

RESEARCH ARTICLE

Population Genomics Reveals Seahorses (*Hippocampus erectus*) of the Western Mid-Atlantic Coast to Be Residents Rather than Vagrants

J. T. Boehm^{1,3*}, John Waldman^{2,3}, John D. Robinson⁴, Michael J. Hickerson^{1,3}

1 Department of Biology, City College of New York, 160 Convent Ave., New York, New York, 10031, United States of America, **2** Biology Department, Queens College, City University of New York, 65-30 Kissena Blvd., Queens, New York, 11367-1597, United States of America, **3** Subprogram in Ecology, Evolution and Behavior, The Graduate Center of the City University of New York, 365 5th Ave, New York, New York, 10016, United States of America, **4** South Carolina Department of Natural Resources, Marine Resources Research Institute, 217 Fort Johnson Rd., Charleston, South Carolina, 29412, United States of America

* jtboehmjr@gmail.com



OPEN ACCESS

Citation: Boehm JT, Waldman J, Robinson JD, Hickerson MJ (2015) Population Genomics Reveals Seahorses (*Hippocampus erectus*) of the Western Mid-Atlantic Coast to Be Residents Rather than Vagrants. PLoS ONE 10(1): e0116219. doi:10.1371/journal.pone.0116219

Academic Editor: Matthias Stöck, Leibniz-Institute of Freshwater Ecology and Inland Fisheries, GERMANY

Received: July 25, 2014

Accepted: October 20, 2014

Published: January 28, 2015

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: Illumina generated sequence reads for each individual are available at NCBI Sequence Read Archive (SRA): SRP048776. Additional data are available from Dryad (doi:10.5061/dryad.qc2qq).

Funding: The National Science Foundation (40C33-00-01) supported the dissertational research of J. T. B. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Understanding population structure and areas of demographic persistence and transients is critical for effective species management. However, direct observational evidence to address the geographic scale and delineation of ephemeral or persistent populations for many marine fishes is limited. The Lined seahorse (*Hippocampus erectus*) can be commonly found in three western Atlantic zoogeographic provinces, though inhabitants of the temperate northern Virginia Province are often considered tropical vagrants that only arrive during warm seasons from the southern provinces and perish as temperatures decline. Although genetics can locate regions of historical population persistence and isolation, previous evidence of Virginia Province persistence is only provisional due to limited genetic sampling (i.e., mitochondrial DNA and five nuclear loci). To test alternative hypotheses of historical persistence versus the ephemerality of a northern Virginia Province population we used a RADseq generated dataset consisting of 11,708 single nucleotide polymorphisms (SNP) sampled from individuals collected from the eastern Gulf of Mexico to Long Island, NY. Concordant results from genomic analyses all infer three genetically divergent subpopulations, and strongly support Virginia Province inhabitants as a genetically diverged and a historically persistent ancestral gene pool. These results suggest that individuals that emerge in coastal areas during the warm season can be considered “local” and supports offshore migration during the colder months. This research demonstrates how a large number of genes sampled across a geographical range can capture the diversity of coalescent histories (across loci) while inferring population history. Moreover, these results clearly demonstrate the utility of population genomic data to infer peripheral subpopulation persistence in difficult-to-observe species.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

In warmer seasons, the waters lining the concrete bulkheads, wooden piers, estuaries, and sandy beaches of the temperate Northeastern United State's mid-Atlantic coast become home to numerous tropical fish species [1,2]. Over a century of research has cataloged the immigration of tropical vagrants or "strays" to these coastal mid-Atlantic waters. The majority of these individuals arrive due to passive planktonic dispersal in summer months, transported by ocean currents that circle north off the warm water mass of the Gulf Stream as it deflects northeast from U.S. towards Europe at roughly 35°N latitude [3,4]. This phenomenon positions Cape Hatteras as a delineation point between the zoogeographic Virginia and Carolina Provinces, each defined by distinct faunal endemism and unique macroclimatic conditions (Fig. 1) [5,6]. Following this observation, studies of species found in both provinces suggest that Cape Hatteras acts as a "barrier" where intraspecific gene flow is reduced between provinces or alternatively acts as a northern latitudinal limit during the winter for species without cold thermal tolerance [6,7]. In this latter case, more sedentary tropical species that passively drift into the temperate Virginia Province during warmer months locally perish after cold winter temperatures advance.

Though many fishes exhibit wide thermal tolerance, ascertaining the true range of marine species can be challenging due to factors that include patchy distributions, cyclical population sizes, and seasonal movement patterns [7,8]. One species often associated with tropical vagrants in the Virginia Province is the Lined seahorse, *Hippocampus erectus* [3]. Its status as a persistent independent gene pool (i.e., subpopulation) is uncertain primarily due to its near-shore absence during cold winter months and a scarcity of direct winter observations of individuals. *H. erectus* is commonly found in coastal zones in three zoogeographic provinces: Caribbean (tropical), Carolina (warm-temperate) and Virginia (temperate) (Fig. 1). Some researchers suggest that long-distance rafting carries migrants northward to temporarily inhabit the Virginia Province as temperatures warm [3], a prediction supported by substantial observational evidence of long-distance rafting migration throughout its range [9,10]. In contrast, other researchers suggest that localized active dispersal directed toward offshore migration for thermal refuge in continental shelf waters during late fall accounts for its winter absence [11]. This hypothesis of seasonal localized migration is partially supported by the observation of in-shore colonization of *H. erectus* as temperatures warm in April to June, characteristic of most temperately adapted fishes [3], and earlier than the July to September arrival typical for the majority of tropical strays [1,12].

Ecologists and evolutionary biologists often focus on questions at different temporal scales, but both fields are increasingly making use of genetic data to test hypotheses about population history, estimate the movement of individuals between local populations, and characterize the spatial distribution of genetic variation for effective species management [8,13,14]. One example is the use of genetic data to examine source-sink dynamics [15,16]. True sink populations, even if annually persistent, require continual immigration from source populations and are expected to exhibit genetic homogeneity with source populations or heterogeneity reflecting multiple sources of immigrants, while over time independent breeding subpopulations through random (genetic drift) or deterministic (natural selection) processes will exhibit distinct genetic divergence [13]. A number of studies have examined the biogeography and genetic divergence of *Hippocampus* species [17]. Most of this research has focused on Indo-Pacific species with genetic variation ranging in spatial scales from among localized South African estuaries [18], to widespread species complexes associated with rafting driven colonization [19], and differing levels of intraspecific divergence attributed to both ecological traits and biogeographic divides [20,21].

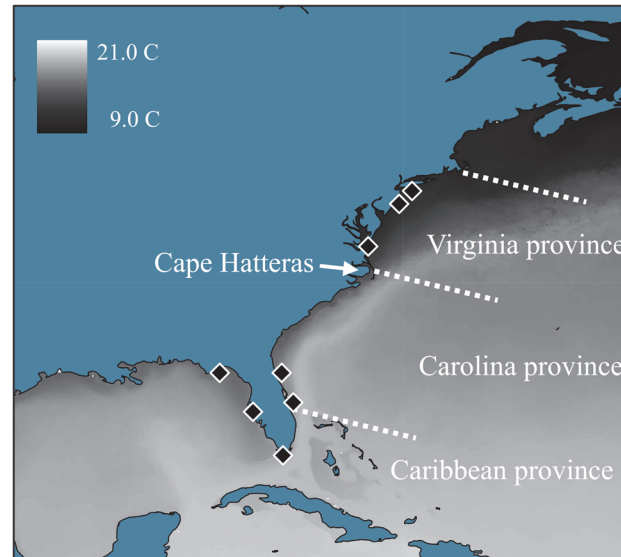


Figure 1. Map of zoogeographic provinces, collection sites, and temperature variance. Contrasting ocean minimum sea surface temperatures across zoogeographic provinces: generated in ARCGIS v.9.3 using the Bio-ORACLE long-term climatic dataset [67]. Collection sites from the northeastern Gulf of Mexico to New York State indicated by diamonds: Apalachicola, Tampa Bay-Charlotte Harbor, Florida Keys, Indian River Lagoon and Jacksonville, FL., Chesapeake Bay, New Jersey-New York.

doi:10.1371/journal.pone.0116219.g001

Here we test whether the presence of *H. erectus* individuals north of Cape Hatteras are the result of an ephemeral deme that is seasonally replenished from demographically persistent southern populations (H1; Hypothesis 1), or in contrast, there are persistent and isolated populations on either side of Cape Hatteras (H2; Hypothesis 2). A previous study of the *H. erectus* complex utilized mitochondrial DNA and five more slowly evolving nuclear loci across many individuals ($n = 115$), yet rejected H2 in favor of H1, with little divergence and evidence of isolation across Cape Hatteras [22]. Now, with the decreasing cost of high-throughput sequencing, data can be sampled from across the autosomal genome to account for variations in mutation, coalescent history, and recombination, thereby facilitating a view of the complexity of a species evolutionary history with the potential to infer more recent divergence and/or populations differentiating in the presence of gene flow [23].

To date, genome wide single nucleotide polymorphism (SNP) datasets generated by restriction site associated DNA sequencing (i.e., RADseq) have been utilized to study several fish species. Examples utilizing RADseq datasets include the support of cryptic differentiation between populations of the Baltic Sea herring (*Clupea herangus*) [24], the detection of hybrid individuals between trout species [25], genetic divergence of various stickleback populations [26–28], and robust phylogenetic resolution between African cichlid species [29]. To test the aforementioned competing hypotheses H1 and H2, we generated a genomic RADseq dataset consisting of 11,708 SNPs across individuals of *H. erectus* from the eastern Gulf of Mexico to Long Island, NY (Fig. 1).

Although we base our inference from only 4–9 individuals per each of the three zoogeographic provinces (total individuals; $n = 23$), data from large numbers of unlinked loci allow highly resolved inference even with few individuals [30–32]. Moreover, given that outbred diploid genomes are comprised of recombining segments of DNA inherited from large pools of ancestors [33], genome-level datasets should capture the diversity of coalescent histories (across loci) that reflects population history, such that information comes more from the

number of loci sampled through the genome than from numbers of individuals per sampling locality [34,35].

Methods and Materials

Sampling and bioinformatics

Samples of *H. erectus* ranged from the eastern Gulf of Mexico to New York State ($n = 23$). Samples were collected from 2009–2013 from the following locations: Apalachicola, FL, Tampa Bay, FL, Charlotte Harbor, FL, the Florida Keys, Jacksonville, FL, and Indian River Lagoon, FL, Chesapeake Bay, New Jersey, the Hudson River and Long Island, NY. The specimens collected in this study were carried out in accordance and approval of the Queens College Institutional Animal Care and Use Committee (IACUC) (Permit # 137), which approved all aspects of specimen use in this study. Domestic fishing of *Hippocampus* is neither under direct regulation within the United States nor under species protection and no specific permissions were required for these locations; however we collaborated with the following authorities for samples, and if standard collection permits were required they were issued for each collection location. The Florida specimens used in our study were collected under the authority of the Florida Fish and Wildlife (FFW) as part of the FFW: Southeast Area Monitoring and Assessment Program. Samples from the Chesapeake Bay were collected in collaboration with the Virginia Institute of Marine Science (VIMS), which is authorized to collect any fishes necessary for research under the Code of Virginia. Lastly, samples collected in New Jersey and New York were collected in collaboration with Rutgers University under the New Jersey Department of Environmental Protection and the New York State Department of Environmental Conservation (DEC) Special Licensing Unit, License No. 1638 with additional samples collected under DEC License No. 1405.

Sequenced samples were randomly chosen from a large number of individuals ($n > 100$) over multiple collection years to ensure genomic similarity was not the result of non-independent relatedness. Total Genomic DNA was extracted using Puregene extraction (Qiagen) from tail muscle tissue and treated with RNAase A following standard protocols. Genomic DNA quality was checked on an agarose gel to ensure that the majority of DNA was $> 10,000$ bp and equalized to 30 ng/uL using Qubit Fluorometric Quantitation (Invitrogen). Library construction and restriction site associated DNA sequencing (RADseq) protocol followed [36,37]. Floragenex carried out library preparation and sequencing. Genomic DNA restriction digestion utilized the SbfI enzyme and individual sequence adapters and barcode identifiers were ligated to genomic DNA prior to sequencing on the Illumina HiSeq platform. All sequences from cut sites resulted in single-end reads, which were demultiplexed and trimmed of adapters to 90bp fragment lengths.

Total reads per individual ranged from 1,264,862–4,736,299. The individual with the largest number of reads was processed to construct a de novo pseudo reference genome, and reads for each individual were aligned using BOWTIE [38], SAMTOOLS algorithms [39] and custom Floragenex perl scripts were used to detect SNPs and call genotypes. SNP datasets were formatted in the variant call format (vcf) [40]. Initial genotyping required a minimum Phred quality score of 15, a minimum of 4 \times sequence coverage, with a minimum of 65% of individuals genotyped. Additional filtering was applied using R v.3 [41] to ensure a Phred score equal to a hard cutoff of $q = 20$ (base call accuracy lower than 99%). To reduce the inclusion of false SNP discovery due to paralogous sequences or low quality genotype calls, vcftools was utilized to remove any sites with a minimum depth of 8 \times sequence coverage and maximum depth calculated in R based on the mean depth + 1.5 standards deviation (= 295) across all sites. The final datasets resulted in a bi-allelic matrix of 11,708 genotypes (5777 90bp sequences) across

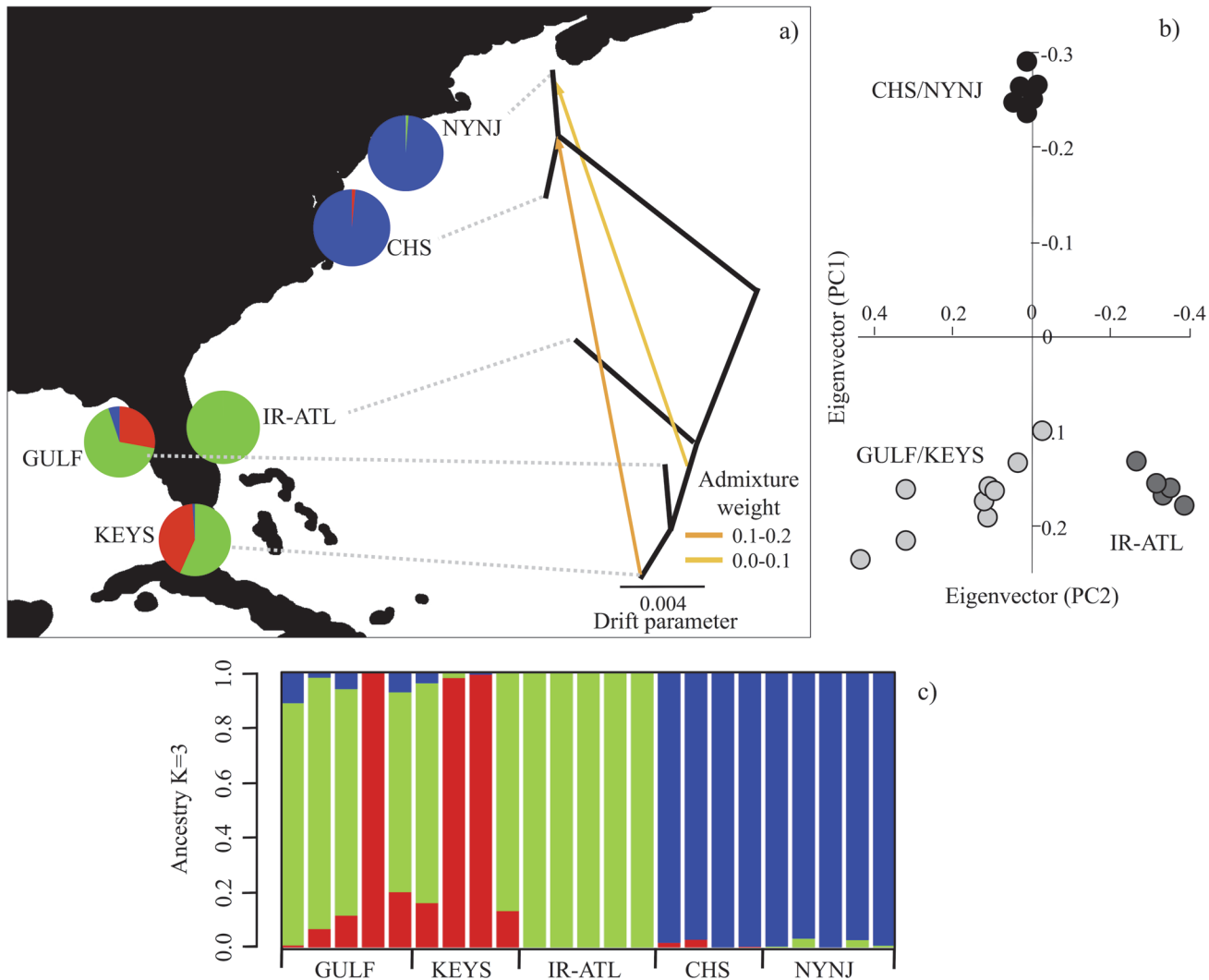


Figure 2. Genomic variation across individuals and subpopulations. (a) *Treemix* population tree with branch lengths scaled to the amount of genetic drift between regions and inferred proportion of genetic admixture ($m = 2$) between southern and northern regions represented by arrows. Dotted lines do not represent branch length. (b) Principle component analysis. Black circles = Chesapeake Bay-New York, dark grey circles = Florida Atlantic coast, and light grey circles = Gulf of Mexico-Florida Keys. Pie diagrams (a) represent ancestry coefficient proportions derived from the sNMF ancestry plot (c). Each line of the sNMF plot represents one individual.

doi:10.1371/journal.pone.0116219.g002

individuals at all sites. For details on per individual raw reads, filtered, and analyzed reads see [S1 Table](#).

Population genomic analyses

A principal components analysis (PCA) was implemented to determine if sampled individuals reflect a history of differentiated populations by outputting individual coordinates along axes of genetic variation within a statistical framework [42] that correspond to the first two principle components in Fig. 2B. To further aide in assigning individuals to differentiated populations by inferring ancestry coefficients representing the proportions of each individual's genome that originated from a specified number of ancestral gene pools (K) we used the program sNMF [43]. The program sNMF estimates individual ancestry and population clustering by utilizing a sparse non-negative matrix factorization algorithm (sNMF) to compute least-squares estimates

of ancestry coefficients. This software is capable of efficiently analyzing large bi-allelic datasets without loss of accuracy when compared with more commonly utilized programs STRUCTURE [44] and ADMIXTURE [45] that use the same underlying model to infer ancestry coefficients. However, in contrast to the aforementioned programs, sNMF has significantly better computational efficiency and is robust to many of the demographic assumptions of Hardy-Weinberg and linkage equilibrium [43,46]. To verify the accuracy of this program Frichot et al. (2014) conducted an in-depth comparison with the software ADMIXTURE using simulated and empirical datasets and found concordant results across trials, while sNMF outperformed ADMIXTURE when population inbreeding (FIS) was high. For our dataset ancestry coefficients (K) were estimated using sNMF to determine subpopulation membership by running 10 replicates of K 2–6 using a cross-entropy criterion (CEC). To evaluate the predictive capability and error of the ancestry estimation algorithm, sNMF employs the CEC, which is comparable to the likelihood value implemented in the program ADMIXTURE. To select the best-supported ancestry coefficient, the lowest CEC value was represented by the K value (K = 3). The ancestry coefficient plot (Fig. 2C) was visualized using R v.3. For information on CEC values, as well as results obtained between sNMF and STRUCTURE on a subset of the total data (SNP = 2000) see S1 Methods.

The program Treemix [47] was utilized to infer the phylogenetic relationships between sampled locations while accounting for ancestral admixture among populations. Specifically, Treemix incorporates a model to allow for population divergence in the presence of post-divergence admixture/migration (m) given that incorporation of this parameter can improve the likelihood fit of a bifurcating phylogeny. More specifically, the m parameter represents the proportion of admixture from one population to another [48]. The resulting phylogeny is based on a composite maximum likelihood of the local optimum tree, determined using a similar approach to Felsenstein [49], with branch lengths proportional to the amount of genetic drift that has occurred per branch.

Population genetic statistics (Table 1; Fig. 2D-2F) were generated using vcftools and calculated across all SNPs per individual. The calculation of F_{ST} utilized between subpopulations [50] specifically accounts for differences in sample size and a small number of sampled individuals, and recent studies have shown that bi-allelic SNPs (>1000) using this approach will result in precise F_{ST} estimates [51].

To investigate the visual similarity between genetic and geographic distance from the PCA analysis (Fig. 2A and 2B), we conducted a test for isolation-by-distance (IBD) to see if this pattern meets the expectation of genetic similarity decaying with geographic distance [52] using the IBD program by Mantel’s test (10,000 randomizations) of linearized F_{ST} (F_{ST}/(1-F_{ST})) and shoreline distance (km) [53]. Pairwise F_{ST}, calculated in vcftools, and distance of coastlines between sampling locations in kilometers was determined using Google Earth Tools. For this Mantel test, the centroid distance between sampling locations for the Gulf-Keys subpopulation

Table 1. Population genetic summary statistics for each subpopulation.

Subpopulations	<i>n</i>	F _{ST}		SNPs	Mean He	Mean S
		North-Atlantic	South-Atlantic			
North-Atlantic	9			11,708	1,104	304
South-Atlantic	5	0.1012		11,708	1,090	191
Gulf-Keys	9	0.083	0.0454	11,708	1,183	309

F_{ST}, number of individuals (*n*), total number of SNPs, observed mean heterozygote alleles per individual/per subpopulation (He) and mean number of singletons (S) per individual/per subpopulation.

doi:10.1371/journal.pone.0116219.t001

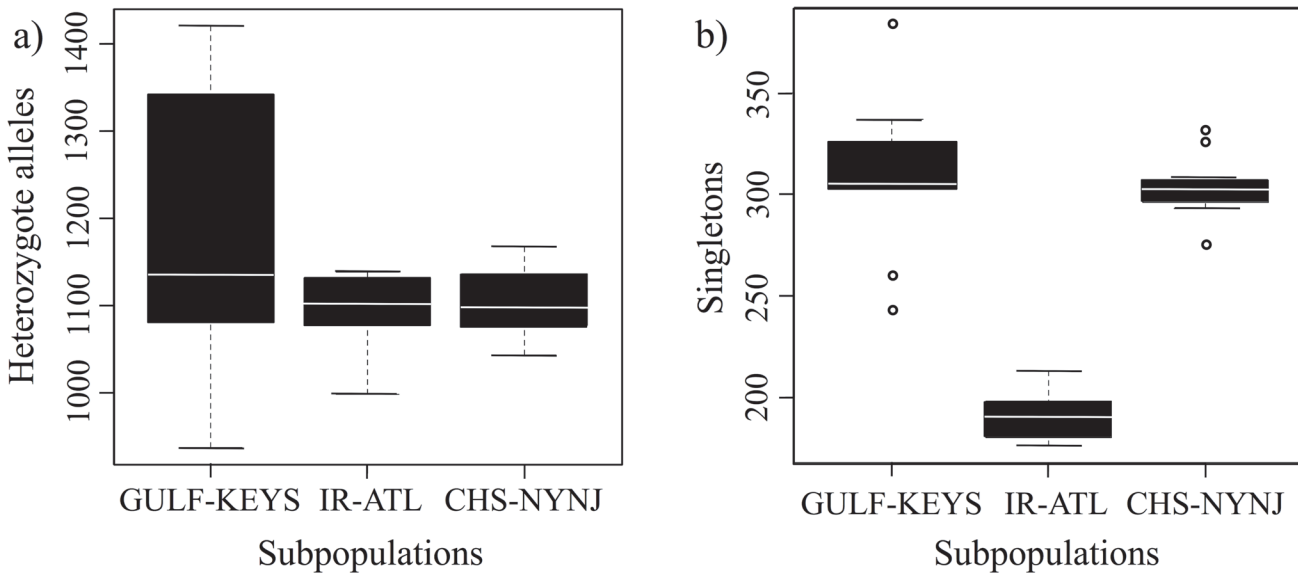


Figure 3. Distribution of heterozygote and singleton genotypes. Boxplots represent the range of observed heterozygote genotypes (a) and singleton genotypes per individual/per subpopulation (b).

doi:10.1371/journal.pone.0116219.g003

was utilized and results indicated a non-significant correlation between geographic and genetic distance ($p = 0.4925$). See [S1 Methods](#) for additional information and regression plots.

Results and Discussion

Support for northern subpopulation divergence and isolation

Our results strongly support H2 over H1 with Virginia Province residents of *H. erectus* coming from a persistently breeding and isolated ancestral gene pool, rejecting the categorization of it being composed of seasonal migrants. The sNMF-based estimates of ancestry coefficients support three distinct subpopulations with limited admixture ($K = 3$) ([Fig. 2C](#)): 1) the eastern Gulf of Mexico-Florida Keys (Gulf-Keys), 2) the eastern Floridian Peninsula (South-Atlantic), and 3) Chesapeake Bay-New York (North-Atlantic). The $K = 3$ value reported in our study is considered robust as it exhibited the lowest CEC value across replicate runs of all K values ($K = 2-6$). This substructure also visually emerges from the first two principle components of the PCA from the total amount of observed genomic variation ([Fig. 2B](#)). Here, the individuals from north of Cape Hatteras form a tight cluster, while individuals sampled from the Gulf-Keys and South-Atlantic form a cline between the tightly clustered South-Atlantic individuals and an admixed set of Gulf-Keys individuals. Consistent with these results is the inferred population history that emerges from *Treemix*, which is concordant with long-term isolation of the North Atlantic sub-population with limited post-divergence admixture with southern subpopulations.

The elevated heterozygosity found in Gulf-Keys individuals ([Fig. 3A](#)) could be the result of admixture from un-sampled western Gulf/Caribbean individuals, which is also indicated from the sNMF analysis ([Fig. 2C](#)). However, this elevated heterozygosity could also be the result of a larger effective population size [54]. In contrast, the northern subpopulation shows a reduction in heterozygosity with an elevated level of singletons ([Fig. 3B](#)). This pattern indicates a possible demographic expansion after the last glacial maximum that is consistent with the likely unsuitable habitat in the Virginia Province during the late Pleistocene. This history of shifting habitat

driven by climate change is suggested by palaeo-climatological research indicating that temperate environments north of Cape Hatteras were displaced southward [6,55], as well as the formation of the Chesapeake Bay 7.4–8.2 kya due to post-last glacial maximum sea level rise [56].

Causes of divergence and isolation of the northern subpopulation

Given the strong evidence we report for Virginia Province inhabitants of *H. erectus* representing a persistently isolated independent subpopulation from other regional ancestral gene pools, there are several conceivable non-mutually exclusive causes of this divergence. First, seagrass is a preferred breeding habitat of *H. erectus* and a long gap without coastal seagrasses exists along the Georgia and South Carolina coastlines (roughly 600km) [57]. This barrier of unsuitable breeding habitat between northern Florida and the Virginia Province therefore likely results in the fish's rarity in this area, thereby increasing genetic isolation of the northern subpopulation [58,59]. The confamilial pipefish *Syngnathus floridae* also shares a similar pattern of genetic divergence across this region of unsuitability, though the area of absence extends from the southern end of the Florida Peninsula to near Cape Hatteras, with the northern population extending from North Carolina to Chesapeake Bay [60]. Secondly, long-distance migration of *H. erectus* is observed to occur via *Sargassum* rafting driven by ocean currents [10]. Under this mode of migration, the northeastern deflection in ocean currents near Cape Hatteras toward the Mid-Atlantic may limit the arrival of southern migrants to the Virginia province. Lastly, individuals that do arrive from southern provinces may have a lower physiological tolerance to temperate conditions, reducing the chance of winter survival and also increasing the amount of genetic isolation. Selection correlated to the shift in macroclimate at Cape Hatteras has been observed in marine fishes [61], and future analysis of northern adaptation in *H. erectus* may help decouple the potential drivers of temperate subpopulation genetic isolation. Although our observed patterns of genetic isolation could have emerged via a continuous isolation-by-distance regime without clear breaks driving the isolation, a Mantel test resulted in a non-significant relationship between genomic and geographic distance ($p = 0.4925$).

Support for local seasonal migration

Our results also support local offshore migration to account for the coastal absence of *H. erectus* from Virginia Province during winter months. While extreme temperature changes influence latitudinal movement of many species [1], substantial seasonal movement to and from provinces for *H. erectus* is unlikely due to its relatively weak swimming ability [62]. To avoid nearshore cold water temperatures, localized inshore-offshore migration has been reported for the confamilial pipefish (*Syngnathus fuscus*), which has similar life history traits to *H. erectus* [63], and has also been suggested for some other species of *Hippocampus* [62]. As a qualitative comparison we examined abundance records from NOAA long-term offshore trawl surveys of *S. fuscus* (1972–2008; >90% 20 km off-coast; depth 10–20m) and found that they closely resemble that of *H. erectus*, further supporting intercontinental shelf overwintering (For additional details see [S2 Methods](#)). Regarding direct observation of this phenomena, a single record from divers in 1968 documented both species “hibernating” on the shelf substrates off Long Island, NY [64], where they resumed swimming several minutes after being brought to the surface. Many fish adapt to winter temperatures by decreasing energy demands and entering semi-torpidity [65], though no research has been conducted on the overwintering physiology of any Syngnathidae species. Nevertheless, localized overwintering in deeper waters may be an important component of *H. erectus*' life history and may also account for their winter absence in estuaries of the warm-temperate eastern Floridian Peninsula (i.e., South-Atlantic).

Conclusions

Overall, our results demonstrate the utility of supplementing life history information with population genomic data when a small number of unlinked genetic loci may be insufficient to discern the range of persistence in difficult-to-observe fishes. Currently, the IUCN (World Conservation Union) Red List categorizes *H. erectus* as “vulnerable” based on it being commonly collected as by-catch and sold by trawl fishermen to supply the aquarium trade [66]. Our results, throughout an extensive range of this species distribution, will help inform conservation, as well as captive breeding efforts, by strongly supporting northern Atlantic seahorses as a genetically distinct subpopulation. More broadly, because genomic data effectively samples a multitude of ancestors, even with a small number of sampled individuals, the approach taken in our study shows the promise of genomic data to infer population genetic structure in rare and/or difficult to obtain species.

Supporting Information

S1 Table. Raw reads, filtered, analyzed reads, and NCBI SRA accession numbers per individual.

(PDF)

S1 Methods. Additional details on sNMF CEC values, sNMF and STRUCTURE comparison, and isolation-by-distance methods.

(PDF)

S2 Methods. Comparison of NOAA Long-term bottom trawl survey of the Mid-Atlantic Bight (i.e., Virginia Province) between *Hippocampus erectus* and *Syngnathus fuscus*.

(PDF)

Acknowledgments

Thank you to the Sackler Institute for Comparative Genomics, American Museum of Natural History and Dr. Rob DeSalle for laboratory space and support; Stephen Harris and Tyler Joseph for assistance with data analysis; Clay Small (University of Oregon), N. Dunham (Florida Fish and Wildlife), T. Tuckey (Virginia Institute of Marine Science), T. Gardner (Atlantis Aquarium, NY), T. M. Grothues (Rutgers University, NJ) and The River Project (riverprojectnyc.org) for help with fish collections and/or providing samples.

Author Contributions

Conceived and designed the experiments: JTB JW MJH. Performed the experiments: JTB. Analyzed the data: JTB JDR. Contributed reagents/materials/analysis tools: JTB MJH. Wrote the paper: JTB JW JDR MJH.

References

1. Briggs PT, Waldman JR (2002) Annotated list of fishes reported from the marine waters of New York. *Northeast Nat* 9: 47–80.
2. Curran M (1989) Occurrence of tropical fishes in New England waters. *AAUS*. 71–82.
3. Teixeira R, Musick J (2001) Reproduction and food habits of the lined seahorse, *Hippocampus erectus* (Teleostei: Syngnathidae) of Chesapeake Bay, Virginia. *Rev Bras Biol* 61: 79–90. PMID: [11340465](https://pubmed.ncbi.nlm.nih.gov/11340465/)
4. Milstein CB, Thomas DL (1976) Fishes New or Uncommon to the New Jersey Coast. *Chesap Sci* 17: 198.
5. Briggs JC, Bowen BW (2012) A realignment of marine biogeographic provinces with particular reference to fish distributions. *J Biogeogr* 39: 12–30. doi: [10.1111/j.1365-2699.2011.02613.x](https://doi.org/10.1111/j.1365-2699.2011.02613.x)

6. McCartney MA, Burton ML, Lima TG (2013) Mitochondrial DNA differentiation between populations of black sea bass (*Centropristis striata*) across Cape Hatteras, North Carolina (USA). *J Biogeogr* 40: 1386–1398. doi: [10.1111/jbi.12103](https://doi.org/10.1111/jbi.12103)
7. McBride RS (2014) Managing a Marine Stock Portfolio: Stock Identification, Structure, and Management of 25 Fishery Species along the Atlantic Coast of the United States. *North Am J Fish Manag* 34: 710–734. doi: [10.1080/02755947.2014.902408](https://doi.org/10.1080/02755947.2014.902408)
8. Grosberg C, Cunningham C (2001) Genetic structure in the sea: from populations to communities. In: Bertness M, Gaines S, Hay M, editors. *Marine Community Ecology*. Sinauer Associates. pp. 61–84.
9. Fish MP and Mowbray MH (1970) *Sounds of western North Atlantic fishes*. The Johns Hopkins University Press.
10. Casazza TL, Ross SW (2008) Fishes associated with pelagic *Sargassum* and open water lacking *Sargassum* in the Gulf Stream off North Carolina. *Fish Bull* 106: 348–363.
11. Able K, Fahay M (1998) *Ecology of Estuarine Fishes: Temperate Waters of the Western North Atlantic*. John Hopkins University Press.
12. Howell P, Auster PJ (2012) Phase Shift in an Estuarine Finfish Community Associated with Warming Temperatures. *Mar Coast Fish* 4: 481–495. doi: [10.1080/19425120.2012.685144](https://doi.org/10.1080/19425120.2012.685144)
13. Lowe WH, Allendorf FW (2010) What can genetics tell us about population connectivity? *Mol Ecol* 19: 3038–3051. doi: [10.1111/j.1365-294X.2010.04688.x](https://doi.org/10.1111/j.1365-294X.2010.04688.x) PMID: [20618697](https://pubmed.ncbi.nlm.nih.gov/20618697/)
14. Hare MP, Nunney L, Schwartz MK, Ruzzante DE, Burford M, et al. (2011) Understanding and estimating effective population size for practical application in marine species management. *Conserv Biol* 25: 438–449. doi: [10.1111/j.1523-1739.2010.01637.x](https://doi.org/10.1111/j.1523-1739.2010.01637.x) PMID: [21284731](https://pubmed.ncbi.nlm.nih.gov/21284731/)
15. Pringle JM, Wares JP (2007) Going against the flow: Maintenance of alongshore variation in allele frequency in a coastal ocean. *Mar Ecol Prog Ser* 335: 69–84. doi: [10.3354/meps335069](https://doi.org/10.3354/meps335069)
16. Martinez-Solano I, Gonzalez EG (2008) Patterns of gene flow and source-sink dynamics in high altitude populations of the common toad *Bufo bufo* (Anura: Bufonidae). *Biol J Linn Soc* 95: 824–839. doi: [10.1111/j.1095-8312.2008.01098.x](https://doi.org/10.1111/j.1095-8312.2008.01098.x)
17. Mobley KB, Small CM, Jones AG (2011) The genetics and genomics of Syngnathidae: pipefishes, seahorses and seadragons. *J Fish Biol* 78: 1624–1646. doi: [10.1111/j.1095-8649.2011.02967.x](https://doi.org/10.1111/j.1095-8649.2011.02967.x) PMID: [21651520](https://pubmed.ncbi.nlm.nih.gov/21651520/)
18. Teske PR, Cherry MI, Matthee CA (2003) Population genetics of the endangered Knysna seahorse, *Hippocampus capensis*. *Mol Ecol* 12: 1703–1715. doi: [10.1046/j.1365-294X.2003.01852.x](https://doi.org/10.1046/j.1365-294X.2003.01852.x) PMID: [12803625](https://pubmed.ncbi.nlm.nih.gov/12803625/)
19. Teske P, Hamilton H, Palsbøll P, Choo C, Gabr H, et al. (2005) Molecular evidence for long-distance colonization in an Indo-Pacific seahorse lineage. *Mar Ecol Prog Ser* 286: 249–260.
20. Lourie SA, Green DM, Vincent ACJ (2005) Dispersal, habitat differences, and comparative phylogeography of Southeast Asian seahorses (Syngnathidae: *Hippocampus*). *Mol Ecol* 14: 1073–1094. doi: [10.1111/j.1365-294X.2005.02464.x](https://doi.org/10.1111/j.1365-294X.2005.02464.x) PMID: [15773937](https://pubmed.ncbi.nlm.nih.gov/15773937/)
21. Lourie SA, Vincent ACJ (2004) A marine fish follows Wallace's Line: the phylogeography of the three-spot seahorse (*Hippocampus trimaculatus*, Syngnathidae, Teleostei) in Southeast Asia. *J Biogeogr* 31: 1975–1985. doi: [10.1111/j.1365-2699.2004.01153.x](https://doi.org/10.1111/j.1365-2699.2004.01153.x)
22. Boehm JT, Woodall L, Teske PR, Lourie SA., Baldwin C, et al. (2013) Marine dispersal and barriers drive Atlantic seahorse diversification. *J Biogeogr* 40: 1839–1849. doi: [10.1111/jbi.12127](https://doi.org/10.1111/jbi.12127) PMID: [17320419](https://pubmed.ncbi.nlm.nih.gov/17320419/)
23. Sousa V, Hey J (2013) Understanding the origin of species with genome-scale data: modelling gene flow. *Nat Rev Genet* 14: 404–414. doi: [10.1038/nrg3446](https://doi.org/10.1038/nrg3446) PMID: [23657479](https://pubmed.ncbi.nlm.nih.gov/23657479/)
24. Corander J, Majander KK, Cheng L, Merilä J (2013) High degree of cryptic population differentiation in the Baltic Sea herring *Clupea harengus*. *Mol Ecol* 22: 2931–2940. doi: [10.1111/mec.12174](https://doi.org/10.1111/mec.12174) PMID: [23294045](https://pubmed.ncbi.nlm.nih.gov/23294045/)
25. Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G (2011) Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Mol Ecol Resour* 11 Suppl 1: 117–122. doi: [10.1111/j.1755-0998.2010.02967.x](https://doi.org/10.1111/j.1755-0998.2010.02967.x) PMID: [21429168](https://pubmed.ncbi.nlm.nih.gov/21429168/)
26. Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson E, et al. (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet* 6: e1000862. doi: [10.1371/journal.pgen.1000862](https://doi.org/10.1371/journal.pgen.1000862) PMID: [20195501](https://pubmed.ncbi.nlm.nih.gov/20195501/)
27. Deagle BE, Jones FC, Absher DM, Kingsley DM, Reimchen TE (2013) Phylogeography and adaptation genetics of stickleback from the Haida Gwaii archipelago revealed using genome-wide single nucleotide polymorphism genotyping. *Mol Ecol* 22: 1917–1932. doi: [10.1111/mec.12215](https://doi.org/10.1111/mec.12215) PMID: [23452150](https://pubmed.ncbi.nlm.nih.gov/23452150/)

28. Catchen J, Bassham S, Wilson T, Currey M, O'Brien C, et al. (2013) The population structure and recent colonization history of Oregon threespine stickleback determined using restriction-site associated DNA-sequencing. *Mol Ecol* 22: 2864–2883. doi: [10.1111/mec.12330](https://doi.org/10.1111/mec.12330) PMID: [23718143](https://pubmed.ncbi.nlm.nih.gov/23718143/)
29. Wagner CE, Keller I, Wittwer S, Selz OM, Mwaiko S, et al. (2013) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Mol Ecol* 22: 787–798. doi: [10.1111/mec.12023](https://doi.org/10.1111/mec.12023) PMID: [23057853](https://pubmed.ncbi.nlm.nih.gov/23057853/)
30. Li S, Jakobsson M (2012) Estimating demographic parameters from large-scale population genomic data using Approximate Bayesian Computation. *BMC Genet* 13: 22. doi: [10.1186/1471-2156-13-22](https://doi.org/10.1186/1471-2156-13-22) PMID: [22453034](https://pubmed.ncbi.nlm.nih.gov/22453034/)
31. Felsenstein J (2006) Accuracy of coalescent likelihood estimates: do we need more sites, more sequences, or more loci? *Mol Biol Evol* 23: 691–700. doi: [10.1093/molbev/msj079](https://doi.org/10.1093/molbev/msj079) PMID: [16364968](https://pubmed.ncbi.nlm.nih.gov/16364968/)
32. Robinson J, Bunnefeld L, Hearn J, Stone G, Hickerson MJ (2014) ABC inference of multi-population divergence with admixture from un-phased population genomic data. *Mol Ecol* 18: 4458–4471. doi: [10.1111/mec.12881](https://doi.org/10.1111/mec.12881) PMID: [25113024](https://pubmed.ncbi.nlm.nih.gov/25113024/)
33. Gronau I, Hubisz MJ, Gulko B, Danko CG, Siepel A (2011) Bayesian inference of ancient human demography from individual genome sequences. *Nat Genet* 43: 1031–1034. doi: [10.1038/ng.937](https://doi.org/10.1038/ng.937) PMID: [21926973](https://pubmed.ncbi.nlm.nih.gov/21926973/)
34. Lohse K, Harrison RJ, Barton NH (2011) A general method for calculating likelihoods under the coalescent process. *Genetics* 189: 977–987. doi: [10.1038/ng.937](https://doi.org/10.1038/ng.937) doi: [10.1534/genetics.111.129569](https://doi.org/10.1534/genetics.111.129569) PMID: [21900266](https://pubmed.ncbi.nlm.nih.gov/21900266/)
35. Hearn J, Stone GN, Bunnefeld L, Nicholls JA, Barton NH, et al. (2014) Likelihood-based inference of population history from low coverage de novo genome assemblies. *Mol Ecol* 23: 198–211. doi: [10.1111/mec.12578](https://doi.org/10.1111/mec.12578) PMID: [24188568](https://pubmed.ncbi.nlm.nih.gov/24188568/)
36. Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, et al. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3: e3376. doi: [10.1371/journal.pone.0003376](https://doi.org/10.1371/journal.pone.0003376) PMID: [18852878](https://pubmed.ncbi.nlm.nih.gov/18852878/)
37. Lozier JD (2014) Revisiting comparisons of genetic diversity in stable and declining species: assessing genome-wide polymorphism in North American bumble bees using RAD sequencing. *Mol Ecol* 23: 788–801. doi: [10.1111/mec.12636](https://doi.org/10.1111/mec.12636) PMID: [24351120](https://pubmed.ncbi.nlm.nih.gov/24351120/)
38. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10: R25. doi: [10.1186/gb-2009-10-3-r25](https://doi.org/10.1186/gb-2009-10-3-r25) PMID: [19261174](https://pubmed.ncbi.nlm.nih.gov/19261174/)
39. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–2079. doi: [10.1093/bioinformatics/btp352](https://doi.org/10.1093/bioinformatics/btp352) PMID: [19505943](https://pubmed.ncbi.nlm.nih.gov/19505943/)
40. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, et al. (2011) The variant call format and VCFtools. *Bioinformatics* 27: 2156–2158. doi: [10.1093/bioinformatics/btr330](https://doi.org/10.1093/bioinformatics/btr330)
41. R Development Core Team (2004) R: A Language and Environment for Statistical Computing.
42. Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genet* 2: e190. doi: [10.1371/journal.pgen.0020190](https://doi.org/10.1371/journal.pgen.0020190)
43. Frichot E, Mathieu F, Trouillon T, Bouchard G, François O (2014) Fast and efficient estimation of individual ancestry coefficients. *Genetics* 196: 973–983. doi: [10.1534/genetics.113.160572](https://doi.org/10.1534/genetics.113.160572) PMID: [24496008](https://pubmed.ncbi.nlm.nih.gov/24496008/)
44. Pritchard JK, Stephens M, Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155: 945–959.
45. Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19: 1655–1664. doi: [10.1101/gr.094052.109](https://doi.org/10.1101/gr.094052.109) PMID: [19648217](https://pubmed.ncbi.nlm.nih.gov/19648217/)
46. Harris SE, O'Neill RJ, Munshi-South J (2014) Transcriptome resources for the white-footed mouse (*Peromyscus leucopus*): new genomic tools for investigating ecologically divergent urban and rural populations. *Mol Ecol Resour: Early View*. doi: [10.1111/1755-0998.12301](https://doi.org/10.1111/1755-0998.12301)
47. Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures from genome-wide allele frequency data. *PLOS Genetics* 8: 28. doi: [10.1371/journal.pgen.1002967](https://doi.org/10.1371/journal.pgen.1002967) PMID: [23166502](https://pubmed.ncbi.nlm.nih.gov/23166502/)
48. Gompert Z, Lucas LK, Buerkle CA, Forister ML, Fordyce JA, et al. (2014) Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. *Mol Ecol* 23: 4555–4573. doi: [10.1111/mec.12811](https://doi.org/10.1111/mec.12811) PMID: [24866941](https://pubmed.ncbi.nlm.nih.gov/24866941/)
49. Felsenstein J (1981) Evolutionary trees from gene frequencies and quantitative characters: finding maximum likelihood estimates. *Evolution* 35: 1229–1242.
50. Weir BS, Cockerhan CC (1984) Estimation of gene flow from F-statistics. *Evolution* 38: 1358–1370.

51. Willing E-M, Dreyer C, van Oosterhout C (2012) Estimates of genetic differentiation measured by F_{ST} do not necessarily require large sample sizes when using many SNP markers. *PLoS One* 7: e42649. doi: [10.1371/journal.pone.0042649](https://doi.org/10.1371/journal.pone.0042649) PMID: [22905157](https://pubmed.ncbi.nlm.nih.gov/22905157/)
52. Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, et al. (2008) Genes mirror geography within Europe. *Nature* 456: 98–101. doi: [10.1038/nature07331](https://doi.org/10.1038/nature07331) PMID: [18758442](https://pubmed.ncbi.nlm.nih.gov/18758442/)
53. Jensen J, AJ B, Kelley S (2005) Isolation by distance, web service. *BMC Genet* 6: <http://ibdws.sdsu.edu/>.
54. Gazave E, Chang D, Clark AG, Keinan A (2013) Population growth inflates the per-individual number of deleterious mutations and reduces their mean effect. *Genetics* 195: 969–978. doi: [10.1534/genetics.113.153973](https://doi.org/10.1534/genetics.113.153973) PMID: [23979573](https://pubmed.ncbi.nlm.nih.gov/23979573/)
55. Cronin TM, Szabo BJ, Ager TA, Hazel JE, Owens JP (1981) Quaternary climates and sea levels of the U.S. Atlantic coastal plain. *Science* 211: 233–240. doi: [10.1126/science.211.4479.233](https://doi.org/10.1126/science.211.4479.233) PMID: [17748008](https://pubmed.ncbi.nlm.nih.gov/17748008/)
56. Bratton JF, Colman SM, Thieler ER, Seal RR (2002) Birth of the modern Chesapeake Bay estuary between 7.4 and 8.2 ka and implications for global sea-level rise. *Geo-Marine Lett* 22: 188–197. doi: [10.1007/s00367-002-0112-z](https://doi.org/10.1007/s00367-002-0112-z)
57. Short F, Carruthers T, Dennison W, Waycott M (2007) Global seagrass distribution and diversity: A bio-regional model. *J Exp Mar Bio Ecol* 350: 3–20. doi: [10.1016/j.jembe.2007.06.012](https://doi.org/10.1016/j.jembe.2007.06.012)
58. Lourie SA, Foster SJ, Cooper EWT, Vincent ACJ (2004) A guide to the identification of seahorses. Project Seahorse and TRAFFIC North America.
59. Wenner CA, Sedberry GR (1989) Species composition, distribution, and relative abundance of fishes in the coastal habitat off the southeastern United States. NOAA Technical Report NMFS 79: 1–78.
60. Mobley KB, Small CM, Jue NK, Jones AG (2010) Population structure of the dusky pipefish (*Syngnathus floridae*) from the Atlantic and Gulf of Mexico, as revealed by mitochondrial DNA and microsatellite analyses. *J Biogeogr* 37: 1363–1377. doi: [10.1111/j.1365-2699.2010.02288.x](https://doi.org/10.1111/j.1365-2699.2010.02288.x)
61. Hice LA, Duffy TA, Munch SB, Conover DO (2012) Spatial scale and divergent patterns of variation in adapted traits in the ocean. *Ecol Lett* 15: 568–575. doi: [10.1111/j.1461-0248.2012.01769.x](https://doi.org/10.1111/j.1461-0248.2012.01769.x) PMID: [22462779](https://pubmed.ncbi.nlm.nih.gov/22462779/)
62. Foster SJ, Vincent ACJ (2004) Life history and ecology of seahorses: implications for conservation and management. *J Fish Biol* 65: 1–61. doi: [10.1111/j.0022-1112.2004.00429.x](https://doi.org/10.1111/j.0022-1112.2004.00429.x)
63. Lazzari MA, Able KW (1990) Northern pipefish, *Syngnathus fuscus*, occurrences over the Mid-Atlantic Bight continental shelf: evidence of seasonal migration. *Environ Biol Fishes* 27: 177–185. doi: [10.1007/BF00001671.x](https://doi.org/10.1007/BF00001671.x)
64. Wicklund R, Wilk S, Ogren L (1968) Observations on wintering locations of the northern pipefish and spotted seahorse. *Underw Nat* 5: 26–28.
65. Ultsch G (1989) Ecology and physiology of hibernation and overwintering among freshwater fishes, turtles, and snakes. *Biol Rev* 64: 435–515. doi: [10.1111/j.1469-185X.1989.tb00683.x](https://doi.org/10.1111/j.1469-185X.1989.tb00683.x)
66. Dias TL, Rosa I, Baum JK (2002) Threatened fishes of the world: *Hippocampus erectus* Perry, 1810 (Syngnathidae). *Environ Biol Fishes* 65: 326. doi: [10.1023/A:1020539222587](https://doi.org/10.1023/A:1020539222587)
67. Tyberghein L, Verbruggen H, Pauly K, Troupin C, Mineur F, et al. (2012) Bio-ORACLE: a global environmental dataset for marine species distribution modelling. *Glob Ecol Biogeogr* 21: 272–281. doi: [10.1111/j.1466-8238.2011.00656.x](https://doi.org/10.1111/j.1466-8238.2011.00656.x)