Received: 15 June 2020 | Revised: 20 January 2021 | Accepted: 27 January 2021

DOI: 10.1111/mec.15827

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Population structure in a continuously distributed coastal marine species, the harbor porpoise, based on microhaplotypes derived from poor-quality samples

Phillip A. Morin[1] | Brenna R. Forester[2] | Karin A. Forney[3,4] | Carla A. Crossman[5,6] | Brittany L. Hancock-Hanser[1] | Kelly M. Robertson[1] | Lance G. Barrett-Lennard[6] | Robin W. Baird[7] | John Calambokidis[7] | Pat Gearin[8] | M. Bradley Hanson[9] | Cassie Schumacher[10] | Timothy Harkins[10] | Michael C. Fontaine[11,12] | Barbara L. Taylor[1] | Kim M. Parsons[8,9]

[1]Southwest Fisheries Science Center, National Marine Fisheries Service, NOAA, La Jolla, CA, USA

[2]Department of Biology, Colorado State University, Fort Collins, CO, USA

[3]Southwest Fisheries Science Center, National Marine Fisheries Service, NOAA, Moss Landing, CA, USA

[4]Moss Landing Marine Laboratories, San Jose State University, Moss Landing, CA, USA

[5]Biology Department, Saint Mary's University, Halifax, NS, Canada

[6]Cetacean Research Program, Vancouver Aquarium, Vancouver, BC, Canada

[7]Cascadia Research Collective, Olympia, WA, USA

[8]Marine Mammal Laboratory, Alaska Fisheries Science Center, National Marine Fisheries Service, NOAA, Seattle, WA, USA

[9]Northwest Fisheries Science Center, National Marine Fisheries Service, NOAA, Seattle, WA, USA

[10]Swift Biosciences, Ann Arbor, MI, USA

[11]MIVEGEC Research Unit (Université de Montpellier, CNRS, IRD) & Centre for Research on the Ecology and Evolution of Diseases (CREES), Centre IRD de Montpellier, Montpellier, France

[12]Groningen Institute for Evolutionary Life Sciences (GELIFES), University of Groningen, Groningen, The Netherlands

Correspondence
Phillip A. Morin, Southwest Fisheries Science Center, National Marine Fisheries Service, NOAA, La Jolla, CA, USA.
Email: phillip.morin@noaa.gov

Abstract

Harbor porpoise in the North Pacific are found in coastal waters from southern California to Japan, but population structure is poorly known outside of a few local areas. We used multiplexed amplicon sequencing of 292 loci and genotyped clusters of single nucleotide polymoirphisms as microhaplotypes (N = 271 samples) in addition to mitochondrial (mtDNA) sequence data (N = 413 samples) to examine the genetic structure from samples collected along the Pacific coast and inland waterways from California to southern British Columbia. We confirmed an overall pattern of strong isolation-by-distance, suggesting that individual dispersal is restricted. We also found evidence of regions where genetic differences are larger than expected based on geographical distance alone, implying current or historical barriers to gene flow. In particular, the southernmost population in California is genetically distinct (F_{ST} = 0.02 [microhaplotypes]; 0.31 [mtDNA]), with both reduced genetic variability and high frequency of an otherwise rare mtDNA haplotype. At the northern end of our study range, we found significant genetic differentiation of samples from the Strait of

Georgia, previously identified as a potential biogeographical boundary or secondary contact zone between harbor porpoise populations. Association of microhaplotypes with remotely sensed environmental variables indicated potential local adaptation, especially at the southern end of the species' range. These results inform conservation and management for this nearshore species, illustrate the value of genomic methods for detecting patterns of genetic structure within a continuously distributed marine species, and highlight the power of microhaplotype genotyping for detecting genetic structure in harbor porpoises despite reliance on poor-quality samples.

KEYWORDS
cetacean, dbRDA, GT-seq, microhaplotype, mtDNA, seascape genetics, SNP, sPCA

1 | INTRODUCTION

Harbor porpoises (*Phocoena phocoena*) are small cetaceans found in temperate and sub-Arctic coastal waters, typically in less than 200 m depth (Fontaine, 2016; Read, 1999; but see Nielsen et al., 2018). Their range is considered continuous along the continental coasts, but there is high variability in density and apparent gaps in suitable habitat (Evenson et al., 2016; Forney et al., 2020). They are susceptible to entanglement in gillnets and disturbance from construction activities for wind farms (Carstensen et al., 2006; Reeves et al., 2013). In addition, the linear coastal distributions with gaps raises the question of how harbor porpoise will respond to climate change. Areas with sufficient data indicate fine-scaled population structure is common (Chivers et al., 2002; Crossman et al., 2014; Fontaine, 2016; Lah et al., 2016; Rosel et al., 1999; Tiedemann et al., 1996; Walton, 1997; Wang & Berggren, 1997). Along the US west coast, early genetic studies based on mitochondrial DNA (mtDNA) provided data to delineate management stocks (Chivers et al., 2002), but data gaps made boundary placement difficult. Because harbor porpoises avoid vessels, which makes dart-biopsy an unviable method to obtain sufficient samples, most samples are from either beach-stranded animals or entangled animals from areas with gillnet fisheries. As a result, sampling gaps existed in areas without gillnet fisheries and samples were often obtained days after death and in moderate to advanced stages of decomposition, and hence had degraded DNA.

Although gillnetting has declined along the US west coast, mitigation of other potential threats, such as development of wind farms, requires an understanding of population structure, often in areas poorly sampled in previous studies. The US Marine Mammal Protection Act (MMPA) manages at the scale of demographically independent populations (DIPs) where allele (or mtDNA haplotype) frequencies will differ but evolutionary differences are not expected. Because our area of interest includes the southernmost part of harbor porpoise distribution in the Pacific where ocean temperatures are already rising, an understanding of evolutionary barriers to dispersal is also of interest to evaluate potential impacts of such temperature shifts.

In our study area along the US West Coast and inland waters of Washington state, samples have slowly accrued to fill most gaps over a period of 30 years, but most samples are of degraded quality, and sample size remains low in some areas. An additional complication requiring consideration for marker choice is that there is evidence of intergeneric hybridization with Dall's porpoise (*Phocoenoides dalli*) in the eastern North Pacific (Baird et al., 1998; Crossman et al., 2014; Willis et al., 2004), which could affect population analyses by skewing allele frequencies, especially if hybridization has been regionally restricted and/or resulted in introgressive gene flow between species.

Genomic methods such as genome resequencing (e.g., Foote et al., 2019) and reduced representation sequencing (e.g., Andrews et al., 2016; Maisano Delser et al., 2016) can produce thousands to tens of thousands of genetic markers, providing unprecedented power to detect population differences (e.g., Candy et al., 2015; Emerson et al., 2010; Leslie & Morin, 2016) and identify candidate loci under selection (Ahrens et al., 2018). However, for studies restricted to the use of poor-quality archived samples, the optimal strategy involves targeting a reduced number of genetic markers that can efficiently and reproducibly be obtained from a large number of samples, such as "genotyping in thousands by sequencing" (GT-seq; Campbell et al., 2015). Typically targeting a few hundred loci, GT-seq relies on multiplexed short amplicon sequencing that requires only ~20 ng of DNA, and locus variability can be increased by targeting highly variable regions with multiple single nucleotide polymorphisms (SNPs) genotyped as microhaplotypes (Baetscher et al., 2017; McKinney et al., 2017).

To increase power to detect population structure from larger numbers of SNPs, while maximizing our ability to use an existing collection of poor-quality samples collected over three decades, we employed GT-seq multiplex sequencing combined with microhaplotype analysis, and traditional mtDNA control region sequencing to incorporate new and previously published data. We describe population structure at both evolutionary and demographic scales through a continuous range in the eastern North Pacific by applying a combination of Bayesian assignment to detect more substantial divergence of evolutionarily significant units (ESUs; Moritz, 1994), and traditional divergence metrics and spatially explicit methods to detect DIPs, the basis of marine mammal population stocks under the U. S. Marine Mammal Protection Act (Martien et al., 2019; Waples & Gaggiotti, 2006). ESUs are expected to have very low gene flow

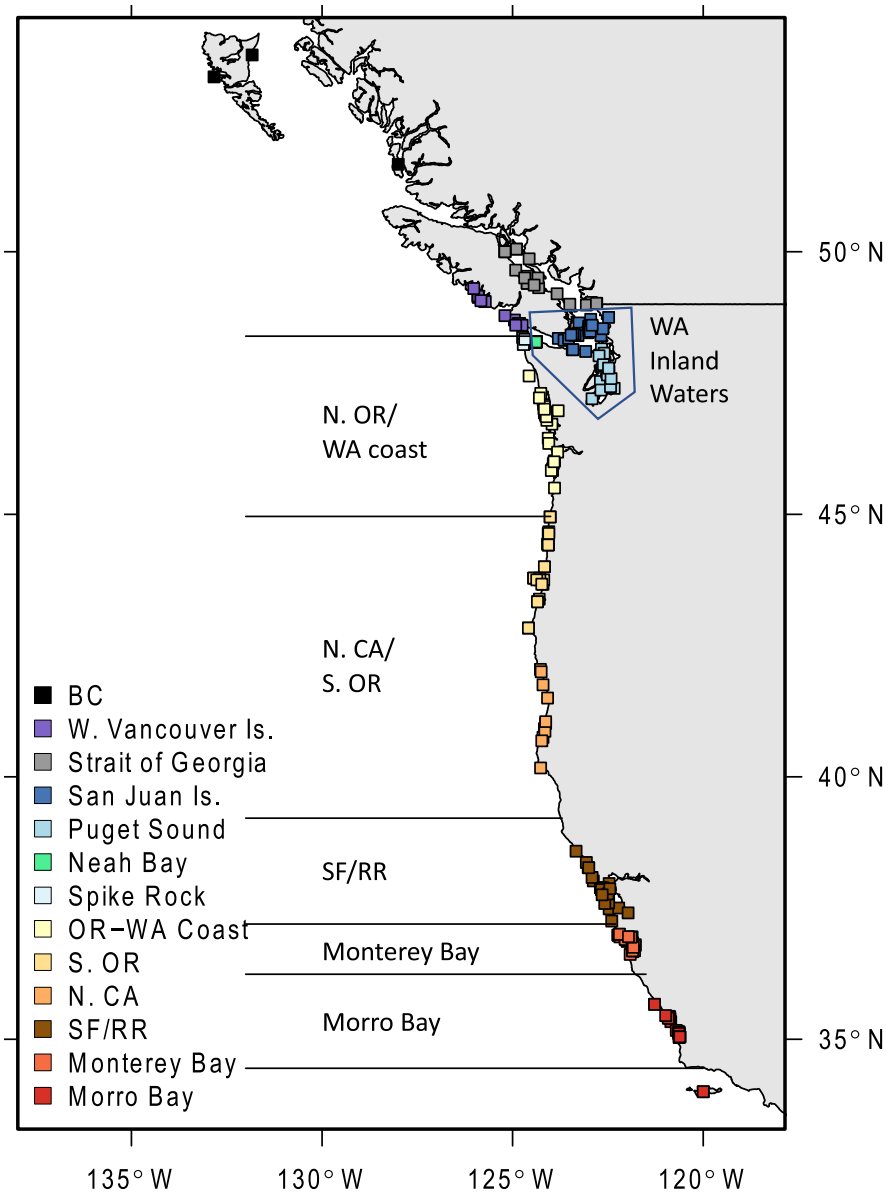

FIGURE 1 Sample distribution coloured by a priori geographical strata used for analyses. Current US stocks are shown with horizontal lines demarcating the boundaries. The Washington Inland Waters stock is shown with the shaded polygon. BC = British Columbia, OR-WA Coast = Northern Oregon/Washington coast, S. OR = Southern Oregon, N. CA = Northern California, SF/RR = San Francisco/Russian River

(on the order of a few successful dispersers per generation) and significant differences in both nuclear and mitochondrial markers. In contrast, demographic independence means that the population dynamics of the affected group is more a consequence of births and deaths within the group (internal dynamics) than immigration or emigration (external dynamics; Martien et al., 2019). DIPs could have gene flow on the order of 1%–2% per year. Besides the large difference in gene flow, DIPs could be based solely on mtDNA because birth and death rates depend on females, though use of larger numbers of nuclear markers may provide higher statistical power to detect population differences.

We also use seascape genotype–environment associations to investigate patterns of localized adaptation. We anticipate that these genomic-scale data will not only provide increased resolution to detect spatial structure among populations, but also allow us to link genotypes to environmental variables that vary both spatially and temporally. This emerging field of seascape genomics (Riginos et al., 2016) is still in its early stages relative to landscape genomics, due to both the difficulty of broad-scale sampling of many marine species and the limited availability of relevant environmental predictors. However, the increasing accessibility of remotely sensed oceanographic variables has improved our ability to identify spatial, temporal and ecological factors that promote population structure and local adaptation in complex and dynamic seascape environments (reviewed by Riginos et al., 2016; Selkoe et al., 2016).

2 | MATERIALS AND METHODS

2.1 | Samples

Harbor porpoise (N = 441), Dall's porpoise (N = 9) and putative harbor/Dall's porpoise hybrids (N = 13) skin samples were collected from beach-cast carcasses, carcasses recovered as fisheries

bycatch, or from animals live-captured for tagging. Tissue samples were preserved in salt-saturated 20% DMSO or 100% ethanol and subsequently stored at −20°C in the U. S. National Marine Fisheries Service (NMFS) Marine Mammal and Sea Turtle Research (MMASTR) Collection at the Southwest Fisheries Science Center (SWFSC). Sample information is given in Table S1 and locations are shown in Figure 1. DNA was extracted from tissue samples using a variety of common extraction methods, including silica-based filter membranes (Qiaxtractor DX reagents, Qiagen), standard phenol/chloroform extraction (modified from Sambrook et al., 1989), lithium chloride (Gemmell & Akiyama, 1996) and sodium chloride protein precipitation (Miller et al., 1988), and quantified using the Quant-iT PicoGreen dsDNA assay kit (Invitrogen) with a Victor X3 fluorospectrometer (Perkin Elmer).

2.2 | mtDNA control region sequencing

New 395-bp mtDNA control region sequences (N = 176) were generated according to previously published methods for harbor porpoise studies (Chivers et al., 2002, 2007; Crossman et al., 2014). Electropherograms of sequences from previous studies (N = 224 from Chivers et al., 2007; N = 91 from Crossman et al., 2014) were evaluated and compared to newly generated sequences, and regions with ambiguities or poor quality were either re-evaluated by a single person to ensure consistent interpretation (N = 224), or the DNA was sequenced again from new PCRs (polymerase chain reactions) using current Sanger sequencing chemistry (N = 91). Haplotype IDs were assigned using the LabelHaplotypes function in the R package STRATAG (version 2.4.905; Archer et al., 2017) and associated with previously published sequences and haplotypes (Table S2).

2.3 | SNP discovery

DNA from 12 North Pacific harbor porpoises was used for indexed genomic library preparation using the Accel-NGS 2S PCR-free genomic library preparation kit (Swift Biosciences). Either 100 ng or 500 ng of genomic DNA, determined by Qubit (Thermo Fisher Scientific) quantification, was used for the library preparation to obtain sufficient library product to pool 15 nmol/L of each sample prior to sequencing. For 12 North Atlantic samples, genomic DNA was used for indexed genomic library preparation at BGI following their proprietary protocol and pooled for next generation sequencing (NGS). The North Pacific pooled library was first sequenced in two 150-bp, paired-end Illumina MiSeq lanes, then both pooled libraries were sequenced in two Illumina HiSeq-4000 lanes each.

NGS read data were trimmed and filtered using TRIMMOMATIC (version 0.36; Bolger et al., 2014) to remove Illumina adapters, reads <50 bp and low-quality bases (Q < 20) at the beginning and end of reads. The sliding window approach was implemented to change internal bases with Q < 15 (sliding window of 4 bp) to N's. Because there was some bias in the nucleotide frequencies, the first 4 bp of all sequences w also removed.

Genome alignment and SNP discovery were conducted as previously described (Morin et al., 2018). The repeat-masked killer whale genome (accession GCA_000331955.2; Foote et al., 2015) was used as the reference for genome assembly. Briefly, the paired-end MiSeq reads from the 12 North Pacific samples were assembled de novo using CLC GENOMICS WORKBENCH (version 4.1; CLCbio) to obtain a complete reference mitochondrial genome as previously described (Hancock-Hanser et al., 2013). The NGS reads from 12 samples from each ocean basin subspecies were aligned separately to the harbor porpoise mitochondrial genome using BWA MEM (version 0.7.5a; Li & Durbin, 2009) and nonaligned reads were extracted using SAMTOOLS (version 1.2; Li et al., 2009). The extracted (nonmitochondrial) reads from each sample were separately aligned to the killer whale reference genome, combined into one alignment, and the consensus harbor porpoise genome sequence was generated. The nuclear DNA reads for each sample were aligned to the new harbor porpoise consensus genome sequence as above, followed by SNP discovery separately from each subspecies using GATK (version 2.5-2; DePristo et al., 2011; McKenna et al., 2010). Potential SNPs were filtered to remove SNPs with mapping quality <30, excessive coverage (>2× mean depth of coverage) and estimated minor allele frequency <0.05. To avoid linked loci, SNPs were selected from contigs that were at least 100 kb in length, and SNPs were at least 100 kb apart on the contigs. SNP loci from the two subspecies were compared to identify loci that were polymorphic in both ocean basins to avoid ascertainment bias in application to either subspecies. Finally, candidate SNPs were compared to GenBank using BLAST+ (Camacho et al., 2009) and filtered to remove loci that were potentially in repeat regions, gene families or a close match to nonmammalian species. The resulting set of filtered SNPs are subsequently referred to as "targeted SNPs," as these were the initial targets for our SNP genotyping effort. Additional SNPs detected in short GT-seq sequences were combined with the targeted SNPs to form microhaplotype genotypes.

2.4 | Multiplex primer design

Primers were designed from a batch of 500 loci selected randomly from the filtered SNPs using the program FASTPCR (Kalendar et al., 2011). Parameters for primer selection (from 300-bp sequences with the target SNP at position 151) were: primer length =15–32, tm =57–62, 3′ Tm =25–50, dimer stringency = 5, synchronized Tm for primer pair = 5, Forward primers between position 40 and 150, Reverse primers between position 152 and the 3′ end, and addition of 5′ GT-seq tails for indexing and library preparation: F-tail = CGACAGGTTCAGAGTTCTACAGTCCGACGATC, R-tail = GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT. After primer design, all loci with both forward and reverse primers were compared using the "Primers list analysis" function in FASTPCR to detect cross-locus primer dimer interactions with Tm > 20°C ("strong" primer dimers). Loci were filtered out of the primer list if either one

or both primers had predicted interactions with Tm > 40°C with more than two other primers, or if the predicted primer interactions were >46°C. Primers for 385 loci were synthesized at 100 μM concentration in 96-well plates by Integrated DNA Technologies.

2.5 | Multiplex PCR optimization

GT-seq primers were pooled and used for multiplex amplification of one sample initially to optimize the locus set prior to genotyping. Optimization consisted of multiple rounds of GT-seq library preparation as described by Campbell et al. (2015), and sequencing a small portion of the library (e.g., 1–10 million reads) to determine the relative abundance of reads per locus, and presence of primer artefacts as determined by the published analysis scripts (https://github.com/GT-seq/GT-seq-Pipeline). Loci were removed at each iteration to eliminate loci represented by disproportionately high read depths, evidence of primer artefacts or low ratios of the probe to primer target sequences. There are no published guidelines for cut-off values, so we removed loci that appeared to be outliers for any of these values, and based on expert advice from experienced users of the GT-seq method (see Acknowledgements). Primer sequences for the final set of 292 loci used for genotyping are given in Table S3.

2.6 | SNP genotyping

Amplicon libraries were prepared following the GT-seq protocol, including the optional Exo-SAP pretreatment of the samples (Campbell et al., 2015), and pooled libraries were sequenced on an Illumina NextSeq500 sequencer, 1 × 150 bp reads. Custom scripts for processing GT-seq data (Campbell et al., 2015) were used to demultiplex the sample files and conduct preliminary genotyping. Genotypes were quality checked for duplicate samples, percentage missing genotypes per locus and sample, and percentage homozygosity using the STRATAG package in R. Replicate samples were used to estimate genotyping error rates, then fastq files from replicates samples were combined to a single file per sample.

Fastq files were checked for standard quality metrics (e.g., per base quality scores, nucleotide composition, sequence duplication level, overrepresented sequences) with FASTQC version 0.11.3 (Babraham Bioinformatics), then trimmed using FASTP (Chen et al., 2018) to remove adapter sequences and poly-A and poly-G 3′ tails that were added during sequencing of amplicons shorter than 150 bp, and to exclude reads shorter than 30 bp after trimming. Reads were mapped to the reference locus sequences using the BWA MEM algorithm (version 0.7.15; Li & Durbin, 2009), and SNPs detected across all samples using FREEBAYES version 1.1.0-54-g49413aa (Garrison & Marth, 2012) after removing sample files smaller than 1 MB (containing <0.1 million filtered reads). FREEBAYES was run with minimal filtering (Supporting Information), followed by additional filtering with VCFTOOLS version 0.1.12b (Danecek et al., 2011) to extract the targeted SNP for each locus (for targeted SNP analysis; see below), and to remove sites with low coverage (minimum depth =10), indels and loci with less than 30% completed genotypes.

For targeted SNPs (position 151 in all reference sequences), custom scripts in R (Supporting Information) were used to extract the genotype data from the vcf file, generate allelic count plots to visualize the genotype distributions of reads for each allele, and re-call genotypes based on minimum depth and allelic ratios. The minimum depth of 10 reads total and default minor allele read proportion for heterozygotes of >0.3 were adjusted as needed until genotypes clearly fell into separate clusters in the allelic plots. Loci with poor resolution of plotted genotypes were removed from the data set.

Microhaplotypes (containing the targeted SNPs and/or newly discovered SNPs) were generated for all loci using the R package MICROHAPLOT (Baetscher et al., 2017). The MICROHAPLOT algorithm inserts N's for missing sequence data at SNPs within haplotypes, so we used a custom R-scripts (Supporting Information) to identify SNPs with >10% N's. The identified SNPs were removed from the original vcf file using VCFTOOLS, and MICROHAPLOT was used to generate new microhaplotypes with the remaining variable SNP positions. The unfiltered haplotypes were exported for subsequent filtering with custom scripts to view and call genotypes similar to the methods described above for targeted SNPs (Supporting Information). The few remaining microhaplotypes with N's in them were excluded from genotypes prior to analysis.

A final combined data set for microhaplotypes and targeted SNPs was created by combining the multi-SNP loci with the single-SNP (targeted) loci. Since the targeted SNP loci had been genotyped using two different methods, we selected the genotype data for each locus from the method that provided the higher quality or quantity of genotypes for the targeted SNP. Some microhaplotype loci that were monomorphic in harbor porpoises, or which had similar genotype quality for the targeted SNP but contained other SNPs present in Dall's porpoise samples, were retained (instead of the targeted locus data) to allow genetic identification of intergeneric hybrids. All loci that were found only in the microhaplotype or targeted SNP data set were then added to the filtered loci to generate a final data set.

2.7 | Quality analysis

Quality analysis and sample and locus filtering were conducted using custom R scripts. Samples missing >80% of the genotypes, and loci missing more than 45% of the genotypes were also removed. Genetic duplicates (>80% identity) were identified and one from each pair of samples identified as duplicates was removed. Analyses of deviations from Hardy–Weinberg equilibrium (HWE) expectations were conducted across all samples using the R package ADEGENET, and loci with a difference between observed and expected heterozygosity >0.2 were removed as extreme outliers (>10× the average difference of 0.02), probably due to non-Mendelian loci (e.g., null alleles, duplicated loci, or high error rates). Remaining loci were tested for significant linkage disequilibrium (LD) and deviations from HWE

TABLE 1 Environmental predictors for distance-based redundancy analysis

Analysis	Environmental predictor	Abbreviation	Predictor resolution	Source
Outer coastal populations	Mean daily sea surface temperate	sst.mn	Averaged over 0.1 degrees around sample location	Regional Ocean Modeling System ROMS/CCSRA 31-year reanalysis (1991–2010) or Near Real Time (2011+), from the U. C. Santa Cruz Ocean Modeling and Data Assimilation group; http://oceanmodeling.ucsc.edu/ (Moore et al., 2011)
	SD of daily sea surface temperate	sst.*SD*	Calculated within 0.3 degrees around sample location	
	Mean daily sea surface height	ssh.mn	Averaged over 0.1 degrees around sample location	
	SD of daily sea surface height	ssh.*SD*	Calculated within 0.3 degrees around sample location	
	Mean daily mixed layer depth (depth at which the temperature is −0.5°C from surface temperature)	ild.mn	Averaged over 0.1 degrees around sample location	
	SD of daily mixed layer depth (depth at which the temperature is −0.5°C from surface temperature)	ild.*SD*	Calculated within 0.3 degrees around sample location	
	Coastal Upwelling Transport Index	CUTI	Calculated every 1 degree latitude along West Coast from 31 to 47°N	http://mjacox.com/upwelling-indices/ (Jacox et al., 2018)
	Biologically Effective Upwelling Transport Index	BEUTI	Calculated every 1 degree latitude along West Coast from 31 to 47°N	
Inland waterways populations	Mean daily sea surface temperate (comparable to sst.mn above)	mSSTday.mn0.1	Averaged over 0.1 degrees around sample location	Multispectral Ultra-high Resolution Sea Surface Temperature (murSST); downloaded from NOAA/ERDDAP (Simons, R. A. 2015. Environmental Research Division Data Access Program ERDDAP. https://upwell.pfeg.noaa.gov/erddap. Data set names: jplMURSST41SST (daily) and jplMURSST41mday (monthly)
	SD of daily sea surface temperature (comparable to sst.*SD* above)	mSSTday.SD0.3	Calculated within 0.3 degrees around sample location	
	Mean monthly sea surface temperature	mSSTmon.mn0.1	Averaged over 0.1 degrees around sample location	
	SD of monthly sea surface temperature	mSSTmon.SD0.3	Calculated within 0.3 degrees around sample location	

Abbreviation: *SD*, standard deviation.

within three discrete, geographically defined strata represented by >20 samples (inland waterways [*N* =88], Neah Bay [*N* =21] and Northern California/Southern Oregon [*N* =35]), after correction for multiple tests using a sequential correction (Holm, 1979). We used the R package DEMERELATE (version 0.9.3; Kraemer & Gerlach, 2017) using relatedness estimators "Wang" and "Mxy" to test for inadvertent sampling of close relatives.

2.8 | Porpoise distribution data

To examine the genetic results in the context of harbor porpoise distribution and relative density along the US West Coast, independent aerial survey data collected during 1991–2017 off California (e.g., Forney et al., 1991), and during 1989–2003 off Oregon and Washington (e.g., Calambokidis et al., 1997) were processed to derive the number of porpoises seen per kilometre surveyed as an index of relative density. Although the transect design differed between these two data sets, the survey protocols, observer team and configuration, and aircraft type were the same. Survey data were truncated spatially to include only the primary porpoise habitat extending from shore out to 90–100 m water depth. Relative densities were calculated for each transect line, assigned to latitude of the transect midpoint, and then smoothed south-to-north using a Loess smoother.

2.9 | Habitat data

A variety of modelled and measured oceanographic variables (Table 1) were extracted to examine potential environmental correlates of genetic patterns. These predictors included (i) sea surface temperature, sea surface height, mixed layer depth, and the standard deviation of these three variables derived from the Regional Ocean Modeling System (ROMS) outputs (Moore et al., 2011); (ii) coastal upwelling indices (Jacox et al., 2018); and (iii) multispectral ultrahigh-resolution sea surface temperature and its standard deviation (Chin et al., 2017).

2.10 | Analytical methods

Several methods were used to estimate the number of populations and population assignment based on the genetic data. Population structure and individual assignment was examined using STRUCTURE (version 2.3), which implements a Bayesian clustering method to identify significant genetic clusters based on HWE allele frequency expectations (Hubisz et al., 2009; Pritchard et al., 2000). We ran 10 replicates for each value of *k* (where *k* is the number of putative populations), using correlated allele frequencies and an admixture model, with location prior using geographically defined units (Table S1). Each analysis consisted of 100,000 burn-in steps followed by 1,000,000 Markov chain Monte Carlo (MCMC) steps, and 10 replicates combined using 100 iterations in CLUMPP (version 1.1.2; Jakobsson & Rosenberg, 2007). We also used CLUMPAK (Kopelman et al., 2015) to assess convergence of the MCMCs and evaluate consistency of replicates across values of *K*. The Δ*K* method (Evanno et al., 2005) and PARSIMONY ESTIMATOR (Wang, 2019) were used to evaluate most likely number of inferred clusters.

STRUCTURE is known to have low power to detect populations (i.e., DIPs) when there is even a very low migration rate ($m \geq 0.005$ per generation), where demographic independence and biologically meaningful differentiation still exist (based on probability plots and the Δ*K* method; Cullingham et al., 2020; Kalinowski, 2011). To better differentiate DIPs, we examined spatially explicit principal components with geographical information using spatial principal components analysis (sPCA; Jombart et al., 2008). Spatial distances were based on type 1 (Delaunay triangulation) connection network. We tested for significant evidence of structure in the sPCA using the Eigenvalue test "spca_randtest" (Montano & Jombart, 2017) with 9999 permutations in the R package ADEGENET (version 2.1.1; Jombart, 2008). Geographical subsets of the data were analysed hierarchically to evaluate structure at decreasing spatial scales.

Given the nearly continuous distribution of harbor porpoises along a coastline, we tested for genetic isolation by distance using Mantel tests for correlation of both individual and population genetic distances with geographical distances using the ADEGENET R package. We used Euclidean distance for individuals (Euclidean distance among vectors of allele frequencies) and pairwise F_{ST} distances between population strata, and straight-line geographical distances between individual samples or average latitude/longitude position of samples clustered into a priori geographical populations. We also used Monmonier's algorithm (Monmonier, 1973) as implemented in ADEGENET. Putative boundaries between populations were inferred based on the default threshold value (third quartile of all distances between neighbours).

For both nuclear and mtDNA data, we tested for a priori population divergence using pairwise estimation of F_{ST} with MCMC resampling implemented in the STRATAG r package. Multiple stratification schemes were tested based on a priori information gained from the Bayesian population genetic and sPCA analyses described above, previously defined management stocks (Carretta et al., 2019) and gaps in harbor porpoise distribution (Forney et al., 1991). For mtDNA, we generated a median joining network (MJN) using the program POPART (Leigh & Bryant, 2015).

We used distance-based redundancy analysis (dbRDA) to investigate genotype–environment associations and identify microhaplotypes potentially under selection (Forester et al., 2018). dbRDA identifies how groups of SNPs or microhaplotypes covary in response to the multivariate environment. It is well suited to isolation-by-distance demographic scenarios, maintaining both high true positive and low false positive rates (Forester et al., 2018). Environmental variables (Table 1) were extracted from longitudinal oceanographic data matched to collection date, latitude and longitude of samples. We used Bray–Curtis dissimilarity (Bray & Curtis, 1957) to calculate the microhaplotype distance matrix. This approach quantifies the

dissimilarity among individuals based on their multilocus genotypes, and is equivalent to the proportion of shared alleles (Shirk et al., 2017). We performed dbRDA separately for porpoises from the outer coastal and inland waterways regions, because data for all environmental predictors were not available for all locations (Table 1).

For dbRDA, we first removed microhaplotype loci with heterozygosity <0.05, then removed individuals with missing data for the retained environmental predictors. We produced three data sets for each region (i.e., outer coastal and inland waterways), representing three thresholds of missing genotype data across individuals: 25%, 20% and 15% (Table 4). Because dbRDA requires complete data frames (i.e., no missing data), we imputed missing values for each data set using *snmf* in the LEA package version 3.1.2 (Frichot & Francois, 2015), testing values of *K* from 1 to 5, and alpha (regularization parameter) values of 10, 100 and 1000. All runs used 25 repetitions, 200 iterations and a 5% cross-entropy withholding. We then performed a dbRDA for each imputed data set, retaining three constrained axes for outlier analysis. We identified candidate microhaplotypes under selection using the robust (e.g., not sensitive to outliers) Mahalanobis distance (Capblancq et al., 2018), which identifies outlier microhaplotypes based on their constrained ordination loadings in multidimensional space. We accounted for confounding factors in the dbRDA, such as population structure and isolation-by-distance, using the genomic inflation factor (Francois et al., 2016), and applied a false discovery rate cutoff of 0.1 (Storey & Tibshirani, 2003) to identify outlier microhaplotypes showing relationships with environmental variation. Finally, we compared detections across missing data thresholds.

3 | RESULTS

We used a total of 431 mtDNA control region sequences (395 bp) from previously published (*N* = 363, including resequenced samples) and newly generated (*N* = 68) sequences, and SNP data for 296 porpoises ranging from the southern extent of the harbor porpoise range in Southern California, USA, to British Columbia, Canada (Table S1, Figure S1). Resequencing of samples with previously published haplotypes resulted in 10 haplotype changes from the Chivers et al. (2002, 2007) studies, and 53 from Crossman et al. (2014), mostly due to resolution of ambiguous positions in the previous electropherograms that resulted in synonymization of multiple haplotypes from each of those studies. We identified 52 harbor porpoise haplotypes, 22 of which were found in only a single individual, and six Dall's porpoise haplotypes (Tables S4 and S5). Our GT-seq locus panel consisted of 340 loci, of which 290 were genotyped in at least 55% of the samples and were polymorphic in harbor porpoises. Two additional loci were polymorphic only in Dall's porpoises and were used to identify hybrids between the two species. Of the 290 loci, 151 (52%) contained a single SNP, while the remaining 139 (48%) contained two or more SNPs, genotyped as microhaplotypes. None of the loci deviated significantly from HWE, and significant LD was only detected in one locus pair in one of the three tested geographical strata; no loci were removed based on these tests of HWE or LD.

DNA quantity and quality varied substantially among samples, resulting in a variable number of completed genotypes, and an inverse correlation of error rate with the number of completed genotypes. Arbitrarily changing the cut-off value for percentage completed genotypes can result in slight changes (see error rates below) to overall data quality, but at the cost of reduction in sample sizes in individual strata, reducing the statistical power to detect structure. To maximize the sample sizes across strata, we used 20% (≥58 of 292) genotyped loci as the minimum cut-off, resulting in 296 genotyped samples (after removal of unintentional duplicates [see below]: 280 harbor porpoise, 11 Dall's porpoise, five hybrids). Of the harbor porpoise samples, 72% were genotyped at >90% of the loci, and 83% were genotyped at >80% of the loci (Table S1).

Average per-allele error rates for SNPs were calculated from 32 samples genotyped in duplicate from separate GT-seq amplicon libraries, based on single SNPs in 270 loci that were genotyped in >50% of samples. Intentionally replicated samples for which there were sufficient data in both replicates (*N* = 32) had matching genotypes at an average of 96% of the loci (range 87%–100%). For samples genotyped at >80% of the loci, the estimated error rate (based on 23 replicated sample pairs) was 0.010 per allele. For samples with lower genotype completion rates (46 in the harbor porpoise data set), the mean error rate estimate increased to 0.045 per allele (based on *N* = 9 replicate pairs). Nine sample pairs (excluding Dall's porpoise samples, which were too homozygous in this data set) were identified as unintentional duplicates based on at least 85% identical genotypes (range 96%–100%), and one from each pair was removed from further analysis. All of the genetic duplicates were inadvertent duplicate samples from the same individual, usually due to samples being stored in different collections with different identification codes. We detected three potential first-order relatives (full siblings or parent–offspring pairs) based on the "wang" estimator, and six (including two of the three from the "wang" estimator) based on the "Mxy" estimator (Table S6; see Kraemer & Gerlach, 2017, for details). All samples were retained for some of the subsequent analyses, but one individual from each putative pair was removed to control for the effects of sampling closely related individuals in STRUCTURE, sPCA and F_{ST} analyses.

Putative hybrids between parapatric Dall's and harbor porpoise have been previously identified based on phenotype and genetic profiles (Crossman et al., 2014). We genotyped nine known Dall's porpoises and had phenotypic or previous genotypic indication (based on microsatellites; Crossman et al., 2014) of 11 putative hybrids in our final data set. Assignment analysis of all samples using STRUCTURE with the number of clusters set to *k* = 2 correctly assigned all nine Dall's porpoises to one group, and additionally assigned two of the putative hybrids to the same group with 100% probability. Five of the putative hybrids were assigned to the harbor porpoise group with ≥99.9% probability, and four samples were identified as F_1 hybrids with assignment probabilities to each species group between 45% and 55%. One additional sample previously identified

as a harbor porpoise was also identified as an F_1 hybrid (49%/51% assignment to the two species groups). The admixture plot from 10 combined structure analyses is presented in Figure S2a. All F_1 intergeneric hybrids were from samples collected in the San Juan Islands and Oregon–Washington coast geographical strata, between latitudes 47° and 49°N. All remaining putative harbor porpoise samples were assigned to the second group with >99.6% probability. As the targeted SNP loci were ascertained only from harbor porpoise samples, Dall's porpoise samples had unsurprisingly low diversity (Table 2), but the use of additional SNPs in microhaplotypes provided variable loci useful for species assignment and hybrid identification. Samples identified as hybrids were removed from subsequent harbor porpoise analyses.

STRUCTURE is useful for identifying population differentiation at the ESU level, where divergence is sufficient to allow high probability of assignment of samples to populations or clusters (Waples & Gaggiotti, 2006). We started with STRUCTURE in a hierarchical analysis of geographically defined strata to identify evolutionarily divergent populations within our harbor porpoise sample distribution. Structure analysis with all samples assigned to geographical strata a priori (Table S1) did not provide strong evidence of multiple divergent groups along the US West Coast, with the number of groups most likely ≤2 (based on Δ*K* and CLUMPAK similarity score >0.993 for 10/10 replicates for *K* = 1–2, and highest parsimony index [PI] for *K* = 1; Figure S2b). However, assignment probabilities among three groups (*K* = 3) differed substantially across geographical regions, especially in the three southernmost geographical strata (Figure 2). Subsets of samples representing only high-quality samples (<80% complete), geographical subsets of the sample (outer West Coast, inland waterways; Figures S2c–e) and samples assigned to population 1 (Figure 2a [purple]) at ≥0.75, which separates all samples in the Morro Bay population from all other samples, did not result in any additional evidence of population structure (highest PI for *K* = 1).

sPCA (Jombart et al., 2008) has been shown to be a useful tool for detecting patterns of genetic variability in harbor porpoise (Fontaine et al., 2017; Lah et al., 2016), and was used to explicitly combine geographical and nuclear genetic data and to investigate spatial patterns of genetic variation without assumptions of Hardy–Weinberg expectations or LD (Montano & Jombart, 2017). Plots of spatial principal components across the study range indicated strong evidence of structure (spca-randtest $p < .001$) and show clear separation between samples representing the two southern-most strata, Morro Bay and Monterey Bay, along the first axis (PC1, Figure 3a,b), with less separation but a north–south gradient along spatial PC axes 2 and 3 (Figure 3a; Figure S3A). Additional geographical population structure is revealed by analysing geographical subsets of the data. Hierarchical analysis of subregions reveals population structure within the inland waterways ($p = .004$; Figure 3g), and within the Washington inland waters management stock (excluding the Strait of Georgia strata samples; $p = .019$; Figure S3H). sPCA of samples from coastal waters (Figure 3b), excluding inland waterways in Washington and British Columbia, shows strong evidence of structure ($p < .001$), with the first two spatial PCs highlighting unique clusters corresponding to Morro Bay (PC1), and to a lesser extent the remaining strata along PC2. Hierarchical analysis of subsets of neighbouring strata indicated significant structure in analyses of all neighbouring strata pairs and within the northern California/southern Oregon management stock ($p = .017$; Figure 3d; Figure S3D). Plots of the first four individual spatial PCs for these stratification sets are shown in Figure S3. Replicate analyses based on the smaller number of high-quality samples (>80% complete genotypes) showed similar patterns, but resulted in nonsignificant *p*-values in some comparisons (Inland waters, southern Oregon/Oregon–Washington, and northern California/southern Oregon; Figure S3).

A Mantel test supported ($p < .001$) isolation by distance (IBD) for nuclear SNP data along the Pacific coast (excluding inland waterways) for individual distances. A scatterplot of genetic and geographical distances (Figure S4) shows discontinuities suggestive of differentiated populations rather than continuous clines of genetic differentiation. The Monmonier's algorithm was used to infer locations of genetic boundaries or discontinuities (Figure 4), and identified the major boundary between coastal Washington and western Vancouver Island at the default threshold level (Figure 4a), with lesser boundaries between Morro Bay and Monterey Bay, and Monterey and SF/RR on the outer coast. Within the inland water ways, the major discontinuities were in the southern Strait of Georgia (Figure 4b). It is not straightforward to interpret the Monmonier plots because the samples are not distributed broadly in two dimensions, but the threshold arrows indicate regions where genetic distances appear to be greater than expected based on geographical distance (see Blair et al., 2012 for a discussion of methods of detecting barriers to gene flow).

Genetically similar groups (Figure 3) and IBD discontinuities (Figure 4) coincided closely with several previously defined geographical strata (Figure 5; Chivers et al., 2002, 2007). Genetic differentiation between adjacent a priori strata was tested using pairwise divergence (F_{ST}), for both nuclear data and mtDNA (Table 3). Results are presented based on inclusion of all samples, but pairwise divergence among adjacent strata after removal of one from each pair of potential first-order relatives, and after removal of samples with <50% or <80% completed genotypes, did not qualitatively alter results (Table S7). As expected when larger numbers of loci are used to calculate F_{ST} (Willing et al., 2012), the F_{ST} point estimates did not change substantially (<0.004), but exclusion of samples reduced the sample sizes in some geographically defined strata, resulting in loss of statistical power to detect structure among some nonadjacent strata (Table S7D–F). Along

TABLE 2 SNP and microhaplotype genotype summary information, based on 292 loci

Species	Samples	H_E	H_O	Monomorphic
P. phocoena	281	0.366	0.356	2
P. dalli	11	0.144	0.145	177
Hybrid	5	0.479	0.324	64

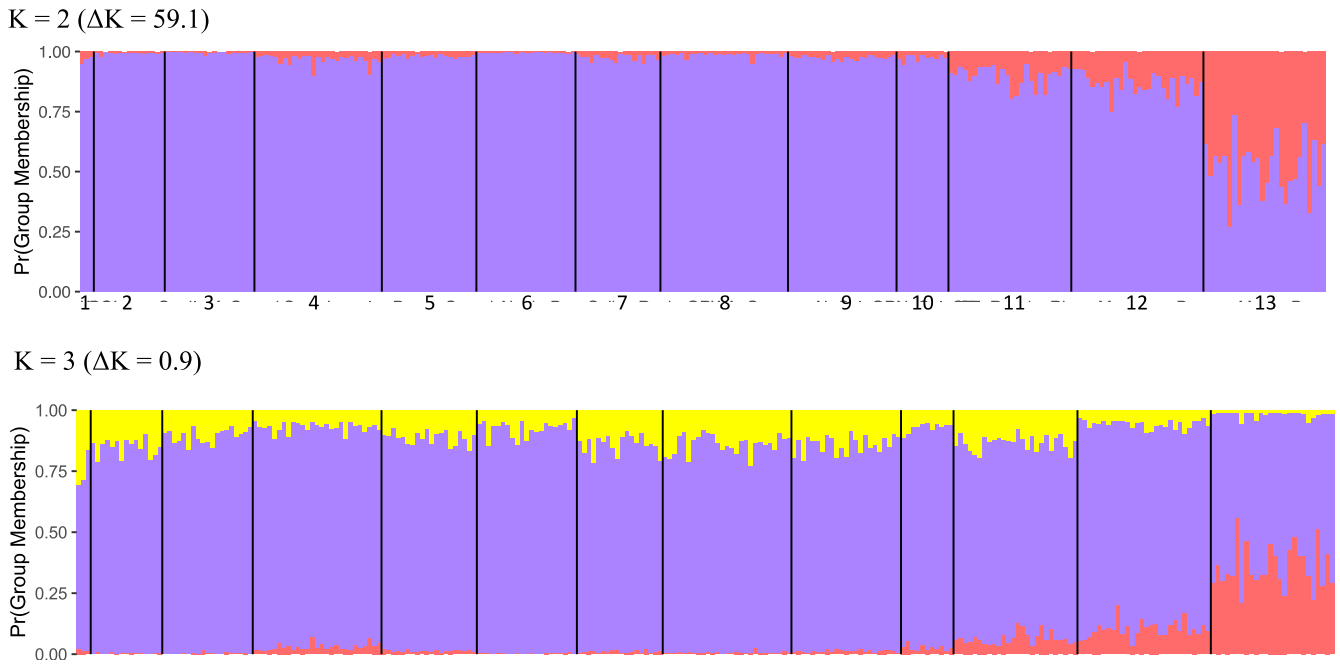

FIGURE 2 STRUCTURE plots for $K = 2$ and $K = 3$ based 290 loci and 264 samples (unrelated) with 10 replicates combined with CLUMPP. Samples assigned to a priori geographical strata are sorted by decreasing latitude in the bar plot from left to right. 1 = BC; 2 = W. Vancouver Is.; 3 = Strait of Georgia; 4 = San Juan Is.; 5 = Puget Sound; 6 = Neah Bay; 7 = Spike Rock; 8 = OR-WA coast; 9 = S. OR; 10 = N. CA; 11 = SF/RR; 12 = Monterey Bay; 13 = Morro Bay

the outer coast the only genetic distances between neighbouring strata pairs that were significant ($p < .05$) for both mtDNA and nuclear DNA (nDNA) were between Morro Bay at the southern end of the distribution and Monterey Bay ($F_{ST} = 0.020$ [nDNA]; 0.310 [mtDNA]). Nuclear locus frequencies were also significantly different between three of the more southern strata pairs (Monterey/SF-RR [$F_{ST} = 0.005$], N. CA/S. OR [$F_{ST} = 0.008$], S. OR/ORWA coast [$F_{ST} = 0.006$]), while mtDNA haplotype frequencies were significantly different between only the two northern-most coastal strata pairs (ORWA coast/Spike Rock [$F_{ST} = 0.049$], and Spike Rock/W. Vancouver Is. [$F_{ST} = 0.128$]). Frequency plots of the common mtDNA haplotypes (excluding haplotypes that occurred in fewer than five samples) show clear differences among a priori geographical strata (Figure 5; Figure S5), with near fixation of haplotypes in Morro Bay to the south, and private haplotypes in BC to the north. Relative density data showed low-density regions along the Big Sur coastline (35.5–36.5°N), just south of 38°N, and just north of 39°N at the existing management stock boundary. These areas of low relative density correspond to existing boundaries between the four southern-most management stocks. The reason for low density for one region, the Big Sur coastline, is probably due to the very narrow shelf, but reasons for density variation in other regions remain unclear.

Inland waters strata were not differentiated by nuclear marker frequencies except for the Strait of Georgia, which was significantly differentiated from all inland waters and outer US coast strata, but not W. Vancouver Is. or BC (Table 3). The inland waters strata were differentiated from the nearest coastal strata in mtDNA (F_{ST} between 0.023 and 0.247, with most point estimates being significant at $p < .05$). Within the inland waters, divergence in mtDNA was significant for Neah Bay, in the outer Strait of Juan de Fuca, vs. San Juan Islands ($F_{ST} = 0.066$), and from both neighbouring coastal strata ($F_{ST} = 0.141$ [Spike Rock]; 0.247 [W. Vancouver Island]). Puget Sound, which has been recolonized in the last 18 years (Evenson et al., 2016), was not significantly differentiated from either the neighbouring San Juan Islands, or Neah Bay in the Strait of Juan de Fuca. The genetic discontinuity identified immediately north of the US–Canadian border at approximately 49.1°N latitude (Figure 4) warranted an adjustment of the boundary between the Strait of Georgia and San Juan Islands strata from 48.8° to 49.1°N, resulting in reassignment of 10 and five samples (mtDNA, nDNA analyses, respectively) from the Strait of Georgia to the San Juan Islands strata (Figure 4b). Genetic divergence remained significant following the reassignment of these samples and F_{ST} increased for both nDNA (0.005 to 0.006) and mtDNA (0.095 to 0.153).

For the genotype–environment association analyses (dbRDA), we analysed three data sets for each region (i.e., outer coastal and inland waterways), representing three thresholds of missing genotype data across individuals: 25%, 20% and 15% (Table 4). We first imputed the data sets with *snmf* using the optimized settings (identified by minimizing the average cross-entropy, with identical optimized parameters for all data sets): alpha =10 and $K = 1$ (Table S8). We modified the default genomic inflation factors to produce *p*-value distributions that better met the uniform distribution assumption (as per Francois et al., 2016, *p*-value histograms provided in Figures S6–S11), providing a balance between true positive detections and false positives driven

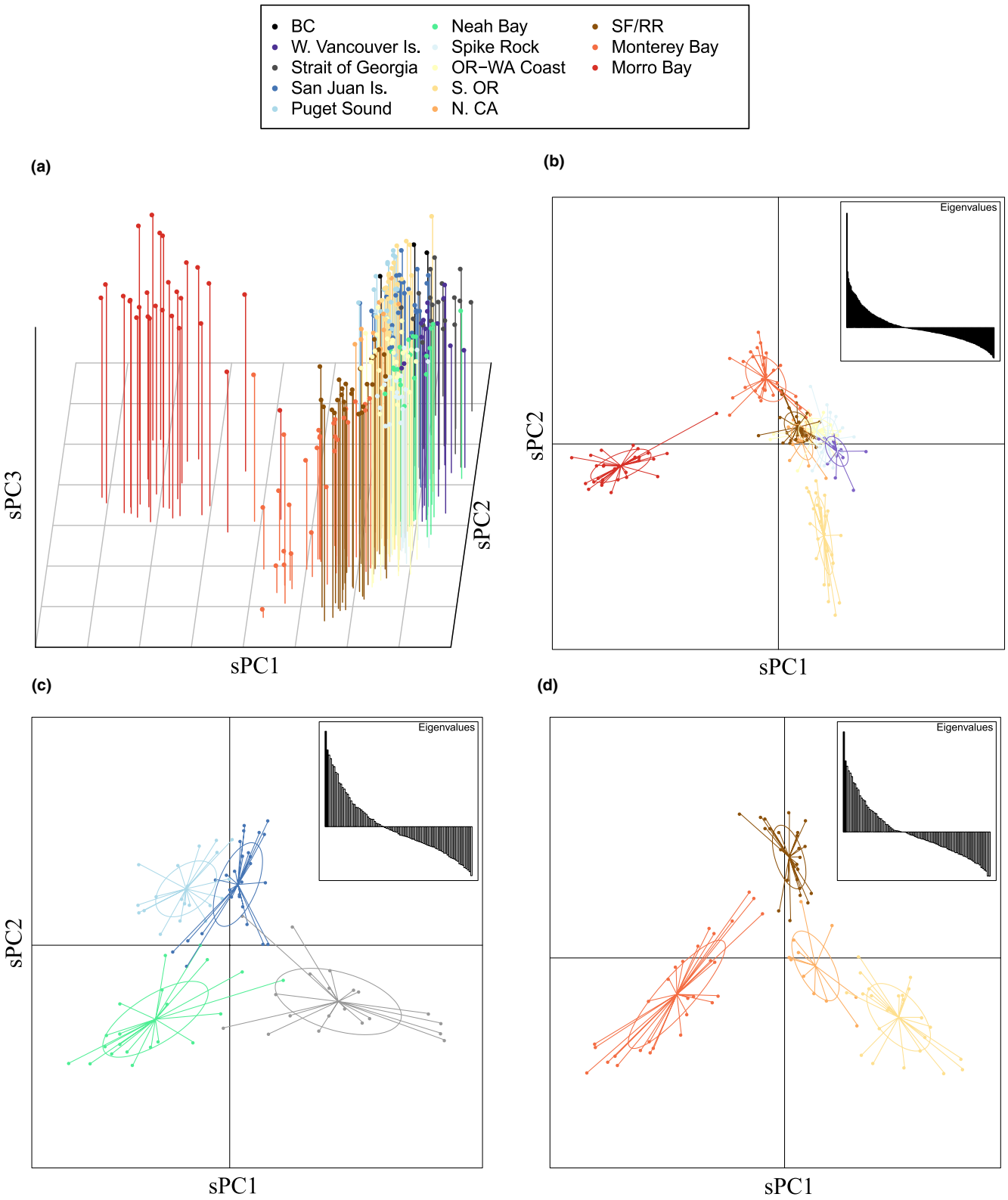

FIGURE 3 sPCA three-dimensional plot of the first three lagged scores of spatial PCs (sPC) for samples from (a) all strata and samples (290 loci), and two-dimensional plots of the first two sPCs for geographical subsets of the data from (b) outer coast strata (all strata except BC and those in panel c; 290 loci), (c) inland waters of Washington and British Columbia (Strait of Georgia, San Juan Is., Puget Sound, Neah Bay; 289 loci), (d) central California/southern Oregon (S. OR, N. CA, SF/RR, Monterey Bay 288 loci). Insert bar-charts show the eigenvalues. The ovals represent ellipses of dispersion. Colours correspond to a priori strata, with sample numbers per stratum as in Table 3. Different numbers of loci are due to removal of monomorphic loci in sample subsets

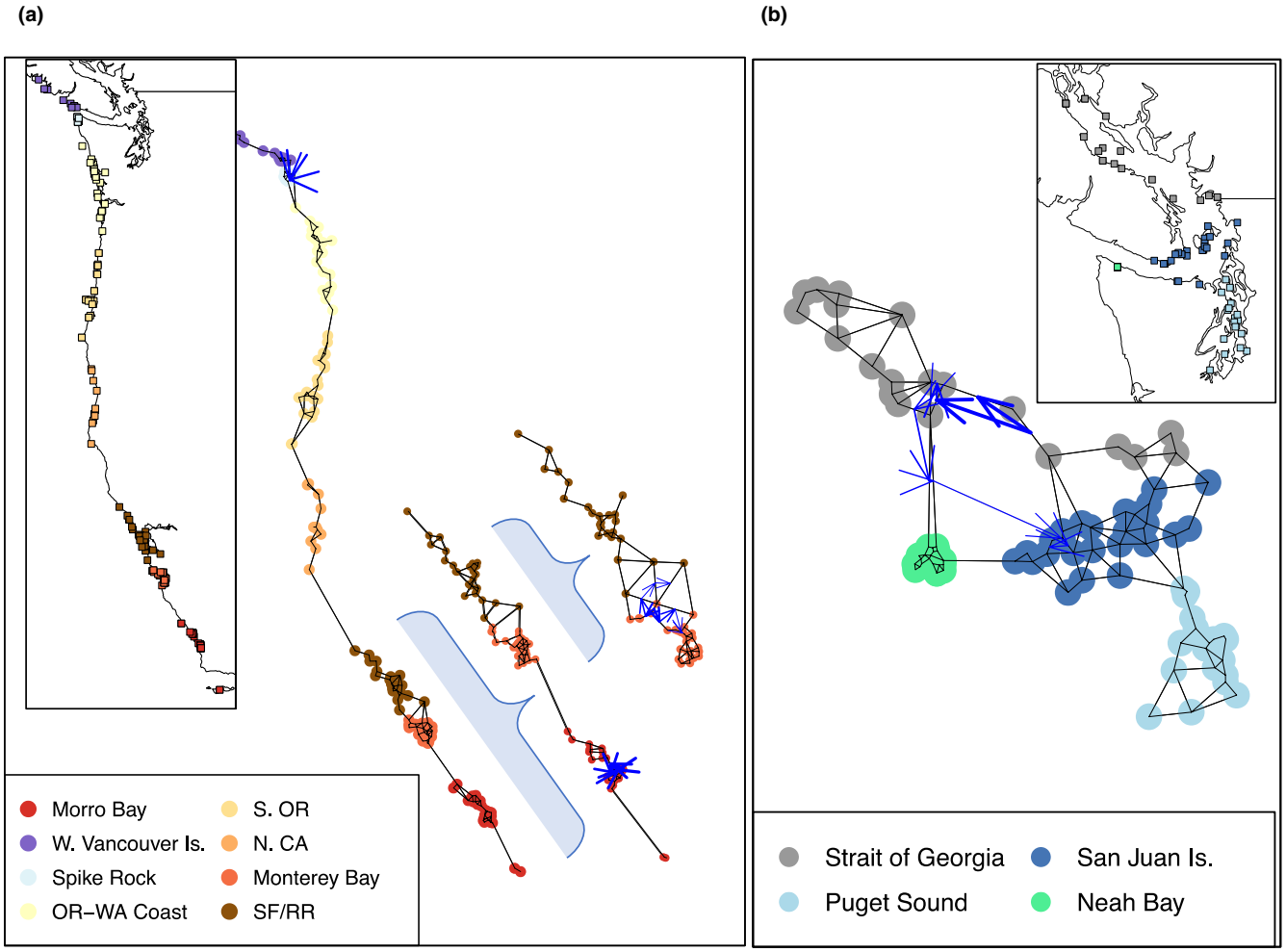

FIGURE 4 Gabriel connection network from a Monmonier analysis of samples (adjusted to prevent sample location overlap) with local nuclear genetic differences above the threshold value indicated by thicker blue lines and arrows between geographical distance edges for (a) the outer coastal samples, with subsets of samples shown for localized thresholds, and (b) inland waterways of Washington and British Columbia. Inset maps are for orientation of sample networks. Thickness of the lines and arrows is proportional to the level of genetic difference above the threshold level. Colours correspond to a priori strata as in Figure 1

by population structure and other potential confounders. We did not include other corrections for confounding factors in the GEA because population structure is low across the study area (e.g., Table 3), and dbRDA is robust to isolation-by-distance demographic scenarios (Forester et al., 2018) such as this one. Environmental predictor variables within each region were not highly correlated (i.e., all pairwise correlations were less than |0.8| and all variance inflation factors, which measure multicollinearity, were less than five, where values >10 can be problematic), and none of the retained predictors were highly correlated with either latitude or longitude (Tables S9 and S10). For the outer strata and inland waterways, we identified as candidates those microhaplotypes that were detected in at least two of the three thresholded data sets. This produced 22 candidates for the outer coastal region and six candidates for the inland waterways region. Candidate microhaplotypes for the outer coastal region showed the strongest relationships with the mean and standard deviation of daily sea surface temperature (sst.mn and sst.*SD*), and the mean of daily sea surface height (ssh.mn). Triplots of the dbRDA results illustrated relationships among an individual's multilocus genotypes, the candidate microhaplotypes and environmental predictors. For example, the southernmost Morro Bay individuals, located in the ordination space as a function of their genotypes at the candidate loci, showed strong relationships with increasing ssh.mn and sst.mn and decreasing ssh.*SD*, indicating potential adaptation to local environmental conditions (Figure 6a). The smaller inland waterways data set showed some differentiation between the Neah Bay and San Juan Islands individuals based on sea surface temperature, with the Puget Sound individuals showing no relationships, possibly due to their most recent recolonization (Figure 6b). Because loci under selection could affect population structure, we confirmed that identification of divergence between adjacent a priori strata were consistent with and without the loci identified by dbRDA analysis (Table S7). Full dbRDA triplots and microhaplotype biplots are provided in Figures S12–S15 for the outer coastal data and Figures S16–S19 for the inner waterways .

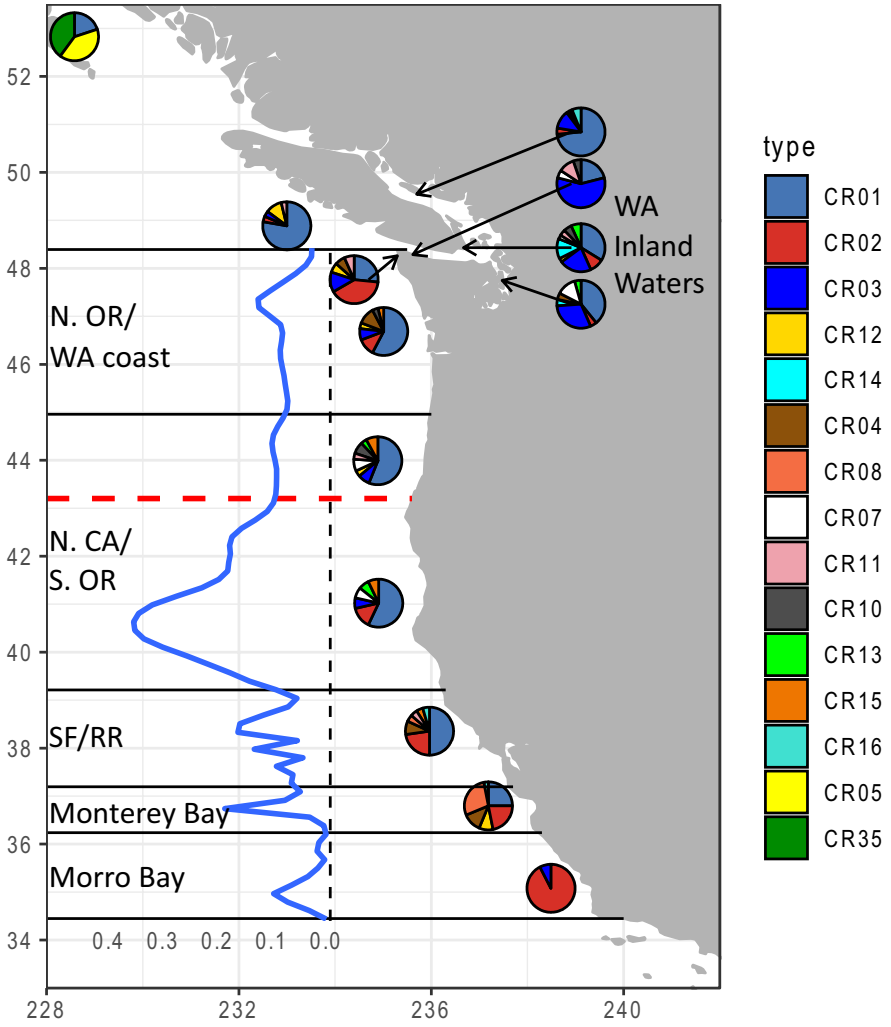

FIGURE 5 Pie charts representing frequencies of haplotypes represented by ≥5 samples in the stratum, plus two haplotypes (CR05, CR35) that were found only in the BC strata, in two samples each (out of five in total). Relative porpoise density estimates (porpoise per kilometre surveyed) is plotted by latitude, averaged across multiple survey periods (blue line). The current US stocks are named and boundaries are represented by horizontal black lines (as in Figure 1). The proposed boundary location for splitting the Northern California/Southern Oregon stock into two stocks is shown with a dashed red line

4 | DISCUSSION

The North Pacific harbor porpoise is one of several geographically and genetically described subspecies of harbor porpoise, some of which consist of multiple ecotypes (Ben Chehida et al., 2020; Fontaine, 2016; Fontaine et al., 2012, 2014). Studies of the North Atlantic subspecies have suggested both historical biogeographical processes (Fontaine, Baird, et al., 2007; Fontaine et al., 2014) and ecological processes (Fontaine et al., 2017; Fontaine, Tolley, et al., 2007; Lah et al., 2016) resulting in divergent populations and limited dispersal and introgression between adjacent types. In the North Pacific, evidence of structure based on both biogeographical (Taguchi et al., 2010) and ecological/spatial divergence (Chivers et al., 2002, 2007; Crossman et al., 2014) has been more limited, but suggests that similar processes may have acted across the ranges of subspecies in both ocean basins.

Previous genetic studies of North Pacific harbor porpoise have been limited by sample availability, sample quality, and low power to detect population genetic differences within this nearly continuously distributed coastal species. We have developed a set of SNP and microhaplotype loci that provide sufficient genetic power to detect low levels of population structure, while allowing us to make use of poor-quality tissue samples available from beach-cast and fishery bycatch carcasses. We used hierarchical partitioning to explore evidence of population structure across the range, and to infer potential barriers to geneflow. Correlation of remotely sensed environmental variables with genetic patterns identified factors potentially influencing local adaptation. Our results indicate that North Pacific harbor porpoises exhibit genetic discontinuity and limited dispersal that may be associated with habitat variability and local adaptation.

Previous evidence indicated intrageneric hybridization between the more pelagic Dall's porpoise and the coastal harbor porpoise, based on morphology and genetic analysis of eight microsatellite loci (Crossman et al., 2014). While the previous genetic analyses suggested bidirectional hybridization and introgression, the more powerful data set used here indicates that many of the putative hybrids (based on genetics) could be assigned with >98% probability to one or the other species (N = 5 *P. phocoena*; N = 2 *P. dalli*), or as F_1 hybrids (N = 5) between male harbor porpoise and female Dall's porpoise. Sightings of putative hybrids in the wild have also all been associated with Dall's porpoises (Baird et al., 1998). Harbor porpoise males have disproportionately large testes and have been observed to exhibit forceful and fast ambush mating (Keener et al., 2018), suggesting a greater reliance on sperm competition rather than mate choice in

TABLE 3 Pairwise F_{ST} for nuclear loci (lower left, based on 290 loci and all samples with >20% completed genotypes), and mtDNA haplotypes (upper right)

	BC (5)	W. Vanc. Is. (29)	Strait of Georgia (55)	San Juan Is. (92)	Puget Sound (24)	Neah Bay (20)	Spike Rock (18)	ORWA Coast (30)	SoOR (29)	NoCal (17)	SF_RR (30)	Monterey Bay (37)	Morro Bay (28)
BC (3)	—	**0.223**	**0.249**	**0.097**	**0.147**	**0.251**	**0.127**	0.151	**0.140**	0.126	**0.108**	**0.125**	**0.604**
W. Vancouver Is. (15)	−0.003	—	−0.005	**0.088**	**0.076**	**0.247**	**0.128**	0.002	0.005	0.004	**0.046**	**0.130**	**0.542**
Strait of Georgia (20)	−0.006	0.005	—	**0.095**	**0.066**	**0.232**	**0.152**	0.011	0.017	0.012	**0.064**	**0.159**	**0.526**
San Juan Is. (27)	0.005	0.000	**0.005**	—	0.008	**0.066**	0.023	**0.043**	**0.037**	0.021	**0.031**	**0.059**	**0.307**
Puget Sound (20)	0.003	0.003	**0.009**	0.001	—	0.039	0.057	0.029	0.020	0.009	**0.046**	**0.093**	**0.454**
Neah Bay (21)	**0.015**	0.000	**0.010**	0.000	0.002	—	**0.141**	**0.176**	**0.157**	**0.160**	**0.174**	**0.192**	**0.531**
Spike Rock (19)	0.002	−0.004	**0.007**	0.002	0.002	−0.001	—	**0.049**	**0.083**	0.033	0.005	0.024	**0.233**
ORWA Coast (28)	0.015	−0.001	**0.014**	**0.004**	**0.005**	**0.004**	−0.001	—	−0.006	−0.023	−0.005	**0.067**	**0.440**
SoOR (24)	0.009	0.002	**0.011**	0.002	−0.001	**0.004**	0.001	**0.006**	—	−0.027	0.019	**0.090**	**0.480**
NoCal (11)	0.014	0.003	**0.015**	0.001	0.005	0.003	0.004	0.002	**0.008**	—	−0.012	**0.059**	**0.448**
SF/RR(26)	0.012	0.004	**0.015**	**0.005**	0.004	**0.005**	−0.002	**0.004**	0.002	0.004	—	0.023	**0.347**
Monterey Bay (29)	**0.021**	**0.006**	**0.016**	**0.007**	**0.011**	**0.008**	0.001	**0.006**	**0.012**	**0.007**	**0.005**	—	**0.310**
Morro Bay (28)	**0.027**	**0.028**	**0.036**	**0.022**	**0.024**	**0.028**	**0.023**	**0.026**	**0.025**	**0.024**	**0.020**	**0.020**	—

Statistically significant values ($p < .05$) are in bold. Sample sizes each population are shown in parentheses for nuclear loci on the left, and mtDNA on the top. The p-values for adjacent strata pairwise comparisons are provided in Table S7.

Population names are as shown in Figure 1.

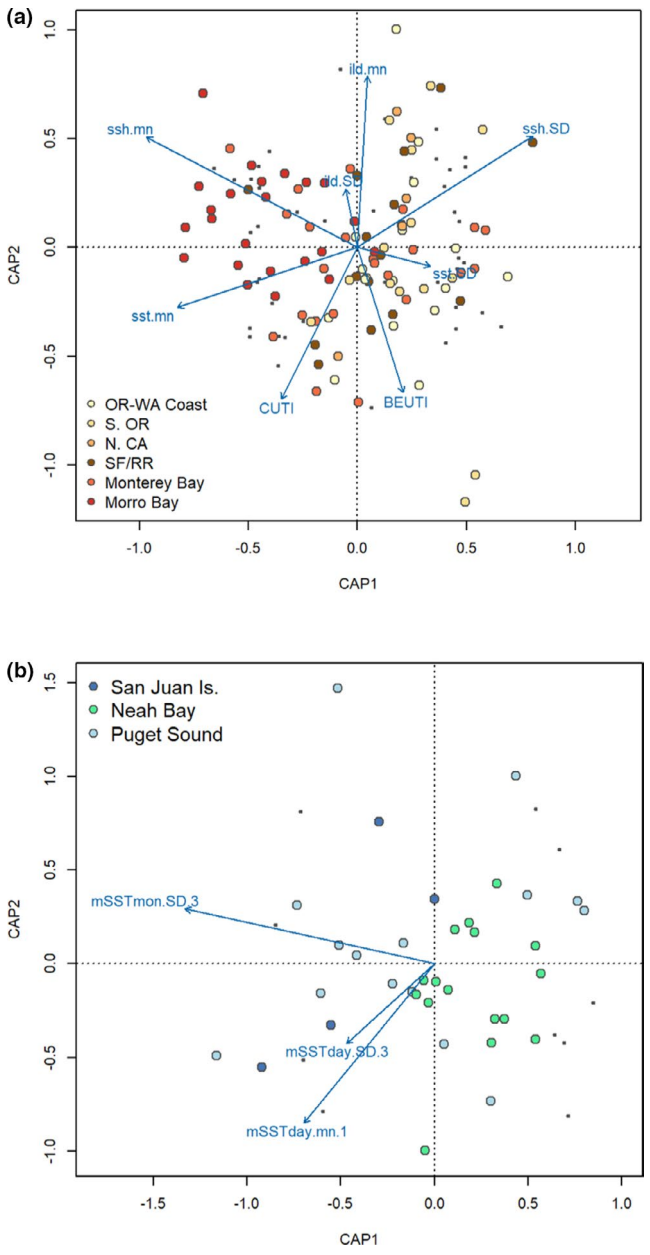

FIGURE 6 Distance-based redundancy analysis triplot of candidate microhaplotypes under inferred selection for (a) the outer coastal strata and (b) inland waterways strata, showing individuals as coloured circles, microhaplotype alleles as grey points, and environmental predictors (see Table 1 for abbreviations) as blue arrows. The location of individuals in the ordination space reflects their relationship with the environmental variables based on their multilocus genotypes at the candidate microhaplotype markers

male harbor porpoise. This mating strategy might result in occasional intergeneric hybrids with Dall's porpoise in areas where their ranges overlap (Baird et al., 1998). Although the lack of individuals exhibiting intermediate probabilities of assignment to both species (between 50% and 100%) suggests hybrids are usually infertile, there have been sightings of putative hybrid females with neonatal calves (Willis et al., 2004).

Within harbor porpoises, our results are consistent with isolation by distance along a mostly linear range from southern California to western Vancouver Island, and through the inland waterways, but also suggest some regions of higher genetic divergence than expected based on geographical distance alone. The two largest population splits are between Morro and Monterey Bays in the south, and between the Strait of Georgia and San Juan Islands in the north, where multiple methods suggest divergence in both mtDNA and nuclear markers. This break between the Strait of Georgia and San Juan Islands, as well as between the San Juan Islands and the outer Strait of Juan de Fuca, is consistent with the limited movements of tagged harbor porpoises in this area (Hanson, 2007). Another population break is suggested between the Monterey Bay and San Francisco sample sets, based on the Monmonier analysis of the connection network (Figure 4), where F_{ST} divergence is significant for nuclear loci but not for mtDNA.

Although the Morro Bay population is designated as a separate management stock based on a hiatus in distribution at the Big Sur coast (Carretta et al., 2019), our results provide the first evidence for genetic differences. The most common mtDNA haplotype (CR01) was found in every a priori geographical group except Morro Bay, and the most common haplotype in Morro Bay (CR02) was found in 86% of the Morro Bay samples, but less than 33% of any other population (range 0%–33%). Two (CR30, CR42) of the three other haplotypes found in Morro Bay were unique to that population, and differed from CR02 by only one nucleotide change. The third haplotype (CR03) was distributed across most of the range, but represents a common haplotype that is more similar to CR02 (3 bp different) than to the cluster including CR01 (7 bp different) and the majority of the other haplotypes distributed throughout the rest of the range (see haplotype network, Figure S5). This suggests long-term isolation of the southernmost population, as well as a persistently small population size, or severe or recurrent bottlenecks. Annual estimates of tens to hundreds of harbor porpoises were killed in gillnet entanglements between approximately 1960 and 2001, resulting in a reduction of the population by 30%–97% (Barlow & Hanan, 1995). The population has since begun recovering, from a low of 571 in 1990 to 4,191 in 2012 (Forney et al., 2020). Our results suggest that recovery has been due to internal recruitment rather than migration from the larger Monterey Bay population to the north.

The northernmost strata were the most differentiated, with high mtDNA divergence metrics and the highest frequencies of the most common control region haplotype (CR01; 66% W. Vancouver Is., 63% Strait of Georgia). The frequency of this haplotype ranged from zero to 50% in other strata. The BC stratum (excluding Strait of Georgia samples) spanned a large geographical range with few samples (N = 5), and included two private mtDNA haplotypes, both found in two samples. These suggest different haplotype composition in this region, but we cannot rule out sampling of related individuals or isolated regional populations, as the sample size is small and shared haplotypes were found in samples that were collected close in space and time. Combined, these data suggest that there is a haplotype gradient between the northern and southern portions of the

TABLE 4 dbRDA sample information for three data sets filtered for different levels of missing genotypes

	Outer coastal populations			Inland waterways populations			
Amount of missing data	25%	20%	15%	25%	20%	15%	
Microhaplotypes retained	280	280	280	274	274	274	
Individuals retained	98	96	92	36	36	31	
OR-WA Coast	*17*	*17*	*14*	*4*	*4*	*4*	*San Juan Is.*
S. OR	*14*	*14*	*14*	*16*	*16*	*14*	*Neah Bay*
N. CA	*5*	*5*	*5*	*16*	*16*	*13*	*Puget Sound*
SF/RR	*16*	*14*	*13*	–	–	–	
Monterey Bay	*23*	*23*	*23*	–	–	–	
Morro Bay	*23*	*23*	*23*	–	–	–	
Modified GIF	1.19	1.20	1.20	1.14	1.14	1.18	
Candidates at FDR =0.1	25	22	4	6	6	9	

Rows in italics report the number of retained individuals in each set of populations.
Abbreviations: FDR, false discovery rate; GIF, genomic inflation factor.

range, possibly representing interglacial range expansion (Taguchi et al., 2010) or the presence of a historical phylogeographical separation that has more recently reconnected, as has been suggested for populations in the eastern Atlantic Ocean (Fontaine, Baird, et al., 2007; Fontaine et al., 2014).

Harbor porpoise habitat in the eastern North Pacific includes the California Current System (CCS), recognized as one of the most productive marine ecosystems on the planet, but spans topographically, oceanographically and temporally complex regions from temperate to Arctic waters. Aside from a preference for coastal waters less than 200 m deep, little is known about the factors influencing suitable habitat for Pacific harbor porpoises, which vary in population density and abundance along the west coast and inland waters of North America. The remotely sensed and modelled oceanographic variables available throughout most of this range (upwelling index, mixed-layer depth, sea surface temperature and sea surface height) are all proxies for habitat variability with complex relationships to wind, currents, topology, isocline depth, productivity, freshwater input, and seasonal and interannual effects (Castelao & Luo, 2018; Hickey et al., 2016; Venegas et al., 2008). Our genotype–environment association analyses provide initial evidence for local adaptation to environmental variability across the study range. Given the limited scale of genomic sampling in this study, these findings should be seen as preliminary, providing a basis for future investigation of local adaptation across the complex seascape environment inhabited by harbor porpoise (see, e.g. Fontaine et al., 2017; Fontaine, Tolley, et al., 2007; Nielsen et al., 2018).

An overall correlation with latitude, which might be expected across a large latitudinal range, appears to strengthen in the regions south of Oregon. In central Oregon, there is an oceanographic shift around Cape Blanco, where the coastal upwelling jet separates from the coast (~15–30 km offshore), becoming an oceanic jet (>100 km offshore) (Castelao & Luo, 2018). Populations in the northern half of the coastal range are associated with several environmental variables (e.g., mean and *SD* of the mixed layer depth, and high variation in both sst and ssh), while those in the central and southern portions are associated with others (e.g., high mean and low variation in ssh; Figure 6). There is also evidence of population structure across that region, with significant nuclear genetic differentiation among populations from southern Washington to northern California (Table 3). There was a particularly strong correlation between increased mean ssh and decreased ssh variability in the southernmost Morro Bay population, which is also the most genetically distinct, suggesting possible local adaptation linked to oceanographic processes (e.g., thermocline and/or upwelling). While these links are indirect, they provide the first evidence of local adaptation in addition to demographic independence among harbor porpoise populations along the west coast of North America. The Morro Bay stock was previously recognized based only on a distribution hiatus and evidence of historical population decline due to extensive fishery bycatch (Barlow & Hanan, 1995), as well as blubber pollutant ratios that differed from other areas (Calambokidis & Barlow, 1991). It is potentially subject to offshore energy production disturbance (Forney et al., 2017), and is also at the southern edge of the species' geographical range, and most likely to be impacted by climate change (Learmonth et al., 2006; Ruiz-Cooley et al., 2017). Our results provide strong support for continued management of this population as a separate management stock, especially in light of potential impacts of coastal development and climate change.

Harbor porpoise populations along the west coast of North America have historically experienced substantial fisheries bycatch in portions of the range (Barlow & Hanan, 1995), and multiple lines of evidence indicate that there is limited dispersal among regions.

Within the USA, this led to identification of five coastal management stocks between southern California and the northern US border (coinciding with four of the geographical strata in this study, plus the combined S. Oregon and N. California strata), and one stock in the inland waterways in Washington (Carretta et al., 2019), with stock boundaries placed to coincide with areas of lower density. The genetic results presented here provide additional evidence of limited movement among the currently defined management stocks, as well as between some geographically defined strata within existing stocks, consistent with previous studies of genetics and contaminant levels (Calambokidis & Barlow, 1991; Chivers et al., 2002), and limited movements of tagged porpoises in the inland waterways of Washington (Hanson, 2007). In particular, sPCA and F_{ST} analyses of nuclear data show separation of the N. California and S. Oregon strata (sPCA randtest p = .017) with a small but significant divergence (F_{ST} =0.008) in nuclear allele frequencies (Table 3; Figure 3d, S3D). We suggest that the lack of statistically significant difference in mtDNA is probably a result of low statistical power due to small sample size in N. California (n = 11) rather than results that contradict the finding of differences between these strata based on nuclear markers. Placement of a potential stock boundary as indicated in Figure 5 is suggested by the drop in density in a region where there is a gap in our sample distribution, though the location of the density minimum in Figure 5 varies through time (Figure S20), either due to sampling differences among surveys or small north–south shifts of populations across years.

Genomic methods are evolving rapidly, increasing both the number and the variety of genetic markers that can be used to understand evolution, population structure, historical demography and ecological adaptations (e.g., Tan et al., 2019). This study of harbor porpoise population structure has built on previous research by expanding the number and geographical distribution of samples and the number and type of genetic markers. We also apply new analytical methods to infer patterns of spatial genetic variation, and investigate their correlation with environmental variables that may drive local adaptation. Genetic analyses have been identified as potentially high value for stock delineation (Martien et al., 2019) and these results will be useful for harbor porpoise management stock structure refinement. Nevertheless, inference remains limited by the uneven distribution of opportunistic samples, potential shifts in populations during multiple decades of sample accumulation, and limited availability of environmental variables across the spatial and temporal scale of this study. Additional studies using genome-wide genetic data from across the range are needed to more fully understand habitat use and local adaptation in the eastern North Pacific, and their importance for harbor porpoise management and conservation.

ACKNOWLEDGEMENTS

We are grateful Andy Foote and Marie Louise for help with SNP discovery methods; Diana Baetscher, Cassie Columbus, Carlos Garza and the SWFSC Santa Cruz genetics laboratory for help with GT-seq optimizations; Nathan Campbell and Ninh Vu for help with GT-seq implementation and optimization; Eric Anderson, Thomas Ng, Anthony Clemento and Neil Thompson for help with MICROHAPLOT software and methods; Philipp Kraemer for help with the DEMERELATE R package; Steven Head and John Shimashita, TSRI Sequencing core for sequencing assistance; David Anderson, Cascadia Research Collective for help with Washington Inland Waterways data and observations; and Samuel Woodman for help with analysis of porpoise relative density data. Samples were generously provided by Lisa Ballance, Barbara Taylor and Jay Barlow at the SWFSC; Jan Straley, University of Alaska Southeast; Marilyn Dahlheim, NOAA National Marine Mammal Lab; Jorge Velez-Juarbe, Los Angeles County Museum of Natural History; Merrill Gosho, Bill Walker, Jeff Foster, Stephen Clausen, The Whale Museum, Olympic Coast National Marine Sanctuary and many contributors through the Marine Mammal Stranding Networks. Samples were collected under various Marine Mammal Protection Act permits, and stored in the SWFSC Marine Mammal and Sea Turtle Research (MMASTR) collection. Funding support for this study was provided by NOAA Office of Protected Resources. We are grateful to Aimee Lang and three anonymous reviewers for helpful comments on previous versions of the manuscript.

AUTHOR CONTRIBUTIONS

P. A. M., K. P. and B. L. T. designed the research project, and P. A. M. and K. P. performed analyses and wrote the paper. K. A. F. and B. R. F. contributed environmental data and analyses, and contributed to writing the paper. B. L. H.-H., K. M. R., C. A. C., L. G. B.-L., C. S., T. H. and M. C. F. contributed samples and/or genetic data used for analyses. All authors contributed to preparation of the manuscript.

DATA AVAILABILITY STATEMENT

Mitochondrial DNA haplotype sequences used in this study are complete or partial sequences of sequences previously available in GenBank (Accession nos. provided in Table S2), and are provided in the Dryad repository (Dryad https://doi.org/10.5061/dryad.4tmpg4f6v). Nuclear genotypes and environmental data associated with the samples: Dryad https://doi.org/10.5061/dryad.4tmpg4f6v).

ORCID

Phillip A. Morin https://orcid.org/0000-0002-3279-1519
Brenna R. Forester https://orcid.org/0000-0002-1608-1904
Karin A. Forney https://orcid.org/0000-0002-9195-4701
Brittany L. Hancock-Hanser https://orcid.org/0000-0002-8532-2329
Kelly M. Robertson https://orcid.org/0000-0002-0090-8604
Robin W. Baird https://orcid.org/0000-0002-9419-6336
Michael C. Fontaine https://orcid.org/0000-0003-1156-4154
Barbara L. Taylor https://orcid.org/0000-0001-7620-0736
Kim M. Parsons https://orcid.org/0000-0003-4580-3651

REFERENCES

Ahrens, C. W., Rymer, P. D., Stow, A., Bragg, J., Dillon, S., Umbers, K. D. L., & Dudaniec, R. Y. (2018). The search for loci under selection: Trends, biases and progress. *Molecular Ecology*, *27*, 1342–1356.

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

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

Baetscher, D. S., Clemento, A. J., Ng, T., Anderson, E. C., & Garza, J. C. (2017). Microhaplotypes provide increased power from short-read DNA sequences for relationship inference. *Molecular Ecology Resources*, *18*, 296–305.

Baird, R. W., Willis, P. M., Guenther, T. J., Wilson, P. J., & White, B. N. (1998). An intergeneric hybrid in the family Phocoenidae. *Canadian Journal of Zoology-Revue Canadienne De Zoologie*, *76*, 198–204.

Barlow, J., & Hanan, D. (1995) An assessment of the status of the harbour porpoise in central California. In A. Bjørge, & G. P. Donovan (Eds.), *Biology of the Phocoenids* (pp. 123–140). Report of the International Whaling Commission, Special Issue 16, Cambridge.

Ben Chehida, Y., Thumloup, J., Schumacher, C., Harkins, T., Aguilar, A., Borrel, A., & Fontaine, M. C. (2020). Mitochondrial genomics reveals evolutionary history of the porpoises (Phocoenidae) across the speciation continuum. *Scientific Reports*, *10*, 15190.

Blair, C., Weigel, D. E., Balazik, M., Keeley, A. T. H., Walker, F. M., Landguth, E., Cushman, S., Murphy, M., Waits, L., & Balkenhol, N. (2012). A simulation-based evaluation of methods for inferring linear barriers to gene flow. *Molecular Ecology Resources*, *12*, 822–833.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*, 2114–2120.

Bray, J. R., & Curtis, J. T. (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs*, *27*, 326–349.

Calambokidis, J., & Barlow, J. (1991). Chlorinated hydrocarbon concentrations and their use for describing population discreteness in harbor porpoises from Washington, Oregon, and California. In J. E. Reynolds, & D. K. Odell (Eds.), *Marine mammals strandings in the United States* (pp. 101–110). NOAA Tech. Rep. NMFS.

Calambokidis, J., Laake, J. L., & Osmek, S. D. (1997). *Aerial surveys for marine mammals in Washington and British Columbia inside waters*. Final report to the National Marine Mammal Laboratory, Seattle, WA, USA. Cascadia Research Collective.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*, 421.

Campbell, N. R., Harmon, S. A., & Narum, S. R. (2015). Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. *Molecular Ecology Resources*, *15*, 855–867.

Candy, J. R., Campbell, N. R., Grinnell, M. H., Beacham, T. D., Larson, W. A., & Narum, S. R. (2015). Population differentiation determined from putative neutral and divergent adaptive genetic markers in Eulachon (*Thaleichthys pacificus*, Osmeridae), an anadromous Pacific smelt. *Molecular Ecology Resources*, *15*, 1421–1434.

Capblancq, T., Luu, K., Blum, M. G. B., & Bazin, E. (2018). Evaluation of redundancy analysis to identify signatures of local adaptation. *Molecular Ecology Resources*, *18*, 1223–1233.

Carretta, J. V., Forney, K. A., Oleson, E. M., Weller, D. W., Lang, A. R., Baker, J., Brownell, R. L. Jr. (2019). *U.S. Pacific marine mammal stock assessments: 2018*. U.S. Department of Commerce, NOAA Technical Memorandum NMFS-SWFSC-617.

Carstensen, J., Henriksen, O. D., & Teilmann, J. (2006). Impacts of offshore wind farm construction on harbour porpoises: Acoustic monitoring of echolocation activity using porpoise detectors (T-PODs). *Marine Ecology Progress Series*, *321*, 295–308.

Castelao, R. M., & Luo, H. (2018). Upwelling jet separation in the California Current System. *Scientific Reports*, *8*, 16004. https://www.nature.com/articles/s41598-018-34401-y#Sec2

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

Chin, T. M., Vazquez-Cuervo, J., & Armstrong, E. M. (2017). A multi-scale high-resolution analysis of global sea surface temperature. *Remote Sensing of Environment*, *200*, 154–169.

Chivers, S. J., Dizon, A. E., Gearin, P. J., & Robertson, K. M. (2002). Small-scale population structure of eastern North Pacific harbour porpoises (*Phocoena phocoena*) indicated by molecular genetic analyses. *Journal of Cetacean Research and Management*, *4*, 111–122.

Chivers, S. J., Hanson, B., Laake, J., Gearin, P. J., Muto, M. M., Calambokidis, J., Duffield, D., McGuire, T., Hodder, J., Greig, D., Wheeler, E., Harvey, J., Robertson, K. M., & Hancock, B. L. (2007). *Additional genetic evidence for population structure of Phocoena phocoena off the coasts of California, Oregon, and Washington*. National Marine Fisheries Service, Administrative Report LJ-07-08, Southwest Fisheries Science Center.

Crossman, C. A., Barrett-Lennard, L. G., & Taylor, E. B. (2014). Population structure and intergeneric hybridization in harbour porpoises *Phocoena phocoena* in British Columbia, Canada. *Endangered Species Research*, *26*, 1–12.

Cullingham, C. I., Miller, J. M., Peery, R. M., Dupuis, J. R., Malenfant, R. M., Gorrell, J. C., & Janes, J. K. (2020). Confidently identifying the correct K value using the DeltaK method: when does K = 2? *Molecular Ecology*, *29*(5), 862–869.

Danecek, P., Auton, A., Abecasis, G., Albers, C., Banks, E., DePristo, M., Handsaker, R., Lunter, G., Marth, G., Sherry, S., McVean, G., & Durbin, R.; Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*, 2156–2158.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*, 491–498.

Emerson, K. J., Merz, C. R., Catchen, J. M., Hohenlohe, P. A., Cresko, W. A., Bradshaw, W. E., & Holzapfel, C. M. (2010). Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Science USA*, *107*, 16196–16200.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, *14*, 2611–2620.

Evenson, J. R., Anderson, D., Murphie, B. L., Cyra, T. A., & Calambokidis, J. (2016). *Disappearance and return of harbor porpoise to Puget Sound: 20 year pattern revealed from winter aerial surveys*. Washington Department of Fish and Wildlife, Wildlife Program and Cascadia Research Collective, Olympia, WA. https://www.cascadiaresearch.org/files/publications/wdfw01787.pdf.

Fontaine, M. C. (2016). Harbour porpoises, *Phocoena phocoena*, in the Mediterranean Sea and adjacent regions: Biogeographic relicts of the last glacial period. *Advances in Marine Biology*, *75*, 333–358.

Fontaine, M. C., Baird, S. J. E., Piry, S., Ray, N., Tolley, K. A., Duke, S., Birkun, A., Ferreira, M., Jauniaux, T., Llavona, Á., Öztürk, B., A Öztürk, A., Ridoux, V., Rogan, E., Sequeira, M., Siebert, U., Vikingsson, G. A., Bouquegneau, J.-M., & Michaux, J. R. (2007). Rise of oceanographic barriers in continuous populations of a cetacean: The genetic structure of harbour porpoises in Old World waters. *BMC Biology*, *5*, 30.

Fontaine, M. C., Roland, K., Calves, I., Austerlitz, F., Palstra, F. P., Tolley, K. A., & Aguilar, A. (2014). Postglacial climate changes and rise of three ecotypes of harbour porpoises, *Phocoena phocoena*, in western Palearctic waters. *Molecular Ecology*, *23*, 3306–3321.

Fontaine, M. C., Snirc, A., Frantzis, A., Koutrakis, E., Ozturk, B., Ozturk, A. A., & Austerlitz, F. (2012). History of expansion and anthropogenic collapse in a top marine predator of the Black Sea estimated from genetic data. *Proceedings of the National Academy of Science USA*, *109*, E2569–E2576.

Fontaine, M. C., Thatcher, O., Ray, N., Piry, S., Brownlow, A., Davison, N. J., Jepson, P., Deaville, R., & Goodman, S. J. (2017). Mixing of porpoise ecotypes in southwestern UK waters revealed by genetic profiling. *Royal Society Open Science*, *4*, 160992.

Fontaine, M. C., Tolley, K. A., Siebert, U., Gobert, S., Lepoint, G., Bouquegneau, J. M., & Das, K. (2007). Long-term feeding ecology and habitat use in harbour porpoises *Phocoena phocoena* from Scandinavian waters inferred from trace elements and stable isotopes. *BMC Ecology*, *7*, 1.

Foote, A. D., Liu, Y., Thomas, G. W. C., Vinař, T., Alföldi, J., Deng, J., Dugan, S., van Elk, C. E., Hunter, M. E., Joshi, V., Khan, Z., Kovar, C., Lee, S. L., Lindblad-Toh, K., Mancia, A., Nielsen, R., Qin, X., Qu, J., Raney, B. J., … Gibbs, R. A. (2015). Convergent evolution of the genomes of marine mammals. *Nature Genetics*, *47*, 272–275.

Foote, A. D., Martin, M. D., Louis, M., Pacheco, G., Robertson, K. M., Sinding, M.-H., Amaral, A. R., Baird, R. W., Baker, C. S., Ballance, L., Barlow, J., Brownlow, A., Collins, T., Constantine, R., Dabin, W., Dalla Rosa, L., Davison, N. J., Durban, J. W., Esteban, R., … Morin, P. A. (2019). Killer whale genomes reveal a complex history of recurrent admixture and vicariance. *Molecular Ecology*, *38*, 3427–3444.

Forester, B. R., Lasky, J. R., Wagner, H. H., & Urban, D. L. (2018). Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. *Molecular Ecology*, *27*, 2215–2233.

Forney, K. A., Hanan, D. A., & Barlow, J. (1991). Detecting trends in harbor porpoise abundance from aerial surveys using analysis of covariance. *Fishery Bulletin*, *89*, 367–377.

Forney, K. A., Moore, J. E., Barlow, J., Carretta, J. V., & Benson, S. R. (2020). A multi-decadal Bayesian trend analysis of harbor porpoise (*Phocoena phocoena*) populations off California relative to past fishery bycatch. *Marine Mammal Science*, *2020*, 1–15. https://doi.org/10.1111/mms.12764

Forney, K. A., Southall, B. L., Slooten, E., Dawson, S., Read, A. J., Baird, R. W., & Brownell, R. L. (2017). Nowhere to go: Noise impact assessments for marine mammal populations with high site fidelity. *Endangered Species Research*, *32*, 391–413.

Francois, O., Martins, H., Caye, K., & Schoville, S. D. (2016). Controlling false discoveries in genome scans for selection. *Molecular Ecology*, *25*, 454–469.

Frichot, E., & Francois, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, *6*, 925–929.

Garrison, E., & Marth, G. (2012) *Haplotype-based variant detection from short-read sequencing*. arXiv preprint arXiv:1207.3907 [q-bio.GN].

Gemmell, N. J., & Akiyama, S. (1996). An efficient method for the extraction of DNA from vertebrate tissues. *Trends in Genetics*, *12*, 338–339.

Hancock-Hanser, B., Frey, A., Leslie, M., Dutton, P. H., Archer, E. I., & Morin, P. A. (2013). Targeted multiplex next-generation sequencing: Advances in techniques of mitochondrial and nuclear DNA sequencing for population genomics. *Molecular Ecology Resources*, *13*, 254–268.

Hanson, M. B. (2007). Using location data from telemetry tagged marine mammals to improve stock assessments. In P. Sheridan, J. W. Ferguson, & S. L. Downing (Eds.), *Report of the National Marine Fisheries Service Workshop on Advancing Electronic Tag Technology and Their Use in Stock Assessments*. U.S. Dept. Commerce, NOAA Tech. Memo. NMFSF/SPO-82, 82.

Hickey, B., Geier, S., Kachel, N., Ramp, S., Kosro, P. M., & Connolly, T. (2016). Alongcoast structure and interannual variability of seasonal midshelf water properties and velocity in the Northern California Current System. *Journal of Geophysical Research-Oceans*, *121*, 7408–7430.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65–70.

Hubisz, M. J., Falush, D., Stephens, M., & Pritchard, J. K. (2009). Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, *9*, 1322–1332.

Jacox, M. G., Edwards, C. A., Hazen, E. L., & Bograd, S. J. (2018). Coastal upwelling revisited: Ekman, Bakun, and improved upwelling indices for the US west coast. *Journal of Geophysical Research-Oceans*, *123*, 7332–7350.

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

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

Jombart, T., Devillard, S., Dufour, A. B., & Pontier, D. (2008). Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity*, *101*, 92–103.

Kalendar, R., Lee, D., & Schulman, A. H. (2011). Java web tools for PCR, in silico PCR, and oligonucleotide assembly and analysis. *Genomics*, *98*, 137–144.

Kalinowski, S. T. (2011). The computer program STRUCTURE does not reliably identify the main genetic clusters within species: Simulations and implications for human population structure. *Heredity*, *106*, 625–632.

Keener, W., Webber, M. A., Szczepaniak, I. D., Markowitz, T. M., & Orbach, D. N. (2018). The sex life of harbor porpoises (*Phocoena phocoena*): Lateralized and aerial behavior. *Aquatic Mammals*, *44*, 620–632.

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, *15*, 1179–1191.

Kraemer, P., & Gerlach, G. (2017). Demerelate: calculating interindividual relatedness for kinship analysis based on codominant diploid genetic markers using R. *Molecular Ecology Resources*, *17*, 1371–1377.

Lah, L., Trense, D., Benke, H., Berggren, P., Gunnlaugsson, Þ., Lockyer, C., Öztürk, A., Öztürk, B., Pawliczka, I., Roos, A., Siebert, U., Skóra, K., Víkingsson, G., & Tiedemann, R. (2016). Spatially explicit analysis of genome-wide snps detects subtle population structure in a mobile marine mammal, the harbor porpoise. *PLoS One*, *11*, e0162792.

Learmonth, J. A., MacLeod, C. D., Santos, M. B., Pierce, G. J., Crick, H. Q. P., & Robinson, R. A. (2006). Potential effects of climate change on marine mammals. *Oceanography and Marine Biology - an Annual Review*, *44*, 431–464.

Leigh, J. W., & Bryant, D. (2015). POPART: Fullfeature software for haplotype network construction. *Methods in Ecology and Evolution*, *6*, 1110–1116.

Leslie, M. S., & Morin, P. A. (2016). Using genome-wide SNPs to detect structure in high-diversity and low-divergence populations of severely impacted eastern tropical Pacific spinner (*Stenella longirostris*) and pantropical spotted dolphins (*S. Attenuata*). *Frontiers in Marine Science*, *3*, 253.

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

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

Maisano Delser, P., Corrigan, S., Hale, M., Li, C., Veuille, M., Planes, S., Naylor, G., & Mona, S. (2016). Population genomics of *C. melanopterus* using target gene capture data: demographic inferences and conservation perspectives. *Science Reports*, *6*, 33753.

Martien, K. K., Lang, A. R., Taylor, B. L., Rosel, P. E., Simmons, E., Oleson, E. M., Hanson, M. B. (2019). *The dip delineation handbook: A guide to using multiple lines of evidence to delineate demographically independent populations of marine mammals*. NOAA Technical

Memorandum NOAA-TM-NMFS-SWFSC-622, https://repository.library.noaa.gov.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*, 1297–1303.

McKinney, G. J., Seeb, J. E., & Seeb, L. W. (2017). Managing mixed-stock fisheries: genotyping multi-SNP haplotypes increases power for genetic stock identification. *Canadian Journal of Fisheries and Aquatic Sciences*, *74*, 429–434.

Miller, S. A., Dykes, D. D., & Polesky, H. F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research*, *16*, 1215.

Monmonier, M. (1973). Maximum-difference barriers: An alternative numerical regionalization method. *Geographical Analysis*, *3*, 245–261.

Montano, V., & Jombart, T. (2017). An Eigenvalue test for spatial principal component analysis. *BMC Bioinformatics*, *18*, 562.

Moore, A. M., Arango, H. G., Broquet, G., Edwards, C., Veneziani, M., Powell, B., Foley, D., Doyle, J. D., Costa, D., & Robinson, P. (2011). The Regional Ocean Modeling System (ROMS) 4-dimensional variational data assimilation systems Part II - Performance and application to the California Current System. *Progress in Oceanography*, *91*, 50–73.

Morin, P. A., Foote, A. D., Hill, C. M., Simon-Bouhet, B., Lang, A. R., & Louise, M. (2018) SNP discovery from single and multiplex genome assemblies of non-model organisms. In S. R. Head, P. Ordoukhanian, & D. Salomon (Eds.), *Next-Generation Sequencing. Methods in Molecular Biology* (pp. 113–144). Humana Press.

Moritz, C. (1994). Defining 'Evolutionarily Significant Units' for conservation. *Trends in Ecology and Evolution*, *9*, 373–375.

Nielsen, N. H., Teilmann, J., Sveegaard, S., Hansen, R. G., Sinding, M. H. S., Dietz, R., & Heide-Jorgensen, M. P. (2018). Oceanic movements, site fidelity and deep diving in harbour porpoises from Greenland show limited similarities to animals from the North Sea. *Marine Ecology Progress Series*, *597*, 259–272.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*, 945–959.

Read, A. J. (1999). Harbour porpoise (*Phocoena phocoena*). In S. Ridgway, & R. Harrison (Eds.), *Handbook of marine mammals* (pp. 323–350). Academic Press.

Reeves, R. R., McClellan, K., & Werner, T. B. (2013). Marine mammal bycatch in gillnet and other entangling net fisheries, 1990 to 2011. *Endangered Species Research*, *20*, 71–97.

Riginos, C., Crandall, E. D., Liggins, L., Bongaerts, P., & Treml, E. A. (2016). Navigating the currents of seascape genomics: How spatial analyses can augment population genomic studies. *Current Zoology*, *62*, 581–601.

Rosel, P. E., France, S. C., Wang, J. Y., & Kocher, T. D. (1999). Genetic structure of harbour porpoise *Phocoena phocoena* populations in the northwest Atlantic based on mitochondrial and nuclear markers. *Molecular Ecology*, *8*, S41–S54.

Ruiz-Cooley, R. I., Gerrodette, T., Fiedler, P. C., Chivers, S. J., Danil, K., & Ballance, L. T. (2017). Temporal variation in pelagic food chain length in response to environmental change. *Science. Advances*, *3*(10), e1701140.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular cloning: A laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press.

Selkoe, K. A., D'Aloia, C. C., Crandall, E. D., Iacchei, M., Liggins, L., Puritz, J. B., von der Heyden, S., & Toonen, R. J. (2016). A decade of seascape genetics: Contributions to basic and applied marine connectivity. *Marine Ecology Progress Series*, *554*, 1–19.

Shirk, A. J., Landguth, E. L., & Cushman, S. A. (2017). A comparison of individual-based genetic distance metrics for landscape genetics. *Molecular Ecology Resources*, *17*, 1308–1317.

Storey, J. D., & Tibshirani, R. (2003). Statistical significance for genome-wide studies. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 9440–9445.

Taguchi, M., Chivers, S. J., Rosel, P. E., Matsuishi, T., & Abe, S. (2010). Mitochondrial DNA phylogeography of the harbour porpoise *Phocoena phocoena* in the North Pacific. *Marine Biology*, *157*, 1489–1498.

Tan, M. P., Wong, L. L., Razali, S. A., Afiqah-Aleng, N., Mohd Nor, S. A., Sung, Y. Y., Van de Peer, Y., Sorgeloos, P., & Danish-Daniel, M. (2019). Applications of next-generation sequencing technologies and computational tools in molecular evolution and aquatic animals conservation studies: A short review. *Evolutionary Bioinformatics*, *15*, 1176934319892284.

Tiedemann, R., Harder, J., Gmeiner, C., & Haase, E. (1996). Mitochondrial DNA sequence patterns of harbour porpoises (*Phocoena phocoena*) from the North and the Baltic sea. *Zeitschrift Fur Saugetierkunde-International Journal of Mammalian Biology*, *61*, 104–111.

Venegas, R. M., Strub, P. T., Beier, E., Letelier, R., Thomas, A. C., Cowles, T., James, C., Soto-Mardones, L., & Cabrera, C. (2008). Satellite-derived variability in chlorophyll, wind stress, sea surface height, and temperature in the northern California Current System. *Journal of Geophysical Research-Oceans*, *113*, C03015.

Walton, M. J. (1997). Population structure of harbour porpoises *Phocoena phocoena* in the seas around the UK and adjacent waters. *Proceedings of the Royal Society B-Biological Sciences*, *264*, 89–94.

Wang, J. (2019). A parsimony estimator of the number of populations from a STRUCTURE-like analysis. *Molecular Ecology Resources*, *19*, 970–981.

Wang, J. Y., & Berggren, P. (1997). Mitochondrial DNA analysis of harbour porpoises (*Phocoena phocoena*) in the Baltic Sea, the Kattegat-Skagerrak Seas and off the west coast of Norway. *Marine Biology*, *127*, 531–537.

Waples, R. S., & Gaggiotti, O. (2006). What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, *15*, 1419–1439.

Willing, E. M., Dreyer, C., & van Oosterhout, C. (2012). Estimates of genetic differentiation measured by F(ST) do not necessarily require large sample sizes when using many SNP markers. *PLoS One*, *7*, e42649.

Willis, P. M., Crespi, B. J., Dill, L. M., Baird, R. W., & Hanson, M. B. (2004). Natural hybridization between Dall's porpoises (*Phocoenoides dalli*) and harbour porpoises (*Phocoena phocoena*). *Canadian Journal of Zoology-Revue Canadienne De Zoologie*, *82*, 828–834.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Morin PA, Forester BR, Forney KA, et al. Population structure in a continuously distributed coastal marine species, the harbor porpoise, based on microhaplotypes derived from poor-quality samples. *Mol Ecol*. 2021;30:1457–1476. https://doi.org/10.1111/mec.15827