Population genomics of harbour seal *Phoca vitulina* from northern British Columbia through California and comparison to the western Atlantic sub-species

Ben J. G. Sutherland[1,2], 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 with important ecological roles as both predator and prey. Two sub-species of harbour seal 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 this species was recently reported, with highest diversity in the Pacific sub-species, 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, for example to provide genetic information for identifying management units. Here we use restriction-site associated DNA (RAD)-sequencing to genotype 146 harbour seals from western and eastern North America (i.e., northern and southern British Columbia (BC), Oregon, and California, Québec, Newfoundland, and Labrador). Using 12,742 identified variants, we confirm the elevated genetic diversity in the Pacific sub-species, and further validate that the difference between the sub-species is independent of reference bias or any other potential technical artefacts. Greatest differentiation occurs between the subspecies, followed by divergence between Canadian and contiguous US populations in the Pacific. Differentiation is also observed between Newfoundland/Labrador and Québec. Gradual separation in samples and lower levels of differentiation are viewed between California and Oregon, and between northern and southern BC. In a PCA based on genotypes, unexpected outlier clusters were observed and comprised of individuals from Burrard Inlet (BC) or southern California. These outliers were comprised of highly related individuals and contain high frequency private alleles. Further, within population spread was observed within the St. Lawrence estuary, where samples separated based on sampling from the Métis or Le Bic harbour seal colonies. The dataset here provides the substrate for future analyses on demographic inference of the harbour seal on both coasts, and identifying differentiating or high heterozygosity markers for high throughput genotyping platforms such as an amplicon panel. Future sampling of the BC central coast, and additional northern samples should help resolve the question of whether a discrete separating line exists between harbour seals in southern and northern BC.

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

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 sub-species 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 sub-species, estimated to have occurred 1.7-2.2 million years ago (Andersen & Olsen 2010). Secondly, strong differentiation exists between western and eastern sub-species within each ocean (Stanley *et al.* 1996). 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 finally the lowest in the eastern Atlantic Ocean (Liu *et al.* 2022), which are known to have experienced significant populations declines and show low genetic diversity (Swart *et al.* 1996).

Harbour seals have been challenged by habitat loss, interspecific competition, 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 sub-species 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) comprises 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 Europe from overhunting, environmental pollution, and epizootics, most notably in the Wadden Sea, may have resulted in bottlenecks and resultant low genetic diversity (Swart *et al.* 1996; Goodman 1998). Genetic diversity differences from the Pacific to the 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 are important to consider for species management and conservation.

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 (Blanchet *et al.* 2021). Genetic diversity has been directly connected to fitness in harbour seals. Lower heterozygosity was associated with increased lungworm burden in juvenile Wadden Sea harbour seals (Rijks *et al.* 2008). Birth weight and survival-to-weaning correlate with genetic diversity in harbour seal pups on Sable Island, Nova Scotia (Coltman *et al.* 1998). Populations in both of these locations are expected to have particularly reduced diversity relative to nearby populations and therefore to be more susceptible to inbreeding depression (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 (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).

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). Females only have a single pup per year (Schaeff *et al.* 1999), and they exhibit even more philopatry than males, with this behaviour increasing as they age (Härkönen & Harding 2001). From eastern North America to Europe, 12 regional populations have been identified, but the true number is likely greater (Andersen & Olsen 2010). Closer inspection identifies more fine-scale structure, for example the United Kingdom being delineated into northwest, northeast, and southeast populations (Olsen *et al.* 2017), and Ireland delineated into northwest/north, eastern, and southwest populations (Steinmetz *et al.* 2023). However, many areas remain to be analyzed at a fine-scale resolution. Currently, eastern North America is considered a single population, which fits with continuous clinal variation in pupping season (Temte *et al.* 1991). Western North America is considered to have three coastal populations: southeast Alaska through northern BC; southern BC; and Washington to California (Burg *et al.* 1999). Harbour seals in Puget Sound, Washington State are genetically distinct from the coastal population (Lamont *et al.* 1996). The inland waters of Washington State may in fact contain three discrete populations, including Puget Sound (Huber *et al.* 2012). 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 (reviewed in Andersen & Olsen 2010). In BC and Washington, perceived impacts of harbour seals on salmon have renewed discussions on regional population control measurements (Trzcinski 2020). Moreover, increasing interest in Food, Social or Ceremonial (FSC)

harvests in BC are compelling managers to consider regional harvest allocations, but there is incomplete information on stock structure or status. Subsequently in BC, harbour seals are managed as one stock (DFO 2022). A coastwide assessment of genetic population structure and diversity would support the definition of units relevant for management and conservation efforts (Waples & Gaggiotti 2006). Such an assessment requires a consistent and high throughput genotyping method, and baseline samples from natal sites, with caution to sample only 'inhabitants' and not 'visitors' (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 on the North American west coast (i.e., northern British Columbia (BC), Strait of Georgia (SOG), Oregon, and California), and east coast (i.e., Newfoundland, Labrador, and Eastern Quebec). We analyze population differentiation and diversity differences, as well as relatedness patterns within the samples, and put the results in the context of 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 sub-species 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; Table 1; Additional File S1). Serum samples were also obtained from harbour seals in Alaska, but extractions failed to yield sufficient DNA and therefore these 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 automated 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 sequence data 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 sub-species) 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 that were found to be extreme outliers in a downstream principal components analysis (PCA) were removed (i.e., NFL_102; NFL_121), and two samples with fewer than 1 M reads were removed (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 the effect of the reference genome, *de novo* genotyping approaches were conducted within Stacks2 for each coast, using a balanced design of 18 samples from each of SOG, ORE (west coast), and EQB, and NFL (east coast). 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

Observed heterozygosity (H_{OBS}), nucleotide diversity (π), and polymorphism rate were calculated considering all genotyped loci in Stacks2. For downstream analysis, a genepop file was generated using only 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 an unsupervised PCA was used to visualize general trends in the data. Relatedness between pairs of individuals was estimated using related (v.1.0; Pew *et al.* 2015). Genetic differentiation between populations was assessed by calculating 95% confidence intervals of F_{ST} (Weir & Cockerham 1984) in hierfstat (v.0.5-11; Goudet & Jombart 2022). Per-locus statistics H_{OBS} and average F_{ST} were calculated using the *summary* function of adegenet and the function *Fst* within pegas (v.1.1; Paradis 2010), respectively. Each locus was tested for significant deviation from Hardy-Weinberg proportions per population using the function *hw.test* of pegas, removing any locus ($p \leq 0.01$) that was non-conforming in any one population. Private alleles were identified using poppr (v.2.9.3; Kamvar *et al.* 2014). Analyses were conducted using functions within the repository *simple_pop_stats* with analysis R scripts available within *ms_harbour_seal* (see Data Availability).

Re-analysis of Liu et al. 2022

A recently published RAD-sequencing dataset (*Pst*I enzyme) investigating 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 mean heterozygosity was calculated by the Stacks2 populations module.

Results

Genotyping and genomic diversity differences

The 146 samples retained in the analysis had on average (± s.d.) 3.8 ± 2.0 M reads each and were comprised in total of 561.6 M reads (see Table 1 for totals per population). Missing data per individual were on average 2.9 ± 3.0% of the genotyped markers. No outlier samples were identified at this stage (but see Methods for earlier outlier removal). Alignment rates against the reference genome, which was assembled from a Pacific sub-species 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 comprised of 488.1 M reads with an average effective coverage of 27.5 ± 27.9x. 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. 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.

Polymorphism rates for collections with more than 15 individuals ranged from 0.180-0.194% for Pacific collections and 0.083-0.087% for Atlantic collections (see Table 2 for population-specific metrics). Observed heterozygosity (H_{OBS}) was 1.9-fold higher in the Pacific than in the Atlantic sub-species for these collections (mean: 0.00045 and 0.00024, respectively), consistent with recently reported results (Liu *et al.* 2022).

Data were separated into west and east coast samples to inspect sub-species trends more closely. Once separated, each coast was filtered again for MAF, conformity to Hardy-Weinberg proportions and $H_{OBS} < 0.5$ (see Methods), which resulted in retention of 7,707 variants in the Pacific and 3,204 variants in the Atlantic. Of these, 57.9% and 47.9% of variants were $0.01 < MAF < 0.10$ in the Pacific and Atlantic harbour seal datasets, respectively (Figure 1).

Consistent diversity difference independent of reference bias

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 sub-species (mean polymorphism rate = 0.183%), and 40,909 loci comprising 3.3 M sites and 3,483 variants for the Atlantic sub-species (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 heterozygosity than SOG (H_{OBS} = 0.00048 and 0.00039, respectively; Additional File S2).

A subset of the data from Liu et al. (2022) was also re-analyzed to confirm the trends viewed here. Using the reference-guided genotyping on the subset of the samples described (see Methods), Pacific H_{OBS} ranged from 0.00031-0.00034 and Atlantic was estimated at 0.00019 (i.e., 1.7-fold higher H_{OBS} in Pacific). The *de novo* approach applied to Liu et al. (2022) 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.

Genetic differentiation between populations

The two sub-species were found to be highly differentiated, separating across PC1 (33.5% of total dataset variation; Figure 2), similar to that recently reported (Liu *et al.* 2022). For example, CAL and NFL 95% confidence interval (c.i.) F_{ST} = 0.395-0.4144 (Additional File S3). The second largest source of variation separated the Pacific sub-species into northern and southern clusters (PC2 explained 3.4% of total dataset variation). In the coast-specific PCAs, following coast-specific MAF filters, 8,933 variants were retained in the Pacific sub-species (n = 97 individuals) and 3,828 variants in the Atlantic (n = 49).

The Pacific dataset shows clear differentiation between the southern and northern samples, explaining 6.7% 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} = 0.053-0.072; 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} = 0.016-0.021 and ORE-CAL F_{ST} = 0.009-0.012 (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 the most northern samplings within the SOG collection, including Nelson Island on the Sunshine Coast (SOG_103, SOG_118), and Anvil Island in Howe Sound (SOG_123). Samples NBC_112, NBC_111, and NBC_113, taken from Prince Rupert and northern Haida Gwaii are the most distant samples geographically, and the most distant samples along PC1 from the SOG samples (Figure 3A; Additional File S4). However, while some geographic trend is therefore apparent, given the existing characterized samples with only a handful of northern BC samples, there is no clear break yet observed in the data between the SOG and the more northern samples (Figure 3A).

The Atlantic sub-species analysis has a generally strong separation between the EQB and the LAB/NFL collections along PC1 (5% of the variation; Figure 3B), although this PC also separates many of the individual EQB samples as well. Interestingly, all of the EQB samples in the positive PC2 axis were sampled at or near the Métis harbour seal colony, and all those in negative PC2 were from the Le Bic colony, which appears to be explaining the separation of these samples across both PC2 and PC1. The LAB

297 and NFL collections were completely overlapping and have an F_{ST} value overlapping zero, suggesting that
298 these are not separate populations. The separation of EQB and NFL is however reflected in relatively high
299 F_{ST} between EQB and NFL (95% c.i. F_{ST} = 0.022-0.028; Table 3B).
300
301 *Outlier clusters, relatedness, and private alleles*
302 A cluster of SOG samples grouped away from the main grouping, and was comprised entirely of individuals
303 sampled from within Burrard Inlet (n = 9; SOG PC2 scores > 3.3; Figure 3A). By contrast, the main
304 grouping of SOG samples (SOG PC2 scores < 1.5; Figure 3A) was comprised of individuals from various
305 locations including East Coast Vancouver Island, Strait of Juan de Fuca, coastal Vancouver, and near the
306 Fraser River (see Additional File S4). Similarly, a cluster of CAL samples (n = 7; CAL PC2 scores > 3.5;
307 Figure 3A) grouped away from the main grouping, and was comprised of six individuals from the
308 southernmost end of the sampling range (i.e., Pismo Beach, latitude: 35.1422 to Monterey, latitude:
309 36.6010), and only one sample from further north (Sausalito; latitude = 37.8588). By contrast, the main
310 CAL cluster was comprised of samples from Moss Landing (latitude: 36.8044) to Fort Bragg (latitude:
311 39.4454) (Additional File S4). Therefore, the two notable outlier clusters were both comprised of samples
312 from restricted geographic locations.
313 The inter-individual relatedness values within a population for all Pacific harbour seals were
314 inspected, which indicated that outlier clusters were comprised of individuals with the highest relatedness
315 values in their respective collections. This was observed in both the microhaplotype analysis (Additional
316 File S5), and the SNP analysis (Figure 4; Additional File S6). Within the SOG comparisons with high
317 relatedness (Ritland > 0.10), there were 33 pairs and 18 unique individuals identified. Of these, 27 pairs
318 and nine unique individuals were from the Burrard Inlet cluster. Within the CAL comparisons, there were
319 20 pairs of individuals with high relatedness, 13 of which were pairs where both individuals were from the
320 southern CAL cluster, and four had a single individual from the southern CAL cluster. The shared
321 microhaplotypes analysis (Additional File S5) indicates two groups of highly related individuals within the
322 Burrard Inlet cluster, with one group comprised of SOG_136, 108, 125, and 131, and the second group
323 comprised of SOG_141, 135, 127, 115, and 102. Therefore, the outlier samples from Burrard Inlet and
324 southern CAL were comprised of harbour seals with elevated levels of relatedness. Although the individuals
325 in the Burrard Inlet and southern CAL outlier groups had higher relatedness than other samples in their
326 respective populations, they did not show significantly elevated inbreeding coefficients (F) for either SOG
327 (mean F ± s.d. in outliers: 0.16896 ± 0.066; all samples: 0.13681 ± 0.044) nor for CAL (mean F ± s.d. in
328 outliers: 0.0965 ± 0.0119; all samples: 0.0809 ± 0.0237; Additional File S7).
329 Private alleles were inspected for the Burrard Inlet cluster. Most notably, a private allele was
330 present in all nine individuals (i.e., 586731_19.04), two private alleles in each of six individuals (i.e.,

729912_41.01, 807825_8.01) and five individuals (i.e., 503353_64.04 and 839341_28.01), and six private alleles in four individuals. By comparison, when a haphazard selection of nine SOG (non-Burrard Inlet cluster) individuals was queried for private alleles, no private alleles were found in any more than three individuals, further indicating the uniqueness of the Burrard Inlet cluster.

Selecting markers for panel design

Markers that may be useful for characterizing population structure or individual identification were identified in the dataset. Elevated F_{ST} markers were identified between the NBC and SOG groups (excluding the Burrard Inlet cluster), between SOG and ORE (excluding Burrard Inlet cluster), and between the SOG and Burrard Inlet cluster (Additional File S8). Per locus H_{OBS} was analyzed for the USA and BC samples separately. Contrasting per-locus H_{OBS} between the USA or BC datasets indicates high correlation in H_{OBS} between the datasets (Figure 5B), and therefore indicating that markers can be selected with high H_{OBS} in both regions.

Discussion

Sub-species divergence and differential genetic diversity

Species conservation and management of harbour seal will likely depend on effective characterization of population structure and contemporary standing genetic variation (including functional variation), that is considered in the context of large-scale historical demographic events. In this study, we confirm the existence of elevated genetic diversity in the Pacific sub-species of harbour seal *P. v. richardii* over the Atlantic sub-species *P. v. concolor* as was recently reported by Liu and co-workers (2022). We identify the elevated diversity of the Pacific sub-species in new regions including British Columbia (BC) and Oregon, whereas the previous analysis included Alaska and California for the Pacific sub-species (Liu *et al.* 2022). Further, we evaluate and find that this trend is not caused by reference genome bias in genotyping, where individuals with increased divergence from the reference genome may exhibit biased population genetic metrics (Bohling 2020). Although reference bias is typically considered in the context of divergent species, given the high divergence between the sub-species here, it was important to test whether *P. v. concolor* metrics were impacted by aligning to the *P. v. richardii* genome. The diversity difference of around 1.8x higher diversity as evaluated by H_{OBS} in the Pacific over the Atlantic North American sub-species remains when genotyping without a reference genome (i.e., *de novo*) and with a normalized number of samples and reads.

The ultimate cause of the diversity difference was proposed by Liu and co-workers (2022) to be due to stepwise founder effects during colonization from the Pacific Ocean to the western Atlantic Ocean, and then from the western Atlantic Ocean to the eastern Atlantic Ocean, with a decrease in genetic diversity

365 at each step. The divergence of the sub-species is expected to have occurred between 1.7-2.2 million years
366 ago (Andersen & Olsen 2010). Proposed west to east colonization within each basin was supported by
367 previous work (Stanley *et al.* 1996). Stanley and co-workers (1996) proposed that the European population
368 may have been eliminated during the last Ice Age (~18,000 y.a.), and that it was recolonized following the
369 recession of the ice sheets. The demographic history of these populations will be important to explore, albeit
370 with caution given the susceptibility of demographic inference to artefacts from pipeline parameters (Shafer
371 *et al.* 2017), and the use of these multiple independent RAD-seq datasets may aid in this analysis. The
372 relatively deep sequencing and therefore high reliability genotypes provided in the present analysis, and the
373 numerous RAD-tags generated through restriction enzyme selection (Lepais & Weir 2014), is expected to
374 enable the analysis of the allele frequency spectra of the two sub-species to characterize fluctuations in
375 effective population size (Ne) and therefore approximate the timing of the reduced genetic diversity event
376 (e.g., Liu & Fu 2020). Uncertainty levels around the demographic inference approaches may not be robust
377 enough for conservation and management decisions, but rather applied towards general understanding of
378 putative and broad biogeographic hypotheses (Shafer *et al.* 2017).
379 The striking diversity differences between the sub-species may provide a model to study the effect
380 of environmental changes and adaptive potential in the two sub-species, and their responses to these
381 challenges. This may be particularly relevant at the edges of the ranges (Blanchet *et al.* 2021). Furthermore,
382 the large divergence between the Pacific and the Atlantic sub-species raises the question of whether the
383 sub-species are in fact separate species (Liu *et al.* 2022), and the very high F_{ST} observed was put into the
384 context of other marine megafaunal sub-species complexes. F_{ST} values observed by Liu et al. (2022) ranged
385 from 0.38 to 0.69, noting that this is higher than most other marine megafaunal sub-species to date. Our
386 results agree with these levels, and although we did not have samples from the eastern Atlantic Ocean, here
387 F_{ST} values ranged from 0.38 to 0.43 between the west and east coast of North America.
388
389 *Population structure in western North America*
390 Western North America is considered to be comprised of three coastal populations: (1) southeast Alaska
391 through northern BC; (2) southern BC; and (3) Washington through California (Burg *et al.* 1999). In the
392 present study, the separation between BC and Oregon/California is clear. Additionally, here some
393 separation is viewed among samples from southern BC and northern BC, and among samples from Oregon
394 and California, but there are no clear break points or strong separations observed between the putative
395 populations. Continuous distributions of breeding animals along coastlines has previously been described
396 to have little local substructuring, although a continued increase in differentiation occurs with distance
397 (Goodman 1998). In addition to the continuous distribution, substructuring may also be impeded with high
398 population numbers (Burg *et al.* 1999). The location of the 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 identified 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, although many of the samples taken throughout the SOG and Strait of Juan de Fuca (JdF) clustered together, a subset of individuals from Burrard Inlet clustered separately. This Burrard cluster was comprised of highly related individuals, which may explain clustering outside of the main grouping given that the animals would have overrepresentation of certain alleles (Elhaik 2022). The additional observation of high frequency private alleles, one of which was identified in all individuals of the cluster, may suggest some genetic separation of this group. In harbour seals, isolation-by-distance may not explain all genetic separation, and significant differentiation can occur within only a few hundred kilometers, potentially being driven by other demographic processes (Stanley *et al.* 1996; Goodman 1998).

Inland populations in Washington State have been found to be genetically distinct from the coastal population and were found to have the most diverse mitochondrial haplotypes (Lamont *et al.* 1996). These populations have different pupping times and cranial morphometrics, although are geographically close, and may be more closely related to harbour seals from northern populations, including the SOG (reviewed by Lamont *et al.* 1996). Late birthing of pups has been observed in Puget Sound and Vancouver Island that is discontinuous with clinal variation observed further south, again suggesting some unique characteristics in this area (Temte *et al.* 1991). Furthermore, Burg and co-workers (1999) identified a unique mitochondrial haplotype signature in harbour seals sampled from southern Vancouver Island, but did not see this group as a separate population in their microsatellite analysis. These unique signatures have been suggested to potentially originate from a separate colonization event for these populations (Lamont *et al.* 1996; Burg *et al.* 1999). Whether the Burrard cluster is simply an artefact of chance sampling of individuals with high relatedness skewing allele frequencies and enriching rare alleles (Elhaik 2022), or in fact is evidence of impeded gene flow from the Burrard colony or colonies, requires further study with

433 systematically obtained samples over time. Additional samples from the inland populations of Washington
434 State (Lamont *et al.* 1996) to compare with the Burrard cluster would be important to include in future
435 analyses.

436 A southern California cluster was also observed, and this group was also found to have high
437 relatedness. This provides additional support for the hypothesis that the family effect is driving the separate
438 clusters. However, it is not clear why related individuals were sampled in this area through the opportunistic
439 sampling method applied. California harbour seals are managed separately from Oregon and coastal
440 Washington State, but this is due to jurisdictional or political separation rather than due to genetic
441 information (Brown *et al.* 2005). In the present results, the level of genetic differentiation between ORE
442 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
443 EQB and NFL (i.e., F_{ST} = 0.019-0.024). The NBC allele frequencies should be taken with caution however,
444 given the low sample size, and the opportunistic sampling method applied. Given the expected gene flow
445 between Oregon and neighbouring states, this may explain how the reduction to approximately 500 harbour
446 seals in Oregon before protections were implemented in the early 1970s (reviewed by Brown *et al.* 2005)
447 did not result in a suppression of genetic diversity in the contemporary population. As described above, the
448 demographic history and Ne fluctuations of Pacific and Atlantic sub-species, including within each basin,
449 would be a valuable approach to understand how recorded historic events compare with the signatures
450 observed in the genetics.
451
452 *Marker selection for an amplicon panel*
453 With the present data, markers identified here could be used to generate a high throughput, low density
454 genotyping platform such as a 500-marker amplicon panel, an approach that is considered to be at the
455 forefront of conservation genetic approaches (Meek & Larson 2019). To select highly differentiating
456 markers that will separate populations along the range of the Pacific harbour seal, markers should be
457 selected based on high F_{ST} both between Northern BC and SOG, and between SOG and Oregon/California.
458 Additional goals of individual identification should be considered as well, and other markers with elevated
459 heterozygosity will be useful in this regard to include in a panel (e.g., markers with elevated H_{OBS} in both
460 the American and Canadian populations). Removal of the highly related outlier cluster may be a good
461 approach for identifying markers, but it is also likely valuable to include private alleles and some markers
462 that differentiate this potential breeding colony-specific signature from the rest of the SOG. Complete
463 removal of these individuals does not seem to be an appropriate approach, in particular if the highly related
464 individuals are simply a characteristic of the isolated population (Waples & Anderson 2017). For baseline
465 characterization, optimally the future sampling efforts should be directed towards sampling individuals with
466 some certainty of breeding colony, as per the example of sampling unweaned pups (Huber *et al.* 2012).

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 seven different collections 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, such as through an amplicon panel, to further characterize population structure or to enable repeat detections of individuals. This study confirms the recent observation of elevated diversity in Pacific harbour seals over the western Atlantic sub-species, 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 in each country as it is between the BC and the contiguous US populations. Notably, two outlier clusters of related individuals were identified, one from Burrard Inlet, and one from southern California. This may be due to enrichment of rare alleles due to sampling of related individuals near breeding colonies, and the genetic separation of individuals from different colonies with the PCA was also observed in the St. Lawrence estuary (Québec). Additional samples, systematically taken over time will be helpful to determine the extent of gene flow between colonies. Future studies using the multiple, independent datasets now available could explore the demographic history of both sub-species to determine timing of Ne fluctuations on either coast to further investigate the causes of the striking differences in genomic diversity between the sub-species.

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 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 Claire Rycroft and others at the Molecular Genetics Lab of DFO Pacific for technical support and discussions on the results.

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

Supplemental Information

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

Additional File S2. Table summary stats with different analytic methods for testing reference bias

Additional File S3. Full population-level F_{ST} comparison for all populations together from the reference-guided genotyping dataset.

Additional File S4. Coast-specific analysis PCA scores, sample IDs, location, GPS coordinates

Additional File S5. Microhaplotype figure

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. Inbreeding coefficient per sample

Additional File S8. Per locus marker statistics for (a) northern BC vs. southern BC, no Burrard Inlet; (b) Burrard Inlet vs. southern BC; (c) southern BC vs. Oregon, no Burrard Inlet.

Figures and Tables

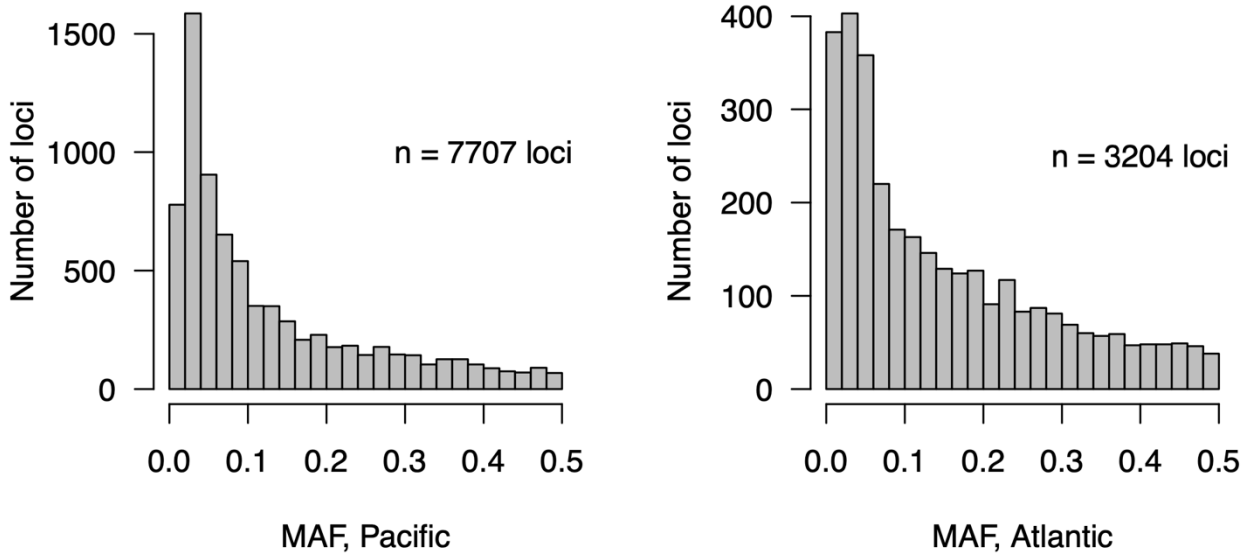


Figure 1. Minor allele frequency distribution for (A) Pacific or (B) Atlantic datasets after removing MAF < 0.01 markers within each dataset.

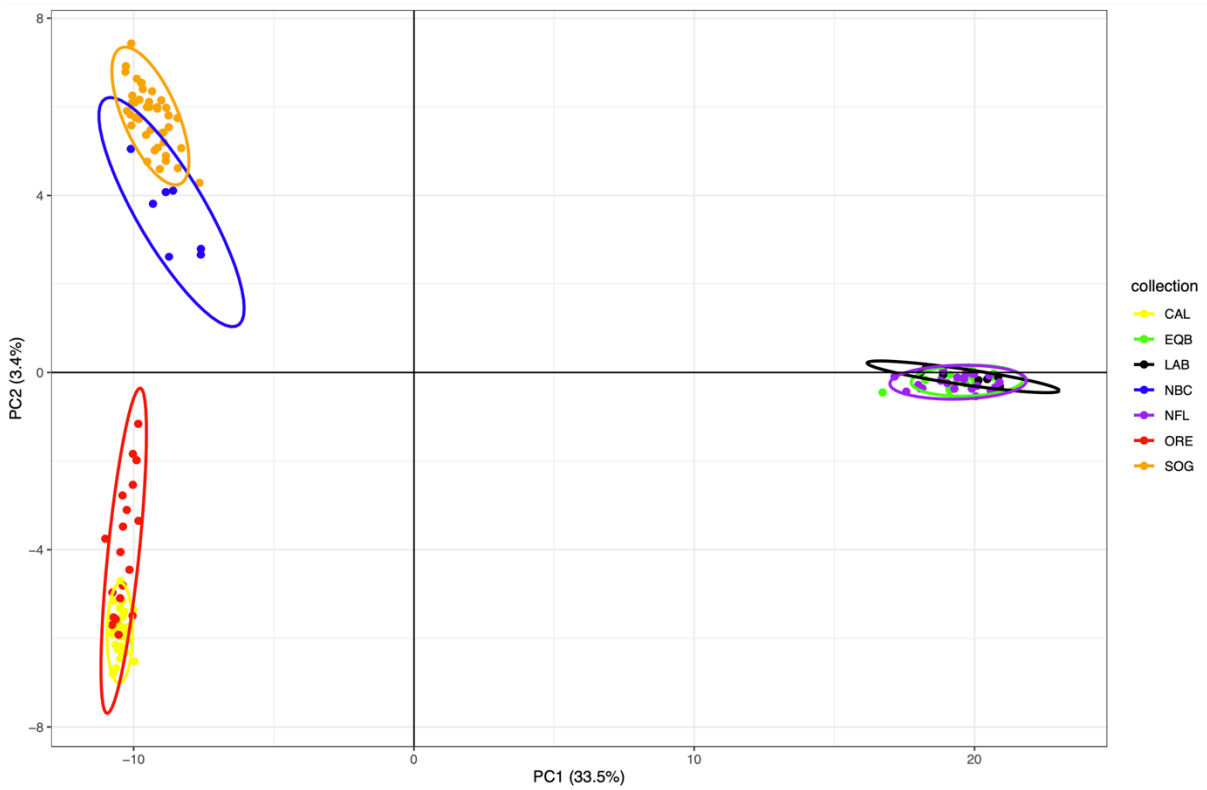
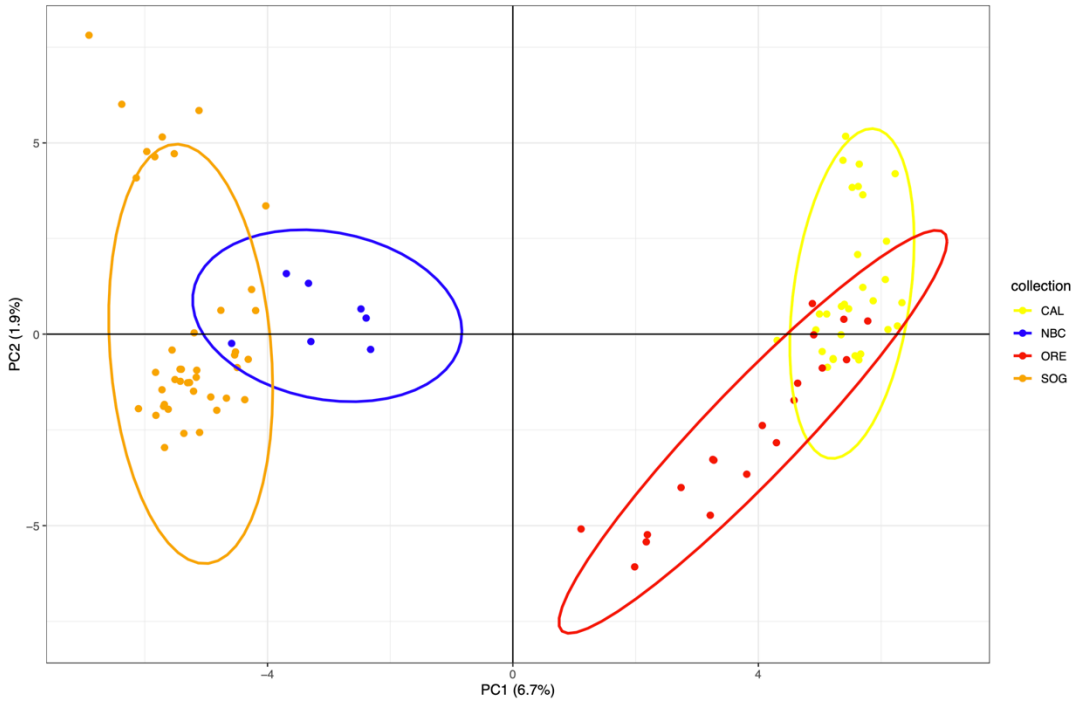
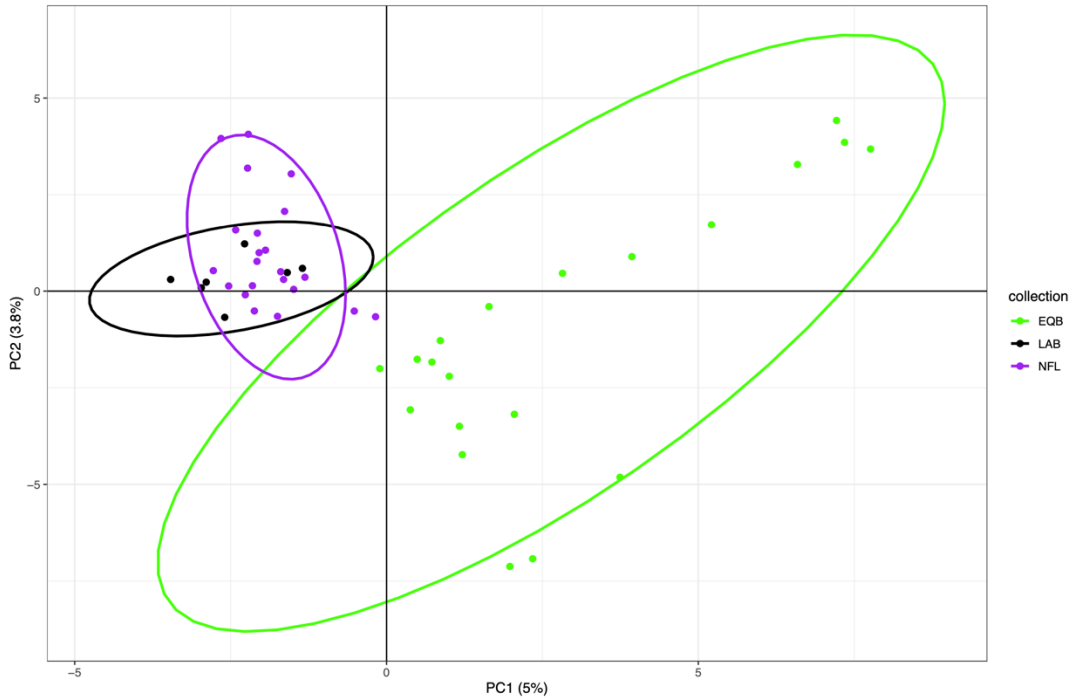


Figure 2. Global analysis of 10,847 SNP variants identified from all populations from the reference genome-guided genotyping dataset indicates strong divergence between Pacific and Atlantic harbour seal sub-species across PC1, explaining 33.5% of the variation. Additional divergence is observable in the Pacific between the more southern populations (California, Oregon) from the British Columbia populations across PC2, explaining 3.4% of the variation. Due to strong divergence, sub-species were analyzed separately after filtering each sub-species separately. Abbreviations: CAL = California; EQB = eastern Quebec; LAB = Labrador; NBC = Northern British Columbia (BC); NFL = Newfoundland; ORE = Oregon; SOG = Strait of Georgia, BC.

536 **(A)**



537
538 **(B)**



539
540 **Figure 3**. Principal components analysis of (A) Pacific sub-species (*P. v. richardii*) collections with 7,707
541 filtered variants; and (B) Atlantic sub-species (*P. v. concolor*) with 3,204 filtered variants. Some separation
542 is viewed between collections within coastal British Columbia (i.e., SOG and NBC) as well as coastal
543 United States (i.e., ORE and CAL), but significant overlap exists and the distribution of both appears
544 continuous. Although LAB and NFL are completely overlapped, EQB is largely separate from the NFL
545 grouping.

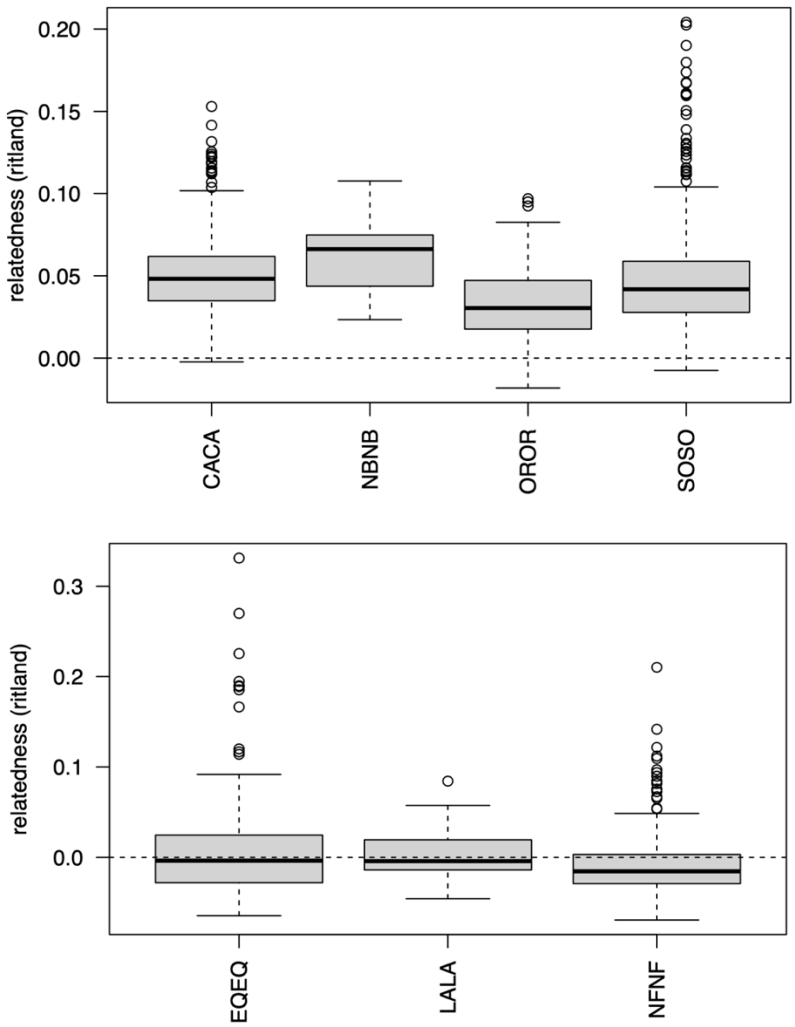


Figure 4. Genetic relatedness of individuals within each collection in the Pacific (A) and Atlantic (B) sub-species datasets calculated using 7,707 and 3,204 variants, respectively. Outliers are indicated as points outside of the 95% confidence limits of the boxplot. Abbreviations indicate same-on-same comparisons: CACA = California; NBNB = Northern British Columbia (BC); OROR = Oregon; SOSO = Strait of Georgia, BC; EQEQ = eastern Quebec; LALA = Labrador; NFNF = Newfoundland.

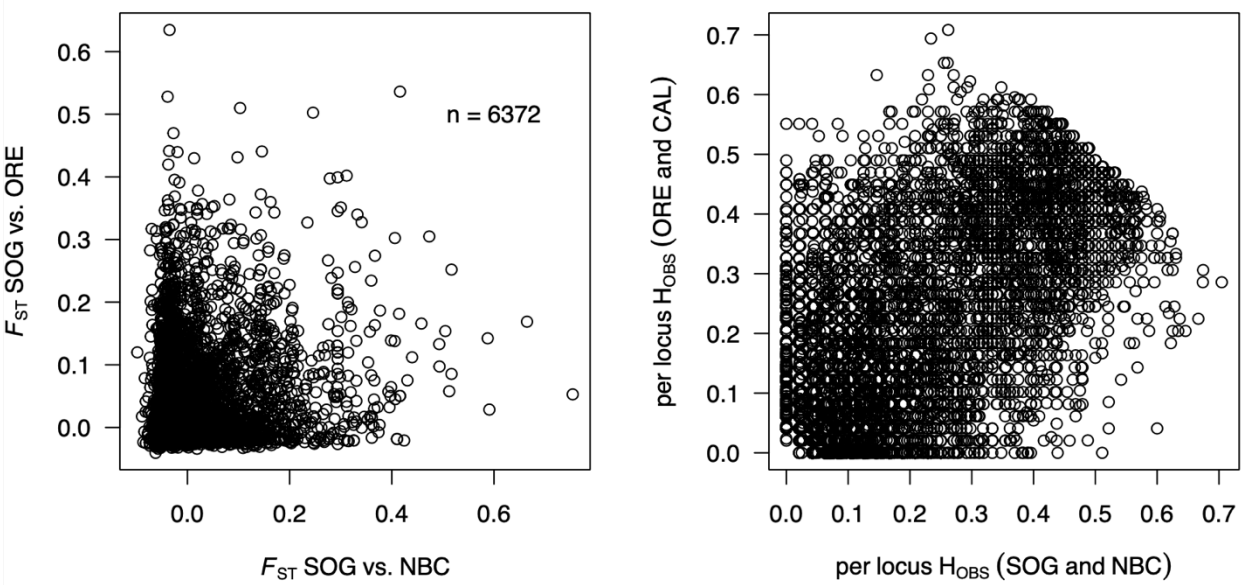


Figure 5. (A) Genetic differentiation (F_{ST}) per locus for contrasts of SOG vs. ORE, and SOG vs. NBC indicates that some markers have high F_{ST} in both comparisons. The Burrard Inlet outlier group was removed from this comparison, and both datasets were filtered for low MAF individually, resulting in a total of 6,372 comparisons. (B) Observed heterozygosity (H_{OBS}) per marker was calculated for the USA populations (ORE and CAL), and separately for the BC populations (SOG, including Burrard Inlet cluster and NBC), then compared to each other to identify elevated H_{OBS} in both regions.

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 sub-species relative to the Atlantic sub-species. 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).

Pop	Avg. samples per locus	Polymorphic variants	Percent polymorphic (%)	H_{OBS}	Nucleotide diversity (π)
NBC	6.8	6400	0.138	0.00044	0.00044
SOG	39.8	9018	0.194	0.00043	0.00044
ORE	18.9	8358	0.180	0.00046	0.00046
CAL	29.7	8528	0.184	0.00045	0.00046
EQB	19.5	3875	0.083	0.00024	0.00024
LAB	6.9	3257	0.070	0.00025	0.00024
NFL	21.4	4044	0.087	0.00024	0.00024

572 **Table 3.** Genetic differentiation (Weir-Cockerham F_{ST}) between populations in the (A) Pacific sub-
573 species dataset comprised of 7,707 variants; and in the (B) Atlantic sub-species dataset comprised of
574 3,204 variants. Values in the upper and lower sections are upper and lower 95% confidence limits of the
575 estimated F_{ST}, respectively. Any negative values should be considered $F_{ST} = 0$, or no differentiation.

576 **(A)**

	NBC	**SOG**	**ORE**	**CAL**
NBC	-	0.021	0.049	0.060
SOG	0.016	-	0.058	0.072
ORE	0.041	0.052	-	0.012
CAL	0.053	0.066	0.009	-

577

578 **(B)**

	EQB	**NFL**	**LAB**
EQB	-	0.028	0.037
NFL	0.022	-	0.005
LAB	0.027	-0.002	-

579

REFERENCES

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

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.

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.

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.

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.

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.

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, 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.

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.

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.

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.