Received: 5 March 2018 | Revised: 25 April 2018 | Accepted: 26 April 2018

DOI: 10.1111/1755-0998.12898

RESOURCE ARTICLE

WILEY MOLECULAR ECOLOGY RESOURCES

Evaluating sample size to estimate genetic management metrics in the genomics era

Elizabeth P. Flesch[1] | Jay J. Rotella[1] | Jennifer M. Thomson[2] | Tabitha A. Graves[3] | Robert A. Garrott[1]

[1]Ecology Department, Montana State University, Bozeman, Montana

[2]Animal and Range Sciences Department, Montana State University, Bozeman, Montana

[3]U.S. Geological Survey Glacier Field Station, Northern Rocky Mountain Science Center, West Glacier, Montana

Correspondence
Elizabeth P. Flesch, Ecology Department, Montana State University, 310 Lewis Hall, Bozeman, 59717 MT.
Email: elizabeth.flesch@gmail.com

Funding information
Montana Fish, Wildlife and Parks; Glacier National Park; Yellowstone National Park; National Science Foundation Graduate Research Fellowship; Montana and Midwest Chapters of the Wild Sheep Foundation; National Geographic Society; University of Wyoming; Montana Agricultural Experiment Station; Canon USA Inc.; Wyoming Governor's Big Game License Coalition; National Park Service; U.S. Geological Survey

Abstract

Inbreeding and relationship metrics among and within populations are useful measures for genetic management of wild populations, but accuracy and precision of estimates can be influenced by the number of individual genotypes analysed. Biologists are confronted with varied advice regarding the sample size necessary for reliable estimates when using genomic tools. We developed a simulation framework to identify the optimal sample size for three widely used metrics to enable quantification of expected variance and relative bias of estimates and a comparison of results among populations. We applied this approach to analyse empirical genomic data for 30 individuals from each of four different free-ranging Rocky Mountain bighorn sheep (*Ovis canadensis canadensis*) populations in Montana and Wyoming, USA, through cross-species application of an Ovine array and analysis of approximately 14,000 single nucleotide polymorphisms (SNPs) after filtering. We examined intra- and interpopulation relationships using kinship and identity by state metrics, as well as F_{ST} between populations. By evaluating our simulation results, we concluded that a sample size of 25 was adequate for assessing these metrics using the Ovine array to genotype Rocky Mountain bighorn sheep herds. However, we conclude that a universal sample size rule may not be able to sufficiently address the complexities that impact genomic kinship and inbreeding estimates. Thus, we recommend that a pilot study and sample size simulation using R code we developed that includes empirical genotypes from a subset of populations of interest would be an effective approach to ensure rigour in estimating genomic kinship and population differentiation.

KEYWORDS
kinship, *Ovis canadensis canadensis*, sampling, single nucleotide polymorphism

1 | INTRODUCTION

The composition of individual genomes can be an important influence on the health and long-term persistence of wildlife populations. Genomes can have profound impacts on individual fitness (Kristensen, Pedersen, Vermeulen, & Loeschcke, 2010; Romanov et al., 2009), population-level demography (Hogg, Forbes, Steele, & Luikart, 2006), resilience to environmental change (Manel et al., 2010) and response to novel pathogens or parasites (Acevedo-Whitehouse et al., 2005; Coltman, Pilkington, Kruuk, Wilson, & Pemberton, 2001; Siddle et al., 2007). Thus, genomic assessment of populations can be an important component of wildlife research and conservation efforts. Two important genetic attributes are inbreeding measured for individuals and kinship among individuals (Blouin, 2003). At the population level, these attributes serve to evaluate gene flow among populations (Morin et al., 1994; Streiff et al., 1999), detect population differentiation (Funk, McKay, Hohenlohe, & Allendorf, 2012) and evaluate demographic history (Li & Durbin, 2011; Sheehan,

Harris, & Song, 2013). On the individual level, inbreeding and relatedness metrics can indicate inbreeding depression effects (Grueber, Laws, Nakagawa, & Jamieson, 2010; Nielsen et al., 2012), heritability of observed phenotypes (Daetwyler et al., 2014; Kruuk, 2004) and life history characteristics, such as propensity for dispersal (Gueijman, Ayali, Ram, & Hadany, 2013; Shafer, Poissant, Côté, & Coltman, 2011).

Researchers and wildlife managers often have limited resources and seek to maximize biological insight derived for the resources invested in wildlife capture and genomic analysis. To make informed decisions regarding study design, biologists require an approach to evaluate the expected level of uncertainty in inbreeding and kinship results and decide on an acceptable sampling intensity. The level of biological insight and uncertainty derived from estimates of inbreeding and kinship can be influenced by many aspects of study design, including the metric employed, marker type, number of markers, number of individuals sampled per population, and composition of the populations and individuals under consideration (Csilléry et al., 2006; Frankham et al., 2017). Despite the potential for these study design decisions to impact biological inferences, few studies evaluate the reliability and precision of relatedness and inbreeding estimates prior to sampling, resulting in potential for imprecise estimates and results being interpreted out of context (Taylor, 2015). Thus, guidelines are needed to promote robust conclusions regarding relatedness and inbreeding metric performance for each data set (Taylor, 2015). As a result, in this study, we sought to conduct a rigorous simulation study to evaluate multiple inbreeding and kinship metrics while accounting for different influences on estimator precision.

There are many different metrics and alternative approaches for estimating inbreeding and kinship using molecular markers, and critical differences exist among their respective inferences when applied to genetic management of populations (Frankham et al., 2017). Three of the main metric types include identity by state, kinship coefficients and *F*-statistics. In terms of single nucleotide polymorphisms (SNPs), identity by state (IBS) means that the same nucleotide is located at the same genomic position in both the maternal and paternal chromosomes (Toro, Villanueva, & Fernández, 2014). The probability of zero identity by state sharing is calculated pairwise between two individuals (dyads) and estimates the probability that two individuals share zero alleles that are identical by state (Manichaikul et al., 2010). Kinship coefficient (ϕ), also termed coancestry, is calculated between two individuals and estimates the probability that two randomly selected alleles, one from each individual, from any locus are identical by descent (Manichaikul et al., 2010). A particularly common *F*-statistic is F_{ST}, which measures differentiation among subpopulations.

The type, number and polymorphism of molecular markers used as inputs for kinship and inbreeding calculations impact the accuracy of resulting estimates (Blouin, 2003). Microsatellite markers, which are short tandem repeats of DNA motifs, have been applied in many wildlife genetic studies, but often include a limited number of markers, resulting in kinship and inbreeding estimates that may correlate poorly with those derived from pedigrees (Slate et al., 2004; Taylor, Kardos, Ramstad, & Allendorf, 2015; Toro et al., 2002). Thus, employing a small number of microsatellite markers may have limited utility to inform management decisions to maintain genetic diversity in conservation programmes (Fernández et al., 2012). Therefore, multiple studies have recommended the use of genomic data over microsatellites for this purpose (Frankham et al., 2017; Saura et al., 2013; Toro et al., 2014). Genomic data are composed of many more markers across the genome and can be generated by RADseq (Thrasher, Butcher, Campagna, Webster, & Lovette, 2018), whole genome sequencing (Pool, Hellmann, Jensen, & Nielsen, 2010), and cross-species application of SNP chips (Haynes & Latch, 2012; Miller, Kijas, Heaton, McEwan, & Coltman, 2012; Miller, Poissant, Kijas, & Coltman, 2011). When SNP data were used instead of microsatellites, inbreeding and kinship estimates were more strongly correlated with genealogical data, and the addition of microsatellite data to SNP data did not improve accuracy (Santure et al., 2010). Mapped genomic data also enable evaluation and management of inbreeding and kinship across specific genomic regions (Roughsedge, Pong-Wong, Woolliams, & Villanueva, 2008). In general, genomic data have the potential to provide stronger inference on patterns of kinship and inbreeding than a limited number of microsatellites and may require 52% fewer samples per population (Jeffries et al., 2016).

Sample size is an important study design factor that influences study cost and inferential strength. There are generally two types of sampling that occur in inbreeding and kinship studies. First, there is process variance, also sometimes termed genetic sampling in the genetic literature, due to variations in allele frequencies caused by natural processes, such as genetic drift and local adaptation (Holsinger & Weir, 2009). Thus, existing composition and demographic history of a considered population can impact precision of results, for example, low variance in kinship can result in lower power to address research questions (Csilléry et al., 2006; Robinson, Simmons, & Kennington, 2013; Taylor, 2015; Van de Casteele, Galbusera, & Matthysen, 2001). Second, there is sampling variance, caused by variation in allele frequencies when a subset of individuals (the sample) is drawn from the population (Holsinger & Weir, 2009). This source of variation can be addressed by increasing the number of animals sampled from each population (Holsinger & Weir, 2009). Despite the influence of sampling variance, actual and recommended sample sizes for evaluating a population have varied widely by study. Evaluations of simulated microsatellite data sets recommended a range of 20–100 individuals per population to evaluate F_{ST} (Kalinowski, 2004), and 50 individuals per population to identify immigrants (Paetkau, Slade, Burden, & Estoup, 2004), while another study that used an empirical microsatellite data set estimated that 25–30 individuals were necessary to accurately estimate allele frequencies (Hale, Burg, & Steeves, 2012).

Evaluations of sample size for genomic data have also varied in their approach and recommendations. In general, studies using high-throughput sequencing have tended to use smaller sample sizes due to expense, in comparison with microsatellite genotyping studies.

However, limited sampling can greatly impact population genetic inferences (Meirmans, 2015). Hoban and Schlarbaum (2014) recommended 25–30 samples per plant population to capture spatially restricted alleles using a simulated microsatellite and SNP data set. In contrast, a simulation using 10,000 bi-allelic loci found that a sample size of four to six could be sufficient for some but not all F_{ST} statistics (Willing, Dreyer, & van Oosterhout, 2012). A study that simulated sequencing data of varying depth estimated that 40 samples with low sequencing depth had the highest accuracy to evaluate population structure (Fumagalli, 2013). Empirical data sets may be even more useful for this evaluation because simulated data sets are unlikely to include all aspects of real systems (May, 2004). A recent empirical study using SNPs from a tree species suggested that increasing sample sizes beyond eight individuals had little impact on estimates of genetic diversity within and among populations (Nazareno, Bemmels, Dick, & Lohmann, 2017). The study was a step forward in contributing to the sample size literature using an empirical genomic data set but was limited to a small number of replicates per simulation (100) and a small number of SNPs (1000) from two populations for a nonmodel plant species. Another empirical simulation study employed 23,057 SNPs to evaluate precision of F_{ST} estimates between Galapagos tortoise populations and determined that three or five samples per population provided more precise estimates than two samples (Gaughran et al., 2017). However, no empirical genomic simulation has been published for free-ranging mammals.

Due to the many factors that can impact population genetic metrics, it can be prudent to evaluate the precision and accuracy of estimators for each unique data set (Taylor, 2015; Van de Casteele et al., 2001; Wang, 2011). This is especially relevant when evaluating populations of conservation concern with past bottlenecks and suspected low genetic diversity (Taylor, 2015). Thus, multiple simulation software options have been developed to address the need to test how a particular method might perform for a given research question, molecular marker data set and study species (Hoban, 2014). For example, the programme "Coancestry" and its associated package "related" (Pew, Muir, Wang, & Frasier, 2015; Wang, 2011) for use with the R statistical software environment (R Core Team, 2017) were developed to allow users to select the best relatedness or inbreeding estimator for a given data set. The software utilizes empirical allele frequencies to conduct a priori simulations and evaluate the reliability of moment and likelihood estimators. However, this tool is limited to seven metrics, all of which estimate relatedness and inbreeding relative to a reference population assumed to include unrelated and noninbred animals (Taylor, 2015). These metrics are based on comparing molecular markers of a specified homogeneous population to those found in individuals or dyads (Purcell et al., 2007). However, detecting population structure depends on correctly identifying individuals that are not related (Zhu, Li, Cooper, & Elston, 2008), which violates the assumption of a homogeneous population and consequently results in less accurate relationship inferences (Manichaikul et al., 2010).

There is a need for a more user-friendly and flexible approach to evaluate sample size for inbreeding and kinship metrics, given different molecular markers and study populations. If researchers implemented an analysis of sample size to derive relationship inferences, reporting of results would become more comparable among studies and allow for more generalizable insights. Thus, we sought to conduct a rigorous simulation study using an empirical genomic data set for wild animals to evaluate inbreeding and kinship metrics, account for different influences on estimator precision and provide sampling guidance for future similar efforts. In a specific manner, we wanted to determine how variance in the population average for metric values related to sample size across a gradient of sample sizes, ranging from inadequate to sufficient for reliable and informative insights. To accomplish this, we conducted simulations employing three selected metrics to evaluate genomic relationship inferences within and between populations of Rocky Mountain bighorn sheep (*Ovis canadensis canadensis*). The genomics of Rocky Mountain bighorn sheep provide an excellent opportunity to evaluate and compare genetic management metrics across population sizes that range from small, isolated herds recovering from population bottlenecks to large metapopulations that can sustain human harvest. Therefore, the herds we examined consisted of different management and demographic histories that likely impacted inbreeding and kinship and are representative of many other wildlife populations of conservation concern. As sampling guidelines might not apply equally well to all situations and species (Hoban & Schlarbaum, 2014), we sought to develop a flexible and transparent approach that could be easily applied by other researchers and managers to other data sets. Thus, we developed well-annotated, straightforward code for R (R Core Team, 2017) that others can modify and implement to make informed sample size decisions and achieve desired biological insights for other populations. We seek to shift from a paradigm of a single sample size recommendation to a more adaptive framework, where researchers employ a similar method to evaluate sample size decisions for specific data sets and metrics, to enhance inference reliability and maximize comparability of studies that estimate inbreeding and kinship.

2 | MATERIALS AND METHODS

2.1 | Genomic data set

Marker density of many agricultural animal SNP chips can provide informative molecular kinship estimates that are better than pedigrees when applied to related, nonmodel species of conservation concern (Gómez-Romano, Villanueva, Rodríguez de Cara, & Fernández, 2013). Thus, we employed a SNP genotype data set generated from the High Density (HD) Ovine array, which contains approximately 606,006 SNPs with a density of 1 SNP per 4.279 kb. The Ovine array (also called a SNP chip) is a new genomic analysis technique originally developed for domestic sheep, but its development

included genotypes from five bighorn sheep and four Dall's sheep (*Ovis dalli*; Kijas et al., 2009, 2014). Species divergence between domestic sheep and bighorn sheep took place around three million years ago (Bunch, Wu, Zhang, & Wang, 2006), but domestic sheep and bighorn sheep can interbreed and produce viable hybrid offspring (Young & Manville, 1960). In addition, the two species have the same number of chromosomes and are expected to have high genomic synteny (Poissant et al., 2010). An estimated 24,000 SNPs on the HD Ovine array are informative for evaluation of Rocky Mountain bighorn sheep (Miller, Moore, Stothard, Liao, & Coltman, 2015). Furthermore, the domestic sheep reference genome enables whole genome genotyping of bighorn sheep and the potential to map informative SNPs to genomic areas of known function (Kohn, Murphy, Ostrander, & Wayne, 2006). However, it is important to consider that the use of SNP chips can result in ascertainment bias, as only select individuals were assessed to construct the panel (Albrechtsen, Nielsen, & Nielsen, 2010), and cross-species application could be biased towards highly conserved markers. Thus, we sought to address this issue by comparing results across sample sizes among herds with different known population attributes.

2.2 | Study populations

We examined four wild populations of bighorn sheep that we expected to differ in kinship both within and between herds, due to a spectrum of population attributes and geographic isolation among herds (Figure 1). Bighorn sheep populations located in Glacier National Park, Montana, and across the Beartooth Absaroka Mountains in Wyoming served as baseline examples of large, native metapopulations with high anticipated connectivity and genetic diversity. The selected samples from Glacier National Park spanned the eastern front of the park inside Glacier and Flathead Counties, with approximately 16 from north of St. Mary Lake and 14 from the southern areas of the park (Flesch & Graves, 2018). The samples from the Beartooth Absaroka metapopulation spanned the eastern front of the Greater Yellowstone Area, across Wyoming hunt units 1, 2, 3, 5 and 22 inside Park, Hot Springs, and Fremont Counties. The Fergus (Fergus County) and Taylor-Hilgard (Gallatin and Madison Counties) herds served as examples of herds with more complex management histories. The Fergus herd is a large population that was reintroduced (43 bighorn sheep reintroduced from 1958 to 1961), experienced a population bottleneck of a limited number of individuals, and was supplemented with additional augmentations. Thus, this population is representative of a herd with a successful reintroduction and a current population size of greater than 200 individuals, as well as a past bottlenecks and augmentations. The Taylor-Hilgard herd represents a native population that experienced multiple augmentations and catastrophic die-offs that reduced the population to several 10s of animals, but has recovered to a moderate size of about 280 individuals. In addition, this herd has been impacted by respiratory disease, which is a major limiting factor to bighorn sheep conservation and management throughout the western United States (Besser et al., 2008, 2012; Cassirer et al., 2013; Cassirer & Sinclair, 2007; Miller, 2008; Monello, Murray, & Cassirer, 2001). Based on a synthesis of these herd history characteristics, we expected inbreeding and kinship to be lower within the Beartooth Absaroka and Glacier National Park herds, in comparison with the Fergus and Taylor-Hilgard herds.

2.3 | Sample collection

Bighorn sheep samples were collected using chemical immobilization, helicopter net-gunning or baited drop-nets. All capture and handling protocols complied with scientific guidelines and permits from the States of Montana and Wyoming, Yellowstone National Park, and Glacier National Park. Animal capture and handling protocols were approved by Institutional Care and Use Committees at Montana State University (Permit # 2011-17, 2014-32), Montana Department of Fish, Wildlife, and Parks (Permit # 2016-005), Wyoming Game and Fish Department (Permit # 854) or the U.S. Geological Survey (Permit #2004-01). Samples were collected at the Fergus, Taylor-Hilgard and Beartooth Absaroka populations from 2013 to 2016; Glacier samples were collected from 2004 to 2011. We collected multiple types of genetic samples, including gene cards, biopsy ear punches, whole blood and tissue. Collection using gene cards involved placing 2–4 drops of whole blood directly onto an FTA Classic gene card. Biopsy punches were obtained from ear cartilage during ear tagging and stored frozen in 90% ethanol. We also collected whole blood samples for a limited number of animals and tissue samples from hunter-harvested animals. We performed the DNA extractions using the Maxwell 16 LEV Blood DNA Kit for blood and gene card samples and the SEV Tissue Kit for biopsy punch and tissue samples. Because of sample availability and the fact that a previous simulation study (Hoban & Schlarbaum, 2014) suggested that 25–30 samples should be used per population to assess population structure, we evaluated 30 samples per herd. We employed stratified random sampling to select six samples from each of five hunt units across the Beartooth Absaroka population, which had a total of 86 samples. However, because one study suggested that more than 30 samples per population should be used (Fumagalli, 2013), we also employed all 86 samples from the Beartooth Absaroka for a separate simulation with a greater sample size to evaluate within-population kinship, compare how results might differ from analyses using a sample size of 30 and assess changes in variance when samples were drawn from a larger pool.

2.4 | Data quality control and analysis

We performed quality control of the 30 samples per herd data set and all 86 samples from the Beartooth Absaroka data set separately. The 86 sample Beartooth Absaroka data set included the 30 samples selected for 30 samples per herd analysis. We completed preliminary filtering for quality control using Golden Helix SNP & Variation Suite v8.6 software (SNP & Variation Suite, n.d.). First, we filtered for sample quality using a call rate threshold of 0.85. We deactivated markers of unknown mappings and on the sex chromosomes. We filtered

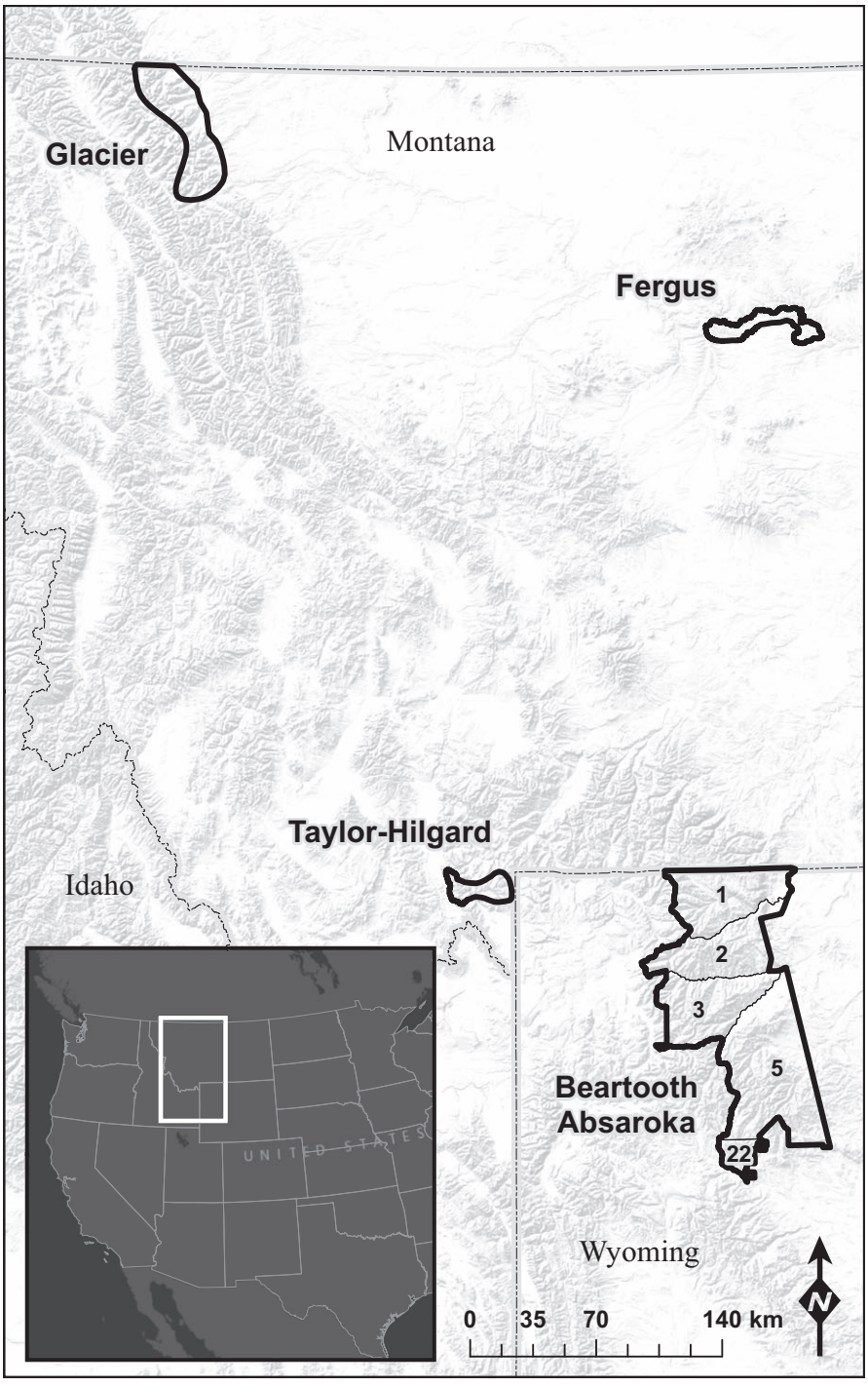

FIGURE 1 Bighorn sheep populations in Montana and Wyoming that were evaluated. Hunting districts are labelled for the Beartooth Absaroka population, but exact hunting district boundaries have been modified to more precisely show known species range

SNPs using a minor allele frequency of less than 0.0001 to remove monomorphic and extremely rare markers (De Cara, Villanueva, Toro, & Fernández, 2013). We removed markers with poor performance by requiring a SNP call rate of greater than 0.99 and this data set was used for input in the simulations. We also generated a principal components analysis (PCA) of the 30 samples per herd data set using Golden Helix after additional filtering using a minor allele frequency threshold of 0.01 and Hardy–Weinberg equilibrium *p*-value less than 0.00001 (SNP & Variation Suite, n.d.). We also performed linkage disequilibrium (LD) pruning of the data set used for the PCA analysis, which removes nonindependent SNPs that inform the presence of nearby variants, using a window size of 100, window increment of 25, LD statistic of r^2, LD threshold of 0.99 and LD computation method CHM (Huisman, Kruuk, Ellis, Clutton-Brock, & Pemberton, 2016).

We conducted simulations of inbreeding and kinship estimates within and among bighorn sheep populations to determine optimal sample size for these analyses. We used program R (R Core Team, 2017) for simulations, modifying the approach by Nazareno et al. (2017); Supporting Information Appendix S1). Our criteria for optimal sample size included an assessment of variance and precision. For variance, we used boxplots to evaluate differences among estimates

with increasing sample size, as high variance in estimates would result in similar or potentially misrepresentative estimates. In addition, we evaluated differentiation among mean estimates using the standard deviation of the mean for all replicates in each simulation. We also used boxplots to compare the distribution of mean estimates provided by the 10,000 replicates with the estimates of other populations with similar or different management histories. We compared the mean estimate for kinship, IBS and F_{ST} generated from each simulation of a certain sample size to the 30-sample estimate, which for these evaluations we assumed represented "truth" and calculated relative bias. We combined both variance and relative bias into root mean squared error (RMSE), which estimates the sample standard deviation of the differences between predicted and observed values. Thus, we looked for decreasing root mean squared error for each sample size simulation, which would indicate lower overall relative bias and variance.

For intrapopulation simulations, we estimated mean proportion of SNPs with zero IBS and mean kinship within each population using 30 samples to provide a benchmark for precision of the simulation replicates. Each simulation consisted of randomly selecting (without replacement) sample sizes of 5, 10, 15, 20 and 25 drawn from 30 possible genotypes for each population using R and PLINK v1.07 or v1.90 software (Purcell et al., 2007; R Core Team, 2017), with 10,000 replicate simulations for each sample size. There were a large number of possibilities for unique combinations of samples for each selected sample size from each herd (ranging from 142,506 possible unique combinations for sample sizes of 5 and 25 to 155,117,520 unique combinations for the sample size of 15), so we used the "sample" command in R to select each of the 10,000 replicates (R Core Team, 2017). Due to a maximum of 30 genotypes, replicates were generally not independent of one another. For example, for the sample size of 25, pairs of replicate data sets shared on average about 20 of the same genotypes. Independence among replicates increased at lower sample sizes, and a pair of replicates for a sample size of 20 shared an average of 13 genotypes in common. Despite these limitations, we were able to evaluate the decrease in variance due to sample size with more independent replicates using the 86-sample Beartooth Absaroka data set, and for a sample size of 25, pairs of replicate data sets shared an average of seven genotypes in common. Populations composed of unrelated individuals are expected to have mean kinship values that are normally distributed around 0 (Manichaikul et al., 2010). Negative mean kinship values can be produced but can be effectively truncated at 0 to indicate low kinship, but we did not truncate values to evaluate the distribution of simulation results (Manichaikul et al., 2010). Identity by state and kinship simulations did not require filtering within the simulation or LD pruning (Manichaikul et al., 2010). Subsetting of samples was implemented in PLINK v1.90 for identity by state and v1.07 for kinship simulations (Purcell et al., 2007). Probability of zero identity by state and kinship estimates were calculated using KING software v2.0 (Manichaikul et al., 2010). For all metrics, we calculated the mean of the estimates produced by each sampling group. We compared variance and bias of the 10,000 replicate estimates for each sample size simulation relative to values obtained when we used all available samples (n = 30 or 86). To evaluate intrapopulation metric estimates across the spectrum of sample sizes, we determined the extent of overlap in the range of estimates provided by each sample size for populations with different management histories.

For interpopulation metrics, we conducted simulations to compare results by sample size for F_{ST}, probability of zero identity by state, and kinship between herds. We estimated mean kinship, IBS and F_{ST} between populations using 30 samples from each population (60 total) to provide a benchmark for precision of the simulation replicates. For the simulations, we randomly selected (without replacement) sample sizes of 5, 10, 15, 20 and 25 drawn from 30 possible genotypes for each population using R and PLINK v1.07 or v1.90 software (Purcell et al., 2007; R Core Team, 2017) using 10,000 replicates. There were a large number of possibilities for unique combinations of samples for each selected sample size (ranging from 285,012 possible unique combinations for sample sizes of 5 and 25 to 310,235,040 unique combinations for the sample size of 15), so we used the "sample" command in R to select each of the 10,000 replicates (R Core Team, 2017). For interpopulation metric estimates, we evaluated if the range of estimates produced by the simulation provided different inferences regarding the relative comparisons of population differentiation among herds for various sample sizes. The F_{ST} simulation required filtering each random sampling group and analysis using PLINK v1.90 (Purcell et al., 2007) to emulate typical F_{ST} analysis procedures, which involved removal of SNPs with a minor allele frequency of less than 0.0001 and those that deviated from Hardy–Weinberg equilibrium using a threshold of $p < 0.00001$. In addition, we employed LD pruning of each sampling group to ensure independence of markers using a window size of 100, window increment of 25, LD statistic of r^2, LD threshold of 0.99 and LD computation method CHM (Huisman et al., 2016).

3 | RESULTS

3.1 | Kinship of bighorn sheep populations

Ovine HD array analysis produced a genotype of 605,898 SNPs from the forward strand. Applying a sample quality filter call rate threshold of 0.85 did not remove any samples from the 30 samples per herd data set and removed one sample from the Beartooth Absaroka large data set, leaving 86 samples. Deactivation of SNPs of unknown mappings and on sex chromosomes removed 29,401 SNPs from both data sets. Application of the minor allele frequency of less than 0.0001 removed 505,235 SNPs in the 30 samples per herd data set and 530,050 SNPs from the Beartooth Absaroka 86 sample data set. Requiring a SNP call rate of greater than 0.99 resulted in removal of 57,038 SNPs from the 30 samples per herd data set and 35,665 SNPs from the Beartooth Absaroka 86 sample data set. These quality control steps resulted in 14,224 SNPs remaining for the 30 samples per herd data set and 10,782 SNPs remaining for the 86 sample Beartooth Absaroka data set, with a SNP density of 1 SNP per 170.404 kb and 1 SNP per 224.002 kb, respectively. We

generated a PCA of the 30 samples per herd data set using 7,688 SNPs after filtering, which suggested four distinct populations (Figure 2).

Mean proportion of SNPs with zero IBS within populations was similar for the two metapopulations using all 30 samples, Glacier (0.023 ± 0.005 *SD*) and Beartooth Absaroka (0.024 ± 0.003 *SD*). Intrapopulation mean kinship was also comparable between Glacier (−0.002 ± 0.067 *SD*) and Beartooth Absaroka (0.003 ± 0.033 *SD*). As expected for a population composed of unrelated individuals, Glacier and the Beartooth Absaroka had average population mean kinship values that were close to 0 (Manichaikul et al., 2010). Estimates based on all 86 samples from the Beartooth Absaroka were 0.032 ± 0.004 *SD* for probability of zero IBS and 0.020 ± 0.033 *SD* for mean kinship, which were slightly higher than values obtained from the subsample of 30 for this herd. The two smaller populations with more complex herd histories had greater within-herd kinship than the metapopulations (Glacier or Beartooth Absaroka) for both mean proportion of SNPs with zero IBS using all 30 samples (Fergus x = 0.019 ± 0.005 *SD*; Taylor-Hilgard x = 0.018 ± 0.005 *SD*) and mean kinship (Fergus x = 0.045 ± 0.063 *SD*; Taylor-Hilgard x = 0.064 ± 0.055 *SD*).

Interpopulation metric estimates using 30 samples from each herd (60 total genotypes) resulted in differences in relative ranking of relationships between herds, depending on the metric employed. Application of F_{ST} resulted in the Fergus versus Beartooth Absaroka comparison estimated to be most distantly related (0.086), followed by Fergus versus Glacier (0.080), Fergus versus Taylor-Hilgard (0.079), Glacier versus Taylor-Hilgard (0.075) and Glacier versus Beartooth Abaroka (0.071). Taylor-Hilgard and Beartooth Absaroka were the most closely related of the F_{ST} estimates (0.032). Mean proportion of SNPs with zero IBS resulted in slightly different ranking of interpopulation estimates, with Glacier versus Beartooth estimated to be the most distantly related (0.0363 ± 0.004 *SD*), followed by Fergus versus Beartooth Absaroka (0.0361 ± 0.004 *SD*), Fergus versus Glacier (0.034 ± 0.005 *SD*), Glacier versus Taylor-Hilgard (0.033 ± 0.005 *SD*), Taylor-Hilgard versus Beartooth Absaroka (0.032 ± 0.004 *SD*). Fergus and Taylor-Hilgard were estimated to be most closely related using IBS (0.029 ± 0.005 *SD*). Mean kinship estimates between populations resulted in similar relative ranking of estimates as the IBS approach, except Fergus versus Beartooth Absaroka (−0.188 ± 0.108 *SD*) was estimated to be more distantly related than Glacier and Beartooth Absaroka (−0.182 ± 0.088 *SD*). All interpopulation mean kinship values were estimated to be negative, indicating relatively low kinship among the populations.

3.2 | Sample size simulations

For intrapopulation metric estimates, sample size influenced whether the range of estimates provided by each sample size overlapped for populations with different management histories (Figure 3). We calculated one minus mean kinship so that relative kinship estimates among herds would be comparable to IBS results. At smaller sample sizes of $n \leq 15$, the two metapopulation (Glacier and Beartooth Absaroka) distributions overlapped mean kinship and IBS distributions of herds with more complex management histories (Fergus and Taylor-Hilgard). Increasing differentiation among herds was noticeable at n = 20 for mean proportion of SNPs with zero IBS, and the two different types of management histories were clearly differentiated at a sample size of 25 for IBS and mean kinship metrics. Increasing sample size also resulted in decreasing RMSE across all populations, regardless of management history. RMSE was primarily influenced by variance, as bias relative to the mean for n = 30 was

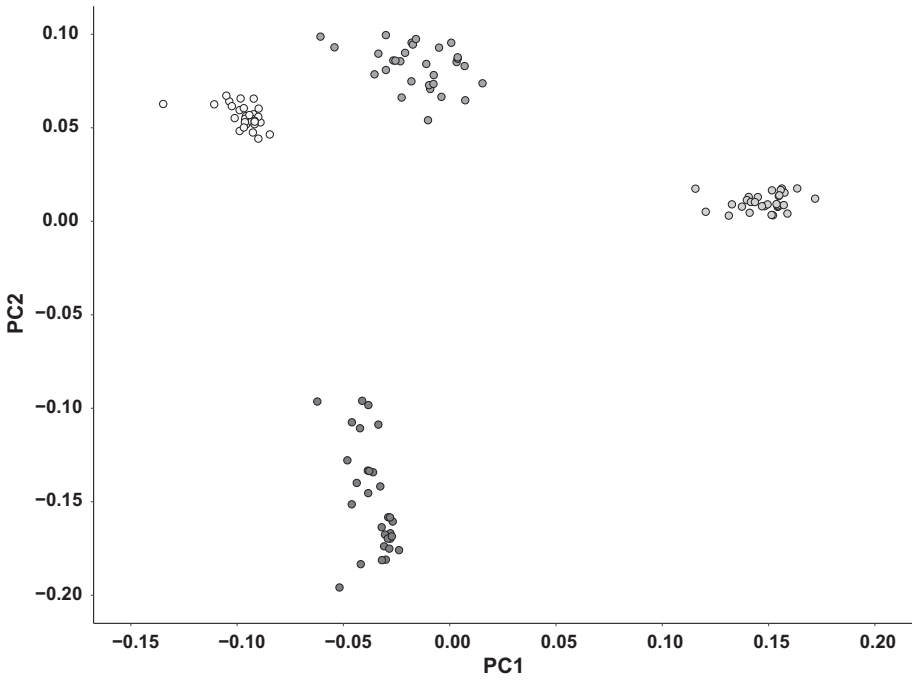

FIGURE 2 Principal component 1 (PC1; eigenvalue = 4.03) plotted against principal component 2 (PC2; eigenvalue = 3.40) for 30 Rocky Mountain bighorn sheep (*Ovis canadensis canadensis*) per population from four different herds. Different populations are indicated by colour, including Beartooth Absaroka (light grey), Fergus (white), Glacier (black) and Taylor-Hilgard (dark grey). PCA analysis included 7,688 SNPs and was completed using Golden Helix SNP & Variation Suite v8.6 software (SNP & Variation Suite, n.d.)

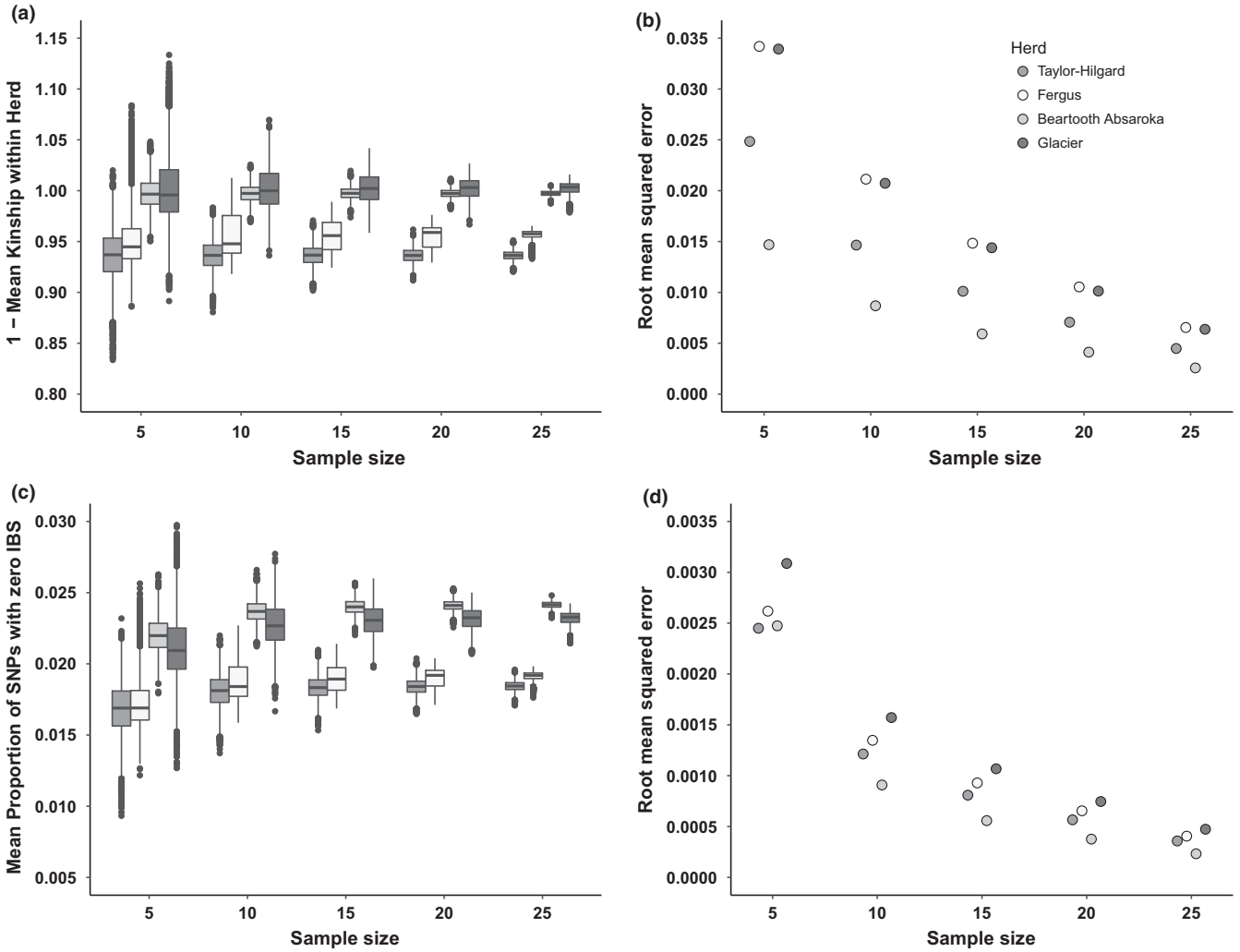

FIGURE 3 Boxplots of intrapopulation metric estimates based on 10,000 replicate simulations using empirical SNP genotypes from populations of Rocky Mountain bighorn sheep (*Ovis canadensis canadensis*), including one minus mean kinship (a) and mean proportion of SNPs with zero identity by state (IBS; c) by increasing sample size. Centre lines represent the median, box limits represent the 25th and 75th percentiles, whiskers indicate 1.5 multiplied by the interquartile range from the 25th and 75th percentiles, points represent outliers. Root mean squared error (RMSE) is plotted for mean kinship within herd (b) and mean proportion of SNPs with zero identity by state (IBS; d). Different populations are indicated by colour, including Beartooth Absaroka (light grey), Fergus (white), Glacier (black) and Taylor-Hilgard (dark grey)

relatively low across all sample sizes. Evaluation of intrapopulation metrics using all 86 samples from the Beartooth Absaroka (Supporting Information Figure S1) showed that RMSE continued to decrease at a slower rate beyond the smaller sample sizes of $n = 5$ to $n = 20$ and declined towards zero at a sample size greater than $n = 25$. In general, uncertainty in estimates indicated that we could not confidently discern differences in IBS and mean kinship between herds of differing population histories at sample size 15.

Similar to our results for the intrapopulation simulation results, all metric estimates based on sample sizes of 5–15 had high variance and similar distributions across population comparisons (Figure 4). Furthermore, RMSE decreased with increasing sample size per herd and was generally not affected by the estimated relative kinship among herds. Increasing sample size changed the mean F_{ST} estimates the most of all three metrics across the 10,000 replicates, likely due to required filtering implemented for each individual replicate. As sample size increased, the number of SNPs removed at each filtering stage generally decreased. Thus, greater sample sizes included a disproportionately greater number of SNPs in the F_{ST} calculation than smaller sample sizes, which resulted in a greater change in the mean estimate for the 10,000 replicates than observed for the other metrics (Supporting Information Figure S2). Within the simulation, we used a filtering process that typically is applied to calculate F_{ST} given a certain sample size. The differences in mean F_{ST} estimates by sample size due to the filtering process demonstrated that a comparison of F_{ST} estimates using different sample sizes has the potential to be especially problematic for this metric. For sample sizes 5–25, the distribution of F_{ST} estimates for many of the population comparisons overlapped, with greater differentiation among estimates at sample sizes of 20 and 25. Estimates of mean proportion of SNPs with zero IBS provided slightly different results and provided more nonoverlapping distributions of estimates than F_{ST} estimates at sample sizes of

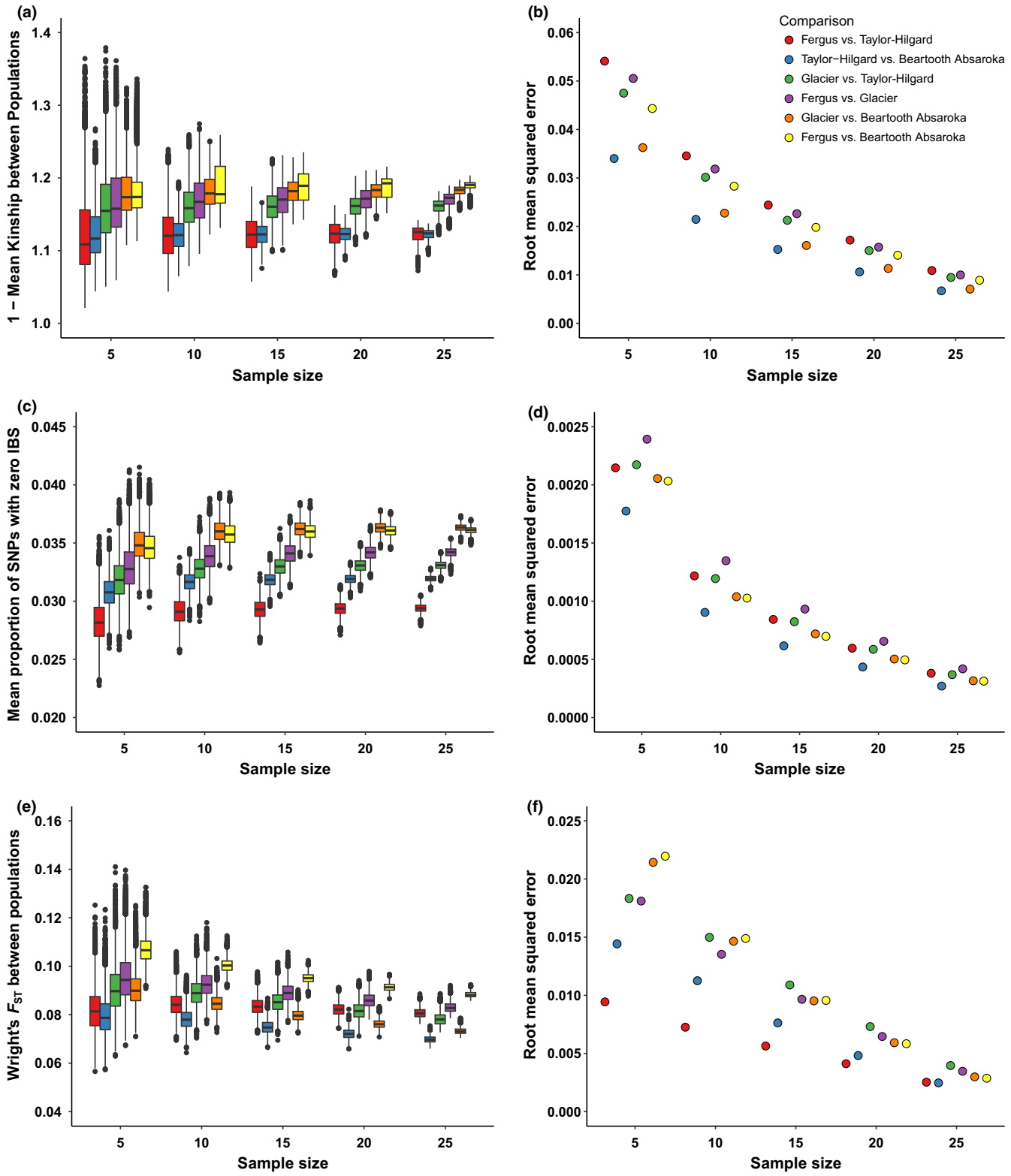

FIGURE 4 Boxplots of interpopulation metric estimates based on 10,000 replicate simulations using empirical SNP genotypes from populations of Rocky Mountain bighorn sheep (*Ovis canadensis canadensis*), including one minus mean kinship (a), mean proportion of SNPs with zero identity by state (IBS; c) and Wright's F_{ST} (e) by increasing sample size per individual population included. Centre lines represent the median, box limits represent the 25th and 75th percentiles, whiskers indicate 1.5 multiplied by the interquartile range from the 25th and 75th percentiles, points represent outliers. Root mean squared error (RMSE) is plotted for mean kinship within herd (b) and mean proportion of SNPs with zero identity by state (IBS; d), and Wright's F_{ST} (f). Different population comparisons are indicated by colour

20 and 25. Mean kinship simulations resulted in similar overall results to IBS estimates, with decreases in overlap with increasing sample size, but distributions overlapped only up to a sample size of 20. At sample size of 25, populations that were most related had distributions of mean kinship estimates that could be distinguished from populations that were estimated to be the least related.

By evaluating our simulation results, we concluded that a sample size of 25 was adequate for intrapopulation and interpopulation genetic management metric assessments of bighorn sheep populations using the HD Ovine array. Clear differences in estimates between herds for intrapopulation metrics and between herd comparisons for interpopulation metrics were apparent across metric types at a sample size of 25. In addition, RMSE consistently decreased with increasing sample size across all metrics and populations, regardless of management history. For the intrapopulation metrics, a sample size of 25 allowed for differentiation in estimate results among herds with different management histories, as the metapopulations (Glacier and Beartooth Absaroka) had lower mean intrapopulation kinship and IBS than the relatively small native herd (Taylor-Hilgard) and the reintroduced herd (Fergus; Figure 3). Interpopulation metric results also supported a sample size of 25 to discern meaningful differences among herd comparisons for all metrics (Figure 4). The relative ranking of herd comparisons differed slightly for F_{ST} estimates, in comparison with mean kinship and IBS, which is likely due to the metric's differing approach and filtering requirements (Supporting Information Figure S2).

As the simulations drawn from a total sample size of 30 had less independence among replicates for larger sample sizes, the standard deviation estimates for the sample size of 25 may be slightly underestimated. To evaluate the extent of this issue, we compared the standard deviation of the Beartooth Absaroka data set at a sample size of 25 drawing from 30 samples and to that drawing from 86 samples. For the 86 sample size pool, pairs of replicate data sets will on average share about 7 of 25 individuals. The standard error is roughly double for $n = 25$ (0.00547) drawing from 86 samples in comparison with $n = 25$ using 30 samples (0.00258; Supporting Information Table S3). Even if the standard error estimates are actually double those presented in that figure, the distributions of estimates for each population would still have little overlap (Figure 3a). The results of intrapopulation simulations for the Beartooth Absaroka using 86 samples (Supporting Information Figure S1) also suggested that a sample size greater than 25 can further reduce RMSE. Furthermore, the inclusion of additional Beartooth Absaroka samples resulted in slightly different within-herd mean kinship and IBS estimates, in comparison with the 30-sample estimate. The slight change in estimates is likely due to the 30 samples being selected across a wide geographic area, whereas the 86-sample approach included more individuals from similar areas, which impacted overall estimated kinship and IBS. If a reduced RMSE is important to study objectives, these results indicate that increased sample size continues to improve RMSE beyond a sample size of 25 for within-herd kinship estimates. Complete tables of our simulation results can be found in Supporting Information Tables S1–S3.

4 | DISCUSSION

The framework outlined here provides an approach to identify the optimal sample size for three different common relationship metrics to facilitate comparing results among different populations, as well as quantify expected variance of estimates. As genomic metrics are estimates, rather than exact measures of quantities of interest, the inherent uncertainty due to sampling is important to consider. Uncertainty can be introduced through selection of a limited number of individuals to represent the population of interest. For the evaluated bighorn sheep populations, a sample size of less than 20–25 would introduce an unacceptable level of uncertainty to estimate both within and between populations for the selected metrics. The sample size of 8–10 recommended by Nazareno et al. (2017) using plant genomic data would not provide clear results to compare any of our examined metrics among herds using our bighorn sheep genotype data set. For example, transformed mean population kinship estimates at a sample size of 10 for Glacier ranged from 0.936 to 1.070 and for Fergus ranged from 0.918 to 1.012. However, differences in within-herd kinship were detected at a sample size of 25, which provided estimates for Glacier ranging from 0.979 to 1.016 and for Fergus ranging from 0.933 to 0.966 (Figure 3a).

Thus, we suggest that a universal sample size rule is unlikely to exist given the complexities that impact genomic kinship and inbreeding estimates. These complexities include not only the molecular markers genotyped, but also the spatial distribution of samples and the genetic characteristics of the individuals and populations examined. The Beartooth Absaroka generally had lower RMSE for within-herd mean kinship (Figure 3b,d), suggesting that a combination of population allele frequencies and the markers genotyped can result in slightly different variance and relative bias by population. Individual genetic samples are often nonrandomly collected at low numbers, due to the cost and logistical difficulties of live animal capture, and a small sample of convenience may influence inferences from genetic management metric estimates. Uncertainty in estimates provided by our bighorn sheep empirical simulations indicated that we cannot confidently discern differences in intrapopulation kinship and IBS between herds of differing population histories at commonly used sample sizes of 15 and lower (Figure 3).

Depending on the samples drawn using a sample size of 15 individuals for each herd, one might incorrectly infer the Glacier metapopulation had higher average within-herd mean kinship than the Fergus and Taylor-Hilgard herds with more complex histories. However, when we increased sampling intensity by a modest amount (5–10 samples), the estimates provided more accurate inferences, such that we detected that the metapopulations had lower within-population mean kinship and IBS than the smaller herds. Similar to that, between population mean kinship, IBS and F_{ST} estimates required higher sample sizes for precise estimates, and relative differences among the comparisons were not clear for sample sizes 15–20 or lower (Figure 4). Small sample sizes have the potential to result in a lack of clarity in the relative comparisons of estimates, which limits our ability to accrue reliable knowledge, effectively

address ecological questions, and make management decisions. For example, mean kinship can be useful for making translocation decisions to maximize genetic diversity, by identifying the best source population for a translocation based on which candidate population is least related to the recipient population (Frankham et al., 2017). Thus, it would be beneficial for researchers to have a resource that can help with the process of deriving meaningful inferences based on comparing genomic estimates of mean kinship among populations within the same study area or among studies.

Researchers can add rigour by increasing sample size per population and using consistent sample sizes across compared populations to ensure comparable precision. To determine how many samples are adequate for study goals, a pilot study and simulations using our provided R code with actual genotype data from a select number of populations can help determine what sample size is required to detect differences among sampled herds that are deemed biologically meaningful. A challenging decision at the simulation stage is to determine the level of investment necessary to create a group of samples to use in a simulation context. One limitation of our work is the use of only 30 genotypes from which replicates were drawn, and this resulted in lack of independence among replicates. We sampled to the extent possible given the expense of captures and genotyping, and similar nonindependence among replicates is a common issue among other empirical simulations (Gaughran et al., 2017; Nazareno et al., 2017). Employing a greater number of samples than 30 would allow us to generate more unique replicates. We can address this limitation by evaluating the intrapopulation simulations for the Beartooth Absaroka using 86 samples (Supporting Information Figure S1). The within-population mean kinship and IBS simulations for the Beartooth Absaroka using 86 samples suggested that a sample size of 25 and greater can further reduce RMSE. Thus, if the largest sample size available in the simulation is unacceptably small, the relative bias and RMSE calculations can be misrepresented. The Beartooth Absaroka results indicated that researchers should look for a decrease in RMSE as sample size increases, which suggests that the maximum sample size available may be an acceptable reference point (Supporting Information Figure S1). If the RMSE decreases dramatically or the mean simulation estimate experiences large changes at all examined sample sizes, as seen from sample sizes 5–15 in our simulation results, it may be necessary to examine a greater maximum sample size for effective sample size decisions. We suggest that the effect of nonindependent replicates in empirical simulations with limited data is an important area for future research.

A simulation approach to evaluate sample size not only provides a procedure to standardize uncertainty as effectively as possible, but also provides more information to the scientific community to draw biological inferences. The standard deviation of the mean for all simulation replicates of a certain sample size serves as an appropriate estimate of the empirical standard error of the mean, as long as it is reasonable to assume that 30 sample estimates for our selected metrics are adequate to represent truth. The extent to which this is true depends on the proportion of the population that was sampled and the amount of variation in kinship and inbreeding found in the population. Thus, our standard errors should be viewed as approximate values for the range of standard error values that can be expected for metric estimates at a specified sample size. For example, the standard deviation or standard error of the mean for mean kinship simulation replicates for Fergus was 0.0211 for a sample size of 10, which could be interpreted to indicate that one could only expect to estimate the mean kinship for the population within $\pm$0.0422 for an estimate of 0.0448. In contrast, with a sample of 25 from the Fergus population, the standard deviation was reduced to 0.0065, so that the mean kinship estimate could be calculated within $\pm$0.013 for an estimate of 0.0445. The level of variance that is acceptable can vary by study goals and inherent differences among populations examined. For making management decisions, Frankham et al. (2017) recommend detecting differences in mean kinship at a scale of 0.10 or finer. However, given our data set and filtering protocols, we found that slightly greater precision was necessary to distinguish between different populations.

Bighorn sheep captures can be expensive and relatively difficult, so we promoted genetic sampling by collaborating agencies of all animals captured for management actions and nongenetic research objectives such as collaring or disease monitoring, given the low cost of collecting gene cards and biopsy punch samples. Our motivation to evaluate sample size was to inform managers as to the number of samples that should be genotyped per bighorn sheep herd for a future large-scale genomic assessment of additional populations, by conducting simulations with a small number of herds with a range of possible management histories. In addition to providing sampling insights for our own study, we think that the herds used in our simulation may capture the range of attributes of bighorn herds throughout western North America, and our findings can inform sample size decisions for population genomic assessment of other bighorn sheep herds when the HD Ovine array is employed.

For other species, our approach can be applied to conduct sample size simulations specific to the examined species and molecular markers to provide information regarding the sampling required for evaluation of population genetic metrics. When animals are routinely captured for research and management purposes, it would be worthwhile for field personnel to collect genetic samples from all captured animals and build an archive of samples at a relatively low expense. In the event that the management agency or research entity has interest in research questions that can be addressed through population genomics, this archive would enable geneticists to be much more efficient with resources and have samples available for future genomic techniques that may be developed. By conducting a small simulation pilot study on a subset of available samples, they can select an optimal number of samples per population to genotype for the larger study. An additional consideration that can be important in a pilot study may be the proportion of the population of interest captured within a sample. When the sample includes $\geq$5% of the actual population with an accurate population size estimate, biologists can use a finite population correction factor to more accurately estimate the standard deviation of mean kinship. Our approach did

not include a finite population correction factor, and thus our standard deviation estimates may be conservative for Taylor-Hilgard, our smallest population examined, with an estimated population size of 280. In regard to very large populations, spatial distribution of the selected sample size may become increasingly important and the reference sample size may need to be larger than 30 to successfully capture a representative sample that accurately describes mean kinship across the area of interest, particularly if the population may exhibit genetic spatial structure.

Selecting an appropriate sample size and making that decision consistent across populations of interest would enhance population genetic inferences and serve as an informative alternative to sampling based on convenience. Our suggested simulation method may not be possible for studies concerning species or populations that are extremely rare or difficult to capture, and in this case, researchers are constrained to the available genetic samples. However, this method can be relatively easily implemented with the use of other genomic marker types for species that are accessible or easily captured. Our annotated R code (Supporting Information Appendix S1) employs commonly used software, data formatting and metrics for straightforward application to any genomic data set. There are many more population genetic metrics than those included in this study, and we expect our workflow can be easily adapted to include almost any alternative genetic metric within the simulation script. When a sample size simulation is not feasible for a particular study, researchers could apply insight from other comparable species with similar marker sets to establish a reasonable sample size target. As genomics continues to become an increasingly important approach to address questions for both ecologists and wildlife managers, we recommend that sample size simulations be conducted to help standardize precision of results across evaluated populations to indicate the necessary sample size for research regarding relationship inferences and population structure. Rigorously evaluating uncertainty and adapting sample size decisions to each unique problem can serve to enhance inference reliability and maximize comparability of studies that estimate inbreeding and kinship.

ACKNOWLEDGEMENTS

We would like to thank individuals that contributed to this work at Montana Fish, Wildlife and Parks: Neil Anderson, Emily Almberg, Jennifer Ramsey, Keri Carson, Justin Gude, Julie Cunningham, Bruce Sterling, Sonja Anderson, and Vanna Boccadori; Wyoming Game and Fish Department: Hank Edwards, Doug McWhirter; U.S. Geological Survey: Kim Keating; Glacier National Park: Mark Biel, Yellowstone National Park: P.J. White; and the University of Wyoming: Holly Ernest and Sierra Love Stowell. Funding sources included Montana Fish, Wildlife and Parks, National Science Foundation Graduate Research Fellowship, Montana and Midwest Chapters of the Wild Sheep Foundation, Glacier National Park, National Geographic Society Young Explorer grant, Holly Ernest at the University of Wyoming, the Montana Agricultural Experiment Station, Yellowstone National Park, Canon USA Inc., and Wyoming Governor's Big Game License Coalition. Danielle Walker conducted laboratory work to extract and assess quality of bighorn sheep genetic samples. Certain computational efforts were performed on the Hyalite High-Performance Computing System, operated and supported by University Information Technology Research Cyberinfrastructure at Montana State University. Katherine Ralls, Mark Miller and two anonymous reviewers provided comments that helped to improve an earlier version of this manuscript. Any use of trade, product or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

AUTHOR CONTRIBUTIONS

R.G. led collaborations to obtain DNA samples, T.G. provided samples, R.G. and E.F. collected DNA samples in the field and designed research, E.F. and J.T. completed laboratory work, J.R. and R.G. contributed to the analysis, E.F. analysed data and wrote the manuscript, and all authors edited the manuscript and approved the final version.

DATA ACCESSIBILITY

SNP genotypes are available from FigShare (https://doi.org/10.6084/m9.figshare.6157691). R code is available on GitHub (https://doi.org/10.5281/zenodo.1229504) and in Supporting Information Appendix S1.

REFERENCES

Acevedo-Whitehouse, K., Vicente, J., Gortazar, C., Hoefle, U., Fernandez-de-Mera, I. G., & Amos, W. (2005). Genetic resistance to bovine tuberculosis in the Iberian wild boar. *Molecular Ecology*, *14*(10), 3209–3217. https://doi.org/10.1111/j.1365-294X.2005.02656.x

Albrechtsen, A., Nielsen, F. C., & Nielsen, R. (2010). Ascertainment biases in SNP chips affect measures of population divergence. *Molecular Biology and Evolution*, *27*(11), 2534–2547. https://doi.org/10.1093/molbev/msq148

Besser, T. E., Cassirer, E. F., Potter, K. A., VanderSchalie, J., Fischer, A., Knowles, D. P., … Srikumaran, S. (2008). Association of Mycoplasma ovipneumoniae infection with population-limiting respiratory disease in free-ranging Rocky Mountain bighorn sheep (Ovis canadensis canadensis). *Journal of Clinical Microbiology*, *46*(2), 423–430. https://doi.org/10.1128/JCM.01931-07

Besser, T. E., Highland, M. A., Baker, K., Cassirer, E. F., Anderson, N. J., Ramsey, J. M., … Jenks, J. A. (2012). Causes of pneumonia epizootics among bighorn sheep, western United States, 2008–2010. Retrieved from http://digitalcommons.unl.edu/icwdm_usdanwrc/1101/

Blouin, M. S. (2003). DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. *Trends in Ecology & Evolution*, *18*(10), 503–511. https://doi.org/10.1016/S0169-5347(03)00225-8

Bunch, T. D., Wu, C., Zhang, Y.-P., & Wang, S. (2006). Phylogenetic analysis of snow sheep (Ovis nivicola) and closely related taxa. *Journal of Heredity*, *97*(1), 21–30. https://doi.org/10.1093/jhered/esi127

Cassirer, E. F., Plowright, R. K., Manlove, K. R., Cross, P. C., Dobson, A. P., Potter, K. A., & Hudson, P. J. (2013). Spatio-temporal dynamics of pneumonia in bighorn sheep. *Journal of Animal Ecology*, *82*(3), 518–528. https://doi.org/10.1111/1365-2656.12031

Cassirer, E. F., & Sinclair, A. R. E. (2007). Dynamics of pneumonia in a bighorn sheep metapopulation. *The Journal of Wildlife Management*, *71*(4), 1080–1088. https://doi.org/10.2193/2006-002

Coltman, D. W., Pilkington, J., Kruuk, L. E. B., Wilson, K., & Pemberton, J. M. (2001). Positive genetic correlation between parasite resistance and body size in a free-living ungulate population. *Evolution*, *55*(10), 2116–2125. https://doi.org/10.1111/j.0014-3820.2001.tb01326.x

Csilléry, K., Johnson, T., Beraldi, D., Clutton-Brock, T., Coltman, D., Hansson, B., … Pemberton, J. M. (2006). Performance of marker-based relatedness estimators in natural populations of outbred vertebrates. *Genetics*, *173*(4), 2091–2101. https://doi.org/10.1534/genetics.106.057331

Daetwyler, H. D., Capitan, A., Pausch, H., Stothard, P., Van Binsbergen, R., Brøndum, R. F., … Grohs, C. (2014). Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nature Genetics*, *46*(8), 858–865. https://doi.org/10.1038/ng.3034

de Cara, M. Á. R., Villanueva, B., Toro, M. Á., & Fernández, J. (2013). Using genomic tools to maintain diversity and fitness in conservation programmes. *Molecular Ecology*, *22*(24), 6091–6099. https://doi.org/10.1111/mec.12560

Fernández, J., Clemente, I., Amador, C., Membrillo, A., Azor, P., & Molina, A. (2012). Use of different sources of information for the recovery and genetic management of endangered populations: Example with the extreme case of Iberian pig Dorado strain. *Livestock Science*, *149*(3), 282–288. https://doi.org/10.1016/j.livsci.2012.07.019

Flesch, E., & Graves, T. A. (2018). Bighorn sheep HD Ovine array genotypes from Glacier National Park, Montana, USA, 2004–2011: U.S. Geological Survey data release. http://doi.org/10.6084/m9.figshare.6207605

Frankham, R., Ballou, J. D., Ralls, K., Eldridge, M. D. B., Dubash, M., Fenster, C. B., … Sunnucks, P. (2017). *Genetic management of fragmented animal and plant populations*. Oxford, UK: Oxford University Press. Retrieved from https://books.google.com/ebooks/app#reader/bEsrDwAAQBAJ/GBS.PA268

Fumagalli, M. (2013). Assessing the effect of sequencing depth and sample size in population genetics inferences. *PLoS ONE*, *8*(11), e79667. https://doi.org/10.1371/journal.pone.0079667

Funk, W. C., McKay, J. K., Hohenlohe, P. A., & Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution*, *27*(9), 489–496. https://doi.org/10.1016/j.tree.2012.05.012

Gaughran, S. J., Quinzin, M. C., Miller, J. M., Garrick, R. C., Edwards, D. L., Russello, M. A., … Caccone, A. (2017). Theory, practice, and conservation in the age of genomics: The Galápagos giant tortoise as a case study. *Evolutionary Applications*, 1–10. https://doi.org/10.1111/eva.12551

Gómez-Romano, F., Villanueva, B., Rodríguez de Cara, M. Á., & Fernández, J. (2013). Maintaining genetic diversity using molecular coancestry: The effect of marker density and effective population size. *Genetics Selection Evolution*, *45*, 38. https://doi.org/10.1186/1297-9686-45-38

Grueber, C. E., Laws, R. J., Nakagawa, S., & Jamieson, I. G. (2010). Inbreeding depression accumulation across life-history stages of the endangered takahe. *Conservation Biology*, *24*(6), 1617–1625. https://doi.org/10.1111/j.1523-1739.2010.01549.x

Gueijman, A., Ayali, A., Ram, Y., & Hadany, L. (2013). Dispersing away from bad genotypes: The evolution of Fitness-Associated Dispersal (FAD) in homogeneous environments. *BMC Evolutionary Biology*, *13*, 125. https://doi.org/10.1186/1471-2148-13-125

Hale, M. L., Burg, T. M., & Steeves, T. E. (2012). Sampling for microsatellite-based population genetic studies: 25 to 30 individuals per population is enough to accurately estimate allele frequencies. *PLoS ONE*, *7*(9), e45170. https://doi.org/10.1371/journal.pone.0045170

Haynes, G. D., & Latch, E. K. (2012). Identification of novel single nucleotide polymorphisms (SNPs) in deer (Odocoileus spp.) using the BovineSNP50 BeadChip. *PLoS ONE*, *7*(5), e36536. https://doi.org/10.1371/journal.pone.0036536

Hoban, S. (2014). An overview of the utility of population simulation software in molecular ecology. *Molecular Ecology*, *23*(10), 2383–2401. https://doi.org/10.1111/mec.12741

Hoban, S., & Schlarbaum, S. (2014). Optimal sampling of seeds from plant populations for ex-situ conservation of genetic biodiversity, considering realistic population structure. *Biological Conservation*, *177*, 90–99. https://doi.org/10.1016/j.biocon.2014.06.014

Hogg, J. T., Forbes, S. H., Steele, B. M., & Luikart, G. (2006). Genetic rescue of an insular population of large mammals. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1593), 1491–1499. https://doi.org/10.1098/rspb.2006.3477

Holsinger, K. E., & Weir, B. S. (2009). Genetics in geographically structured populations: Defining, estimating and interpreting FST. *Nature Reviews Genetics*, *10*(9), 639–650. https://doi.org/10.1038/nrg2611

Huisman, J., Kruuk, L. E. B., Ellis, P. A., Clutton-Brock, T., & Pemberton, J. M. (2016). Inbreeding depression across the lifespan in a wild mammal population. *Proceedings of the National Academy of Sciences*, *113*(13), 3585–3590. https://doi.org/10.1073/pnas.1518046113

Jeffries, D. L., Copp, G. H., Lawson Handley, L., Olsén, K. H., Sayer, C. D., & Hänfling, B. (2016). Comparing RADseq and microsatellites to infer complex phylogeographic patterns, an empirical perspective in the Crucian carp, Carassius carassius. *L. Molecular Ecology*, *25*(13), 2997–3018. https://doi.org/10.1111/mec.13613

Kalinowski, S. T. (2004). Do polymorphic loci require large sample sizes to estimate genetic distances? *Heredity*, *94*(1), 33–36. https://doi.org/10.1038/sj.hdy.6800548

Kijas, J. W., Porto-Neto, L., Dominik, S., Reverter, A., Bunch, R., McCulloch, R., … McEwan, J. (2014). Linkage disequilibrium over short physical distances measured in sheep using a high-density SNP chip. *Animal Genetics*, *45*(5), 754–757. https://doi.org/10.1111/age.12197

Kijas, J. W., Townley, D., Dalrymple, B. P., Heaton, M. P., Maddox, J. F., McGrath, A., … for the International Sheep Genomics Consortium (2009). A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS ONE*, *4*(3), e4668. https://doi.org/10.1371/journal.pone.0004668

Kohn, M. H., Murphy, W. J., Ostrander, E. A., & Wayne, R. K. (2006). Genomics and conservation genetics. *Trends in Ecology & Evolution*, *21*(11), 629–637. https://doi.org/10.1016/j.tree.2006.08.001

Kristensen, T. N., Pedersen, K. S., Vermeulen, C. J., & Loeschcke, V. (2010). Research on inbreeding in the 'omic' era. *Trends in Ecology & Evolution*, *25*(1), 44–52. https://doi.org/10.1016/j.tree.2009.06.014

Kruuk, L. E. (2004). Estimating genetic parameters in natural populations using the 'animal model'. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *359*(1446), 873–890. https://doi.org/10.1098/rstb.2003.1437

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, *475*(7357), 493–496.

Manel, S., Joost, S., Epperson, B. K., Holderegger, R., Storfer, A., Rosenberg, M. S., … Fortin, M.-J. (2010). Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Molecular Ecology*, *19*(17), 3760–3772. https://doi.org/10.1111/j.1365-294X.2010.04717.x

Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., & Chen, W.-M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics*, *26*(22), 2867–2873. https://doi.org/10.1093/bioinformatics/btq559

May, R. M. (2004). Uses and abuses of mathematics in biology. *Science*, *303*(5659), 790–793. https://doi.org/10.1126/science.1094442

Meirmans, P. G. (2015). Seven common mistakes in population genetics and how to avoid them. *Molecular Ecology*, *24*(13), 3223–3231. https://doi.org/10.1111/mec.13243

Miller, M. W. (2008). Pasteurellosis. In E. S. Williams, & I. K. Barker (Eds.), *Infectious disease of wild mammals* (pp. 330–339). Ames, IA: Iowa State University Press.

Miller, J. M., Kijas, J. W., Heaton, M. P., McEwan, J. C., & Coltman, D. W. (2012). Consistent divergence times and allele sharing measured from cross-species application of SNP chips developed for three domestic species. *Molecular Ecology Resources*, *12*(6), 1145–1150. https://doi.org/10.1111/1755-0998.12017

Miller, J. M., Moore, S. S., Stothard, P., Liao, X., & Coltman, D. W. (2015). Harnessing cross-species alignment to discover SNPs and generate a draft genome sequence of a bighorn sheep (Ovis canadensis). *BMC Genomics*, *16*(1), 397. https://doi.org/10.1186/s12864-015-1618-x

Miller, J. M., Poissant, J., Kijas, J. W., Coltman, D. W., & the International Sheep Genomics Consortium (2011). A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. *Molecular Ecology Resources*, *11*(2), 314–322. https://doi.org/10.1111/j.1755-0998.2010.02918.x

Monello, R. J., Murray, D. L., & Cassirer, E. F. (2001). Ecological correlates of pneumonia epizootics in bighorn sheep herds. *Canadian Journal of Zoology*, *79*(8), 1423–1432. https://doi.org/10.1139/z01-103

Morin, P. A., Moore, J. J., Chakraborty, R., Jin, L., Goodall, J., & Woodruff, D. S. (1994). Kin selection, social structure, gene flow, and the evolution of chimpanzees. *Science*, *265*, 1193–1201. https://doi.org/10.1126/science.7915048

Nazareno, A. G., Bemmels, J. B., Dick, C. W., & Lohmann, L. G. (2017). Minimum sample sizes for population genomics: An empirical study from an Amazonian plant species. *Molecular Ecology Resources*, *17*, 1136–1147. https://doi.org/10.1111/1755-0998.12654

Nielsen, J. F., English, S., Goodall-Copestake, W. P., Wang, J., Walling, C. A., Bateman, A. W., … Thavarajah, N. K. (2012). Inbreeding and inbreeding depression of early life traits in a cooperative mammal. *Molecular Ecology*, *21*(11), 2788–2804. https://doi.org/10.1111/j.1365-294X.2012.05565.x

Paetkau, D., Slade, R., Burden, M., & Estoup, A. (2004). Genetic assignment methods for the direct, real-time estimation of migration rate: A simulation-based exploration of accuracy and power. *Molecular Ecology*, *13*(1), 55–65. https://doi.org/10.1046/j.1365-294X.2004.02008.x

Pew, J., Muir, P. H., Wang, J., & Frasier, T. R. (2015). Related: An R package for analysing pairwise relatedness from codominant molecular markers. *Molecular Ecology Resources*, *15*(3), 557–561. https://doi.org/10.1111/1755-0998.12323

Poissant, J., Hogg, J. T., Davis, C. S., Miller, J. M., Maddox, J. F., & Coltman, D. W. (2010). Genetic linkage map of a wild genome: Genomic structure, recombination and sexual dimorphism in bighorn sheep. *BMC Genomics*, *11*, 524. https://doi.org/10.1186/1471-2164-11-524

Pool, J. E., Hellmann, I., Jensen, J. D., & Nielsen, R. (2010). Population genetic inference from genomic sequence variation. *Genome Research*, *20*(3), 291–300. https://doi.org/10.1101/gr.079509.108

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., … Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

R Core Team (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Robinson, S. P., Simmons, L. W., & Kennington, W. J. (2013). Estimating relatedness and inbreeding using molecular markers and pedigrees: The effect of demographic history. *Molecular Ecology*, *22*(23), 5779–5792. https://doi.org/10.1111/mec.12529

Romanov, M. N., Tuttle, E. M., Houck, M. L., Modi, W. S., Chemnick, L. G., Korody, M. L., … Ryder, O. A. (2009). The value of avian genomics to the conservation of wildlife. *BMC Genomics*, *10*(2), 1–19. https://doi.org/10.1186/1471-2164-10-S2-S10

Roughsedge, T., Pong-Wong, R., Woolliams, J. A., & Villanueva, B. (2008). Restricting coancestry and inbreeding at a specific position on the genome by using optimized selection. *Genetics Research*, *90*(2), 199–208. https://doi.org/10.1017/S0016672307009214

Santure, A. W., Stapley, J., Ball, A. D., Birkhead, T. R., Burke, T., & Slate, J. O. N. (2010). On the use of large marker panels to estimate inbreeding and relatedness: Empirical and simulation studies of a pedigreed zebra finch population typed at 771 SNPs. *Molecular Ecology*, *19*(7), 1439–1451. https://doi.org/10.1111/j.1365-294X.2010.04554.x

Saura, M., Fernández, A., Rodríguez, M. C., Toro, M. A., Barragán, C., Fernández, A. I., & Villanueva, B. (2013). Genome-wide estimates of coancestry and inbreeding in a closed herd of ancient iberian pigs. *PLoS ONE*, *8*(10), e78314. https://doi.org/10.1371/journal.pone.0078314

Shafer, A. B. A., Poissant, J., Côté, S. D., & Coltman, D. W. (2011). Does reduced heterozygosity influence dispersal? A test using spatially structured populations in an alpine ungulate. *Biology Letters*, *7*(3), 433–435. https://doi.org/10.1098/rsbl.2010.1119

Sheehan, S., Harris, K., & Song, Y. S. (2013). Estimating variable effective population sizes from multiple genomes: A sequentially Markov conditional sampling distribution approach. *Genetics*, *194*(3), 647–662.

Siddle, H. V., Kreiss, A., Eldridge, M. D., Noonan, E., Clarke, C. J., Pyecroft, S., … Belov, K. (2007). Transmission of a fatal clonal tumor by biting occurs due to depleted MHC diversity in a threatened carnivorous marsupial. *Proceedings of the National Academy of Sciences*, *104*(41), 16221–16226. https://doi.org/10.1073/pnas.0704580104

Slate, J., David, P., Dodds, K. G., Veenvliet, B. A., Glass, B. C., Broad, T. E., & McEwan, J. C. (2004). Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: Theoretical expectations and empirical data. *Heredity*, *93*(3), 255–265. https://doi.org/10.1038/sj.hdy.6800485

SNP & Variation Suite (n.d.). (Version 8.6.0). Bozeman, MT: Golden Helix, Inc. Retrieved from http://www.goldenhelix.com

Streiff, R., Ducousso, A., Lexer, C., Steinkellner, H., Gloessl, J., & Kremer, A. (1999). Pollen dispersal inferred from paternity analysis in a mixed oak stand of Quercus robur L. and Q. petraea(Matt.) Liebl. *Molecular Ecology*, *8*(5), 831–841. https://doi.org/10.1046/j.1365-294X.1999.00637.x

Taylor, H. R. (2015). The use and abuse of genetic marker-based estimates of relatedness and inbreeding. *Ecology and Evolution*, *5*(15), 3140–3150. https://doi.org/10.1002/ece3.1541

Taylor, H. R., Kardos, M. D., Ramstad, K. M., & Allendorf, F. W. (2015). Valid estimates of individual inbreeding coefficients from marker-based pedigrees are not feasible in wild populations with low allelic diversity. *Conservation Genetics*, *16*(4), 901–913. https://doi.org/10.1007/s10592-015-0709-1

Thrasher, D. J., Butcher, B. G., Campagna, L., Webster, M. S., & Lovette, I. J. (2018). Double-digest RAD sequencing outperforms microsatellite loci at assigning paternity and estimating relatedness: A proof of concept in a highly promiscuous bird. *Molecular Ecology Resources*, 1–13. https://doi.org/10.1111/1755-0998.12771

Toro, M., Barragán, C., Óvilo, C., Rodrigañez, J., Rodriguez, C., & Silió, L. (2002). Estimation of coancestry in Iberian pigs using molecular markers. *Conservation Genetics*, *3*(3), 309–320. https://doi.org/10.1023/A:1019921131171

Toro, M. A., Villanueva, B., & Fernández, J. (2014). Genomics applied to management strategies in conservation programmes. *Livestock Science*, *166*, 48–53. https://doi.org/10.1016/j.livsci.2014.04.020

Van de Casteele, T., Galbusera, P., & Matthysen, E. (2001). A comparison of microsatellite-based pairwise relatedness estimators. *Molecular Ecology*, *10*(6), 1539–1549. https://doi.org/10.1046/j.1365-294X.2001.01288.x

Wang, J. (2011). Coancestry: A program for simulating, estimating and analysing relatedness and inbreeding coefficients. *Molecular Ecology*

Resources, *11*(1), 141–145. https://doi.org/10.1111/j.1755-0998.2010.02885.x

Willing, E.-M., Dreyer, C., & van Oosterhout, C. (2012). Estimates of genetic differentiation measured by FST do not necessarily require large sample sizes when using many SNP markers. *PLoS ONE*, *7*(8), e42649. https://doi.org/10.1371/journal.pone.0042649

Young, S. P., & Manville, R. H. (1960). Records of bighorn hybrids. *Journal of Mammalogy*, *41*(4), 523–525. https://doi.org/10.2307/1377561

Zhu, X., Li, S., Cooper, R. S., & Elston, R. C. (2008). A unified association analysis approach for family and unrelated samples correcting for stratification. *The American Journal of Human Genetics*, *82*(2), 352–365. https://doi.org/10.1016/j.ajhg.2007.10.009

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Flesch EP, Rotella JJ, Thomson JM, Graves TA, Garrott RA. Evaluating sample size to estimate genetic management metrics in the genomics era. *Mol Ecol Resour*. 2018;00:1–15. https://doi.org/10.1111/1755-0998.12898