



**HAL**  
open science

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

Phillip A. Morin, Brenna R. Forester, Karin A. Forney, Carla A. Crossman, Brittany L. Hancock-Hanser, Kelly M. Robertson, Lance G. Barrett-Lennard, Robin W. Baird, John Calambokidis, Pat Gearin, et al.

### ► To cite this version:

Phillip A. Morin, Brenna R. Forester, Karin A. Forney, Carla A. Crossman, Brittany L. Hancock-Hanser, et al.. Population structure in a continuously distributed coastal marine species, the harbor porpoise, based on microhaplotypes derived from poor quality samples. *Molecular Ecology*, 2021, 30 (6), pp.1457-1476. 10.1111/mec.15827 . hal-03088621

**HAL Id: hal-03088621**

**<https://hal.science/hal-03088621>**

Submitted on 28 Oct 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

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

8  
9 <sup>1</sup> Southwest Fisheries Science Center, National Marine Fisheries Service, NOAA, 8901  
10 La Jolla Shores Dr., La Jolla, CA, USA

11 <sup>2</sup> Department of Biology, Colorado State University, Fort Collins, CO, USA

12 <sup>3</sup> Southwest Fisheries Science Center, National Marine Fisheries Service, NOAA, 7544  
13 Sandholdt Rd, Moss Landing CA 95039 USA

14 <sup>4</sup> Moss Landing Marine Laboratories, San Jose State University, 7544 Sandholdt Rd,  
15 Moss Landing CA 95039 USA

16 <sup>5</sup> Cetacean Research Program, Vancouver Aquarium, PO Box 3232, Vancouver, British  
17 Columbia V6B 3X8, Canada

18 <sup>6</sup> Cascadia Research Collective, Olympia, Washington

19 <sup>7</sup> Marine Mammal Laboratory, Alaska Fisheries Science Center, National Marine  
20 Fisheries Service, NOAA, 7600 Sand Point Way NE, Seattle, WA 98115

21 <sup>8</sup> Northwest Fisheries Science Center, National Marine Fisheries Service, NOAA, 2725  
22 Montlake Blvd. E, Seattle, WA, USA

23 <sup>9</sup> Swift Biosciences, 674 S. Wagner Rd., Suite 100, Ann Arbor, MI 48103, USA.

24 <sup>10</sup> MIVEGEC Research Unit (Université de Montpellier, CNRS, IRD), Centre IRD de  
25 Montpellier, Montpellier, France.

26 <sup>11</sup> Groningen Institute for Evolutionary Life Sciences (GELIFES), University of  
27 Groningen, PO Box 11103 CC, Groningen, The Netherlands

28  
29  
30 \*Current address: Biology Department, Saint Mary's University, Halifax, NS, B3H 3C3,  
31 Canada

32  
33 Abstract: 246

34 Main text: 7849 words

35

36 **Abstract**

37 Harbor porpoise in the North Pacific are found in coastal waters from southern California  
38 to Japan, but population structure is poorly known outside of a few local areas. We used  
39 multiplexed amplicon sequencing of 292 loci and genotyped clusters of SNPs as  
40 microhaplotypes (N=271 samples) in addition to mtDNA sequence data (N=413  
41 samples), to examine the genetic structure from samples collected along the Pacific coast  
42 and inland waterways from California to southern British Columbia. We confirmed an  
43 overall pattern of strong isolation-by-distance, suggesting that individual dispersal is  
44 restricted. We also found evidence of regions where genetic differences are larger than  
45 expected based on geographic distance alone, implying current or historical barriers to  
46 gene flow. In particular, the southernmost population in California is genetically distinct  
47 ( $F_{ST} = 0.02$  (microhaplotypes); 0.31 (mtDNA)), with both reduced genetic variability and  
48 high frequency of an otherwise rare mtDNA haplotype. At the northern end of our study  
49 range, we found significant genetic differentiation of samples from the Strait of Georgia,  
50 previously identified as a potential biogeographic boundary or secondary contact zone  
51 between harbor porpoise populations. Association of microhaplotypes with remotely-  
52 sensed environmental variables indicated potential local adaptation, especially at the  
53 southern end of the species' range. These results inform conservation and management  
54 for this nearshore species, illustrate the value of genomic methods for detecting patterns  
55 of genetic structure within a continuously distributed marine species, and highlight the  
56 power of microhaplotype genotyping for detecting genetic structure in harbor porpoises  
57 despite reliance on poor-quality samples.

58

59 Key Words: sPCA, dbRDA, seascape genetics, mtDNA, SNP, microhaplotype, GT-seq,  
60 cetacean

61

62 **Introduction**

63

64 Harbor porpoises (*Phocoena phocoena*) are small cetaceans found in temperate and sub-  
65 Arctic coastal waters, typically in less than 200m depth (Fontaine, 2016; Read, 1999) (but  
66 see Nielsen et al., 2018). Their range is considered continuous along the continental  
67 coasts, but there is high variability in density and apparent gaps in suitable habitat  
68 (Evenson, Anderson, Murphie, Cyra, & Calambokidis, 2016; Forney, Moore, Barlow,  
69 Carretta, & Benson, in press). They are susceptible to entanglement in gillnets and  
70 disturbance from construction activities for wind farms (Carstensen, Henriksen, &  
71 Teilmann, 2006; Reeves, McClellan, & Werner, 2013). In addition, the linear coastal  
72 distributions with gaps raises the question of how harbor porpoise will respond to climate  
73 change. Areas with sufficient data indicate fine-scaled population structure is common  
74 (Chivers, Dizon, Gearin, & Robertson, 2002; Crossman, Barrett-Lennard, & Taylor,  
75 2014; Fontaine, 2016; Lah et al., 2016; Rosel, France, Wang, & Kocher, 1999;  
76 Tiedemann, Harder, Gmeiner, & Haase, 1996; Walton, 1997; Wang & Berggren, 1997).  
77 Along the U.S. west coast, early genetic studies based on mitochondrial DNA (mtDNA)  
78 provided data to delineate management stocks (Chivers et al., 2002), but data gaps made  
79 boundary placement difficult. Because harbor porpoises avoid vessels, which makes dart-  
80 biopsy an unviable method to obtain sufficient samples, most samples are from either  
81 beach-stranded animals or entangled animals from areas with gillnet fisheries. As a result,  
82 sampling gaps existed in areas without gillnet fisheries and samples were often obtained  
83 days after death and in moderate to advanced stages of decomposition, and hence had  
84 degraded DNA.

85

86 Although gillnetting has declined along the U.S. west coast, mitigation of other potential  
87 threats, such as development of wind farms, requires an understanding of population  
88 structure, often in areas poorly sampled in previous studies. The U.S. Marine Mammal  
89 Protection Act (MMPA) manages at the scale of demographically independent  
90 populations (DIPs) where allele (or mtDNA haplotype) frequencies will differ but  
91 evolutionary differences are not expected. Because our area of interest includes the  
92 southernmost part of harbor porpoise distribution in the Pacific where ocean temperatures

93 are already rising, an understanding of evolutionary barriers to dispersal is also of interest  
94 to evaluate potential impacts of such temperature shifts.

95

96 In our study area along the U.S. West Coast and inland waters of Washington, samples  
97 have slowly accrued to fill most gaps over a period of 30 years, but most samples are of  
98 degraded quality and sample size remains low in some areas. An additional complication  
99 requiring consideration for marker choice is that there is evidence of intergeneric  
100 hybridization with Dall's porpoise (*Phocoenoides dalli*) in the eastern North Pacific  
101 (Baird, Willis, Guenther, Wilson, & White, 1998; Crossman et al., 2014; Willis, Crespi,  
102 Dill, Baird, & Hanson, 2004), which could affect population analyses, especially if  
103 hybridization has been regionally restricted and/or resulted in introgressive gene flow  
104 between species.

105

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

120

121 To increase power to detect population structure from larger numbers of SNPs, while  
122 maximizing our ability to use an existing collection of poor-quality samples collected  
123 over three decades, we employed GT-seq multiplex sequencing combined with

124 microhaplotype analysis, and traditional mtDNA control region sequencing to incorporate  
125 new and previously published data. We describe population structure at both evolutionary  
126 and demographic scales through a continuous range in the eastern North Pacific by  
127 applying a combination of Bayesian assignment to detect more substantial divergence of  
128 evolutionarily significant units (ESUs; Moritz, 1994), and traditional divergence metrics  
129 and spatially explicit methods to detect demographically independent populations (DIPs),  
130 the basis of marine mammal population stocks under the U.S. Marine Mammal Protection  
131 Act (Martien et al., 2019; Waples & Gaggiotti, 2006). Evolutionarily significant units are  
132 expected to have very low gene flow (on the order of a few successful dispersers per  
133 generation) and significant differences in both nuclear and mitochondrial markers. In  
134 contrast, a demographic independence means that the population dynamics of the affected  
135 group is more a consequence of births and deaths within the group (internal dynamics)  
136 than immigration or emigration (external dynamics) (Martien et al., 2019). DIPs could  
137 have gene flow on the order of 1-2% per year. Besides the large difference in gene flow,  
138 DIPs could be based solely on mtDNA because birth and death rates depend on females,  
139 though use of larger numbers of nuclear markers may provide higher statistical power to  
140 detect population differences.

141

142 We also use seascape genotype-environment association to investigate patterns of  
143 localized adaptation. We anticipate that these genomic scale data will not only provide  
144 increased resolution to detect spatial structure among populations, but also allow us to  
145 link genotypes to environmental variables that vary both spatially and temporally. This  
146 emerging field of seascape genomics (Riginos, Crandall, Liggins, Bongaerts, & Treml,  
147 2016) is still in its early stages relative to landscape genomics, due to both the difficulty  
148 of broad scale sampling of many marine species and the limited availability of relevant  
149 environmental predictors. However, the increasing accessibility of remotely-sensed  
150 oceanographic variables has improved our ability to identify spatial, temporal, and  
151 ecological factors that promote population structure and local adaptation in complex and  
152 dynamic seascape environments (reviewed by Riginos et al., 2016; Selkoe et al., 2016).

153

154 **Materials and Methods**

155

156 **Samples**

157 Harbor porpoise (N=441), Dall's porpoise (N=9) and putative Harbor/Dall's porpoise  
158 hybrids (N=13) skin samples were collected from beach-cast carcasses, carcasses  
159 recovered as fisheries bycatch, or from animals live-captured for tagging. Tissue samples  
160 were preserved in salt-saturated 20% DMSO or 100% ethanol and subsequently stored at  
161 -20°C in the U.S. National Marine Fisheries Service (NMFS) Marine Mammal and Sea  
162 Turtle Research (MMASTR) Collection at the Southwest Fisheries Science Center  
163 (SWFSC). Sample information is in supplemental Table S1 and locations are shown in  
164 Figure 1. DNA was extracted from tissue samples using a variety of common extraction  
165 methods, including silica-based filter membranes (Qiaextractor® DX reagents, Qiagen,  
166 Valencia, CA, USA), standard phenol/chloroform extraction (modified from Sambrook,  
167 Fritsch, & Maniatis, 1989), lithium chloride (Gemmell & Akiyama, 1996) and sodium  
168 chloride protein precipitation (Miller, Dykes, & Polesky, 1988), and quantified using the  
169 Quant-iT PicoGreen dsDNA assay kit (Invitrogen, Paisley, UK) with a Victor X3  
170 fluorospectrometer (Perkin Elmer, Waltham, MA, USA).

171

172 **mtDNA control region sequencing**

173 New 395bp mitochondrial DNA control region sequences (N=176) were generated  
174 according to previously published methods for harbor porpoise studies (Chivers et al.,  
175 2002; Chivers et al., 2007; Crossman et al., 2014). Electropherograms of sequences from  
176 previous studies (N=224 from Chivers et al. 2007; N=91 from Crossman et al. 2014)  
177 were evaluated and compared to newly generated sequences, and regions with  
178 ambiguities or poor quality were either re-evaluated by a single person to ensure  
179 consistent interpretation (N=224), or the DNA was sequenced again from new PCR  
180 reactions using current Sanger sequencing chemistry (N=91). Haplotype IDs were  
181 assigned using the LabelHaplotypes function in the R package *strataG* (v2.4.905; Archer,  
182 Adams, & Schneiders, 2017) and associated with previously published sequences and  
183 haplotypes (supplemental Table S2).

184

185 **SNP discovery**

186 DNA from 12 North Pacific harbor porpoises was used for indexed genomic library  
187 preparation using the Accel-NGS 2S PCR-free genomic library preparation kit (Swift  
188 Biosciences, Ann Arbor, MI, USA). Either 100 ng or 500 ng of genomic DNA,  
189 determined by Qubit (Thermo Fisher Scientific, Waltham, MA, USA) quantification, was  
190 used for the library prep to obtain sufficient library product to pool 15 nmol/l of each  
191 sample prior to sequencing. For 12 North Atlantic samples, genomic DNA was used for  
192 indexed genomic library preparation at BGI (Hong Kong) following their proprietary  
193 protocol and pooled for next generation sequencing (NGS). The North Pacific pooled  
194 library was first sequenced in two 150 bp, paired-end Illumina, Inc. (La Jolla, CA, USA)  
195 MiSeq lanes, then both pooled libraries were sequenced in two Illumina HiSeq-4000  
196 lanes each.

197

198 NGS read data were trimmed and filtered using Trimmomatic (v. 0.36; Bolger, Lohse, &  
199 Usadel, 2014) to remove Illumina adapters, reads <50bp and low-quality bases (Q<20) at  
200 the beginning and end of reads. The sliding window approach was implemented to  
201 change internal bases with Q<15 (sliding window of 4bp) to N's. Because there was  
202 some bias in the nucleotide frequencies, the first 4 bp of all sequences were also  
203 removed.

204

205 Genome alignment and SNP discovery were conducted as previously described (Morin et  
206 al., 2018). The repeat-masked killer whale genome (accession GCA\_000331955.2; Foote  
207 et al., 2015) was used as the reference for genome assembly. Briefly, the paired-end  
208 MiSeq reads from the 12 North Pacific samples were assembled *de novo* using CLC  
209 Genomics Workbench (v.4.1; CLCbio) to obtain a complete reference mitochondrial  
210 genome as previously described (Hancock-Hanser et al., 2013). The NGS reads from 12  
211 samples from each ocean basin subspecies were aligned separately to the harbor porpoise  
212 mitochondrial genome using BWA mem (v. 0.7.5a; Li & Durbin, 2009) and non-aligned  
213 reads were extracted using samtools (v. 1.2; Li et al., 2009). The extracted (non-  
214 mitochondrial) reads from each sample were separately aligned to the killer whale  
215 reference genome, combined into one alignment, and the consensus harbor porpoise  
216 genome sequence generated. The nuclear DNA reads for each sample were aligned to the



217 new harbor porpoise consensus genome sequence as above, followed by SNP discovery  
218 separately from each subspecies using GATK (v. 2.5-2; DePristo et al., 2011; McKenna  
219 et al., 2010). Potential SNPs were filtered to remove SNPs with mapping quality <30,  
220 excessive coverage (>2x mean depth of coverage), and estimated minor allele frequency  
221 <0.05. To avoid linked loci, SNPs were selected from contigs that were at least 100 kb in  
222 length, and SNPs were at least 100 kb apart on the contigs. SNP loci from the two  
223 subspecies were compared to identify loci that were polymorphic in both ocean basins to  
224 avoid ascertainment bias in application to either subspecies. Finally, candidate SNPs  
225 were compared to GenBank using BLAST+ (Camacho et al., 2009) and filtered to  
226 remove loci that were potentially in repeat regions, gene families, or a close match to  
227 non-mammalian species. The resulting set of filtered SNPs are subsequently referred to  
228 as "targeted SNPs", as these were the initial targets for our SNP genotyping effort.  
229 Additional SNPs detected in short GT-seq sequences were combined with the targeted  
230 SNPs to form microhaplotype genotypes.

231

### 232 **Multiplex primer design**

233 Primers were designed from a batch of 500 loci selected randomly from the filtered SNPs  
234 using the program FastPCR (Kalendar, Lee, & Schulman, 2011). Parameters for primer  
235 selection (from 300 bp sequences with the target SNP at position 151) were: primer  
236 length = 15-32, tm = 57 – 62, 3' Tm = 25-50, dimer stringency = 5, synchronized Tm for  
237 primer pair = 5, Forward primers between position 40 and 150, Reverse primers between  
238 position 152 and the 3' end, and addition of 5' GT-seq tails for indexing and library  
239 preparation: F-tail = CGACAGGTTTCAGAGTTCTACAGTCCGACGATC, R-tail =  
240 GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT. After primer design, all loci  
241 with both forward and reverse primers were compared using the "Primers list analysis"  
242 function in FastPCR to detect cross-locus primer dimer interactions with Tm > 20°C  
243 ("strong" primer dimers). Loci were filtered out of the primer list if either one or both  
244 primers had predicted interactions with Tm > 40°C with >2 other primers, or if the  
245 predicted primer interactions were >46°C. Primers for 385 loci were synthesized at 100  
246 µM concentration in 96-well plates by Integrated DNA Technologies (Coralville, Iowa,  
247 USA).

248

### 249 **Multiplex PCR optimization**

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

262

### 263 **SNP genotyping**

264 Amplicon libraries were prepared following the GT-seq protocol, including the optional  
265 Exo-SAP pre-treatment of the samples (Campbell et al., 2015), and pooled libraries were  
266 sequenced on an Illumina NextSeq500 sequencer, 1x150 bp reads. Custom scripts for  
267 processing GT-seq data (Campbell et al., 2015) were used to demultiplex the sample files  
268 and conduct preliminary genotyping. Genotypes were quality checked for duplicate  
269 samples, percent missing genotypes per locus and sample, and percent homozygosity  
270 using the *strataG* package in R. Replicate samples were used to estimate genotyping error  
271 rates, then fastq files from replicates samples were combined to a single file per sample.

272

273 Fastq files were checked for standard quality metrics (e.g., per base quality scores,  
274 nucleotide composition, sequence duplication level, overrepresented sequences) with  
275 FASTQC v0.11.3 (Babraham Bioinformatics), then trimmed using FASTP (Chen, Zhou,  
276 Chen, & Gu, 2018) to remove adapter sequences and poly-A and poly-G 3' tails that were  
277 added during sequencing of amplicons shorter than 150 bp, and to exclude reads shorter  
278 than 30bp after trimming. Reads were mapped to the reference locus sequences using the

279 BWA MEM algorithm (v. 0.7.15; Li & Durbin, 2009), and SNPs detected across all  
280 samples using FREEBAYES v1.1.0-54-g49413aa (Garrison & Marth, 2012) after  
281 removing sample files smaller than 1 MB (containing <0.1 M filtered reads).  
282 FREEBAYES was run with minimal filtering (supplemental materials), followed by  
283 additional filtering with vcfTools v0.1.12b (Danecek et al., 2011) to extract the targeted  
284 SNP for each locus (for targeted SNP analysis; see below), and to remove sites with low  
285 coverage (minimum depth = 10), indels, and loci with less than 30% completed  
286 genotypes.

287

288 For targeted SNPs (position 151 in all reference sequences), custom scripts in R  
289 (supplemental materials) were used to extract the genotype data from the vcf file,  
290 generate allelic count plots to visualize the genotype distributions of reads for each allele,  
291 and re-call genotypes based on minimum depth and allelic ratios. The minimum depth of  
292 10 reads total and default minor allele read proportion for heterozygotes of >0.3 were  
293 adjusted as needed until genotypes clearly fell into separate clusters in the allelic plots.  
294 Loci with poor resolution of plotted genotypes were removed from the data set.

295

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

306

307 A final combined data set for microhaplotypes and targeted SNPs was created by  
308 combining the multi-SNP loci with the single-SNP (targeted) loci. Since the targeted SNP  
309 loci had been genotyped using two different methods, we selected the genotype data for

310 each locus from the method that provided the higher quality or quantity of genotypes for  
311 the targeted SNP. Some microhaplotype loci that were monomorphic in harbor porpoises,  
312 or which had similar genotype quality for the targeted SNP but contained other SNPs  
313 present in Dall's porpoise samples, were retained (instead of the targeted locus data) to  
314 allow genetic identification of intergeneric hybrids. All loci that were found only in the  
315 microhaplotype or targeted SNP data set were then added to the filtered loci to generate a  
316 final data set.

317

### 318 **Quality analysis**

319 Quality analysis and sample and locus filtering were conducted using custom R scripts.  
320 Samples missing >80% of the genotypes, and loci missing more than 45% of the  
321 genotypes were also removed. Genetic duplicates (>80% identity) were identified and  
322 one from each pair of samples identified as duplicates was removed. Analyses of  
323 deviations from Hardy-Weinberg equilibrium (HWE) expectations were conducted across  
324 all samples using the R package *adegenet*, and loci with a difference between observed  
325 and expected heterozygosity >0.2 were removed as extreme outliers (>10x the average  
326 difference of 0.02), most likely due to non-mendelian loci (e.g., null alleles, duplicated  
327 loci, or high error rates). Remaining loci were tested for significant linkage  
328 disequilibrium (LD) and deviations from HWE within three discrete, geographically-  
329 defined strata represented by greater than 20 samples (inland waterways (N=88), Neah  
330 Bay (N=21), and Northern California/Southern Oregon (N=35)), after correction for  
331 multiple tests using a sequential correction (Holm, 1979). We used the R package  
332 *Demerelate* (v. 0.9.3; Kraemer & Gerlach, 2017) using relatedness estimators "Wang"  
333 and "Mxy" to test for inadvertent sampling of close relatives.

334

### 335 **Porpoise distribution data**

336 To examine the genetic results in the context of harbor porpoise distribution and relative  
337 density along the U.S. West Coast, independent aerial survey data collected during 1991-  
338 2017 off California (e.g., Forney, Hanan, & Barlow, 1991), and during 1989-2003 off  
339 Oregon and Washington (e.g., Calambokidis, Laake, & Osmeck, 1997) were processed to  
340 derive the number of porpoise seen per kilometer surveyed as an index of relative

341 density. Although the transect design differed between these two data sets, the survey  
342 protocols, observer team and configuration, and aircraft type were the same. Survey data  
343 were truncated spatially to include only the primary porpoise habitat extending from  
344 shore out to 90-100 m water depth. Relative densities were calculated for each transect  
345 line, assigned to latitude of the transect mid-point, and then smoothed south-to-north  
346 using a Loess smoother.

347

### 348 **Habitat data**

349 A variety of modeled and measured oceanographic variables (Table 1) were extracted to  
350 examine potential environmental correlates of genetic patterns. These predictors included  
351 1) sea surface temperature, sea surface height, mixed layer depth, and the standard  
352 deviation of these three variables derived from the Regional Ocean Modeling System  
353 (ROMS) outputs (Moore et al., 2011), 2) coastal upwelling indices (Jacox, Edwards,  
354 Hazen, & Bograd, 2018), and 3) multispectral ultra-high resolution sea surface  
355 temperature and its standard deviation (Chin, Vazquez-Cuervo, & Armstrong, 2017).

356

### 357 **Analytical methods**

358 Several methods were used to estimate the number of populations and population  
359 assignment based on the genetic data. Population structure and individual assignment was  
360 examined using STRUCTURE (v. 2.3), which implements a Bayesian clustering method  
361 to identify significant genetic clusters based on Hardy-Weinberg equilibrium allele  
362 frequency expectations (Hubisz, Falush, Stephens, & Pritchard, 2009; Pritchard,  
363 Stephens, & Donnelly, 2000). We ran ten replicates for each value of k (where k is the  
364 number of putative populations), using correlated allele frequencies and an admixture  
365 model, with location prior using geographically defined units (supplemental Table S1).  
366 Each analysis consisted of 50,000 burn-in steps followed by 100,000 MCMC steps, and  
367 10 replicates combined using 100 iterations in CLUMPP (v. 1.1.2; Jakobsson &  
368 Rosenberg, 2007). We also used CLUMPAK (Kopelman, Mayzel, Jakobsson, Rosenberg,  
369 & Mayrose, 2015) to assess convergence of the MCMCs and evaluate consistency of  
370 replicates across values of K. The  $\Delta K$  method (Evanno, Regnaut, & Goudet, 2005) was  
371 used to evaluate most likely number of inferred clusters.

372

373 STRUCTURE is known to have low power to detect populations (i.e., DIPs) when there  
374 is even a very low migration rate ( $m \geq 0.005$ /generation), where demographic  
375 independence and biologically meaningful differentiation still exist (e.g., Cullingham et  
376 al., 2020; Kalinowski, 2011). To better differentiate DIPs, we examined spatially explicit  
377 principal components with geographic information using sPCA (Jombart, Devillard,  
378 Dufour, & Pontier, 2008). Spatial distances were based on type 1 (Delaunay  
379 triangulation) connection network. We tested for significant evidence of structure in the  
380 sPCA using the Eigenvalue test “spca\_randtest” (Montano & Jombart, 2017) with 9999  
381 permutations in the R package *adegenet* (v. 2.1.1; Jombart, 2008). Geographical subsets  
382 of the data were analyzed hierarchically to evaluate structure at decreasing spatial scales.  
383

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

394

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

402

403 We used distance-based redundancy analysis (dbRDA) to investigate genotype-  
404 environment associations and identify microhaplotypes potentially under selection  
405 (Forester, Lasky, Wagner, & Urban, 2018). dbRDA identifies how groups of SNPs or  
406 microhaplotypes covary in response to the multivariate environment. It is well-suited to  
407 isolation-by-distance demographic scenarios, maintaining both high true positive and low  
408 false positive rates (Forester et al., 2018). Environmental variables (Table 1) were  
409 extracted from longitudinal oceanographic data matched to collection date, latitude and  
410 longitude of samples. We used Bray-Curtis dissimilarity (Bray & Curtis, 1957) to  
411 calculate the microhaplotype distance matrix. This approach quantifies the dissimilarity  
412 among individuals based on their multilocus genotypes, and is equivalent to proportion of  
413 shared alleles (Shirk, Landguth, & Cushman, 2017). We performed dbRDA separately  
414 for porpoises from the outer coastal and inland waterways regions, because data for all  
415 environmental predictors were not available for all locations (Table 1).

416

417 For dbRDA, we first removed microhaplotype loci with heterozygosity less than 0.05,  
418 then removed individuals with missing data for the retained environmental predictors. We  
419 produced three data sets for each region (i.e., outer coastal and inland waterways),  
420 representing three thresholds of missing genotype data across individuals: 25%, 20%, and  
421 15%. We imputed missing values for each data set using *snmf* in the LEA package v.  
422 3.1.2 (Frichot & Francois, 2015), testing values of K from 1-5, and alpha (regularization  
423 parameter) values of 10, 100, and 1000. All runs used 25 repetitions, 200 iterations, and a  
424 5% cross entropy withholding. We then performed a dbRDA for each imputed data set,  
425 retaining three constrained axes for outlier analysis. We identified candidate  
426 microhaplotypes under selection using robust (e.g., not sensitive to outliers) Mahalanobis  
427 distance (Capblancq, Luu, Blum, & Bazin, 2018), which identifies outlier  
428 microhaplotypes based on their constrained ordination loadings in multidimensional  
429 space. We accounted for confounding factors in the dbRDA, such as population structure  
430 and isolation-by-distance, using the genomic inflation factor (Francois, Martins, Caye, &  
431 Schoville, 2016), and applied a false discovery rate cutoff of 0.1 (Storey & Tibshirani,  
432 2003) to identify outlier microhaplotypes showing relationships with environmental  
433 variation. Finally, we compared detections across missing data thresholds.

434

435 **RESULTS**

436

437 We used a total of 431 mtDNA control region sequences (395bp) from previously  
438 published (N=363, including re-sequenced samples) and newly generated (N=68)  
439 sequences, and SNP data for 296 porpoises ranging from the southern extent of the  
440 harbor porpoise range in Southern California, USA, to British Columbia, Canada  
441 (Supplemental Table S1, Figure S1). Resequencing of samples with previously published  
442 haplotypes resulted in ten haplotype changes from the Chivers et al. (2002; 2007) studies,  
443 and 53 from Crossman et al. (2014), mostly due to resolution of ambiguous positions in  
444 the previous electropherograms that resulted in synonymization of multiple haplotypes  
445 from each of those studies. We identified 52 harbor porpoise haplotypes, 22 of which  
446 were found in only a single individual, and six Dall's porpoise haplotypes (supplemental  
447 Tables S4, S5). Our GT-seq locus panel consisted of 340 loci, of which 290 were  
448 genotyped in at least 55% of the samples and were polymorphic in harbor porpoises. Two  
449 additional loci were polymorphic only in Dall's porpoises and were used to identify  
450 hybrids between the 2 species. Of the 290 loci, 151 (52%) contained a single SNP, while  
451 the remaining 139 (48%) contained  $\geq 2$  SNPs, genotyped as microhaplotypes. None of the  
452 loci deviated significantly from HWE, and significant LD was only detected in one locus  
453 pair in one of the three tested geographic strata; no loci were removed based on these  
454 tests of HWE or LD.

455

456 DNA quantity and quality varied substantially among samples, resulting in variable  
457 number of completed genotypes, and an inverse correlation of error rate with the number  
458 of completed genotypes. Arbitrarily changing the cut-off value for percent completed  
459 genotypes can result in slight changes (see error rates below) to overall data quality, but  
460 at the cost of reduction in sample sizes in individual strata, reducing statistical power to  
461 detect structure. To maximize the sample sizes across strata, we used 20% ( $\geq 58$  of 292)  
462 genotyped loci as the minimum cut-off, resulting in 296 genotyped samples (after  
463 removal of unintentional duplicates (see below): 280 harbor porpoise, 11 Dahl's porpoise,



464 5 hybrids). Of the harbor porpoise samples, 72% were genotyped at >90% of the loci, and  
465 83% were genotyped at >80% of the loci (supplemental Table S1).

466

467 Average per-allele error rates for SNPs were calculated from 32 samples genotyped in  
468 duplicate from separate GT-seq amplicon libraries, based on single SNPs in 270 loci that  
469 were genotyped in >50% of samples. Intentionally replicated samples for which there  
470 was sufficient data in both replicates (N=32) had matching genotypes at an average of  
471 96% of the loci (range 87-100%). For samples genotyped at >80% of the loci, the  
472 estimated error rate (based on 23 replicated sample pairs) was 0.010/allele. For samples  
473 with lower genotype completion rates (46 in the harbor porpoise data set), the mean error  
474 rate estimate increased to 0.045/allele (based on N=9 replicate pairs). Nine sample pairs  
475 (excluding Dall's porpoise samples, which were too homozygous in this data set) were  
476 identified as unintentional duplicates based on at least 85% identical genotypes (range  
477 96-100%), and one from each pair was removed from further analysis. All of the genetic  
478 duplicates were inadvertent duplicate samples from the same individual, usually due to  
479 sample being stored in different collections with different identification codes. We  
480 detected three potential first order relatives (full siblings or parent-offspring pairs) based  
481 on the "wang" estimator, and six (including two of the three from the "wang" estimator)  
482 based on the "Mxy" estimator (supplemental Table S6) (see Kraemer & Gerlach, 2017,  
483 for details). All samples were retained for some of the subsequent analyses, but one  
484 individual from each putative pair was removed to control for the effects of sampling  
485 closely related individuals in STRUCTURE, sPCA, and  $F_{ST}$  analyses.

486

487 Putative hybrids between parapatric Dall's and harbor porpoise species have been  
488 previously identified based on phenotype and genetic profiles (Crossman et al., 2014).  
489 We genotyped nine known Dall's porpoises and had phenotypic or previous genotypic  
490 indication (based on microsatellites; Crossman et al., 2014) of 11 putative hybrids in our  
491 final data set. Assignment analysis of all samples using STRUCTURE with the number  
492 of clusters set to  $k = 2$  correctly assigned all nine Dall's porpoises to one group, and  
493 additionally assigned two of the putative hybrids to the same group with 100%  
494 probability. Five of the putative hybrids were assigned to the harbor porpoise group with

495  $\geq 99.9\%$  probability, and four samples were identified as F1 hybrids with assignment  
496 probabilities to each species group between 45% and 55%. One additional sample  
497 previously identified as a harbor porpoise was also identified as an F1 hybrid (49%/51%  
498 assignment to the 2 species groups). The admixture plot from ten combined structure  
499 analyses is presented in supplemental Figure S2a. All F1 intergeneric hybrids were from  
500 samples collected in the San Juan Islands and Oregon-Washington coast geographic  
501 strata, between latitudes 47° and 49° N. All remaining putative harbor porpoise samples  
502 were assigned to the second group with  $>99.6\%$  probability. As the targeted SNP loci  
503 were ascertained only from harbor porpoise samples, Dall's porpoise samples had  
504 unsurprisingly low diversity (Table 2), but the use of additional SNPs in microhaplotypes  
505 provided variable loci useful for species assignment and hybrid identification. Samples  
506 identified as hybrids were removed from subsequent harbor porpoise analyses.

507

508 STRUCTURE is useful for identifying population differentiation at the ESU level, where  
509 divergence is sufficient to allow high probability of assignment of samples to populations  
510 or clusters (Waples & Gaggiotti, 2006). We started with STRUCTURE in a hierarchical  
511 analysis of the SNP data to identify evolutionarily divergent populations within our  
512 harbor porpoise sample distribution. Structure analysis with all samples assigned to  
513 geographic strata *a priori* (supplemental Table S1) did not provide strong evidence of  
514 multiple divergent groups along the U.S. west coast, with the number of groups most  
515 likely  $\leq 2$  (based on  $\Delta K$  and CLUMPAK similarity score  $>0.993$  for 10/10 replicates for  
516 K1-2; supplemental Figure S2b). However, assignment probabilities among three groups  
517 differed substantially across geographic region, especially in the three southernmost  
518 geographic strata (Figure 2). Subsets of samples representing only high-quality samples  
519 ( $<80\%$  complete) and smaller geographic regions did not result in any additional  
520 evidence of population structure (supplemental Figures S2c-e).

521

522 Spatial principle component analysis (sPCA; Jombart et al., 2008) has been shown to be a  
523 useful tool for detecting patterns of genetic variability in harbor porpoise (Fontaine et al.,  
524 2017; Lah et al., 2016), and was used to explicitly combine geographic and nuclear  
525 genetic data and to investigate spatial patterns of genetic variation without assumptions of

526 Hardy-Weinberg expectations or linkage disequilibrium (Montano & Jombart, 2017).  
527 Plots of spatial principle components across the study range indicated strong evidence of  
528 structure (spca-randtest  $p < 0.001$ ) and show clear separation between samples  
529 representing the two southern-most strata, Morro Bay and Monterey Bay, along the first  
530 axis (PC1, Figure 3A, 3B), with less separation but a north-south gradient along spatial  
531 PC axes two and three (Figure 3A; supplemental Figure S3A). Additional geographic  
532 population structure is revealed by analyzing geographic subsets of the data. Hierarchical  
533 analysis of sub-regions reveals population structure within the inland waterways ( $p =$   
534  $0.004$ ; Figure 3G), and within the Washington inland waters management stock (excluding  
535 the Strait of Georgia strata samples) ( $p = 0.019$ ; supplemental Figure S3H). sPCA of  
536 samples from coastal waters (Figure 3B), excluding inland waterways in Washington and  
537 British Columbia, shows strong evidence of structure ( $p < 0.001$ ), with the first two  
538 spatial PC's highlighting unique clusters corresponding to Morro Bay (PC1), and to a  
539 lesser extent the remaining strata along PC2. Hierarchical analysis of subsets of  
540 neighboring strata indicated significant structure in analyses of all neighboring strata  
541 pairs and within the northern California/southern Oregon management stock ( $p = 0.017$ ;  
542 Figure 3D; supplemental Figure S3D). Plots of the first four individual spatial PCs for  
543 these stratification sets are in supplemental Figure S3. Replicate analyses based on the  
544 smaller number of high-quality samples (>80% complete genotypes) showed similar  
545 patterns, but resulted in non-significant p-values in some comparisons (Inland waters,  
546 southern OR/OR-WA, and northern California/southern Oregon; Supplemental Figure  
547 S3).

548

549 A Mantel test supported ( $p < 0.001$ ) isolation by distance (IBD) for nuclear SNP data  
550 along the Pacific coast (excluding inland waterways) for individual distances. A  
551 scatterplot of genetic and geographic distances (Supplemental Figure S4) shows  
552 discontinuities suggestive of differentiated populations rather than continuous clines of  
553 genetic differentiation. The Monmonier's algorithm was used to infer locations of genetic  
554 boundaries or discontinuities, and identified the major boundary between coastal  
555 Washington and western Vancouver Island at the default threshold level (Figure 4A),  
556 with lesser boundaries between Morro Bay and Monterey Bay, and Monterey and SF-RR

557 on the outer coast. Within the inland water ways, the major discontinuities were in the  
558 southern Strait of Georgia (Figure 4B). It is not straightforward to interpret the  
559 Monmonier plots because the samples are not distributed broadly in two dimensions, but  
560 the threshold arrows indicate regions where genetic distances appear to be greater than  
561 expected based on geographic distance (see Blair et al., 2012 for discussion of methods of  
562 detecting barriers to gene flow).

563

564 Genetically similar groups (Figure 3) and IBD discontinuities (Figure 4) coincided  
565 closely with several previously defined geographical strata (Figure 5; Chivers et al.,  
566 2002; Chivers et al., 2007). Genetic differentiation between adjacent *a priori* strata was  
567 tested using pairwise divergence ( $F_{ST}$ ), for both nuclear data and mtDNA (Table 3).  
568 Results are presented based on inclusion of all samples, but pairwise divergence among  
569 adjacent strata after removal of one from each pair of potential first order relatives, and  
570 after removal of samples with <50% or <80% completed genotypes, did not qualitatively  
571 alter results (supplemental Table S7). As expected when larger numbers of loci are used  
572 to calculate  $F_{ST}$  (Willing, Dreyer, & van Oosterhout, 2012), the  $F_{ST}$  point estimates did  
573 not change substantially (<0.004), but exclusion of samples reduced the sample sizes in  
574 some geographically-defined strata, resulting in loss of statistical power to detect  
575 structure among some non-adjacent strata (supplemental Table S7D-F). Along the outer  
576 coast the only genetic distances between neighboring strata pairs that were significant ( $p$   
577 < 0.05) for both mtDNA and nuclear DNA were between Morro Bay at the southern end  
578 of the distribution and Monterey Bay ( $F_{ST} = 0.020$  (nDNA); 0.310 (mtDNA)). Nuclear  
579 locus frequencies were also significantly different between three of the more southern  
580 strata pairs (Monterey/SF-RR ( $F_{ST} = 0.005$ ), N. CA/S. OR ( $F_{ST} = 0.008$ ), S. OR/ORWA  
581 coast ( $F_{ST} = 0.006$ ), while mtDNA haplotype frequencies were significantly different  
582 between only the two northern-most coastal strata pairs (ORWA coast /Spike Rock ( $F_{ST} =$   
583 0.049), and Spike Rock/W. Vancouver Is. ( $F_{ST} = 0.128$ )). Frequency plots of the common  
584 mtDNA haplotypes (excluding haplotypes that occurred in fewer than 5 samples) show  
585 clear differences among *a priori* geographical strata (Figure 5, supplemental Figure S5),  
586 with near fixation of haplotypes in Morro Bay to the south, and private haplotypes in BC  
587 to the north. Relative density data showed low-density regions along the Big Sur

588 coastline (35.5-36.5° N), just south of 38° N, and just north of 39° N at the existing  
589 management stock boundary. These low relative density areas correspond to existing  
590 boundaries between the four southern-most management stocks. The reason for low  
591 density for one region, the Big Sur coastline, is likely due to the very narrow shelf, but  
592 reasons for density variation in other regions remain unclear.

593

594 Inland waters strata were not differentiated by nuclear marker frequencies except for the  
595 Strait of Georgia, which was significantly differentiated from all inland waters and outer  
596 U.S. coast strata, but not W. Vancouver Is. or BC (Table 3). The inland waters strata were  
597 differentiated from the nearest coastal strata in mtDNA ( $F_{ST}$  between 0.023 and 0.247,  
598 with most point estimates being significant at  $p < 0.05$ ). Within the inland waters,  
599 divergence in mtDNA was significant for Neah Bay, in the outer Strait of Juan de Fuca,  
600 versus San Juan Islands ( $F_{ST} = 0.066$ ), and from both neighboring coastal strata ( $F_{ST} =$   
601 0.141 (Spike Rock); 0.247 (W. Vancouver Island)). Puget Sound, which has been re-  
602 colonized in the last 18 years (Evenson et al., 2016), was not significantly differentiated  
603 from either the neighboring San Juan Islands, or Neah Bay in the Strait of Juan de Fuca.  
604 The genetic discontinuity identified immediately north of the U.S.-Canadian border at  
605 approximately 49.1°N. latitude (Figure 4) warranted an adjustment of the boundary  
606 between the Strait of Georgia and San Juan Islands strata from 48.8° to 49.1°, resulting in  
607 reassignment of ten and five samples (mtDNA, nDNA analyses, respectively) from the  
608 Strait of Georgia to the San Juan Islands strata (Figure 4B). Genetic divergence remained  
609 significant following the reassignment of these samples and  $F_{ST}$  increased for both nDNA  
610 (0.005 to 0.006) and mtDNA (0.095 to 0.153).

611

612 For the genotype-environment association analyses (dbRDA), we analyzed three data sets  
613 for each region (i.e., outer coastal and inland waterways), representing three thresholds of  
614 missing genotype data across individuals: 25%, 20%, and 15%. We first imputed the data  
615 sets with *snmf* using the optimized settings (identified by minimizing the average cross  
616 entropy, with identical optimized parameters for all data sets):  $\alpha=10$ , and  $K=1$   
617 (supplemental Table S8). We modified the default genomic inflation factors to produce p-  
618 value distributions that better met the uniform distribution assumption (per Francois et

619 al., 2016, p-value histograms provided in supplemental Figures S6-S11), providing a  
620 balance between true positive detections and false positives driven by population  
621 structure and other potential confounders. We did not include other corrections for  
622 confounding factors in the GEA since population structure is low across the study area  
623 (e.g., Table 3), and dbRDA is robust to isolation-by-distance demographic scenarios  
624 (Forester et al., 2018) such as this one. Environmental predictor variables within each  
625 region were not highly correlated (i.e., all pairwise correlations were less than  $|0.8|$  and all  
626 variance inflation factors, which measure multicollinearity, were less than five, where  
627 values  $>10$  can be problematic), and none of the retained predictors were highly  
628 correlated with either latitude or longitude (supplemental Tables S9, S10). For the outer  
629 strata and inland waterways, we identified as candidates those microhaplotypes that were  
630 detected in at least two of the three thresholded data sets. This produced 22 candidates for  
631 the outer coastal region and six candidates for the inland waterways region. Candidate  
632 microhaplotypes for the outer coastal region showed the strongest relationships with the  
633 mean and standard deviation of daily sea surface temperature (sst.mn and sst.SD), and the  
634 mean of daily sea surface height (ssh.mn). Triplots of the dbRDA results illustrated  
635 relationships among individual's multilocus genotypes, the candidate microhaplotypes,  
636 and environmental predictors. For example, the southernmost Morro Bay individuals,  
637 located in the ordination space as a function of their genotypes at the candidate loci,  
638 showed strong relationships with increasing ssh.mn and sst.mn and decreasing ssh.SD,  
639 indicating potential adaptation to local environmental conditions (Figure 6A). The  
640 smaller inland waterways data set showed some differentiation between the Neah Bay  
641 and San Juan Islands individuals based on sea surface temperature, with the Puget Sound  
642 individuals showing no relationships, possibly due to their most recent re-colonization  
643 (Figure 6B). Because loci under selection could affect population structure, we confirmed  
644 that identification of divergence between adjacent a priori strata were consistent with and  
645 without the loci identified by dbRDA analysis (supplemental Table S7). Full dbRDA  
646 triplots and microhaplotype biplots are provided in supplemental Figures S12-S15 for the  
647 outer coastal data and supplemental Figures S16-S19 for the inner waterways.  
648  
649

650 **Discussion**

651

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

663

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

675

676 Previous evidence indicated intrageneric hybridization between the more pelagic Dall's  
677 porpoise and the coastal harbor porpoise, based on morphology and genetic analysis of  
678 eight microsatellite loci (Crossman et al., 2014). While the previous genetic analyses  
679 suggested bi-directional hybridization and introgression, the more powerful dataset used  
680 here indicates that many of the putative hybrids (based on genetics) could be assigned

681 with >98% probability to one or the other species (N=5 *P. phocoena*; N=2 *P. dalli*), or as  
682 F1 hybrids (N=5) between male harbor porpoise and female Dall's porpoise. Sightings of  
683 putative hybrids in the wild have also all been associated with Dall's porpoises (Baird et  
684 al., 1998). Harbor porpoise males have disproportionately large testes and have been  
685 observed to exhibit forceful and fast ambush mating (Keener, Webber, Szczepaniak,  
686 Markowitz, & Orbach, 2018), suggesting a greater reliance on sperm competition rather  
687 than mate choice in male harbor porpoise. This mating strategy might result in occasional  
688 intergeneric hybrids with Dall's porpoise in areas where their ranges overlap (Baird et al.,  
689 1998). Although the lack of individuals exhibiting intermediate probabilities of  
690 assignment to both species (between 50% and 100%) suggests hybrids are usually  
691 infertile, there have been sightings of putative hybrid females with neonatal calves  
692 (Willis et al., 2004).

693

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

706

707 Although the Morro Bay population is designated as a separate management stock based  
708 on a hiatus in distribution at the Big Sur coast (Carretta et al., 2019), our results provide  
709 the first evidence for genetic differences. The most common mtDNA haplotype (CR01)  
710 was found in every *a priori* geographic group except Morro Bay, and the most common  
711 haplotype in Morro Bay (CR02) was found in 86% of the Morro Bay samples, but less



712 than 33% of any other population (range 0 – 33%). Two (CR30, CR42) of the three other  
713 haplotypes found in Morro Bay were unique to that population, and differed from CR02  
714 by only one nucleotide change. The third haplotype (CR03) was distributed across most  
715 of the range, but represents a common haplotype that is more similar to CR02 (3bp  
716 different) than to the cluster including CR01 (7bp different) and the majority of the other  
717 haplotypes distributed throughout the rest of the range (see haplotype network,  
718 supplemental figure S5). This suggests long-term isolation of the southernmost  
719 population, as well as a persistently small population size, or severe or recurrent  
720 bottlenecks. Annual estimates of tens to hundreds of harbor porpoises were killed in  
721 gillnet entanglements between approximately 1960 and 2001, resulting in a reduction of  
722 the population by 30-97% (Barlow & Hanan, 1995). The population has since begun  
723 recovering, from a low of 560 in 1990 to 4255 in 2012 (Carretta et al., 2019; Forney et  
724 al., in press). Our results suggest that recovery has been due to internal recruitment rather  
725 than migration from the larger Monterey Bay population to the north.

726

727 The northernmost strata were the most differentiated, with high mtDNA divergence  
728 metrics and the highest frequencies of the most common control region haplotype (CR01;  
729 66% W. Vancouver Is., 63% Strait of Georgia). The frequency of this haplotype ranged  
730 from zero to 50% in other strata. The BC stratum (excluding Strait of Georgia samples)  
731 spanned a large geographic range with few samples (N=5), and included two private  
732 mtDNA haplotypes, both found in two samples. These suggest different haplotype  
733 composition in this region, but we cannot rule out sampling of related individuals or  
734 isolated regional populations, as the sample size is small and shared haplotypes were  
735 found in samples that were collected close in space and time. Combined, these data  
736 suggest that there is a haplotype gradient between the northern and southern portions of  
737 the range, possibly representing interglacial range expansion (Taguchi et al., 2010) or the  
738 presence of a historical phylogeographic separation that has more recently re-connected,  
739 as has been suggested for populations in the eastern Atlantic Ocean (Fontaine, Baird, et  
740 al., 2007; Fontaine et al., 2014).

741

742 Harbor porpoise habitat in the eastern North Pacific includes the California Current  
743 System (CCS), recognized as one of the most productive marine ecosystems on the  
744 planet, but spans topographically, oceanographically and temporally complex regions  
745 from temperate to Arctic waters. Aside from preference for coastal waters less than 200  
746 m deep, little is known about factors influencing suitable habitat for Pacific harbor  
747 porpoises, which vary in population density and abundance along the west coast and  
748 inland waters of North America. The remotely sensed and modeled oceanographic  
749 variables available throughout most of this range (upwelling index, mixed-layer depth,  
750 sea surface temperature and sea surface height) are all proxies for habitat variability with  
751 complex relationships to wind, currents, topology, isocline depth, productivity, fresh-  
752 water input and seasonal and interannual effects (Castelao & Luo, 2018; Hickey et al.,  
753 2016; Venegas et al., 2008). Our genotype-environment association analyses provide  
754 initial evidence for local adaptation to environmental variability across the study range.  
755 Given the limited scale of genomic sampling in this study these findings should be seen  
756 as preliminary, providing a basis for future investigation of local adaptation across the  
757 complex seascape environment inhabited by harbor porpoise (see, e.g., Fontaine et al.,  
758 2017; Fontaine, Tolley, et al., 2007; Nielsen et al., 2018).

759

760 An overall correlation with latitude, which might be expected across a large latitudinal  
761 range, appears to strengthen in the regions south of Oregon. In central Oregon, there is an  
762 oceanographic shift around Cape Blanco, where the coastal upwelling jet separates from  
763 the coast (~15-30 km offshore), becoming an oceanic jet (>100 km offshore) (Castelao &  
764 Luo, 2018). Populations in the northern half of the coastal range are associated with  
765 several environmental variables (e.g., mean and SD of the mixed layer depth, and high  
766 variation in both sst and ssh), while those in the central and southern portions are  
767 associated with others (e.g., high mean and low variation in ssh) (Figure 6). There is also  
768 evidence of population structure across that region, with significant nuclear genetic  
769 differentiation among populations from southern Washington to northern California  
770 (Table 3). There was a particularly strong correlation between increased mean ssh and  
771 decreased ssh variability in the southernmost Morro Bay population, which is also the  
772 most genetically distinct, suggesting possible local adaptation linked to oceanographic

773 processes (e.g., thermocline and/or upwelling). While these links are indirect, they  
774 provide the first evidence of local adaptation in addition to demographic independence  
775 among harbor porpoise populations along the west coast of North America. The Morro  
776 Bay stock was previously recognized based only on a distribution hiatus and evidence of  
777 historical population decline due to extensive fishery bycatch (Barlow & Hanan, 1995),  
778 as well as blubber pollutant ratios that differed from other areas (Calambokidis &  
779 Barlow, 1991). It is potentially subject to offshore energy production disturbance (Forney  
780 et al., 2017), and is also at the southern edge of the species' geographic range, and most  
781 likely to be impacted by climate change (Learmonth et al., 2006; Ruiz-Cooley et al.,  
782 2017). Our results provide strong support for continued management of this population as  
783 a separate management stock, especially in light of potential impacts of coastal  
784 development and climate change.

785

786 Harbor porpoise populations along the west coast of North America have historically  
787 experienced substantial fisheries bycatch in portions of the range (Barlow & Hanan,  
788 1995), and multiple lines of evidence indicate that there is limited dispersal among  
789 regions. Within the U.S., this led to identification of five coastal management stocks  
790 between southern California and the northern U.S. border (coinciding with four of the  
791 geographic strata in this study, plus the combined S. Oregon and N. California strata),  
792 and one stock in the inland waterways in Washington (Carretta et al., 2019), with stock  
793 boundaries placed to coincide with areas of lower density. The genetic results presented  
794 here provide additional evidence of limited movement among the currently defined  
795 management stocks, as well as between some geographically defined strata within  
796 existing stocks, consistent with previous studies of genetics and contaminant levels  
797 (Calambokidis & Barlow, 1991; Chivers et al., 2002), and limited movements of tagged  
798 porpoises in the inland waterways of Washington (Hanson, 2007). In particular, sPCA  
799 and  $F_{ST}$  analyses of nuclear data show separation of the N. California and S. Oregon  
800 strata (sPCA randtest  $p = 0.017$ ) with a small but significant divergence ( $F_{ST}=0.008$ ) in  
801 nuclear allele frequencies (Table 3, Figure 3D, supplemental Figure S3D). We suggest  
802 that the lack of statistically significant difference in mtDNA is likely a result of low  
803 statistical power due to small sample size in N. California ( $n=11$ ) rather than results that

804 contradict the finding of differences between these strata based on nuclear markers.  
805 Placement of a potential stock boundary as indicated in Figure 5 is suggested by the drop  
806 in density in a region where there is a gap in our sample distribution, though the location  
807 of the density minimum in Figure 5 varies through time (supplemental Figure S20), either  
808 due to sampling differences among surveys or small north-south shifts of populations  
809 across years.

810

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

827

828

## 829 **Acknowledgements**

830 We are grateful Andy Foote and Marie Louise for help with SNP discovery methods;  
831 Diana Baetscher, Cassie Columbus, Carlos Garza and the SWFSC Santa Cruz genetics  
832 lab for help with GT-seq optimizations, and Nathan Campbell and Ninh Vu for help with  
833 GT-seq method implementation and optimization; Eric Anderson, Thomas Ng, Anthony  
834 Clemento and Neil Thompson for help with MicrohaPlot software and methods; Philipp

835 Kraemer for help with the Demerelate R package; Steven Head and John Shimashita,  
836 TSRI Sequencing core for sequencing assistance; David Anderson, Cascadia Research  
837 Collective for help with Washington Inland Waterways data and observations; Samuel  
838 Woodman for help with analysis of porpoise relative density data. Samples were  
839 generously provided by Lisa Ballance, Barbara Taylor, Jay Barlow at the SWFSC; Jan  
840 Straley, University of Alaska Southeast; Marilyn Dahlheim, NOAA National Marine  
841 Mammal Lab; Jorge Velez-Juarbe, Los Angeles County Museum of Natural History;  
842 Merrill Gosho, Bill Walker, Jeff Foster, Stephen Clausen, The Whale Museum, Olympic  
843 Coast National Marine Sanctuary and many contributors through the Marine Mammal  
844 Stranding Networks. Samples were collected under various Marine Mammal Protection  
845 Act permits, and stored in the SWFSC Marine Mammal and Sea Turtle Research  
846 (MMASTR) collection. Funding support for this study was provided by NOAA Office of  
847 Protected Resources. We are grateful to Aimee Lang and 3 anonymous reviewers for  
848 helpful comments on previous versions of this manuscript.

849

## 850 **References**

- 851 Ahrens, C. W., Rymer, P. D., Stow, A., Bragg, J., Dillon, S., Umbers, K. D. L., &  
852 Dudaniec, R. Y. (2018). The search for loci under selection: trends, biases and  
853 progress. *Molecular Ecology*, 27(6), 1342-1356. doi:10.1111/mec.14549
- 854 Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016).  
855 Harnessing the power of RADseq for ecological and evolutionary genomics.  
856 *Nature Reviews Genetics*, 17(2), 81-92. doi:10.1038/nrg.2015.28
- 857 Archer, F. I., Adams, P. E., & Schneiders, B. B. (2017). STRATAG: An R package for  
858 manipulating, summarizing and analysing population genetic data. *Molecular*  
859 *Ecology Resources*, 17, 5-11. doi:10.1111/1755-0998.12559
- 860 Baetscher, D. S., Clemento, A. J., Ng, T., Anderson, E. C., & Garza, J. C. (2017).  
861 Microhaplotypes provide increased power from short-read DNA sequences for  
862 relationship inference. *Molecular Ecology Resources*, 18(2), 296-305.  
863 doi:10.1111/1755-0998.12737
- 864 Baird, R. W., Willis, P. M., Guenther, T. J., Wilson, P. J., & White, B. N. (1998). An  
865 intergeneric hybrid in the family Phocoenidae. *Canadian Journal of Zoology-*  
866 *Revue Canadienne De Zoologie*, 76(1), 198-204. doi:10.1139/cjz-76-1-198

- 867 Barlow, J., & Hanan, D. (1995). An assessment of the status of the harbour porpoise in  
868 central California. In A. Bjørge & G. P. Donovan (Eds.), *Biology of the*  
869 *Phocoenids* (pp. 123-140). Cambridge: Report of the International Whaling  
870 Commission, Special Issue 16.
- 871 Ben Chehida, Y., Thumloup, J., Schumacher, C., Harkins, T., Aguilar, A., Borrel, A., . . .  
872 Fontaine, M. C. (2020). Evolutionary history of the porpoise family  
873 (Phocoenidae) across the speciation continuum: a mitogenome phylogeographic  
874 perspective. *Scientific Reports*, *10*(BioRxiv doi 10.1101/851469), 15190.  
875 doi:10.1101/851469
- 876 Blair, C., Weigel, D. E., Balazik, M., Keeley, A. T., Walker, F. M., Landguth, E., . . .  
877 Balkenhol, N. (2012). A simulation-based evaluation of methods for inferring  
878 linear barriers to gene flow. *Molecular Ecology Resources*, *12*(5), 822-833.  
879 doi:10.1111/j.1755-0998.2012.03151.x
- 880 Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for  
881 Illumina sequence data. *Bioinformatics*, *30*(15), 2114-2120.  
882 doi:10.1093/bioinformatics/btu170
- 883 Bray, J. R., & Curtis, J. T. (1957). An Ordination of the Upland Forest Communities of  
884 Southern Wisconsin. *Ecological Monographs*, *27*(4), 326-349.
- 885 Calambokidis, J., & Barlow, J. (1991). Chlorinated hydrocarbon concentrations and their  
886 use for describing population discreteness in harbor porpoises from Washington,  
887 Oregon, and California. In J. E. Reynolds III & D. K. Odell (Eds.), *Marine*  
888 *Mammals Strandings in the United States* (Vol. 98, pp. 101-110): NOAA Tech.  
889 Rep. NMFS.
- 890 Calambokidis, J., Laake, J. L., & Osmeck, S. D. (1997). *Aerial surveys for marine*  
891 *mammals in Washington and British Columbia inside waters. Final report to the*  
892 *National Marine Mammal Laboratory, Seattle, WA, USA*. Retrieved from  
893 [https://www.cascadiaresearch.org/publications/aerial-surveys-marine-mammals-](https://www.cascadiaresearch.org/publications/aerial-surveys-marine-mammals-washington-and-british-columbia-inside-waters)  
894 [washington-and-british-columbia-inside-waters](https://www.cascadiaresearch.org/publications/aerial-surveys-marine-mammals-washington-and-british-columbia-inside-waters)
- 895 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., &  
896 Madden, T. L. (2009). BLAST+: architecture and applications. *BMC*  
897 *Bioinformatics*, *10*, 421. doi:10.1186/1471-2105-10-421
- 898 Campbell, N. R., Harmon, S. A., & Narum, S. R. (2015). Genotyping-in-Thousands by  
899 sequencing (GT-seq): A cost effective SNP genotyping method based on custom  
900 amplicon sequencing. *Molecular Ecology Resources*, *15*(4), 855-867.  
901 doi:10.1111/1755-0998.12357

- 902 Candy, J. R., Campbell, N. R., Grinnell, M. H., Beacham, T. D., Larson, W. A., &  
903 Narum, S. R. (2015). Population differentiation determined from putative neutral  
904 and divergent adaptive genetic markers in Eulachon (*Thaleichthys pacificus*,  
905 Osmeridae), an anadromous Pacific smelt. *Molecular Ecology Resources*, 15(6),  
906 1421-1434. doi:10.1111/1755-0998.12400
- 907 Capblancq, T., Luu, K., Blum, M. G. B., & Bazin, E. (2018). Evaluation of redundancy  
908 analysis to identify signatures of local adaptation. *Molecular Ecology Resources*,  
909 18(6), 1223-1233. doi:10.1111/1755-0998.12906
- 910 Carretta, J. V., Forney, K. A., Oleson, E. M., Weller, D. W., Lang, A. R., Baker, J., . . .  
911 Brownell Jr., R. L. (2019). *U.S. Pacific marine mammal stock assessments: 2018*.  
912 (NOAA Technical Memorandum NMFS-SWFSC-617). U.S. Department of  
913 Commerce.
- 914 Carstensen, J., Henriksen, O. D., & Teilmann, J. (2006). Impacts of offshore wind farm  
915 construction on harbour porpoises: acoustic monitoring of echolocation activity  
916 using porpoise detectors (T-PODs). *Marine Ecology Progress Series*, 321, 295-  
917 308. doi:10.3354/meps321295
- 918 Castelao, R. M., & Luo, H. (2018). Upwelling jet separation in the California Current  
919 System. *Scientific Reports*, 8. doi:10.1038/s41598-018-34401-y
- 920 Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ  
921 preprocessor. *BioRxiv*, 274100. doi:doi.org/10.1101/274100
- 922 Chin, T. M., Vazquez-Cuervo, J., & Armstrong, E. M. (2017). A multi-scale high-  
923 resolution analysis of global sea surface temperature. *Remote Sensing of*  
924 *Environment*, 200, 154-169. doi:10.1016/j.rse.2017.07.029
- 925 Chivers, S. J., Dizon, A. E., Gearin, P. J., & Robertson, K. M. (2002). Small-scale  
926 population structure of eastern North Pacific harbour porpoises (*Phocoena*  
927 *phocoena*) indicated by molecular genetic analyses. *Journal of Cetacean*  
928 *Research and Management*, 4(2), 111-122.
- 929 Chivers, S. J., Hanson, B., Laake, J., Gearin, P. J., Muto, M. M., Calambokidis, J., . . .  
930 Hancock, B. L. (2007). *Additional genetic evidence for population structure of*  
931 *Phocoena phocoena off the coasts of California, Oregon, and Washington*.  
932 (Administrative Report LJ-07-08). Southwest Fisheries Science Center: National  
933 Marine Fisheries Service, Administrative Report LJ-07-08.
- 934 Crossman, C. A., Barrett-Lennard, L. G., & Taylor, E. B. (2014). Population structure  
935 and intergeneric hybridization in harbour porpoises *Phocoena phocoena* in British

- 936 Columbia, Canada. *Endangered Species Research*, 26(1), 1-12.  
937 doi:10.3354/esr00624
- 938 Cullingham, C. I., Miller, J. M., Peery, R. M., Dupuis, J. R., Malenfant, R. M., Gorrell, J.  
939 C., & Janes, J. K. (2020). Confidently identifying the correct K value using the  
940 DeltaK method: when does K = 2? *Molecular Ecology*. doi:10.1111/mec.15374
- 941 Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., . . .  
942 Genomes Project Analysis Group. (2011). The variant call format and VCFtools.  
943 *Bioinformatics*, 27(15), 2156-2158. doi:10.1093/bioinformatics/btr330
- 944 DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., . . .  
945 Daly, M. J. (2011). A framework for variation discovery and genotyping using  
946 next-generation DNA sequencing data. *Nature Genetics*, 43(5), 491-498.  
947 doi:10.1038/ng.806
- 948 Emerson, K. J., Merz, C. R., Catchen, J. M., Hohenlohe, P. A., Cresko, W. A., Bradshaw,  
949 W. E., & Holzapfel, C. M. (2010). Resolving postglacial phylogeography using  
950 high-throughput sequencing. *Proceedings of the National Academy of Science*  
951 *USA*, 107(37), 16196-16200. doi:10.1073/pnas.1006538107
- 952 Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of  
953 individuals using the software STRUCTURE: a simulation study. *Molecular*  
954 *Ecology*, 14(8), 2611-2620. doi:10.1111/j.1365-294X.2005.02553.x
- 955 Evenson, J. R., Anderson, D., Murphie, B. L., Cyra, T. A., & Calambokidis, J. (2016).  
956 *Disappearance and return of harbor porpoise to Puget Sound: 20 year pattern*  
957 *revealed from winter aerial surveys*. Olympia, WA: Washington Department of  
958 Fish and Wildlife, Wildlife Program and Cascadia Research Collective.
- 959 Fontaine, M. C. (2016). Harbour porpoises, *Phocoena phocoena*, in the Mediterranean  
960 Sea and adjacent regions: Biogeographic relicts of the last glacial period.  
961 *Advances in Marine Biology*, 75, 333-358. doi:10.1016/bs.amb.2016.08.006
- 962 Fontaine, M. C., Baird, S. J., Piry, S., Ray, N., Tolley, K. A., Duke, S., . . . Michaux, J. R.  
963 (2007). Rise of oceanographic barriers in continuous populations of a cetacean:  
964 the genetic structure of harbour porpoises in Old World waters. *BMC Biology*, 5,  
965 30. doi:10.1186/1741-7007-5-30
- 966 Fontaine, M. C., Roland, K., Calves, I., Austerlitz, F., Palstra, F. P., Tolley, K. A., . . .  
967 Aguilar, A. (2014). Postglacial climate changes and rise of three ecotypes of  
968 harbour porpoises, *Phocoena phocoena*, in western Palearctic waters. *Molecular*  
969 *Ecology*, 23(13), 3306-3321. doi:10.1111/mec.12817



- 970 Fontaine, M. C., Snirc, A., Frantzis, A., Koutrakis, E., Ozturk, B., Ozturk, A. A., &  
971 Austerlitz, F. (2012). History of expansion and anthropogenic collapse in a top  
972 marine predator of the Black Sea estimated from genetic data. *Proceedings of the*  
973 *National Academy of Science USA*, 109(38), E2569-2576.  
974 doi:10.1073/pnas.1201258109
- 975 Fontaine, M. C., Thatcher, O., Ray, N., Piry, S., Brownlow, A., Davison, N. J., . . .  
976 Goodman, S. J. (2017). Mixing of porpoise ecotypes in southwestern UK waters  
977 revealed by genetic profiling. *R Soc Open Sci*, 4(3), 160992.  
978 doi:10.1098/rsos.160992
- 979 Fontaine, M. C., Tolley, K. A., Siebert, U., Gobert, S., Lepoint, G., Bouquegneau, J. M.,  
980 & Das, K. (2007). Long-term feeding ecology and habitat use in harbour  
981 porpoises *Phocoena phocoena* from Scandinavian waters inferred from trace  
982 elements and stable isotopes. *BMC Ecology*, 7, 1. doi:10.1186/1472-6785-7-1
- 983 Foote, A. D., Liu, Y., Thomas, G. W., Vinar, T., Alfoldi, J., Deng, J., . . . Gibbs, R. A.  
984 (2015). Convergent evolution of the genomes of marine mammals. *Nature*  
985 *Genetics*, 47(3), 272-275. doi:10.1038/ng.3198
- 986 Foote, A. D., Martin, M. D., Louis, M., Pacheco, G., Robertson, K. M., Sinding, M.-H.  
987 S., . . . Morin, P. A. (2019). Killer whale genomes reveal a complex history of  
988 recurrent admixture and vicariance. *Molecular Ecology*, 38, 3427-3444.  
989 doi:10.1111/mec.15099
- 990 Forester, B. R., Lasky, J. R., Wagner, H. H., & Urban, D. L. (2018). Comparing methods  
991 for detecting multilocus adaptation with multivariate genotype-environment  
992 associations. *Molecular Ecology*, 27(9), 2215-2233. doi:10.1111/mec.14584
- 993 Forney, K. A., Hanan, D. A., & Barlow, J. (1991). Detecting trends in harbor porpoise  
994 abundance from aerial surveys using analysis of covariance. *Fishery Bulletin*,  
995 89(3), 367-377.
- 996 Forney, K. A., Moore, J. E., Barlow, J., Carretta, J. V., & Benson, S. R. (in press). A  
997 multi-decadal Bayesian trend analysis of harbor porpoise (*Phocoena phocoena*)  
998 populations off California relative to past fisher bycatch. *Marine Mammal*  
999 *Science*.
- 1000 Forney, K. A., Southall, B. L., Slooten, E., Dawson, S., Read, A. J., Baird, R. W., &  
1001 Brownell, R. L. (2017). Nowhere to go: noise impact assessments for marine  
1002 mammal populations with high site fidelity. *Endangered Species Research*, 32,  
1003 391-413. doi:10.3354/esr00820

- 1004 Francois, O., Martins, H., Caye, K., & Schoville, S. D. (2016). Controlling false  
 1005 discoveries in genome scans for selection. *Molecular Ecology*, 25(2), 454-469.  
 1006 doi:10.1111/mec.13513
- 1007 Frichot, E., & Francois, O. (2015). LEA: An R package for landscape and ecological  
 1008 association studies. *Methods in Ecology and Evolution*, 6(8), 925-929.  
 1009 doi:10.1111/2041-210x.12382
- 1010 Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read  
 1011 sequencing. *arXiv preprint arXiv:1207.3907 [q-bio.GN]*.
- 1012 Gemmell, N. J., & Akiyama, S. (1996). An efficient method for the extraction of DNA  
 1013 from vertebrate tissues. *Trends Genet*, 12(9), 338-339.
- 1014 Hancock-Hanser, B., Frey, A., Leslie, M., Dutton, P. H., Archer, E. I., & Morin, P. A.  
 1015 (2013). Targeted multiplex next-generation sequencing: Advances in techniques  
 1016 of mitochondrial and nuclear DNA sequencing for population genomics.  
 1017 *Molecular Ecology Resources*, 13, 254-268. doi:10.1111/1755-0998.12059
- 1018 Hanson, M. B. (2007). Using location data from telemetry tagged marine mammals to  
 1019 improve stock assessments. In P. Sheridan, J. W. Ferguson, & S. L. Downing  
 1020 (Eds.), *Report of the National Marine Fisheries Service Workshop on Advancing*  
 1021 *Electronic Tag Technology and Their Use in Stock Assessments*: U.S. Dept.  
 1022 Commerce, NOAA Tech. Memo. NMFSF/SPO-82, 82.
- 1023 Hickey, B., Geier, S., Kachel, N., Ramp, S., Kosro, P. M., & Connolly, T. (2016).  
 1024 Alongcoast structure and interannual variability of seasonal midshelf water  
 1025 properties and velocity in the Northern California Current System. *Journal of*  
 1026 *Geophysical Research-Oceans*, 121(10), 7408-7430. doi:10.1002/2015jc011424
- 1027 Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian*  
 1028 *Journal of Statistics*, 6(2), 65-70.
- 1029 Hubisz, M. J., Falush, D., Stephens, M., & Pritchard, J. K. (2009). Inferring weak  
 1030 population structure with the assistance of sample group information. *Molecular*  
 1031 *Ecology Resources*, 9(5), 1322-1332. doi:10.1111/j.1755-0998.2009.02591.x
- 1032 Jacox, M. G., Edwards, C. A., Hazen, E. L., & Bograd, S. J. (2018). Coastal upwelling  
 1033 revisited: Ekman, Bakun, and improved upwelling indices for the US west coast.  
 1034 *Journal of Geophysical Research-Oceans*, 123(10), 7332-7350.  
 1035 doi:10.1029/2018jc014187

- 1036 Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and  
 1037 permutation program for dealing with label switching and multimodality in  
 1038 analysis of population structure. *Bioinformatics*, 23(14), 1801-1806.  
 1039 doi:10.1093/bioinformatics/btm233
- 1040 Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic  
 1041 markers. *Bioinformatics*, 24(11), 1403-1405. doi:10.1093/bioinformatics/btn129
- 1042 Jombart, T., Devillard, S., Dufour, A. B., & Pontier, D. (2008). Revealing cryptic spatial  
 1043 patterns in genetic variability by a new multivariate method. *Heredity (Edinb)*,  
 1044 101(1), 92-103. doi:10.1038/hdy.2008.34
- 1045 Kalendar, R., Lee, D., & Schulman, A. H. (2011). Java web tools for PCR, in silico PCR,  
 1046 and oligonucleotide assembly and analysis. *Genomics*, 98(2), 137-144.  
 1047 doi:10.1016/j.ygeno.2011.04.009
- 1048 Kalinowski, S. T. (2011). The computer program STRUCTURE does not reliably  
 1049 identify the main genetic clusters within species: simulations and implications for  
 1050 human population structure. *Heredity (Edinb)*, 106(4), 625-632.  
 1051 doi:10.1038/hdy.2010.95
- 1052 Keener, W., Webber, M. A., Szczepaniak, I. D., Markowitz, T. M., & Orbach, D. N.  
 1053 (2018). The sex life of harbor porpoises (*Phocoena phocoena*): Lateralized and  
 1054 aerial behavior. *Aquatic Mammals*, 44(6), 620-632.  
 1055 doi:10.1578/Am.44.6.2018.620
- 1056 Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015).  
 1057 Clumpak: a program for identifying clustering modes and packaging population  
 1058 structure inferences across K. *Molecular Ecology Resources*, 15(5), 1179-1191.  
 1059 doi:10.1111/1755-0998.12387
- 1060 Kraemer, P., & Gerlach, G. (2017). Demerelate: calculating interindividual relatedness  
 1061 for kinship analysis based on codominant diploid genetic markers using R.  
 1062 *Molecular Ecology Resources*, 17(6), 1371-1377. doi:10.1111/1755-0998.12666
- 1063 Lah, L., Trense, D., Benke, H., Berggren, P., Gunnlaugsson, T., Lockyer, C., . . .  
 1064 Tiedemann, R. (2016). Spatially explicit analysis of genome-wide snps detects  
 1065 subtle population structure in a mobile marine mammal, the harbor porpoise.  
 1066 *PLoS One*, 11(10), e0162792. doi:10.1371/journal.pone.0162792
- 1067 Learmonth, J. A., MacLeod, C. D., Santos, M. B., Pierce, G. J., Crick, H. Q. P., &  
 1068 Robinson, R. A. (2006). Potential effects of climate change on marine mammals.  
 1069 *Oceanography and Marine Biology - an Annual Review*, 44, 431-464.  
 1070 doi:10.1201/9781420006391

- 1071 Leigh, J. W., & Bryant, D. (2015). POPART: fullfeature software for haplotype network  
1072 construction. *Methods in Ecology and Evolution*, 6(9), 1110-1116.
- 1073 Leslie, M. S., & Morin, P. A. (2016). Using genome-wide SNPs to detect structure in  
1074 high-diversity and low-divergence populations of severely impacted eastern  
1075 tropical Pacific spinner (*Stenella longirostris*) and pantropical spotted dolphins (*S.*  
1076 *Attenuata*). *Frontiers in Marine Science*, 3, 253. doi:10.3389/fmars.2016.00253
- 1077 Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-  
1078 Wheeler transform. *Bioinformatics*, 25(14), 1754-1760.  
1079 doi:10.1093/bioinformatics/btp324
- 1080 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Genome Project  
1081 Data, P. (2009). The Sequence alignment/map format and SAMtools.  
1082 *Bioinformatics*, 25(16), 2078-2079. doi:10.1093/bioinformatics/btp352
- 1083 Maisano Delsler, P., Corrigan, S., Hale, M., Li, C., Veuille, M., Planes, S., . . . Mona, S.  
1084 (2016). Population genomics of *C. melanopterus* using target gene capture data:  
1085 demographic inferences and conservation perspectives. *Science Reports*, 6, 33753.  
1086 doi:10.1038/srep33753
- 1087 Martien, K. K., Lang, A. R., Taylor, B. L., Rosel, P. E., Simmons, E., Oleson, E. M., . . .  
1088 Hanson, M. B. (2019). The dip delineation handbook: A guide to using multiple  
1089 lines of evidence to delineate demographically independent populations of marine  
1090 mammals. *NOAA Technical Memorandum, NOAA-TM-NMFS-SWFSC-622*,  
1091 <https://repository.library.noaa.gov>.
- 1092 McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., . . .  
1093 DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework  
1094 for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9),  
1095 1297-1303. doi:10.1101/gr.107524.110
- 1096 McKinney, G. J., Seeb, J. E., & Seeb, L. W. (2017). Managing mixed-stock fisheries:  
1097 genotyping multi-SNP haplotypes increases power for genetic stock  
1098 identification. *Canadian Journal of Fisheries and Aquatic Sciences*, 74(4), 429-  
1099 434. doi:10.1139/cjfas-2016-0443
- 1100 Miller, S. A., Dykes, D. D., & Polesky, H. F. (1988). A simple salting out procedure for  
1101 extracting DNA from human nucleated cells. *Nucleic Acids Research*, 16(3),  
1102 1215. doi:10.1093/nar/16.3.1215
- 1103 Monmonier, M. (1973). Maximum-difference barriers: an alternative numerical  
1104 regionalization method. *Geographical Analysis*, 3, 245-261.

- 1105 Montano, V., & Jombart, T. (2017). An Eigenvalue test for spatial principal component  
1106 analysis. *BMC Bioinformatics*, 18(1), 562. doi:10.1186/s12859-017-1988-y
- 1107 Moore, A. M., Arango, H. G., Broquet, G., Edwards, C., Veneziani, M., Powell, B., . . .  
1108 Robinson, P. (2011). The Regional Ocean Modeling System (ROMS) 4-  
1109 dimensional variational data assimilation systems Part II - Performance and  
1110 application to the California Current System. *Progress in Oceanography*, 91(1),  
1111 50-73. doi:10.1016/j.pocean.2011.05.003
- 1112 Morin, P. A., Foote, A. D., Hill, C. M., Simon-Bouhet, B., Lang, A. R., & Louise, M.  
1113 (2018). SNP discovery from single and multiplex genome assemblies of non-  
1114 model organisms. In S. R. Head, P. Ordoukhanian, & D. Salomon (Eds.), *Next-  
1115 Generation Sequencing. Methods in Molecular Biology* (Vol. 1712, pp. 113-144):  
1116 Humana Press.
- 1117 Moritz, C. (1994). Defining 'Evolutionarily Significant Units' for conservation. *Trends in  
1118 Ecology and Evolution*, 9(10), 373-375. doi:10.1016/0169-5347(94)90057-4
- 1119 Nielsen, N. H., Teilmann, J., Sveegaard, S., Hansen, R. G., Sinding, M. H. S., Dietz, R.,  
1120 & Heide-Jorgensen, M. P. (2018). Oceanic movements, site fidelity and deep  
1121 diving in harbour porpoises from Greenland show limited similarities to animals  
1122 from the North Sea. *Marine Ecology Progress Series*, 597, 259-272.  
1123 doi:10.3354/meps12588
- 1124 Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure  
1125 using multilocus genotype data. *Genetics*, 155(2), 945-959.
- 1126 Read, A. J. (1999). Harbour porpoise (*Phocoena phocoena*). In S. Ridgway & R.  
1127 Harrison (Eds.), *Handbook of Marine Mammals* (pp. 323-350). London, San  
1128 Diego: Academic Press.
- 1129 Reeves, R. R., McClellan, K., & Werner, T. B. (2013). Marine mammal bycatch in gillnet  
1130 and other entangling net fisheries, 1990 to 2011. *Endangered Species Research*,  
1131 20, 71-97. doi:10.3354/esr00481
- 1132 Riginos, C., Crandall, E. D., Liggins, L., Bongaerts, P., & Treml, E. A. (2016).  
1133 Navigating the currents of seascape genomics: how spatial analyses can augment  
1134 population genomic studies. *Current Zoology*, 62(6), 581-601.  
1135 doi:10.1093/cz/zow067
- 1136 Rosel, P. E., France, S. C., Wang, J. Y., & Kocher, T. D. (1999). Genetic structure of  
1137 harbour porpoise *Phocoena phocoena* populations in the northwest Atlantic based  
1138 on mitochondrial and nuclear markers. *Molecular Ecology*, 8(12 Suppl 1), S41-  
1139 54.

- 1140 Ruiz-Cooley, R. I., Gerrodette, T., Fiedler, P. C., Chivers, S. J., Danil, K., & Ballance, L.  
 1141 T. (2017). Temporal variation in pelagic food chain length in response to  
 1142 environmental change. *Science Advances*, 3(10). doi:10.1126/sciadv.1701140
- 1143 Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular Cloning: A Laboratory*  
 1144 *Manual, 2nd edn*. New York: Cold Spring Harbor Laboratory Press.
- 1145 Selkoe, K. A., D'Aloia, C. C., Crandall, E. D., Iacchei, M., Liggins, L., Puritz, J. B., . . .  
 1146 Toonen, R. J. (2016). A decade of seascape genetics: contributions to basic and  
 1147 applied marine connectivity. *Marine Ecology Progress Series*, 554, 1-19.  
 1148 doi:10.3354/meps11792
- 1149 Shirk, A. J., Landguth, E. L., & Cushman, S. A. (2017). A comparison of individual-  
 1150 based genetic distance metrics for landscape genetics. *Molecular Ecology*  
 1151 *Resources*, 17(6), 1308-1317. doi:10.1111/1755-0998.12684
- 1152 Storey, J. D., & Tibshirani, R. (2003). Statistical significance for genomewide studies.  
 1153 *Proceedings of the National Academy of Sciences of the United States of America*,  
 1154 100(16), 9440-9445. doi:10.1073/pnas.1530509100
- 1155 Taguchi, M., Chivers, S. J., Rosel, P. E., Matsuishi, T., & Abe, S. (2010). Mitochondrial  
 1156 DNA phylogeography of the harbour porpoise *Phocoena phocoena* in the North  
 1157 Pacific. *Marine Biology*, 157(7), 1489-1498. doi:10.1007/s00227-010-1423-7
- 1158 Tan, M. P., Wong, L. L., Razali, S. A., Afiqah-Aleng, N., Mohd Nor, S. A., Sung, Y. Y., .  
 1159 . . Danish-Daniel, M. (2019). Applications of Next-Generation Sequencing  
 1160 Technologies and Computational Tools in Molecular Evolution and Aquatic  
 1161 Animals Conservation Studies: A Short Review. *Evol Bioinform Online*, 15,  
 1162 1176934319892284. doi:10.1177/1176934319892284
- 1163 Tiedemann, R., Harder, J., Gmeiner, C., & Haase, E. (1996). Mitochondrial DNA  
 1164 sequence patterns of harbour porpoises (*Phocoena phocoena*) from the North and  
 1165 the Baltic sea. *Zeitschrift Fur Saugetierkunde-International Journal of*  
 1166 *Mammalian Biology*, 61(2), 104-111.
- 1167 Venegas, R. M., Strub, P. T., Beier, E., Letelier, R., Thomas, A. C., Cowles, T., . . .  
 1168 Cabrera, C. (2008). Satellite-derived variability in chlorophyll, wind stress, sea  
 1169 surface height, and temperature in the northern California Current System.  
 1170 *Journal of Geophysical Research-Oceans*, 113(C3). doi:10.1029/2007jc004481
- 1171 Walton, M. J. (1997). Population structure of harbour porpoises *Phocoena phocoena* in  
 1172 the seas around the UK and adjacent waters. *Proceedings of the Royal Society B-*  
 1173 *Biological Sciences*, 264(1378), 89-94. doi:DOI 10.1098/rspb.1997.0013

- 1174 Wang, J. Y., & Berggren, P. (1997). Mitochondrial DNA analysis of harbour porpoises  
1175 (*Phocoena phocoena*) in the Baltic Sea, the Kattegat-Skagerrak Seas and off the  
1176 west coast of Norway. *Marine Biology*, 127(4), 531-537. doi:DOI  
1177 10.1007/s002270050042
- 1178 Waples, R. S., & Gaggiotti, O. (2006). What is a population? An empirical evaluation of  
1179 some genetic methods for identifying the number of gene pools and their degree  
1180 of connectivity. *Molecular Ecology*, 15(6), 1419-1439. doi:10.1111/j.1365-  
1181 294X.2006.02890.x
- 1182 Willing, E. M., Dreyer, C., & van Oosterhout, C. (2012). Estimates of genetic  
1183 differentiation measured by F(ST) do not necessarily require large sample sizes  
1184 when using many SNP markers. *PLoS One*, 7(8), e42649.  
1185 doi:10.1371/journal.pone.0042649
- 1186 Willis, P. M., Crespi, B. J., Dill, L. M., Baird, R. W., & Hanson, M. B. (2004). Natural  
1187 hybridization between Dall's porpoises (*Phocoenoides dalli*) and harbour  
1188 porpoises (*Phocoena phocoena*). *Canadian Journal of Zoology-Revue  
1189 Canadienne De Zoologie*, 82(5), 828-834. doi:10.1139/Z04-059  
1190

1191 Data accessibility  
1192 - Mitochondrial DNA haplotype sequences used in this study are complete or partial  
1193 sequences of sequences previously available in Genbank (Accession numbers  
1194 provided in Supplemental Table S2).  
1195 Nuclear genotypes and environmental data associated with the samples: Dryad doi:  
1196 <https://doi.org/10.5061/dryad.4tmpg4f6v>  
1197  
1198  
1199  
1200 Author contributions  
1201 P.A.M. K.P. and B.L.T designed the research project, and P.A.M and K.P. performed  
1202 analyses and wrote the paper. K.A.F., B.R.F. contributed environmental data and  
1203 analyses, and contributed to writing the paper. B.L.H.-H., K.M.R., C.A.C., L.G.B.-L.,  
1204 C.S., T.H. and M.C.F. contributed samples and/or genetic data used for analyses. All  
1205 authors contributed to preparation of the manuscript.



Figure 1. Sample distribution colored by *a priori* geographic strata used for analyses. Current U.S. stocks are shown with horizontal lines demarcating the boundaries. The Washington Inland Waters stock is shown with the shaded polygon.

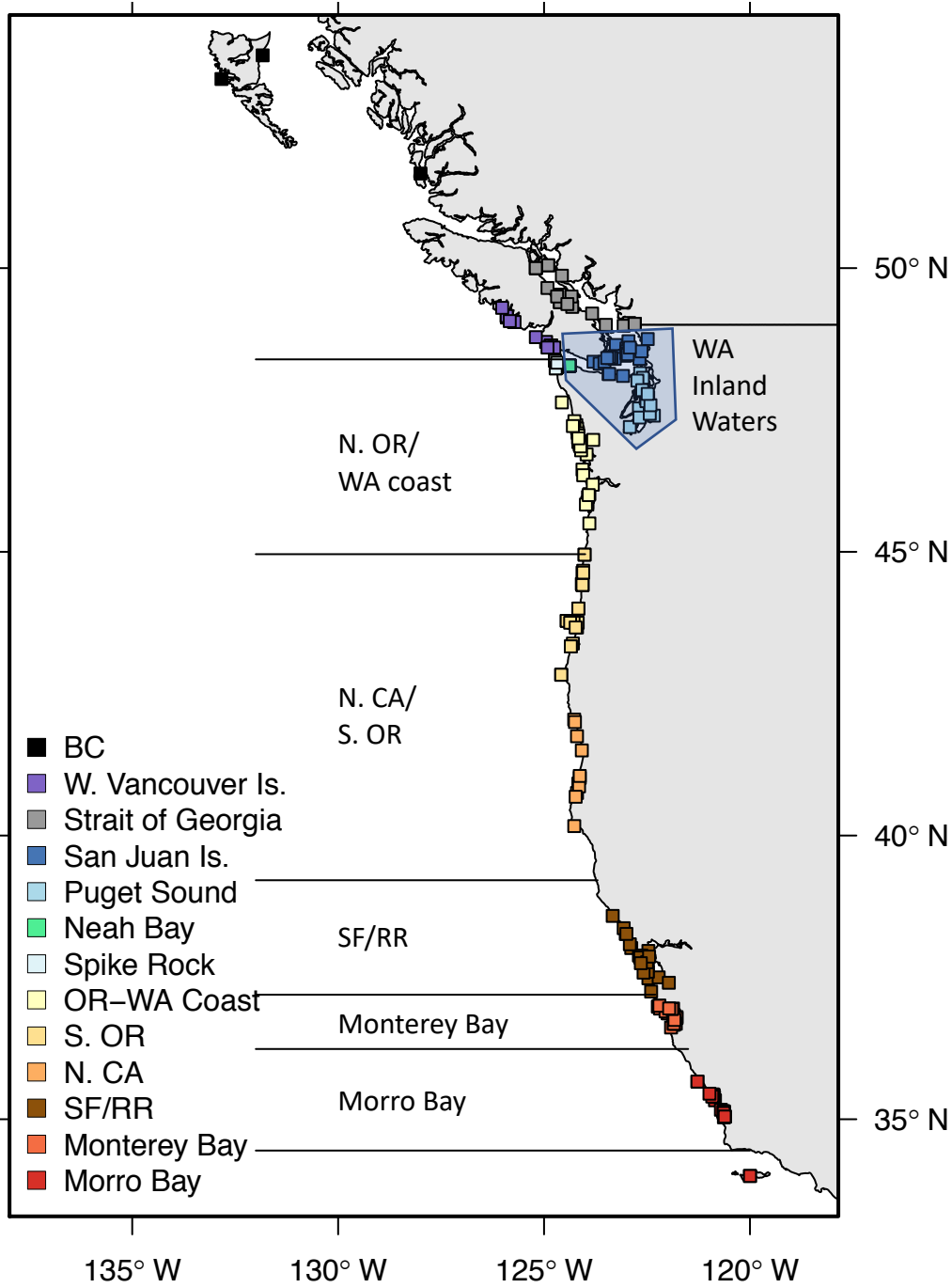
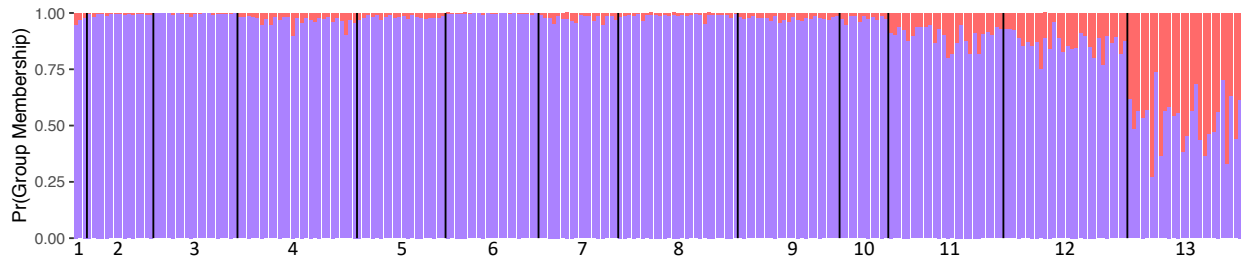


Figure 2. STRUCTURE plots for  $K = 2$  and  $K = 3$  based 290 loci and 264 samples (unrelated) with 10 replicates combined with CLUMPP. Samples assigned to *a priori* geographic strata are sorted by decreasing latitude in the bar plot from left to right. 1 = BC; 2 = W. Vancouver Is.; 3 = Strait of 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.

$K = 2$  ( $\Delta K = 59.1$ )



$K = 3$  ( $\Delta K = 0.9$ )

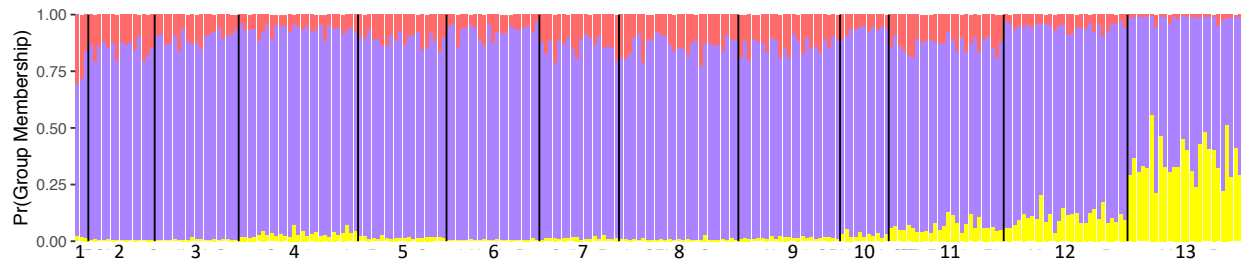


Figure 3. sPCA three-dimensional plot of the first three lagged scores of spatial PCs (sPC) for samples from A) all strata and samples (290 loci), and two-dimensional plots of the first two sPCs for geographical subsets of the data from B) outer coast strata (all strata except BC and those in (C); 290 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. Colors 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.

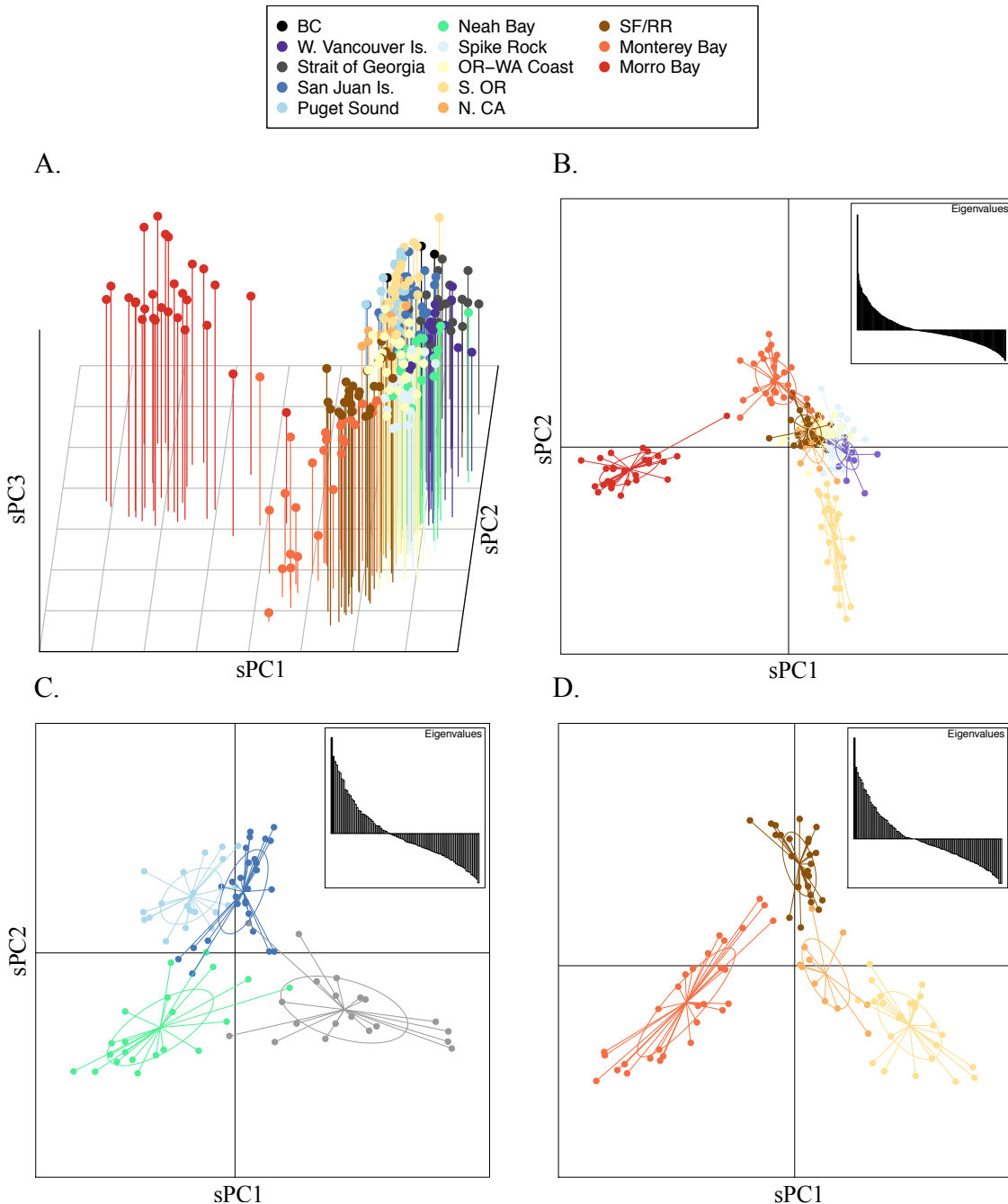




Figure 4. A) Gabriel connection network from a Monmonier analysis of samples (adjusted to prevent sample location overlap) with local nuclear genetic differences above the threshold value indicated by thicker blue lines and arrows between geographic distance edges for A) the outer coastal samples, with 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. Colors correspond to *a priori* strata as in Figure 1.

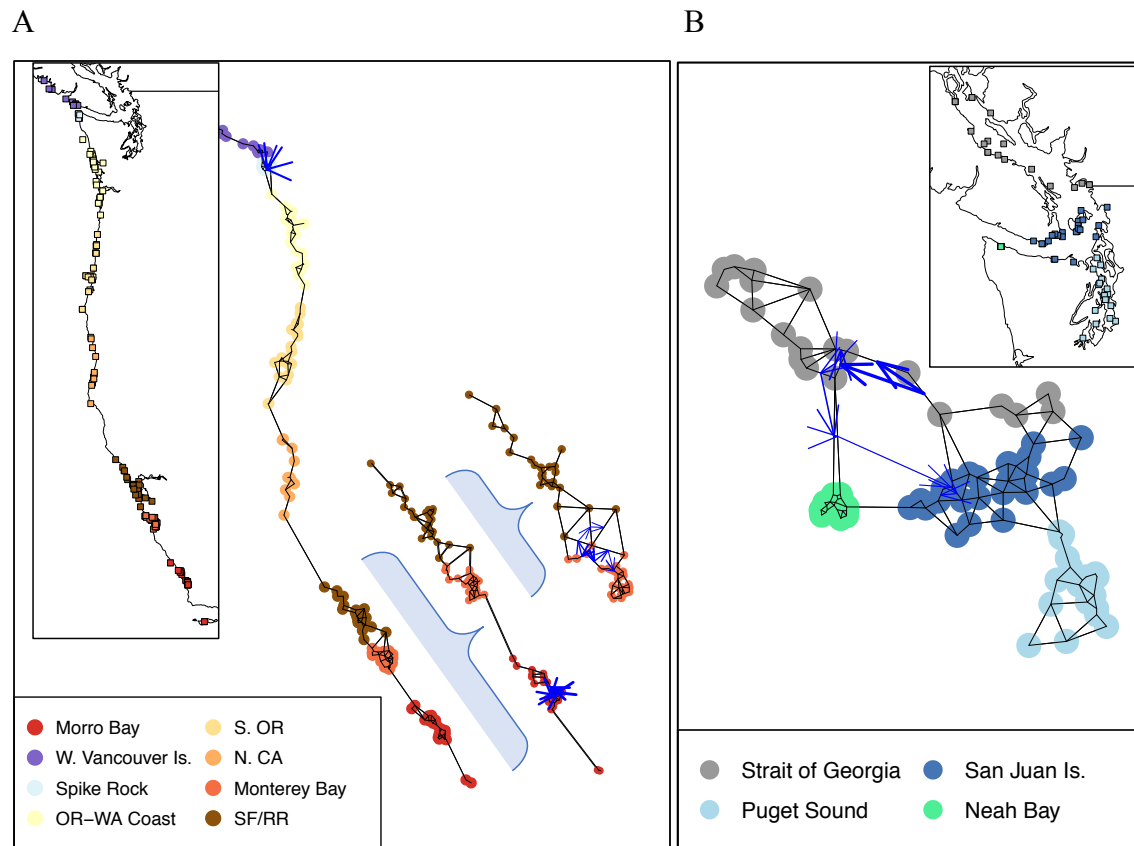


Figure 5. Pie charts represent frequencies of haplotypes represented by  $\geq 5$  samples in the stratum, plus two haplotypes (CR05, CR35) that were found only in the BC strata, in two samples each (out of 5 total). Relative porpoise density estimates (porpoise per kilometer surveyed) are plotted by latitude, averaged across multiple survey periods (blue line). The current U.S. 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.

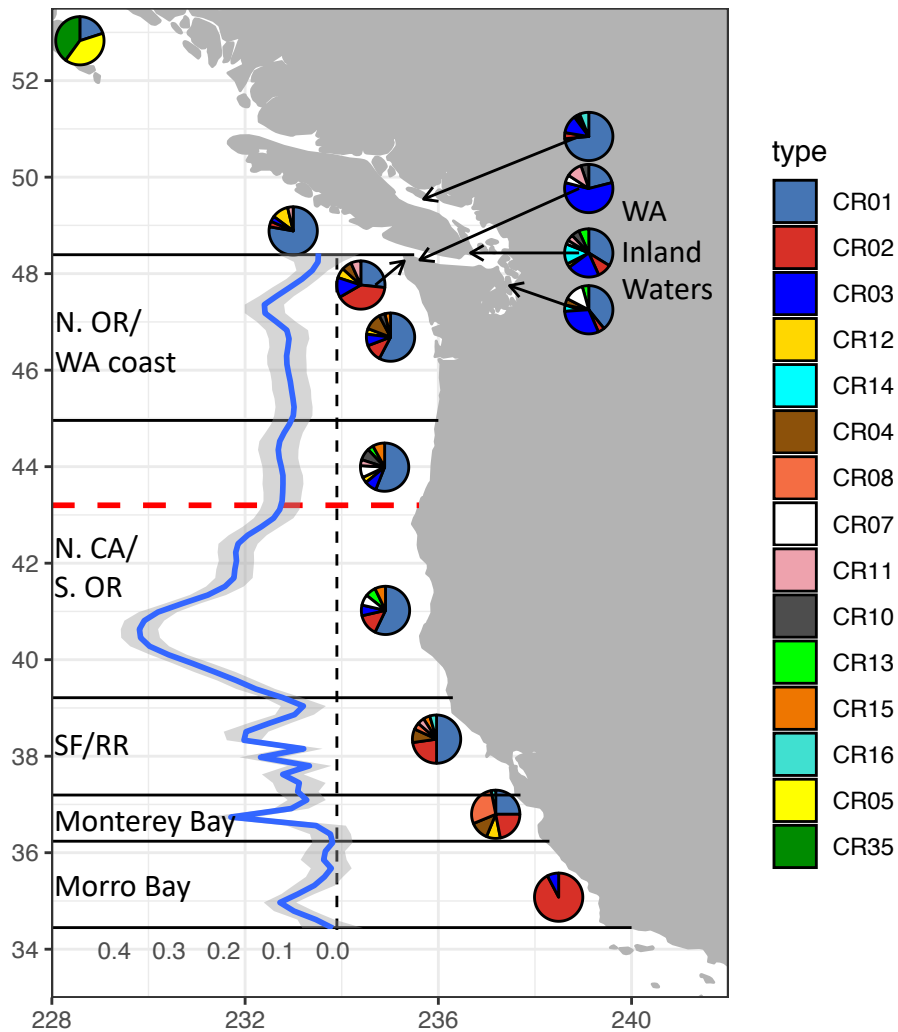
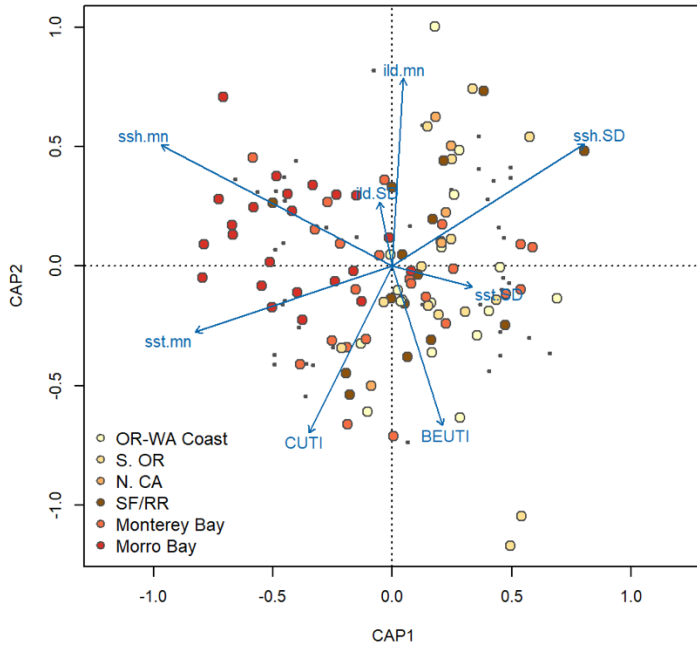


Figure 6. Distance-based redundancy analysis triplot of candidate microhaplotypes under selection for A) the outer coastal strata and B) inland waterways strata, showing individuals as colored circles, microhaplotype alleles as grey points, and environmental predictors (see Table 1 for abbreviations)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

A)



B)

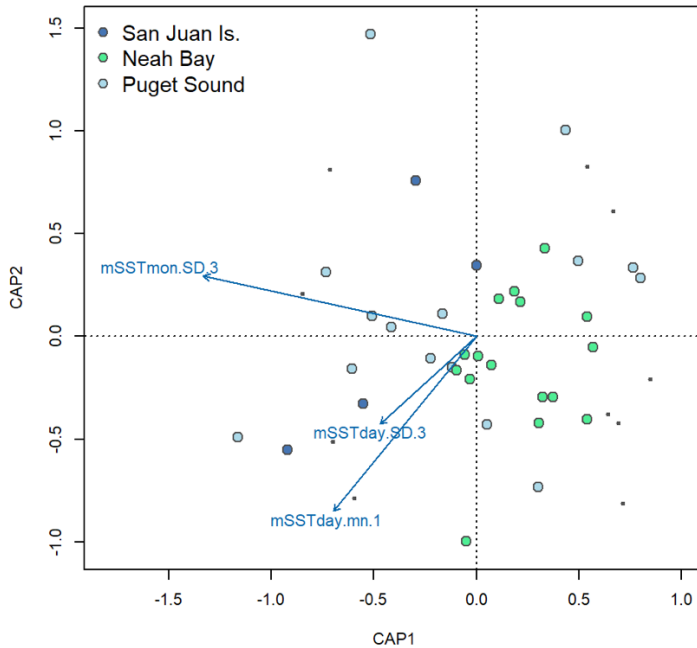


Table 1. Environmental predictors for distance-based redundancy analysis. SD = standard deviation.

<b>Analysis</b>	<b>Environmental predictor</b>	<b>Abbreviation</b>	<b>Predictor resolution</b>	<b>Source</b>
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; <a href="http://oceanmodeling.ucsc.edu/">http://oceanmodeling.ucsc.edu/</a> ; (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 31N-47N	<a href="http://mjacox.com/upwelling-indices/">http://mjacox.com/upwelling-indices/</a> (Jacox et al., 2018)
Biologically Effective Upwelling Transport Index	BEUTI	Calculated every 1 degree latitude along West Coast from 31N-47N		
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. <a href="https://upwell.pfeg.noaa.gov/erddap">https://upwell.pfeg.noaa.gov/erddap</a> . 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	



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

Species	Samples	Hexp.	Hobs.	Monomorphic
<i>P. phocoena</i>	281	0.366	0.356	2
<i>P. dalli</i>	11	0.144	0.145	177
hybrid	5	0.479	0.324	64

Table 3. Pairwise  $F_{ST}$  for nuclear loci (lower left) and mtDNA haplotypes (upper right). Statistically significant values ( $p < 0.05$ ) are in bold. Sample sizes each population is 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 supplemental Table S9.

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

Table 4. dbRDA sample information for three data sets filtered for different levels of missing genotypes. Rows in italics report the number of retained individuals in each set of populations. GIF = genomic inflation factor; FDR = false discovery rate.

	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	
<i>OR-WA Coast</i>	<i>17</i>	<i>17</i>	<i>14</i>	<i>4</i>	<i>4</i>	<i>4</i>	<i>San Juan Is.</i>
<i>S. OR</i>	<i>14</i>	<i>14</i>	<i>14</i>	<i>16</i>	<i>16</i>	<i>14</i>	<i>Neah Bay</i>
<i>N. CA</i>	<i>5</i>	<i>5</i>	<i>5</i>	<i>16</i>	<i>16</i>	<i>13</i>	<i>Puget Sound</i>
<i>SF/RR</i>	<i>16</i>	<i>14</i>	<i>13</i>	-	-	-	
<i>Monterey Bay</i>	<i>23</i>	<i>23</i>	<i>23</i>	-	-	-	
<i>Morro Bay</i>	<i>23</i>	<i>23</i>	<i>23</i>	-	-	-	
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	