Conservation Genetics Resources (2022) 14:69–80
https://doi.org/10.1007/s12686-021-01237-0

METHODS AND RESOURCES ARTICLE

Evaluation of novel genomic markers for pedigree construction in an isolated population of Weddell Seals (*Leptonychotes weddellii*) at White Island, Antarctica

Joshua M. Miller[1] · Erin O. Campbell[1] · Jay J. Rotella[2] · Kaitlin R. Macdonald[2] · Thomas S. Gelatt[3] · Corey S. Davis[1]

Received: 12 May 2021 / Accepted: 6 September 2021 / Published online: 23 September 2021

Abstract
Pedigrees have a long history in classical genetics, agronomics, evolutionary ecology, and ex situ captive breeding. Use of molecular techniques has expanded the variety of species for which pedigrees can be constructed. However, molecular pedigrees almost exclusively consider microsatellite loci, despite advances in high-throughput sequencing allowing development of genomic marker sets in nearly any organism. Here we generate a novel set of genomic SNPs derived from ddRAD sequencing in two populations of Weddell seals (*Leptonychotes weddellii*) and describe the diversity and differentiation between them. We then compare and contrast parentage assignment rates and accuracy in one population that has been the subject of long-term monitoring. Specifically, we consider pedigrees constructed using two sets of markers (microsatellites and SNPs), two pedigree construction software (CERVUS than Sequoia), as well as varying the groupings of candidate parents (either all individuals simultaneously, only individuals born before a focal year, or only individuals known to have survived to a focal year). ddRAD sequencing returned between 1568 and 3240 loci depending on whether both populations were considered simultaneously or individually. Parentage assignment rates were always higher using CERVUS than Sequoia, with the latter at times either not assigning parentage or creating "inferred parents". In all cases, "polarizing" the datasets (e.g., including year of birth) significantly improved assignments. This represents one of the first direct comparisons of pedigree construction using different markers in the same set of individuals, and the SNPs described here will be a resource for continued pedigree construction, and future research in Weddell seals.

Keywords SNPs · ddRAD · CERVUS · Molecular pedigree

Introduction

Pedigrees have long been used in classical genetics, agronomics, evolutionary ecology, and ex situ captive breeding efforts. They are the basis for classical trait mapping (Lynch and Walsh 1998), have been used to increase yields in crops and livestock (Piepho et al. 2008; Mrode and Thompson 2014), and are used to manage stud books for domesticated and companion animals (Leroy 2011). Pedigrees also play a key role in conservation efforts (Nielsen et al. 2007; Fienieg and Galbusera 2013; Miller et al. 2015; Jiménez-Mena et al. 2016) including equalizing family sizes among individuals in ex situ populations to prevent adaptation to captivity (Allendorf 1993; Williams and Hoffman 2009).

Although most often applied to model organisms or those in captivity, pedigrees have also been constructed for a variety of wild populations (Pemberton 2008; Jones and Wang 2010). Such efforts were aided by the advent of affordable genetic tools coupled with statistical software for parentage assignment (Jones et al. 2010; Flanagan and Jones 2019). Unlike pedigrees based solely on field observations, molecular pedigrees can assign parentage in species with no parental care, as well as highlight incorrect assignments due to multiple mating, nonmonogamy, or cryptic female choice

✉ Joshua M. Miller
jmm1@ualberta.ca

[1] Department of Biological Sciences, University of Alberta, Edmonton, AB, Canada

[2] Ecology Department, Montana State University, Bozeman, MT, USA

[3] NOAA Fisheries, Alaska Fisheries Science Center, Seattle, WA, USA

(Reid et al. 2014; Farquharson et al. 2019). Pedigrees in wild populations allow for examination of questions relating to evolutionary ecology, including mating system dynamics (Szulkin et al. 2013; Richardson et al. 2020) and monitoring of inbreeding (Keller 1998; Malenfant et al. 2016; Chen et al. 2018), as well as questions related to quantitative genetics, such as estimating heritability (Kruuk 2004; Malenfant et al. 2018) and genetic mapping of traits through linkage mapping (Backström et al. 2006; Jaari et al. 2009; Poissant et al. 2012; Nietlisbach et al. 2015). In the latter cases, pedigrees are especially important for species where controlled crosses are not feasible due to litter/clutch size, longevity, or conservation concerns.

However, use of pedigrees in wild populations has been relatively rare as their construction can be logistically complicated, usually requiring long-term monitoring to collect samples and identify candidate parents (Pemberton 2008; Jones et al. 2010). In the absence of long-term monitoring, attempting to place individuals into age groups or cohorts (e.g., year of birth or known recruitment into the population) can help “polarize” the data, thereby reducing and focusing the pool of candidate parents making the calculations more powerful and tractable. Intuitively, some of these factors could be mitigated by studying small, isolated populations where the pool of candidate parents is reduced relative to large, outbred populations. However, pedigree construction in such populations is challenging given that these small populations often have reduced genetic diversity and increased relatedness among individuals, both of which result in the need for more genetic markers to achieve the same amount of power as in a large, outbred population.

Molecular pedigrees have traditionally been built using data from microsatellite loci given their high allelic diversity and thereby power, abundance in the genome, and ability to be applied across species (Jones and Wang 2010; Guichoux et al. 2011; Flanagan and Jones 2019). However as genomic techniques have continued to improve, allowing rapid discovery and typing of 100 s to 1000 s of single nucleotide polymorphism (SNP) loci, using SNPs for pedigree construction has become feasible (Anderson and Garza 2006; Anderson 2012; Huisman 2017; Flanagan and Jones 2019). Although information content per SNP is lower, orders of magnitude more markers returned by most genomic methods could lead to more confident assignments, and improved estimates of relatedness to infer distant relationships (e.g., half-sibs) in the absence of complete sampling. As such, new software for SNP-based pedigree construction have been developed (Anderson 2012; Huisman 2017).

Here we present the development and application of SNP markers for Weddell seals (*Leptonychotes weddellii*) from White Island (WI) and Erebus Bay (EB) in the Ross sea, Antarctica (Fig. 1). These populations represent a unique study system where the WI population is a small, isolated group that was established after as few as five individuals from EB were cut off from the larger population by a large expanse of sea ice and has been the subject of long-term monitoring (Stirling 1972; Testa and Scotton 1999). Specifically, we compare the diversity and divergence between the two populations using newly developed SNP loci, update the WI pedigree with 17 years of new samples collected between 2001 and 2017, and compare and contrast pedigree construction for the WI population using SNP loci and the set of microsatellite loci previously used. This is the first comparison of pedigree construction using two different marker sets in the same individuals. We also examine how results differ when construction incorporates covariates which can “polarize” the pedigree compared to when “unpolarized” data is used. The latter representing similar conditions to pedigree construction from a single sampling event.

Methods

Study system

The population of Weddell Seals at WI was hypothesized to have been founded in the 1950’s by immigration of individuals from the adjacent population at EB (Fig. 1). Despite their physical proximity, intensive mark-recapture work each year at EB and typically two annual visits to WI (Siniff et al. 1977; Rotella et al. 2016; Paterson et al. 2018), migration between EB and WI has only been observed once (one immature female born in EB was sighted at WI in one year), and is typically prevented due to the ice shelf blocking passage to open water in Erebus Bay. Since the late 1960s the WI population has been the subject of continuous monitoring (Stirling 1972; Testa and Scotton 1999). Beginning in 1993 the entire adult population has been individually marked, allowing collection of detailed life-history metrics and construction of a microsatellite-based pedigree (Gelatt et al. 2010).

Sample collection and DNA extraction

Samples were collected as in Gelatt et al. (2010). Briefly, trips from McMurdo to White Island were made 2–4 times a season (more commonly two trips in recent years) between late October and mid-February. This period corresponds to when adult females and pups are observed together on the surface of the fast ice. Seals were uniquely marked with livestock ear tags in the rear flippers. All tissue samples were collected under Marine Mammal Protection Act Permits and Antarctic Conservation Act Permits and with methods approved by Montana State University’s Institutional Animal Care and Use Committee. Total genomic DNA was extracted from tissue samples collected

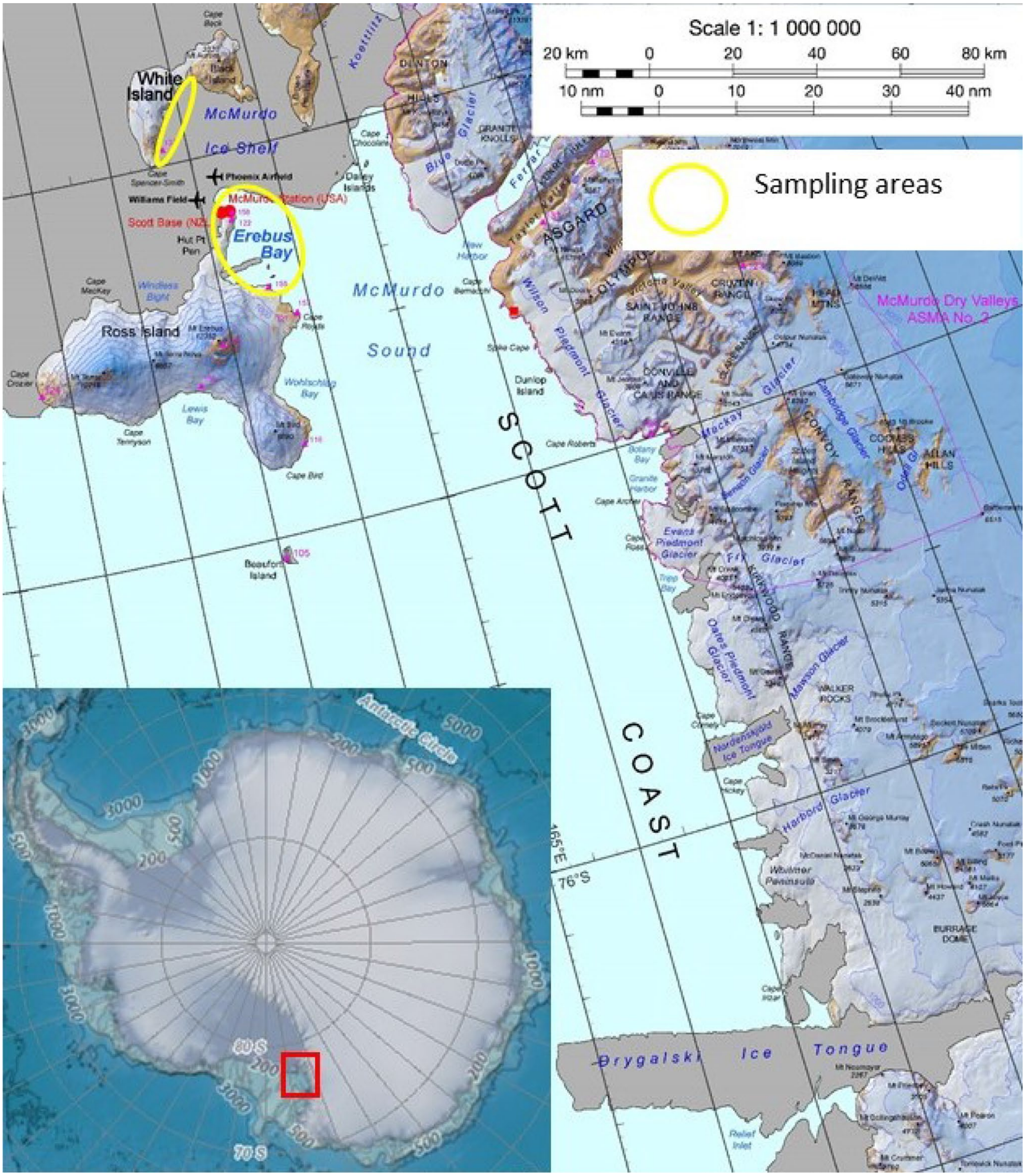

Fig. 1 A map of the study area. The detailed map was adapted from: Polar Geospatial Center, 2018, "PGC Map Catalog", https://doi.org/10.7910/DVN/6R8F7U, Harvard Dataverse, V1, Map 16: Victoria Land, August 16, 2021. The inset Antarctic map was obtained from: NOAA National Centers for Environmental Information (NCEI); International Bathymetric Chart of the Southern Ocean (IBCSO); General Bathymetric Chart of the Oceans (GEBCO)

during tagging using DNeasy spin columns (QIAGEN) following the manufacturer's protocols.

Microsatellite genotyping

For microsatellite-based pedigree construction we genotyped the newly collected samples at 41 loci as in Gelatt et al. (2010). Briefly, multiplex PCRs were performed as previously described (Gelatt et al. 2010) and resolved on a 3730 DNA Analyzer (Applied Biosystems). Genotypes were then scored with GeneMapper version 2.0 (Applied Biosystems).

ddRAD genotyping

Double digest restriction-site associated DNA sequencing (ddRAD; Peterson et al. 2012) libraries were prepared for both EB (N = 29) and all WI samples for which DNA was available (N = 144). Library construction followed the protocols outlined by MacDonald et al. (2020). However, in our case the restriction enzymes used were SbfI and EcoRI. We sequenced two pooled libraries containing 168 and 32 individually barcoded samples with single-end, 75 bp sequencing on a high output flowcell of an Illumina NextSeq 500. Here the latter library contained novel individuals as well as individuals included in the first run but for which additional sequencing depth was desired.

Following sequencing, reads that failed Illumina chastity filtering were removed, and then demultiplexed using STACKS 2.0b (Catchen et al. 2011, 2013). Adapters were removed and reads were quality trimmed using cutadapt version 1.9.1 (Martin 2011). Specifically, we trimmed the 5' end of the demultiplexed reads to remove the PstI cut site, as well as to remove any remnant Illumina adapter sequence. Reads from each individual were aligned to the Weddell Seal reference genome (LepWed1.0, GenBank assembly accession: GCA_000349705.1) using the mem algorithm in bwa 0.7.17 (Li and Durbin 2009, 2010) with default parameters. SNPs were then called using the ref_map pipeline within STACKS 2.0b, run on ComputeCanada Cedar cluster. When calling SNPs we considered four different sets of individuals: (1) All WI and EB (172 individuals; one WI individual was removed following genotyping thresholds, see Results); (2) WI only (143 individuals); (3) the "founding generation" of WI (18 individuals); and 4) EB only (29 individuals). Across all sets, we filtered loci to include only those that had a minimum minor allele frequency (MAF) of 0.01, minimum genotype quality score of 30, and were present in 80% of individuals in a given population. We output a single SNP per RAD-tag to reduce linkage within a tag, though acknowledge there may be linkage among tags.

Diversity and divergence between EB and WI using SNP loci

For each of the datasets outlined above, we calculated per-population allelic diversity, heterozygosity, and G_{is} (an estimate of inbreeding), as well as a measure of differentiation between populations (F_{st}) when appropriate (Weir and Cockerham 1984). All calculations were conducted in GenoDive version 2.0b27 (Meirmans and Van Tienderen 2004) using default settings. For the all WI and EB dataset we conducted a principal component analysis (PCA) of allele frequencies among individuals to visualize differentiation between the populations, and search for evidence of migrants (Patterson et al. 2006). Previous research has shown that the filtering parameters used when generating RAD genotypes can influence estimates of diversity and divergence (Paris et al. 2017; Shafer et al. 2017; Rochette and Catchen 2017). These parameters include the within population minor allele frequency and the minimum number of populations in which a locus must be present. Given the large sample size difference between WI and EB we generated several subsets of the full dataset with equal sampling between the two populations to more directly compare diversity and divergence. Specifically, we generated 10 datasets containing a random 18 individuals from EB and the original 18 founders of the WI population. In theory the WI founders represent a random subset of unrelated EB individuals that should capture diversity in the population before the effects of genetic drift and inbreeding in subsequent generations. We then recalculated genetic diversity and divergence statistics for each of these subsets using the same methods as above.

Pedigree reconstruction

When building the pedigree for WI we initially updated the existing pedigree with the newly collected samples using the microsatellite loci and previously described methods (Gelatt et al. 2010). Briefly, parentage was assigned using CERVUS (Marshall et al. 1998; Kalinowski et al. 2007) where the candidate parent file was adjusted for each cohort to include only individuals that had reached breeding age (at least 4 years) and were not assumed to have died (4 years since last sighting/inference of parentage) (Gelatt et al. 2010). Maternal assignments were verified against recorded mother–pup pairs at time of tagging.

We then compared and contrasted pedigrees constructed using the two sets of markers (microsatellites and SNPs) and two pedigree construction software (CERVUS and Sequoia) as applied to all WI samples. CERVUS (Marshall et al. 1998; Kalinowski et al. 2007) calculates log-likelihood ratios for trios between a focal offspring and candidate parents. However, CERVUS is not able to explicitly incorporate covariates to aid in pedigree construction, such as cohort or year of birth. Sequoia (Huisman 2017) is also a likelihood-based pedigree construction method, but was explicitly made to consider large panels of SNP markers and incorporate covariates such as year of birth to aid in assignments. In addition, the program will attempt to assign full or half siblings through generation of unsampled (inferred) parents (Huisman 2017). However, Sequoia cannot analyze microsatellite data. While not designed for large sets of SNP loci, CERVUS can analyze such datasets. Therefore, we applied CERVUS for both microsatellite and SNP based pedigree construction.

Molecular markers used in CERVUS were either all microsatellite loci (N = 41), or the SNP loci discovered in WI that were genotyped in 90% of individuals and had a MAF greater than or equal to 0.01 (N = 1303). For the SNP analyses, individuals had to have less than 25% missing data to be retained (N = 142). We performed parentage assignment in three ways using CERVUS: (1) Considering all individuals simultaneously (All); (2) In a sequential fashion where the candidate parent file was adjusted to include all "older" individuals, e.g. for pup born in 1999, all individuals born before 1998 were included as parents (Older); and (3) In a sequential fashion where the candidate parent file was adjusted to include only individuals that had reached breeding age (at least 4 years old) and were not assumed to have died (4 years since last sighting/inference of parentage) (Gelatt et al. 2010); (Survivor). These last two methods were undertaken to account for the fact that CERVUS cannot "polarize" assignments through time when all individuals are considered simultaneously. Such lack of polarization does not account for births or death of candidates, and therefore may miss-assign relationships among closely related individuals. For all three methods allele frequencies (for both microsatellites and SNPs) were calculated from all individuals sampled as adults in the population (N = 26) and used in simulations of 10,000 offspring from 82 females and 55 males (maximum number of individuals at least 1 year old in 1997) with the empirical proportion of loci typed (0.998 for microsatellites and 0.983 for SNPs) with 1% of loci mistyped and a 1% likelihood error. Assignments were assessed for offspring-mother-father assignments using LOD scores and a 99% confidence level for assignment.

For pedigree construction using Sequoia we created two sets of SNP loci. The first was as described above for use in CERVUS, and the second had the MAF threshold raised to 0.3 as recommended in the documentation for Sequoia (Huisman 2017). For both sets of loci we implemented a single assignment method where all individuals were considered simultaneously, incorporating sex and year of birth. We increased the number of iterations for sibship clustering to 10 and set the log10-likelihood ratio (LLR) threshold for differentiating between a proposed relationship versus unrelated to − 20. These analyses used Sequoia 1.3.3 in R 3.6 (R Core Team 2019).

Following pedigree construction, we compared the assignments made among the two marker types and two programs in terms of the number of assignments and the number of concordant assignments among the datasets and methods.

Results

Sample collection, microsatellite genotyping, and ddRAD genotyping

Between 2001 and 2017 67 pups were observed in WI. Tissue samples were collected from 61 of these, and for 5 out of the 6 pups without tissue samples, mothers could be assigned based on field observations.

Microsatellite genotyping of the newly collected samples resulted in nearly complete genotypes for all loci in all individuals. When combined with the long-term database there were only 6 missing genotypes across 145 individuals and 41 loci (0.1% missing data). Only one individual is missing data at more than one locus (4 loci).

After demultiplexing the two sequencing runs produced a total of 290,242,376 reads. Post quality filtering 269,029,186 reads were retained with a mean of 1,564,123 reads per individual (SD = 1,190,225) among 172 individuals (one individual from WI being removed by quality filters in STACKS).

Diversity and divergence between EB and WI

When all individuals were considered together 2096 SNP loci were retained following filtering (Table 1). In this sample allelic diversity was higher in EB (1.352) than WI (1.337), as was observed heterozygosity EB = 0.232 vs WI = 0.223. Similarly, G_{is} in EB was larger and not significantly different than 0, whereas in WI the estimate was negative including 95% CI. F_{st} values between the two populations was 0.140, with this differentiation reflected in the PCA which clearly separated the two populations with no evidence of migrants (Fig. 2).

When each population was analyzed separately the number of SNPs discovered was higher in EB and lower in WI (Table 1). However, when compared to the combined dataset, observed heterozygosity was slightly higher in WI and appreciably lower in EB. G_{is} estimates remained negative in WI and were not different from 0 in EB.

Across the 10 subsets considering the 18 WI founders and a randomized equal number of EB individuals the number of loci retained was higher than when all samples were analyzed together (mean ± SD = 3194.3 ± 42.7),

Table 1 Genetic diversity statistics and differentiation estimates for White Island (WI) and Erebus Bay (EB) Weddell Seal populations

	WI						EB					
Subset	No Loci	No Alleles	Effective No Alleles	Ho	He	G_{is}	No Alleles	Effective No Alleles	Ho	He	G_{is}	F_{ST}
1	3104	1.542	1.221	0.159	0.139	−0.147 (−0.164 to −0.129)	1.873	1.237	0.163	0.162	−0.002 (−0.019 to 0.015)	0.100
2	3186	1.537	1.221	0.158	0.138	−0.149 (−0.166 to −0.131)	1.871	1.242	0.169	0.166	−0.019 (−0.035 to −0.003)	0.107
3	3195	1.532	1.214	0.155	0.135	−0.150 (−0.167 to −0.132)	1.873	1.231	0.161	0.159	−0.012 (−0.028 to 0.005)	0.102
4	3217	1.523	1.213	0.154	0.134	−0.154 (−0.171 to −0.136)	1.878	1.231	0.156	0.159	0.020 (0.004 to 0.037)	0.102
5	3236	1.527	1.213	0.154	0.134	−0.150 (−0.167 to −0.133)	1.873	1.234	0.161	0.161	−0.002 (−0.018 to 0.014)	0.105
6	3234	1.531	1.217	0.156	0.136	−0.149 (−0.167 to −0.132)	1.879	1.241	0.166	0.164	−0.007 (−0.023 to 0.009)	0.103
7	3186	1.526	1.217	0.156	0.136	−0.149 (−0.167 to −0.132)	1.875	1.236	0.158	0.162	0.024 (0.009 to 0.041)	0.103
8	3240	1.519	1.211	0.151	0.132	−0.146 (−0.164 to −0.128)	1.875	1.231	0.159	0.159	−0.002 (−0.019 to 0.015)	0.102
9	3198	1.527	1.213	0.154	0.134	−0.147 (−0.164 to −0.129)	1.871	1.231	0.159	0.159	0.002 (−0.014 to 0.019)	0.109
10	3147	1.543	1.220	0.159	0.138	−0.150 (−0.167 to −0.131)	1.875	1.240	0.166	0.164	−0.014 (−0.03 to 0.001)	0.105
All Seals	2096	1.798	1.337	0.223	0.203	−0.100 (−0.114 to −0.087)	1.916	1.352	0.232	0.228	−0.019 (−0.034 to −0.002)	0.140
WI 18	1767	2.000	1.415	0.299	0.26	−0.151 (−0.168 to −0.133)						
All EB	3142						2.000	1.223	0.152	0.152	0.001 (−0.014 to 0.016)	
All WI	1568	2.000	1.436	0.29	0.263	−0.102 (−0.115 to −0.089)						

Estimates are made separately for loci called based on all samples together, the 18 WI founders, all individuals from each population independently, and subsets of a random 18 individuals from EB and the original 18 founders of the WI population. Metrics include the number of loci, number of alleles, effective number of alleles, observed heterozygosity (Ho), expected heterozygosity (He), inbreeding (G_{is}) along with the 95% Cis in parenthesis, and differentiation (F_{ST})

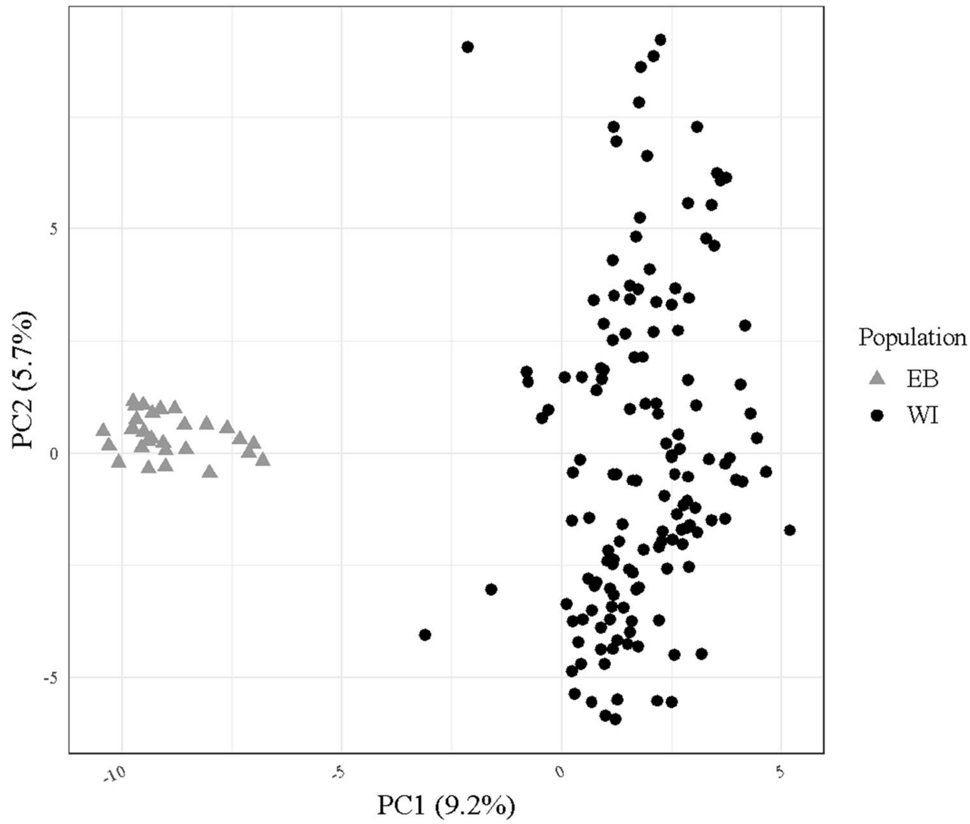

Fig. 2 PCA plot of allele frequency differences between the Erebus Bay (EB) and White Island (WI) Weddell seal populations based on 2096 SNP loci. Individuals are represented by points. Percentages show the amount of variation attributed to each axis

values that were more on par with when EB was considered on its own. In these subsets allelic diversity was always higher in EB (mean ± SD = 1.235 ± 0.004) than WI (mean ± SD = 1.216 ± 0.004), as was observed heterozygosity (EB mean ± SD = 0.162 ± 0.003 versus WI mean ± SD = 0.156 ± 0.003). Similarly, G_{is} values in EB (mean ± SD = − 0.001 ± 0.014) were always larger than those in WI, which were always less than 0 (mean ± SD = − 0.149 ± 0.002). F_{st} values between the populations averaged 0.104 (SD = 0.003) and were not correlated with the number of markers in the dataset (Pearson's product-moment correlation = 0.17; t_8 = 0.496, p = 0.633).

Pedigree reconstruction and comparison

All of the 61 individuals with tissue samples born between 2001 and 2017 had parents successfully assigned when we used microsatellite loci and the pedigree construction methods of Gelatt et al. (2010). The assigned maternities matched the field-observed mother in all but 2 cases, where the discrepancy may have arisen from a pup-switch in the year 2000. The addition of these individuals brings the total number of unique individuals in the WI pedigree to 160. Of this, 145 individuals comprise the genetic-based pedigree which includes 127 individuals with maternal and paternal genetic assignments, and the 18 founding individuals.

For the SNP dataset, filtering with MAF 0.01 and per-individual missingness threshold of 75% resulted in 1303 loci genotyped in 142 individuals of which 124 represent "non-founding" individuals. Filtering to loci with MAF greater than 0.30 resulted in retention of 394 loci. Mean polymorphic information content (PIC) was higher with the microsatellite dataset than with SNPs (Table 2), however non-exclusion probabilities were lower for SNPs, though all values were below $4.7*10^{-4}$ for all marker sets.

When considering the entirety of the WI population, results of pedigree construction differed depending on the program (CERVUS or Sequoia), method (All, Older, or Survivor), and marker set being considered (microsatellites or SNPs). In CERVUS, assignment rates were consistent for the microsatellite dataset. Here, 127 progenies (100% of those considered) were assigned a sire and dam regardless of the analytical method being used. However, there were differences in the specific individuals assigned as parents (see below). Using the MAF 0.01 dataset, assignment rates in CERVUS were equally robust to analytical method with 124 progeny (100% of those considered) assigned sires and dams. However, when Sequoia was applied to the full SNP dataset the program assigned only 54 individuals (43.2%) both a sire and dam, 7 individuals (5.6%) were assigned only a dam, and 4 (3.2%) were assigned only a sire. With the MAF 0.30 dataset assignment rates in Sequoia increased,

Table 2 Diversity statistics and exclusion probabilities for marker sets used in parentage assignment of Weddell seals from White Island

	Microsatellites	All SNPs	High MAF SNPs
Number of loci:	41	1303	394
Mean proportion of loci typed:	1	0.982	0.975
Mean expected heterozygosity:	0.527	0.286	0.478
Mean polymorphic information content (PIC):	0.461	0.227	0.358
Combined non-exclusion probability (first parent):	4.72E-04	5.73E-32	9.25E-21
Combined non-exclusion probability (second parent):	3.40E-07	2.77E-70	1.68E-34
Combined non-exclusion probability (parent pair):	1.33E-11	3.61E-117	7.93E-55
Combined non-exclusion probability (identity):	6.60E-24	1.32E-321	2.35E-160
Combined non-exclusion probability (sib identity):	3.89E-11	1.01E-164	3.13E-84

Calculations were done in CERVUS

with 75 individuals (60.0%) assigned a sire and dam, 13 individuals (10.4%) assigned only a dam, and 3 individuals (2.4%) assigned only a sire. Note that the previous two results do not include assignments to inferred parents.

For the microsatellite-based CERVUS results, parentage assignments were 100% concordant for the Older and Survivor datasets. Therefore, these pedigrees will serve as the base against which all other comparisons will be made. When considering the All dataset and microsatellite loci, assignments differed for 10 progeny: for 6 individuals both parents were different, for 2 individuals only the dams differed, and for 2 individuals only the sires differed. The Older and Survivor datasets with SNP loci showed differences involving assignments for 4 offspring: for 3 individuals, only the dam assigned was different; for the remaining individual, only the sire differed. In contrast, the All dataset with SNP loci differed in assignments for 10 individuals: for 5 individuals both parents were different, for 2 individuals only the dams differed, and for 3 individuals the sires differed.

Note, for comparisons involving Sequoia we did not consider cases where the program was unable to assign either parent (no-calls), only cases where the program made an assignment and it differed from the reference microsatellite pedigree. When considering the MAF 0.01 SNP dataset the Sequoia pedigree differed in assignments for 38 progenies: for 2 individuals both parents were different, and for the remaining 36 individuals the sire assigned was different. The major source of discrepancy was when an inferred parent was assigned to a sample (both parent offspring trios, as well as for 32 sires). In the remaining 4 cases a different sire was assigned compared to the reference pedigree. In this dataset a total of 4 inferred dams and 8 inferred sires were created. Three of the 4 dams of which were assigned to individuals in the oldest individuals (Generation G_1 of Gelatt et al. 2010) that did not have assignments in the microsatellite pedigree. The remaining dam was assigned to two offspring that had different dams in the reference pedigree. Three of the 8 inferred sires were assigned to individuals in G_1 that did not have assignments in the microsatellite pedigree. Of the remaining inferred sires 4 could be consistently associated to a corresponding male in the microsatellite pedigree (though these males were also assigned paternities in Sequoia) while 1 inferred sire was assigned to two offspring that had different sires in the microsatellite pedigree.

When considering the MAF 0.3 SNP dataset the Sequoia pedigree differed in assignments for 27 progenies: for 3 individuals both parents were different, and for the remaining 24 individuals the sire assigned was different. Again, the major discrepancy was when an inferred parent was assigned to a sample (all 3 parent offspring trios and 21 of the sires). In this pedigree there were only 4 cases where Sequoia assigned a different sire than the microsatellite pedigree, and no cases where a different dam was assigned. In total, 4 inferred dams and 10 inferred sires were created. Two of the 4 dams were assigned to individuals in the G_1 generation that did not have assignments in the microsatellite pedigree, while 1 of the remaining inferred dams was assigned to two offspring that had different dams in the microsatellite pedigree. The remaining inferred dam was assigned both to individuals in the G_1 generation that did not have assignments in the microsatellite pedigree as well as one individual that did have an assignment in the microsatellite dataset. Two of the inferred sires were exclusively assigned to individuals in G_1 that did not have assignments in the microsatellite pedigree. Four sires could be consistently associated to a corresponding male in the reference pedigree (though these males were also assigned paternities in this pedigree). Two inferred sires were associated with a single sire in the reference pedigree (who was also assigned paternities in this pedigree). One inferred sire was assigned to two offspring that had different sires in the reference pedigree. The remaining inferred sire was assigned both to individuals in the G_1 generation that did not have assignments in the reference pedigree as well as one individual that did have an assignment in the reference dataset.

Discussion

Here we developed a new set of SNP loci for Weddell seals discovered through ddRAD sequencing of individuals from two populations. When comparing genetic diversity between populations, WI showed consistent signs of reduced diversity compared to EB. This result is consistent with a population bottleneck after the founding of WI from EB individuals (Stirling 1972; Testa and Scotton 1999; Gelatt et al. 2010). It is noteworthy that the number of loci returned was diminished when all WI and EB individuals we sampled were analyzed simultaneously. In this situation, "rare" loci in the outbred EB population were not retained after frequency-based filters were applied. However, when the sample sizes were equalized the number of loci retained increased, especially in EB. We note that our choice to focus on the 18 WI founders and a matching number of EB individuals in the subsampling analyses did not detect as large a reduction in genetic diversity as was found in previous work (Gelatt et al. 2010). This is likely because the latter study, in addition to using multi-allelic microsatellites, included multiple generations of related WI individuals. Such related individuals would have depressed observed heterozygosity even more than the initial founder effect. However, in these cases our primary interest was in investigating the number of loci that were discovered and not genetic diversity between the two populations per se. We observed moderate genetic divergence between WI and EB and found no genetic evidence of migration between the two populations. In 2017 a yearling female born in EB was recorded during both surveys conducted at WI, establishing a connection between the EB and WI populations (Jay Rotella, personal communication); however, as this sole female seal has not been observed as a mother, there continues to be no effective dispersal between the two populations. Therefore, it appears that enough uninterrupted ice cover exists to continue preventing sizable movement between the two populations.

We applied this new marker set to pedigree construction in WI, comparing the results to those from microsatellite loci and two pedigree construction software (CERVUS and Sequoia). CERVUS had higher rates of assignment regardless of marker type. However, having some sort of "polarization" significantly improved assignment using both SNPs and microsatellites. In these cases, polarization reduces and focuses the pool of candidate parents making the calculations more powerful and tractable (Pemberton 2008; Jones et al. 2010; Flanagan and Jones 2019). In the absence of polarization, when all individuals were simultaneously assessed, the majority of assignment differences were for "middle tier" individuals (i.e., those with both parents and offspring in the dataset). Without polarization, a focal "middle tier" individual could be assigned as the parent of its parent. In many cases polarization comes from long-term observation of a population. However, other methods exist to group individuals into cohorts based on single sampling events; for example, tooth ageing (Hewison et al. 1999; Gipson et al. 2000; Blundell and Pendleton 2008), counting horn annuli in wild sheep (Geist 1966; Hemming 1969), or using molt patterns in birds (Mulvihill 1993; Wolfe et al. 2010; Johnson et al. 2011). These covariates assist in determining reproductive tenure, but additional information such as spatial proximity of candidate parents can also help in pedigree construction (e.g. Hadfield et al. 2006).

When built in CERVUS pedigrees based on yearly cohorts and SNP loci differed in only 4 assignments compared to the reference microsatellite pedigree. One of these is likely a true error of the microsatellite data where the original analysis had two equally probable candidate sires (0 trio mismatches), but the SNP genotypes showed that the originally selected sire had 25 mismatches while the alternate one had 0. In 2 of the remaining 3 cases there are almost no mismatches between offspring and either the candidate dam or candidate sire when considered independently, but when a trio is suggested there are > 13 mismatches found. To us, this suggests that one of the parents is incorrectly called as homozygous for alleles that should be heterozygous. This phenomenon seems similar to how null alleles in microsatellite genotypes can influence assignments (Paetkau and Strobeck 1995; Dakin and Avise 2004).

We agree with Pemberton (2008) that pedigrees constructed in wild populations will remain an important tool in evolution and conservation research. As new pedigrees are developed researchers will likely move to construction based on SNPs for a number of reasons including accuracy, repeatability, and ability to incorporate non-invasive samples which may be necessary for getting full stock of candidate parents in illusive or rare species. Therefore, we expect that the methods of assigning parentage using SNPs will likely develop with their increased use (Huisman 2017; Flanagan and Jones 2019). However, we acknowledge that development of genomic SNPs allows for pedigree-free assessments of some of the same questions. In particular, genomic estimates of relatedness can be used to calculate the heritability of traits (Evans et al. 2018; Perrier et al. 2018), and genome-wide association studies (GWAS) can link genotypes to phenotypes for trait mapping (Santure and Garant 2018).

In addition to being a resource for continued pedigree construction in WI, the SNPs described here will be used to address a number of research questions. This includes linking genomic diversity with phenotypic or life-history characteristics (Huisman et al. 2016), examining reproductive tenure and correlates with success (Charpentier et al.

2005; Wikberg et al. 2017), and searching for evidence of inbreeding avoidance or tolerance in this isolated population (Rioux-Paquette et al. 2010; Szulkin et al. 2013). Furthermore, the loci can help examine the demographic histories of both the WI and EB populations (Cabrera and Palsbøll 2017; Nunziata and Weisrock 2018).

Acknowledgements Field work to collect tissue samples was supported by the National Science Foundation, Division of Polar Programs under grant numbers ANT-1141326 and ANT-1640481 to J. J. Rotella, R. A. Garrott and Donald B. Siniff and prior NSF Grants to R. A. Garrott, J. J. Rotella, D. B. Siniff and J. Ward Testa. We would like to thank students in Biol392/592 at the University of Alberta for assistance with preparing samples for sequencing, as well as the staff of the Molecular Biology Services Unit (MBSU) at the University of Alberta for conducting the sequencing.

Author Contributions Conceptualization: JMM, CSD, Data curation: JMM, EOC, CSD, Formal analysis and investigation: JMM, EOC, CSD, Writing—original draft preparation: JMM, Writing—review and editing: EOC, TG, KRM, JJR CSD, Funding acquisition: JR, CSD, Resources: JR, KRM, CSD.

Funding Field work to collect tissue samples was supported by the National Science Foundation, Division of Polar Programs under grant numbers ANT-1141326 and ANT-1640481 to J. J. Rotella, R. A. Garrott and Donald B. Siniff and prior NSF Grants to R. A. Garrott, J. J. Rotella, D. B. Siniff and J. Ward Testa.

Data Availability Sequencing reads are archived on NCBI (Accessions: SAMN19069034-SAMN19069205). VCF files for whole dataset as well as each of the 10 random subsets are included as supplementary materials.

Code availability Not applicable.

Declarations

Conflict of interest All authors declare that they have no conflict of interest.

Ethical approval All tissue samples were collected under Marine Mammal Protection Act Permits and Antarctic Conservation Act Permits and with methods approved by Montana State University's Institutional Animal Care and Use Committee.

Consent to participate Not applicable.

Consent for publication All authors have read the manuscript and approve of its submission for publication.

References

Allendorf F (1993) Delay of adaptation to captive breeding by equalizing family size. Conserv Biol 7:416–419

Anderson EC (2012) Large-scale parentage inference with SNPs: an efficient algorithm for statistical confidence of parent pair allocations. Statistical Applications in Genetics and Molecular Biology. https://doi.org/10.1515/1544-6115.1833

Anderson EC, Garza JC (2006) The power of single-nucleotide polymorphisms for large-scale parentage inference. Genetics 172:2567–2582

Backström N, Brandstrom M, Gustafsson L et al (2006) Genetic mapping in a natural population of collared flycatchers (Ficedula albicollis): Conserved synteny but gene order rearrangements on the avian Z chromosome. Genetics 174:377–386

Blundell GM, Pendleton GW (2008) Estimating age of harbor seals (Phoca vitulina) with incisor teeth and morphometrics. Mar Mamm Sci 24:577–590

Cabrera AA, Palsbøll PJ (2017) Inferring past demographic changes from contemporary genetic data: A simulation-based evaluation of the ABC methods implemented in diyabc. Mol Ecol Resour 17:e94–e110

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. *G3: Genes|Genomes|Genetics*, **1**, 171–182.

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

Charpentier M, Setchell JM, Prugnolle F et al (2005) Genetic diversity and reproductive success in mandrills (Mandrillus sphinx). Proc Natl Acad Sci USA 102:16723–16728

Chen N, Cosgrove EJ, Bowman R, Fitzpatrick JW, Clark AG (2018) Genomic consequences of population decline in the endangered florida scrub-jay. Curr Biol 26:2974–2979

Dakin EE, Avise JC (2004) Microsatellite null alleles in parentage analysis. Heredity 93:504–509

Evans LM, Tahmasbi R, Jones M et al (2018) Narrow-sense heritability estimation of complex traits using identity-by-descent information. Heredity 121:616–630

Farquharson KA, Hogg CJ, Grueber CE (2019) A case for genetic parentage assignment in captive group housing. Conserv Genet 20:1187–1193

Fienieg ES, Galbusera P (2013) The use and integration of molecular DNA information in conservation breeding programmes: a review. Journal of Zoo and Aquarium Research 1:44–51

Flanagan SP, Jones AG (2019) The future of parentage analysis: From microsatellites to SNPs and beyond. Mol Ecol 28:544–567

Geist V (1966) Validity of horn segment counts in aging bighorn sheep. J Wildl Manag 30:634–635

Gelatt TS, Davis CS, Stirling I et al (2010) History and fate of a small isolated population of Weddell seals at White Island, Antarctica. Conserv Genet 11:721–735

Gipson PS, Ballard WB, Nowak RM, Mech LD (2000) Accuracy and precision of estimating age of gray wolves by tooth wear. J Wildl Manag 64:752–758

Guichoux E, Lagache L, Wagner S et al (2011) Current trends in microsatellite genotyping. Mol Ecol Resour 11:591–611

Hadfield JD, Richardson DS, Burke T (2006) Towards unbiased parentage assignment: combining genetic, behavioural and spatial data in a Bayesian framework. Mol Ecol 15:3715–3730

Hemming JE (1969) Cemental deposition, tooth succession, and horn development as criteria of age in dall sheep. J Wildl Manag 33:552–558

Hewison AJM, Vincent JP, Angibault JM et al (1999) Tests of estimation of age from tooth wear on roe deer of known age: variation within and among populations. Can J Zool 77:58–67

Huisman J (2017) Pedigree reconstruction from SNP data: parentage assignment, sibship clustering and beyond. Mol Ecol Resour 17:1009–1024

Huisman J, Kruuk LEB, Ellis PA, Clutton-Brock T, Pemberton JM (2016) Inbreeding depression across the lifespan in a wild mammal population. Proc Natl Acad Sci 113:3585–3590

Jaari S, Li M-H, Merilä J (2009) A first-generation microsatellite-based genetic linkage map of the Siberian jay (Perisoreus infaustus): insights into avian genome evolution. BMC Genomics. https://doi.org/10.1186/1471-2164-10-1

Jiménez-Mena B, Schad K, Hanna N, Lacy RC (2016) Pedigree analysis for the genetic management of group-living species. Ecol Evol 6:3067–3078

Johnson EI, Wolfe JD, Brandt Ryder T, Pyle P (2011) Modifications to a molt-based ageing system proposed by Wolfe et al. (2010). J Field Ornithol 82:422–424

Jones AG, Small CM, Paczolt KA, Ratterman NL (2010) A practical guide to methods of parentage analysis. Mol Ecol Resour 10:6–30

Jones OR, Wang J (2010) Molecular marker-based pedigrees for animal conservation biologists. Anim Conserv 13:26–34

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program cervus accommodates genotyping error increases success in paternity assignment. Mol Ecol 16:1099–1106

Keller LF (1998) Inbreeding and Its Fitness Effects in an Insular Population of Song Sparrows (Melospiza melodia). Evolution 52:240–250

Kruuk LEB (2004) Estimating genetic parameters in natural populations using the "animal model". *Philosophical Transactions of the Royal Society of London*. Series b: Biological Sciences 359:873–890

Leroy G (2011) Genetic diversity, inbreeding and breeding practices in dogs: Results from pedigree analyses. Vet J 189:177–182

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

Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589–595

Lynch M, Walsh B (1998) *Genetics and analysis of quantitative traits*. Sinauer Sunderland MA.

MacDonald ZG, Dupuis JR, Davis CS et al (2020) Gene flow and climate-associated genetic variation in a vagile habitat specialist. Mol Ecol 29:3889–3906

Malenfant RM, Coltman DW, Richardson ES et al (2016) Evidence of adoption, monozygotic twinning, and low inbreeding rates in a large genetic pedigree of polar bears. Polar Biol 39:1455–1465

Malenfant RM, Davis CS, Richardson ES, Lunn NJ, Coltman DW (2018) Heritability of body size in the polar bears of Western Hudson Bay. Mol Ecol Resour 18:854–866

Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. Mol Ecol 7:639–655

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. Embnet.journal 17:10–12

Meirmans PG, Van Tienderen PH (2004) GENOTYPE and GENODIVE: Two programs for the analysis of genetic diversity of asexual organisms. Mol Ecol Notes 4:792–794

Miller EJ, Eldridge MDB, Morris K, Thomas N, Herbert CA (2015) Captive management and the maintenance of genetic diversity in a vulnerable marsupial, the greater bilby. Australian Mammalogy 37:170–181

Mrode RA, Thompson R (2014) *Linear models for the prediction of animal breeding values, 3rd Edition*. CAB

Mulvihill RS (1993) Using wing molt to age passerines. North American Bird Bander 18:10

Nielsen RK, Pertoldi C, Loeschcke V (2007) Genetic evaluation of the captive breeding program of the Persian wild ass. J Zool 272:349–357

Nietlisbach P, Camenisch G, Bucher T et al (2015) A microsatellite-based linkage map for song sparrows (Melospiza melodia). Mol Ecol Resour 15:1486–1496

Nunziata SO, Weisrock DW (2018) Estimation of contemporary effective population size and population declines using RAD sequence data. Heredity 120:196–207

Paetkau D, Strobeck C (1995) The molecular basis and evolutionary history of a microsatellite null allele in bears. Mol Ecol 4:519–520

Paris JR, Stevens JR, Catchen JM (2017) Lost in parameter space: a road map for stacks. Methods Ecol Evol 8:1360–1373

Paterson JT, Rotella JJ, Link WA, Garrott R (2018) Variation in the vital rates of an Antarctic marine predator: the role of individual heterogeneity. Ecology 99:2385–2396

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

Pemberton JM (2008) Wild pedigrees: the way forward. Proceedings of the Royal Society b: Biological Sciences 275:613–621

Perrier C, Delahaie B, Charmantier A (2018) Heritability estimates from genomewide relatedness matrices in wild populations: Application to a passerine, using a small sample size. Mol Ecol Resour 18:838–853

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

Piepho HP, Möhring J, Melchinger AE, Büchse A (2008) BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161:209–228

Poissant J, Davis CS, Malenfant RM, Hogg JT, Coltman DW (2012) QTL mapping for sexually dimorphic fitness-related traits in wild bighorn sheep. Heredity 108:256–263

R Core Team (2019) R: A language and environment for statistical computing, reference index version 3.2.2.

Reid JM, Keller LF, Marr AB et al (2014) Pedigree error due to extra-pair reproduction substantially biases estimates of inbreeding depression. Evolution 68:802–815

Richardson ES, Davis C, Stirling I et al (2020) Variance in lifetime reproductive success of male polar bears. Behav Ecol 31:1224–1232

Rioux-Paquette E, Festa-Bianchet M, Coltman DW (2010) No inbreeding avoidance in an isolated population of bighorn sheep. Anim Behav 80:865–871

Rochette NC, Catchen JM (2017) Deriving genotypes from RAD-seq short-read data using Stacks. Nat Protoc 12:2640

Rotella JJ, Paterson JT, Garrott RA (2016) Birth dates vary with fixed and dynamic maternal features, offspring sex, and extreme climatic events in a high-latitude marine mammal. Ecol Evol 6:1930–1941

Santure AW, Garant D (2018) Wild GWAS—association mapping in natural populations. Mol Ecol Resour 18:729–738

Shafer ABA, Peart CR, Tusso S et al (2017) Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. Methods Ecol Evol 8:907–917

Siniff DB, DeMaster DP, Hofman RJ, Eberhardt LL (1977) An Analysis of the dynamics of a weddell seal population. Ecol Monogr 47:319–335

Stirling I (1972) Regulation of numbers of an apparently isolated population of weddell seals (Leptonychotes weddelli). J Mammal 53:107–115

Szulkin M, Stopher KV, Pemberton JM, Reid JM (2013) Inbreeding avoidance, tolerance, or preference in animals? Trends Ecol Evol 28:205–211

Testa JW, Scotton BD (1999) Dynamics of an Isolated Population of Weddell Seals (Leptonychotes weddellii) at White Island, Antarctica. J Mammal 80:82–90

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

Wikberg EC, Jack KM, Fedigan LM et al (2017) Inbreeding avoidance and female mate choice shape reproductive skew in capuchin monkeys (Cebus capucinus imitator). Mol Ecol 26:653–667

Williams SE, Hoffman EA (2009) Minimizing genetic adaptation in captive breeding programs: A review. Biol Cons 142:2388–2400

Wolfe JD, Ryder TB, Pyle P (2010) Using molt cycles to categorize the age of tropical birds: an integrative new system. J Field Ornithol 81:186–194