

Data on hybridizing species also give researchers the opportunity to study adaptation that might impact a species' resilience to climate change (Chunco, 2014; Taylor et al., 2015; Thomas et al., 2006). For instance, the expansion or contraction of hybrid zones can result from changes to the oceanic environment (Hilbish et al., 2012; Taylor et al., 2015). In some cases, range retractions due to habitat loss could lead to reductions in range overlap between closely related species and thus reduced gene flow (Thomas et al., 2006). In other cases, the likelihood of hybridization could increase if one or both hybridizing species expand their distributions (Garroway et al., 2010; Vallejo-Marín & Hiscock, 2016).

Sea stars (Asteroidea) are ecologically important keystone species in rocky intertidal communities; their population fluctuations can have rippling effects on whole communities (Lubchenco & Menge, 1978; MacKenzie Jr & Pikanowski, 1999; Paine, 1966). They are thus valuable models for studying organismal responses to the environment. Importantly, invertebrates inhabiting rocky intertidal communities have a high potential for local adaptation due to selection imposed by strong gradients and a complex mosaic of biotic and abiotic conditions (Sanford & Kelly, 2011). Thus, sea stars are also excellent models for understanding how the environment shapes genomic variation as well as how changes to environmental conditions could affect coastal communities on the whole.

Asterias rubens and *A. forbesi* are sister species that occur in the North Atlantic intertidal (Clark & Downey, 1992), and presently overlap in an area from northern New England to the Canadian Maritimes (Menge, 1979). Atlantic *Asterias* are estimated to have initially diverged into these two sister lineages about 2–3 mya across the North Atlantic (Wares, 2001; Wares & Cunningham, 2001). Initially, *A. forbesi* was restricted to the North American coast and *A. rubens* to Europe. Subsequently, trans-Atlantic colonization of *A. rubens* from Europe to North America is thought to have occurred soon after the last glacial maximum (Wares, 2001; Wares & Cunningham, 2001) leading to the present pattern of co-occurrence in northeastern North America (Harper et al., 2007).

Within this zone of overlap, the two *Asterias* species have largely overlapping spawning seasons (Menge 1986), and there is evidence that hybridization may be occurring. Although some authors have suggested that hybridization in *Asterias* is rare or absent (Schopf & Murphy, 1973; Worley & Franz, 1983), studies combining morphological and mitochondrial DNA (mtDNA) have suggested the potential for more widespread admixture (Harper & Hart, 2007). Moreover, an experimental study of *Asterias* showed that the two species are able to produce fertile offspring in a laboratory setting (Harper & Hart, 2005).

Important questions remain about the maintenance of this hybrid zone, including the width of the hybrid cline, the degree of introgression and the relative importance of the environmental niches of the two parental species in determining and predicting where hybrids are found. Among the studies that have documented evidence of hybridization in this system, geographic sampling was relatively sparse and inferences were based on only two mitochondrial gene fragments (Harper & Hart, 2007). Furthermore, although hybrids have been shown to be viable, many factors useful for predicting future hybrid zone movements are unknown, such as the environmental factors that may influence the geographic distribution of hybridization. Thus, determining the geographic and genomic extent of hybridization and finding the relative importance of the parental ecological niches on cline maintenance requires a combined approach that integrates population genomics data with spatially explicit environmental data.

To this end, we use RADseq data from 171 geo-referenced individuals across the ranges of both *A. rubens* and *A. forbesi* and the Pacific outgroup, *A. amurensis*, along with species distribution models (SDMs) to (1) characterize the spatial distribution of genomic variation in Atlantic *Asterias* and subsequently test the hypothesis that *A. rubens* and *A. forbesi* experience hybridization (2) quantify and compare the environmental space and level of environmental specialization of the two parental species and their hybrids, (3) test the alternative hypotheses that (a) the environmental overlap between

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17118 by University Of Georgia Libraries, Wiley Online Library on [24/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

the two species predicts the observed geographical distribution of hybridization as predicted under a mosaic or tension hybrid zone, or (b) that hybrids persist outside the parental environmental niches and (4) characterize the genome-environmental associations (GEA) in *Asterias* that could be driving locally adaptive divergence across their genomes and potentially limit hybridization beyond areas of environmental overlap.

2 | METHODS

2.1 | DNA extraction and sequencing

We visited 33 sites across the North Atlantic to collect samples across both species' ranges between 2015 and 2019 (Table 1). For each sample, we collected 15–30 tube feet or a portion of one leg and placed them either in RNAlater or flash frozen on liquid N_2 at −190°C. We also photographed and assigned georeferenced localities to each sample.

We extracted DNA using the DNEasy Blood & Tissue Kit or the MagAttract High Molecular-Weight DNA kit, both from Qiagen (Valencia). Library preparation and sequencing for 190 samples was completed by Floragenex using restriction site-associated DNA sequencing (RADseq) with established protocols (Lozier, 2014). For RADseq, genomes are digested with restriction enzymes which allows for a short DNA fragment flanking the restriction site to be sequenced. This results in thousands of small, independent fragments across the whole genome (Andrews et al., 2016; Miller et al., 2007). These samples were processed into a single RAD library using the enzyme SbfI-HF (NEB). After sonication to shear the DNA, initial size selection was ~300–500 bp and final size selection ~400–600 bp (including adapters). Size selection was performed using an agarose gel and the Qiagen MinElute Gel Extraction Kit, and the completed library was sequenced on an Illumina HiSeq 4000 sequencer. Sequencing included eight samples of *Asterias amurensis* collected from Tasmania as an outgroup. *Asterias amurensis* is found in the Pacific and is considered invasive to many regions in the Southern hemisphere (Byrne et al., 1997).

2.2 | RADseq read processing and filtering

We processed RadSeq reads using iPyrad v0.9.81 (Eaton & Overcast, 2020). We first demultiplexed sequence data, then mapped cleaned reads to the *Asterias rubens* reference genome (https://www.ncbi.nlm.nih.gov/assembly/GCA_902459465.3), and finally called single nucleotide polymorphisms (SNPs). iPyrad utilizes bwa for read-mapping (Li & Durbin, 2009). As parameters in the iPyrad pipeline, we required base calls with a minimum Phred quality score above 20, up to five low-quality bases per read, and a maximum fraction of 0.05 Ns and heterozygous sites in each consensus sequence, and a minimum of four samples per locus for output. For the clustering depth for assembly, we set a minimum of six reads per locus for statistical base-calling and a minimum depth of four for majority-rule base-calling. We set the clustering threshold to 0.85 similarity, which is the standard recommended by iPyrad. Samples with fewer than 150 recovered loci after filters were applied were removed from downstream analyses.

2.3 | Population structure and estimation of hybrid indices using genome polarization

To calculate summary statistics for each population, including π, Watterson's Theta (θ) and Tajima's D, we used DnaSP v6.12.03 (Rozas et al., 2017). We used STRUCTURE, a model-based clustering method, to assign individuals into genetic clusters based on their inferred ancestry (Pritchard et al., 2000). STRUCTURE assigns each individual probabilistically to one assumed population (out of K assumed populations) for each iteration. We tested *K* values 2–5 for the dataset that included the outgroup, *A. amurensis* and *K* values of 1 through 5 without the outgroup. For each value of *K*, we ran 10 independent replicates of STRUCTURE, with each replicate evaluating 100,000 MCMC generations after an initial burn-in of 25,000. We selected an appropriate *K* by choosing the highest $\triangle K$-value (Evanno et al., 2005) and by assessing the mean log probability of the model at each value of *K* (Figure S1). We then used CLUMPP, an algorithm for aligning multiple replicate analyses of the same dataset (Jakobsson & Rosenberg, 2007), to optimally align runs by permuting 105 replicates for each *K* value. As an additional method for visualizing and assessing population genomic structure, we also computed and visualized a Principal Components Analysis (PCA) both with and without the outgroup included, using the extended iPyrad analysis toolkit with sampling imputation, which imputes missing data by randomly sampling genotypes from predefined populations.

To characterize the potential hybrid cline between the *Asterias* sister species, we used genome polarization implemented in *diem* (Baird et al., 2022) to calculate each individual's genome-wide hybrid index (or genome admixture proportions) across the diploid genome blocks obtained from the RAD sequences. This genome painting by association, coupled with sufficient geographic sampling, allows for quantification of the width and barrier strength of a putative hybrid zone, and with sufficient genome sampling, the locations and size distribution of introgressing blocks. Introgression is identified by polarizing the labelling of bistate markers with respect to their association with the sides of a barrier, if one exists (Baird et al., 2022). Unlike the post-STRUCTURE Evanno procedure, this approach allows the $K=1$ versus $K=2$ cases to be distinguished, as no coherent polarization occurs if $K=1$. The *diem* polarization is applied, as with the RAD analysis, after mapping the reads to the *A. rubens* reference genome (https://www.ncbi.nlm.nih.gov/assembly/GCA_902459465.3). This has two advantages: First, it provides an ordering of the loci along the genome; second it provides a measure of relevance for all SNP variants within RAD blocks—the diagnostic index (DI). This allows an objective choice of which SNPs to use in downstream analyses. In

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17118 by University Of Georgia Libraries, Wiley Online Library on [24/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Sampling locations, coordinates and number of samples used in downstream analyses for *Asterias*.

Location and sample code	Geographic coordinates (latitude, longitude)	Number of samples	Hybrid index range
Le Croisic, France (LEC)	47.3014, −2.5211	7	0.0094–0.0174
Hoek van Holland, The Netherlands (HHND)	51.9850, 4.0927	1	0.0149
Scheveningen, Netherlands (SCH)	52.1026, 4.2588	8	0–0.0278
Drøbak Marine Field Station, Norway (DRNO)	59.6630, 10.6257	3	0.0176–0.0396
Skateraw near Edinburgh, UK (EDUK)	55.9726, −2.4207	3	0.0142–0.0213
Hvalfjörður, SW Iceland (ISL)	64.3927, −21.5566	11	0.0111–0.0232
Newman's Cove, Newfoundland, Canada (NCNL)	48.5835, −53.1961	2	0.0201
Amherst Cove, Newfoundland (ACNL)	48.5695, −53.2196	2	0.0231–0.0245
Melrose, Newfoundland, Canada (MLNL)	48.4850, −53.0632	2	0.0177–0.0215
New Melbourne, Newfoundland, Canada (NMNL)	48.0473, −53.1528	2	0.0201–0.0211
Sibleys Cove, Newfoundland, Canada (SCNL)	48.0436, −53.1054	2	0.0156–0.0188
Prince Edward Island, Canada (PEI)	46.1863, −63.1352	7	0.8657–0.8968
Halifax, Nova Scotia, Canada (HFNS)	44.6246, −63.5640	9	0.0236–0.9128
Lubec, Maine (LBME)	44.8627, −66.9834	1	0.8624
Eastport, Maine (EPME)	44.9065, −66.9843	2	0.0217–0.0244
Harriman Point Preserve, Brooklin, Maine (HPME)	44.2971, −68.5313	6	0.4118–0.9090
Belfast, Maine (BFME)	44.4286, −69.0041	4	0.8236–0.8940
Rockport, Maine (RPME)	−69.0669, 44.1793	8	0.0228–0.8995
Owl's Head, Maine (OHME)	44.0927, −69.0452	10	0.0204–0.8986
Kennebunkport, Maine (KBME)	43.3465, −70.4742	2	0.0216–0.8943
Portsmouth, New Hampshire (NH)	43.0473, −70.7154	6	0.0208–0.4140
Folly's Cove, Rockport, Massachusetts (FCMA)	42.6859, −70.6428	8	0.8964–0.9178
Gloucester, Massachusetts (GLMA)	42.6200, −70.6240	3	0.4214–0.8882
Nahant, Massachusetts (NAMA)	42.4197, −70.9065	10	0.8959–0.9149
Marshfield, Massachusetts (MRMA)	42.1190, −70.6710	1	0.8976
Brant Rock, Marshfield, Massachusetts (BRMA)	42.0952, −70.6472	9	0.8873–0.9157
Provincetown, Cape Cod, Massachusetts (PTMA)	42.0360, −70.1963	6	0.8944–0.9073
Chatham Pier, Cape Cod, Massachusetts (CCMA)	41.6804, −69.9477	7	0.85–0.9884
Wood's Hole, Massachusetts (WHMA)	41.5275, −70.6794	9	0.9015–0.9411
Ponquogue Bridge, Long Island, New York (PBNY)	40.8448, −72.5005	4	0.9136–0.9393
Pine Knoll, Bogue Sound, North Carolina (PKNC)	34.7223, −76.7573	1	0.9578
Charleston, South Carolina (CHSC)	32.7510, −79.8970	4	0.9556–0.9585
Tasmania (TAS)	−43.0044, 147.3244	8	–

Note: The range of hybrid indices of samples from each location is also included. A hybrid index range between 0–0.124 is considered *A. rubens*, a hybrid index range between 0.876–1 is considered *A. forbesi*, and a hybrid index range between 0.125 and 0.875 is considered a true hybrid.

particular, RAD blocks that might otherwise be rejected due to high coverage (signalling over-merging) may still contain highly barrier-diagnostic SNP states that need not be discarded for barrier analyses. Mapping to a reference itself reduces over-merging without introducing reference bias such that the reference state is ignored when calling the mapped data (Eaton & Overcast, 2020). For *diem* analysis, we used mpileup in samtools (Li et al., 2009) to compile mapped RAD reads which was then converted to matrix, or sync, format using Popoolation2 (Kofler et al., 2011), and the two most common states at each variant site were identified. Importantly, the commonest means occurring most frequently across individuals, not 'with the highest total pileup count' (which can be biased by individual coverage variation). All variant sites were then encoded with 0 and 2 representing the homozygotes of the two commonest SNP states, 1 representing their heterozygote, and U for diplotypes Un-encodable using the 2 most common states (this naturally included the missing data state). The Mathematica implementation of the *diem* algorithm was then used to polarize the encoded data, and all sites were filtered post-*diem* on the per-site DI output. This step discards both data-poor sites (with few reads mapped) and sites with variation orthogonal to the barrier such as shared polymorphisms.

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17118 by University Of Georgia Libraries, Wiley Online Library on [24/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

For much of the downstream analyses that required assigning genotypes to discrete categories, we designated individuals into one of three subsets (*A. rubens*, *A. forbesi* or hybrids) based on their hybrid index (HI) as calculated by *diem*. Any HI value between 0.125 and 0.875 was considered a hybrid (see Table S1 for HI values for each sample and Table 1 for the range in hybrid index values per location), while individuals with HI between 0 and 0.124 were designated as *A. rubens* and those with HI between 0.876 and 1 were designated as *A. forbesi*.

2.4 | Spatially explicit population connectivity

We assessed samples for spatial patterns of migration and diversity with the Estimated Effective Migration Surfaces (EEMS) program (Petkova et al., 2016), which uses Bayesian inference to model the relationship between geographic and genetic distance. EEMS provides an estimate of effective migration by assessing the areas in which genetic similarity decays or increases more quickly than expected from a null model. We ran EEMS on all samples of *A. rubens* and *A. forbesi*. We calculated genetic distance using the iPyrad toolkit, which samples one randomly selected variable site per locus to calculate pairwise genetic distance across samples based on the proportion of shared SNPs. The polygon for the entire *Asterias* range contained 1500 demes, or subpopulations, while the polygon for the North American portion of the range contained 1000 demes to account for a smaller area. We performed three runs of the Markov chain Monte Carlo (MCMC) for each of these polygons for 2,000,000 iterations, thinning every 9999 iterations, and excluding a burn-in of 1000,000 iterations. We assessed MCMC convergence by plotting the log posterior probabilities of each run and comparing results among the runs for each polygon.

2.5 | Species distribution models

We built species distribution models (SDMs) for *A. rubens* and *A. forbesi* using WALLACE2 (Kass et al., 2023), which implements the Maximum Entropy (Maxent) machine-learning method with the maxnet R package (Phillips et al., 2006, 2017) to model the distribution of each species using presence-only data. These SDMs build on previous work predicting the distribution of *Asterias* using SDM techniques that used a smaller set of environmental variables that included both the terrestrial and marine domain (Waltari & Hickerson, 2013), but which lacked many relevant marine environmental layers available today, as well as the genetic context of hybridization in *Asterias*.

We obtained occurrence data from each species from the Global Biodiversity Information Facility (GBIF.org, 2021a; GBIF.org, 2021b) and our own additional sampling locations (Table 1). Using the STRUCTURE analysis as a guide, we conservatively chose only areas with no evidence of admixture for each species to avoid mis-identified occurrence points. Therefore, the occurrences for *A. rubens* only included presence data from Newfoundland and Europe, and the occurrences for *A. forbesi* only included presence data from Cape Cod, Massachusetts and further South. We filtered *A. rubens* and *A. forbesi* occurrence data from GBIF to include only georeferenced points from 1981 to 2021, to capture a robust number of data points while also partially matching the temporal range of our contemporary environmental layers from Bio-Oracle 2.0 (Assis et al., 2018), which range from 2000 to 2014. For *A. rubens*, we further filtered our data to include only points with photographs for taxonomic verification. We did not perform this filtering for *A. forbesi* since the southern range of this species occurs in an area with fewer similar-looking sea stars, which allows for more confidence in species identification. With this filtering strategy, we retained 490 occurrence points for *A. rubens* and 105 occurrence points for *A. forbesi*.

We spatially thinned occurrence points to 10 km using spThin (Aiello-Lammens et al., 2015), which matches the scale of the environmental variables (5 arcmin) and which is shown to improve species distribution models by reducing sampling bias (Boria et al., 2014). We downloaded marine environmental variables from Bio-Oracle 2.0 (Assis et al., 2018). Maxent has no issues with collinearity because it deals with it internally in model training, and the exclusion of highly correlated variables does not have a real impact on Maxent model performance (Feng et al., 2019). We therefore chose which environmental variables to use for model construction (Table S2) based on biological relevance and used the mean value of those variables. For temperature we also used the long-term minimum, long-term maximum, and range.

We calibrated the models within a study region composed of a 1.25 point buffer around each occurrence point. For model evaluation, we spatially partitioned the points using the block method (Muscarella et al., 2014; Wenger & Olden, 2012). To account for model complexity, we explored different combinations of feature classes (linear [L], quadratic [Q], hinge [H], product [P]) to obtain several types of modelled responses from explanatory environmental variables: L, LQ, H, LQH, LQHP. We also used regularization multipliers, which penalize model complexity, between one and five separated by steps of 0.5. We chose optimal models for each species based on a lowest Akaike Information Criterion (AIC), and checked for a low 10th Percentile training omission rate and high validation Area Under the Curve (AUC) as a second measure of fit. Finally, we then transferred the models for both species into the area of confirmed genomic admixture, to determine overlapping and differential suitability in this area of hybridization. We visualized each model, as well as the transfer into the area of admixture, using the 10th percentile training presence. This is a binary visualization method to identify the areas where each species is most likely to occur and have the highest levels of environmental suitability.

To estimate the environmental niche overlap between the two *Asterias* species, we employed the ellipsenm R package (https://github.com/marlonecobos/ellipsenm). This package represents the species' environmental niche as a simple convex shape (Jiménez et al., 2019) that contains the environmental conditions in which each species was observed. Using the same occurrence data that were used to build the

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17118 by University Of Georgia Libraries, Wiley Online Library on [24/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

SDMs, we generated a minimum-volume ellipsoid (MVE) for each species that included 95% of the occurrences. For this analysis, instead of using the 18 environmental variables in their raw forms, we reduced dimensionally by obtaining the first three principal components with the *kuenm* package (Cobos et al., 2019). Niche breadth corresponded to the volume of each species' ellipsoid. We then performed a niche overlap test using the background points of the union between niche ellipsoids. A rejected null hypothesis of the niche overlap test (p-value < .05) means that the observed overlap value between ellipsoids is lower than the 95% values obtained from random ellipsoids generated by background points (1000 replicates). Finally, environmental conditions of the sampling localities from which we have genomic data (including the hybrid population with evidence of admixture) were plotted within the niche ellipsoids of species.

2.6 | Genome-environment associations and detection of loci under environmentally divergent selection

We took two complementary genome-environment associations (GEAs) approaches to uncover how the different environmental niches of the two species could be leading to adaptive ecological divergence across parts of their genomes and leading to a hybrid cline maintained by environmental barriers (i.e. exogenous selection). First, we performed a redundancy analysis (RDA), which is a multivariate method for genome-environment associations. Second, we flagged regions of the genome associated with F_{ST} outliers, and then used Latent Factor Mixed Models (LFMM) on our set of environmental variables to uncover which ones could be driving divergent selection at these loci while explicitly accounting for population structure.

The first approach, RDA, uses linear regression to detect candidate adaptive loci that are associated with our chosen 18 environmental variables. RDA, as compared to univariate methods, are highly effective in detecting multilocus selection since they consider how genomic markers covary in response to environmental predictors (Rellstab et al., 2015). RDA also uses constrained ordination, and has been found to maintain a balance of true and false positive rates and is robust across sampling designs, demographics and selection levels (Capblancq & Forester, 2021; Forester et al., 2018). This method performs a PCA on allele frequencies while constraining the PCA axes as linear combinations of uncorrelated environmental predictor variables. We assessed the collinearity of the marine environmental variables from Bio-Oracle2.0 and removed any variables with correlations higher than 70%, and re-assessed this collinearity by examining variance inflation factors. We performed a global redundancy analysis (RDA) using the rda function in the R package vegan (Oksanen et al., 2020) on the entire dataset to detect candidate loci that were correlated with environmental variation. We also ran an RDA on each of the two species independently and on hybrids, to assess the influence on each group separately. While it is possible to perform a partial RDA which accounts for population structure, it has been shown that global RDAs perform better (Forester et al., 2018; Xuereb et al., 2018), so global RDAs utilizing the entire dataset were used for this analysis.

For the second approach, we obtained F_{ST} outliers that signal reduced diversity within genome regions (Cruickshank & Hahn, 2014) or alternatively excessive divergence between the two species that potentially include regions of the genome that have reduced gene flow due to selection driven by adaptation to local environmental conditions (Gosset & Bierne, 2013). For this, we used sNMF in the R package LEA (Frichot & François, 2015), which uses a clustering method to find differential levels of ancestry in individual samples. We used an extended function of this method to find significant F_{ST} outlier SNPs between *Asterias rubens* and *A. forbesi* that could indicate divergent selection. We used a K of 3 from the sNMF analysis in LEA to account for inherent population genetic structure in the data. The best run among the 100 sNMF repetitions was chosen based on the minimum cross entropy score and subsequently used to infer population differentiation statistics across the entire *A. rubens* and *A. forbesi* dataset, complete with p-values for all loci. This method uses a genomic inflation factor (λ) to rescale the chi-squared statistics and associated p values for all loci, and after exploring different values as recommended by the creators of the software, we used a $\lambda = 2.5$ and a false discovery rate of 0.01.

To understand which of these F_{ST} outliers were associated with specific environmental variables, we then conducted GEA tests using Latent Factor Mixed Models (LFMM) implemented in LEA (Frichot & François, 2015). This was done on each of our 18 environmental variables separately, such that SNPs that are flagged as both F_{ST} outliers and associated with environmental variation suggest putative islands of genomic differentiation driven by local adaptation to species-specific conditions.

The 18 environmental variables were the same ones used to build SDMs (Table S2). LFMM uses MCMC to model the effects of environmental variables and population structure on the frequency of alleles across populations. SNPs flagged in these analyses reflect correlations between genomic variation and environmental variation, with SNPs that are highly correlated with environmental variation (e.g. the environment is highly predictive of genomic variants) flagged as significant by LFMM. Generally, environmental variables with a high level of correlated SNPs can be considered important for shaping genomic variation. This allows detection of loci that are associated with environment variation across the range of both species, while removing the influence of population structure.

LFMM has been shown to have low rates of false positives and negatives (Frichot et al., 2013) and is robust to the effects of population structure and sampling (Rellstab et al., 2015). To control for false discoveries, we corrected p-values with empirical genomic inflation factors (λ) of 10^{-5} after inspection of histograms, as recommended by program developers and similar to previous studies using this program (Prates et al., 2018). To reduce the effects of missing data, we imputed haplotype-based sNMF ancestry coefficients using the "mode" method as recommended in the LEA manual, wherein the most likely genotype is used in matrix completion. A list of candidate SNPs

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17118 by University Of Georgia Libraries, Wiley Online Library on [24/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

was generated using the Benjamini–Hochberg algorithm (Benjamini & Hochberg, 1995) and assuming a false discovery rate (FDR) of 0.01.

3 | RESULTS

3.1 | RADseq processing and filtering

After filtering and quality control, including removal of PCR duplicates, our RADseq dataset included a total of 171 samples containing 31,604 loci. The number of loci per individual ranged from 172 to 12,360 with a mean ± SD = 5307.31 ± 1945.96, and we obtained a SNP matrix in which 73.03% of samples were missing sites.

Average number of nucleotide differences between sampled individuals, or π, is similar between the two species, estimated at 0.0842 for *A. rubens* and 0.0856 for *A. forbesi*. Watterson's θ, another estimate of diversity based on the number of segregating sites found in a sample, is 0.1526 in *A. rubens* and 0.0764 in *A. forbesi*, suggesting that *A. rubens*, with its large range across the Atlantic Ocean, has higher levels of genetic diversity than the range-restricted *A. forbesi*. The Tajima's D summary statistic was −0.8584 for *Asterias rubens* and −0.0001 for *A. forbesi*, indicative of more low-frequency polymorphisms in *A. rubens* than in *A. forbesi*, potentially suggesting stronger population expansion in *A. rubens* that could be a consequence of recent colonization from Europe to North America. Additional summary statistics, as well as calculations for the small set of *A. amurensis* and hybrid samples, can be found in Table S3.

3.2 | Population structure

Both PCA and STRUCTURE are suggestive of two distinct population clusters with a zone of admixture along the coastline of New England and Nova Scotia (Figures 1 and 2a,b), a general result that was also consistent with the distribution of hybrid indices obtained from the *diem* method (Figures S2 and S4). The STRUCTURE results were consistent with a three-population model ($K = 3$) when run with the Pacific outgroup species *A. amurensis*, and though the Evanno

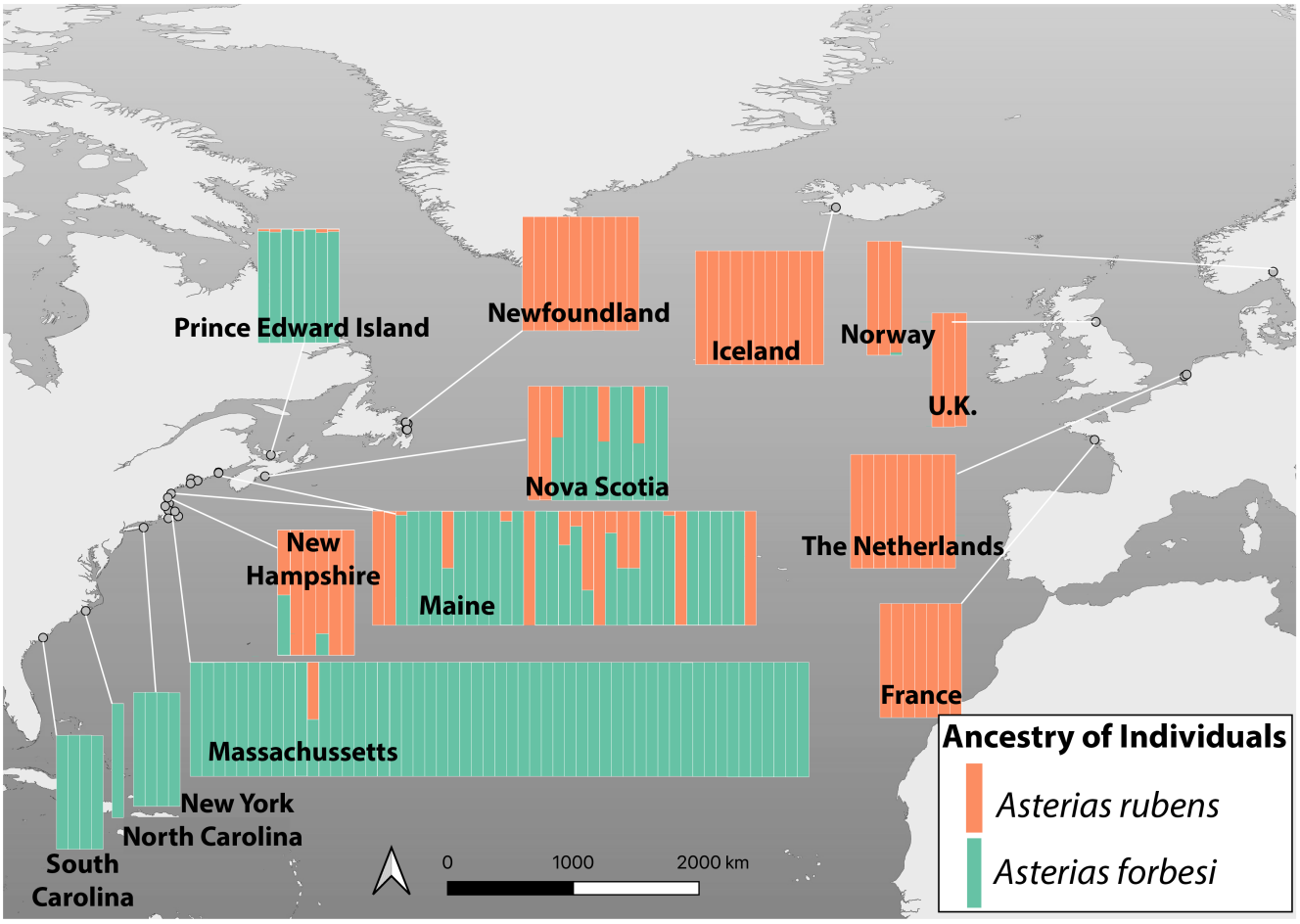

FIGURE 1 Estimates of population structure as inferred from STRUCTURE, overlaid across the *Asterias* distribution in the North Atlantic Ocean, and showing our sampling locations. Orange denotes ancestry from *A. rubens* and green denotes ancestry from *A. forbesi*. Individuals showing evidence of admixture (or, hybrids) are those with both colours. Data used in this visualization were run without the outgroup (*A. amurensis*), with a *K* of 2. Each vertical bar represents a single individual, and each group of bars is a sampling region (see Table S2 for a full list of locations and their coordinates).

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17118 by University Of Georgia Libraries, Wiley Online Library on [24/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

method is unable to distinguish between a 1 versus 2 population model for STRUCTURE data, when viewed in conjunction with the *diem* results we were able to reasonably infer that the two-population model ($K=2$) is the likeliest when analysed without the outgroup (Figure 1; Figures S2 and S5). In both STRUCTURE and the PCA, *A. amurensis* clusters on its own, with no apparent admixture between this Pacific species and the two Atlantic members of the genus (Figures 1, 2a,b and Figure S5). The PCAs show the two Atlantic *Asterias* sister species clustering as a continuous geographic cline along the first PC when the outgroup is present (Figure 2a). Without the outgroup, they cluster along two continuous clines, with the samples from Maine, Massachusetts and New Hampshire spreading between two clusters that are associated with the two species (Figure 2b). This pattern is indicative of a north–south hybrid cline in New England, centred in Maine. Similarly in the STRUCTURE plot, *Asterias rubens* and *A. forbesi* show evidence of admixture beginning in northern Massachusetts through Nova Scotia and Prince Edward Island. *Asterias rubens* individuals from Newfoundland cluster entirely with European individuals (Figure 1).

3.3 | Geographic cline analysis and hybrid indices

The hybrid indices obtained from *diem* are largely consistent with our STRUCTURE analysis with regards to the geographic patterns of admixture proportions and the location of the hybrid zone spanning from Cape Cod to Nova Scotia albeit excluding the Prince Edward Island samples which were all strongly *A. forbesi* (HI > 0.875; Figure 3; Table S1; Figures S2 and S4). These geographic trends of hybrid proportions were robust to coverage bias imposed by a reference genome from one of the two sister species (*A. rubens*; Figures S2 and S4). Consistent with either exogenous or endogenous selection being insufficiently strong to prevent F1 hybrids from back-crossing or mating with other F1 hybrid individuals, *diem* uncovered a large number of individual genotypes with HI values deviating from 0.5 while individuals with HI values >0.125 and <0.875 suggested a unimodal distribution (Table S1; Figure 3 and Figure S6). The plotting of coastal-distance and hybrid indices suggests a strong cline in the Gulf of Maine, along with some heterogeneity (Figure 3 and Figure S6). For instance, when plotted against coastal distance the hybrid indices

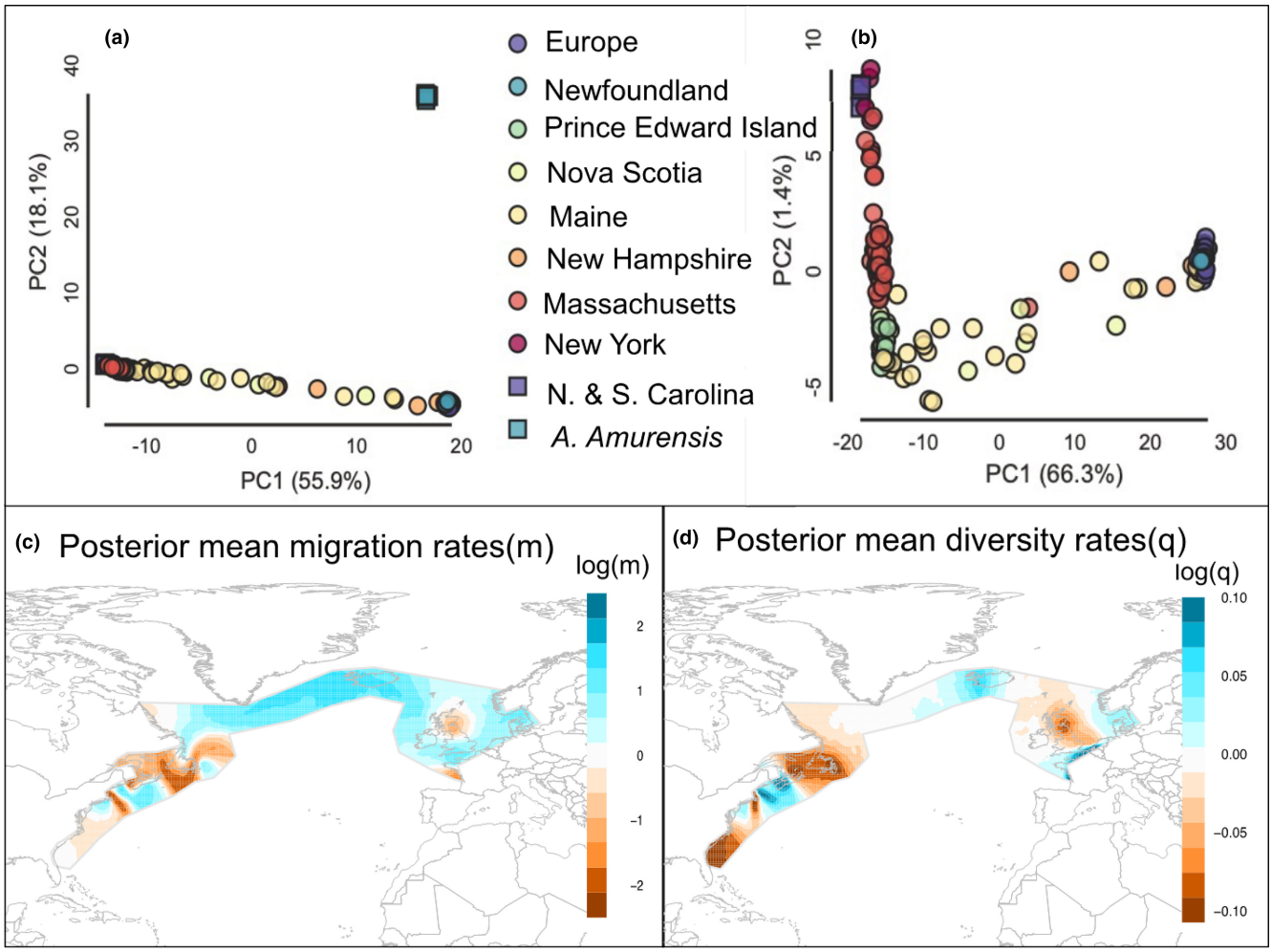

FIGURE 2 Principal components analysis (PCA) plots for RADseq data, including the outgroup *Asterias amurensis* (a) and excluding it (b). PCAs were calculated and visualized in the ipyrad-analysis PCA toolkit, using a sampling imputation method. Panels c and d show visualized results from EEMS. The mean migration rates (c) and diversity rates (d) are presented on a log scale. This includes all *A. rubens* and *A. forbesi* samples.

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17118 by University Of Georgia Libraries, Wiley Online Library on [24/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

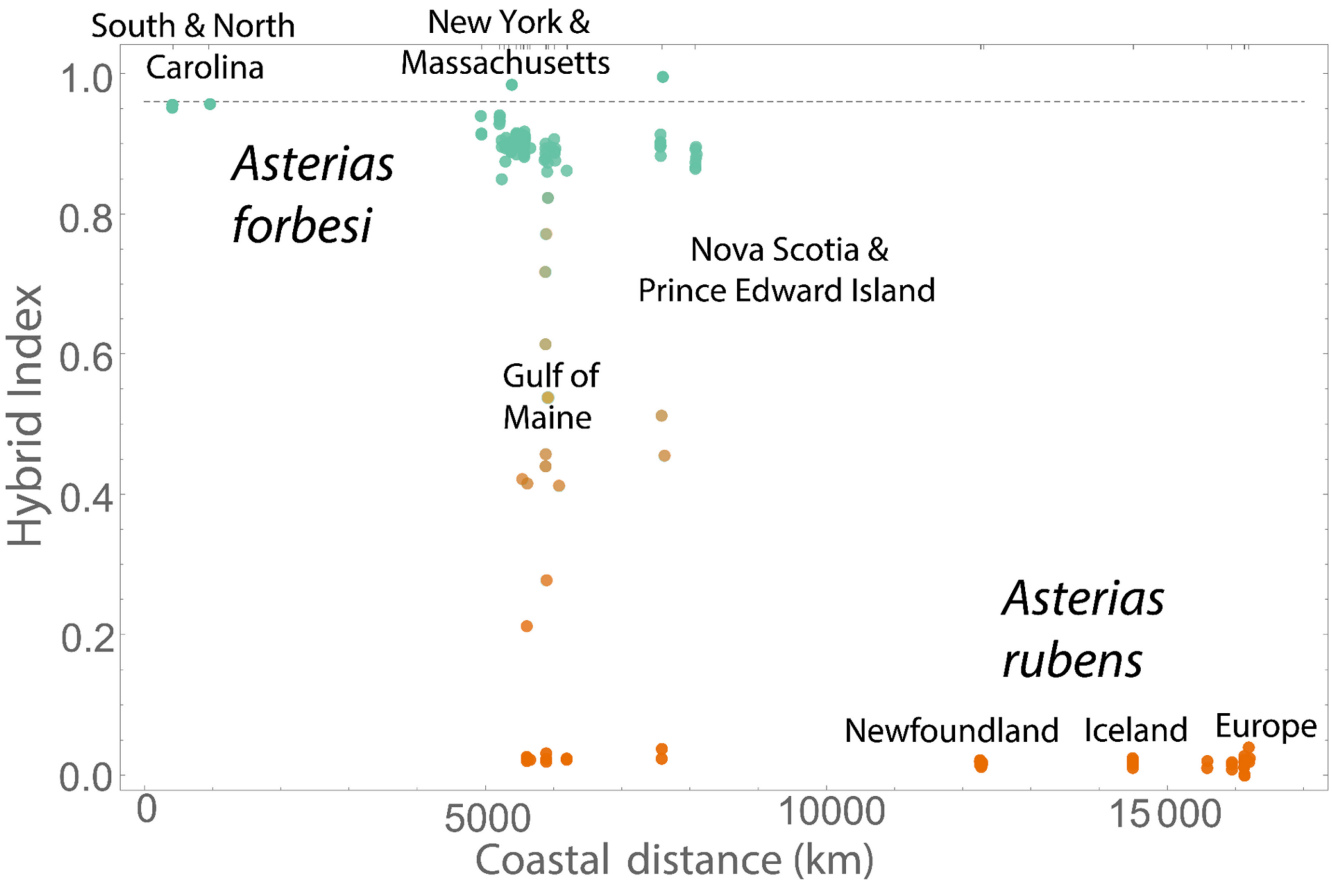

FIGURE 3 *diem* results showing the hybrid indices of North American *Asterias* across geographical space, with coastal distance on the *x*-axis and hybrid index on the *y*-axis. In orange are samples classified as *A. rubens* (HI < 0.125), and in green are samples classified *A. forbesi* (HI > 0.875). Intermediate samples are hybrids (HI between 0.125 and 0.875) and are coloured on a corresponding gradient between green and orange.

change as a step function between Newfoundland (*A. rubens*) and Prince Edward Island (*A. forbesi*), localities with the most hybrid genomes are found much further south-west, on the coast of New England (Figure 3 and Figure S4). The area between Prince Edward Island and Newfoundland, however, has not been sampled, and the overall pattern in this area can therefore not be verified with genomic data.

3.4 | Effective migration in the North Atlantic

The EEMS analysis indicated gene flow across the North Atlantic with an area of reduced gene flow and elevated isolation between Newfoundland *A. rubens* samples and all other samples in North America, a region associated with the Cabot Strait and Laurentian Channel (Figure 2c,d). In addition to this zone of isolation, there was a break in gene flow roughly delineated at Cape Cod separating admixed samples in the Gulf of Maine and *A. forbesi* samples to the south of the Cape. Genetic diversity patterns recovered from EEMS reflect the pattern of hybridization inferred by STRUCTURE and *diem* (Figure 1; Figure S2) with high diversity in the Gulf of Maine hybrid zone and low diversity in areas that have low levels of admixture between species both in North America and in Europe. The combination of gene flow and high diversity in the zone of contact are consistent with our hypothesis of hybridization on a portion of the east coast of North America.

3.5 | Species distribution models

The best-fitting SDMs for each species differed in regularization multiplier, feature class and environmental variables used. For *A. forbesi*, we selected a hinge model with a regularization multiplier of 2. This model had an AIC of 730.221, as well as a training AUC of 0.905, average testing AUC value of 0.859 and a testing omission rate of 0.162. For *A. rubens*, we selected an LQH model with a regularization multiplier of 2.5. This model had an AIC of 4276.328, a training AUC of 0.841, an average testing AUC of 0.824, and a testing omission rate of 0.143. We used each of the two species-specific SDMs to transfer their respective potential range distribution across the North American coastline (Figure 4a). These range transfers were consistent with our STRUCTURE analyses and ground-truth localities in which *Asterias* were found across several field seasons. Variables used by Maxent to build each model are listed in Table S2, along with the percent contribution to the model.

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17118 by University Of Georgia Libraries, Wiley Online Library on [24/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

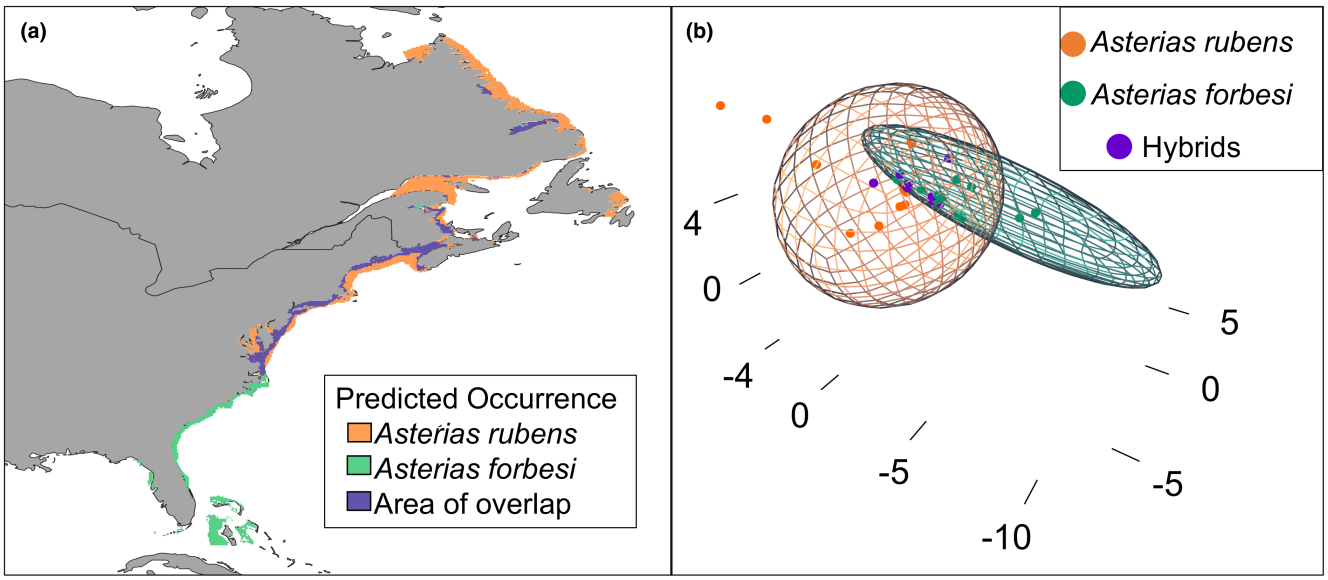

FIGURE 4 (a) Species distribution models, built using Maxent, for *A. rubens* (orange), *A. forbesi* (green), and area of overlap (purple). These models are thresholded with the 10 percentile training presence with the area of overlap depicting areas where both species' 10th percentile training presence are projected. (b) Climate niche breadth, shown as minimum volume envelopes (MVE), or ellipsoids, for each species. Ellipsoids are shown in orange for *A. rubens* and green for *A. forbesi* along with their occurrence points. Points at which hybrids occur are plotted in purple.

In the area of admixture delineated by the STRUCTURE and *diem* methods (Figure 1 and Figure S2), the models show high suitability for both *A. forbesi* and *A. rubens*, with a wide projected area of overlap between Delaware and Nova Scotia. The niche overlap test performed on the two respective models was significant ($p = .011$; Figure S3), indicating that the two species' environmentally defined niches are different. Analysis of niche breadth and overlap (Figure 4; Figure S3) also showed approximately 9.02% overlap in the environmental space of the *A. rubens* and the *A. forbesi* niches. The volume of the environmental space ellipsoid for *A. forbesi* (210.6327), is less than half of the volume of the *A. rubens* ellipsoid (459.6536). *Asterias rubens*, with a much wider distribution across Europe in addition to Newfoundland, shows a correspondingly larger range of environmental tolerance and overall larger niche breadth. However, *A. forbesi* shows a much smaller range of environmental values.

The hybrids recovered from *diem* ($0.125 < HI < 0.875$) were almost entirely within the area of environmental niche overlap predicted by our SDM and the associated analysis of ecological breadths of the two species (Figure 4). This is consistent with exogenous selection having an impact on the geographic distribution of the observed hybrid individuals, if we assume that effective dispersal is large and hybridization is not recent. This is highlighted in the Prince Edward Island samples, which genetically ($HI < 0.125$) most resemble *A. forbesi* from the Carolinas, rather than samples found in Newfoundland, which instead resemble European *A. rubens* ($HI > 0.875$). Strikingly, Prince Edward Island lies in the environmental ellipse of *A. forbesi* rather than *A. rubens* (Figure 4b).

3.6 | Genome-environment associations and detection of loci under environmentally divergent selection

The F_{ST} outlier procedure implemented in LEA uncovered 734 out of 25,148 loci to be significantly differentiated between *A. rubens* and *A. forbesi* according to our chosen threshold values, including a false discovery rate of 0.01. Differentiated SNPs and the histogram of *p*-values for all SNPs are visualized in Figure S7. Given our strict parameters for testing, these figures show under 2.9% of tested SNPs to be highly differentiated between *A. rubens* and *A. forbesi*.

When performing the RDA, we began by investigating correlations between the 18 Bio-Oracle 2.0 variables used in our SDM analysis. Several of the variables were found to be correlated (>70%), and were therefore removed (Table 2; Figure S8), leaving one correlated variable per set. All of the remaining predictor variables had a variance inflation factor < 10 (Table 2), which is satisfactorily low, and indicates no issues of collinearity in the predictors for the RDA model (Capblancq & Forester, 2021). The RDA with all 25,148 SNPs, including all sampling locations and samples, was globally significant ($p = .001$), explaining 11.6% of the variance (with an adjusted R^2 of .058) thereby demonstrating that the divergent environmental niches we uncover in our SDM analysis could be driving divergent selection across parts of the genome (Figure 5). The RDA for *A. forbesi* was significant (*p*-value of .045, R^2 of .132, adjusted R^2 of .0197), but was not significant for *A. rubens* (R^2 of .198 and an adjusted R^2 of .016 and a *p*-value of .095) or for hybrids (R^2 of .520 and an adjusted R^2 of .0875 and a *p*-value of .152). The results for these data subsets are available in Table S4. The RDA found 1128 SNPs significantly associated with environmental variation. RDA ordination

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17118 by University Of Georgia Libraries, Wiley Online Library on [24/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Genome-environment associations tested with RDA in the R package vegan.

Environmental layer	Correlated variables	Variance inflation factor (VIF)	Number of associated loci
Mean calcite (CAL)	–	4.4009	0
Mean currents velocity (CRV)	–	1.9673	10 (0.034%)
Mean iron (IRO)	–	3.6007	0
Mean nitrate (NIT)	PHY, CHL, PRO	6.8349	232 (0.92%)
Mean photosynthetic available radiation (RAD)	–	4.3541	10 (0.034%)
Mean pH (PH)	–	4.0644	201 (0.80%)
Mean salinity (SAL)	–	9.6952	1 (0.004%)
Mean silicate (SIL)	–	2.2232	2 (0.008%)
Mean temperature (TAV)	TMAX, TMIN, OXY, CLU, PHO	4.2214	48 (0.191%)
Temperature range (TRA)	–	5.0012	18 (0.072%)

Note: This table lists the number of loci most associated with each environmental variable, as well as the variables that are highly correlated with the tested variable. In total, 25,148 loci were tested. For definitions of the correlated environmental variable abbreviations, see Table S4.

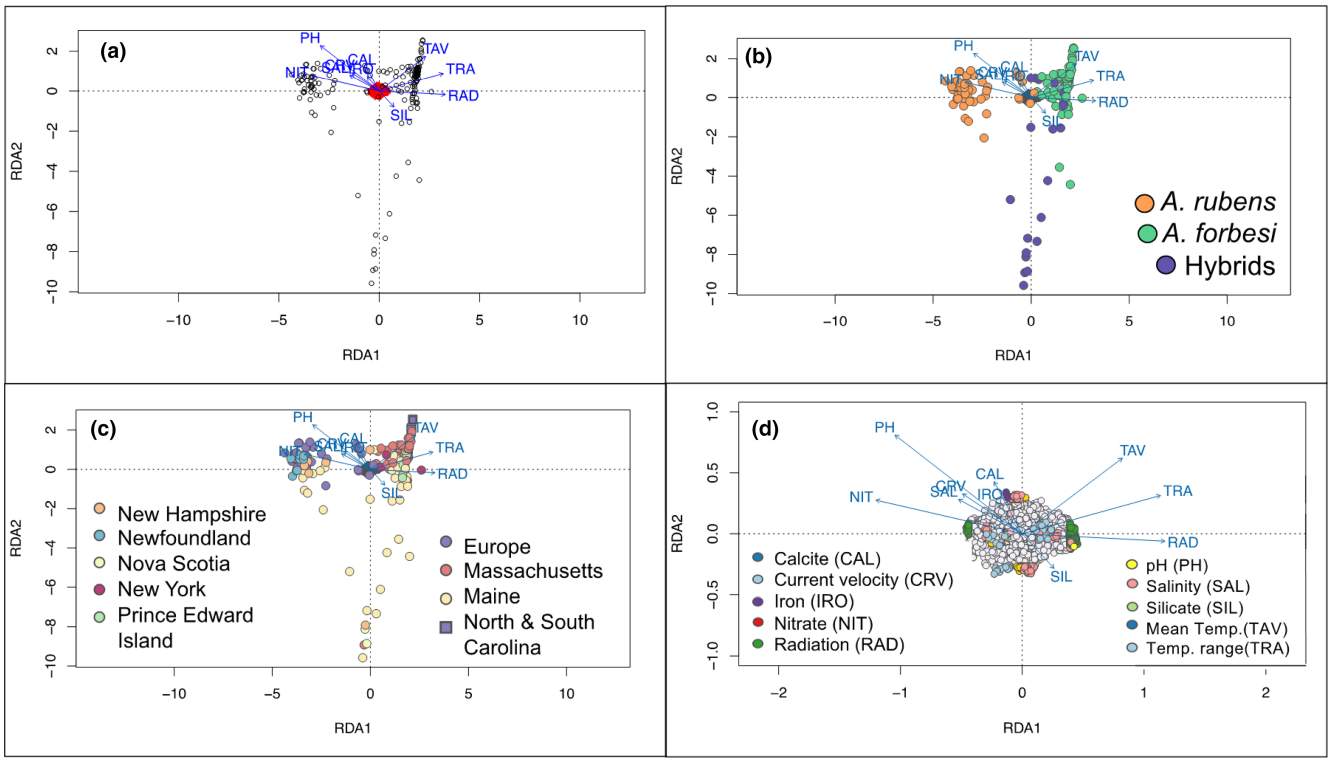

FIGURE 5 RDAs for the global dataset. (a) The global RDA, with genomic variation shaped by the environmental variables (see Table 2 for abbreviations of these variables). (b) The RDA with points coloured by population (R = *Asterias rubens*, F = *Asterias forbesi*, H = hybrids). (c) The RDA with points coloured by sampling location. (d) The RDA by locus. Loci not significantly associated with environmental variation are shown in grey, and loci associated with environmental variables are coloured by those variables.

plots have been coloured by locus, location and population (Figure 5) and the relative proportions of environmental variants most associated with SNPs, according to the RDA, is available in Table 2 and Table S4.

Genotype-environment association tests were run in LFMM for 18 Bio-Oracle 2.0 environmental variables used in our SDM analysis and for all 25,148 recovered and imputed SNPs for the (1) entire dataset, and subsequently, (2) *A. rubens*, (3) *A. forbesi* and (4) hybrids separately. The total number of loci significantly associated with each variable for each species is shown in Table S5. Generally, *A. forbesi* had more SNPs that were significantly associated with environmental variables than *A. rubens*, which had relatively few SNPs flagged as correlative to the environment, and hybrid samples occupy a distinct space in the RDA as well. The global analyses from both the RDA and the LFMM show significant correlations with

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17118 by University Of Georgia Libraries, Wiley Online Library on [24/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

several of the same environmental variables, highlighting their putative role in contributing to genomic variation.

We were also interested in understanding which of these environmentally linked variants uncovered with LFMM were *also* flagged in our analysis as being highly differentiated between species in our F_{ST} outlier tests; if a SNP is both highly differentiated between the two species *and* highly correlated with environmental variation, it could suggest that this SNP resides in a genomic region that has undergone ecologically driven isolation between the two species. Of the 734 F_{ST} outlier SNPs, 143 were also associated with environmental variation in *Asterias* as a whole in our pooled analysis and are visualized in Figure S9. Some F_{ST} outlier SNPs were associated with more than one environmental variable; the numbers provided are the number of unique SNPs that are differentiated across species and are associated with the environment. Reference maps showing the distribution of variation in the environmental variables are provided in Figures S11A–R.

4 | DISCUSSION

4.1 | *Asterias* hybridization is extensive

Although it has long been suggested that *A. forbesi* and *A. rubens* hybridize in the wild as well as in the laboratory (Harper & Hart, 2005), this study provides the first genome-wide evidence of introgression in these taxa across a broad contact zone along the northeastern North American coastline. Our study shows that hybrids likely survive in nature to produce late generation backcrosses, and we see no evidence of hybridization beyond this region. Indeed, the edges of the zone of hybridization are abrupt (Figure 3) despite evidence suggesting high larval dispersal potential (Loosanoff 1964; Menge 1986) and a history of long-distance colonization (Ilves et al., 2010; Ingolfsson, 1992; Wares & Cunningham, 2001).

Rather than forming a distinct region that bisects pure populations of both parental species, the observed hybrid zone in the Gulf of Maine disjoins the distribution of pure *A. forbesi* populations into one main group of sampling localities along the Mid-Atlantic Bight (MAB), as well as a lone sampling locality from the Prince Edward Island (PEI) 1000s of kilometres to the north (Figures 1 and 3). The genetic clustering of the PEI samples with the MAB samples could reflect contemporary larval dispersal coupled with environmental filtering that may be influencing the distribution of both species as well as areas of persistent hybridization. Indeed, environmental filtering of larval settlement has been observed in *Asterias* (Casties et al., 2015; Sameoto & Metaxas, 2008).

4.2 | Selection leading to a mosaic or tension hybrid zone

In mosaic hybrid zones, selection is environmentally driven; it can favour hybrids within the area of environmental overlap, there can be selection for or against hybrids outside areas of environmental niche overlap between the parental species (Harrison & Larson, 2014; Rieseberg et al., 1999). Mosaic hybrid zones are typically broader in area than those predicted under a "tension zone" (Barton & Hewitt, 1985) with the former often consisting of a patchwork of hybrid and parental forms often associated with gradual environmental or ecological transitions. Differential selection between hybrids and parental genotypes could be driven by a variety of factors, such as ecological, behavioural or physiological differences between the two parent species that result in reduced fitness of hybrids in environments that are not shared by both parental species (Moore & Price, 1993).

Our test of environmental niche divergence performed on the SDMs was significant (Figure S3B), indicating that the two species' environmental niches are different with little overlap in the environmental space (approximately 9.02%) occupied by *A. rubens* and *A. forbesi* based on our ellipsoid analyses (Figure 4b). One of our central questions is if hybrids experience reduced fitness except in areas of environmental conditions associated with both species. Indeed, these predictions are matched with the observation that all individuals with HI's ranging from 0.125–0.875 are almost entirely found at locations that fall within the two overlapping MVEs of the two species (Figure 4b). This area corresponds with the Gulf of Maine, an oceanographic region that has a different climate envelope than the rest of these species' distributions (Loder, 1998).

Selection against hybrids outside the zone of MVE overlap might also be consistent with our finding F_{ST} outlier loci between *A. rubens* and *A. forbesi* (Figures S7 and S9) suggesting divergent selection, as well as our *Asterias*-wide RDA results. However, interpretation of these values warrants caution as elevated F_{ST} can emerge in genomic regions of reduced diversity due to the effects of background selection and low recombination (Charlesworth, 1998; Cruickshank & Hahn, 2014), and finding localized areas of the genome with restricted gene flow is also consistent with endogenous selection. Interestingly, hybrids invade a separate space in the global RDA (Figure 5b), suggesting that there may be an environmental space that leads to positive selection for hybrid genotypes, and even potentially indicative of transgressive segregation (Rieseberg et al., 1999).

Indeed, it is challenging and perhaps impossible to completely discriminate between exogenous selection consistent with a mosaic hybrid zone and mechanisms of endogenous selection predicted under a tension zone, as outliers could be linked to endogenous genomic incompatibilities that coincide with environmental boundaries (Bierne et al., 2011; Fraïsse et al., 2016; Ryan et al., 2017) and the predicted shapes of clines generated by these two selection mechanisms are indistinguishable if a hybrid zone is stable and in long-term equilibrium (Kruuk et al., 1999; Schneemann et al., 2020). Fitness consequences for first-generation hybrids may also be slight, with diminishing effects through backcross generations, further limiting means for distinguishing the two mechanisms. Even with the limited modes of distinguishing the two mechanisms (Bierne et al., 2013; Simon et al., 2018), the spatial resolution of our samples also makes it difficult to clarify (Harrison & Larson, 2016). Greater spatial and

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17118 by University Of Georgia Libraries, Wiley Online Library on [24/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

numeric density, complemented by experimental evidence, could be promising in future efforts.

4.3 | Historical fisheries' impacts on hybridization

In addition to selective processes, hybrid zones in marine systems are likely to be shaped by ocean currents as well as historical contingencies. For example, the southerly coastal currents could be a source for *A. rubens* genomes into the Gulf of Maine before this current diverts from the coast as it passes between Cape Cod and the Georges Bank (Pringle et al., 2011). Unidirectional larval transport in ship ballast water could also potentially affect observed admixture and/or gene flow patterns, as the Pacific outgroup species of *Asterias* (*A. amurensis*) is known to have invaded Tasmania from northern Pacific source populations via larval transport in ballast water (Ling et al., 2012), and there are also examples of *A. rubens* invading the Black Sea in a similar manner (Ünsal Karhan et al., 2008).

Could the historical fisheries of the Northwestern Atlantic have led to some of the observed patterns of admixture? Our localities with hybrids (Harper & Hart, 2007) coincide with the eastern seaboard hub for large ground-fishing fleets that targeted the Grand Banks off of Newfoundland. At its peak, millions of tons of biomass from shallow Newfoundland waters were transported to the Gulf of Maine each year (McFarland, 1911), raising the possibility that the observed hybrid zone in the Gulf of Maine is partially the product of a large recent pulse of *A. rubens* (larvae or adults) as by-catch from Grand-Banks ground-fishing, which grew exponentially from the 1850s to the 1920s, before collapsing in the 1950s and finally closing in the mid-1990s (Starkey & Heidbrink, 2009). In this case, the mosaic configuration of the observed hybrid zone might also be consistent with a hybrid swarm associated with these putative pulses of human-mediated movement of *A. rubens* larvae (Nielsen et al., 2003; Nikula et al., 2008; Wang et al., 2017).

4.4 | Genome-environmental associations

The volume of the environmental space of *A. rubens*' ellipsoid (MVE) is over double the size of *A. forbesi*'s, suggesting a much larger range of environmental tolerance for *A. rubens* (Figure 4b). This large difference in MVE volumes is consistent with the much broader geographic range of *A. rubens* which extend across both western European and eastern North American coastlines. This key difference between *A. rubens* and *A. forbesi* is also consistent with the two genome-environment association (GEA) methods we implemented. First, the broader environmental niche breadth of *A. rubens* evident from comparing MVEs is corroborated by *Asterias rubens* showing fewer genotype-environment associations than *A. forbesi* obtained from LFMM (Table S5) and an RDA model for *A. forbesi* that is significant, in contrast with the *A. rubens* RDA model that is not significant (Table S4). The larger geographic range and niche breadth coupled with fewer GEAs could indicate that *A. rubens* has greater environmental niche plasticity.

For both *Asterias* species, environmental variables associated with specific nutrients were strongly associated with allelic variation (Table S5 for LFMM results and Table 2 and Table S4 for RDA results). In particular, mean nitrate showed relatively high associations with allelic variation (Table 2; Tables S4 and S5). This is consistent with the findings of a previous study that found that sea star species richness in the Gulf of California was correlated with nitrate concentrations (Cintra-Buenrostro et al., 2005). Post-metamorphic sea stars cannot obtain nitrate directly, but instead acquire it from the environment (Lawrence & Lane, 1982), suggesting that sea star genomic variation could be shaped in part by the availability of this nutrient. Interestingly, in early experiments involving echinoderm utilization of dissolved organic material, *A. forbesi* was shown to be highly adept at this type of amino acid absorption through epidermal tissue (Ferguson, 1963a, 1963b, 1967; Stephens & Schinske, 1961). Similarly, allelic components in *A. forbesi* are correlated with the spatial distributions of salinity (Table S5) and ocean pH (Table 2). Both factors are known to be key to survival in sea stars (Held & Harley 2009; Clark & Downey, 1992), and again known to be changing in this region (Salisbury & Jönsson, 2018; Brickman et al, 2021).

Perhaps an expected outcome (Clark & Downey, 1992) is that genomic variation in *Asterias* seems driven by temperature. Both sets of GEA analyses show a large influence of temperature on distribution of genomic diversity. In the LFMM analysis, *A. forbesi* harbours many SNPs that are statistically associated with long-term minimum temperatures (Table S5), although this variable did not contribute to the SDM of *A. forbesi*. Conversely, long-term minimum temperatures do appear to be a driver of the SDM for *A. rubens*, with a distribution well into the Arctic along the Norwegian coast. *A. rubens* also exhibits a negative relationship between maximum temperature and predicted presence (Figure S10). Our results suggest that each species faces oppositional selection pressure regarding temperature, such that *A. forbesi* may be limited by colder habitats (again making the PEI population of interest), *A. rubens* by warmer sites, and hybrid individuals persisting in regional patches within the tolerances of both species. As the Gulf of Maine is rapidly warming (Pershing et al., 2015), we expect further change in these distributional boundaries.

4.5 | The future of *Asterias* sea stars in the North Atlantic

As sea stars are keystone species in the threatened North Atlantic intertidal (Pershing et al., 2015; Petraitis & Dudgeon, 2020), understanding the population dynamics of these hybridizing species can drive our interpretation of a rapidly changing ecosystem (Pershing et al., 2015; Petraitis & Dudgeon, 2020) with continued genomic monitoring. In general, the broader consequences of hybridization in a changing climate are poorly known and variable; in some cases, hybridization may lead to extinction from genetic swamping of rare species and replacement by hybrid individuals (Todesco et al., 2016), whereas in other cases, it holds the promise of evolutionary rescue by introducing adaptive alleles from one population into another, reducing the

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17118 by University Of Georgia Libraries, Wiley Online Library on [24/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

probability of extinction (Seehausen, 2004; Willis et al., 2006; Fitzpatrick et al., 2020). While our study demonstrates that a substantial hybrid swarm has emerged from sister *Asterias* species hybridizing, how hybridization will affect the future of *Asterias* in the context of the rapid climate changes now occurring, such as the extreme warming of the Gulf of Maine during summer months, is less clear. Will the swarm follow a moving environmental gradient as changes in sea surface temperatures accelerate to the point of one species replacing the other? Alternatively, will the hybrid swarm expand in area while the two parental populations persist in their respective areas of origin? Another possible scenario consistent with our finding of loci under environmentally driven selection is where the current hybrid swarm could result in evolutionary rescue (Baskett & Gomulkiewicz, 2011; Brauer et al., 2023; Hansen, 2023), a process by which gene pools are enriched by intraspecific hybridization, allowing for greater resilience to climatic changes due to increased adaptive variation. In either of these cases, *Asterias* population genomic data will be worth monitoring as climate changes, and we expect studies that use whole genome sequence data to further augment such efforts by providing information regarding the timing and directionality of admixture along with probabilistic model-based inferences of demographic history and linked selection. Hence, *Asterias* could represent a useful system for understanding the dynamics of range shifts and introgressive hybridization in marine environments subject to climate change.

AUTHOR CONTRIBUTIONS

This study was conceived of and designed by MG and MJH. The majority of samples were collected by MG (primarily), LJM and MJH. MG conducted all laboratory work. Data analysis and visualization by MG, GEP-B, SJEB and LJM. MG wrote the first draft of the manuscript and performed all revisions. All authors contributed to editing the manuscript.

ACKNOWLEDGEMENTS

We are grateful to invaluable advice and assistance from past and present members of the Hickerson lab: Dr. Isaac Overcast, Dr. Laura Bertola and Connor French. Dr. Ana Carnaval, Dr. Elizabeth Alter, Dr. Mary Blair and Dr. Stephen Gaughran all provided feedback and assistance in project design and manuscript editing. We thank the staff at the American Museum of Natural History's Institute for Comparative Genomics for use of laboratory facilities and facilitation of sample collection and accession, including Dr. Anthony Caragiulo, Dr. Cheryl Hayashi, Lauren Audi, Svetlana Katanova and Mohammad Faiz. We also thank Dr. Naoko Kurata, Dr. Maria Byrne, Dr. Ó. Sindri Gíslason, and Dr. Graham Stone, Dr. Jason Brown, Dr. Alexander Xue, Dr. Ben Flanagan, Dr. Ăki Laruson, Norah Robertson, Anabel Carter, the staff of the Blue Ocean Society of New Hampshire and the staff of Nahant Marine Science Center of Northeastern University for contributing crucial samples for this study. Funding was provided by a grant from the National Science Foundation (DEB-1253710 to MJH). MG was supported by fellowships provided through the CUNY's Graduate Center and the City College of New York as well as the AMNH Lerner-Grey fund for marine research.

CONFLICT OF INTEREST STATEMENT

The authors state no conflict of interest.

DATA AVAILABILITY STATEMENT

Raw demultiplexed sequences are deposited on NCBI SRA (Accession number PRJNA984954) and will be made publicly available at https://www.ncbi.nlm.nih.gov/sra/PRJNA984954 upon publication of this manuscript. R scripts and intermediate files for analyses are available on the corresponding author's Github (https://github.com/MelinaGia/Asterias_RADseq).

ORCID

Melina Giakoumis https://orcid.org/0000-0003-3569-1668
Gonzalo E. Pinilla-Buitrago https://orcid.org/0000-0002-0065-945X
Lukas J. Musher https://orcid.org/0000-0003-3490-0341
John P. Wares https://orcid.org/0000-0003-1827-0669
Stuart J. E. Baird https://orcid.org/0000-0002-7144-9919
Michael J. Hickerson https://orcid.org/0000-0002-5802-406X

REFERENCES

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C. A., Buggs, R., Butlin, R. K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S. H., Hermansen, J. S., Hewitt, G., Hudson, A. G., Jiggins, C., … Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, *26*, 229–246.

Aiello-Lammens, M. E., Boria, R. A., Radosavljevic, A., Vilela, B., & Anderson, R. P. (2015). spThin: An R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography*, *38*, 541–545.

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews. Genetics*, *17*, 81–92.

Assis, J., Tyberghein, L., Bosch, S., Verbruggen, H., Serrão, E. A., de Clerck, O., & Tittensor, D. (2018). Bio-ORACLE v2.0: Extending marine data layers for bioclimatic modelling. *Global Ecology and Biogeography: A Journal of Macroecology*, *27*, 277–284.

Baird, S. J. E., Petružela, J., Jaroň, I., & Škrabánek, P. (2022). Genome polarisation for detecting barriers to geneflow. *Methods in Ecology and Evolution*, *14*, 512–528.

Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual Review of Rcology and Systematics*, *16*, 113–148.

Barton, N. H., & Hewitt, G. M. (1989). Adaptation, speciation and hybrid zones. *Nature*, *341*, 497–503.

Barton, N. H., & Hewitt, G. (1981). Hybrid zones and speciation. In: W. Atchley, & D. Woodruff (Eds.), *Evolution and Speciation* (pp. 109–145). Cambridge University Press.

Barton, N. H., Hewitt, G. M., Atchley, W. R., & Woodruff, D. S. (1981). Hybrid zones and speciation. *Evolution and Speciation*, *109–145*.

Baskett, M. L., & Gomulkiewicz, R. (2011). Introgressive hybridization as a mechanism for species rescue. *Theoretical Ecology*, *4*, 223–239.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, *57*, 289–300.

Bernatchez, L. (2016). On the maintenance of genetic variation and adaptation to environmental change: considerations from population genomics in fishes. *Journal of fish biology*, *89*, 2519–2556.

Bierne, N., Borsa, P., Daguin, C., Jollivet, D., Viard, F., Bonhomme, F., & David, P. (2003). Introgression patterns in the mosaic hybrid zone

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17118 by University Of Georgia Libraries, Wiley Online Library on [24/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

between Mytilus edulis and M. galloprovincialis. *Molecular Ecology*, *12*, 447–461.
Bierne, N., Gagnaire, P.-A., & David, P. (2013). The geography of introgression in a patchy environment and the thorn in the side of ecological speciation. *Current Zoology*, *59*, 72–86.
Bierne, N., Welch, J., Loire, E., Bonhomme, F., & David, P. (2011). The coupling hypothesis: Why genome scans may fail to map local adaptation genes. *Molecular Ecology*, *20*, 2044–2072.
Boria, R. A., Olson, L. E., Goodman, S. M., & Anderson, R. P. (2014). Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecological Modelling*, *275*, 73–77.
Brauer, C. J., Sandoval-Castillo, J., Gates, K., Hammer, M. P., Unmack, P. J., Bernatchez, L., & Beheregaray, L. B. (2023). Natural hybridization reduces vulnerability to climate change. *Nature Climate Change*, *13*, 282–289.
Brickman, D., Alexander, M. A., Pershing, A., Scott, J. D., & Wang, Z. (2021). *Projections of physical conditions in the Gulf of Maine in 2050* (p. 9). Elementa.
Bronson, C. L., Grubb, T. C., Jr., & Braun, M. J. (2003). A test of the endogenous and exogenous selection hypotheses for the maintenance of a narrow avian hybrid zone. *Evolution*, *57*, 630–637.
Byrne, M., Morrice, M. G., & Wolf, B. (1997). Introduction of the northern Pacific asteroid *Asterias amurensis* to Tasmania: Reproduction and current distribution. *Marine Biology*, *127*, 673–685.
Capblancq, T., & Forester, B. R. (2021). Redundancy analysis: A swiss army knife for landscape genomics. *Methods in Ecology and Evolution*, *12*, 2298–2309.
Casties, I., Clemmesen, C., Melzner, F., & Thomsen, J. (2015). Salinity dependence of recruitment success of the sea star *Asterias rubens* in the brackish western Baltic Sea. *Helgoland Marine Research*, *69*, 169–175.
Charlesworth, B. (1998). Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution*, *15*, 538–543.
Chown, S. L., Hodgins, K. A., Griffin, P. C., Oakeshott, J. G., Byrne, M., & Hoffmann, A. A. (2015). Biological invasions, climate change and genomics. *Evolutionary Applications*, *8*, 23–46.
Chunco, A. J. (2014). Hybridization in a warmer world. *Ecology and Evolution*, *4*, 2019–2031.
Cintra-Buenrostro, C. E., Reyes-Bonilla, H., & Herrero-Pérezrul, M. D. (2005). Oceanographic conditions and diversity of sea stars (Echinodermata: Asteroidea) in the Gulf of California, México. *Revista de Biologia Tropical*, *53*(Suppl 3), 245–261.
Clark, A. M., & Downey, M. E. (1992). *Starfishes of the Atlantic*. Chapman & Hall.
Cobos, M. E., Peterson, A. T., Barve, N., & Osorio-Olvera, L. (2019). Kuenm: An R package for detailed development of ecological niche models using maxent. *PeerJ*, *7*, e6281.
Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, *23*, 3133–3157.
Dobzhansky, T. (1934). Studies on hybrid sterility. *Cell and Tissue Research*, *21*, 169–223.
Eaton, D. A. R., & Overcast, I. (2020). Ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics*, *36*(8), 2592–2594.
El Ayari, T., Trigui El Menif, N., Hamer, B., Cahill, A. E., & Bierne, N. (2019). The hidden side of a major marine biogeographic boundary: A wide mosaic hybrid zone at the Atlantic-Mediterranean divide reveals the complex interaction between natural and genetic barriers in mussels. *Heredity*, *122*, 770–784.
Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, *14*, 2611–2620.
Fenberg, P. B., Posbic, K., & Hellberg, M. E. (2014). Historical and recent processes shaping the geographic range of a rocky intertidal gastropod: Phylogeography, ecology, and habitat availability. *Ecology and Evolution*, *4*, 3244–3255.
Feng, X., Park, D. S., Liang, Y., Pandey, R., & Papeş, M. (2019). Collinearity in ecological niche modeling: Confusions and challenges. *Ecology and Evolution*, *9*, 10365–10376.
Ferguson, J. C. (1963a). An autoradiographic study of the distribution of ingested nutrients in the starfish, *Asterias forbesi*. *American Zoologist*, *3*, 524.
Ferguson, J. C. (1963b). *The physiological mechanisms of nutrient transport in the starfish, Asterias forbesi*. Cornell Univ.
Ferguson, J. C. (1967). Utilization of dissolved exogenous nutrients by the starfishes, *Asterias forbesi* and *Henricia sanguinolenta*. *The Biological Bulletin*, *132*, 161–173.
Fitzpatrick, S. W., Bradburd, G. S., Kremer, C. T., Salerno, P. E., Angeloni, L. M., & Funk, W. C. (2020). Genomic and fitness consequences of genetic rescue in wild populations. *Current Biology: CB*, *30*, 517–522.e5.
Forester, B. R., Lasky, J. R., Wagner, H. H., & Urban, D. L. (2018). Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. *Molecular Ecology*, *27*, 2215–2233.
Fraïsse, C., Gunnarsson, P. A., Roze, D., Bierne, N., & Welch, J. J. (2016). The genetics of speciation: Insights from Fisher's geometric model. *Evolution*, *70*, 1450–1464.
Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies (B O'Meara, ed). *Methods in Ecology and Evolution*, *6*, 925–929.
Frichot, E., Schoville, S. D., Bouchard, G., & François, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, *30*, 1687–1699.
Garroway, C. J., Bowman, J., Cascaden, T. J., Holloway, G. L., Mahan, C. G., Malcolm, J. R., Steele, M. A., Turner, G., & Wilson, P. J. (2010). Climate change induced hybridization in flying squirrels. *Global Change Biology*, *16*, 113–121.
GBIF.org. (2021a). *GBIF Occurrence Download*. Accessed 20 September 2021. https://doi.org/10.15468/dl.eapfp4
GBIF.org. (2021b). *GBIF Occurrence Download*. Accessed 08 December 2021. https://doi.org/10.15468/dl.fw5cj9
Gosset, C. C., & Bierne, N. (2013). Differential introgression from a sister species explains high F(ST) outlier loci within a mussel species. *Journal of Evolutionary Biology*, *26*, 14–26.
Hansen, M. M. (2023). Prepping for climate change by introgressive hybridization. *Trends in Genetics*, *39*, 524–525.
Hare, M. P., Guenther, C., & Fagan, W. F. (2005). Nonrandom larval dispersal can steepen marine clines. *Evolution*, *59*, 2509–2517.
Harley, C. D. G., & Helmuth, B. S. T. (2003). Local- and regional-scale effects of wave exposure, thermal stress, and absolute versus effective shore level on patterns of intertidal zonation. *Limnology and Oceanography*, *48*, 1498–1508.
Harper, F. M., Addison, J. A., & Hart, M. W. (2007). Introgression versus immigration in hybridizing high-dispersal echinoderms. *Evolution*, *61*, 2410–2418.
Harper, F. M., & Hart, M. W. (2005). Gamete compatibility and sperm competition affect paternity and hybridization between sympatric Asterias Sea stars. *The Biological Bulletin*, *209*, 113–126.
Harper, F. M., & Hart, M. W. (2007). Morphological and phylogenetic evidence for hybridization and introgression in a sea star secondary contact zone. *Invertebrate Biology*, *126*, 373–384.
Harrison, R. G., & Larson, E. L. (2014). Hybridization, introgression, and the nature of species boundaries. *The Journal of Heredity*, *105*(Suppl 1), 795–809.

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17118 by University Of Georgia Libraries, Wiley Online Library on [24/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Harrison, R. G., & Larson, E. L. (2016). Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Molecular Ecology*, *25*, 2454–2466.

Held, M. B. E., & Harley, C. D. G. (2009). Responses to low salinity by the sea star Pisaster ochraceus from high- and low-salinity populations. *Invertebrate biology: a quarterly journal of the American Microscopical Society and the Division of Invertebrate Zoology/ASZ*, *128*, 381–390.

Hilbish, T. J., Lima, F. P., Brannock, P. M., Fly, E. K., Rognstad, R. L., & Wethey, D. S. (2012). Change and stasis in marine hybrid zones in response to climate warming. *Journal of Biogeography*, *39*, 676–687.

Howard, D. J., Waring, G. L., Tibbets, C. A., & Gregory, P. G. (1993). Survival of hybrids in a mosaic hybrid zone. *Evolution*, *47*, 789–800.

Ilves, K. L., Huang, W. E. N., Wares, J. P., & Hickerson, M. J. (2010). Colonization and/or mitochondrial selective sweeps across the North Atlantic intertidal assemblage revealed by multi-taxa approximate Bayesian computation. *Molecular Ecology*, *19*, 4505–4519.

Ingolfsson, A. (1992). The origin of the rocky shore fauna of Iceland and the Canadian maritimes. *Journal of Biogeography*, *19*, 705–712.

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, *23*, 1801–1806.

Jiménez, L., Soberón, J., Christen, J. A., & Soto, D. (2019). On the problem of modeling a fundamental niche from occurrence data. *Ecological Modelling*, *397*, 74–83.

Kass, J. M., Pinilla-Buitrago, G. E., Paz, A., Johnson, B. A., Grisales-Betancur, V., Meenan, S. I., Attali, D., Broennimann, O., Galante, P. J., Maitner, B. S., & Owens, H. L. (2023). wallace 2: A shiny app for modeling species niches and distributions redesigned to facilitate expansion via module contributions. *Ecography*, *3*, e06547.

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, *7*, 1225–1241.

Kelly, R. P., & Palumbi, S. R. (2010). Genetic structure among 50 species of the northeastern Pacific rocky intertidal community. *PLoS One*, *5*, e8594.

Kofler, R., Pandey, R. V., & Schlötterer, C. (2011). PoPoolation2: Identifying differentiation between populations using sequencing of pooled DNA samples (Pool-seq). *Bioinformatics*, *27*, 3435–3436.

Kruuk, L. E., Baird, S. J., Gale, K. S., & Barton, N. H. (1999). A comparison of multilocus clines maintained by environmental adaptation or by selection against hybrids. *Genetics*, *153*, 1959–1971.

Lawrence, J. M., & Lane, J. M. (1982). The utilization of nutrients by post-metamorphic echinoderms. In M. Jangoux & J. M. Lawrence (Eds.), *Echinoderm Nutrition*. CRC press.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, *25*, 1754–1760.

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

Ling, S. D., Johnson, C. R., Mundy, C. N., Morris, A., & Ross, D. J. (2012). Hotspots of exotic free-spawning sex: Man-made environment facilitates success of an invasive seastar. *The Journal of Applied Ecology*, *49*, 733–741.

Loder, J. W. (1998). The coastal ocean off northeastern North America : A large-scale view. *The Sea*, *11*, 105–138.

Loosanoff, V. L. (1964). Variations in time and intensity of setting of the starfish, Asterias forbesi, in Long Island Sound during a twenty-five year period. *The Biological bulletin*, *126*, 423–439.

Lozier, J. D. (2014). Revisiting comparisons of genetic diversity in stable and declining species: Assessing genome-wide polymorphism in north American bumble bees using RAD sequencing. *Molecular Ecology*, *23*, 788–801.

Lubchenco, J., & Menge, B. A. (1978). Community development and persistence in a low rocky intertidal zone. *Ecological Monographs*, *48*, 67–94.

MacKenzie, C. L., Jr., & Pikanowski, R. (1999). A decline in starfish, *Asterias forbesi*, abundance and a concurrent increase in northern quahog, Mercenaria, abundance and landings in the northeastern United States. *Marine Fisheries Review*, *61*, 66–71.

McFarland, R. (1911). *A history of the New England fisheries: With maps*. University of Pennsylvania.

Menge, B. A. (1979). Coexistence between the seastars *Asterias vulgaris* and *A. Forbes*; in a heterogeneous environment: A non-equilibrium explanation. *Oecologia*, *41*, 245–272.

Menge, B. A. (1986). A preliminary study of the reproductive ecology of the seastars Asterias vulgaris and A. forbesi in New England. *Bulletin of marine science*, *39*, 467–476.

Miller, M. R., Dunham, J. P., Amores, A., Cresko, W. A., & Johnson, E. A. (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research*, *17*, 240–248.

Moore, W. S., & Price, J. T. (1993). Nature of selection in the northern flicker hybrid zone and its implications for speciation theory. *Hybrid Zones and the Evolutionary Process*, *196*, 225.

Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. A., Kass, J. M., Uriarte, M., & Anderson, R. P. (2014). ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for maxent ecological niche models (J McPherson, ed). *Methods in Ecology and Evolution*, *5*, 1198–1205.

Nielsen, E. E., Hansen, M. M., Ruzzante, D. E., Meldrup, D., & Grønkjaer, P. (2003). Evidence of a hybrid-zone in Atlantic cod (*Gadus morhua*) in the Baltic and the Danish Belt Sea revealed by individual admixture analysis. *Molecular Ecology*, *12*, 1497–1508.

Nikula, R., Strelkov, P., & Väinölä, R. (2008). A broad transition zone between an inner Baltic hybrid swarm and a pure North Sea subspecies of *Macoma balthica* (Mollusca, Bivalvia). *Molecular Ecology*, *17*, 1505–1522.

Nosil, P. (2012). *Ecological Speciation*. OUP Oxford.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, P., Legendre, D., & McGlinn, P. R. (2020). *Vegan: Community ecology package*. https. cran. r-project. org/package= vegan.

Paine, R. T. (1966). Food web complexity and species diversity. *The American Naturalist*, *100*, 65–75.

Palumbi, S. R. (1994). Genetic divergence, reproductive isolation, and marine speciation. *Annual Review of Ecology and Systematics*, *25*, 547–572.

Pershing, A. J., Alexander, M. A., Hernandez, C. M., Kerr, L. A., le Bris, A., Mills, K. E., Nye, J. A., Record, N. R., Scannell, H. A., Scott, J. D., Sherwood, G. D., & Thomas, A. C. (2015). Slow adaptation in the face of rapid warming leads to collapse of the Gulf of Maine cod fishery. *Science*, *350*, 809–812.

Petkova, D., Novembre, J., & Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics*, *48*, 94–100.

Petraitis, P. S., & Dudgeon, S. R. (2020). Declines over the last two decades of five intertidal invertebrate species in the western North Atlantic. *Communications Biology*, *3*, 591.

Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E., & Blair, M. E. (2017). Opening the black box: An open-source release of maxent. *Ecography*, *40*, 887–893.

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*, 231–259.

Prates, I., Penna, A., Rodrigues, M. T., & Carnaval, A. C. (2018). Local adaptation in mainland anole lizards: Integrating population history and genome-environment associations. *Ecology and Evolution*, *8*, 11932–11944.

Pringle, J. M., Blakeslee, A. M. H., Byers, J. E., & Roman, J. (2011). Asymmetric dispersal allows an upstream region to control

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17118 by University Of Georgia Libraries, Wiley Online Library on [24/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

population structure throughout a species' range. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 15288–15293.
Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*, 945–959.
Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, *24*, 4348–4370.
Rieseberg, L. H., Archer, M. A., & Wayne, R. K. (1999). Transgressive segregation, adaptation and speciation. *Heredity*, *83*(Pt 4), 363–372.
Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., et al. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution*, *34*, 3299–3302.
Ryan, S. F., Fontaine, M. C., Scriber, J. M., Pfrender, M. E., O'Neil, S. T., & Hellmann, J. J. (2017). Patterns of divergence across the geographic and genomic landscape of a butterfly hybrid zone associated with a climatic gradient. *Molecular Ecology*, *26*, 4725–4742.
Salisbury, J. E., & Jönsson, B. F. (2018). Rapid warming and salinity changes in the Gulf of Maine alter surface ocean carbonate parameters and hide ocean acidification. *Biogeochemistry*, *141*, 401–418.
Sameoto, J. A., & Metaxas, A. (2008). Interactive effects of haloclines and food patches on the vertical distribution of 3 species of temperate invertebrate larvae. *Journal of Experimental Marine Biology and Ecology*, *367*, 131–141.
Sanford, E., & Kelly, M. W. (2011). Local adaptation in marine invertebrates. *Annual Review of Marine Science*, *3*, 509–535.
Schneemann, H., de Sanctis, B., Roze, D., Bierne, N., & Welch, J. J. (2020). The geometry and genetics of hybridization. *Evolution*, *74*, 2575–2590.
Schopf, T. J. M., & Murphy, L. S. (1973). Protein polymorphism of the hybridizing sea stars *Asterias forbesi* and *Asterias vulgaris* and implications for their evolution. *The Biological Bulletin*, *145*, 589–597.
Seehausen, O. (2004). Hybridization and adaptive radiation. *Trends in Ecology & Evolution*, *19*, 198–207.
Simon, A., Bierne, N., & Welch, J. J. (2018). Coadapted genomes and selection on hybrids: Fisher's geometric model explains a variety of empirical patterns. *Evolution Letters*, *2*, 472–498.
Starkey DJ, Heidbrink I (2009) A history of the North Atlantic fisheries.
Stephens, G. C., & Schinske, R. A. (1961). Uptake of amino acids by marine invertebrates. *Limnology and Oceanography*, *6*, 175–181.
Svedin, N., Wiley, C., Veen, T., Gustafsson, L., & Qvarnström, A. (2008). Natural and sexual selection against hybrid flycatchers. *Proceedings of the Royal Society B: Biological Sciences*, *275*, 735–744.
Taylor, S. A., Larson, E. L., & Harrison, R. G. (2015). Hybrid zones: Windows on climate change. *Trends in Ecology & Evolution*, *30*, 398–406.
Thomas, C. D., Franco, A. M. A., & Hill, J. K. (2006). Range retractions and extinction in the face of climate warming. *Trends in Ecology & Evolution*, *21*, 415–416.
Todesco, M., Pascual, M. A., Owens, G. L., Ostevik, K. L., Moyers, B. T., Hübner, S., Heredia, S. M., Hahn, M. A., Caseys, C., Bock, D. G., & Rieseberg, L. H. (2016). Hybridization and extinction. *Evolutionary Applications*, *9*, 892–908.
Ünsal Karhan, S., Kalkan, E., & Baki Yokeş, M. (2008). First record of the Atlantic starfish, *Asterias rubens* (Echinodermata: Asteroidea) from the Black Sea. *Marine Biodiversity Records*, *1*, e63.
Vallejo-Marín, M., & Hiscock, S. J. (2016). Hybridization and hybrid speciation under global change. *The New Phytologist*, *211*, 1170–1187.
Waltari, E., & Hickerson, M. J. (2013). Late Pleistocene species distribution modelling of North Atlantic intertidal invertebrates (C McClain, ed). *Journal of Biogeography*, *40*, 249–260.
Wang, H., Vieira, F. G., Crawford, J. E., Chu, C., & Nielsen, R. (2017). Asian wild rice is a hybrid swarm with extensive gene flow and feralization from domesticated rice. *Genome Research*, *27*, 1029–1038.
Wares, J. P. (2001). Biogeography of Asterias: North Atlantic climate change and speciation. *The Biological Bulletin*, *201*, 95–103.
Wares, J. P., & Cunningham, C. W. (2001). Phylogeography and historical ecology of the North Atlantic intertidal. *Evolution*, *55*, 2455–2469.
Wenger, S. J., & Olden, J. D. (2012). Assessing transferability of ecological models: An underappreciated aspect of statistical validation: Model transferability. *Methods in Ecology and Evolution*, *3*, 260–267.
Willis, B. L., van Oppen, M. J. H., Miller, D. J., Vollmer, S. V., & Ayre, D. J. (2006). *The role of hybridization in the evolution of reef corals* (Vol. *37*, pp. 489–517). Annual Review of Ecology, Evolution, and Systematics.
Worley, E. K., & Franz, D. R. (1983). A comparative study of selected skeletal structures in the sea stars *Asterias forbesi* (Desor), *A. vulgaris Verrill*, and *A. rubens* L., with a discussion of possible relationships. *Proceedings of the Biological Society of Washington*, *96*, 524–547.
Xuereb, A., Kimber, C. M., Curtis, J. M. R., Bernatchez, L., & Fortin, M.-J. (2018). Putatively adaptive genetic variation in the giant California Sea cucumber (*Parastichopus californicus*) as revealed by environmental association analysis of restriction-site associated DNA sequencing data. *Molecular Ecology*, *27*, 5035–5048.
Zardi, G. I., Nicastro, K. R., Canovas, F., Costa, J. F., Serrao, E. A., & Pearson, G. A. (2011). Adaptive traits are maintained on steep selective gradients despite gene flow and hybridization in the intertidal zone. *PLoS One*, *6*(6), e19402.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Giakoumis, M., Pinilla-Buitrago, G. E., Musher, L. J., Wares, J. P., Baird, S. J. E., & Hickerson, M. J. (2023). Evidence of introgression, ecological divergence and adaptation in *Asterias* sea stars. *Molecular Ecology*, *00*, 1–17. https://doi.org/10.1111/mec.17118

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17118 by University Of Georgia Libraries, Wiley Online Library on [24/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License