Population genomic dynamics of mesopelagic lanternfishes *Diaphus dumerilii, Lepidophanes guentheri,* and *Ceratoscopelus warmingii* (Family: Myctophidae) in the Gulf of Mexico

Andrea M. Bernard[a] (andrbern@nova.edu)

Kimberly A. Finnegan[a] (atwaterk5@gmail.com)

Tracey T. Sutton[b] (tsutton1@nova.edu)

Ron Eytan[c] (eytanr@tamug.edu)

Max Weber[c] (mweb68@hotmail.com)

Mahmood S. Shivji[a] (mahmood@nova.edu)

Highlights

First genomic characterization of Atlantic deep-sea lanternfishes

Lanternfishes showed low genetic diversity despite likely large populations

Lanternfishes showed signals of high inbreeding despite likely large populations

Genetic division exists between Gulf of Mexico and western North Atlantic lanternfish

Population genomic dynamics of mesopelagic lanternfishes *Diaphus dumerilii, Lepidophanes guentheri,* and *Ceratoscopelus warmingii* (Family: Myctophidae) in the Gulf of Mexico

Andrea M. Bernard[a] (andrbern@nova.edu)

Kimberly A. Finnegan[a] (atwaterk5@gmail.com)

Tracey T. Sutton[b] (tsutton1@nova.edu)

Ron Eytan[c] (eytanr@tamug.edu)

Max Weber[c] (mweb68@hotmail.com)

Mahmood S. Shivji[a] (mahmood@nova.edu)

[a]Save Our Seas Foundation Shark Research Center and Guy Harvey Research Institute, Nova Southeastern University, 8000 North Ocean Drive, Dania Beach, Florida, USA 33004

[b]Nova Southeastern University, 8000 North Ocean Drive, Dania Beach, Florida, USA 33004

[c]Texas A&M University at Galveston, 1001 Texas Clipper Road, Galveston, Texas, USA 77550

Corresponding author:

Mahmood S. Shivji (mahmood@nova.edu), Save Our Seas Foundation Shark Research Center and Guy Harvey Research Institute, Nova Southeastern University, 8000 North Ocean Drive, Dania Beach, Florida, USA 33004.

Abstract

Assessing the impacts of the Deepwater Horizon oil spill (DWHOS) on deep-sea fish assemblages of the Gulf of Mexico (GOM) has been hindered by an absence of baseline (pre-spill) data concerning the population genetic dynamics of these fishes. The lanternfishes (Myctophidae) are a speciose, yet understudied, taxonomic group, that comprise a significant portion of the global deep-sea biomass, making them integral members of meso- and bathy-pelagic food webs. Herein, we used a genomic approach (double digest restriction site associated DNA sequencing) to investigate the temporal genetic dynamics of three species of lanternfishes within the northern GOM in the region of the oil spill: *Ceratoscopelus warmingii* (N = 65, SNP = 1804), *Diaphus dumerilii* (N = 42, SNP = 2577), and *Lepidophanes guentheri* (N = 44, SNP = 3462). Fishes were sampled in 2011, and then again in 2015, and 2016, and genotyped using single nucleotide polymorphisms (SNP) markers. An additional 22 *C. warmingii* samples collected in 2014 from the western North Atlantic were used to assess genetic connectivity between these two regions. Overall, all three species were characterized by low levels of genetic diversity and high inbreeding coefficients, and within two of the three species (*C. warmingii* and *L. guentheri*), little (if any) evidence of population genetic structure was found within northern GOM waters. Conversely, significant intra-GOM genetic population structure was found for *D. dumerilii*, highlighting the need for a more robust population genetic survey of these fishes within the GOM to understand how populations of these fishes may respond to future environmental perturbations. In addition, significant genetic population structure was also found between sub-populations of *C. warmingii* from the northern GOM and western North Atlantic; with evidence of a mixture of two genetic populations co-occurring within the western North Atlantic. Given the potential for future environmental perturbations caused by expanding oil and

gas extraction and climate change on GOM mesopelagic communities, rigorous population genetic assessments are required to understand the population dynamics of fishes and to safeguard the genetic diversity and resilience of this functionally important group of deep-sea fishes.

Keywords: Gulf of Mexico, single nucleotide polymorphism, genetic connectivity, genetic diversity, Myctophid

1. Introduction

Eukaryotic taxa inhabiting the deep-pelagic ocean—earth's largest ecosystem—represent enormous biomass, high diversity, novel evolutionary adaptations, and play key functional roles in food web dynamics of the global oceans (Turner et al., 2009; Widder, 2010; Irigoien et al., 2014; Porter et al., 2016; Sutton et al., 2017; Gaither et al., 2018). These taxa, including the mesopelagic fishes, have been historically understudied due to the high cost and logistical difficulties of investigator access to these ecosystems (Robinson et al., 2009; Webb et al., 2010; St. John et al., 2016). Consequently, despite the vast biomass and integral ecological importance of mesopelagic fishes in oceanic food webs, major knowledge gaps and uncertainties persist regarding the structure and state of their assemblages (St. John et al., 2016; Hidalgo and Browman, 2019), including their genetic population dynamics and diversity.

The lanternfishes (Myctophiformes: Myctophidae) are a circumglobal family of mesopelagic fishes known for the bioluminescent photophores that occur in species-specific mosaics along their heads and bodies (de Busserolles and Marshall, 2017). This species rich family (ca. 254 known species; Fricke et al., 2021) forms a dominant proportion of the world's

mesopelagic fish biomass (Gjøsæter and Kawaguchi, 1980; Catul et al., 2011). In general, lanternfishes are strong diel vertical migrators (Watanabe et al., 1999; D'Elia et al., 2016), and coupled with their high biomass, form an essential driver of organic matter transport throughout the water column via active (diel vertical migration) and passive (sinking of organic detritus) processes (Davison et al., 2013; Sutton et al., 2013). Despite a central ecological role of lanternfishes in marine ecosystems and the growing interest in commercial exploitation of these and other mesopelagic fishes (Catul et al., 2011; Hidalgo and Browman, 2019), fewer than a dozen investigations employing genetic techniques of any type have been performed on this speciose group (e.g., Suneetha and Salvanes, 2001; Kojima et al., 2009; Florence et al., 2010; Kristoffersen and Salvanes, 2009; Gordeeva, 2011; 2014; Van de Putte et al., 2012; Christiansen et al., 2018; Denton, 2018; Martin et al., 2018; Batta-Lona et al. 2019). Of these studies, only a handful have specifically assessed myctophid genetic diversity and connectivity, finding genetic homogeneity across large, horizontal, oceanic distances in some species (Kojima et al., 2009; Van de Putte et al., 2012), and significant genetic differentiation at much more reduced horizontal spatial scales in others (Gordeeva, 2011; 2014). The limited information particularly on the population genetic statuses and dynamics of lanternfishes leaves little insight into their evolution and hinders assessment of impacts caused by anticipated future environmental perturbations.

Anthropogenic stressors to the deep-pelagic oceans include climate change, ocean acidification, noise, pollution, and chemical spills (Robinson et al., 2009; Ramirez-Llodra et al., 2011; Choy et al., 2019; Drazen et al., 2020). The Deepwater Horizon Oil Spill (DWHOS) at 1500 m depth in the northern Gulf of Mexico (GOM) between April and July, 2010 presented a major environmental perturbation to deep-ocean marine populations and potentially their genetic

dynamics (Portnoy et al., 2020). However, quantifying the impacts of the DWHOS or any future anthropogenic threats on mesopelagic fish populations is complicated by the fact that baseline (in this case, pre-spill) data on genetic diversity, genetic population structure, and demographic spatial linkages of these fishes (i.e., information directly allied to the resilience and evolutionary potential and trajectory of species) does not exist. Since deep-water petroleum extraction activities in this region are not only common but expanding (Murawski et al., 2020), knowledge of the population genetic dynamics of mesopelagic fishes here will be useful for gaining a basic understanding of the population biology of these understudied fishes generally and assessing impacts of future anticipated anthropogenic events to the deep-pelagic ecosystem of the GOM.

In this study, we undertook analysis of the population genomic dynamics of three species of deep-sea lanternfishes (*Diaphus dumerilii*, *Lepidophanes guentheri*, and *Ceratoscopelus warmingii*) (Fig. 1A) that are numerically dominant in the northern GOM (Gartner et al., 1987). While *D. dumerilii* and *L. guentheri* are widespread in the Atlantic, *C. warmingii* occurs circumglobally. These lanternfish species possess short life spans (1-2 years; Gartner, 1991; 1993), relatively long larval durations of 20-35 days (Gartner, 1991; Takagi et al., 2006), and are diel vertical migrators (Gartner et al., 1987; Milligan and Sutton, 2020). While in general, little is known about the vertical diel migratory patterns of these fishes, surveys within the GOM and Southwestern Tropical Atlantic indicate that *C. warmingii* and *L. guentheri* are predominant between 400-1000 m in the day, while *D. dumerili* occurs at shallower depths (200-600 m) (Gartner et al., 1987; Eduardo et al., 2021). At night, all three species occupy depths of 0-100 m. However, some evidence of non-migrators, consisting mainly of juveniles, also exists within the species *C. warmingii and L. guentheri* (Gartner et al., 1987). Their life history features, together with inter-species variations in some other biological parameters, including frequency and

duration of reproduction (Gartner, 1993) and relative primary depth distribution (Gartner et al., 1987; Milligan and Sutton, 2020; Eduardo et al., 2021), make them potentially useful models for studying the comparative, multi-generational, temporal population genetic dynamics of mesopelagic fishes inhabiting the GOM. Here, we applied a genomics approach using single nucleotide polymorphism (SNP) markers (which provide a high-resolution view of the genetic properties of populations; see Helyar et al., 2011 for a review of SNP markers), to address the following objectives: (1) establish a baseline of temporal and spatial genetic diversity and population structure in these three species of lanternfishes within the northern GOM across multiple generations and to test the hypothesis that genomic selection may be driving patterns of any identified population genetic structure, and (2) investigate the genetic connectivity of *C. warmingii* between the northern GOM and western North Atlantic (WNA) with the aim of understanding population linkages and resilience of this important deep-sea community member across large geographic scales.

2. Methods

2.1. Sample collection

To assess baseline population genetic dynamics of lanternfishes within the GOM and WNA, specimens of the three lanternfish species—*Diaphus dumerillii*, *Lepidophanes guentheri*, and *Ceratoscopelus warmingii*—were collected from the northern GOM and WNA between 2011 and 2016 as part of three research initiatives (Table 1, Fig. 1B): The 2011 samples were collected as part of the NOAA supported Offshore Nekton Sampling and Analysis Program cruises conducted one year after final capping of the DWHOS oil well (see Sutton et al. (2020) for detailed sampling information). The 2015 and 2016 samples were collected by the Deep Pelagic

Nekton Dynamics of the Gulf of Mexico (DEEPEND) Consortium in our effort to characterize the deep (200-1500 m) oceanic ecosystem and dynamics of the northern GOM (see Cook et al. (2020) for detailed sampling information). To assess lanternfish spatial linkages, an additional 24 samples (*C. warmingii* only) were collected in 2014 from the WNA in the vicinity of Bear Seamount, a deep-sea, inactive volcano that is part of the New England Seamount Chain (NES) (Moore et al., 2004), by the NOAA National Marine Fisheries Service (NMFS) Northeast Fisheries Science Center as part of their Deepwater Systematics Survey (Table 1, Fig. 1B). Following collection, all specimens were morphologically identified at sea and tissues stored in 95-99% ethanol.

2.2. Genomic DNA extraction and DNA Barcoding

Genomic DNA was extracted from preserved tissue by using either DNeasy Blood & Tissue Kits (QIAGEN Inc.) or the modified lysis extraction method of Wilson et al. (2007). In-house modifications to the latter extraction protocol included: (i) lysis of each tissue sample in 180-µL of the QIAGEN Buffer ATL with 20-µL of Proteinase K, and (ii) final suspension of the DNA pellet in 75-µL of the QIAGEN Buffer AE. Genomic DNA extractions obtained via the modified lysis approach were subsequently purified using Agencourt AMPure XP beads (Beckman Coulter Inc.) to remove contaminants and were eluted to a final volume of 75-µL (QIAGEN Buffer EB). For both DNA isolation methods, 4-µL of RNAse A (QIAGEN) was added to each sample following initial tissue lysis. All genomic DNA extracts were quantified using a Qubit 3 Fluorometer (Invitrogen Inc.).

To confirm the species identity of all samples, we sequenced the 5' end of the mitochondrial Cytochrome *c* oxidase subunit I gene (COI) (i.e., the DNA barcode gene; see

Steinke and Hanner (2010) for general methods and Suppl. Mat. A, and Tables S1 and S2). Sequences were deposited in the Barcode of Life Data System (BOLD, http://www.boldsystems.org) and the National Center for Biotechnology Information (NCBI) databases. The overall species-level maximum pairwise distance between individual COI sequences from each of the morphologically identified lanternfishes was estimated using the BOLD v4 Sequence Analysis module (Distance Summary). Species-level (pooled temporal and spatial samples for each species) haplotype (h) and nucleotide (π) diversity estimates were generated using the program Arlequin 3.5 (Excoffier and Lischer, 2010) and input files generated using Fabox 1.61 (Villesen, 2007).

2.3. Single nucleotide polymorphism laboratory methods

Double digest restriction-site associated DNA (ddRAD) libraries were generated following Peterson et al.'s (2012) protocol. For each fish sample, 200-500 ng of genomic DNA was digested with the restriction enzymes SphI (Life Technologies Inc.) and MluC1 (New England Biolabs Inc.) in 50-µL reaction volumes. Samples were incubated for three hours at 37°C and digests were purified using Agencourt AMPure XP beads. Following digestion, individual barcode (P1; Sigma-Aldrich Inc.) and universal adapters (P2; Sigma-Aldrich Inc.) were ligated to the SphI and MluC1 restriction enzyme cut sites, respectively. A total of eight unique P1 adapters were employed, allowing the pooling of eight samples per final indexed library. Adapter ligation occurred in reaction volumes of 40-µL and contained: 4-µl of 10X T4 DNA Ligase Buffer (Life Technologies Inc.), 0.3-µL of 5U/µL T4 DNA Ligase (Life Technologies Inc.), 2-µL each of 4 µM P1 and P2 adapters, and equimolar amounts of digested samples. Reactions were incubated in a Veriti thermal cycler (Applied Biosystems Inc.) at 23°C for 30 minutes

(mins), 65°C for 10 mins, and 23°C for 1.5 mins. Final temperature ramping of samples from 65°C to 23°C did not exceed 2°C/1.5 mins. Following adapter ligation, samples were pooled into groups of eight and purified using Agencourt AMPure XP beads. Size selection of fragments was performed using a Pippin Prep DNA Size Selection System (Sage Science Inc.) targeting fragments 300-400 bp in length.

Final library preparation for Illumina sequencing was performed in 20-µL reaction volumes using New England Biolabs (NEB) Phusion® High-Fidelity PCR Kits, and contained: 4-µL 5X HF Buffer, 2-µL of 20 µM universal Illumina flow-cell adapter primers, 2-µL of 20 µM Illumina Index primer, 0.4-µL of the dNTP Mix, 0.4-µL of Phusion High-Fidelity DNA Taq Polymerase, and 20 ng of library-pooled DNA. Polymerase Chain Reactions (PCRs) were performed in a Veriti thermal cycler using the following thermal profile: initial denaturation at 98°C for 30 seconds (sec), followed by 8-12 cycles of 98°C for 10 sec, 62°C for 30 sec, and 72°C for 30 sec, and a final extension of 72°C for 10 mins. Following thermal cycling, amplifications were purified using Agencourt AMPure XP beads. To ensure quality and the proper size distribution of resultant fragments of individual library pools, each library was examined using an Agilent TapeStation 2200 (Agilent Technologies Inc.). Individual Illumina Indexed libraries were pooled in equal molar ratios and normalized to 25 nM. Paired-end (PE) sequencing (2 X 150 PE) was performed by GENEWIZ (GENEWIZ Inc.) on four lanes of an Illumina HiSeq 4000.

2.4.1. Bioinformatic analysis: raw read quality filtering and Stacks parameter optimization

Prior to data analysis, the quality of raw sequence reads was assessed using the program FastQC 0.11.8 (Andrews, 2010). Processing, filtering and downstream *de novo* SNP discovery of

Forward Illumina sequence reads was performed using Catchen et al.'s (2013) Stacks 1.48 pipeline. The Stacks module *process_radtags* was used to demultiplex raw Illumina sequences, perform quality filtering [c, q, r (option: 2 nucleotides), E (option: phred33)], rescue barcodes and rad tags, and trim sequences to a final read length of 141 bp. To optimize the *de novo* assembly of loci using Stacks, we investigated a range of Stacks parameter values on a data subset of each species (n = 12) (see Rochette and Catchen, 2017; Paris et al., 2017). Per Rochette and Catchen (2017), we explored a range of values (1-9) for the parameters M (maximum number of nucleotide distance allowed between stacks within a single individual) and n (the number of nucleotide mismatches allowed between sample loci when building the locus catalog), where $M = n$ was fixed across runs, and the parameter m (minimum stack depth coverage) was held at 3. Resultant loci were filtered as outlined in using the Stacks *populations* module to include only those loci occurring in 80% of all sequenced individuals (r = 0.80) and those possessing a minimum average sequence depth of 10X. The optimal Stacks parameters for each species were identified as those values where both the number of loci shared by 80% of samples and the number of SNPs per locus stabilized.

2.4.2. Bioinformatic analysis: Stacks de novo pipeline and SNP quality filtering

Following *de novo* assembly optimization, short read sequences were aligned and assembled using the previously optimized set of species-specific Stacks parameter values (*D. dumerilii*; m = 3, $M = n = 5$; *L. guentheri*, $m = 3$, $M = n = 6$; *C. warmingii*, $m = 3$, $M = n = 6$). Catalogs were built using a subset of 24 samples per species and genotyping errors were corrected using the module *rxstacks* (conf_lim 0.25, prune_haplo, model_type bounded, bounded_high 0.1, and ln_lim -20). Using the *populations* module, raw variant datasets were filtered to retain only those

loci: (1) possessing a minimum read depth of 5X coverage; (2) that were genotyped within all *a priori* defined temporal or spatial sample collections [*D. dumerilii* and *L. guentheri* (p = 3; representing GOM temporal samples 2011, 2015, and 2016) or *C. warmngii* (p = 4; representing GOM temporal samples 2011, 2015, and 2016 and the WNA sample)]; (3) that were genotyped in a minimum of 70% of all individuals (r = 0.70); (4) possessed a minor allele frequency of at least 5% (min_maf = 0.05); and (5) possessed a maximum observed heterozygosity of 70% (max_obs_het = 0.7). To avoid the potential of linkage among loci, the 'write_random_snp' parameter was also employed to ensure only a single SNP was written per stack. Secondary filtering to include only those individuals genotyped at a minimum of 70% of all loci was performed using Tassel 5.2.28 (Bradbury et al., 2007).

Testing for conformation to Hardy-Weinberg equilibrium (HWE) within individual temporal or spatial collections was performed using the R (3.5.0; R Core Team, 2018) package pegas 0.11 (Paradis, 2010) and the 'hw.test' function. For each species, loci found to deviate from HWE expectations at $P < 0.01$ in more than one sample collection were discarded.

2.5. SNP outlier identification

To test the hypothesis that genomic selection may be driving patterns of population genetic structure across the three lanternfishes, we used the programs BayeScan 2.1 (Foll and Gaggiotti, 2008) and OutFLANK (Whitlock and Lotterhos, 2015) to detect candidate loci under selection. BayeScan runs possessed a 50, 000 iteration chain length (+ 50, 000 iteration burn-in), assumed a prior odds for the neutral model of 10, an FDR of 5%, and implemented the remaining default program settings. OutFLANK analyses were performed assuming default settings and those loci possessing q-values of less 0.05 were identified as candidates for selection. For each assessment

method, analyses were performed by grouping each lanternfish species by collection year (2011, 2015, 2016) and/or location (for *C. warmingii*: GOM pooled and Bear Seamount). Only those SNP loci identified by both outlier detection methods were deemed candidates for selection. Where candidate outliers were detected, a subset of the population genetic analyses (F_{ST} and Principal Component Analysis; see section 2.6 below) were repeated in triplicate to allow for analysis of: 1) the putatively neutral only SNP loci dataset (i.e., dataset with candidate outlier SNP loci removed), 2) the candidate outlier only SNP loci dataset (i.e., dataset with putatively neutral SNP loci removed), 3) all SNPs combined (putatively neutral and candidate outliers). All other population genetics analyses were performed solely for the combined (putatively neutral and candidate outliers) datasets.

2.6. Genetic diversity and population structure

Following HWE filtering and outlier identification, genetic diversity statistics of observed (H_O) and expected heterozygosities (H_E), percent polymorphic loci, number of private alleles, and population-level inbreeding coefficients (F_{IS}) were estimated using the Stacks module *populations*.

To assess the genetic population structure among the temporal collections (2011, 2015, and 2016) of the three GOM myctophid species and the two *C. warmingii* spatial collections (GOM pooled and Bear Seamount), we used the R packages strataG 2.0.2 (Archer et al., 2017; function ‘pairwiseTest’) and diveRsity (function: ‘diffCalc’; Keenan et al., 2013) for estimating overall and pairwise values of genetic differentiation [F_{ST} (Weir and Cockerham, 1984)]. Using strataG, statistical significance of F_{ST} estimates was determined using 1000 permutations and all *P*-values were adjusted using Benjamini and Hochberg’s (1995) False Discovery Rate (FDR)

correction (implemented in R; ‘p.adjust’ function). Significance of diveRsity estimates was determined via the estimation of 95% confidence intervals (CIs) by generating 1000 bootstraps across both loci and individuals. Population genetic structure was further assessed among temporal (GOM: 2011, 2015, and 2016) and spatial [*C. warmingii*: GOM (pooled) vs. WNA] collections by means of three different individual-based clustering methods: First, multivariate species-level principal component analyses (PCAs) were performed for each lanternfish dataset using the R package adegenet 2.1.2 (‘dudi.pca’; Jombart, 2008; Jombart and Ahmed, 2011), and by retaining the first three principal components. Second, the Maximum Likelihood program Admixture (Alexander et al., 2009) was adopted along with its cross-validation (CV) error approach, to determine the most appropriate numbers of genetic clusters (i.e., *K*) for each lanternfish dataset. Using default settings, values spanning $K = 1$-5 were assessed and the value of *K* with the lowest comparative CV was determined to be the most appropriate number of genetic clusters. And third, we used a *K*-means clustering approach, as implemented in the R package adegenet (‘find.clusters’), to determine the optimal number of genetic clusters (*K*) *de novo* in each lanternfish SNP dataset. The optimal *K* was selected by retaining all identified principal components, assuming a maximum value of $K = 10$, and implementing the Bayesian Information Criterion (BIC). The value of *K* with the lowest BIC was identified as the optimal value of *K*. Multiple clustering approaches were adopted herein, given the varying model assumptions and limitations of each method (see Alexander et al., 2009; Miller et al., 2020).

3. Results

3.1. COI barcodes and diversity

Of those samples included in the downstream SNP analyses, the COI barcodes showed high within species sequence similarity (GenBank Accession Numbers and sequence lengths in Suppl. Mat. Tables S3-S5). All individuals within species shared the same Barcode Index Number (BIN) (*D. dumerilii*, BIN = BOLD:AAF5612; *L. guentheri*, BIN = BOLD:AAD1806; *C. warmingii*, BIN = BOLD:AAC3131; Ratnasingham and Hebert, 2013), and possessed a maximum intraspecific pairwise distance of less than 2% (*D. dumerilli* = 0.78%, *L. guentheri* = 1.69%, *C. warmingii* = 1.69%). Overall species-level mitochondrial genetic diversity was high for all three lanternfishes: *D. dumerilii* (h = 0.9013, π = 0.0122; N = 42, 610-652 bp), *L. guentheri* (h = 0.8858, π = 0.0087; N = 44; 631-652 bp), *C. warmingii* (h = 0.9858; π = 0.0243; N = 87; 597-663 bp).

3.2.1. Diaphus dumerilii: candidate outlier loci, genomic diversity, and temporal population genetic structure

We obtained 275,472,213 PE reads across all 48 genotyped samples and used only Forward reads for all downstream analyses. Four individuals were discarded from *de novo* assembly, as the species identity of these samples could not be confirmed via DNA barcoding. *De novo* assembly and filtering using Stacks and Tassel yielded a dataset containing 2617 genotyped SNPs across 42 individuals (N_{2011} = 14, N_{2015} = 15, N_{2016} = 13). An additional 40 SNPs were removed due to deviations from HWE ($P < 0.01$), resulting in a final dataset containing 2577 SNPs. BayeScan and OutFLANK analyses of the *D. dumerilii* temporal SNP dataset yielded no evidence of candidate outlier loci; thus, only a single dataset containing all SNP loci was utilized for all downstream analyses.

Small differences in genomic diversity metrics were found in *D. dumerilii* across the three GOM temporal sample sets (2011, 2015, 2016), with the notable exception that the 2011 samples possessed more than three times the number of private alleles than found in 2015 or 2016 collections; however, 65% of the 2011-collection private alleles were restricted to just three of the 14 genotyped 2011 fish (data not shown). All three temporal collections demonstrated low levels of observed heterozygosity (H_O) compared to expected values (H_O = 0.144-0.152; Table 2), and high inbreeding coefficients (F_{IS} = 0.241-0.336; Table 2).

The overall estimate of genomic differentiation across temporal *D. dumerilii* samples was low (F_{ST} = 0.011), but statistically significant across two of three assessment methods (Table 3). Pairwise F_{ST} estimates across sampling years were mostly low (F_{ST} = 0.005-0.020) and showed limited and inconsistent statistical significance across sampling years and assessment methods (Table 3). The first three components of the *D. dumerilii* PCA represented 20.04% of the total genetic variance in the SNP dataset, and visualization of the first two components showed two distinct clusters of individuals (with separation along the x-axis or first component)—one cluster containing four individuals (N = 3 from 2011 and N = 1 from 2015) and a second cluster containing the remaining 38 *D. dumerilii* individuals, with only moderate separation along the second component (y-axis) (Fig. 2A). Admixture's CV method identified no population structure, as the optimal value of *K* for the *D. dumerilii* SNP dataset was *K* = 1 (CV = 0.675) (Suppl. Mat. Fig. S1A). Adegenet's *K*-means clustering method indicated the likely presence of two genetic clusters (adegenet: BIC = 219.95; Suppl. Mat. Fig. S2A)—one of these clusters contained the bulk of all samples (N = 38), and the other contained the same four fish identified as y-axis outliers by the PCA.

3.2.2. Lepidophanes guentheri: candidate outlier loci, genomic diversity, and temporal population genetic structure

We obtained 360,101,978 PE reads across all 48 genotyped samples and used only Forward reads for all downstream analyses. Four individuals were discarded from *de novo* assembly, as the species identity of these samples could not be confirmed via DNA barcoding. *De novo* assembly and filtering using Stacks and Tassel yielded a dataset containing 3482 genotyped SNPs across 44 individuals (N_{2011} = 18, N_{2015} = 13, N_{2016} = 13). An additional 20 loci were discarded due to deviations from HWE ($P < 0.01$), resulting in a final dataset containing 3462 SNPs. BayeScan and OutFLANK analyses of the *L. guentheri* temporal SNP dataset yielded no evidence of candidate outlier loci; thus, only a single dataset containing all SNP loci was retained for downstream analysis.

Genomic diversity metrics across the *L. guentheri* temporal sample collections (2011, 2015, 2016) were largely similar, and each collection demonstrated high F_{IS} values (0.194-0.212) and low observed heterozygosity estimates (0.185-0.186) (compared to H_E; Table 2). The number of private alleles was highest within the 2011 sample set, as was the percent polymorphic loci (Table 2).

The overall estimate of genomic differentiation across *L. guentheri* temporal samples was low and not statistically significant (F_{ST} = 0.002; Table 3), and pairwise tests showed largely negligible differentiation between temporal samples (Table 3). The first three components of the *L. guentheri* PCA represented 10.98% of the total variance in the SNP dataset and visualization of the first two components showed no marked clustering of individuals, and high overlap of temporal samples in multivariate space (Fig. 2B). Admixture's CV method and adegenet's *K*-

means clustering algorithm identified the optimal value of *K* for the *L. guentheri* SNP dataset as *K* = 1 (Admixture: CV = 0.684; adegenet: BIC = 243.04) (Suppl. Mat. Fig. S1B and Fig. S2B).

3.2.3. Ceratoscopelus warmingii: candidate outlier loci, genomic diversity, and temporal and spatial population genetic structure

We obtained 715,460,103 PE reads across all 96 genotyped samples with only the Forward reads used for all downstream analyses. Four individuals were discarded from the *de novo* assembly to assure taxonomic certainty. *De novo* assembly and filtering using Stacks and Tassel yielded a dataset containing 1942 genotyped SNPs across the combined 87 individuals (N_{2011} = 22, N_{2015} = 19, N_{2016} = 24, N_{BEAR} = 22). An additional 138 SNPs were removed due to deviations from HWE ($P < 0.01$), resulting in a final dataset containing 1804 SNPs. BayeScan and OutFLANK analyses of the GOM *C. warmingii* temporal SNP dataset yielded no evidence of candidate outlier loci; notably however, analysis of the pooled GOM (2011, 2015, 2016) and Bear Seamount samples identified 20 and 52 candidate outlier SNP loci using the programs BayeScan and OutFLANK, respectively, and 20 of these loci were identified as candidate outliers by both methods. Thus, three *C. warmingii* GOM-Bear Seamount SNP datasets were generated: Cwar-SNP-all, containing all 1804 SNPs, Cwar-SNP-outlier, containing only candidate outlier loci (20 SNPs), and Cwar-SNP-neutral, which contained only putatively neutral SNPs (1784 SNPs).

With respect to the *C. warmingii* GOM temporal (2011, 2015, 2016) samples only minor differences were identified in genomic diversity indices among the three temporal sample collections, and all temporal collections possessed observed heterozygosity deficits (compared to H_E) and high associated inbreeding coefficients (Table 2). Comparison of the pooled GOM (2011, 2015, 2016) *vs.* Bear Seamount samples (Cwar-SNP-all) showed similar levels of

genomic diversity; however, the GOM samples (pooled, N = 65) possessed 71 private alleles compared to the Bear Seamount samples (N = 22; private alleles = 17).

Across GOM temporal sample collections of *C. warmingii*, the overall (F_{ST} = 0.001) and pairwise estimates of differentiation had inconsistent statistical support and were largely negligible and non-significant (Table 3). The first three principal components of the *C. warmingii* GOM-only PCA represented 12.82% of the total variance in the SNP dataset and visualization of the first two components showed no clustering of individual genotypes (Fig. 2C). Likewise, Admixture and *de novo* *K*-means clustering (as implemented in adegenet) identified a single genetic cluster in the *C. warmingii* GOM-only SNP dataset (Admixture: CV = 0.600; adegenet: BIC = 318.60) (Suppl. Mat. Fig. S1C and Fig. S2C).

Comparison of pooled GOM *C. warmingii* samples to conspecifics collected from Bear Seamount identified low (F_{ST} = 0.022; Cwar-SNP-all), but statistically significant, genetic differentiation using all three significance determination methods (i.e., permutation testing and both 95% CIs; Table 3). Upon analysis of all 1804 SNPs generated from the *C. warmingii* samples (i.e., Cwar-SNP-all), the first three principal components of the *C. warmingii* GOM-Bear Seamount PCA represented 13.79% of the total variance in the genetic dataset, and visualization of the first two components indicated the presence of two distinct genetic clusters: one cluster containing eight Bear Seamount fish (separated along the first component; x-axis), and a second genetic cluster comprising the remaining 79 fish which had a mixture of both Bear Seamount and GOM individuals (Fig. 2D). Analysis of the *C. warmingii* GOM-Bear Seamount samples (Cwar-SNP-all) using Admixture identified a single cluster as the optimal value of genetic groupings (CV = 0.601; Suppl. Mat. Fig. S1D), whereas results of adegenet's *K*-means

clustering method mirrored those of the PCA and identified the same two distinct genetic clusters (BIC = 427.48; Suppl. Mat. Fig. S2D).

Pairwise estimates of genetic differentiation using the *C. warmingii* candidate outlier (20 loci; Cwar-SNP-outlier) and neutral (1784 loci; Cwar-SNP-neutral) datasets each identified highly significant genetic differentiation between the GOM and WNA sample collections ($F_{ST\text{-}OUTLIER}$ = 0.357, P <0.001; $F_{ST\text{-}NEUTRAL}$ = 0.018, P <0.001), with levels of differentiation ~20X higher across candidate outliers. Patterns of differentiation identified via PCA of the neutral and candidate outlier datasets were comparable to those identified through analysis of the overall dataset—with separation of eight Bear Seamount *C. warmingii* individuals along the first principal component, albeit the magnitude of this differentiation (Cwar-SNP-outlier *vs.* Cwar-SNP-all dataset) was much more pronounced (Suppl. Mat. Fig. S3).

4. Discussion

Herein, we examined the population genomics of three numerically dominant, mesopelagic lanternfish species within the GOM following one of the largest marine oil spills in history, providing the first genome-scale view of the genetic diversity and population structure of this ecologically key group of fishes. Overall, this work showed: (1) that all three lanternfish species demonstrated low SNP genetic diversity (H_O) but high mitochondrial genetic diversity, large heterozygosity deficits at SNP loci (H_O compared to H_E), and high inbreeding coefficients, (2) mixed indications of temporal or spatial population genetic structure within species in the GOM, and (3) statistically significant spatial genetic differentiation in *C. warmingii* between the northern GOM and WNA.

4.1. Genetic diversity

The high abundance and biomass of lanternfishes in the deep-sea in general and the GOM in particular (Gartner et al., 1987; Ross et al., 2010; Catul et al., 2011; Davison et al., 2015; Milligan and Sutton, 2020), coupled with our data indicating high levels of mitochondrial genetic diversity in all three species, made the ubiquitous findings of low overall genetic diversity and high inbreeding coefficients unexpected. This finding raised the possibility of a suboptimal RAD loci assembly and/or the presence of genotyping errors within our SNP dataset (see Hendricks et al., 2018); we suggest that the likelihood of such issues is low for the following reasons. First, we used the program Stacks for our *de novo* RAD loci assembly and followed the recommendations of Rochette and Catchen (2017) and Paris et al. (2017), abiding by established guidelines for parameter optimization. Second, to directly test whether any of the observed heterozygosity deficits across species might be due to SNP calling errors at low sequencing coverage loci (i.e., the unintentional omission of heterozygotes due to low sequencing depth), we re-filtered each dataset *a posteriori* to remove genotypes possessing less than 10X depth coverage and re-analyzed the data. Estimation of H_O and F_{IS} values following 10X coverage filtering using the resultant slightly reduced SNP datasets (from the increased presence of invariant sites), however, showed little change in either diversity metric [*D. dumerilii*, SNPs = 2610, H_O = 0.149-0.155, F_{IS} = 0.203-0.322; *L. guentheri*, SNPs = 3467, H_O = 0.196-0.199, F_{IS} = 0.159-0.175; *C. warmingii* (GOM only; SNPs = 1942), H_O = 0.202-207, F_{IS} = 0.208-0.231], indicating that insufficient sequencing depth was not likely driving the observed low diversity values.

Findings of low observed heterozygosity (compared to expected heterozygosity) and/or high inbreeding coefficients have been noted in studies of at least three other evolutionarily divergent mesopelagic and bathypelagic species that used similar RADseq SNP discovery

methods: hatchetfish (Rodriguez-Ezpeleta et al., 2017), cephalopods (Timm et al., 2020a) and crustaceans (Timm et al., 2020b). Across these studies, it was suggested that cryptic genetic population structure (the Wahlund effect), inadequate, uneven, or non-random sampling, purifying selection, non-random mating or inbreeding, and/or population size declines may be responsible for these values. Herein, akin to Rodriguez-Ezpeleta et al. (2017), we suggest that given the apparent large census numbers of lanternfishes within the GOM and greater North Atlantic, the hypothesis of inbreeding as an explanatory factor seems unlikely. We note, however, lanternfishes are pelagic spawners with extended larval durations (>20 days), and some species possess exceptionally low batch fecundities compared to other pelagic fishes (lanternfish batch fecundity: 100 to 1000s; Catul et al., 2011), including the species of myctophids studied herein (Gartner, 1993). We speculate that this combination of life history characters could lead to high variance in reproductive success among individuals and thus low effective population sizes despite high population census sizes (i.e., low Ne/Nc ratios), a characteristic often seen in marine teleosts (Hauser et al., 2002; Hauser and Carvalho, 2008; Hare et al., 2011). Such life history traits may also result in high levels of genetic drift and ultimately non-random mating between individuals (i.e., Sweepstakes Reproductive Success hypothesis; Hedgecock and Pudovkin, 2011), which might explain the low observed heterozygosity and/or high inbreeding coefficients seen here. As other factors such as selection, unknown population genetic structure, and non-random or limited sampling may also yield similar signatures of high inbreeding coefficients observed here, we recommend increased sampling (both in numerical abundance and geographic distribution) of these lanternfishes to further test any and all the above biological hypotheses about the genetic diversity statuses of these deep-sea fishes.

Despite variations in some life history parameters among the three lanternfish species, including frequency and duration of reproduction (Gartner, 1993) and relative depth of diel vertical migrations (Gartner et al., 1987), genomic diversity estimates were comparable across taxa. Furthermore, few (if any) temporal diversity differences were noted within species, albeit the high number of private alleles in a subset of the *D. dumerilii* individuals sampled in 2011 excepted (see discussion below), and variations in sample size or location did not affect the magnitude of observed F_{IS} values. To date, little data exist concerning the genetic diversity of lanternfishes to assess if our results are emblematic of this group in general, or a reflection of just the species and region of our study. We note that high microsatellite marker-based allelic variation was reported for the Southern Ocean lanternfish, *Electrona antarctica* (Van de Putte et al., 2012), albeit this inference was based on a relatively small number (seven) of microsatellite markers.

4.2. Temporal genetic dynamics

Given the short life span of our three lanternfish species (1-2 years), which allowed for two to four generations to pass between the 2011 and 2016 sampling events, some variation in allele frequencies due to genetic drift, variance in reproductive success and/or finite sampling is expected. Within species, the largest temporal change in allele frequencies, as measured by pairwise F_{ST}, was between the 2011 and 2015/2016 collections of *D. dumerilii*; however, these comparisons were not consistently statistically significant across metrics. Likewise, little variation was noted across GOM temporal replicates of *L. guentheri* and *C. warmingii* ($F_{ST} <$ 0.004), indicating that allele frequencies were relatively stable over the time period and generations assessed for these species.

Within the GOM, physical oceanography is largely dominated by the Loop Current, a warm water mass that originates south as the Caribbean Current and flows north into the Gulf via the Yucatan Channel, as reviewed in Lindo-Atichati et al. (2012). Once in the GOM, the Loop Current flows eastward or northward. When moving northward, it periodically sheds large anti-cyclonic eddies as it flows clockwise and southward, passing through the Florida Straits as the Florida Current and exiting into the Atlantic proper as the Gulf Stream Current. As lanternfishes are small, pelagic spawners with buoyant eggs (Catul et al., 2011) and possess extended larval durations, the circulation in the WNA, including the GOM, has the potential to facilitate lanternfish dispersal across large horizontal distances. Consistent with this expected wide dispersal, the GOM populations of lanternfishes appeared to be largely panmictic. However, the *D. dumerilii* diversity metrics, PCA and the *K*-means cluster analyses, indicated the likely presence of at least two genetic clusters, which may be reflected in the allelic variation seen across temporal samples. Given that the *D. dumerilii* novel genotypes were found in only four of 42 sequenced individuals (N = 3, 2011; N =1, 2015) and that the species identity of these individuals was confirmed via DNA barcoding, we offer two competing hypotheses to explain their occurrence. First, these novel genotypes may in fact be endemic to the GOM, but are comparatively rare, and our 2016 sample size (N = 13) was simply insufficient to capture all the variation present within this region. Second, these novel *D. dumerilii* genotypes may have been the result of a rare dispersal event from an up-stream and genetically diverse source. As *D. dumerilii*'s range extends broadly throughout the Atlantic, dispersal of this species via larval entrainment in the Caribbean Current or some other downstream source from the Loop Current may have occurred. Such dispersal may have allowed for the transient introduction of genetically diverse migrants into GOM waters, thereby increasing its local genetic diversity. Past surveys

have demonstrated that all three studied lanternfish species are found within the Gulf Loop, albeit it at a reduced magnitude compared to the surrounding Gulf Common Water (Gartner et al., 1987; Milligan and Sutton, 2020). Thus, GOM mesoscale oceanography supports the hypothesis that the Loop Current may serve to facilitate the dispersal, and ultimately, shape the genetic connectivity of lanternfishes in this region.

4.3. Ceratoscopelus warmingii spatial genetic connectivity

Our ability to draw inferences about the extent of genetic connectivity between GOM and WNA lanternfishes (and mesopelagic fishes in general) is limited to results from *C. warmingii*, since tissue samples from Bear Seamount were available for only this species. The lanternfish *C. warmingii* exhibits a circumglobal distribution (Gaither et al., 2016), and previous work has indicated the presence of population genetic structure across relatively limited geographic scales (Gordeeva et al., 2011; 2014). Here, we found statistically significant genetic differentiation between *C. warmingii* from the GOM and Bear Seamount, when the entire SNP dataset (Cwar-SNP-all) was analyzed. Notably, this partitioning of genetic variation was not exclusive to the geographic capture location of individuals, indicating that asymmetric gene flow between geographic sites may have occurred. For instance, while a single genetically homogeneous population of *C. warmingii* was found within the GOM, individuals collected from Bear Seamount comprised two distinct genetic groups: one whose genotypes clustered with those from the GOM and a second that consisted of eight lone Bear Seamount individual genotypes (see Fig. 2D and Suppl. Mat. Fig. S2D). Per other regional faunal (Markle et al., 1980; Moore et al., 2003) and population genetic surveys (Timm et al., 2020a; 2020b), we offer that the Loop Current/Gulf Stream complex, in conjunction with the strong diel, vertical migratory propensity

of *C. warmingii* (Gartner et al., 1987; Eduardo et al., 2021) may help facilitate larval and/or adult dispersal of this species from low (subtropical/tropical) to high WNA latitudinal habitats (Timm et al., 2020b). Alternatively, findings of an ancillary, differentiated genetic cluster of novel genotypes found exclusively at Bear Seamount, indicate the presence of a second genetic population of *C. warmingii* in the Atlantic, which (1) may be endemic and perhaps locally adapted to Bear Seamount, or (2) may be a result of past and/or ongoing gene flow from one or more distant sources (Moore et al., 2004). The identification of a set of candidate outlier SNP loci differentiating fish inhabiting the GOM and the WNA is consistent with both above hypotheses and suggests that despite the high dispersal potential of *C. warmingii*, local adaptation may potentially be serving to differentiate populations of this circumglobally distributed species. Additional sampling, including the collection of temporal replicates at Bear Seamount, is needed to better resolve the connectivity dynamics and the potential for adaptive differentiation between these two highly environmentally diverse western Atlantic habitats.

5. Conclusions

Characterizing the population genetic dynamics of mesopelagic species is an increasing conservation management priority given the growing exploitation of hydrocarbon resources and imminent extraction of mineral deposits (Cuyvers et al., 2018). Herein, we present the first perspective on the genomic diversity of three dominant lanternfishes within the GOM, and for lanternfishes anywhere. We also provide preliminary data concerning the genomic connectivity and diversity of *C. warmingii,* a highly abundant, circumglobal lanternfish species, in the western Atlantic. These data serve as baseline genetic knowledge for these species, which could prove useful in future assessments of anthropogenic impacts in a region with large-scale ongoing

and increasing hydrocarbon industrial exploration activities. These data also provide some of the earliest views of the population genomics of mesopelagic fishes (we are only aware of one other published population genomics study on a mesopelagic fish: the sternoptychid, *Maurolicus muelleri*; Rodriguez-Ezpeleta et al. 2017). It will be of interest to determine if the unexpected findings of low genetic diversity and high inbreeding in the three lanternfishes (this study) and *Maurolicus muelleri* is a general feature of mesopelagic fishes, despite their presumed large, regional population sizes. These findings suggest that there remain many unknown factors shaping the evolutionary history of deep-ocean fishes.

Acknowledgements

This research was made possible in part by a grant from the National Oceanic and Atmospheric Administration's RESTORE Science Program under award NA19NOS4510193 to Nova Southeastern University, in part by The Gulf of Mexico Research Initiative (SA 15-21), and accessory financial support from the Save Our Seas Foundation Program Grant 157 and Guy Harvey Ocean Foundation. Data are publicly available through the Gulf of Mexico Research Initiative Information & Data Cooperative (GRIIDC) at https://data.gulfresearchinitiative.org [doi: 10.7266/n7-67wg-mz19]. We thank Laura Timm for her technical advice when designing our SNP protocol, Jon Moore for his collection and taxonomic efforts, and April Cook, Nina Pruzinsky and Natalie Slayden for their invaluable help curating tissues samples and associated data. We would also like to thank the NOAA/NMFS Northeast Fisheries Science Center and the officers, crew, and science team on the NOAA ship *Pisces* for the collection of the Bear Seamount samples during their Deepwater Systematics Survey. Portions of this research were

conducted with the advanced computing resources provided by the Texas A&M High Performance Research Computing Cluster.

Appendix A. Supplemental Materials: Cytochrome *c* oxidase subunit I gene (COI) DNA Barcoding

Supplementary data for this work may be found online.

Appendix B. Supplemental Materials: Single Nucleotide Polymorphism Results

Supplementary data for this work may be found online.

References

Alexander, D.H., Novembre, J., Lange, K. (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Research, 19(9): 1655-1664.

Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Archer, F.I., Adams, P.E., Schneiders, B.B. (2017) STRATAG: An R package for manipulating, summarizing and analyzing population genetic data. Molecular Ecology Resources, 17(1): 5-11.

Batta-Lona, P.G., Galindo-Sánchez, C.E., Arteaga, M.C., Robles-Flores, J., Jiménez-Rosenberg, S.P.A. (2019) DNA barcoding and morphological taxonomy: identification of lanternfish (Myctophidae) larvae in the Gulf of Mexico, Mitochondrial DNA Part A, 30(2): 375-383.

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

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., Buckler, E.S. (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics, 23(19): 2633-2635.

Catchen, J., Hohenlohe, P., Bassham, S., Amores, A., Cresko, W. (2013) Stacks: an analysis tool set for population genomics. Molecular Ecology, 22: 3124-3140.

Catul, V., Gauns, M., Karuppasamy, P.K. (2011) A review of mesopelagic fishes belonging to family Myctophidae. Reviews in Fish Biology and Fisheries, 21: 339-354.

Choy, C.A., Robinson, B.H., Gagne, T.O., Erwin, B., Firl, E., Halden, R.U., Hamilton, J.A., Katija, K., Lisin, S.E., Rolsky, C., Van Houten, K.S. (2019) The vertical distribution and biological transport of marine microplastics across the epipelagic and mesopelagic water column. Scientific Reports, 9(1): 1-9.

Christiansen, H., Dettai, A., Heindler, F.M., Collins, M.A., Duhamel, G., Hautecoeur, M., Steinke, D., Volckaert, F.A.M., Van de Putte, A.P. (2018) Diversity of mesopelagic fishes in the southern ocean – a phylogeographic perspective using DNA barcoding. Frontiers in Ecology and Evolution, 6: 120.

Cook, A., Bernard, A.M., Boswell, K.M., Bracken-Grissom, H., D'Elia, M., Derada, S., Easson, C.G., English, D., Eytan, R.I., Frank, T., Hu, C., Johnston, M.W., Judkins, H., Lembke, C., Lopez, J.V., Milligan, R.J., Moore, J.A., Penta, B., Pruzinsky, N.M., Quinlan, J.A., Richards, T., Romero, I.C., Shivji, M., Vecchione, M., Weber, M.D., Wells, D., Sutton, T.T. (2020) A Multidisciplinary Approach to Investigate Deep-Pelagic Ecosystem Dynamics in the Gulf of Mexico following Deepwater Horizon. Frontiers in Marine Science, 7: 548880.

Cuyvers, L., Berry, W., Gjerde, K.M., Thiele, T., Wilhem, C. (2018) Deep seabed mining: a rising environmental challenge, IUCN and Gallifrey Foundation, Gland, Switzerland (2018), pp. 1-74.

Davison, P., Lara-Lopez, A., Koslow, J.A. (2015) Mesopelagic fish biomass in the southern California current ecosystem. Deep-Sea Research II, 112: 129-142.

Davison, P.C., Checkley Jr., D.M., Koslow, J.A., Barlow, J. (2013) Carbon export mediated by mesopelagic fishes in the northeast Pacific Ocean. Progress in Oceanography, 116: 14-30.

de Busserolles, F., Marshall, N.J. (2017) Seeing in the deep-sea: visual adaptations in lanternfishes. Philosophical Transactions of the Royal Society B, 372: 20160070.

D'Elia, M., Warren, J.D., Rodriquez-Pinto, I., Sutton, T.T., Cook, A., Boswell, K.M. (2016) Diel variation in the vertical distribution of deep-water scattering layers in the Gulf of Mexico. Deep-Sea Research I, 115: 91-102.

Denton, J.S.S. (2018) Diversification patterns of lanternfishes reveal multiple rate shifts in a critical mesopelagic clade targeted for human exploitation. Current Biology, 28: 933-940.

Drazen, J.C., Smith, C.R., Gjerde,, K.M., Haddock, S.H.D., Carter, G.S., Choy, C.A., Clark, M.R., Dutrieux, P., Goetze, E., Hauton, C., Hatta, M., Koslow, J.A., Leitner, A.B., Pacini, A., Perelman, J.N., Peacock, T., Sutton, T.T., Watling, L., Yamamoto, H. (2020) Opinion: Midwater ecosystems must be considered when evaluating environmental risks of deep-sea mining. Proceedings of the National Academy of Sciences, 117(30): 17455-17460.

Eduardo, L.N., Bertrand, A., Mincarone, M.M., Martins, J.R., Frédou, T., Assunção, R.V., Lima, R.S., Ménard, F., Le Loc'h, F., Lucena-Frédou, F. (2021) Distribution, vertical migration, and trophic ecology of lanternfishes (Myctophidae) in the Southwestern Tropical Atlantic. Progress in Oceanography, 199: 102695.

Excoffier, L., Lischer, H.E.L. (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetic analyses under Linux and Windows. Molecular Ecology Resources, 10(3): 567-567.

Florence, W.K., Hulley, P.A., Stewart, B.A., Gibbons, M.J. (2002) Genetic and morphological variation on the lanternfish *Lampanyctodes hectoris* (Myctophiformes: Myctophidae) off southern Africa. South Africa Journal of Marine Science, 24: 193-203.

Foll, M., Gaggiotti, O. (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics, 180(2): 977-993.

Fricke, R., Eschmeyer, W.N., Fong, J.D. (2021). Genera/Species by Family/Subfamily in Eschmeyer's Catalog of Fishes. [Online] Available online at: https://researcharchive.calacademy.org/research/ichthyology/catalog/SpeciesByFamily.asp (Updated 01 Mar 2021).

Gaither, M.R., Bowen, B.W., Rocha, L.A., Briggs, J.C. (2016) Fishes that rule the world: circumglobal distributions revisited. Fish and Fisheries, 17: 664-679.

Gaither, M.R., Gkafas, G.A., de Jong, M., Sarigol, F., Neat, F., Regnier, T., Moore, D., Gröcke, D.R., Hall, N., Liu, X., Kenny, J., Lucaci, A., Hughes, M., Haldenby, S., Hoelzel, A.R.

(2018) Genomics of habitat choice and adaptive evolution in a deep-sea fish. Nature Ecology & Evolution, 2: 680-687.

Gartner, J.V., Hopkins, T.L., Baird, R.C., Milliken, D.M. (1987) The lanternfishes (Pisces: Myctophidae) of the eastern Gulf of Mexico. Fishery Bulletin, 85: 81-98.

Gartner, J.V. (1991) Life histories of three species of lanternfishes (Pisces: Myctophidae) from the eastern Gulf of Mexico. II. Age and growth patterns. Marine Biology, 111: 21-27.

Gartner, J.V. (1993) Patterns of reproduction in the dominant lanternfish species (Pisces: Myctophidae) of the eastern Gulf of Mexico, with a review of reproduction among tropical-subtropical Myctophidae. Bulletin of Marine Science, 52: 721-750.

Gjøsæter, J., Kawaguchi, K. (1980). A review of the world resources of mesopelagic fish, 193, FAO Fish Tech Pap, FAO, Rome.

Gordeeva, N.V. (2011) On structure of species in pelagic fish: the results of populational-genetic analysis of four species of lanternfish (Myctophidae) from the southern Atlantic. Journal of Ichthyology, 51: 152-165.

Goordeeva, N.V. (2014) Phylogeography, genetic isolation, and migration of deep-sea fishes in the South Atlantic. Journal of Ichthyology, 54: 642-659.

Hare, M.P., Nunney, L., Schwartz, M.K., Ruzzante, D.E., Burford, M., Waples, R.S., Ruegg, K., Palstra, F. (2011) Understanding and estimating effective population size for practical application in marine species management. Conservation Biology, 25(3): 438-449.

Hauser, L., Adcock, G.J., Smith, P.J., Bernal Ramirez, J.H., Carvalho, G.R. (2002) Loss of microsatellite diversity and lowe effective population size in an overexploited population of New Zealand snapper (*Pagrus auratus*). Proceedings of the National Academy of Sciences of the United States of America, 99: 11742-11747.

Hauser, L., Carvalho, G.R. (2008) Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. Fish and Fisheries, 9: 333-362.

Hedgecock, D., Pudovkin, A.I. (2011) Sweepstakes reproductive success in highly fecund marine fish and shellfish: a review and commentary. Bulletin of Marine Science, 87(4): 971-1002.

Helyar, S.J., Hemmer-Hansen, J., Bekkevold, D., Taylor, M.I., Ogden R., Limborg, M.T., Cariani, A., Maes, G.E., Diopere, E., Carvalho, G.R., Nielsen, E.E. (2011) Applications of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. Molecular Ecology Resources, 11(s1): 123-136.

Hendricks, S., Anderson, E.C., Antao, T., Bernatchez, L., Forester, B.R., Garner, B., Hand, B.K., Hohenlohe, P.A., Kardos, M., Koop, B., Sethuraman, A., Waples, R.S., Luikart, G. (2018) Recent advances in conservation and population genomics data analysis. Evolutionary Applications, 11: 1197-1211.

Hidalgo, M., and Browman, H.I. (2019) Developing the knowledge base needed to sustainably manage mesopelagic resources. ICES Journal of Marine Science, 76: 609-615.

Irigoien, X., Klevjer, T.A., Røstad, A., Martinez, U., Boyra, G., Acuña, J.L., Bode, A., Echevarria, F. Gonzalez-Gordillo, J.I., Hernadez-Leon, S., Agusti, S., Aksnes, D.L.,

Duarte, C.M., Kaartvedt, S. (2014) Large mesopelagic fishes biomass and trophic efficiency in the open ocean. Nature Communications, 5: 3271.

Jombart, T. (2008) adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics, 24: 1403-1405.

Jombart, T., and Ahmed, I. (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics, 27(21): 3070-3071.

Keenan, K., McGinnity, P., Cross, T.F., Crozier, W.W., Prodöhl, P.A. (2013) diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. Methods in Ecology and Evolution, 4(8): 782-788.

Kojima, S., Moku, M., Kawaguchi, K. (2009) Genetic diversity and population structure of three dominant myctophid fishes (*Diaphus theta*, *Stenobrachius leucopsarus*, and *S. nannochir*) in the North Pacific Ocean. Journal of Oceanography, 65: 187-193.

Kristoffersen, J.B., Salvanes, A.G.V. (2009) Distribution, growth, and population genetics of the glacier lanternfish (*Benthosema glaciale*) in Norwegian waters: contrasting patterns in fjords and the ocean. Marine Biology Research, 5: 596-604.

Lindo-Atichati, D., Bringas, F., Goni, G., Mulhing, B., Muller-Karger, F.E., Habtes, S. (2012) Varying mesoscale structures influence larvalfish distribution in the northern Gulf of Mexico. Marine Ecology Progress Series, 463: 245-257.

Markle, D.F., Scott, W.B. (1980) New and rare records of Canadian fishes and the influence of hydrography on resident and nonresident Scotian shelf Icthyofauna. Canadian Journal of Fisheries and Aquatic Sciences, 37: 49-65.

Martin, R.P., Olson, E.E., Girard, M.G., Smith, W.L., Davis, M.P. (2018) Light in the darkness: new perspective on lanternfish relationships and classification using genomic and morphological data. Molecular Phylogenetics and Evolution, 121: 71-85.

Miller, J.M., Cullingham, C.I., Peery, R.P. (2020) The influence of a priori grouping on inference of genetic clusters: simulation study and literature review of the DAPC method. Heredity, 125: 269-280.

Milligan, R.J., Sutton, T.T. (2020) Dispersion overrides environmental variability as a primary driver of the horizontal assemblage structure of the mesopelagic fish family Myctophidae in the northern Gulf of Mexico. Frontiers in Marine Science, 7:15, doi: 10.3389/fmars.2020.00015.

Moore, J.A., Vecchione, M., Collette, B.B., Gibbons, R., Hartel, K.E., Galbraith, J.K., Turnipseed, M., Southworth, M., Watkins, E. (2003) Biodiversity of Bear Seamount, New England Seamount chain: results of exploratory trawling. Journal of Northwest Atlantic Fishery Science, 31: 363-372.

Moore, J.A., Vecchione, M., Collette, B.B., Gibbons, R., Hartel, K.E. (2004) Selected fauna of Bear Seamount (New England Seamounts chain), and the presence of “natural invader” species. Archive of Fishery and Marine Research, 51(1-3): 241-250.

Murawski, S. A., Ainsworth, C. H., Gilbert, S., Hollander, D. J., Paris, C. B., Schlüter, M., Wetzel, D. (2020) Scenarios and Responses to Future Deep Oil Spills - Fighting the Next War. Springer, doi: 10.1007/978-3-030-12963-7

Paradis E. (2010) pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics, 26: 419–420.

Paris, J.R., Stevens, J.R., Catchen, J.M. (2017) Lost in parameter space: a road map for STACKS. Methods in Ecology and Evolution, 8(10): 1360-1373.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E. (2012) Double digest RADSeq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PloS One, 7(5): e37135.

Porter, M.L., Roberts, N.W., Partridge, J.C. (2016) Evolution under pressure and the adaptation of visual pigment compressibility in deep-sea environments. Molecular Phylogenetics and Evolution, 105: 160-165.

Portnoy, D.S., Fields, A.T., Greer, J.B., Schlenk, D. (2020) Genetics and oil: transcriptomics, epigenetics, and population genomics as tools to understand animal responses to exposure across different time scales. In: Murawski, S.A., Ainsworth, C.H., Gilbert, S., Hollander, D.J., Paris, C.B., Schlüter, M., Wetzel, D.L. (Eds.), Deep oil spills: facts, fate, and effects. Springer Nature, Switzerland, pp. 515-532.

R Core Team. (2018) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

Ramirez-Llodra, E., Tyler, P.A., Baker, M.C., Bergstad, O.A., Clark, M.R., Escobar, E., Levin, L.A., Menot, L., Rowden, A.A., Smith, C.R., Van Dover, C.L. (2011) Man and the last great wilderness: human impact on the deep sea. PLoS ONE, 6: e22588.

Ratnasingham, S., Hebert, P.D.N. (2013) A DNA-based registry for all animal species: the barcode index number (BIN) system. PLoS ONE, 8(7): e66213.

Robinson, B.H. (2009) Conservation of deep pelagic biodiversity. Conservation Biology, 23: 847-858.

Rochette, N.C., Catchen, J.M. (2017) Deriving genotypes from RAD-seq short-read data using Stacks. Nature Protocols, 12: 2640-2659.

Rodriguez-Ezpeleta, N., Álvarez, P., Irigoien, I. (2017) Genetic diversity and connectivity in *Maruolicus muelleri* in the Bay of Biscay inferred from thousands of SNP markers. Frontiers in Genetics, 8: 195.

Ross, S.W., Quattrini, A.M., Roa-Varón, A.Y., McClain, J.P. (2010) Species composition and distributions of mesopelagic fishes over the slope of the north-central Gulf of Mexico. Deep-Sea Research II, 57: 1926-1956.

St. John, M.A., Borja, A., Chust, G., Heath, M., Grigorov, I., Mariani, P., Martin, A.P., Santos, R.S. (2016) A dark hole in our understanding of marine ecosystems and their services: perspectives from the mesopelagic community. Frontiers in Marine Science, 3: 31.

Steinke, D., Hanner, R. (2010) The FISH-BOL collaborators' protocol. Mitochondrial DNA, 21(S2): 1-5.

Suneetha, K.B., Salvanes, A.G.V. (2001) Population genetic structure of the glacier lanternfish, *Benthosema glaciale* (Myctophidae) in Norwegian waters. Sarsia, 86: 203-212.

Sutton, T.T. (2013) Vertical ecology of the pelagic ocean: classical patterns and new perspectives. Journal of Fish Biology, 83: 1508-1527.

Sutton, T.T., Clark, M.R., Dunn, D.C., Halpin, P.N., Rogers, A.D., Guinotte, J., Bograd, S.J., Angel, M.V., Perez, J.A.A., Wishner, K., Haedrich, R.L., Lindsay, D.J., Drazen, J.C., Vereshchaka, A., Piatkowski, U., Morato, T., Błachowiak-Samołyk, K., Robison, B.H., Gjerde, K.M., Pierrot-Bults, A., Bernal, P., Reygondeau, G., Heino, M. (2017) A global biogeographic classification of the mesopelagic zone. Deep-Sea Research Part I, 126: 85-102.

Sutton, T.T., Frank, T.M., Romero, I.C., Judkins, H. (2020) Chapter 24: As Gulf oil extraction goes deeper, who is at risk? Community structure, distribution, and connectivity of the deep-pelagic fauna. In: Murawski, S.A., Ainsworth C., Gilbert, S., Hollander, D., Paris, C.B., Schlüter, M. and Wetzel, D. (Eds.) Scenarios and responses to future deep oil spills – Fighting the Next War. Springer, pp. 403-418.

Takagi, K., Yatsu, A., Moku, M., Sassa, C. (2006). Age and growth of lanternfishes, *Symbolophorus californiensis* and *Ceratoscopelus warmingii* (Myctophidae), in the Kuroshio-Oyashio Transition Zone. Ichthyological Research, 53: 281-289.

Timm, L.E., Bracken-Grissom, H.D., Sosnowski, A., Breitbart, M., Vecchione, M., Judkins, H. (2020a) Population connectivity of three deep-sea cephalopod species between the Gulf of Mexico and northwestern Atlantic Ocean. Deep-sea Research I, 158: 103222.

Timm, L.E., Isma, L.M., Johnston, M.W., Bracken-Grissom, H.D. (2020b) Comparative population genomics and biophysical modeling of shrimp migration in the Gulf of Mexico reveals current-mediated connectivity. Frontiers in Marine Science, 7: 19.

Turner, J.R., White, E.M., Collins, M.A., Partridge, J.C., Douglas, R.H. (2009). Vision in lanternfish (Myctophidae): adaptations for viewing bioluminescence in the deep-sea. Deep-Sea Research I, 56: 1003-1017.

Van de Putte, A.P., Van Houdt, J.K.J., Maes, G.E., Hellemans, B., Collins, M.A., Volckaert, F.A.M. (2012) High genetic diversity and connectivity in a common mesopelagic fish of the Southern Ocean: the myctophid *Electrona antarctica*. Deep-Sea Research II, 59-60: 199-207.

Villesen, P. (2007) FaBox: an online toolbox for fasta sequences. Molecular Ecology Notes, 7(6): 965968.

Watanabe, H., Masatoshi, M., Kawaguchi, K., Ishimaru, K., Ohno, A. (1999). Diel vertical migration of myctophid fishes (Family Myctophidae) in the transitional waters of the western North Pacific. Fisheries Oceanography, 8: 115-127.

Webb, T.J., Berghe, E.V., O'Dor, R. (2010) Biodiversity's big wet secret: the global distribution of marine biological records reveals chronic underexploration of the deep pelagic ocean. PLoS ONE 5(8): e10223.

Weir, B.S., Cockerham, C.C. (1984). Estimating F-statistics for the analysis of population structure. Evolution, 38(6): 1358-1370.

Whitlock, M.C., Lotterhos, K.E. (2015) Reliable detection of loci responsible for local adaptation: inference of a null model through trimming the distribution of F_{ST}. American Naturalist, 186(S1): S24-S36.

Widder, E.A. (2010). Bioluminescence in the ocean: origins of biological, chemical, and ecological diversity. Science, 328: 704-708.

Wilson, C.C., Lavender, M., Black, J. (2007). Genetic assessment of walleye (*Sander vitreus*) restoration efforts and options in Nipigon Bay and Black Bay, Lake Superior. Journal of Great Lakes Research, 33(Suppplement 1): 133-144.

1 Table 1. Lanternfish species, sample number genotyped (N; prior to bioinformatic filtering), sample collection dates, sampling region,

2 and collection location GPS (°N, °W) range.

Species	N	Collection Date	Sampling Region	GPS Location (°N, °W) (Decimal Degrees)
Diaphus	17	07/26/2011 – 09/28/2011	Northern Gulf of Mexico	26.98-27.85, 85.44-90.86
dumerilii	16	05/01/2015 – 05/07/2015	Northern Gulf of Mexico	27.00-29.00, 86.00-90.00
	15	05/01/2016 – 05/13/2016	Northern Gulf of Mexico	27.00-29.00, 86.00-90.00
Lepidophanes	20	07/20/2011 – 09/28/2011	Northern Gulf of Mexico	26.85-28.06, 85.47-92.56
guentheri	14	05/01/2015 – 05/07/2015	Northern Gulf of Mexico	27.00-29.00, 86.00-90.00
	14	05/01/2016 – 08/18/2016	Northern Gulf of Mexico	27.00-29.00, 86.00-90.00
Ceratoscopelus	24	07/12/2011 – 09/25/2011	Northern Gulf of Mexico	27.32-27.84, 89.45-92.46
warmingii	24	05/01/2015 – 05/07/2015	Northern Gulf of Mexico	27.00-29.00, 86.00-90.00
	24	08/05/2016 – 08/18/2016	Northern Gulf of Mexico	27.00-29.00, 86.00-90.00
	24	10/20/2014 – 10/21/2014	Bear Seamount, Western North Atlantic	39.94-39.95, 67.23-67.27

3

4 Table 2. Summary sample and genetic diversity statistics of three lanternfish species (*Diaphus dumerilii*, *Lepidophanes guentheri*, and

5 *Ceratoscopelus warmingii*) for single nucleotide polymorphisms (SNPs) across temporal sample collections from the northern Gulf of Mexico

6 (2011, 2015, and 2016) and Bear Seamount (*C. warmingii only*).

Species (# SNPs)	Collection Year and Region	N	H_O	H_E	# private	% poly	F_{IS}
Diaphus	GOM 2011	14	0.144	0.244	81	89.7	0.336
dumerilii	GOM 2015	15	0.152	0.226	21	89.6	0.298
(2577)	GOM 2016	13	0.147	0.211	25	80.5	0.241
Lepidophanes	GOM 2011	18	0.186	0.241	26	94.8	0.212
guentheri	GOM 2015	13	0.185	0.237	16	90.3	0.201
(3462)	GOM 2016	13	0.186	0.236	11	89.9	0.194
Ceratoscopelus	GOM 2011	22	0.206	0.260	8	94.9	0.208
warmingii	GOM 2015	19	0.202	0.255	7	93.3	0.209
(Cwar-SNP-all; 1804)	GOM 2016	24	0.202	0.261	19	96.6	0.230
	GOM Pooled (2011, 2015, 2016)	65	0.201	0.263	71	99.1	0.252
	Bear (2014)	22	0.181	0.259	17	96.1	0.264

7 Abbreviations: # SNPs = number of single nucleotide polymorphisms; N = sample size in final filtered SNP data set; H_O = observed
8 heterozygosity; H_E = expected heterozygosity; # private = number of private alleles per collection; % poly = percentage of polymorphic loci; F_{IS}
9 = inbreeding coefficient; GOM = northern Gulf of Mexico; Bear = Bear Seamount.

10

11

12 Table 3. Overall and pairwise estimates of temporal and spatial genetic differentiation (F_{ST}) and associated significance (P-values) and 95%

13 confidence intervals (CIs) for three lanternfish species: *Diaphus dumerilii*, *Lepidophanes guentheri*, and *Ceratoscopelus warmingii*. Temporal

14 and spatial (Gulf of Mexico (GOM) *vs*. Bear Seamount, *C. warmingii* only) samples were genotyped using single nucleotide polymorphisms

15 (SNPs).

Species (# SNPs)	Temporal-Spatial Comparison	F_{ST}	P-value	95% CI [bs loci]	95% CI [bs ind]
D. dumerilii	GOM Overall	0.011	**0.036**	**(0.0078, 0.0135)**	(-0.0149, 0.0549)
(2577)	GOM 2011 *vs*. GOM 2015	0.008	0.104	**(0.0042, 0.0116)**	(-0.0250, 0.0644)
	GOM 2011 *vs*. GOM 2016	0.020	0.101	**(0.0196, 0.0153)**	(-0.0168, 0.0866)
	GOM 2015 *vs*. GOM 2016	0.005	**0.050**	**(0.0013, 0.0090)**	(-0.0207, 0.0409)
L. guentheri	GOM Overall	0.002	0.100	(0.0000, 0.0039)	(-0.0173, 0.0246)
(3462)	GOM 2011 *vs*. GOM 2015	0.002	0.176	(-0.0010, 0.0044)	(-0.0191, 0.0292)
	GOM 2011 *vs*. GOM 2016	0.001	0.241	(-0.0014, 0.0037)	(-0.0202, 0.0297)
	GOM 2015 *vs*. GOM 2016	0.003	0.100	**(0.0004, 0.0105)**	(-0.0227, 0.0354)
C. warmingii	GOM Overall	0.001	0.218	(-0.0004, 0.0032)	(-0.0096, 0.0161)
(1804)	GOM 2011 *vs*. GOM 2015	0.004	0.077	**(0.0013, 0.0072)**	(-0.0111, 0.0252)
	GOM 2011 *vs*. GOM 2016	0.003	0.128	**(0.0007, 0.0053)**	(-0.0094, 0.0194)
	GOM 2015 *vs*. GOM 2016	-0.003	0.944	(-0.0058, -0.0011)	(-0.0171, 0.0156)
(Cwar-SNP-all; 1804)	GOM (2011, 2015, 2016) *vs*. Bear Seamount (2014)	0.022	**0.001**	**(0.0179, 0.0254)**	**(0.0018, 0.0520)**

16 Abbreviations: # SNPs = number of single nucleotide polymorphisms; bs loci = bootstrapping over loci; bs ind = bootstrapping over individuals.

17 Bold values indicate significant differentiation at $P<0.05$ or 95% confidence intervals that do not overlap zero.

18

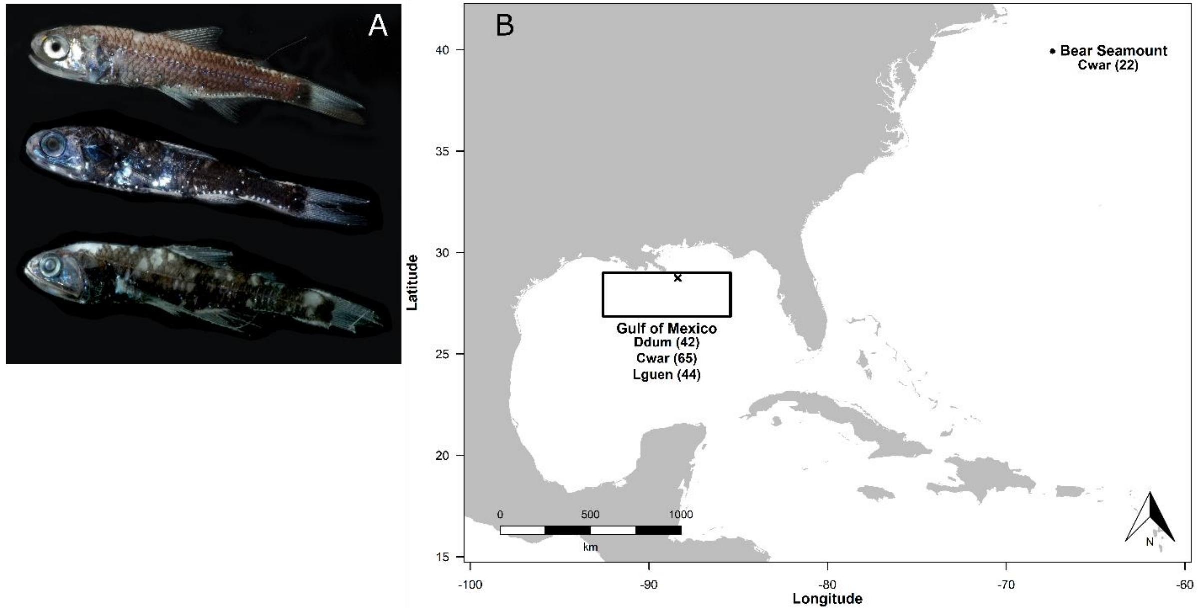

Fig. 1. A. The three myctophid study species sampled in the Gulf of Mexico (© 2022 Danté Fenolio / DEEPEND-RESTORE) from top to bottom in photograph: *Diaphus dumerilii*, Ddum*;* *Ceratoscopelus warmingii*, Cwar*; Lepidophanes guentheri*, Lguen. **B.** Map of sampling locations of three lanternfish species. x, represents the location of the Deepwater Horizon Oil Rig, ●, represents the location of Bear Seamount, western North Atlantic. The rectangle represents the area of sampling in the northern Gulf of Mexico. Bioinformatics post-filtering final sample sizes of each lanternfish species shown in brackets.

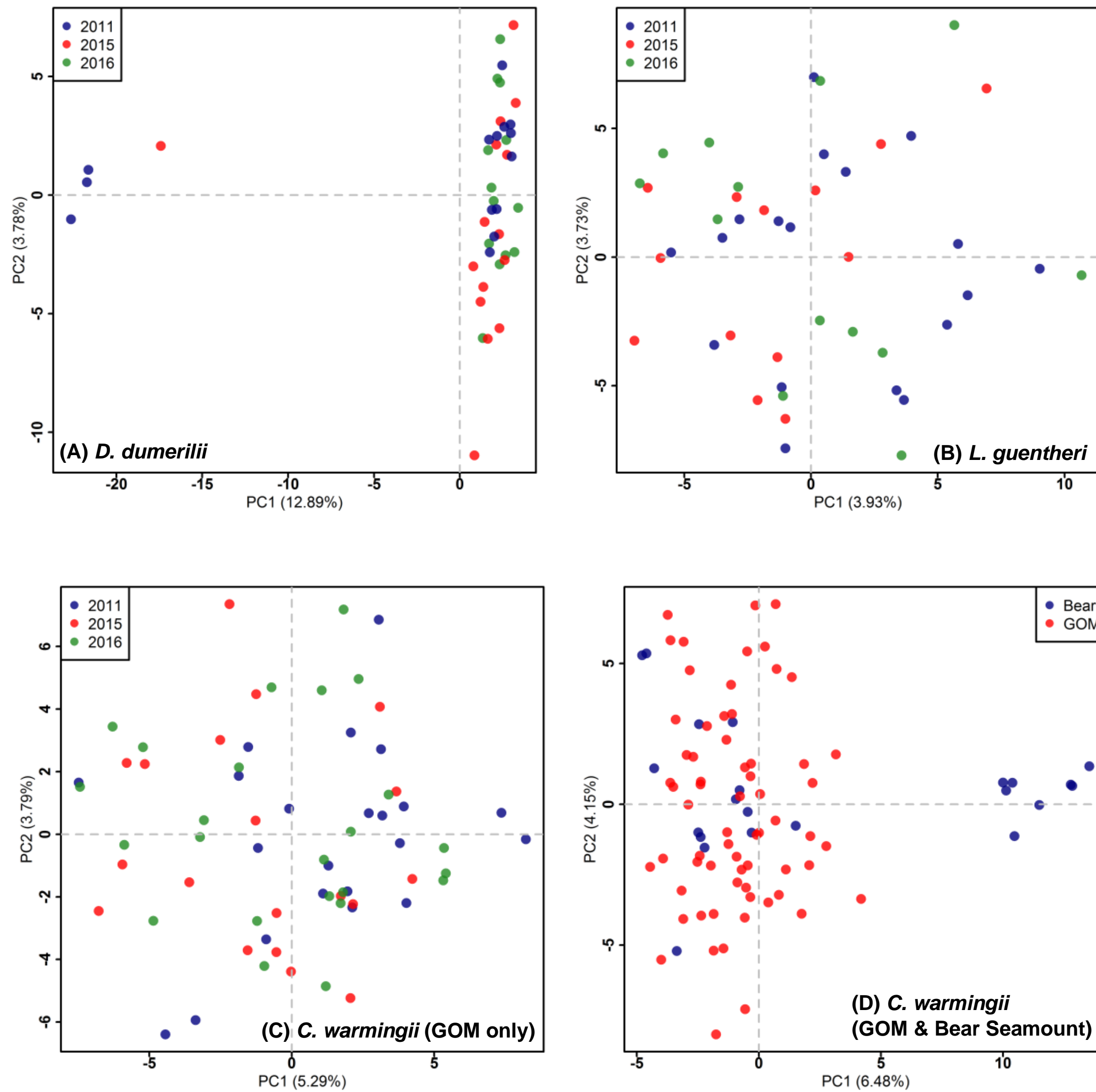

Fig. 2. Biplot displaying the first two principal components (x-axis = component 1, y-axis = component 2) of a Principal Component Analysis of: (**A**) 42 samples of *Diaphus dumerilii* collected from the Gulf of Mexico (GOM) during three temporal periods (2011, 2015, and 2016) and genotyped at 2577 SNPs; (**B**) 44 samples of *Lepidophanes guentheri* collected from the GOM during 2011, 2015, and 2016 and genotyped at 3462 SNPs; (**C**) 65 samples of

Ceratoscopelus warmingii collected from the GOM during 2011, 2015, and 2016 and genotyped at 1804 SNPs; and (**D**) a total 87 samples of *C. warmingii* (Cwar-SNP-all) collected from the GOM (N = 65; red circles) and Bear Seamount (Bear), western North Atlantic (N = 22; blue circles) at 1804 SNPs (Cwar-SNP-all).

Appendix A. Supplemental Materials: Cytochrome *c* oxidase subunit I gene (COI) DNA Barcoding

The 5' end of the mitochondrial Cytochrome *c* oxidase subunit I gene (COI; i.e., the DNA barcode gene, ~650 base pairs (bp)) was sequenced from all individuals included in the final filtered SNP datasets (*Diaphus dumerilli*, N = 42; *Lepidophanes guentheri*, N = 44; *Ceratoscopelus warmingii*, N = 87). Genomic DNA extraction, polymerase chain amplification (PCR), and Sanger Sequencing was performed for most samples (n = 151) by the Canadian Centre for DNA Barcoding (CCDB) (Steinke and Hanner, 2010) and deposited in the Barcode of Life Data System (BOLD, http://www.boldsystems.org). An additional three *D. dumerilii* (Set 1) and 15 *C. warmingii* (Set 2) samples were DNA barcoded in-house (see amplification reaction, thermal cycling, and sequencing conditions below, Table S1 and Table S2).

Table S1. PCR conditions for amplification of the mitochondrial Cytochrome *c* oxidase subunit I gene for three *D. dumerilii* samples.

Reactant [conc]	Volume (µL)
dH_20	34.3
dNTP [10mM]	8.0
10X Buffer	5.0
Forward primer [10µM]	1.25
Reverse primer [10µM]	1.25
HotStar Taq Polymerase	0.2
Template DNA	1.0

The COI gene of three *D. dumerilii* individuals (MX0640-5, MX0640-6, DPND 3294) were amplified in a final volume of 50-µL, per the reaction conditions outlined in Table S1 (Forward

primer: 5’-TCAACCAACCACAAAGACATTGGCAC-3’; Reverse primer: 5’-TAGACTTCTGGGTGGCCAAAGAATCA-3’; Ward et al., 2005) and with the following thermal profile: 15 minute (min) denaturation at 95°C, followed by 35 cycles of 94°C for 1 min, 50°C for 1 min, and 72°C for 2 min, with a final extension of at 72°C for 5 min. Following amplification, electrophoresis of PCR products was performed using a 1.2% TBE agarose gel to check for successful amplification and proper sizing of resultant amplicons. Following amplification, PCR products were purified in-house using QIAquick PCR purification Kits (Qiagen Inc.) or sent to GENEWIZ, Inc. (South Plainfield, NJ) for enzymatic PCR clean-up. All resultant products were sequenced by GENEWIZ using BigDye Terminator v3.1 cycle sequencing reactions on an ABI 3730xl. Sequences were checked for errors and trimmed using the program CLC Sequence Viewer 8.0 (http://www.qiagenbioinformatics.com/).

Table S2. PCR conditions for amplification of the mitochondrial Cytochrome *c* oxidase subunit I gene for 15 *C. warmingii* samples.

Reactant [conc]	Volume (µL)
dH_20	10.2
100X BSA	0.1
$MgCL_2$ [25mM]	2.0
dNTP [10mM]	1.5
5X GoTaq Green PCR buffer	4.0
Forward primer [10µM]	0.5
Reverse primer [10µM]	0.5
GoTaq Green Polymerase	0.2
Template DNA	1.0

The COI gene of 15 *C. warmingii* individuals (RIE 0129, RIE 0198, RIE 0607, RIE 0616, RIE 0629, RIE 0630, RIE 0631, RIE 0632, RIE 0666, RIE 0667, RIE 0669, RIE 0714, RIE 0759, RIE 0768, RIE 0801) were amplified in a final volume of 20-µL, per the reaction conditions outlined in Table S2 (Forward primer: 5’-TTCTCCACCAACCACAARGAYATYGG-3’; Reverse primer: 5’-CACCTCAGGGTGTCCGAARAATCARAA-3’) and with the following thermal

profile: 2 min denaturation at 94°C, followed by 30 cycles of 94°C for 1 min, 50°C for 1 min, and 72°C for 1 min, with a final extension of at 72°C for 5 min. Following amplification, electrophoresis of PCR products was performed using a 1% TBE agarose gel to check for successful amplification and proper sizing of resultant amplicons. PCR clean-up was performed using a standard PEG protocol. Purified PCR products were double-strand sequenced using the BigDye Terminator v3.1 cycle sequencing kit (Applied Biosystems, Inc), followed by sequencing on a ABI 3730 at the Yale Keck School of Medicine. Sequences were checked for errors and trimmed using the program Sequencher 5.1.

Table S3. *Diaphus dumerillii* specimen identification number (#), NCBI GenBank Accession Number (#), and DNA sequence length in base pairs (bp) of the mitochondrial Cytochrome *c* oxidase subunit I gene.

Specimen ID #	GenBank Accession #	Length (bp)
MX0609_1	MN621521	652
MX0609_2	MN621515	652
MX0609_3	MN621531	652
MX0611_1	MN621518	652
MX0611_4	MN621532	652
MX0611_5	MN621517	652
MX0613_1	MN621511	652
MX0619_2	MN621530	652
MX0636_2	MN621525	652
MX0640_1	MN621513	652
MX0640_2	MN621533	652
MX0640_3	MN621527	610
MX0640_5	MN621512	626
MX0640_6	MN621524	626
RIE_1027	MN621516	652
RIE_1028	MN621535	652
RIE_1029	MN621514	652
RIE_1030	MN621528	652
RIE_1031	MN621529	652
RIE_109	MN621509	650
RIE_110	MN621534	652
RIE_111	MN621510	652
RIE_112	MN621522	652

RIE_587	MF041178	652
RIE_608	MF041514	652
RIE_711	MF041495	652
RIE_664	MF041263	652
RIE_808	MN621520	652
RIE_995	MN621519	652
DPND_2571	MG856556	652
DPND_2572	MG856739	652
DPND_3294	MN621526	622
DPND_3297	MG856395	640
DPND_3333	MG856741	652
DPND_3334	MG856780	652
DPND_3335	MG856695	652
DPND_3336	MG856627	652
DPND_3337	MG856769	652
DPND_3382	MG856536	652
DPND_3383	MG856602	652
DPND_3464	MG856534	652
DPND_3532	MG856870	652

Table S4. *Lepidophanes guentheri* specimen identification number (#), NCBI GenBank Accession Number (#), and DNA sequence length in base pairs (bp) of the mitochondrial Cytochrome *c* oxidase subunit I gene.

Specimen ID #	GenBank Accession #	Length (bp)
MX0614_1	MN621492	652
MX0614_2	MN621489	651
MX0617_4	MN621503	652
MX0631_1	MN621494	652
MX0631_2	MN621485	652
MX0631_3	MN621505	631
MX0631_4	MN621487	652
MX0631_5	MN621491	648
MX0649_1	MN621501	652
MX0654_1	MN621499	652
MX0659_1	MN621484	652
MX0659_2	MN621508	652
MX0664_1	MN621502	652
MX0664_2	MN621486	652
MX0664_3	MN621490	652
MX0696_1	MN621506	652
MX0696_2	MN621507	652
MX0697_1	MN621496	652
RIE_227	MF041063	652

RIE_228	MF041427	652
RIE_250	MF040975	652
RIE_318	MF041403	652
RIE_319	MF041575	652
RIE_320	MF041473	641
RIE_321	MF041278	652
RIE_565	MF041483	652
RIE_600	MF041508	643
RIE_601	MF041103	652
RIE_640	MN621488	652
RIE_641	MN621495	652
RIE_710	MN621493	652
DPND_3414	MG856460	652
DPND_3415	MG856724	652
DPND_3416	MG856793	652
DPND_3417	MG856902	652
DPND_3418	MG856784	652
DPND_3419	MG856773	652
DPND_3420	MG856591	652
DPND_3421	MG856683	652
DPND_3535	MG856436	652
DPND_3561	MG856443	652
DPND_3562	MG856607	652
DPND_3563	MG856652	652
DPND_3726	MG856524	652

Table S5. *Ceratoscopelus warmingii* specimen identification number (#), NCBI GenBank Accession Number (#), and DNA sequence length in base pairs (bp), and field identification (ID; if different than the specimen ID) of the mitochondrial gene Cytochrome *c* oxidase subunit I gene.

Specimen ID #	GenBank Accession #	Length (bp)	Additional Field ID
BS_400	MN621545	652	P14-400
BS_401	MN621543	652	P14-401
BS_402	MN621595	645	P14-402
BS_403	MN621589	652	P14-403
BS_404	MN621602	652	P14-404
BS_405	MN621597	652	P14-405
BS_406	MN621569	652	P14-406
BS_411	MN621605	652	P14-411
BS_414	MN621577	652	P14-414
BS_415	MN621549	652	P14-415
BS_416	MN621572	652	P14-416

BS_417	MN621598	652	P14-417
BS_418	MN621586	652	P14-418
BS_420	MN621564	652	P14-420
BS_423	MN621587	652	P14-423
BS_424	MN621551	652	P14-424
BS_426	MN621580	621	P14-426
BS_429	MN621568	652	P14-429
BS_430	MN621596	652	P14-430
BS_431	MN621600	652	P14-431
BS_434	MN621556	652	P14-434
BS_435	MN621609	652	P14-435
DPND_3863	MG856440	652	N/A
DPND_3879	MG856757	652	N/A
DPND_3881	MG856691	652	N/A
DPND_3918	MG856728	652	N/A
DPND_4027	MG856654	652	N/A
DPND_4028	MG856862	652	N/A
DPND_4070	MG856847	652	N/A
DPND_4071	MG856637	652	N/A
DPND_4072	MG856830	652	N/A
DPND_4073	MG856822	652	N/A
DPND_4074	MG856473	652	N/A
DPND_4089	MG856467	652	N/A
DPND_4090	MG856850	652	N/A
DPND_4095	MG856833	652	N/A
DPND_4164	MG856576	652	N/A
DPND_4165	MG856493	652	N/A
DPND_4171	MG856538	652	N/A
DPND_4217	MG856422	652	N/A
DPND_4260	MG856873	652	N/A
DPND_4261	MG856854	652	N/A
DPND_4262	MG856686	652	N/A
DPND_4288	MG856514	652	N/A
DPND_4289	MG856441	652	N/A
DPND_4290	MG856727	652	N/A
PS_036	MN621604	652	PS0036
PS_041	MN621608	652	MX0579-01
PS_042	MN621607	652	MX0579-02
PS_043	MN621566	652	MX0581-01
PS_045	MN621555	652	MX0581-03
PS_049	MN621599	652	MX0583-01
PS_050	MN621573	652	MX0583-02
PS_051	MN621576	652	MX0583-03
PS_052	MN621588	652	MX0583-04
PS_053	MN621550	652	MX0583-05
PS_055	MN621557	652	MX0580-01

PS_056	MN621552	652	MX0580-02
PS_057	MN621547	652	MX0580-03
PS_058	MN621546	652	MX0580-04
PS_059	MN621553	652	MX0580-05
PS_060	MN621565	652	MX0580-06
PS_061	MN621562	652	MX0580-07
PS_062	MN621592	652	MX0580-08
PS_063	MN621571	652	MX0580-09
PS_064	MN621584	652	MX0580-10
PS_065	MN621575	652	MX0580-11
PS_066	MN621601	652	MX0580-12
RIE_1020	MN621603	652	N/A
RIE_1021	MN621559	652	N/A
RIE_1022	MN621544	652	N/A
RIE_129	MN621561	663	N/A
RIE_198	MN621583	597	N/A
RIE_496	MN621606	652	N/A
RIE_607	MN621563	597	N/A
RIE_616	MN621574	597	N/A
RIE_629	MN621539	597	N/A
RIE_630	MN621567	597	N/A
RIE_631	MN621591	663	N/A
RIE_632	MN621582	597	N/A
RIE_666	MN621593	597	N/A
RIE_667	MN621540	597	N/A
RIE_669	MN621585	597	N/A
RIE_759	MN621558	663	N/A
RIE_768	MN621579	663	N/A
RIE_801	MN621554	663	N/A
RIE_933	MN621581	652	N/A

Appendix B. Supplemental Materials: Single Nucleotide Polymorphism Results

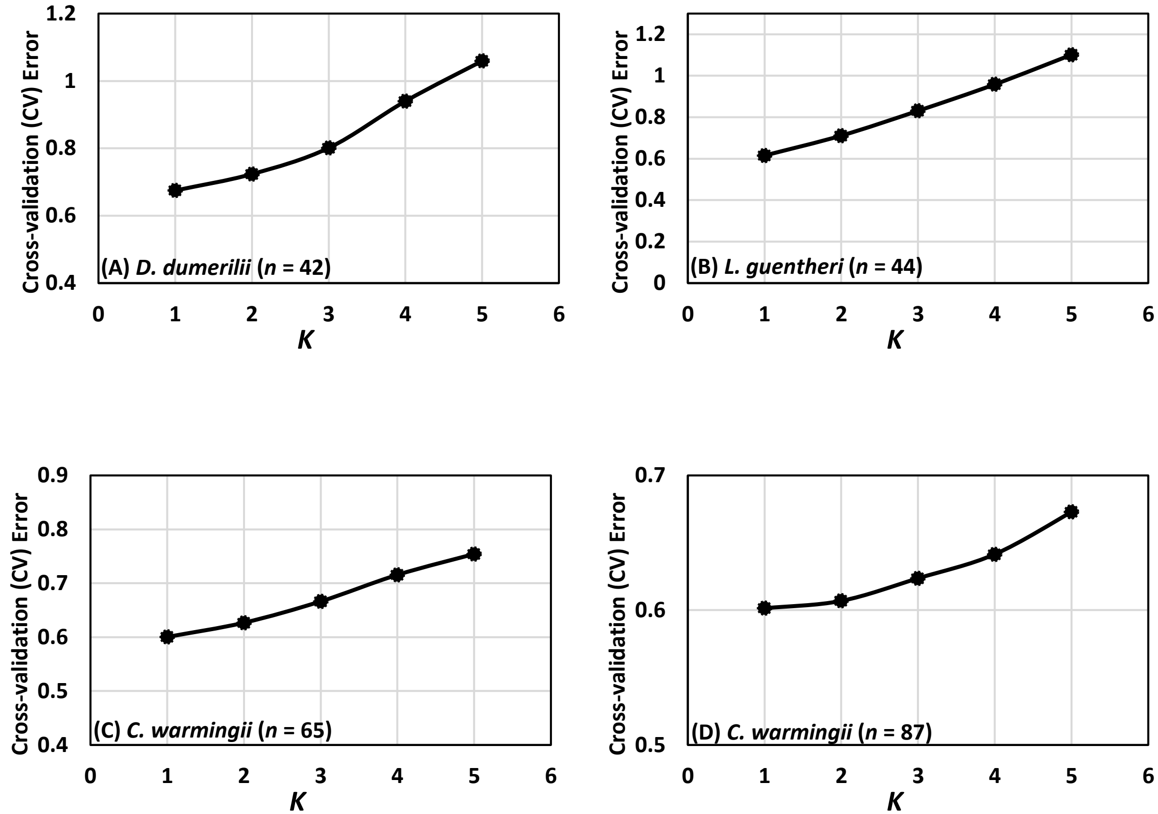

Fig. S1. Plot of Admixture's cross-validation error values for K = 1-5 for: (**A**) 42 samples of *Diaphus dumerilii* collected from the Gulf of Mexico during three temporal periods (2011, 2015, and 2016) and genotyped at 2577 SNPs; (**B**) 44 samples of *Lepidophanes guentheri* collected from the Gulf of Mexico during three temporal periods (2011, 2015, and 2016) and genotyped at 3462 SNPs; (**C**) 65 samples of *Ceratoscopelus warmingii* collected from the Gulf of Mexico during three temporal periods (2011, 2015, and 2016) and genotyped at 1804 SNPs; (**D**) 87 total samples of *C. warmingii* collected from the Gulf of Mexico (N = 65) and from Bear Seamount, western North Atlantic (2014; N = 22) genotyped at 1804 SNPs (Cwar-SNP-all).

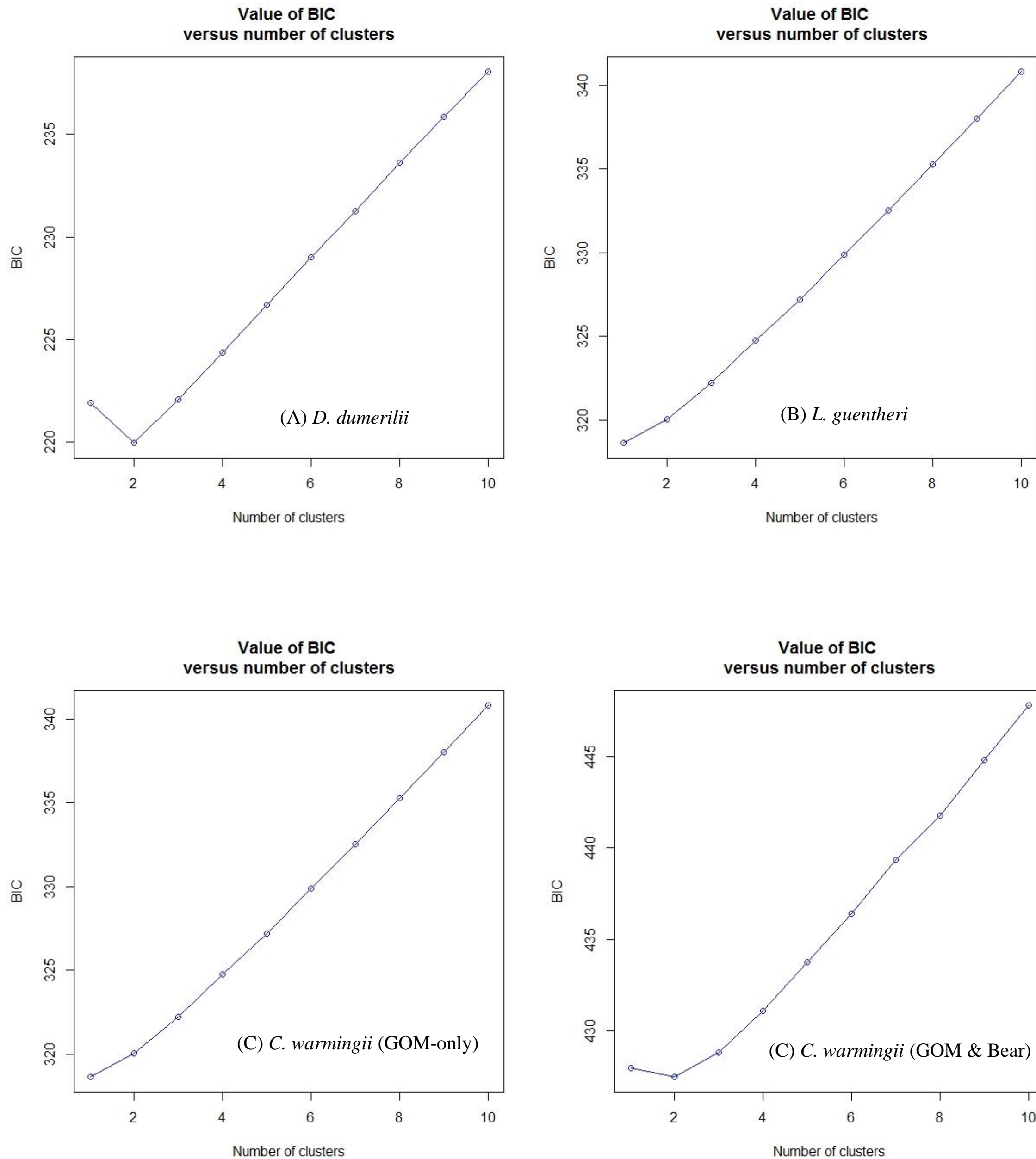

Fig. S2. Plot of the Bayesian Information Criterion (BIC) vs. number of clusters (*K*) of: (**A**) 42 samples of *Diaphus dumerilii* collected from the Gulf of Mexico (GOM) during three temporal periods (2011, 2015, and 2016) and genotyped at 2577 SNPs; (**B**) 44 samples of *Lepidophanes guentheri* collected from the GOM during 2011, 2015, and 2016 and genotyped at 3462 SNPs; (**C**) 65 samples of *Ceratoscopelus warmingii* collected from the GOM during 2011, 2015, and 2016 and genotyped at 1804 SNPs; and (**D**) 87 total samples of *C. warmingii* collected from the GOM (N = 65) and Bear Seamount, western North Atlantic (N = 22) and genotyped at 1804 SNPs (Cwar-SNP-all).

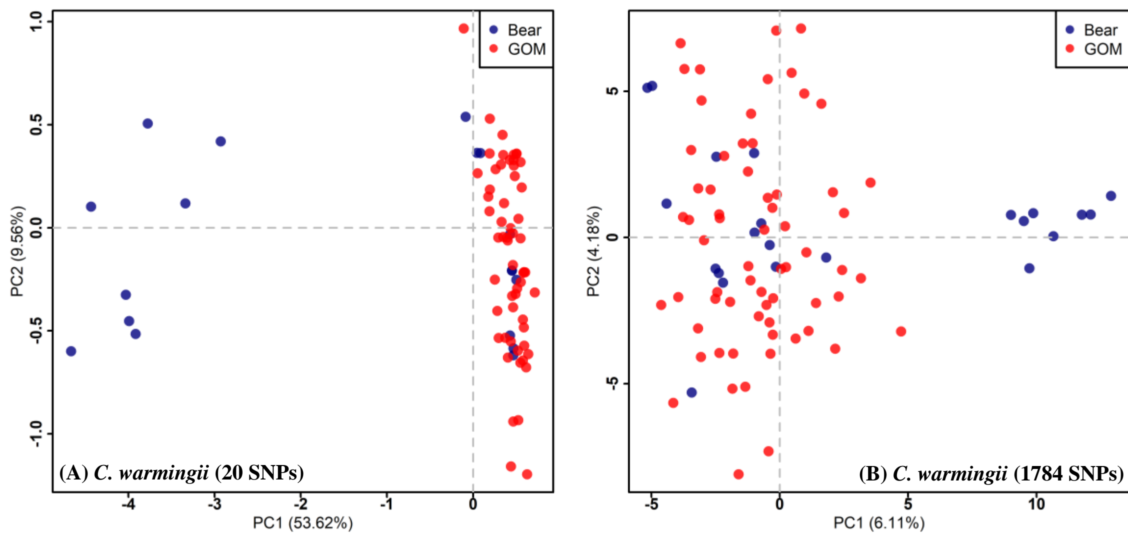

Fig. S3. Biplot displaying the first two principal components (x-axis = component 1, y-axis = component 2) of a Principal Component Analysis of *Ceratoscopelus warmingii* samples collected from the GOM (N = 65; red circles) and Bear Seamount, Western North Atlantic (N = 22; blue circles) at (**A**) 20 candidate outlier SNPs (Cwar-SNP-outlier), and (**B**)1784 putatively neutral SNPs (Cwar-SNP-neutral).
Abbreviations: GOM: Gulf of Mexico; Bear: Bear Seamount, western North Atlantic.

References

Steinke, D., and Hanner, R. (2010) The FISH-BOL collaborators' protocol. Mitochondrial DNA, 21(S2):1-5.

Ward, R.D., Zemlak, T.S., Innes, B.H., Last, P.R., Hebert, P.D. (2005) DNA barcoding Australia's fish species. Philosophical Transactions of The Royal Society B Biological Sciences, 360(1462): 1847-1857.