Population genomics of harbour seal *Phoca vitulina* from northern British Columbia through California and comparison to the Atlantic subspecies

Ben J. G. Sutherland[1,2], Claire Rycroft[2], Ashtin Duguid[2], Terry D. Beacham[2], Strahan Tucker[2]

[1] Sutherland Bioinformatics, Lantzville, British Columbia, Canada, V0R 2H0.

[2] Fisheries and Oceans Canada, Pacific Biological Station, 3190 Hammond Bay Road, Nanaimo, BC V9T 6N7, Canada.

Author for correspondence: BJGS

Current Address:

Sutherland Bioinformatics, Lantzville BC, Canada, V0R 2H0

Email: Sutherland.Bioinformatics@protonmail.com

Author for correspondence: ST

Fisheries and Oceans Canada, Pacific Biological Station, 3190 Hammond Bay Road, Nanaimo, BC V9T 6N7, Canada

Email: Strahan.Tucker@dfo-mpo.gc.ca

Data Deposition: Raw sequence data have been uploaded to SRA under BioProject PRJNA948428, BioSamples SAMN33902654-SAMN33902801.

Abstract

The harbour seal *Phoca vitulina* is a marine mammal predator with potentially significant ecological impacts. Two subspecies exist in North America, *P. v. richardii* in the Pacific Ocean, and *P. v. concolor* in the Atlantic. Across the coastal ranges of the harbour seal, strong natal philopatry results in fine-scale genetic structure and isolation-by-distance. A stepwise decrease in genetic diversity in the subspecies was recently reported, with highest diversity in the eastern Pacific Ocean, medium diversity in the western Atlantic Ocean, and lowest diversity in the eastern Atlantic Ocean. Management of harbour seals is expected to benefit from improved resolution of fine-scale population structure and dynamics. Here we use genotyping-by-sequencing to genotype 146 harbour seals from the eastern Pacific Ocean (i.e., British Columbia, Oregon, and California) and the western Atlantic Ocean (i.e., Québec, Newfoundland, and Labrador). Using 12,742 identified variants, we confirm elevated genetic diversity in the eastern Pacific, and determine that this is independent of reference genome bias or other potential technical artefacts. The largest differentiation occurs between the subspecies, followed by divergence between British Columbia and Oregon-California, and between Québec and Newfoundland-Labrador. Gradual separation in samples and lower differentiation is viewed between California and Oregon, and between northern and southern BC. Unexpected outlier clusters were observed in two populations due to cryptic relatedness of individuals, which were subsequently removed from the analysis. Additional sampling is needed in the central and north coast of British Columbia to determine whether a discrete separation of populations exists within the region.

Keywords: breeding colony; genetic diversity; genotyping; harbour seal; marine mammal; *Phoca vitulina*; population genomics; RADseq; relatedness

Introduction

The harbour seal (*Phoca vitulina* Linnaeus 1758) is the most widely distributed pinniped species globally (Andersen & Olsen 2010; Blanchet *et al.* 2021). The reported number of subspecies varies from three to five, but typically the recognized subspecies include *P. v. stejnegeri* (Allen 1902) in the western Pacific Ocean, *P. v. richardii* (Gray 1864) in the eastern Pacific Ocean, *P. v. concolor* (De Kay 1842) in the western Atlantic, and *P. v. vitulina* (Linnaeus 1758) in the eastern Atlantic (Andersen & Olsen 2010). Deep divergence exists between the Pacific and Atlantic subspecies (Liu *et al.* 2022), where the separation is estimated to have occurred 1.7-2.2 million years ago (Andersen & Olsen 2010). Secondly, strong differentiation exists between western and eastern subspecies within each ocean (Stanley *et al.* 1996; Liu *et al.* 2022). Historical events and glaciations have had large influences on the present day population genetics of the species (Stanley *et al.* 1996), and a step-wise reduction in genetic diversity has been observed, with the highest diversity in the eastern Pacific Ocean, followed by the western Atlantic and western Pacific, and finally the lowest in the eastern Atlantic Ocean (Liu *et al.* 2022).

Harbour seals have been challenged by habitat loss, interspecific competition (Cordes *et al.* 2017), predation, epizootics, and human hunting or culling activity; different stressors have impacted different regions, but generally most regions have seen historical population declines, many of which have rebounded in recent years (Jeffries *et al.* 2003; Brown *et al.* 2005; Olesiuk 2010; Muto *et al.* 2021; DFO 2022). In the eastern Pacific Ocean, large-scale predator control programs and harvests in the late 1800s to mid 1900s reduced the subspecies down to approximately 10,000 animals (Olesiuk 2010). Harbour seals were then protected in the early 1970s, dramatically increasing to over 100,000 animals, which are now stable at the expected carrying capacity (Brown *et al.* 2005). By comparison, the population in eastern Canada (excluding Newfoundland) is comprised of 8,000-12,000 animals (reviewed by Blanchet *et al.* 2021). Throughout the range, declines and bottlenecks may have left impacts on the genetic diversity of the species (Burg *et al.* 1999). Severe reductions in population size in the eastern Atlantic Ocean from overhunting, environmental pollution, and epizootics, may have resulted in bottlenecks and resultant low genetic diversity (Swart *et al.* 1996; Goodman 1998; Andersen *et al.* 2011), leading to calls for immediate monitoring in some areas (Andersen *et al.* 2011). Genetic diversity differences from the eastern Pacific to the western Atlantic Ocean, and across the Atlantic Ocean from west to east suggests founder effects of colonization (Liu *et al.* 2022) and/ or recolonization events (Stanley *et al.* 1996). Therefore historical population dynamics and demographic history have shaped the contemporary diversity patterns in this species, and along with population structure, are important to consider for species management and conservation (Olsen *et al.* 2014).

Standing genetic variation is important for rapid adaptation (Barrett & Schluter 2008). In harbour seals, this may preserve the evolutionary potential for handling disease outbreaks or environmental

perturbations (Andersen & Olsen 2010), especially at the edges of the range where changes in climate may have a more drastic impact (Andersen *et al.* 2011; Blanchet *et al.* 2021). Genetic diversity has been directly connected to fitness in harbour seals, where lower heterozygosity was associated with increased lungworm burden in juvenile Wadden Sea harbour seals (Rijks *et al.* 2008), and birth weight and survival-to-weaning correlate with genetic diversity in harbour seal pups on Sable Island, Nova Scotia (Coltman *et al.* 1998). Quantifying genetic diversity of populations may help estimate adaptive potential (Barrett & Schluter 2008), and preserving this diversity is an important conservation goal that is expected to guide management decisions. Population structure is also expected to support these efforts, for example through designating management units with reliable and integrated data, potentially including demographic information and population viability analysis (Olsen *et al.* 2014; Olsen *et al.* 2017). Population structure and management unit designation are particularly important to understand at a fine-scale if the population is subject to harvest (Andersen & Olsen 2010). Additionally, the health of populations is important to consider in a metapopulation context, where some healthy populations may act as sources of harbour seals into other areas, and other populations in decline may act as sinks (Carroll *et al.* 2020).

Isolation-by-distance (IBD) is a main driver of population structure in harbour seals (Stanley *et al.* 1996; Goodman 1998; Huber *et al.* 2012; Liu *et al.* 2022), with increased genetic separation occurring with discontinuous distribution of colonies (Goodman 1998). However, some unexpectedly high differentiation has been observed in relatively close proximities (e.g., western and eastern Baltic Sea; Stanley *et al.* 1996). Even though harbour seals travel great distances, high philopatry to breeding colony occurs (Goodman 1998; Härkönen & Harding 2001). From eastern North America to Europe, originally 12 regional populations were identified, but the authors considered that the true number is likely greater (Andersen & Olsen 2010), which has been further explored recently (Liu *et al.* 2022). Closer inspection of regions identifies more fine-scale structure; the United Kingdom is now delineated into northwest, northeast, and southeast populations (Olsen *et al.* 2017), and Ireland is delineated into northwest/north, eastern, and southwest populations (Steinmetz *et al.* 2023). Similarly, eastern North America was originally considered a single population, fitting with continuous clinal variation in pupping season (Temte *et al.* 1991), but has now been further characterized (see Liu *et al.* 2022). Western North America has previously been considered to be comprised of three coastal harbour seal populations, delineated as southeast Alaska through northern BC; southern BC; and Washington to California (Burg *et al.* 1999), as well as a genetically distinct population in Puget Sound, Washington State (Lamont *et al.* 1996). However, inland waters of Washington State may in fact contain three discrete populations, including Puget Sound (Huber *et al.* 2012). More generally, most areas still require characterization of localized regions to understand population dynamics at this finer scale (Blanchet *et al.* 2021).

Clear identification of genetic units is essential for understanding population status and implementing management schemes (see Andersen & Olsen 2010), and this is likely to benefit from an integrative approach including demographic analysis (Olsen *et al.* 2014). In BC and Washington, perceived impacts of harbour seals on salmon have renewed discussions on regional population control measures (Trzcinski 2020). Moreover, increased interest in Food, Social or Ceremonial (FSC) harvests in BC is leading managers to consider regional harvest allocations, but there is incomplete information on stock structure or status, and in BC, harbour seals are managed as one stock (DFO 2022). A coastwide assessment of population genetic structure and diversity could support the definition of units relevant for management and conservation efforts (Waples & Gaggiotti 2006), although other considerations, such as local demography, can also be integrated into such analyses (Olsen *et al.* 2014). A broad-scale population genetic assessment would benefit from a consistent and high throughput genotyping method, as well as systematically sampled natal sites, sampling inhabitants if possible (Huber *et al.* 2012).

Genetic studies on harbour seals have typically used allozymes (Swart *et al.* 1996), mitochondrial DNA (e.g., Stanley *et al.* 1996), and microsatellites (e.g., Goodman 1998; Huber *et al.* 2012). Recently, single nucleotide polymorphisms (SNPs) by restriction site-associated DNA sequencing (RAD-seq; Baird *et al.* 2008; Peterson *et al.* 2012) were analyzed at a global scale (Liu *et al.* 2022). SNPs have some known benefits over microsatellites (Hauser *et al.* 2011), such as ease and automation of genotyping, interoperability between labs, and fewer individuals required to estimate allele frequencies (Beacham *et al.* 2011). Although high throughput methods such as RAD-sequencing require some bioinformatic expertise (Shafer *et al.* 2017), once variants are identified they can be integrated into amplicon panels, which can be easier to analyze to provide consistent and effective results for conservation genomics applications (Meek & Larson 2019).

In the present study we genotype harbour seals from broad geographic regions in the eastern Pacific Ocean (i.e., northern British Columbia (BC), Strait of Georgia (SOG), Oregon, and California), and in the western Atlantic Ocean (i.e., Newfoundland, Labrador, and eastern Quebec). We analyze population differentiation and diversity differences, as well as within-population relatedness, and discuss the results in relation to the recently published global analysis (Liu *et al.* 2022), testing both datasets for any impacts of reference bias on diversity differences. The high quality multilocus genotypes generated here are expected to be used to develop a high throughput genotyping tool for the Pacific harbour seal subspecies for future work in characterizing the populations of this species.

Methods

Sample collection, DNA extraction, library preparation and sequencing

Harbour seal samples were obtained as skin or blood samples from collections in northern and southern British Columbia (NBC, SBC), Oregon (ORE), California (CAL), Quebec (QUE), Newfoundland (NFL), and Labrador (LAB) (Figure 1; Table 1; Additional Files S1-S3). Serum samples were also obtained from harbour seals in Alaska, but extractions failed to yield sufficient DNA and therefore were not included. Samples were often opportunistically collected and came from a variety of sources that included mortalities, strandings, rehabilitating captive seals, and research surveys.

DNA was extracted using magnetic bead extraction on a BioSprint automated DNA extraction system, following manufacturer's instructions (QIAGEN). Extracted purified DNA was then quantified by Qubit (Thermo Fisher) and normalized to 20 ng/µl in 10 µl volume. The normalized DNA was submitted to the Institute of Integrative Biology and Systems (IBIS) Genomic Analyses Platform of Université Laval, where libraries were prepared using double-digest restriction-site associated DNA sequencing (ddRAD-seq). The enzymes *Nsi*I and *Msp*I were used based on *in silico* cut site assessment using SimRAD (Lepais & Weir 2014). Libraries were sequenced on an Ion Torrent sequencer (Thermo Fisher) in batches of 20 samples per sequencing chip. Raw single-end sequence data of variable read lengths with a mode of 177 bp were obtained for downstream analyses.

Bioinformatics

Raw sequence data were inspected for quality using fastqc (v.0.11.4; Andrews 2010) and visualized using multiqc (v.1.14; Ewels *et al.* 2016). In general, the pipeline stacks_workflow (commit 868b410; E. Normandeau, GitHub) was followed, as explained below in brief. Adapters and any reads shorter than 50 bp were removed using cutadapt (v.1.9.1; Martin 2011). Sequence data were then de-multiplexed using *process_radtags* of Stacks2 (v.2.62; Rochette *et al.* 2019) in parallel (Tange 2022) trimming to a uniform length of 80 bp and de-multiplexing based on cut sites from the appropriate restriction enzymes. Samples were re-inspected for quality, then aligned to the harbour seal reference genome (i.e., GCA_004348235.1_GSC_HSeal_1.0; sourced from a female Pacific subspecies) using bwa (v.0.7.12-r1039; Li 2013) and converted to bam format using samtools (v.1.17; Danecek *et al.* 2021). Numbers of reads and alignments per sample were quantified using custom scripts (see Data Availability). Two samples were found to be extreme outliers in a downstream principal components analysis (PCA) and were removed (i.e., NFL_102; NFL_121), as were two samples with fewer than 1 M reads (i.e., NBC_110, ORE_104).

Several approaches to genotyping were applied. First, the reference genome-based approach was used within *gstacks* with software defaults. The *populations* module was then used to filter, retaining any locus that was present in at least 70% of individuals, in all seven regional groupings, and at a minor allele

frequency (MAF) ≥ 0.01. Second, to evaluate possible effects of the reference genome on genotyping of the different subspecies, *de novo* genotyping approaches were conducted within Stacks2 for each coast, using a balanced design of 18 samples from each of SOG, ORE (eastern Pacific), and EQB, and NFL (western Atlantic). The datasets were further balanced by normalizing all samples to 1.25 M reads each using seqtk's sample function (v1.3-r106; Heng Li). The analytic pipeline for the project with all code is available on GitHub (see Data Availability, *ms_harbour_seal*).

Population genetics

Polymorphism rate, average observed heterozygosity (H_{OBS}) and nucleotide diversity (π) per-population were calculated considering all genotyped loci in Stacks2. A VCF was generated using a single SNP exported per locus, and the mean genotype depth per locus per sample and per population was calculated from the VCF using the *extract.gt* function of vcfR (Knaus & Grünwald 2017). For downstream analysis, a genepop file was also generated using a single SNP per locus using the Stacks2 populations module, and this was read into the R environment (R Core Team 2023) using adegenet (v.2.1.8; Jombart & Ahmed 2011). The data were converted to genlight format using dartR (v.2.0.4; Gruber *et al.* 2018) and coast-specific datasets were generated (i.e., Atlantic or Pacific), each of which included a coast-specific MAF filter (i.e., MAF > 0.01). Unsupervised PCAs were generated for the global and coast-specific datasets. Due to unexpected clustering of samples within regional collections, inter-individual relatedness was estimated between samples using related (v.1.0; Pew *et al.* 2015) and the Ritland statistic (Ritland 1996) to characterize potential cryptic relatedness. Within-population outlier pairs were identified (i.e., pairs with Ritland relatedness values above the third quartile + 1.5x the interquartile range for that population). For each outlier pair, one of the individuals was removed from the dataset until no more outlier pairs remained above the population-specific cutoff. After purging the putative local relatives, each dataset was filtered again for MAF (i.e., MAF > 0.01). Loci were then evaluated for Hardy-Weinberg proportions within each population using the function *hw.test* of pegas, removing any variant that was non-conforming to HWP ($p \leq 0.01$) in any one population. H_{OBS} and F_{ST} were then calculated per locus for each dataset using the *summary* function of adegenet and the function *Fst* within pegas (v.1.1; Paradis 2010), respectively, and any locus with excess observed heterozygosity ($H_{OBS} > 0.5$) was removed. These datasets were then used as inputs to coast-specific PCAs, genetic differentiation assessment between populations with 95% confidence intervals of F_{ST} (Weir & Cockerham 1984) in hierfstat (v.0.5-11; Goudet & Jombart 2022), and a neighbour-joining tree using the edwards.dist distance metric (Cavalli-Sforza & Edwards 1967) was generated using 10,000 bootstraps with the aboot function of poppr (v.2.9.3; Kamvar *et al.* 2014). Private alleles were also identified between specific groupings of samples using poppr. Analyses were conducted

using functions within the GitHub repository *simple_pop_stats* with analysis R scripts available within the repository *ms_harbour_seal* (see Data Availability).

Re-analysis of Liu et al. 2022

A recently published RAD-sequencing dataset (*Pst*I enzyme) that used genotype likelihood calls to investigate global trends in harbour seal genomic diversity (Liu *et al.* 2022) was downloaded using the RunSelector from the Short Read Archive (SRA) of NCBI and the SRA Toolkit (NCBI 2023). Specifically, 14 samples from each of the following collections were downloaded: Endicott Arm (END) and Kodiak, Alaska (KOD); Île du Bic (BIC) in the St. Lawrence River, Quebec; Newfoundland (NFL); Orkney, Scotland (ORK); and Wadden Sea, Netherlands (WNL). Samples were analyzed using the reference-based genotyping methods described above, as well as by region-specific, individual *de novo* analyses. Methods were as described above, except that all reads were truncated to 70 bp and any read shorter than 70 bp was removed from the analysis, due to Stacks2 developers' recommendation to keep all reads the same length for *de novo* analyses. In brief, the all-region approach retained genotypes present in all six populations in at least 70% of individuals per population, with global MAF ≥ 0.01. The three separate *de novo* analyses were also filtered requiring both populations in each dataset to be genotyped in at least 70% of the individuals in each population, and all retained variants to have global MAF ≥ 0.01. Population-level average heterozygosity was calculated by the Stacks2 populations module.

Results

Genotyping

The 146 samples retained in the analysis had an average (± s.d.) of 3.8 ± 2.0 M reads each and were comprised in total of 561.6 M reads (see Table 1 for totals per population). No outlier samples were identified at this stage (but see Methods for earlier outlier removal). Alignment rates against the reference genome, which was generated from a Pacific subspecies female, ranged from 88.5-95.1% for Pacific samples and 85.9-87.3% for Atlantic samples. Genotyping guided by reference genome alignments used 96% of the 508.6 M primary alignments, discarding 2.8% due to low mapping quality and 1.2% due to excessive soft clipping. This resulted in the formation of 1.11 M loci (i.e., RAD-tags) comprised of 488.1 M reads with an average effective coverage of 27.5 ± 27.9x prior to filtering. Filtering for missing data and minor allele frequency (MAF; see Methods) resulted in 53,684 loci retained with a mean length of 86.6 bp. For the genotyped markers, per individual missing data was on average 2.9 ± 3.0%. These final loci were comprised of 4.6 M genomic sites, and 12,742 variants were observed, or 10,847 variants if only retaining a single SNP per locus. The mean genotype depth for all samples and loci was 35.1x, with a wide variance

between samples (Figure S1), with each population ranging from a mean genotype depth of 13.3x (SOG) to 83.3x (ORE; Figure S2).

Subspecies genetic differentiation and diversity differences

The two subspecies were found to be highly differentiated (Figure 1C), separating across PC1 (33.3% of total dataset variation; Figure S3), similar to that recently reported (Liu *et al.* 2022). As an example, CAL and NFL 95% confidence interval (c.i.) F_{ST} was 0.40-0.42 (Table S1).

Polymorphism rates also differed greatly between the subspecies. The polymorphism rate considering populations with at least 15 individuals ranged from 0.180-0.194% in Pacific populations and 0.083-0.087% in Atlantic populations (see Table 2 for population-specific metrics). Observed heterozygosity (H_{OBS}) was 1.9-fold higher in the Pacific than in the Atlantic subspecies for these collections (mean: 0.00045 and 0.00024, respectively), consistent with recently reported results (Liu *et al.* 2022). Data were then separated into west and east coast datasets, and once separated, each coast was filtered again for MAF, retaining 8,933 variants in the Pacific dataset and 3,828 variants in the Atlantic.

To further evaluate the large differences in polymorphism rate, we re-analyzed the data per coast using a consistent sample size and read depth per sample, and without the reference genome (i.e., *de novo*; see Methods). For each of the two coasts, 45 M reads were used in total, and after filters, 17,490 loci comprising 1.4 M sites and 3,373 variants were observed for the Pacific subspecies (mean polymorphism rate = 0.183%), and 40,909 loci comprising 3.3 M sites and 3,483 variants for the Atlantic subspecies (mean polymorphism rate = 0.085%). This demonstrates that the cause of the elevated diversity was not due to reference genome bias. Notably, in the balanced and normalized dataset, ORE showed higher observed heterozygosity than SOG (H_{OBS} = 0.00048 and 0.00039, respectively; Additional File S4).

A subset of the data from Liu et al. (2022) was also re-analyzed to confirm the trends viewed here, and as previously reported, are independent of significant reference genome bias. Using the reference-guided genotyping on the subset of the samples described (see Methods), Pacific H_{OBS} ranged from 0.00031 to 0.00034, whereas the western Atlantic was estimated at 0.00019 (i.e., 1.7-fold higher H_{OBS} in Pacific), and 0.00009-0.00010 in the eastern Atlantic (Additional File S4). The *de novo* approach applied to a subset of the Liu et al. (2022) data resulted in a Pacific H_{OBS} of 0.00074 and Atlantic H_{OBS} of 0.00035-0.00037, respectively (i.e., 1.8-fold higher H_{OBS} in Pacific). Collectively, this indicates that the differences in diversity reported previously and observed here are not due to differences in sample size, read depth, nor genetic divergence of individuals from the reference genome.

Cryptic relatedness within population and genetic differentiation between populations

Coast-specific datasets were each inspected by PCA, and this showed an outlier cluster with nine individuals in the SOG, all of which were from Burrard Inlet, and an outlier grouping comprised of harbour seals from the southernmost sampled region in CAL (Figure S4). Each population was then inspected for relatedness (Figure S5), and it was determined that both PCA outlier clusters were comprised of individuals with elevated relatedness (outlier pairs Ritland metric ≥ 0.0975 and 0.1027 in CAL and SOG, respectively).

Private alleles were inspected for the Burrard Inlet cluster, which identified a private allele observed nine times in seven of the individuals (i.e., 586731_19.04), two private alleles observed six times (i.e., 729912_41.01, 807825_8.01), and three private alleles observed five times (i.e., 71641_26.04, 503353_64.04, 839341_28.01). By comparison, analyzing a haphazard selection of nine SOG (non-Burrard Inlet cluster) individuals found no private alleles observed more than three times.

A dendrogram based on genetic similarity between individuals by microhaplotypes indicated two shallow branches for the Burrard Inlet cluster (Additional File S5), but there was no difference between or within each of the samples in the shallow branches in the relatedness statistic (Additional File S6), and therefore individuals within the Burrard Inlet cluster were considered approximately equally related, with no evidence for further family structure. Within-population outlier pairs were then identified and one individual from each outlier pair was removed to remove putative close relatives (see Methods; Table 2; Table S2). Following the removal of putative relatives, loci were kept when MAF > 0.01 within each coast, retaining 8,721 and 3,658 variants in the Pacific and Atlantic coasts, respectively. Hardy-Weinberg outliers were then identified per population and removed (Table 2), as were any loci with dataset-level $H_{OBS} > 0.5$, resulting in 7,699 and 3,134 variants retained in the Pacific and Atlantic datasets, respectively. Of these remaining variants, 56.3% and 46.5% were between 0.01-0.10 MAF in the Pacific and Atlantic harbour seal datasets, respectively (Figure 2).

The Pacific dataset shows clear differentiation between the southern and northern samples, explaining 7% of the variation across PC1 and separating SOG/NBC from ORE/CAL (Figure 3A), with significant genetic differentiation (e.g., SOG-CAL 95% c.i. F_{ST} = 0.065-0.071; Table 3A). Within BC, overlap of confidence interval ellipses is observed for NBC and SOG, as well as for ORE and CAL in the USA. These collections show similar levels of differentiation, with NBC-SOG F_{ST} = 0.014-0.021 and ORE-CAL F_{ST} = 0.006-0.009 (Table 3A). Due to the low sample size of NBC, the SOG-NBC F_{ST} value should be taken with caution. Notably, the closest SOG samples in the PCA to the NBC samples were some of the most northern samplings within the SOG collection, including Nelson Island on the Sunshine Coast (SOG_118), and Anvil Island in Howe Sound (SOG_123). Sample NBC_111, sampled at northern Haida Gwaii was one of the most distant samples geographically, and the most distant sample along PC1 from the SOG samples (Figure 3A; Additional File S7). Putative relatives (removed in final analysis), NBC_112 and

316 NBC_113, collected from Prince Ruper and Massett (Haida Gwaii), respectively, were also grouped
317 together with NBC_111 distantly from the SOG, but this is also expected due to the estimated relatedness
318 to NBC_111. However, while some geographic trend is therefore apparent, given the low number of
319 characterized samples from northern BC, there is no clear break yet observed in the data between the SOG
320 and the more northern samples (Figure 3A).

321 The Atlantic subspecies analysis has a generally strong separation between the EQB and the
322 LAB/NFL collections along PC1 (5.3% of the variation; Figure 3B). After removing putative siblings, the
323 three samples that were highest on PC2 were sampled at or near the Métis harbour seal colony, and all those
324 lower on PC2 were from the Le Bic colony, which appears to be explaining the separation of these samples
325 across both PC2 and PC1. The LAB and NFL collections were completely overlapping and have an F_{ST}
326 value overlapping zero, suggesting that these are not separate populations. The separation of EQB and NFL
327 is however reflected in the significant F_{ST} between EQB and NFL (95% c.i. F_{ST} = 0.016-0.023; Table 3B).
328

329 *Selecting markers for panel design*

330 Markers that may be useful for characterizing population structure or individual identification were
331 identified in the dataset. Following data subsetting and filtering, elevated F_{ST} markers were identified
332 between SOG and NBC, and between SOG and ORE (Additional File S8). Per locus H_{OBS} was also analyzed
333 for the BC and USA samples separately, and contrasting per-locus H_{OBS} between the BC and USA datasets
334 indicates high correlation in H_{OBS} between the datasets (Figure S6), therefore indicating that markers can
335 be selected with high H_{OBS} in both regions.
336

337 # Discussion

338 *Subspecies divergence and differential genetic diversity*

339 Species conservation and management of harbour seal will likely depend on effective characterization of
340 population structure and contemporary standing genetic variation (including functional variation), that is
341 considered in the context of large-scale historical demographic events (as per Beichman *et al.* 2023), as
342 well as contemporary and local demography (Olsen *et al.* 2014). In this study, we confirm the existence of
343 elevated genetic diversity in the eastern Pacific subspecies of harbour seal *P. v. richardii* over the western
344 Atlantic subspecies *P. v. concolor* as was recently reported by Liu and co-workers (2022). We observe the
345 elevated diversity of the Pacific subspecies in new regions including British Columbia (BC) and Oregon,
346 whereas the previous analysis included Alaska and California for the eastern Pacific subspecies (Liu *et al.*
347 2022). Further, we evaluate and find that this trend is not caused by reference genome bias in genotyping,
348 where individuals with increased divergence from the reference genome may exhibit biased population
349 genetic metrics (Bohling 2020). Reference bias is typically considered in the context of divergent species,

350 but as more local genomes are produced, these issues may become more relevant within species (Beichman
351 *et al.* 2023; Thorburn *et al.* 2023). Further, given the high divergence between the subspecies here, it was
352 important to test whether *P. v. concolor* metrics were impacted by the alignment to the *P. v. richardii*
353 genome. The diversity difference of around 1.8x higher diversity as evaluated by H_{OBS} in the Pacific over
354 the Atlantic North American subspecies remains when genotyping without a reference genome (i.e., *de*
355 *novo*) and with a normalized number of samples and reads.

356 The ultimate cause of the diversity difference was proposed by Liu and co-workers (2022) to be
357 due to stepwise founder effects during colonization from the Pacific Ocean to the western Atlantic Ocean,
358 and then from the western Atlantic Ocean to the eastern Atlantic Ocean, with a decrease in genetic diversity
359 at each step. The divergence of the subspecies is expected to have occurred between 1.7-2.2 million years
360 ago (Andersen & Olsen 2010). Proposed west to east colonization within each basin was supported by
361 previous work (Stanley *et al.* 1996). Stanley and co-workers (1996) proposed that the European population
362 may have been eliminated during the last Ice Age (~18,000 y.a.), and that it was recolonized following the
363 recession of the ice sheets. These estimates and additional demographic history of these populations will
364 be important to explore with next-generation sequencing data (Beichman *et al.* 2023), albeit with caution
365 given the susceptibility of demographic inference to artefacts from pipeline parameters (Shafer *et al.* 2017).
366 The use of these multiple independent RAD-seq datasets may aid in this analysis. The relatively deep
367 sequencing and therefore high reliability genotypes expected in the present analysis, and the numerous
368 RAD-tags generated through restriction enzyme selection (Lepais & Weir 2014), may enable the analysis
369 of the allele frequency spectra of the two subspecies to characterize fluctuations in effective population size
370 (Ne) and therefore approximate the timing of the reduced genetic diversity event (e.g., Liu & Fu 2020;
371 Beichman *et al.* 2023). Uncertainty levels around the demographic inference approaches may not be robust
372 enough for conservation and management decisions, but rather applied towards general understanding of
373 putative and broad biogeographic hypotheses (Shafer *et al.* 2017).

374 The striking diversity differences between the subspecies may provide a model to study the effect
375 of environmental changes and adaptive potential in the two subspecies, and their responses to these
376 challenges. This may be particularly relevant at the edges of the ranges (Andersen *et al.* 2011; Blanchet *et*
377 *al.* 2021). Furthermore, the large divergence between the Pacific and the Atlantic subspecies raises the
378 question of whether the subspecies are in fact separate species (Liu *et al.* 2022). F_{ST} values observed by Liu
379 et al. (2022) ranged from 0.38 to 0.69, noting that this is higher than most other marine megafaunal
380 subspecies to date. Our results agree with these levels, and although we did not have samples from the
381 eastern Atlantic Ocean, here 95% lower and upper F_{ST} values for all contrasts between the west and east
382 coast of North America (excluding the low sample size NBC) ranged from F_{ST} = 0.36 to 0.43.
383

Population structure in the eastern Pacific

Harbour seals from the eastern Pacific Ocean along North America have been considered to be comprised of three coastal populations: (1) southeast Alaska through northern BC; (2) southern BC; and (3) Washington through California (Burg *et al.* 1999). In the present study, the separation between BC and Oregon/California is clear. Additionally, here some separation is viewed among samples from southern BC and northern BC, and among samples from Oregon and California, but there are no clear break points or strong separations observed between the putative populations. Continuous distributions of breeding animals along coastlines has previously been described to have little local substructuring, although a continued increase in differentiation occurs with distance (Goodman 1998). In the case of clinal isolation-by-distance, sampling at extreme ends of a geographical region may result in apparent clusters, whereas if the sampling was continuous throughout, a cline would be observed (Patterson *et al.* 2006). In this regard, continuous sampling along coastlines would help refine whether breakpoints exist or if a continuous distribution explains the coastal population structure. In addition to the continuous distribution of breeding animals occurring along the coast, substructuring may also be impeded with high population numbers (Burg *et al.* 1999). The location of a putative dividing line between southern and northern BC has yet to be determined (Burg *et al.* 1999). In the present study we also do not identify this line, which may be due to having too few northern BC samples genotyped here, and the location may be only identifiable with a larger survey of coastal breeding colonies. Furthermore, the samples used in the present study were mainly opportunistically sampled, and this type of sampling has been proposed to blur the lines between populations as opposed to careful sampling of unweaned pups to ensure exact natal locations (Huber *et al.* 2012). In any case, markers used here did show some spread of samples between northern and southern BC, and between Oregon and California. These markers, applied to a higher number of individuals in future sample collections using amplicon sequencing approaches, may help resolve some of these questions. A focus on systematic sampling of natal sites within the Central Coast region may be key in identifying a potential dividing line (Burg *et al.* 1999).

The Strait of Georgia (SOG) has the highest density of harbour seals on the BC coast with nearly 40,000 animals (i.e., 37% of BC population), and is used for population surveys since the implementation of protections (DFO 2022; Majewski & Ellis 2022). Notably, of the 12 harbour seals sampled in the Burrard Inlet area, nine were found to be highly related. These harbour seals were all pups, and were all sampled in the same summer. As female harbour seals only produce a single pup per year (Schaeff *et al.* 1999), it is possible that the samples could be half-sibs, sharing a sire. However, harbour seals are not expected to mate in harems, but rather in-water in a lek-type mating system (Boness *et al.* 2006), which makes the observation of so many putatively related pups surprising. Different levels of relatedness in colonies has been previously reported (Olsen *et al.* 2014). Regardless, the unexpected relatedness (i.e., cryptic relatedness) had

418 significant effects on the initial PCA (Yao & Ochoa 2023), highlighting the importance of removing
419 putative relatives to avoid erroneous identification of population clusters (Patterson *et al.* 2006), given their
420 overrepresentation of certain alleles (Elhaik 2022). This feature of population characterization, which is
421 expected in other methods and not only PCA (Patterson *et al.* 2006), should be considered with caution in
422 future sampling to determine whether structure exists along the coastline. This may be particularly relevant
423 for harbour seals, given the potential for variable levels of relatedness among colonies (Olsen *et al.* 2014).
424 The presence or extent of any genetic separation existing between this colony and others is unknown, and
425 would need follow-up sampling to understand further. In harbour seals, isolation-by-distance may not
426 explain all genetic separation, and significant differentiation can occur within only a few hundred
427 kilometers, potentially being driven by other demographic processes (Stanley *et al.* 1996; Goodman 1998).
428 Inland populations in Washington State that are genetically distinct from the coastal population had
429 the most diverse mitochondrial haplotypes (Lamont *et al.* 1996). These populations have different pupping
430 times and cranial morphometrics, although are geographically close, and may be more closely related to
431 harbour seals from northern populations, including the SOG (reviewed by Lamont *et al.* 1996). Late birthing
432 of pups has been observed in Puget Sound and Vancouver Island that is discontinuous with clinal variation
433 observed further south, again suggesting some unique characteristics in this area (Temte *et al.* 1991).
434 Furthermore, Burg and co-workers (1999) identified a unique mitochondrial haplotype signature in harbour
435 seals sampled from southern Vancouver Island, but did not see this group as a separate population in their
436 microsatellite analysis. Similarly, in the present work, we noted high frequency private alleles in the Burrard
437 cluster of putatively related individuals. Unique signatures were suggested to potentially originate from a
438 separate colonization event for these populations (Lamont *et al.* 1996; Burg *et al.* 1999). Whether the
439 Burrard cluster is simply an artefact of chance sampling of individuals with high relatedness from a fairly
440 genetically homogenous colony in the inlet, skewing allele frequencies and enriching rare alleles (Elhaik
441 2022), or in fact is evidence of impeded gene flow from the Burrard colony or colonies, requires further
442 study with systematically obtained samples over time. Additional samples from the inland populations of
443 Washington State (Lamont *et al.* 1996) to compare with the Burrard cluster would be important to include
444 in future analyses. A southern California cluster was also observed, and similarly this group was also found
445 to have high relatedness, and therefore was required to have close relatives removed to not appear as
446 artefactual population structure (Patterson *et al.* 2006). This provides additional support for the hypothesis
447 that the family effect is driving the separate clusters. However, it is not clear why related individuals were
448 specifically sampled in this area through the opportunistic sampling method applied.
449 California harbour seals are managed separately from Oregon and coastal Washington State, but
450 this is due to jurisdictional or political separation rather than due to genetic information (Brown *et al.* 2005).
451 In the present results, the level of genetic differentiation between ORE and CAL (i.e., F_{ST} = 0.008-0.011)

was lower than that between NBC and SOG (i.e., F_{ST} = 0.016-0.021), or EQB and NFL (i.e., F_{ST} = 0.019-0.024). The NBC allele frequencies should be taken with caution however, given the low sample size, and the opportunistic sampling method applied. Given the expected gene flow between Oregon and neighbouring states, this may explain how the reduction to approximately 500 harbour seals in Oregon before protections were implemented in the early 1970s (reviewed by Brown *et al.* 2005) did not result in a suppression of genetic diversity in the contemporary population. Similarly, in sea otter *Enhydra lutris*, simulations suggest that accumulated genetic load through bottlenecks and low Ne decreases more swiftly with restoration of gene flow (Beichman *et al.* 2023). As described above, the demographic history and Ne fluctuations of Pacific and Atlantic subspecies, including within each basin, would be a valuable approach to understand how recorded historic events compare with the signatures observed in the genetics. Undertaking an analysis of contemporary demography alongside this information would be expected to further improve the resolution and information for the management of the subspecies (Olsen *et al.* 2014).

Marker selection for an amplicon panel and future work

Marker panels, such as a low-density 500 marker panel, are a highly useful tool for conservation genetics (Meek & Larson 2019), including for characterizing population structure, conducting genetic assignments and parentage, as well as for providing some estimates of genetic diversity. With the present data, markers identified here could be used to generate such a panel. To select highly differentiating markers that will separate populations along the range of the Pacific harbour seal, markers should be selected based on high F_{ST} both between Northern BC and SOG, and between SOG and Oregon/California. Additional goals of individual identification should be considered as well, and other markers with elevated heterozygosity will be useful in this regard to include in a panel (e.g., markers with elevated H_{OBS} in both the USA and Canadian populations). Removal of the highly related outlier cluster is likely to be a good approach for identifying markers, as was done here, but it is also likely valuable to include several of the high frequency private alleles identified in the Burrard cluster in the panel for broader interpretation. A caution regarding the removal of the putative close relatives is that we are unable to determine whether this is an isolated colony without significant gene flow to neighbouring colonies; the highly related nature of the animals may simply be a characteristic of an isolated population (Waples & Anderson 2017). Obtaining samples from through the Central Coast and North Coast may aid in determining whether a dividing line occurs in BC, as this has yet to be identified, including with the data generated here. These genetic efforts are expected to be useful for future planning of management units for the Pacific harbour seal.

Conclusions

The present study characterizes harbour seals collected from seven different regions in the east and west coast of North America and identifies 12,742 genetic variants. Top differentiating or high heterozygosity variants can be used for future studies, for example using an amplicon panel, to further characterize population structure, for parentage analysis, or to enable repeat detections of individuals. This study confirms the recent observation of elevated diversity in Pacific harbour seals over the western Atlantic subspecies, and verifies that this trend is not due to analytic artefacts such as reference bias, sequencing depth, or sample size. Continuous distributions in multivariate analyses were observed within the southern and northern BC samples (albeit with a low northern BC sample size), and within the Oregon and California samples, with no discrete separation observed within each country. Clusters of related individuals were observed in the PCA for Burrard Inlet and southern California, and as a result, putative relatives were removed from the dataset; this may have broader implications for future sampling efforts from colonies to characterize discrete populations. Additional samples, systematically characterized over time will be helpful to determine the extent of gene flow between colonies. Samples of putative parents would also help to resolve the nature of the relationships among samples by parentage analysis. Future studies using the multiple, independent datasets now available could explore the demographic history of both subspecies to determine timing of Ne fluctuations on either coast to further investigate the causes of the striking differences in genomic diversity between the subspecies.

Acknowledgements

This work was supported by Fisheries and Oceans Canada Pinniped Research Program and the Species At Risk Program. The authors are grateful to sample providers from a wide range of locations, including Garry Stenson of DFO Atlantic (Newfoundland and Labrador), Mike Hammill and Xavier Bordeleau of DFO Québec (eastern Québec), the Marine Mammal Rescue Centre, Vancouver and DFO Pacific (SOG, NBC), James Rice of the Oregon Marine Mammal Stranding Network (Oregon), Barbie Halaska of the Marine Mammal Center of Sausalito (California), and Lauri Jemison and Lori Polasek from Alaska Department of Fish and Game (ADFG). Thanks to the Molecular Genetics Lab of DFO Pacific for technical support and discussions on the results. Thanks to Drs. Thierry Gosselin, Morten Tange Olsen, an anonymous reviewer, and the subject Editor for their constructive comments on an earlier version of the manuscript.

Competing Interests

Ben Sutherland is affiliated with Sutherland Bioinformatics. The author has no competing financial interests to declare. The other authors declare that no competing interests exist.

Data Accessibility

All code required to analyze the project: https://github.com/bensutherland/ms_harbour_seal

Demultiplexed sequence data is available on NCBI through the SRA within BioProject PRJNA948428, BioSamples SAMN33902654- SAMN33902801.

Associated files on FigShare including compressed GitHub analysis pipeline, sample information file needed for *stacks_workflow*, genotypes for all populations from the reference-guided analysis (all samples or normalized) in genepop and VCF formats (both are single SNP per locus) as well as RADpainter input (microhaplotypes), and the balanced, normalized, *de novo* genotyping genepops (single SNP per locus) and VCFs (multiple SNP per locus): https://doi.org/10.6084/m9.figshare.22290472

Benefit-sharing statement: This research addresses the population genetics of harbour seals in the eastern Pacific and western Atlantic. We have ensured that all raw data, output genotype data, and metadata is provided, as well as well-documented scripts and instructions for data analysis. All individuals who provided samples for the project are included in the Acknowledgements section.

Supplemental Information

In addition to the Additional Files shown below, the Supplemental Results section contains Figure S1-S6 and Table S1 and Table S2.

Additional File S1. All sample metadata including location, sex, and age information.

Additional File S2. Detailed sample location map for southern BC, highlighting sampling locations for the Burrard Inlet cluster.

Additional File S3. Detailed sample location map for Eastern Quebec.

Additional File S4. Table summary stats with different analytic methods for testing reference bias

Additional File S5. Inter-individual relatedness as evaluated by shared microhaplotypes showing the strong dissimilarity between subspecies, and also clusters of highly related individuals.

Additional File S6. Pairwise relatedness statistics displaying histogram of all pairwise values and the top interindividual relatedness pairs for both Pacific and Atlantic datasets.

Additional File S7. Coast-specific analyses PCA scores, sample IDs, location, and GPS coordinates.

Additional File S8. Per locus marker statistics for (a) Pacific populations and (b) Atlantic populations, both with all putative relatives removed, as well as between (c) SOG and NBC and (d) SOG and ORE.

Figures and Tables

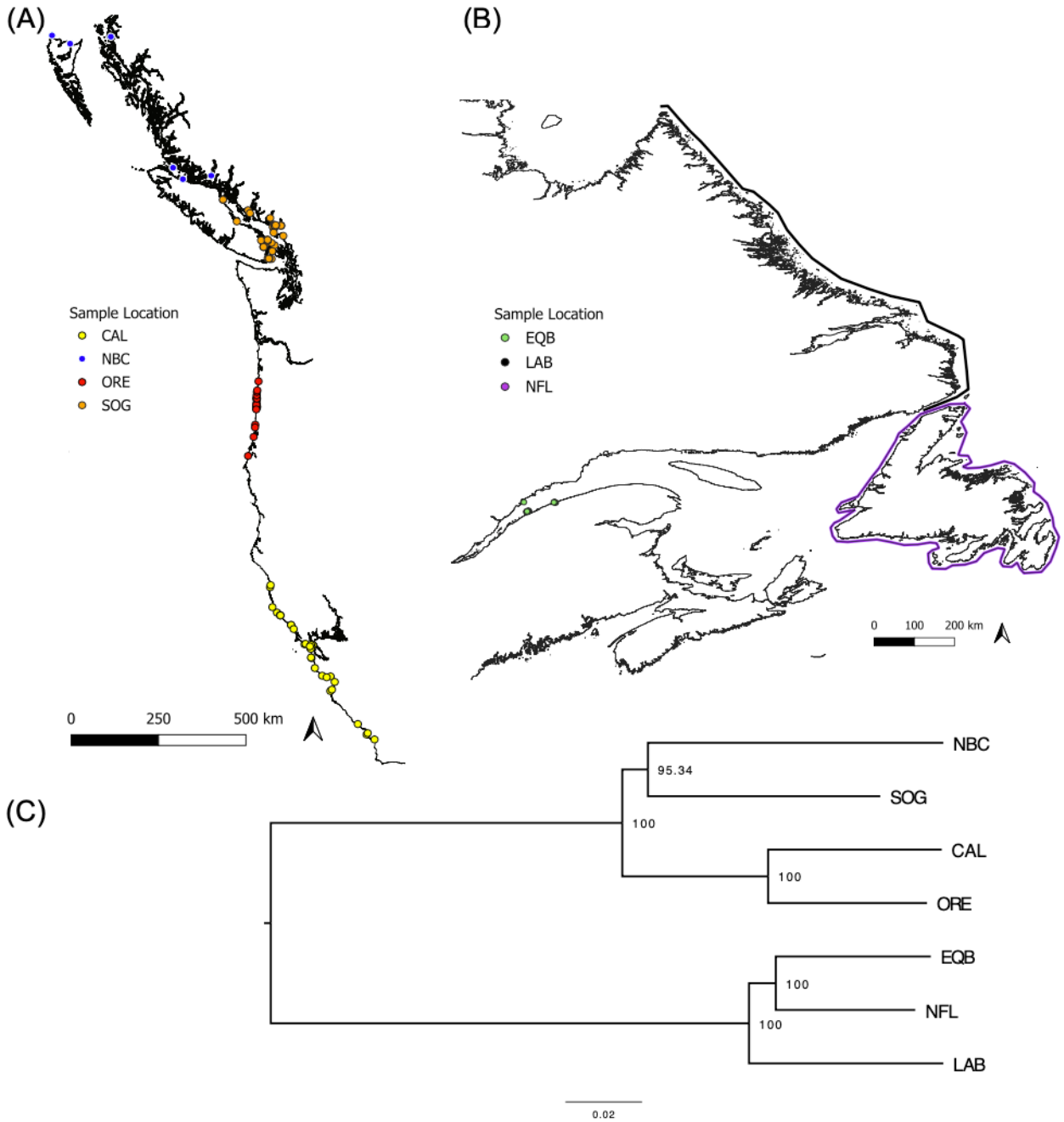

Figure 1. Sampling locations for the eastern Pacific (A), and western Atlantic (B), where each individual harbour seal sampled is shown with a coloured dot or the region highlighted with the respective population colour in cases where specific location information was not available for the sample. Genetic similarity among the populations is shown in the dendrogram (C) along with bootstrap support at the nodes. Acronyms: NBC = northern British Columbia; SOG = Strait of Georgia; CAL = California; ORE = Oregon; EQB = eastern Quebec; NFL = Newfoundland; and LAB = Labrador.

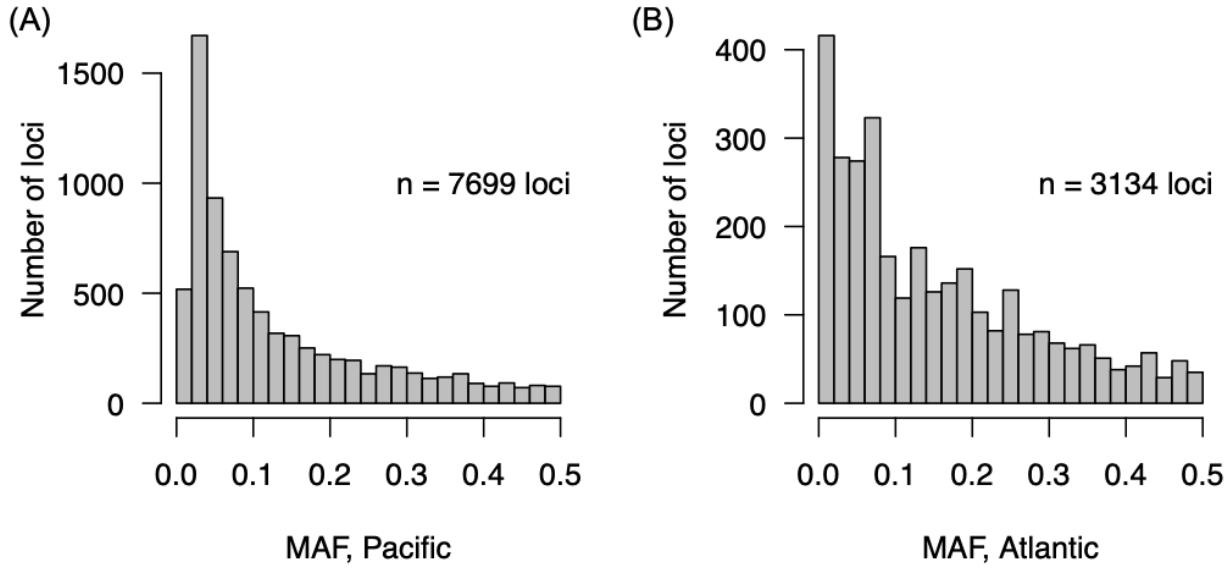

Figure 2. Minor allele frequency distribution for (A) Pacific or (B) Atlantic datasets after removing low MAF markers within each coast-specific dataset (i.e., retain MAF ≥ 0.01). The Pacific dataset is comprised of 56.3% of the markers between 0.01-0.10 MAF, whereas the Atlantic dataset is comprised of 46.5%.

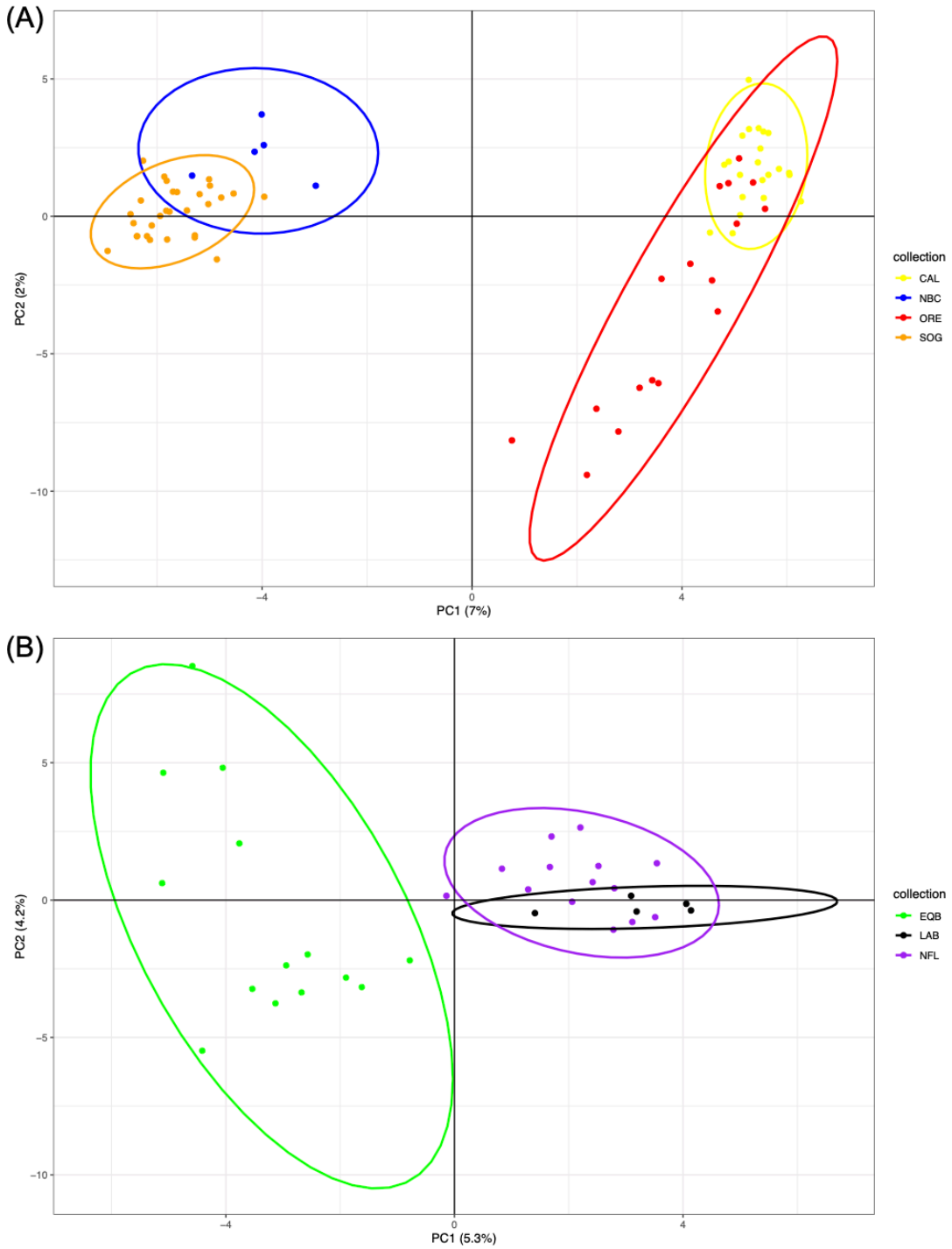

Figure 3. Principal components analysis following removal of putative close relatives for (A) Pacific subspecies (*P. v. richardii*) collections with 7,699 filtered variants; and (B) Atlantic subspecies (*P. v. concolor*) with 3,134 filtered variants. Some separation is viewed between collections within coastal British Columbia (i.e., SOG and NBC) as well as coastal United States (i.e., ORE and CAL), but significant overlap exists and the distribution of both appears continuous. Although LAB and NFL are completely overlapped, EQB is largely separate from the NFL grouping.

Table 1. Per population sample size (n), total reads (M) and per sample average (± s.d.), total alignments (M) and per sample average (± s.d.), and average alignment rate (%; ± s.d.). Values were calculated after extreme outliers (n = 2) and low coverage outliers were removed (n = 2).

Pop	n	Reads, total (M)	Per sample average	Alignments, total (M)	Per sample average	Alignment rate (%)
NBC	7	30.2	4.3 ± 3.0	27.5	3.9 ± 2.8	90.0 ± 2.2
SOG	41	138.6	3.4 ± 1.4	122.7	3.0 ± 1.2	88.5 ± 1.2
ORE	19	104.5	5.5 ± 2.4	99.5	5.2 ± 2.3	95.1 ± 0.8
CAL	30	108.7	3.6 ± 1.9	102.8	3.4 ± 1.8	94.1 ± 1.8
EQB	20	73.8	3.7 ± 2.0	63.9	3.2 ± 1.8	86.2 ± 2.0
LAB	7	29.0	4.1 ± 2.3	25.6	3.7 ± 2.1	87.3 ± 3.3
NFL	22	76.8	3.5 ± 2.0	66.5	3.0 ± 1.9	85.9 ± 2.7

Table 2. Summary statistics from genotyping 146 individuals at 12,742 variant sites in 4.6 Mbp indicates elevated polymorphism rate and observed heterozygosity (H_{OBS}) in the Pacific subspecies relative to the Atlantic subspecies. This dataset is not restricted to a single SNP per RAD locus. The differential genomic diversity is consistent with Liu et al. (2022) and was confirmed to not be due to reference bias (see Additional File S2). The number of close-relatives that were removed from each population is shown along with the number of loci removed in the final single-SNP per locus dataset that did not conform to Hardy-Weinberg proportions. The total variants remaining in Pacific and Atlantic after all filters were 7,699 and 3,134, respectively.

Pop	Avg. n per locus	Polymorph. variants	Percent polymorph. (%)	H_{OBS}	Nucleotide diversity (π)	Related cutoff (Ritland)	Related indiv. rem.	Loci non-HWP
NBC	6.8	6400	0.138	0.00044	0.00044	NA*	2	0
SOG	39.8	9018	0.194	0.00043	0.00044	0.1027	12	360
ORE	18.9	8358	0.180	0.00046	0.00046	0.0921	2	289
CAL	29.7	8528	0.184	0.00045	0.00046	0.0975	8	281
EQB	19.5	3875	0.083	0.00024	0.00024	0.1061	6	123
LAB	6.9	3257	0.070	0.00025	0.00024	0.0433	2	0
NFL	21.4	4044	0.087	0.00024	0.00024	0.0461	8	131

*no outlier pairs, applied same cutoff as SOG

Table 3. Genetic differentiation (Weir-Cockerham F_{ST}) between populations in the (A) Pacific subspecies dataset comprised of 7,699 variants; and in the (B) Atlantic subspecies dataset comprised of 3,134 variants. Values in the upper and lower sections are upper and lower 95% confidence limits of the estimated F_{ST}, respectively. Any negative values should be considered $F_{ST} = 0$, or no differentiation.

(A)

	NBC	**SOG**	**ORE**	**CAL**
NBC	-	0.021	0.052	0.064
SOG	0.014	-	0.057	0.071
ORE	0.043	0.052	-	0.009
CAL	0.055	0.065	0.006	-

(B)

	EQB	**NFL**	**LAB**
EQB	-	0.023	0.032
NFL	0.016	-	0.002
LAB	0.020	-0.007	-

REFERENCES

Andersen L, Olsen MT (2010) Distribution and population structure of North Atlantic harbour seals (*Phoca vitulina*). *NAMMCO Scientific Publications* **8**, 15-35.

Andersen LW, Lydersen C, Frie AK, Rosing-Asvid A, Hauksson E, Kovacs KM (2011) A population on the edge: genetic diversity and population structure of the world's northernmost harbour seals (*Phoca vitulina*). *Biological Journal of the Linnean Society* **102**, 420-439.

Andrews S (2010) FastQC: A quality control tool for high throughput sequence data.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLOS ONE* **3**, e3376.

Barrett RDH, Schluter D (2008) Adaptation from standing genetic variation. *Trends in Ecology & Evolution* **23**, 38-44.

Beacham TD, McIntosh B, Wallace CG (2011) A comparison of polymorphism of genetic markers and population sample sizes required for mixed-stock analysis of sockeye salmon (*Oncorhynchus nerka*) in British Columbia. *Canadian Journal of Fisheries and Aquatic Sciences* **68**, 550-562.

Beichman AC, Kalhori P, Kyriazis CC, DeVries AA, Nigenda-Morales S, Heckel G, Schramm Y, Moreno-Estrada A, Kennett DJ, Hylkema M, Bodkin J, Koepfli K-P, Lohmueller KE, Wayne RK (2023) Genomic analyses reveal range-wide devastation of sea otter populations. *Molecular Ecology* **32**, 281-298.

Blanchet M-A, Vincent C, Womble JN, Steingass SM, Desportes G (2021) Harbour seals: population structure, status, and threats in a rapidly changing environment. *Oceans* **2**, 41-63.

Bohling J (2020) Evaluating the effect of reference genome divergence on the analysis of empirical RADseq datasets. *Ecology and Evolution* **10**, 7585-7601.

Boness DJ, Bowen WD, Buhleier BM, Marshall GJ (2006) Mating tactics and mating system of an aquatic-mating pinniped: the harbor seal, *Phoca vitulina*. *Behavioral Ecology and Sociobiology* **61**, 119-130.

Brown RF, Wright BE, Riemer SD, Laake J (2005) Trends in abundance and current status of harbor seals in Oregon: 1977-2003. *Marine Mammal Science* **21**, 657-670.

Burg TM, Trites AW, Smith MJ (1999) Mitochondrial and microsatellite DNA analyses of harbour seal population structure in the northeast Pacific Ocean. *Canadian Journal of Zoology* **77**, 930-943.

Carroll EL, Hall A, Olsen MT, Onoufriou AB, Gaggiotti OE, Russell DJ (2020) Perturbation drives changing metapopulation dynamics in a top marine predator. *Proceedings of the Royal Society B: Biological Sciences* **287**, 20200318.

Cavalli-Sforza LL, Edwards AW (1967) Phylogenetic analysis. Models and estimation procedures. *American Journal of Human Genetics* **19**, 233-257.

Coltman DW, Bowen WD, Wright JM (1998) Birth weight and neonatal survival of harbour seal pups are positively correlated with genetic variation measured by microsatellites. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **265**, 803-809.

Cordes LS, O'Corry-Crowe G, Small RJ (2017) Surreptitious sympatry: Exploring the ecological and genetic separation of two sibling species. *Ecology and Evolution* **7**, 1725-1736.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H (2021) Twelve years of SAMtools and BCFtools. *GigaScience* **10**.

DFO (2022) Stock assessment of Pacific harbour seals (*Phoca vitulina richardsi*) in Canada in 2019. (ed. Canada FaO). DFO Can. Sci. Advis. Sec. Sci. Advis. Rep.

Elhaik E (2022) Principal component analyses (PCA)-based findings in population genetic studies are highly biased and must be reevaluated. *Scientific Reports* **12**, 14683.

Ewels P, Magnusson M, Lundin S, Käller M (2016) MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047-3048.

Goodman SJ (1998) Patterns of extensive genetic differentiation and variation among European harbor seals (*Phoca vitulina vitulina*) revealed using microsatellite DNA polymorphisms. *Molecular Biology and Evolution* **15**, 104-118.

Goudet J, Jombart T (2022) hierfstat: estimation and tests of hierarchical F-statistics.

Gruber B, Unmack PJ, Berry OF, Georges A (2018) dartr: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Molecular Ecology Resources* **18**, 691-699.

Härkönen T, Harding KC (2001) Spatial structure of harbour seal populations and the implications thereof. *Canadian Journal of Zoology* **79**, 2115-2127.

Hauser L, Baird M, Hilborn R, Seeb LW, Seeb JE (2011) An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (*Oncorhynchus nerka*) population. *Molecular Ecology Resources* **11**, 150-161.

Huber HR, Dickerson BR, Jeffries SJ, Lambourn DM (2012) Genetic analysis of Washington State harbor seals (*Phoca vitulina richardii*) using microsatellites. *Canadian Journal of Zoology* **90**, 1361-1369.

Jeffries S, Huber H, Calambokidis J, Laake J (2003) Trends and status of harbor seals in Washington State: 1978-1999. *The Journal of Wildlife Management* **67**, 207-218.

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

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

Knaus BJ, Grünwald NJ (2017) vcfr: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources* **17**, 44-53.

Lamont MM, Vida JT, Harvey JT, Jeffries S, Brown R, Huber HH, Delong R, Thomas WK (1996) Genetic substructure of the Pacific harbor seal (*Phoca vitulina richardsi*) off Washington, Oregon, and California. *Marine Mammal Science* **12**, 402-413.

Lepais O, Weir JT (2014) SimRAD: an R package for simulation-based prediction of the number of loci expected in RADseq and similar genotyping by sequencing approaches. *Molecular Ecology Resources* **14**, 1314-1321.

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM, p. arXiv:1303.3997.

Liu X, Fu Y-X (2020) Stairway Plot 2: demographic history inference with folded SNP frequency spectra. *Genome Biology* **21**, 280.

Liu X, Rønhøj Schjøtt S, Granquist SM, Rosing-Asvid A, Dietz R, Teilmann J, Galatius A, Cammen K, O'Corry-Crowe G, Harding K, Härkönen T, Hall A, Carroll EL, Kobayashi Y, Hammill M, Stenson G, Kirstine Frie A, Lydersen C, Kovacs KM, Andersen LW, Hoffman JI, Goodman SJ, Vieira FG, Heller R, Moltke I, Tange Olsen M (2022) Origin and expansion of the world's most widespread pinniped: Range-wide population genomics of the harbour seal (*Phoca vitulina*). *Molecular Ecology* **31**, 1682-1699.

Majewski SP, Ellis GM (2022) Abundance and distribution of harbour seals (*Phoca vitulina*) in the Strait of Georgia, British Columbia; synthesis of the 2014 aerial survey and long-term trends. (ed. DFO), p. iv + 59 p. Canadian Science Advisory Secretariat (CSAS).

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 3.

Meek MH, Larson WA (2019) The future is now: Amplicon sequencing and sequence capture usher in the conservation genomics era. *Molecular Ecology Resources* **19**, 795-803.

Muto MM, Helker VT, Delean BJ, Young NC, Freed JC, Angliss RP, Friday NA, Boveng PL, Breiwick JM, Brost BM, Cameron MF, Clapham PJ, Crance JL, Dahle SP, Dahlheim ME, Fadely BS, Ferguson MC, Fritz LW, Goetz KT, Hobbs RC, Ivashchenko YV, Kennedy AS, London JM, Mizroch SA, Ream RR, Richmond EL, Shelden KEW, Sweeney KL, Towell RG, Wade PR,

Waite JM, Zerbini AN (2021) Alaska marine mammal stock assessments, 2020. (ed. Commer. USD), p. 398. NOAA Tech. Memo. NMFS-AFSC.

NCBI (2023) SRA Toolkit. National Center for Biotechnology Information (NCBI).

Olesiuk PF (2010) An assessment of population trends and abundance of harbour seals (*Phoca vitulina*) in British Columbia (ed. Canada FaO). Canadian Science Advisory Secretariat, Nanaimo, B.C.

Olsen MT, Andersen LW, Dietz R, Teilmann J, Härkönen T, Siegismund HR (2014) Integrating genetic data and population viability analyses for the identification of harbour seal (*Phoca vitulina*) populations and management units. *Molecular Ecology* **23**, 815-831.

Olsen MT, Islas V, Graves JA, Onoufriou A, Vincent C, Brasseur S, Frie AK, Hall AJ (2017) Genetic population structure of harbour seals in the United Kingdom and neighbouring waters. *Aquatic Conservation: Marine and Freshwater Ecosystems* **27**, 839-845.

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

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genetics* **2**, e190.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLOS ONE* **7**, e37135.

Pew J, Muir PH, Wang J, Frasier TR (2015) related: an R package for analysing pairwise relatedness from codominant molecular markers. *Mol Ecol Resour* **15**, 557-561.

R Core Team (2023) R: A Language and Environment for Statistical Computing, Vienna, Austria.

Rijks JM, Hoffman JI, Kuiken T, Osterhaus ADME, Amos W (2008) Heterozygosity and lungworm burden in harbour seals (*Phoca vitulina*). *Heredity* **100**, 587-593.

Ritland K (1996) Estimators for pairwise relatedness and individual inbreeding coefficients. *Genetics Research* **67**, 175-185.

Rochette NC, Rivera-Colón AG, Catchen JM (2019) Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology* **28**, 4737-4754.

Schaeff CM, Boness DJ, Bowen WD (1999) Female distribution, genetic relatedness, and fostering behaviour in harbour seals, *Phoca vitulina*. *Animal Behaviour* **57**, 427-434.

Shafer ABA, Peart CR, Tusso S, Maayan I, Brelsford A, Wheat CW, Wolf JBW (2017) Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods in Ecology and Evolution* **8**, 907-917.

Stanley HF, Casey S, Carnahan JM, Goodman S, Harwood J, Wayne RK (1996) Worldwide patterns of mitochondrial DNA differentiation in the harbor seal (*Phoca vitulina*). *Molecular Biology and Evolution* **13**, 368-382.

Steinmetz K, Murphy S, Cadhla OÓ, Carroll EL, Onoufriou AB, Russell DJF, Cronin M, Mirimin L (2023) Population structure and genetic connectivity reveals distinctiveness of Irish harbour seals (*Phoca vitulina*) and implications for conservation management. *Aquatic Conservation: Marine and Freshwater Ecosystems* **33**, 160-178.

Swart JAA, Reijnders P, J. H. , Van Delden W (1996) Absence of genetic variation in harbor seals (*Phoca vitulina*) in the dutch Wadden Sea and the British Wash. *Conservation Biology* **10**, 289-293.

Tange O (2022) GNU Parallel 20220822. Zenodo.

Temte JL, Bigg MA, Wiig Ø (1991) Clines revisited: The timing of pupping in the harbour seal (*Phoca vitulina*). *Journal of Zoology* **224**, 617-632.

Thorburn D-MJ, Sagonas K, Binzer-Panchal M, Chain FJJ, Feulner PGD, Bornberg-Bauer E, Reusch TBH, Samonte-Padilla IE, Milinski M, Lenz TL, Eizaguirre C (2023) Origin matters: Using a local reference genome improves measures in population genomics. *Molecular Ecology Resources* **n/a**.

Trzcinski MK (2020) Synthesizing scientific knowledge about population dynamics and diet preferences of harbour seals, Stellar sea lions and California sea lions, and their impacts on salmon in the Salish Sea Workshop 2: November 20-21, 2019, Bellingham, WA. (ed. Canada FaO), p. x + 50 p.

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

Waples RS, Anderson EC (2017) Purging putative siblings from population genetic data sets: a cautionary view. *Molecular Ecology* **26**, 1211-1224.

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

Yao Y, Ochoa A (2023) Limitations of principal components in quantitative genetic association models for human studies. *eLife* **12**, e79238.