Genetic evidence from embryos suggests a new species of skate related to *Bathyraja parmifera* (Rajiformes: Arhynchobatidae) in the Bering Sea

Ingrid Spies[1],*, James W. Orr[2], Duane E. Stevenson[2], Pamela Goddard[2], Gerald R. Hoff[2], Jared Guthridge[3], Christopher N. Rooper[4]

[1]**Resource Ecology and Fisheries Management Division, Alaska Fisheries Science Center, 7600 Sand Point Way NE, Seattle, WA 98115, USA**

[2]**Resource Assessment and Conservation Engineering Division, Alaska Fisheries Science Center, 7600 Sand Point Way NE, Seattle, WA 98115, USA**

[3]**Alaska SeaLife Center, PO Box 1329, 301 Railway Ave, Seward, AK 99664, USA**

[4]**Stock Assessment and Research Division, Pacific Biological Station, Fisheries and Oceans Canada, 3190 Hammond Bay Road, Nanaimo, BC V9T 6N7, Canada**

ABSTRACT: Several new species of oviparous skates of the genus *Bathyraja* have been identified over the past 2 decades, yet it is possible that a complete understanding of species diversity among this group has not been achieved. We used genetics and morphology to screen for the presence of species-level differentiation among embryos from nursery areas in the eastern Bering Sea that were initially identified as the Alaska skate *B. parmifera*. A sample of 57 embryos from Pervenets Canyon differed significantly at single nucleotide polymorphism loci from 297 other *B. parmifera* samples, and differences were on the order of those observed between *B. parmifera* and the leopard skate *B. panthera*. This sample of embryos was similar to *B. parmifera* at the cytochrome *c* oxidase subunit I (COI) mitochondrial locus. We discuss whether this group of embryos may represent an undescribed cryptic species. Although morphological description of adults is required for a complete species description, only embryo samples were available. Even so, results presented here are a step toward understanding, managing, and conserving the diversity of skate species in the North Pacific Ocean.

KEY WORDS: Elasmobranch · Speciation · Genetic identification · Skate

1. INTRODUCTION

Accurate recognition of species diversity and associated geographic ranges are essential steps towards management and conservation. Understanding species diversity is particularly important in regions facing threats due to fishing and climate change, especially for species whose life history renders them particularly vulnerable. Research on morphologically cryptic species (2 or more distinct species classified under a single species name) has increased exponentially with the availability of DNA sequence data (Bickford et al. 2007). Sequence diversity in the mitochondrial gene cytochrome *c* oxidase (COI), coined the 'Barcode of Life' (Hebert et al. 2004), can be used to discriminate a high proportion of species within diverse groups (e.g. 92.1% of beetles, Pentinsaari et al. 2014; 93% of Canadian freshwater fishes, Hubert et al. 2008; 98% of Pacific Canadian marine fishes, Steinke et al. 2009). However, this tool is not universally diagnostic (e.g. Spies et al. 2006), and species identification may require morphological examination and taxonomic evaluation (Will & Rubinoff 2004).

*Corresponding author: ingrid.spies@noaa.gov

Currently, 14 skate species are known to occur in the Bering Sea and Aleutian Islands, 11 of which are in the genus *Bathyraja* (Stevenson et al. 2008, Orr et al. 2011). The Alaska skate *B. parmifera* (Bean 1881) is the most abundant skate in the eastern Bering Sea, ranging across the North Pacific from the Bering Sea to southeast Alaska. It is an oviparous elasmobranch that deposits egg cases in nursery areas that persist across years (Stevenson et al. 2008, Hoff 2016). Female *B. parmifera* carrying yolked eggs have been found during all months of the year (Matta 2015), although egg case deposition rates are higher during summer months, indicating annual deposition (Hoff 2008). The presence of multiple cohorts developing simultaneously at nursery areas indicates that eggs incubate for at least 3–4 yr without parental care, the longest recorded development time for oviparous chondrichthyans (Hoff 2008, 2010). In 2015, the North Pacific Fishery Management Council identified 6 nursery sites as Habitat Areas of Particular Concern (see Table 1 for designation names). This designation recognizes the importance of skate egg nursery areas but confers no protection from trawling. Species of the genus *Bathyraja* and other skates with similar life histories are considered vulnerable due to their dependence on nursery areas for embryo development. They are also vulnerable to commercial fisheries trawl gear, by which they are commonly encountered roughly in proportion to the species' abundance (Stevenson et al. 2019). In addition, the temperature recorded at nursery areas during summer surveys falls within a 1°C range, indicating that these species may be affected by thermal shifts under climate change (see Table 1).

The presence of morphologically cryptic speciation is not unprecedented in *B. parmifera* and other species of skates. Smith et al. (2008) found evidence for a cryptic species previously recognized as *B. eatonii* using mtDNA sequence divergence. Similarly, Griffiths et al. (2010) found evidence for cryptic speciation in the common skate *Dipturus batis* with mtDNA and microsatellite DNA markers. *Bathyraja panthera*, the leopard skate, was formally described in 2011 as a congener of *B. parmifera* that resides along the Aleutian Islands chain west of Amchitka Pass (Orr et al. 2011, Spies et al. 2011). This species is distinguished by morphological characters, a unique color pattern, and its geographic range, as well as a unique COI sequence (Orr et al. 2011, Spies et al. 2011). However, differences at COI are not always present among skates and rays, even between known species with distinct morphologies. The Commander and white-blotched skates, *B. lindbergi* and *B. maculata*, exhibit identical COI se-

quences (Spies et al. 2006), as do the yellow and crossback stingarees, *Urolophus sufflavus* and *U. cruciatus* (Ward et al. 2008). Sharks present slower rates of mitochondrial DNA evolution than mammals (Martin et al. 1992); if this is the case for other elasmobranchs, then there may be less diversity at the COI gene than in other species groups. Hybridization and introgression among species should also be considered in studies that examine population structure; for example, the sister skate species *Raja polystigma* and *R. montagui* that co-occur in the Mediterranean have been known to hybridize (Frodella et al. 2016).

Substitutions have been observed at 2 nucleotide sites previously published for *B. parmifera*, positions 366 and 453, resulting in 2 haplotypes (Spies et al. 2011). Specimens of *B. panthera* previously sequenced at the COI gene had a single haplotype distinct from the COI haplotype of *B. parmifera* (Spies et al. 2011). Without differences among COI sequences, and in the absence of obvious morphological characters, cryptic diversity within species could easily go undetected. This problem is exacerbated by the large size of skates such as *B. parmifera*, whose wingspan can reach over 1 m (Matta & Gunderson 2007), hindering collection of adult individuals for morphological analysis.

We applied restriction-site-associated DNA sequencing (RADseq), COI sequencing, and morphology to examine whether there may be evidence for species-level differentiation among *B. parmifera* embryos from nursery areas throughout the eastern Bering Sea (see Fig. 1). RADseq produces thousands of single nucleotide polymorphism (SNP) loci throughout the genome and has served as a complement to COI sequence data (Hebert et al. 2004, Baird et al. 2008, Orr et al. 2019). Samples of *B. panthera* were included for comparison with a closely related species (Orr et al. 2011) (see Fig. 1).

2. MATERIALS AND METHODS

2.1. Sample collection and individual data

Egg cases sampled for skate embryos identified as *Bathyraja parmifera* were collected during Alaska Fisheries Science Center (AFSC) standard eastern Bering Sea bottom trawl surveys in 2006, 2007, and 2016 (see Table 1, Fig. 1). Tissue samples of skates were preserved in 95% ethanol at sea. For reasons described in Section 3, the sample originally identified as *B. parmifera* from Pervenets Canyon in 2016 will be referred to hereafter as *B.* cf. *parmifera*.
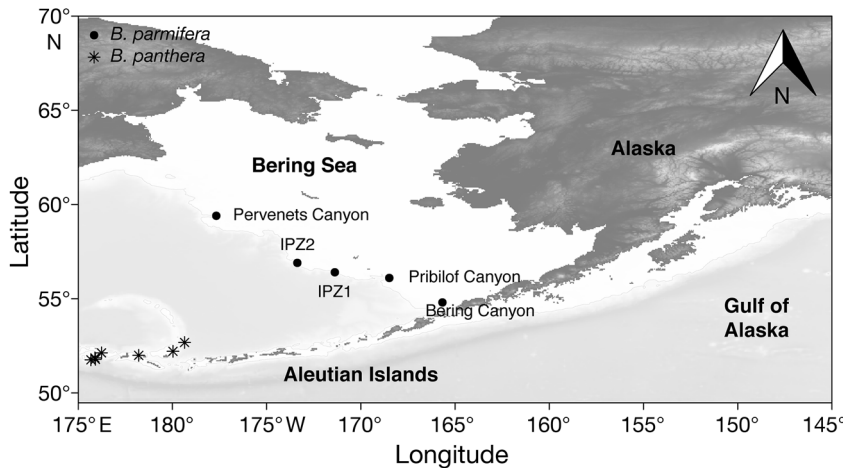
Fig. 1. Nursery site locations in the eastern Bering Sea from which *Bathyraja* embryos were collected: Pervenets Canyon, Inter-Pribilof-Zhemchug-Canyon-1 (IPZ1), Inter-Pribilof-Zhemchug-Canyon-2 (IPZ2), Pribilof Canyon, and Bering Canyon. Collection sites for *B. parmifera* were along the eastern Bering Sea slope, *B.* cf. *parmifera* was collected only at Pervenets Canyon, and *B. panthera* adults (asterisks) were collected in the Aleutian Islands. The 500 m isobath is shown as a light grey line

Collection sites were documented egg case nurseries (Hoff 2008, 2010) (Fig. 1, see Table 1). Bottom trawl gear for collection of embryo samples consisted of the 83–112 eastern otter trawl net (Lauth & Acuna 2007), and tows were limited to 10 min to minimize impacts to the nursery area. We included for comparison tissues of 19 adults of *Bathyraja panthera* collected from the western Aleutian Islands (Rooper 2008) (Fig. 1, Table S1 in Supplement 1 at www.int-res.com/articles/suppl/m670p155_suppl1.pdf).

Embryo developmental stages were assigned based on known embryonic developmental events described by Hoff (2009) (Table S2 in Supplement 2 at www.int-res.com/articles/suppl/m670p155_supp2.xlsx). The total length was measured for all embryos to the nearest mm. Sex was identified for embryos in stages 3–5 (claspers are not visible in stage 2) and only for samples collected in the 2016 survey (Table S2). Vouchered whole embryos and tissues were preserved at the University of Washington Fish Collection: 37 whole embryos of *B.* cf. *parmifera* collected from Pervenets Canyon in 2016 (UW157700–UW157736) and 19 embryos of *B. parmifera* collected from Pervenets Canyon in 2006 (UW155609–UW155627).

2.2. COI sequencing

The COI region was sequenced in a subsample of embryos of *B.* cf. *parmifera* to compare with previ-

ously sequenced specimens of the other known species of skate in the Bering Sea (Spies et al. 2006, 2011), using methodology described therein. Sequencing was performed to identify whether a fixed difference existed between *B. parmifera* and *B.* cf. *parmifera*; therefore, a subsample was considered sufficient. Sequencing was performed at the Molecular Cloning Laboratory (www.mclab.com). Sequences were generated in reverse and forward directions, and trimmed contigs were submitted to GenBank (www.ncbi.com). The sequenced embryos were vouchered at the Burke Museum University of Washington Fish Collection with continuous numbers UW157700–UW157736 corresponding to GenBank Accession numbers MH411166–MH411200. A haplotype network was generated to compare COI haplotypes of *B. parmifera*, *B.* cf. *parmifera*, and *B. panthera* using the R package 'pegas' (Paradis 2010) in R v.4.0.3 (R Core Team 2019).

2.3. Morphological examination

We examined counts of thorns from 4 embryos from Pervenets Canyon collected in 2006 and 5 embryos from Pervenets Canyon collected in 2016. Counts were limited to the largest vouchered and genetically identified embryos. We examined orbital, nasal, scapular, middorsal, trail, and interdorsal thorns following the methods of Orr et al. (2011). Significance was tested using a 2-sample Student's *t*-test with an alpha level of 0.05.

2.4. Generation of SNP data set using RADseq

Restriction-site associated DNA sequencing libraries were constructed for 479 embryo samples and 19 adult *B. panthera* (see Tables 1 & S1). Genomic DNA was extracted from embryo and adult wing tissue using the 96-well format Qiagen DNeasy Blood and Tissue Kit (Qiagen). DNA was quantified using the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen). Dilute DNA was concentrated using a MultiScreen-PCR$_{96}$ filter plate (MilliporeSigma) and eluted in 50 µl H$_2$O. RAD library construction followed the protocol of Baird et al. (2008) and Drinan et al. (2018).

A single library containing 24 embryos of *B. parmifera* from Pervenets Canyon collected in 2006 and 24 embryos from Bering Canyon collected in 2007 (see Table 1) was sequenced on an Illumina HiSeq 2500 (100 bp, single read SR100) at the University of Oregon Genomics and Cell Characterization Core Facility (GC3F). Ten additional libraries, including the *B.* cf. *parmifera* collection, were sequenced on an Illumina HiSeq 4000 (150 bp, single read SR150). A maximum of 48 individuals was used per DNA library due to the large size of the skate genome, which is 3–8 times the size of the genome in teleost fishes (Venkatesh et al. 2014, Tørresen et al. 2017). Raw data were archived at the Sequence Read Archive, which is accessible from the National Center for Biotechnology Information (NCBI; www.ncbi.nlm.nih.gov/bioproject/4306953, BioProject ID: SUB4306953).

Data were downloaded from the GC3F website and processed using Stacks v.1.46 (Catchen et al. 2013). Raw data were quality filtered and demultiplexed using the 'process_radtags' subroutine. All sequences were trimmed to 95 bp to accommodate the difference in sequence lengths (100 and 150 bp), and a population map ('popmap') was used to specify which individuals belonged to each collection location and date to ensure that loci were equally represented in data sequenced on the HiSeq 2500 and the HiSeq 4000.

For all samples, a 2 step process was performed to establish a de novo baseline set of SNPs without duplicates and to effectively combine data that had been run on different platforms. A de novo analysis pipeline was selected that has been found to increase the number of loci and reduce the SNP calling error rates (Mastretta-Yanes et al. 2015). The 10 individuals with the most sequence data from each unique egg case nursery site per year were used in catalog construction during the first and second steps to reduce the detection of false polymorphisms. In the first step, sequence alignment, SNP identification, and construction of a catalog containing all SNPs were performed using the subroutines 'ustacks', 'cstacks', 'sstacks', and 'populations' with default values, except -m 3 (minimum depth of coverage required to create a stack) was used as the minimum depth of coverage required to create a set of aligned sequences, or stack. Default parameter options included -M 2, which was selected to avoid large, repetitive sequences and paralogs (Mastretta-Yanes et al. 2015), where -M is the maximum distance in nucleotides allowed between stacks.

The fasta file of all selected loci was converted to a database using Bowtie software v.2.3.4.1 (Langmead & Salzberg 2012). The Bowtie database was aligned to itself to check for duplicate or inadvertently split loci (-v 3, -k 2), where -v is the number of mismatches regardless of quality, -k is the number of valid alignments to report. Identical loci were culled so that all loci in the database were unique using the function 'RemoveDups.R' (https://github.com/31ingrid/RemoveDups). In the second round of analysis, data from all individuals were aligned to the new database using 'pstacks' with flags (-m 3, -$bound\_high$ 0.05, -max_locus_stacks 3), where $bound\_high$ is the upper bound for the error parameter epsilon, and max_locus_stacks is the maximum number of stacks allowed per single locus. These were followed by subroutines *cstacks* (-g) and *sstacks*. Finally, the subroutine 'populations' was run with no restrictions on the number of SNPs per locus (-m 10, min_maf 0.05), where min_maf is the minimum minor allele frequency. The $r = 0.5$ parameter was also incorporated, which required that 50% of individuals in a population were genotyped to process a locus for that population.

The SNP with the highest minimum allele frequency was then selected from each locus. The output was converted to PLINK software v.1.90b5.3 format (Purcell et al. 2007). Loci genotyped in less than 75% of individuals (-geno 0.25) and individuals with less than 70% genotyping rate (-mind 0.3) were removed because this provided sufficient loci genotyped in a high proportion of individuals in the data set.

Conformation of genotype proportions to Hardy-Weinberg expectations (HWE) was examined for each population using 'hw.test' in the R package 'pegas'. A chi-squared test compared expected genotype frequencies calculated from allelic frequencies to actual genotypes. The false discovery rate method (Benjamini & Hochberg 1995) was used to correct for multiple comparisons at a significance level of 0.05. Loci were discarded if they were out of HWE in 2 or more of the 8 populations.

2.5. Testing for genetic differentiation

Several tests for genetic differentiation among cohorts and egg case nursery sites were performed. The R package StAMPP (Pembleton et al. 2013) was used to estimate pairwise F_{ST}, a commonly used measure of genetic differentiation, among *B. panthera*, *B.* cf. *parmifera*, and all *B. parmifera* calculated according to Weir & Cockerham (1984) with 1000 bootstrap comparisons for significance testing. Fisher's exact test, which examines the hypothesis that alleles are drawn from the same distribution among populations,

was performed among *B. panthera*, *B.* cf. *parmifera*, and all *B. parmifera* across all loci and combined using Fisher's method in the R package GENEPOP (Rousset 2008). StAMPP was also used to perform analysis of molecular variance (AMOVA) among developmental stages in the sample of *B.* cf. *parmifera*.

Discriminant analysis by principal components (DAPC) was performed to identify de novo genetic clusters in the data according to best practices, using the R package 'adegenet' (Jombart 2008, Jombart & Ahmed 2011, Miller et al. 2020). Cross-validation was carried out using the 'xvalDAPC' function in 'adegenet' to determine the optimal number of principal components with the lowest mean squared error. The function 'find.clusters' was used to select the optimal number of clusters without *a priori* information on collection site or year based on the lowest Bayesian information criterion (BIC) value. The inbreeding coefficient (F_{IS}) was calculated for each unique year and egg case nursery site collection and samples of *B. panthera* using the function 'genediv-Fis' in GENEPOP. F_{IS} was examined for statistical significance using $\chi^2 \approx n \times F_{IS}^2$, which follows the chi-square distribution with a single degree of freedom, and n represents the number of individuals (Nei 1987, p. 156, Eq. 7.52). Heterozygosity was calculated per locus and population in 'adegenet'.

2.6. Kinship/relatedness

Relatedness among individuals was determined using the KING v.2.1.3 software (Manichaikul et al. 2010) and COLONY v.2.0.6.5 (Jones & Wang 2010). Using KING, we estimated the kinship coefficient and the proportion of SNP loci, $N_{AA,aa}$, at which pairs of individuals are homozygotes for different alleles ($Pr[IBD=0]$). The kinship coefficient takes into account genetic relatedness using κ_1, and κ_2; the probability that 2 individuals, *i* and *j*, share 1 (κ_1) or 2 (κ_2) alleles identical through common ancestry (Manichaikul et al. 2010). This estimator is defined as:

$$\phi_{i,j} = \frac{\kappa_1}{2} + \kappa_2 = \frac{N_{Aa,Aa} - 2N_{AA,aa}}{2N_{Aa}^i} + \frac{1}{2} - \frac{1}{4}\frac{N_{Aa}^{(i)} + N_{Aa}^{(j)}}{N_{Aa}^{(i)}} \quad (1)$$

where $N_{Aa,Aa}$ is the number of SNP loci at which both individuals are heterozygotes, $N_{AA,aa}$ is the number of SNP loci where individuals are homozygotes for different alleles, and $N_{AA,aa}^{(i)}$ is the number of markers at which an individual *i* is heterozygous (Manichaikul et al. 2010). Dependency among $Pr(IBD=0)$, κ_1, and κ_2 is expected within groups of individuals that interbreed. Therefore, the relationship between $Pr(IBD=0)$ and the kinship coefficient ($\phi_{i,j}$) is expected to be influenced by patterns of relatedness and has been used elsewhere to delineate differences among interrelated individuals (Manichaikul et al. 2010, Kling & Tillmar 2019). These statistics were estimated within and among collections of *B. parmifera*, *B. panthera*, and *B.* cf. *parmifera*.

COLONY considers sibship (the presence of full or half siblings) among all sampled individuals jointly, while KING considers relationships among pairs of individuals. If 2 or more siblings (sharing the same father, or mother, or both) were present in a collection, COLONY uses a maximum likelihood method to find the probability of full or half sibship using multilocus genotypes (Jones & Wang 2010). COLONY was run with polygamy for males and females, medium precision for full likelihood, and no inbreeding. COLONY can only accept a limited number of loci (250), so the number of SNPs was reduced to 243 by selecting SNPs with a minor allele frequency of 40% or greater and a genotyping rate of at least 95%.

The kinship coefficient (KING) and the probability of sibship (COLONY) among collections of *B. parmifera* were used as thresholds for sibship, as sibship among nursery areas is unlikely. Over 200 million individuals of *B. parmifera* were estimated in the Bering Sea in recent stock assessment models (Ormseth 2018). It is not impossible for a female skate to deposit eggs in more than one nursery area, but given the large population size and the low expected fecundity of each skate (between 21 and 37 eggs yr^{-1}; Matta 2015), sibship across nursery areas is exceedingly unlikely.

2.7. Effective population size

We used the program LDNE (Waples & Do 2008) to estimate effective population size (N_e) for *B.* cf. *parmifera* with a random mating model and a jackknife method for estimating confidence intervals, as N_e was estimated for embryos of *B. parmifera* in (Spies et al. 2021). This method uses the amount of linkage disequilibrium within a population to estimate N_e and corrects for bias due to sample size. We selected a minor allele frequency of 0.05 to ensure that the results were not influenced by the presence of rare alleles.

2.8. UPGMA tree

We built an unweighted pair group method with arithmetic mean (UPGMA) tree with the complete SNP data set, including all embryo samples and *B.*

Table 1. Collection information for embryos identified as *Bathyraja parmifera* and *B.* cf. *parmifera*, each collected in a single trawl haul: nursery site name, designation, coordinates: latitude (°N) and longitude (°W), and date of collection, average measured temperature, depth, and number of individuals sequenced (N) before [and after] filtering, the developmental stages within each collection, the number of embryos corresponding to each stage (n), the inbreeding coefficient (F_{IS}) for each collection, and the number of single nucleotide polymorphisms in each collection. IPZ1 and IPZ2: Inter-Pribilof-Zhemchug-Canyon-1 and Inter-Pribilof-Zhemchug-Canyon-2; AK: *B. parmifera*; LS: *B. panthera*. Information on *B. panthera* is provided where applicable; –: not applicable

Site name	Site designation	Coordinates collected	Date	Species	Temp. (°C)	Depth (m)	N	Stage(s)	n	F_{IS}	Alleles
Bering Canyon	AK-10-54.8	54.8, 165.67	3 Oct 2007	AK	4.5	145	24[18]	3, 5	14, 4	0.0189	13317
Bering Canyon	AK-10-54.8	54.8, 165.67	29 Jul 2016	AK	4.5	145	96[69]	2, 4, 5	19, 11, 39	0.0351	13555
IPZ1	AK-31-56.4	56.4, 171.38	3 Jul 2016	AK	4.3	207	96[81]	2, 3, 4, 5	20, 16, 31, 14	0.1091	13570
IPZ2	AK-31-56.9	56.9, 173.37	3 Jul 2016	AK	3.9	217	47[22]	2, 5	1, 21	0.1834	13345
Pervenets Canyon	AK-61-59.4	59.4, 177.67	3 Aug 2006	AK	3.9	316	24[17]	3, 4	15, 2	0.0085	13315
Pervenets Canyon	AK-61-59.4	59.4, 177.67	3 Jul 2016	AK[a]	4.0	316	96[57]	2, 4, 5	11, 22, 24	0.0398	9090
Pribilof Canyon	AK-21-56.1	56.1, 168.49	8 Aug 2016	AK	3.8	205	96[90]	2, 3, 4, 5	23, 5, 37, 25	0.0231	13580
–	–	–	2006	LS	–		19[11]	–	–	0.2075	10572

[a]*B.* cf. *parmifera*

panthera, to show relative branch lengths. The tree was based on Nei's genetic distance (Nei 1987) and was generated using the R package 'poppr' (Kamvar et al. 2014). The mutation rate is not known; therefore, the tree was ultrameric, assuming that the rate of mutation was the same across all lineages. The significance of each node was calculated with 1000 replicate trees.

3. RESULTS

3.1. Sample collection and individual data

Embryo collections contained stages 2 through 4, but not all stages were observed at each nursery area, and embryos of stage 3 were relatively uncommon among the collections taken in 2016 (Tables 1 & S2, Fig. 2). The length ranges are shown for the 354 embryo samples that passed RADseq filtering criteria and were relatively similar across nursery areas and years, although there was some variation in the length at each stage (Table 1, Fig. 2). There were 138 males, 107 females, and 109 unsexed embryos in the final SNP data set (Table S2).

3.2. COI sequencing

The COI region was sequenced in a subsample of 35 embryos from the 2016 Pervenets collection and trimmed to 613 bp. No new COI sequences were present in the embryos of *Bathyraja* cf. *parmifera* that had not been previously observed in *B. parmifera* (Fig. 3) (Spies et al. 2011).

3.3. Morphological examination

Morphological characteristics of embryos of *B. parmifera* collected in 2006 from Pervenets Canyon were limited because these specimens were small, stages 3 and 4. We selected the 5 largest vouchered specimens identified as *B.* cf. *parmifera* collected from Pervenets Canyon in 2016 for comparison. No significant differences in counts of thorns were observed between embryos of *B.* cf. *parmifera* collected from Pervenets Canyon in 2016 and those collected in 2006 (Table 2).

3.4. Generation of SNP data set using RADseq

Libraries were built for the 479 embryos and an additional 19 adult *Bathyraja panthera* collected from the western Aleutian Islands. A total of 34 048 SNPs were present following the stacks analysis, 33 342 SNPs remained following selection of the single SNP with the highest minimum allele frequency, and 8410 SNPs remained after the selection of loci genotyped in at least 75 % of individuals. A total of 125 individuals had >30 % missing data and were removed from the final data set. Eight additional individuals were removed during the stacks pipeline due to poor barcoding and 2 were removed due to suspected contamination. An additional 777 SNPs were removed that were not in HWE in at least 7 of the 8 populations (Table 1). The final data set contained 7633 loci and 365 individuals from tissue samples of 354 embryo samples of *B. parmifera* and *B.* cf. *parmifera* and 11 adults of *B. panthera*. The total genotyping rate was 96.6 % and mean stack depth of coverage was 66.
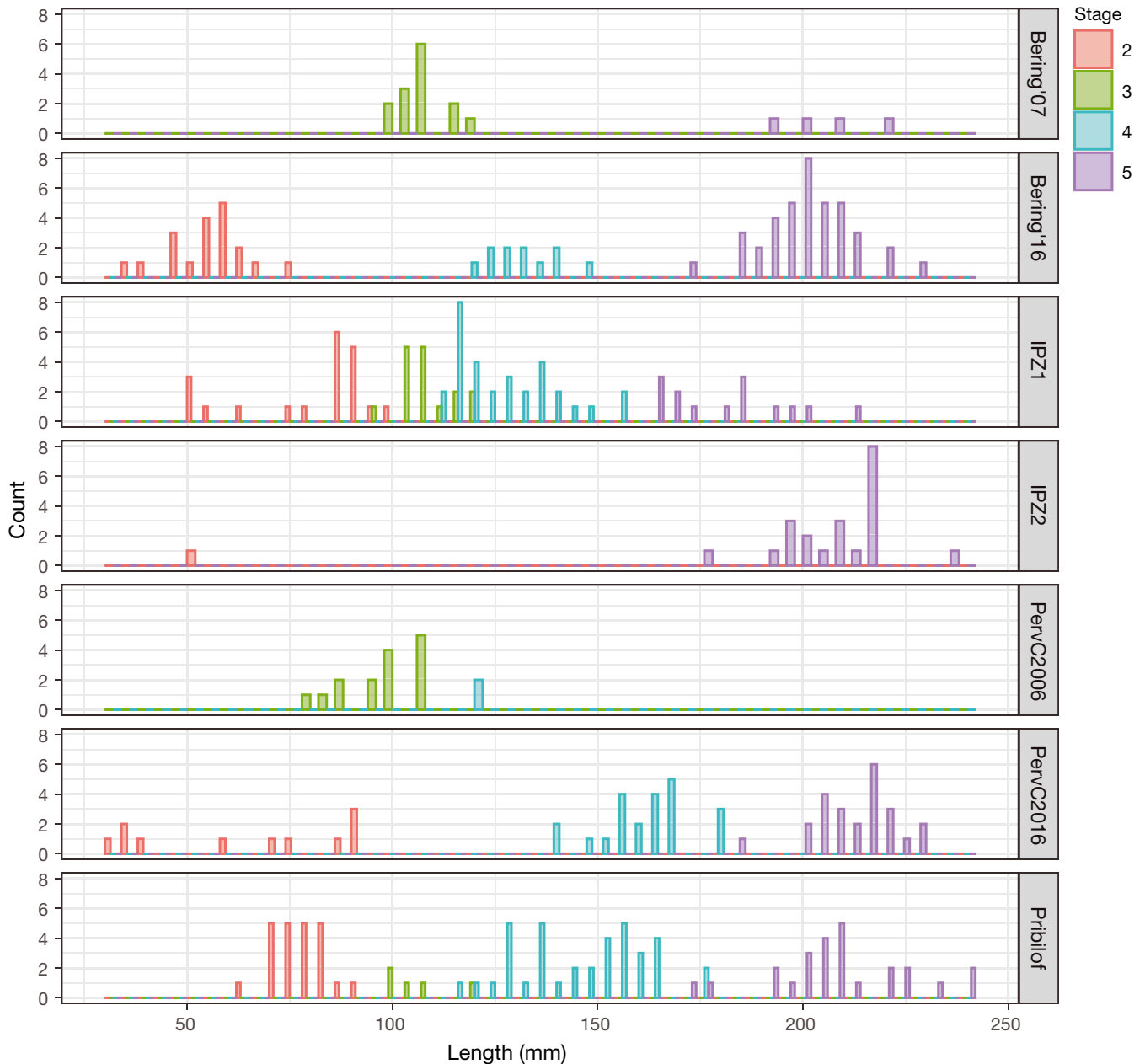
Fig. 2. Counts of estimated stage by length of all *Bathyraja* embryos sequenced using RADseq from the 7 unique nursery area–year combinations: Bering Canyon in 2007 and 2016, Inter-Pribilof-Zhemchug-Canyon-1 (IPZ1, 2016), Inter-Pribilof-Zhemchug-Canyon-2 (IPZ2, 2016), Pervenets Canyon (PervC) in 2006 and 2016, and Pribilof (2016)

3.5. Testing for genetic differentiation

All pairwise Fisher's exact tests were highly significant among *B.* cf. *parmifera*, *B. panthera*, and *B. parmifera* from all other collections (p < 0.0001). Pairwise F_{ST} estimates were congruent with Fisher's exact test results; all were significant and large. The highest pairwise F_{ST} was 0.8331 between *B.* cf. *parmifera* and *B. panthera.* The F_{ST} between *B.* cf. *parmifera* and *B. parmifera* was also high at 0.6039,

while the pairwise F_{ST} between *B. panthera* and *B. parmifera* was 0.1573. AMOVA results indicated no significant differences among the 3 developmental stages observed in the embryos of *B.* cf. *parmifera* ($\sigma^2 = 2.0345 \times 10^{-6}$, error = 5.6939×10^{-4}, p = 0.2128).

DAPC cluster analysis of embryos among nursery sites highlighted the distinctiveness of the embryos of *B.* cf. *parmifera* from those of *B. parmifera* and *B. panthera* (Fig. 4). Cross-validation selected 200 principal components, resulting in a mean squared error of zero.
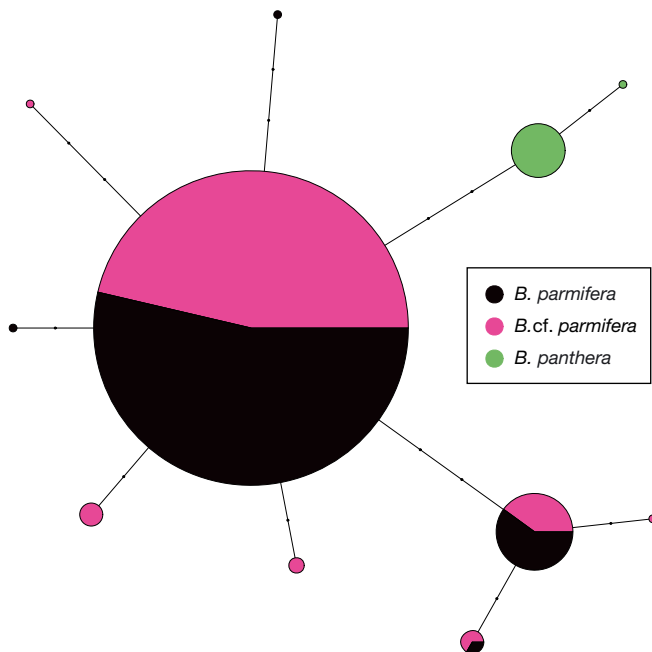
Fig. 3. Haplotype network of 35 *Bathyraja* cf. *parmifera*, 8 *B. panthera*, and 31 *B. parmifera* cytochrome oxidase subunit I sequences, trimmed to 613 bp

The 'find.clusters' algorithm found that *k* = 3 groups was associated with the lowest BIC value (Table S3 in Supplement 1) and placed *B. parmifera*, *B. cf. parmifera*, and *B. panthera* into separate groups with no *a priori* information. The DAPC clustered *B. panthera* and *B.* cf. *parmifera* in different quadrants, with *B. parmifera* located centrally. The first axis, which primarily separated *B.* cf. *parmifera* from *B. parmifera*, explained 41.1% of the variance and the second axis, which primarily separated *B. panthera* from *B. parmifera*, explained 1.3% of the variance. Samples of *B. parmifera* grouped closely together (Fig. 4).

F_{IS} was positive but small (<0.1) for all embryo collections except for IPZ1 and IPZ2 (Inter-Pribilof-Zhemchug-Canyon-1 and -2, respectively), which were between 0.1 and 0.2 (Table 1). None of the ob-

served F_{IS} estimates were significantly different from zero. F_{IS} was highest in *B. panthera* (>0.2), and lack of relatedness (Wahlund effect) is consistent with a collection of randomly sampled adults (Fig. 1, Table 1).

3.6. Kinship/relatedness

Three measures of relatedness were used to examine various aspects of the data. The kinship coefficient of KING vs. the *Pr*(IBD=0), intended to provide information on genetic differences among groups, showed cohesive patterns within embryos of *B.* cf. *parmifera*, embryos of *B. parmifera*, and *B. panthera*. Differences among these groups were demarcated by physical spacing between them (Fig. 5). *B. panthera* was centrally placed between *B.* cf. *parmifera* and *B. parmifera*.

The mean *Pr*(IBD=0), or probability of zero identity by state, was a measure of the number of SNPs loci with fixed differences among pairwise sample comparisons. *Pr*(IBD=0) within *B.* cf. *parmifera* was 1.6%; within *B. parmifera*, 5.2%; and within *B. panthera*, 4.4%. The average proportion of SNPs with zero identity by state between *B. panthera* and *B. parmifera* was 8.5%. In comparison, the probability of zero identity by state among pairs of *B. parmifera* and *B.* cf. *parmifera* was 32.7% and among pairs of *B.* cf. *parmifera* and *B. panthera* was 37.1%.

Finally, the highest kinship coefficient (KING) among nursery collections was 0.15, and higher values were not recorded within any nursery area. Similarly, the highest probability of sibship among nursery areas was 0.703 (COLONY), and higher values were not found within nursery areas. Therefore, close kin relationships were considered unlikely within the *B.* cf. *parmifera* collection or within any of the *B. parmifera* collections because no pairwise comparisons were higher than among *B. parmifera* nursery areas.

3.7. N_e

The N_e over 3 cohorts for *B.* cf. *parmifera* was estimated to be 610.9 (95% CI = 496.4–791.6).

3.8. UPGMA Tree

The UPGMA tree with the longest branch length belonged to *B.* cf.

Table 2. Thorn counts in embryos of *Bathyraja parmifera* (n = 4) and *B.* cf. *parmifera* (n = 5) identified by genetics

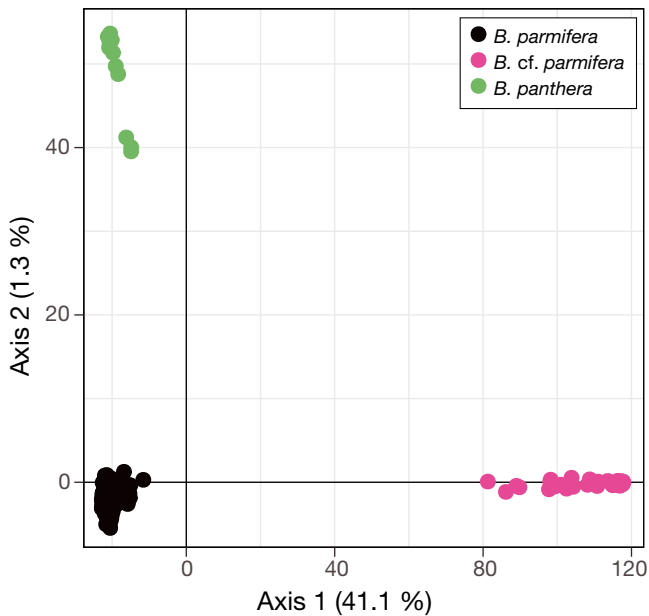| Character	*B. parmifera*		*B.* cf. *parmifera*	
	Range	Mean ± SD	Range	Mean ± SD
Total length (mm)	190–200	196 ± 4.3	207–220	212 ± 5.4
Orbital thorns	1–10	5.5 ± 3.9	8–10	9.2 ± 1.1
Nasal thorns	3–4	3.3 ± 0.5	2–3	2.8 ± 0.4
Scapular thorns	1–2	1.5 ± 0.6	2–2	2 ± 0
Middorsal thorns	5–7	6.5 ± 1	0–7	5.2 ± 2.9
Tail thorns	16–21	18.5 ± 2.1	19–23	20.4 ± 1.9
Interdorsal thorns	0–1	0.5 ± 0.6	1–2	1.2 ± 0.4

Fig. 4. Discriminant analysis of principal components based on genotypes at 7633 single nucleotide polymorphism loci for embryos of *Bathyraja parmifera* (n = 297), *B.* cf. *parmifera* (n = 57), and adult *B. panthera* (n = 11) using 200 principal components

parmifera, followed by *B. panthera* (Fig. 6). The shortest branch lengths were associated with samples of *B. parmifera* (represented as IPZ1, Pribilof, Bering 2016, Pervenets 2006, Bering 2007, and IPZ2). Bootstrap support for *B. panthera*, *B. parmifera*, and *B.* cf. *parmifera* were all 100%, and support for nodes within embryo samples of *B. parmifera* ranged from 60–100%.

4. DISCUSSION

Morphologically cryptic species likely represent a significant proportion of undetected biodiversity (Struck et al. 2018), and new species of chondrichthyans continue to be discovered (White & Last 2012, Borsa et al. 2018). Identification of species diversity is critical to informing conservation and management policy (Delić et al. 2017). The most abundant oviparous skate in the Bering Sea, *Bathyraja parmifera*, was shown to represent a species complex in 2011 with the identification of the allopatric congener *Bathyraja panthera* (Orr et al. 2011, Spies et al. 2011).

Here, we found evidence of previously undetected genetic diversity among embryos identified as *B. parmifera.*

We used RADseq to examine broad-scale differences among embryos identified as *B. parmifera* followed by the traditional approaches of mtDNA sequencing and morphology, for examination of specimens with anomalous SNP genotypes. The main finding of this study was the detection of a genetically distinct group of embryos, referred to as *B.* cf. *parmifera*, that was collected in Pervenets Canyon in the eastern Bering Sea in 2016. While multiple skate species use nursery sites synchronously (Hoff 2010), finding *B.* cf. *parmifera* was unexpected, as earlier sampling in 2006 in Pervenets Canyon confirmed the presence of embryos of *B. parmifera.*

In contrast to SNP genotypes, no unique COI sequences were observed among *B.* cf. *parmifera* embryos that had not been observed in previously sequenced specimens of *B. parmifera* (Fig. 3) (Spies et al. 2006). Two haplotypes previously observed in *B. parmifera* were also observed in *B.* cf. *parmifera*, at similar frequencies. Other species of the subgenus *Arctoraja* recognized from the western North Pacific, *B. smirnovi* and *B. simoterus*, differ from *B. parmifera* at no fewer than 5 positions in COI, and *B. panthera* of the western Aleutian Islands differs at no less than 2 positions (Orr et al. 2011, Spies et al. 2011).

Although genetic results indicate that *B.* cf. *parmifera* may represent a morphologically cryptic species, we also considered other explanations. Mitochondrial DNA is typically passed from mother to offspring (Luo et al. 2018), so it is conceivable that hybridization between 2 distinct groups could show differences among nuclear loci but not mtDNA. However, admixture among distinct groups was not
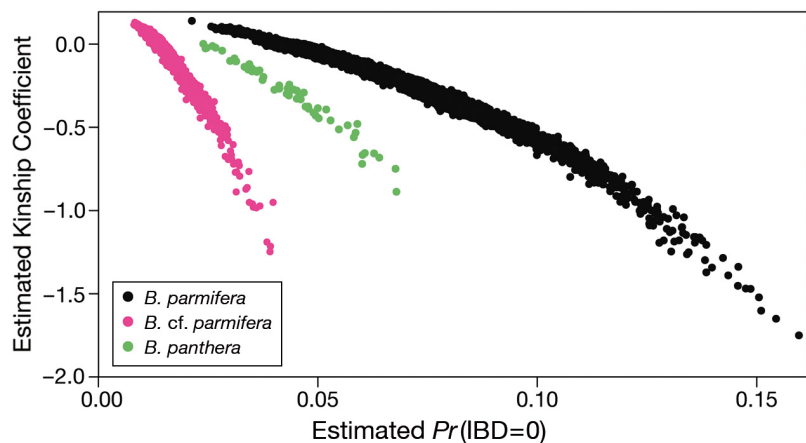

Fig. 5. Probability that identity by state is zero, *Pr*(IBD=0), vs. the kinship coefficient for *Bathyraja parmifera*, *B.* cf. *parmifera*, and *B. panthera*
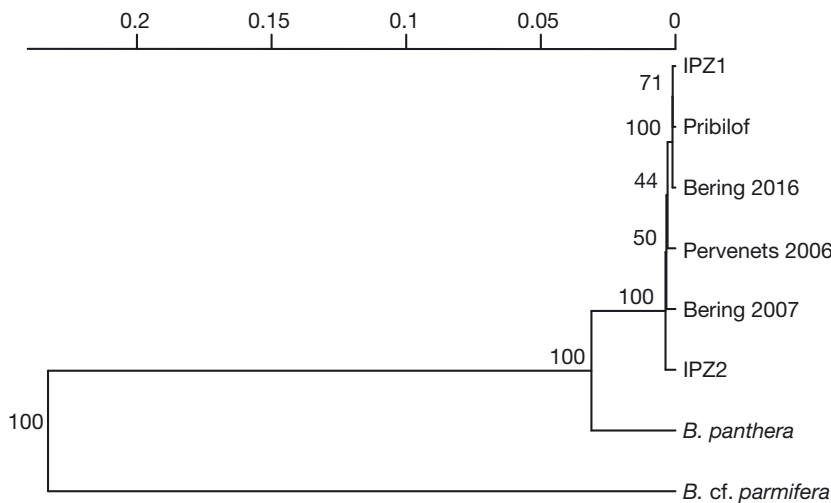
Fig. 6. Rooted unweighted pair group method with arithmetic mean (UPGMA) tree of unique *Bathyraja* embryo collection and year combinations and *B. panthera*, based on single nucleotide polymorphism data set. Distances were calculated according to Nei's genetic distance. Note that collections IPZ1, Pribilof, Bering 2016, Pervenets 2006, Bering 2007, and IPZ2 represent *B. parmifera*

consistent with our finding of >30% of loci fixed for distinct alleles (*Pr*[IBD=0]) because hybridization would result in heterozygous genotypes rather than widespread fixed differences. A second explanation for a distinct population could be recent expansion from a limited number of individuals. Indeed, this may be consistent with evolution of a distinct group of skates, but a recent origin is unlikely because of the slow rate of evolution of elasmobranchs. In fact, the longest branch node in the UPGMA tree belonged to *B.* cf. *parmifera*, contradicting the idea of a recent origin of this group.

While the lack of adult specimens for morphological comparison represented a significant limitation to this study, the genetic results were compelling. We restricted our morphological comparison to genetically identified embryos of similar development, and therefore our sample size of embryos was not substantial. Embryos of *B. parmifera* were morphologically indistinguishable from *B.* cf. *parmifera* but we may have missed variability that may have been discovered if more specimens were available. Typically, morphological characters that distinguish among species and sexes of skates develop with growth and maturity (Braccini & Chiaramonte 2002, Geniz et al. 2007). Morphology is generally conserved among closely related species of skates and is demonstrably so among species of *Arctoraja* (Orr et al. 2011). In contrast to our unremarkable morphological results, pairwise F_{ST} between *B.* cf. *parmifera* and *B. pan-*

thera and between *B.* cf. *parmifera* and *B. parmifera* were considerably higher than typically observed within marine species. Waples (1998) found a mean within-species F_{ST} of 0.062 in a survey of 57 marine species whereas we observed $F_{ST} > 0.6$. Further, simulation studies indicate de novo DAPC clustering success to be highly (>90%) accurate when pairwise F_{ST} values are comparable to those observed here and population sizes consist of 100 individuals (Miller et al. 2020). Finally, RADseq has been used elsewhere to provide information on phylogeny and discriminate among species of turtles (Georges et al. 2018) and snailfishes (Orr et al. 2019).

The naming conventions for *B. parmifera* and *B.* cf. *parmifera* merit some explanation. Which genotype shares genetic identity with the type material for *B. parmifera* may never be known, since the type specimens were collected in the 1880s and are partially disintegrated (Bean 1881, Orr et al. 2011). The *B. parmifera* genotype embryos appear to be much more abundant and widespread than the *B.* cf. *parmifera* genotype embryos. Furthermore, the type locality of *B. parmifera* is in the Aleutian Islands, and the *B.* cf. *parmifera* embryos have only been identified from Pervenets Canyon, over 600 km from the type locality, while multiple nursery sites (which apparently include only the *B. parmifera* genotype embryos) are much closer. In summary, it seems more plausible that the abundant genotype matches the original type material of *B. parmifera*. Thus, it is logical to call the more abundant genotype *B. parmifera* and the other *B.* cf. *parmifera*, as it conveys this informal hypothesis.

Overall, this work provides evidence for the presence of previously undetected genetic diversity in embryos that are morphologically indistinguishable from the closely related *B. parmifera*. Future work will require the collection and examination of adult specimens genetically matching the embryos of *B.* cf. *parmifera*, possibly through genetic testing in the field, to conduct the thorough morphological analysis required to diagnose and describe this distinct group of skates. Examination of the biology, habitat preferences, and range of *B.* cf. *parmifera* will clarify the ecological and evolutionary implications that led to its genetic distinctiveness. More research on the N_e

of this group is important, as the harmonic mean N_e over 3 cohorts of the *B.* cf. *parmifera* collection was 611, on the order of values observed for *B.* cf. *parmifera*; N_e = 1257 (95% CI = 1190−1332) from the IPZ1 nursery area and N_e = 890 (95% CI = 890−905) from the Bering/Pribilof nursery area (Spies et al. 2021). An N_e of 1000 is a known threshold for conservation (Franklin 1980, Lande 1994), while N_e > 100 may prevent short-term genetic erosion (Domingues et al. 2018). Further research on *B.* cf. *parmifera* will improve the conservation and management of the diversity of skate species in the Bering Sea.

LITERATURE CITED

Baird NA, Etter PD, Atwood TS, Currey MC and others (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLOS ONE 3:e3376

Bean TH (1881) Descriptions of new fishes from Alaska and Siberia. Proc U S Natl Mus 4:144−159

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 57:289−300

Bickford D, Lohman DJ, Sodhi NS, Ng PK and others (2007) Cryptic species as a window on diversity and conservation. Trends Ecol Evol 22:148−155

Borsa P, Arlyza IS, Hoareau TB, Shen KN (2018) Diagnostic description and geographic distribution of four new cryptic species of the blue-spotted maskray species complex (Myliobatoidei: Dasyatidae; *Neotrygon* spp.) based on DNA sequences. J Oceanol Limnol 36:827−841

Braccini JM, Chiaramonte GE (2002) Intraspecific variation in the external morphology of the sand skate. J Fish Biol 61:959−972

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. Mol Ecol 22:3124−3140

Delić T, Trontelj P, Rendoš M, Fišer C (2017) The importance of naming cryptic species and the conservation of endemic subterranean amphipods. Sci Rep 7:1−12

Domingues RR, Hilsdorf AWS, Gadig OBF (2018) The importance of considering genetic diversity in shark and ray conservation policies. Conserv Genet 19:501−525

Drinan DP, Gruenthal KM, Canino MF, Lowry D, Fisher MC, Hauser L (2018) Population assignment and local adaptation along an isolation-by-distance gradient in Pacific cod (*Gadus macrocephalus*). Evol Appl 11:1448−1464

Franklin IR (1980) Evolutionary change in small populations. In: Soule ME, Wilcox BA (eds) Conservation biology: an evolutionary ecological perspective. Sinauer Associates, Sunderland, MA, p 135−140

Frodella N, Cannas R, Velonà A, Carbonara P and others (2016) Population connectivity and phylogeography of the Mediterranean endemic skate *Raja polystigma* and evidence of its hybridization with the parapatric sibling *R. montagui*. Mar Ecol Prog Ser 554:99−113

Geniz JLC, Nishizaki OS, Jiménez JCP (2007) Morphological variation and sexual dimorphism in the California skate, *Raja inornata* Jordan and Gilbert, 1881 from the Gulf of California, Mexico. Zootaxa 1545:1−16

Georges A, Gruber B, Pauly GB, White D and others (2018) Genomewide SNP markers breathe new life into phylogeography and species delimitation for the problematic short-necked turtles (Chelidae: *Emydura*) of eastern Australia. Mol Ecol 27:5195−5213

Griffiths AM, Sims DW, Cotterell SP, El Nagar A and others (2010) Molecular markers reveal spatially segregated cryptic species in a critically endangered fish, the common skate (*Dipturus batis*). Proc R Soc B 277:1497−1503

Hebert PD, Stoeckle MY, Zemlak TS, Francis CM (2004) Identification of birds through DNA barcodes. PLOS Biol 2:e312

Hoff GR (2008) A nursery site of the Alaska skate (*Bathyraja parmifera*) in the eastern Bering Sea. Fish Bull (Wash DC) 106:233−244

Hoff GR (2009) Embryo developmental events and the egg case of the Aleutian skate *Bathyraja aleutica* (Gilbert) and the Alaska skate *Bathyraja parmifera* (Bean). J Fish Biol 74:483−501

Hoff GR (2010) Identification of skate nursery habitat in the eastern Bering Sea. Mar Ecol Prog Ser 403:243−254

Hoff GR (2016) Identification of multiple nursery habitats of skates in the eastern Bering Sea. J Fish Biol 88: 1746−1757

Hubert N, Hanner R, Holm E, Mandrak NE and others (2008) Identifying Canadian freshwater fishes through DNA barcodes. PLOS ONE 3:e2490

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

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

Jones OR, Wang J (2010) COLONY: a program for parentage and sibship inference from multilocus genotype data. Mol Ecol Resour 10:551−555

Kamvar ZN, Tabima JF, Grünwald NJ (2014) poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2:e281

Kling D, Tillmar A (2019) Forensic genealogy — a comparison of methods to infer distant relationships based on dense SNP data. Forensic Sci Int Genet 42:113−124

Lande R (1994) Risk of population extinction from fixation of new deleterious mutations. Evolution 48:1460−1469

Langmead B, Salzberg S (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357−359

Lauth B, Acuna E (2007) Results of the 2006 eastern Bering Sea continental shelf bottom trawl survey of groundfish and invertebrate resources. NOAA Tech Memo NMFS-AFSC-176

Luo S, Valencia CA, Zhang J, Lee NC and others (2018) Biparental inheritance of mitochondrial DNA in humans. Proc Natl Acad Sci USA 115:13039–13044

Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM (2010) Robust relationship inference in genome-wide association studies. Bioinformatics 26:2867–2873

Martin AP, Naylor GJ, Palumbi SR (1992) Rates of mitochondrial DNA evolution in sharks are slow compared with mammals. Nature 357:153–155

Mastretta-Yanes A, Arrigo N, Alvarez N, Jorgensen TH, Piñero D, Emerson BC (2015) Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. Mol Ecol Resour 15:28–41

Matta ME (2015) Reproductive biology of the Alaska skate *Bathyraja parmifera*, with comments on an intersexual individual. J Fish Biol 87:664–678

Matta ME, Gunderson DR (2007) Age, growth, maturity, and mortality of the Alaska skate, *Bathyraja parmifera*, in the eastern Bering Sea. In: Ebert D, Sulikowski DA (eds) Biology of skates. Springer, Dordrecht, p 203–217

Miller JM, Cullingham CI, Peery RM (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

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York, NY

Ormseth O (2018) Assessment of the skate stock complex in the Bering Sea and Aleutian Islands. NPFMC Bering Sea and Aleutian Islands Stock Assessment and Fishery Evaluation. https://www.afsc.noaa.gov/REFM/Stocks/assessments.htm

Orr JW, Stevenson DE, Hoff GR Spies IB, McEachran JD (2011) *Bathyraja panthera*, a new species of skate (Rajidae: Arhynchobatinae) from the western Aleutian Islands, and resurrection of the subgenus *Arctoraja* Ishiyama. NOAA Prof Pap NMFS 11

Orr JW, Spies I, Stevenson DE, Longo GC, Kai Y, Ghods S, Hollowed M (2019) Molecular phylogenetics of snailfishes (Cottoidei: Liparidae) based on mtDNA and RAD-seq genomic analyses, with comments on selected morphological characters. Zootaxa 4642:1–79

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

Pembleton LW, Cogan NOI, Forster JW (2013) StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. Mol Ecol Resour 13:946–952

Pentinsaari M, Hebert PD, Mutanen M (2014) Barcoding beetles: a regional survey of 1872 species reveals high identification success and unusually deep interspecific divergences. PLOS ONE 9:e108651

Purcell S, Neale B, Todd-Brown K, Thomas L and others (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575

Core Team (2019) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. www.r-project.org

Rooper CN (2008) Data report: 2006 Aleutian Islands bottom trawl survey. NOAA Tech Memo NMFS-AFSC-179

Rousset F (2008) GENEPOP′007: a complete reimplementation of the GENEPOP software for Windows and Linux. Mol Ecol Resour 8:103–106

Smith PJ, Steinke D, McVeagh SM, Stewart AL, Struthers CD, Roberts CD (2008) Molecular analysis of Southern Ocean skates (*Bathyraja*) reveals a new species of Antarctic skate. J Fish Biol 73:1170–1182

Spies IB, Gaichas S, Stevenson DE, Orr JW, Canino MF (2006) DNA-based identification of Alaska skates (*Amblyraja*, *Bathyraja* and *Raja*: Rajidae) using cytochrome *c* oxidase subunit I (coI) variation. J Fish Biol 69:283–292

Spies IB, Stevenson DE, Orr JW, Hoff GR (2011) Molecular systematics of the skate subgenus *Arctoraja* (*Bathyraja*: Rajidae) and support for an undescribed species, the leopard skate, with comments on the phylogenetics of *Bathyraja.* Ichthyol Res 58:77–83

Spies I, Orr JW, Stevenson DE, Goddard P and others (2021) Skate egg nursery areas support genetic diversity of Alaska and Aleutian skates in the Bering Sea. Mar Ecol Prog Ser 669:121–138

Steinke D, Zemlak TS, Boutillier JA, Hebert PD (2009) DNA barcoding of Pacific Canada′s fishes. Mar Biol 156: 2641–2647

Stevenson DE, Orr JW, Hoff GR, McEachran JD (2008) Emerging patterns of species richness, diversity, population density, and distribution in the skates (Rajidae) of Alaska. Fish Bull (Wash DC) 106:24–39

Stevenson DE, Hoff GR, Orr JW, Spies I, Rooper CN (2019) Interactions between fisheries and early life stages of skates in nursery areas of the eastern Bering Sea. Fish Bull 117:8–14

Struck TH, Feder JL, Bendiksby M, Birkeland S and others (2018) Finding evolutionary processes hidden in cryptic species. Trends Ecol Evol 33:153–163

Tørresen OK, Star B, Jentoft S, Reinar WB and others (2017) An improved genome assembly uncovers prolific tandem repeats in Atlantic cod. BMC Genomics 18:95

Venkatesh B, Lee AP, Ravi V, Maurya AK and others (2014) Elephant shark genome provides unique insights into gnathostome evolution. Nature 505:174–179

Waples RS (1998) Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. J Hered 89:438–450

Waples RS, Do C (2008) LDNE: a program for estimating effective population size from data on linkage disequilibrium. Mol Ecol Resour 8:753–756

Ward RD, Holmes BH, White WT, Last PR (2008) DNA barcoding Australasian chondrichthyans: results and potential uses in conservation. Mar Freshw Res 59:57–71

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. Evolution 38: 1358–1370

White WT, Last PR (2012) A review of the taxonomy of chondrichthyan fishes: a modern perspective. J Fish Biol 80: 901–917

Will KW, Rubinoff D (2004) Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. Cladistics 20:47–55