

# Genotyping-in-thousands by sequencing (GT-seq) of non-invasive fecal and degraded samples: a new panel to enable ongoing monitoring of Canadian polar bear populations

Kristen Hayward<sup>1</sup>, Rute Clemente-Carvalho<sup>1</sup>, Evelyn Jensen<sup>2</sup>, Peter Van Coeverden de Groot<sup>1</sup>, Marsha Branigan<sup>3</sup>, Markus Dyck<sup>4</sup>, Christina Tschirter<sup>1</sup>, Zhengxin Sun<sup>1</sup>, and Stephen Lougheed<sup>1</sup>

<sup>1</sup>Queen's University

<sup>2</sup>Yale University

<sup>3</sup>Government of the Northwest Territories

<sup>4</sup>Government of Nunavut

July 4, 2021

## Abstract

Genetic monitoring using non-invasive samples provides a complement or alternative to traditional population monitoring methods. However, Next Generation Sequencing approaches to monitoring typically require high quality DNA and the use of non-invasive samples (e.g. scat) is often challenged by poor DNA quality and contamination by non-target species. One promising solution is a highly multiplexed sequencing approach called Genotyping-in-thousands by sequencing (GT-seq), which can enable cost-efficient genomics-based monitoring for populations based on non-invasively collected samples. Here, we develop and validate a GT-seq panel of 324 single nucleotide polymorphisms (SNPs) optimized for genotyping of polar bears based on DNA from non-invasively collected fecal samples. We demonstrate 1) successful GT-seq genotyping of DNA from a range of sample sources, including successful genotyping of 85.7% of non-invasively collected fecal samples determined to contain polar bear DNA, and 2) that we can reliably differentiate individuals, ascertain sex, assess relatedness, and resolve population structure of Canadian polar bear subpopulations based on a GT-seq panel of 324 SNPs. Our GT-seq data reveal similar spatial-genetic patterns as previous polar bear studies but at lesser cost per sample and using non-invasively collected samples, indicating the potential of this approach for population monitoring. This GT-seq panel provides the foundation for a non-invasive toolkit for polar bear monitoring and contribute to community-based programs – a framework which may serve as a model for wildlife management and contribute to conservation and policy for species worldwide.

## INTRODUCTION

Informed wildlife management requires accurate demographic data that include recurring, reliable population estimates and an understanding of the factors that shape population dynamics (Durner, Laidre, & York, 2018; Hamilton & Derocher, 2019; Laidre et al., 2015). Population monitoring is especially urgent for species that are being impacted by rapid climate change, are harvested or poached, or reside in habitats that have been heavily altered or fragmented by human activity (Durner et al., 2018; Laidre et al., 2015; Robinson, Morrison, & Baillie, 2014). Such monitoring can be challenging for species that have large territories or occupy inaccessible habitat, are cryptic or elusive (e.g. nocturnal, fossorial), or are of heightened conservation concern. For such species, capture, direct handling, and invasive sampling may be impractical, inappropriate, or culturally undesirable. Moreover, traditional methods for monitoring of animal populations (e.g. aerial censusing, mark-recapture, and radiotelemetry) can be expensive, time-consuming, and stressful for the focal animals (Solberg, Bellemain, Drageset, Taberlet, & Swenson, 2006; Stapleton, Atkinson, Hedman, &

Garshelis, 2014; Van Coeverden de Groot et al., 2013). When populations have low densities, mark-recapture and aerial surveys may also be hindered by low probabilities of capture and detection, respectively (Garshelis & Noyce, 2006; Hayward, Miquelle, Smirnov, & Nations, 2002).

Genetic monitoring using non-invasive samples, such as scat, hair, feathers, or skin, affords an alternative that can mitigate some of the challenges of traditional monitoring. Non-invasive monitoring can be deployed on a large scale and with greater frequency, potentially enabling larger sample sizes and improved temporal monitoring. Other benefits of non-invasive monitoring include ease of collection, reduced disturbance to study species, potentially decreased spatial or temporal biases, and diminished physical risk to collectors (Carroll et al., 2018; Morin, Waits, McNitt, & Kelly, 2018; Steyer et al., 2016; Waits & Paetkau, 2005). Further, non-invasive genetic monitoring can provide robust and repeatable data for individual identification (including sex), movement, and population trends (e.g. Aziz et al., 2017; Quinn, Alden, & Sacks, 2019; Schmidt, Campbell, Govindarajulu, Larsen, & Russello, 2020; Schultz, Cristescu, Littleford-Colquhoun, Jaccoud, & Frere, 2018). Scat samples in particular can be useful for assessing the health of individuals and populations, as they offer a range of other information, including data on parasite and pathogen presence (Bergner et al., 2019; Cristescu et al., 2019; Weese, Salgado-Bierman, Rupnik, Smith, & van Coeverden de Groot, 2019), diet composition (e.g. via quantitative PCR or DNA metabarcoding – Iversen et al., 2013; Nelms et al., 2019; Ogurtsov, 2018), hormone profiles (Morden et al., 2011; Vynne, Baker, Breuer, & Wasser, 2012), and contaminant loads (Lundin et al., 2016; Lundin, Riffell, & Wasser, 2015).

Despite the purported advantages of non-invasive genetic monitoring, it poses challenges that have limited its widespread implementation. A major impediment is that the DNA from non-invasive samples may be degraded due to environmental exposure (Bourgeois et al. 2019; Poinar, Höss, Bada, & Pääbo, 1996; Schultz et al., 2018) and contaminated by non-target species (Carroll et al., 2018; Taberlet, Waits, & Luikart, 1999). Often less than 5% of the total DNA that scat contains is host DNA, with most DNA coming from pathogens, parasites, commensal bacteria, and prey species (Han et al., 2019; Perry, Marioni, Melsted, & Gilad, 2010; Snyder-Mackler et al., 2016). Due to low quality and quantity of host DNA, accurate quantification and genotyping of non-invasive samples has often proved difficult when using Next Generation sequencing (NGS) methods, which require high DNA concentration and quality (Maroso et al., 2018; Graham et al., 2015; Taberlet et al., 1999). Such reduced genotyping accuracy due to contaminated and degraded DNA may lead to increased processing efforts and costs, and complicate inferences from collected data.

Multiple genotyping methods that have been developed to improve NGS of low-quality samples (e.g. scat, hair, or archaeological samples) often use selectively targeted, species-specific arrays of single nucleotide polymorphism markers or SNPs (Carroll et al., 2018). Traditional NGS methods (e.g. double-digest restriction-associated DNA sequencing: ddRADseq) can be used to identify large SNP panels across a focal species genome, from which smaller, informative panels can be selected (Andrews et al., 2018; Blåhed, Königsson, Ericsson, & Spong, 2018; Hess et al., 2015). DNA capture, SNP genotyping assays (e.g. TaqMan), or amplicon sequencing methods can then be used to genotype the reduced panel with high coverage (reviewed in Meek & Larson, 2019). Indeed, SNP genotyping has been applied to non-invasive samples from a range of wildlife species and has yielded high genotyping success and low genotyping error, reducing the need for systematic replicates that increase cost and effort (Fitak, Naidu, Thompson, & Culver, 2016; Kleinman-Ruiz et al., 2017; Kraus et al., 2015; Schultz et al., 2018; von Thaden et al., 2017).

Genotyping-in-Thousands by sequencing (GT-seq) is among the most promising approaches for genotyping SNP panels from non-invasively collected DNA. It uses highly multiplexed PCR to amplify short amplicons, followed by individual barcoding that allows rapid, high-quality genotyping of targeted SNP panels across thousands of individuals (Campbell, Harmon, & Narum, 2015). GT-seq library preparation can be performed with standard lab equipment and shows decreased genotyping error and genotyping costs relative to other NGS-based genotyping methods, including TaqMan assays (Campbell et al., 2015). The technique uses suites of multiplexed, species-specific primers, which may also mitigate some of the challenges presented by exogenous DNA and degraded host DNA in non-invasive fecal samples. GT-seq has been successfully applied to minimally invasive cloacal swab DNA samples collected from Western Rattlesnakes (*Crotalus oreganus*

) with low rates of genotyping error and discordance relative to RADseq (Schmidt et al., 2020). Natesh et al. (2019) also found high genotyping success using GT-seq of non-invasive scat samples for Indian Bengal Tigers (*Panthera tigris tigris*). Thus, GT-seq has the potential to enable efficient and economical genetic monitoring of populations based on non-invasively collected samples, but for implementation, requires a clear guide for development, testing, and validation.

There is a long-expressed desire by Northern communities in Canada for monitoring practices to use non-invasively collected samples, particularly for polar bears (*Ursus maritimus*). Current polar bear monitoring is based primarily on surveys of 19 subpopulations (also called management units, MUs), designated largely using mark-recapture, radio collaring, and aerial surveys (e.g. Stapleton et al., 2014; Taylor, Laake, McLoughlin, Cluff, & Messier, 2009). Northern communities, including the Inuit - for whom polar bears are of key cultural and economic importance - have voiced concern about the invasiveness of some of these monitoring methods (e.g. mark-recapture), potential negative impacts on polar bear health and behaviour, lack of inclusion of Traditional Ecological Knowledge in monitoring and management, and lack of collaboration in management activities (Wong, Dyck, & Murphy, 2017; York, Dowsley, Cornwell, Kuc, & Taylor, 2016). Two-thirds (13) of the 19 polar bear subpopulations are fully or partially found in Canada (Figure 3), highlighting Canada's need to lead on polar bear management. However, surveys for many subpopulations are infrequent due to logistical and economic constraints, and 11 of 19 subpopulations are data deficient (Government of Canada, 2018; Hamilton & Derocher, 2019). As of 2019, only 6 subpopulations had population estimates less than 10 years old (Hamilton & Derocher, 2019). These data deficiencies preclude robust estimates of population parameters and have limited implementation of effective management strategies. Thus, as polar bears continue to be impacted by climate change and face limitations in range and prey availability as rapid sea ice decline continues (Fonturbel, Lara, Lobos, & Little, 2018; Hamilton & Derocher, 2019; Hunter et al., 2010; Rode et al., 2014), there is an ever-growing, urgent need for new monitoring approaches.

Non-invasive scat surveys may enable more direct community participation, and provide a cost-effective complement to traditional polar bear monitoring methods. Non-invasive scat samples could be obtained through community-level monitoring programs, in which Inuit hunters are remunerated for field sampling efforts and surveys are repeated regularly to better track the trajectory of polar bear subpopulations. Non-invasive scat surveys have already been used in Brown bears (*Ursus arctos*) as an alternative source of DNA to high quality samples, such as muscle (e.g. Giangregorio, Norman, Davoli, & Spong, 2019). We have also established that sufficient quantities of DNA for GT-seq protocols can be obtained and quantified from field-collected polar bear scat (Hayward et al., 2020). Thus, polar bears present an opportunity to demonstrate clearly GT-seq panel development and validation, application, and usefulness for population monitoring. Importantly, this application of GT-seq will be in collaboration with Northern Canadian communities and will have real socio-economic benefits for the communities involved.

In our study, we have two primary goals: 1. To test the practicality of using GT-seq for SNP genotyping of non-invasively collected fecal samples from polar bears and 2. To apply GT-seq to population monitoring and expand our understanding of Canadian polar bear genetic structure by using degraded samples that heretofore could not be genotyped with NGS-based methods. To identify an optimized GT-seq panel, we filtered and called SNPs from a ddRADseq dataset previously generated from high-quality, geographically representative tissue samples (Jensen et al., 2020). We validate the ability of our panel to differentiate individuals and compare its ability to diagnose population structure to a fuller suite of SNPs (i.e. all high quality SNPs identified from the ddRADseq dataset). To validate the GT-seq assay itself, we estimate missing data and genotyping success for multiple DNA sources, compare hunter reports of bear sex and GT-seq-determined sex, and estimate genotyping error for non-invasive scat samples by comparing genotypic data from harvest-derived muscle and fecal samples collected from the colons of the same individuals. Finally, we combine our new GT-seq data with samples previously genotyped at our GT-seq panel loci using ddRADseq and use this combined dataset to examine relatedness, genetic diversity, and population structure. We demonstrate an 85.7% genotyping success rate (individual must have >50% SNPs genotyped) for non-invasively collected fecal samples determined to have detectable polar bear DNA, and show that one can comprehensively characterize polar bear population structure using an optimized, cost-efficient GT-seq panel of 324 SNPs.

## METHODS

### GT-seq panel development

To identify potential SNPs for our GT-seq panel, we screened 411,094 SNP loci identified from ddRADseq data generated from 327 polar bears in Jensen et al. (2020). We filtered this dataset to retain loci that were successfully genotyped in [?] 85% of individuals and had a minor allele frequency of at least 0.25. Retained loci were thinned to one SNP per 50,000 bp using the polar bear reference genome (GenBank GCA\_000687225.1; Liu et al., 2014) to reduce possibility of linkage. These filters resulted in 442 SNP loci being retained (See *Table S1* for full filtering workflow). To allow for sex identification, we added two SNPs previously recognized as sex-linked in polar bears (Pagès et al., 2009). We validated our GT-seq panel of 442 SNPs and designed primers as described in *Supplemental Information: GT-seq Panel Development and Validation*.

### Sample collection

Our work draws upon multiple sample sources, including archived harvest tissue samples (HV;  $n = 38$ ) from bears taken by Inuit hunters, and biopsy samples (BP;  $n = 138$ ) that are housed in collections by the Nunavut and Northwest Territories governments in accordance with Government of the Northwest Territories and Government of Nunavut research permits. We also received “sample sets” from hunted bears, which contained fat, tissue, liver, and lower intestine with feces, and were collected in accordance with wildlife research permits ARI #WL 500540 to MB and WL-2019-061 to SCL. These were used to estimate fecal genotyping error by comparing genotypes from set tissue (MS;  $n = 108$ ) and feces from the colon (CF;  $n = 78$ ). Non-invasive, field-collected fecal samples (FF;  $n = 72$ ) were also located and collected by Inuit hunters in accordance with wildlife research permit WL-2018-006 to SCL. All sample types were stored in  $-20^{\circ}\text{C}$  or  $-80^{\circ}\text{C}$  at Queen’s university until subsampling. Many of these samples were too degraded for use in previous NGS-based studies and thus, represent opportunities to expand sampling coverage and draw new inferences regarding population structure. As the ultimate goal is to use our GT-seq method on field-collected scat samples, we took duplicate subsamples from 21 FF samples to assess within-sample genotyping error and variation in genotype quality. Subsamples of MS, HV, and BP were stored in 100% ethanol at  $-20^{\circ}\text{C}$  or  $-80^{\circ}\text{C}$  and CF and FF subsamples were stored without ethanol at  $-20^{\circ}\text{C}$  until DNA extraction.

All samples were collected across 11 of the 13 Canadian subpopulations between 1998 and 2019, with mean year of sample collection ranging from 1999 (M’Clintock Channel) to 2018 (Southern Hudson Bay). For 9 subpopulations, we had at least 10 sampled individuals (range: 10-95). However, there were only 4 samples from Davis Strait and 3 samples from Western Hudson Bay, and no samples from either Norwegian Bay or Kane Basin. While this represents a sampling limitation, we supplement our GT-seq data with ddRADseq data (Jensen et al., 2020) for our final assessment of Canadian polar bear structure and diversity, and note that census population size estimates for the two subpopulations for which we have no samples are small (KB = 357 individuals, NW = 203 individuals; Hamilton & Derocher, 2019).

### DNA extraction

Whole genomic DNA was extracted from fecal samples using the QIAamp Fast DNA Stool Mini Kit (Qiagen) according to manufacturer’s protocols. For BP, HV, and MS samples, genomic DNA was extracted using a modified salt extraction protocol (Aljanabi & Martinez, 1997), with an RNaseA (ThermoFisher Scientific) step included. Once extracted, DNA extracts from all tissue samples were run on 1.5% agarose gel stained with RedSafe<sup>TM</sup> Nucleic Acid Staining Solution (iNtRON Biotechnology) to assess quality and quantified using a Nanodrop ND\_1000 spectrophotometer (Nanodrop Technologies Inc.). As feces contains DNA from multiple sources, we used a polar bear-specific qPCR assay targeting the *F2* gene to quantify polar bear DNA from both CF and FF samples (Hayward et al., 2020). To gauge the value of running fecal samples through this qPCR assay as a screening tool, we devised a small double-blind experiment in which we randomly divided CF and FF samples into two subsets: 1. 55 samples for which we assayed DNA quantity before GT-seq library construction and sequencing; and 2. 116 samples for which we assayed DNA quantity only after sequencing had already been performed (See *Supplemental Information: qPCR Experiment* for qPCR

experiment methods and results).

### GT-seq genotyping and genotype calling methods

A full description of GT-seq panel optimization can be found in *Supplemental Information: Panel Optimization*. Using optimized GT-seq conditions and primers for our final GT-seq panel, we prepared two libraries for fecal samples (CF, FF) and another three for tissue samples (HV, BP, MS). The fecal libraries contained all 150 fecal samples, as well as the 21 FF duplicates taken to assess subsample variation in genotyping error. We included 284 tissue samples in the latter libraries and also included two technical replicates (BP) to assess genotyping error for GT-seq. Genotyping error was measured as percent discordance between technical replicates. Within our total of 457 samples, we included 65 paired sets of MS and CF collected from the same individuals. This allowed us to compare GT-seq genotyping error across sample types and obtain a “best case scenario” estimate of genotyping error for FF samples. Here, we assume that the muscle genotype is “correct” and calculate the genotyping error (percent discordance) for CF as a proxy for FF samples. Prior to calculating genotyping error within and among sample types, samples with >50% missing data were removed from the GT-seq dataset, as we considered samples with >50% data (i.e. genotypes for at least half of panel loci) to have been successfully sequenced. To assess the validity of the sex identities provided by our two GT-seq loci, we compared the GT-seq determined polar bear sex to hunter-provided sex identification for 293 samples that were successfully genotyped at >50% loci and for which we had field data.

Library preparation followed the original protocols of Campbell et al. (2015), modified based on our pilot tests (See *Supplemental Information: Panel Optimization*), and using only primers for a final optimized panel of 327 SNPs. Libraries were sequenced using an Illumina MiSeq at Queen’s University. We used the GT-seq pipeline available on GitHub (<https://github.com/GTseq/GTseq-Pipeline>) for filtering to a minimum depth of 10 and genotype calling, as suggested by Campbell et al. (2015). However, we were also interested in examining genotype discordance between different SNP calling models, as some discrepancy has been found between the GT-seq pipeline and other workflows used to call RADseq data (Schmidt et al., 2020). As BCFTOOLS (Li, 2011) is a common genotyping tool and was used to call genotypes for our original ddRADseq dataset (Jensen et al., 2020), we used this same workflow to process our raw GT-seq sequencing reads. Briefly, reads from GT-seq were aligned to the polar bear reference genome (assembly version UrsMar.-1.0, PMID: 24813606) using the BWA-MEM v0.7.17 aligner (Li & Durbin, 2009). Alignments were sorted, indexed, and read pairs were fixed using tools from the SAMtools v1.9 suite (Li et al., 2009). A target file of the GT-seq assay SNP positions was used with BCFtools mpileup to call and produce a VCF file of the targeted sites. The VCF was filtered in VCFtools (Danecek et al., 2011) for a minimum depth cut off of either 6 (dataset BCF-6) or 10 (dataset BCF-10). Output files for each workflow (GT-seq, BCF-6, BCF-10) were converted to genpop format to compare percent missing data by locus. Average mismatch rates for all 457 samples were also calculated between the GT-seq pipeline and the BCFTOOLS workflows. The dataset resulting from the calling method with the least missing data (BCF-6) was used to calculate genotyping error, percent missing data, and for all subsequent population genetic analyses.

### Population genetic analyses

To evaluate the usefulness and power of GT-seq for population genetic analysis and further extend our understanding of polar bear population structure, we estimated several metrics of genetic diversity and structure using a combined dataset called and filtered with our BCF-6 workflow. This dataset comprised individuals genotyped at 322 autosomal SNPs using GT-seq, as well as individuals previously genotyped at our 322 GT-seq loci using ddRADseq (Jensen et al., 2020). After combining GT-seq and ddRADseq data, we removed all replicates (similarity >0.8, known duplicate subsamples, tissue from tissue-colon sets), and all individuals with >50% missing data. Our final GT-seq+ddRADseq dataset contained a total of 642 individuals genotyped at 322 autosomal loci and 2 sex-linked loci.

For the combined GT-seq+ddRADseq dataset, we estimated observed and expected heterozygosity ( $H_o/H_E$ ) and inbreeding coefficients ( $G_{is}$ ) for each subpopulation, as implemented in GENODIVE v2.0b27 (Meirmans & Van Tienderen, 2004). Using the *related* package in R (Pew, Muir, Wang, & Frasier, 2015), we used

our empirical allele frequencies to simulate pairs of individuals with known relatedness. One hundred pairs were simulated for each of the following categories: unrelated individuals, half siblings, full siblings, and parent-offspring pairs. We used this simulated dataset to test the ability of multiple relatedness estimators to distinguish among relatedness categories. Based on the results, we calculated pairwise relatedness between all 642 individuals in our combined dataset.

To assess population substructure, we used Bayesian clustering analysis as implemented in STRUCTURE 2.3.4 (Pritchard, Stephens, & Donnelly, 2000). We evaluated the number of clusters in the data (K) from 1 to 10, with 20 iterations of each, and with a run length of 300,000 MCMC following a burn-in period of 100,000 MCMC. The most likely value of K was identified using the DeltaK method of Evanno, Regnaut, & Goudet (2005) in Structure Harvester (Earl & vonHoldt, 2012). To complement the STRUCTURE analysis, we also used discriminant analysis of principal components (DAPC; Jombart, Devillard, & Balloux, 2010), implemented in *adegenet* in R (Jombart, & Ahmed, 2011) to evaluate the number of genetic clusters in the data. Principal component analysis was also performed using the R package *ade4* v. 1.7-16 (Dray & Dufour, 2007) with *a priori* groups assigned based on sample type (MS, HV, BP, CF, FF) and data type (GT-seq, ddRADseq) to confirm structuring was independent of sample and data types (Figure S5). The best-fit value of K was tested using the `find.clusters` function and Bayesian information criterion (BIC), and the chosen value of K was selected based on the lowest BIC value.

To estimate assignment accuracy to each subpopulation and STRUCTURE-derived genetic cluster, we used principal components analyses and Monte Carlo cross-validation procedures implemented in the *AssignPOP* package in R (Chen et al., 2018). For the subpopulation version of analysis, we removed subpopulations with <10 samples (Norwegian Bay), whereas for the genetic cluster version, we only retained individuals with >0.7 membership to a single genetic cluster. From these data we built a predictive model using a support vector machine (model `svm`) classification based on training sets composed of the most informative 75% loci and a random sample of 75% of individuals in the dataset. The remaining 25% of individuals were used to test the rate of assignment, which was then averaged across 30 iterations.

## RESULTS

### Calling workflow comparisons

Our final GT-seq dataset, called using the published GT-seq pipeline (Campbell et al., 2015), included 325 autosomal SNPs and 2 sex-linked markers. An additional 3 loci were removed from the BCFTOOLS workflow datasets after filtering for minimum depth (depth 6, 10), leaving each with 322 autosomal loci and 2 sex-linked markers. We removed the same 3 loci from our GT-seq pipeline dataset to enable direct comparison of genotypes and missing data by locus across the calling methods. Based on all 457 samples, there was an average of 25.4% missing data for the GT-seq calling pipeline, whereas missing data were 23.9% and 21.3% for BCF-10 and BCF-6 calling workflows, respectively. Regardless of BCFTOOLS calling workflow, genotype mismatch with the GT-seq workflow was 1.1% on average. Based on these results and the potential for easy comparison with existing ddRADseq data, we chose to use the dataset generated from the BCF-6 calling workflow to assess genotyping error and analyze population structure.

### Genotyping Success

After removing individuals with >50% missing data, a total of 365 samples, including 3 duplicate samples, remained in our GT-seq dataset. Average missing data dropped to 3.2% from 21.3% when these individuals were removed. Percent individuals successfully genotyped ranged from 30.6% to 97.4% depending on sample type (Table 1), recalling that most of the FF samples were not screened before sequencing to allow for testing of the qPCR screening tool (Hayward et al., 2020). When considering only the FF samples that would have passed screening (>0 ng/ $\mu$ L polar bear DNA), percent individuals successfully genotyped increases to 85.7%. See *Supplemental Information: qPCR Experiment* for detailed results.

For retained individuals with >50% data, average missing data per locus was similar across sample types, with FF samples missing the most at 14.9% of loci (Table 1). Both BP replicates were successfully genotyped

at >50% loci and gave a mean genotyping error rate of 0.17%. Of the 21 FF samples for which duplicate subsamples were included in the GT-seq runs without qPCR pre-screening, one or both subsamples failed to amplify in 20 of 21 cases. In most cases (15/21), both subsamples failed to amplify, although there were a few cases where we witnessed large differences in genotyping success between FF duplicates. For example, one FF subsample was genotyped at only 8 loci, whereas its corresponding subsample was genotyped at all 322 loci. For the one FF sample for which both subsamples successfully genotyped, the genotyping error was 10.7%. 56 of the 65 muscle-colon feces sets had both sample types from the same individual successfully genotyped, giving a mean genotyping error estimate of 6.8% for fecal samples. Of the 322 retained autosomal loci, 2 had >50% missing data across individuals. Removal of these loci in future GT-seq runs may increase genotyping success, particularly for field-collected feces.

We had a total of 293 bear samples for which hunters provided a sex identity. Of these, 78.6% were genotyped successfully at one or both sex-linked loci. There were 18 instances where a sample had only a genotype at one of the sex-linked loci, and only one case of allelic dropout. For samples with genotypes at both sex loci, genotype concordance was 99.7%. Hunter identification of bear sex matched the sex genotype provided by the first sex locus for 95.9% of samples, matched the second sex locus for 91.1% of samples, and matched both sex loci for 90.1% of samples.

**Table 1.** GT-seq genotyping success for five different sources of polar bear DNA, including 365 total samples collected across 10 subpopulations. Shown are percent individuals successfully genotyped before and after samples with no detectable polar bear DNA were removed (based on qPCR results), and average percent missing data by locus for each sample type (individuals with >50% missing data removed).

Sample type	n	% individuals >50% data	% individuals >50% data	% individuals >50% data
Set muscle (MS)	101	96.2	-	-
Colon feces (CF)	69	88.5	90.1	90.1
Biopsy (BP)	134	95.7	-	-
Harvest muscle (HV)	38	97.4	-	-
Field feces (FF)	23	30.6	85.7	85.7

*Set muscle* : tissue from corresponding muscle and colon sets; *Colon feces* : feces removed from the colon of corresponding muscle and colon sets; *Biopsy* : biopsy tissue sample; *Harvest muscle* : tissue from annual polar bear harvest; *Field feces* : non-invasively collected scat from the field

#### Population structure analyses

Based on the final combined dataset of 642 individuals (GT-seq+ddRADseq data at 324 GT-seq loci), all subpopulations display similar levels of genetic diversity (Table 2). Expected heterozygosity values ranged from 0.42 to 0.45 and  $G_{IS}$  values were mostly negative, ranging from -0.006 to 0.004.

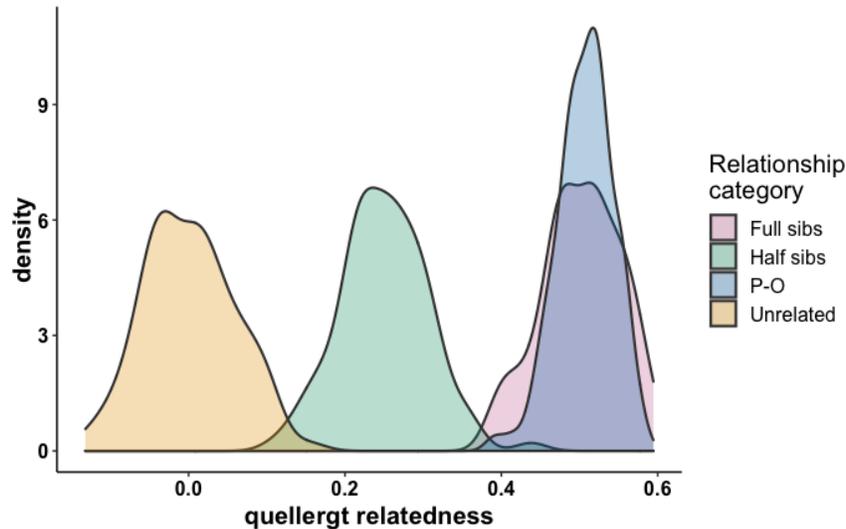
**Table 2.** Diversity metrics for 11 Canadian polar bear subpopulations (see Figure S1) based on a combined GT-seq+ddRADseq dataset consisting of 642 individuals genotyped at 322 autosomal loci.

Subpopulation	n	$H_o$	$H_E$	$G_{IS}$	Self-assignment	Main genetic cluster
Baffin Bay (BB)	50	0.45	0.45	-0.006	0	Arctic Archipelago
Davis Strait (DS)	40	0.45	0.45	0.004	0.01	Arctic Archipelago
Foxe Basin (FB)	82	0.45	0.44	-0.025	0.52	Hudson Complex
Gulf of Boothia (GB)	110	0.45	0.45	0.003	0.73	Arctic Archipelago
Lancaster Sound (LS)	83	0.45	0.45	-0.009	0.03	Arctic Archipelago
M'Clintock Channel (MC)	80	0.45	0.44	-0.023	0.56	M'Clintock Channel
Northern Beaufort Sea (NB)	55	0.44	0.43	-0.019	0.04	Polar Basin
Norwegian Bay (NW)	1	-	-	-	-	Arctic Archipelago

Southern Beaufort Sea (SB)	12	0.45	0.42	-0.070	0	Polar Basin
Southern Hudson Bay (SH)	80	0.44	0.43	-0.017	0.65	Hudson Complex
Viscount Melville Sound (VM)	11	0.45	0.44	-0.018	0	M'Clintock Channel /Polar Basin
Western Hudson Bay (WH)	38	0.44	0.43	-0.017	0	Hudson Complex

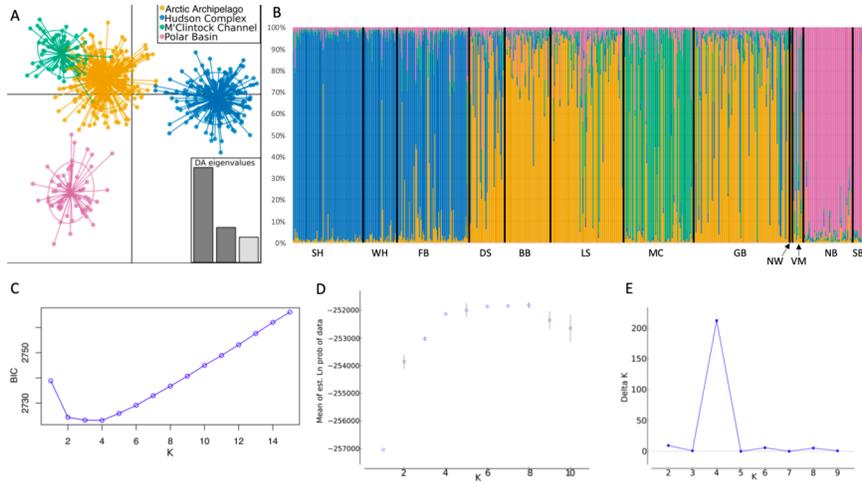
$n$ , sample size;  $H_o$ , observed heterozygosity;  $H_E$ , expected heterozygosity;  $G_{IS}$ , inbreeding coefficient

We discovered 13 true recaptures in our combined dataset (relatedness  $>0.80$ ), which were removed from all subsequent analysis. Recaptures spanned sample types, with pairs existing among or within sample types: FF ( $n = 8$ ), CF ( $n = 4$ ), MS ( $n = 5$ ), BP ( $n = 4$ ), and HV ( $n = 4$ ). For example, one bear identified from a BP sample (non-lethal) was later recaptured with a CF sample (lethal). Based on comparison of multiple relatedness estimators, we determined that the *quellertg* correlation coefficient was best for our combined dataset (correlation between observed and expected values = 0.97; Queller & Goodnight, 1989). Results from simulations with the *quellertg* coefficient yielded pairwise relatedness density plots with marked separation for full siblings, half siblings and parent-offspring pairs, and unrelated individuals (Figure 1). Although relatedness among individuals in our dataset was mostly unknown prior to analysis, we can see multiple cases where samples from the same area show relatedness values close to 0.5, as would be expected with parent- offspring or sibling pairs travelling together.

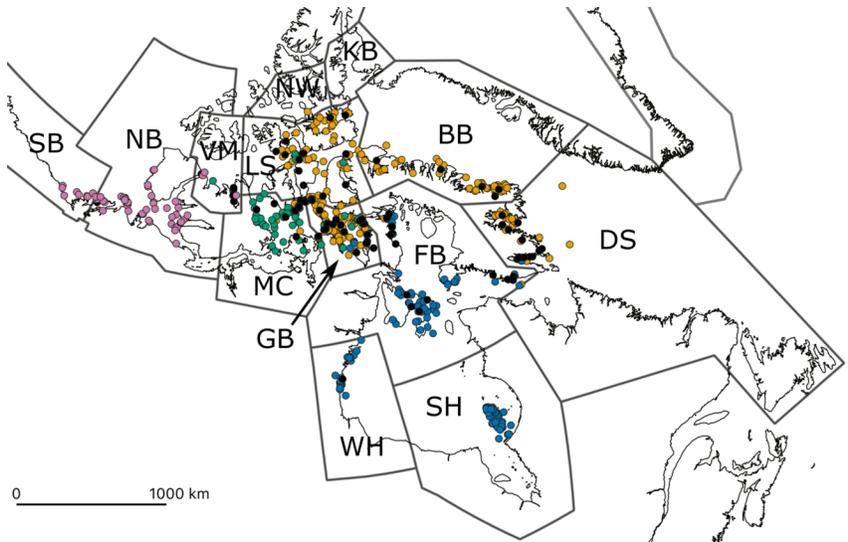


**Figure 1.** Density plot of pairwise *quellertg* relatedness values generated from simulations (100 per relatedness category) based on allele frequencies from our combined GT-seq+ddRADseq dataset of 642 individuals genotyped at 322 autosomal loci. Each colour represents a different relatedness category (Full sibs = full siblings, Half sibs = half siblings, P-O = parent-offspring, Unrelated = unrelated pair).

Our STRUCTURE and DAPC analyses revealed similar patterns in population genetic structure. STRUCTURE analysis suggested an optimal value of  $K = 4$ , with  $\ln \Pr(X | K)$  plateauing around  $K = 4$  (Figure 2D) and  $\Delta K$  greatest at  $K = 4$  (Figure 2E). For the find.clusters DAPC analysis, the lowest BIC scores occurred at  $K = 3$  and  $K = 4$  (Figure 2C). For both these analyses, the genetic clusters at  $K = 4$  correspond to three geographic regions typically referred to as the Hudson Complex, the Arctic Archipelago, and the Polar Basin, with an additional cluster corresponding to the subpopulation of M'Clintock Channel (Figure 2A, B, Figure 3). Barplots showing  $K = 3$  and  $K = 5$  are provided in Figure S6.



**Figure 2.** Results of two clustering analyses performed with combined GT-seq+ddRADseq dataset of 642 bears genotyped at 322 autosomal loci. A) find.clusters Discriminant Analysis of Principal Components (DAPC) displaying inferred clustering at  $K = 4$ . Each point represents an individual bear and inertial ellipses surround each genetic cluster. B) STRUCTURE barplot showing inferred clustering at  $K = 4$ . Each colour corresponds to a distinct genetic cluster and each vertical bar represents an individual and their proportional membership in each cluster. C) Plot of Bayesian Information Criterion (BIC) from DAPC analysis for each number of clusters evaluated. D) Plot of  $\ln P(K)$  for each number of clusters evaluated in the STRUCTURE analysis. E) Plot of  $\Delta K$  for each number of clusters evaluated in the STRUCTURE analysis.



**Figure 3.** Map of the Canadian Arctic with samples used for analysis. Outlined regions are existing subpopulations that are fully or partially in Canada. Points correspond to sampled individuals and are coloured according to genetic cluster assignment based on STRUCTURE analysis (pink = Polar Basin, green = M'Clintock Channel, orange = Arctic Archipelago, Blue = Hudson Complex). Black points represent individuals with membership  $< 0.7$  to a genetic cluster. Subpopulation abbreviations are the same as in Table 2.

Despite 90 individuals having values of less than 0.7 membership to a particular genetic cluster (black dots in Figure 3), most of the 642 individuals in our combined ddRADseq and GT-seq dataset were assigned to a single cluster. One exception was Viscount Melville (VM), which comprised individuals with high assignment probabilities to both the Polar Basin and M’Clintock Channel clusters. The self-assignment test performed by subpopulation (Table 2) suggests that few subpopulations are highly genetically distinguishable, as multiple have self-assignment rates low or near zero. The Gulf of Boothia (GB) and Southern Hudson Bay (SH) subpopulations displayed the highest self-assignment rates at 0.73 and 0.65, respectively. Our second self-assignment test shows that the genetic clusters suggested by our STRUCTURE analysis are typically more highly distinguishable than the subpopulations, with self-assignment rates  $>0.80$  for three of the clusters (Table 3). Self-assignment was lowest for the Polar Basin cluster (0.35).

**Table 3.** Results for a self-assignment test performed in *AssignPOP* in R for the four genetic clusters suggested by our STRUCTURE analysis. Individuals with  $<0.7$  membership to one cluster were removed from our combined GT-seq+ddRADseq dataset of 642 bears genotyped at 322 autosomal loci.

Genetic cluster	n	Self-assignment
Arctic Archipelago	230	0.98
Polar Basin	66	0.35
Hudson Complex	188	1.00
M’Clintock Channel	68	0.83

$n$ , sample size

## DISCUSSION

Non-invasive samples (e.g. feces, hair snags) are viable sources of DNA for genetic monitoring and may help mitigate some of the limitations of traditional monitoring methods, while also providing opportunities to integrate Traditional Ecological Knowledge and western science perspectives. Here, we developed a final GT-seq assay of 324 SNPs optimized for genotyping of polar bears based on DNA from non-invasively collected fecal samples. We demonstrated 1) successful GT-seq genotyping of DNA from field-collected polar bear feces, and 2) the practicality of GT-seq for distinguishing individuals, assessing relatedness, and expanding our understanding of genetic population structure and diversity for Canadian polar bear subpopulations. Here, we determined that all existing unrelated bears could be distinguished with as few as 34 SNPs (Figure S2). Given the current global polar bear population estimate of 23,315 (Hamilton & Derocher, 2019), our panel of 322 autosomal and 2 sex-linked SNPs is more than sufficient to reliably discern bears among relationship categories (e.g. siblings, unrelated) with high certainty. This is supported by the marked separation among density plots for full siblings, half siblings, parent-offspring pairs or unrelated individuals (Figure 1). Given this ability to differentiate individuals based on both tissue and fecal samples, our GT-seq assay may be particularly useful in genetic mark-recapture studies making use of non-invasive scat surveys.

We have also shown that our GT-seq panel provides sufficient statistical power to assess polar bear population structure and diversity. Diversity metrics based on our GT-seq dataset resemble those of previous studies based on SNPs, mtDNA, or microsatellites (e.g. Jensen et al., 2020; Malenfant, Davis, Cullingham, & Coltman, 2016), although our estimates of population diversity may have been slightly higher than expected due to our selection of high diversity SNPs during panel selection. Levels of genetic diversity did not vary markedly among subpopulations, corroborating previous suggestions that these existing subpopulations do not reflect the subtle patterns of polar bear dispersal and genetic structure in the Canadian Arctic (Jensen et al., 2020; Malenfant et al., 2016). Although many subpopulations have only subtle differences in genetic diversity, high assignment rates for a few of the subpopulations (e.g. GB, SH, MC) may have implications for conservation and management planning. In our original selection of SNPs for our GT-seq panel, we focused on selecting high diversity SNPs to maximize resolution of individuals and population genetic structure. However, other characteristics of a species may also influence the number of markers required to resolve differences among individuals and subpopulations, including its mating system and linkage patterns within

its genome. For example, mating systems and small populations with high relatedness may require a larger panel of loci for individual identification. Final panel selection also depends on intended application. Here, we aimed to examine baseline polar bear population structure and thus, we selected putatively neutral SNPs with high diversity by filtering for minor allele frequency and departure from Hardy-Weinberg Equilibrium. In the future, we may wish to develop a panel to assess adaptation in real time by including SNPs that are flagged as potentially being under selection, or perhaps a panel to distinguish between bear species (e.g. grizzly and polar bears) through selection of species diagnostic SNPs, similar to application of GT-seq to invasive brown (*Rattus norvegicus*) and black rats (*R. rattus*; Sjodin, Irvine, & Russello, 2020). Thus, the GT-seq assay can be modified and widely adapted to other study systems to address a range of genetic questions.

GT-seq uses a reduced panel of SNPs that allows genotyping of samples with degraded DNA. It thus enables us to combine GT-seq data with data from other methods (e.g. ddRADseq), as long as there are loci in common. However, differences in genotyping calling methods between GT-seq and RADseq datasets may contribute to genotype discordance among datasets (Schmidt et al., 2020). To address this possibility, we called genotypes from our GT-seq data using both the original calling pipeline for GT-seq (Campbell et al., 2015) and a BCFTOOLS workflow previously used to call genotypes from ddRADseq data (Jensen et al., 2020). Genotypic discordance between the two calling workflows was low (1.1%) but still suggested some variation between calling methods for GT-seq and ddRADseq data. In addition, levels of missing data were moderate for all workflows (21.3-25.4%). Based on these results, we acknowledge that workflow choice is primarily a matter of preference with discordance among workflows being negligible. Regardless, we recommend being consistent and using the same workflow while working with both GT-seq and ddRADseq data.

Genotyping success estimated using the BCF-6 workflow was high across all sample types, with FF samples in particular being genotyped successfully (>50% data) for 85.7% of individuals with qPCR-detectable polar bear DNA, with an average of 14.9% missing data. We estimated genotyping error for field scat to be around 10% based on MS:CF comparisons and FF sample replicates not screened before sequencing, and much lower for other sample types (e.g. 0.17% for biopsy). Genotyping success using GT-seq was expected to be relatively low for field feces compared to tissue samples, as there is an unknown length of environmental exposure and each sample contains a unique mix of DNA from pathogens, gut microflora and microfauna, diet species, and the host itself. High genotyping error rates may introduce significant bias into estimates of population parameters, and greater uncertainty into individual identification and relatedness estimates (Bonin et al., 2004; Gagneux, Boesch, & Woodruff, 1997; Kleinman-Ruiz et al., 2017; von Thaden et al., 2017). To address this limitation and minimize the use of resources on fecal samples that are unlikely to have sufficient polar bear DNA, we evaluated a qPCR screening assay (Hayward et al., 2020). We found that the qPCR screening assay substantially increased the percentage of field fecal samples successfully genotyped by GT-seq (30.6 to 85.7%). In one case, duplicate subsamples retrieved from the same fecal sample demonstrated a large difference in genotyping success (1 failure, 1 success), suggesting that samples can be heterogenous in amount of host DNA and samples for which genotyping is critical may be repeatedly subsampled to achieve success. Thus, we recommend using a similar sample quality screening step for any future research making use of GT-seq and non-invasive samples.

By combining our new GT-seq data with our ddRADseq data (Jensen et al., 2020), we were able to expand our sample sizes for certain subpopulations (e.g. GB, MC). This allowed us to reassess Canadian polar bear population structure using a larger dataset (642 vs. 358 individuals from Jensen et al., 2020) and make use of samples that were unusable using ddRADseq methodology (i.e. fecal, degraded biopsy and muscle samples), which requires high quality DNA. These combined data, including genotypes from non-invasively collected feces, allowed us to resolve clear genetic differentiation within the Canadian Arctic. Sample clustering patterns and geographic distributions were largely consistent with genetic groups called the “Hudson Complex,” “Arctic Archipelago,” and “Polar Basin” described previously both in studies using microsatellites and SNPs (Jensen et al., 2020; Malenfant, Coltman, & Davis, 2015; Malenfant et al., 2016; Paetkau et al., 1999). Boundaries between our clusters were not discrete, with the existence of admixed

individuals ( $n = 90$ ) and individuals that belonged to one genetic cluster being found within the geographic range of another. Due to our small sample size for Norwegian Bay subpopulation ( $n = 1$ ), we were unable to confirm its status as its own genetic cluster, as has been proposed (Malenfant et al., 2015; Malenfant et al., 2016; Paetkau et al., 1999). However, a new genetic cluster coincident with the M’Clintock Channel subpopulation emerged in both our STRUCTURE and DAPC analyses (Figures 2 & 3), a pattern not evident from analysis of ddRADseq data alone (Jensen et al., 2020). To rule out technical reasons for this new cluster, we confirmed that our data do not cluster based on sample type (e.g. FF vs. MS vs. BP) or data type (e.g. GT-seq vs. ddRADseq; Figure S5). Rather than M’Clintock Channel existing as a clear genetic cluster comparable to “Hudson Complex,” “Arctic Archipelago,” and “Polar Basin”, it seems more likely that it is the result of more subtle genetic differentiation within the Arctic Archipelago or possibly a by-product of the timespan over which our samples were collected.

Based on mitochondrial DNA and microsatellite data, the Gulf of Boothia and M’Clintock Channel subpopulations have been recommended as significantly differentiated conservation units with both historical and contemporary barriers to gene flow (Campagna et al., 2013). Similarly, analyses of SNP data have suggested 6 genetic clusters, including Western and Eastern clusters within both the Arctic Archipelago and Polar Basin major clusters (Malenfant et al., 2016). Preprint SNP data (Malenfant et al., in press) also imply that the Arctic Archipelago cluster may comprise 3 genetically differentiated groups: a Western Archipelago cluster that includes Viscount Melville Sound and M’Clintock Channel, a Gulf of Boothia cluster, and an Eastern Archipelago cluster including Lancaster Sound, Kane Basin, and Baffin Bay. Thus, it may be that our M’Clintock Channel subcluster is part of a larger Western Arctic Archipelago cluster. Our small sample size for Viscount Melville Sound may have precluded clear resolution of a Western Archipelago cluster, as there exists an admixture zone between a M’Clintock Channel-Viscount Melville Sound subcluster and the Polar Basin (Malenfant et al., in press).

Our dataset contains samples collected between 1998 and 2018 (a 20-year timespan) which presents a potential temporal confound that is not uncommon to studies investigating polar bear population structure (e.g. Malenfant et al., 2016; Paetkau et al., 1999; Paetkau, Calvert, Stirling, & Strobeck, 1995). Considering that the mean sample collection date for M’Clintock Channel was 1999 and the rapidness of environmental changes in the Canadian Arctic, it is possible that these data do not represent the contemporary genetic structuring of the Canadian polar bear population. Regardless, we were able to discern individual bears and to resolve genetic patterns comparable to other key studies of polar bears (e.g. Malenfant et al., 2016; Paetkau et al., 1995, 1999), implying that our panel does adequately capture these known patterns. Importantly, we are able to do so with much greater cost-efficiency and based on genetic data collected from non-invasive and degraded samples. Thus, GT-seq presents a real opportunity to efficiently and iteratively assess the genetic structure of Canada’s polar bear populations – and a flexible and cost-effective way to update our understanding of subpopulation structure as continued sampling is performed.

We originally intended to include at least 30 individuals from each Canadian subpopulation in our ddRADseq dataset (Jensen et al., 2020) used to develop our GT-seq SNP panel, but some subpopulations had fewer than 30 samples of suitable quality. Ascertainment bias may thus be of concern as the GT-seq panel did not represent the full array of subpopulations, nor did it include data from other Arctic nations. Thus, application to the global polar bear population may require some panel modifications. Establishing baseline global polar bear population structure will be especially important for examining trends in dispersal patterns, diversity and structure, and behaviour in the context of a rapidly changing climate. High-quality genotyping of non-invasive samples enabled by GT-seq will facilitate monitoring of admixture, dispersal, and temporal trends in structure.

The need for monitoring using genetic methods that can yield results quickly in the midst of quick environmental change, as well as the need to integrate Indigenous ways of knowing and western science, is not unique to polar bear monitoring. Here, we respond to a long-expressed desire by northern communities for a means to non-invasively monitor polar bears. We demonstrate that we can use GT-seq to differentiate individuals with a high degree of certainty and comprehensively discern population structure at levels comparable to

other methods of genetic monitoring (e.g. Malenfant et al., 2016; Paetkau et al., 1995, 1999), but with much greater efficiency (Campbell et al., 2015) and based on the inclusion of non-invasive and degraded samples. The GT-seq assay that we developed here can provide a foundation for new community-based programs that aim to improve temporal monitoring of the global polar bear population and directly inform conservation efforts and government policy. Importantly, envisioned community-based monitoring will incorporate Traditional Ecological Knowledge and the perspectives of Indigenous communities throughout the planning and monitoring processes, while also providing both social and economic benefits to northern communities. This GT-seq protocol can also serve as the foundation for a comprehensive toolkit to assess important aspects of polar bear and ecosystem health (e.g. contaminants, parasite load, diet). With GT-seq at the core of the toolkit, a suite of data would be provided to communities and territorial governments that can be incorporated into a community-driven framework that will ultimately have societal and management implications. As this framework and our GT-seq protocol can be adapted to other species and for other research questions, the monitoring methodology we propose here may be adapted and applied as a model for inclusive wildlife monitoring worldwide.

### *Acknowledgements*

This study would not be possible without the support and collaboration of the Inuvialuit Game Council, the Gjoa Haven and Coral Harbour Hunters and Trappers Association, the Canadian Rangers, and all the Northwest Territories and Nunavut communities that contributed to the project. N. Campbell, L. Waits, and R. Malenfant provided valuable advice on study design and interpretation. Thank you to everyone who helped with lab work and have provided unwavering support for the project: M. Navarrete Bedolla, D. Lougheed, S. Vanderluit, H. Wainwright, E. Landon, S. Edie, S. Maracle, A. Siew, T. Frappier-Brinton, M. Harwood, and the rest of the BEARWATCH team and Lougheed Lab. This work was funded by the Government of Canada through Genome Canada and the Ontario Genomics Institute (OGI-123), and the Natural Sciences and Engineering Research Council of Canada (NSERC). Computational resources were provided by Compute Canada through the Resources for Research Groups program. K. Hayward's Master's thesis is funded by an NSERC Alexander Graham Bell Canada Graduate Scholarship.

### REFERENCES

- Aljanabi, S.M., & Martinez, I. (1997). Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research* , **25** , 4692-4693.
- Andrews, K.R., Adams, J.R., Cassirer, E.F., Plowright, R.K., Gardner, C., Dwire, M., ... Waits, L.P. (2018). A bioinformatic pipeline for identifying informative SNP panels for parentage assignment from RADseq data. *Molecular Ecology Resources* , **18** , 1263-1281.
- Aziz, M.A., Tollington, S., Barlow, A., Greenwood, C., Goodrich, J.M., Smith, O., ... Groombridge, J.J. (2017). Using non-invasively collected genetic data to estimate density and population size of tigers in the Bangladesh Sundarbans. *Global Ecology and Conservation* , **12** , 272-282.
- Bergner, L.M., Orton, R.J., da Silva Filipe, A., Shaw, A.E., Becker, D.J., Tello, C., ... Streicker, D.G. (2019). Using noninvasive metagenomics to characterize viral communities from wildlife. *Molecular Ecology Resources* , **19** , 128-143.
- Blåhed, I.-M., Königsson, H., Ericsson, G., & Spong, G. (2018). Discovery of SNPs for

- individual identification by reduced representation sequencing of moose (*Alces alces*). *PLoS ONE* , **13** , e0197364.
- Bonin, A., Bellemain, E., Bronken Eidesen, P., Pompanon, F., Brochmann, C., & Taberlet, P. (2004). How to track and assess genotyping errors in population genetics studies. *Molecular Ecology* , **13** , 3261–3273.
- Bourgeois, S., Kaden, J., Senn, H., Bunnefeld, N., Jeffery, K.J., Akomo-Okoue, E.F., ... McEwing, R. (2019). Improving cost-efficiency of faecal genotyping: New tools for elephant species. *PLoS ONE* , **14** , e0210811.
- Campagna, L., Van Coeverden de Groot, P.J., Saunders, B.L., Atkinson, S.N., Weber, D.S., Dyck, M.G., ... Lougheed, S.C. (2013). Extensive sampling of polar bears (*Ursus maritimus*) in the Northwest Passage (Canadian Arctic Archipelago) reveals population differentiation across multiple spatial and temporal scales. *Ecology and Evolution* , **9** , 3152-65.
- Campbell, N. R., Harmon, S. A., & Narum, S. R. (2015). Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. *Molecular Ecology Resources* , **15** , 855–867.
- Carroll, E.L., Bruford, M.W., DeWoody, J.A., Leroy, G., Strand, A., Waits, L., & Wang, J. (2018). Genetic and genomic monitoring with minimally invasive sampling methods. *Evolutionary Applications* , **11** , 1094-1119.
- Chen, K.-Y., Marschall, E.A., Sovic, M.G., Fries, A.C., Gibbs, H.L., & Ludsin, S.A. (2018). assignPOP: An R package for population assignment using genetic, non-genetic, or integrated data in a machine-learning framework. *Methods in Ecology and Evolution* , **9** , 439-446.
- Cristescu, R.H., Miller, R.L., Schultz, A.J., Hulse, L., Jaccoud, D., Johnston, S., ... Frere, C.H. (2019). Developing noninvasive methodologies to assess koala population health through detecting *Chlamydia* from scats. *Molecular Ecology Resources* , **19** , 957-969. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., ... 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics* , **27** , 2156-2158.
- Dray, S., & Dufour, A.-B. (2007). The ade4 package: Implementing the Duality Diagram for ecologists. *Journal of Statistical Software* , **22** , 1-20.
- Durner, G.M., Laidre, K.L., & York, G.S. (2018). Polar bears: Proceedings of the 18th working meeting of the IUCN/SSC Polar Bear Specialist Group, Anchorage, Alaska, 7-11 June 2016. Occasional Paper No. 63. of the IUCN Species Survival Commission.

Retrieved from <https://portals.iucn.org/library/node/47667>

Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* , **4** , 359– 361.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology* , **14** , 2611-2620.

Fitak, R.R., Naidu, A., Thompson, R.W., & Culver, M. (2016). A new panel of SNP Markers for the individual identification of North American Pumas. *Journal of Fish and Wildlife Management* , **7** , 13–27.

Fonturbel, F.E., Lara, A., Lobos, D., & Little, C. (2018). The cascade impacts of climate change could threaten key ecological interactions. *Ecosphere* ,**9** , e02485.

Gagneux, P., Boesch, C., & Woodruff, D.S. (1997). Microsatellite scoring errors associated with noninvasive genotyping based on nuclear DNA amplified from shed hair. *Molecular Ecology* , **6** , 861–868.

Garshelis, D.L, & Noyce, K.V. (2006). Discerning biases in large scale mark-recapture population estimate for black bears. *The Journal of Wildlife Management* , **70** , 1634-1643.

Giangregorio, P., Norman, A.J., Davoli, F., & Spong, G. (2019). Testing a new SNP-chip on the Alpine and Apennine brown bear (*Ursus arctos* ) populations using non-invasive samples. *Conservation Genetics Resources* , **11** , 355-363.

Government of Canada. (2018). Maps of subpopulations of polar bears and protected areas. Retrieved from <https://www.canada.ca/en/environment-climate-change/services/biodiversity/maps-sub-populations-polar-bears-protected.html>

Graham, C.F., Glenn, T.C., McArthur, A.G., Boreham, D.R., Kieran, T., Lance, S., ... Somers, C.M. (2015). Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). *Molecular Ecology Resources* , **15** , 1304–1315.

Hamilton, S.G., & Derocher, A.E. (2019). Assessment of global polar bear abundance and vulnerability. *Animal Conservation*, **22** , 83–95.

Han, S., Guan, Y., Dou, H., Yang, H., Yao, M., Ge, J., & Feng, L. (2019). Comparison of the fecal microbiota of two free-ranging Chinese subspecies of the leopard (*Panthera pardus* ) using high-throughput sequencing. *PeerJ* , **7** , e6684.

Hayward, G.D., Miquelle, D.G., Smirnov, E.N., & Nations, C. (2002). Monitoring Amur Tiger populations: Characteristics of track surveys in snow. *Wildlife Society Bulletin (1973-2006)* , **30** , 1150-1159.

Hayward, K.M., Harwood, M.P., Lougheed, S.C., Sun, Z., Van Coeverden de Groot, P., &

- Jensen, E.L. (2020). A real-time PCR assay to accurately quantify polar bear DNA in fecal extracts. *PeerJ* , **8** , e8884.
- Hess, J.E., Campbell, N.R., Docker, M.F., Baker, C., Jackson, A., Lampman, R., ... Narum, S.R. (2015). Use of genotyping by sequencing data to develop a high-throughput and multifunctional SNP panel for conservation applications in Pacific lamprey. *Molecular Ecology Resources* , **15** , 187–202.
- Hunter, C.M., Caswell, H., Runge, M.C., Regehr, E.V., Amstrup, S.C., & Stirling, I. (2010). Climate change threatens polar bear populations: A stochastic demographic analysis. *Ecology* , **91** , 2883-2897.
- Iversen, M., Aars, J. Haug, T., Alsos, I.G., Lydersen, C., Bachmann, L., & Kovacs, K.M. (2013). The diet of polar bears (*Ursus maritimus* ) from Svalbard, Norway, inferred from scat analysis. *Polar Biology* , **36** , 561–571.
- Jensen, E.L., Tschirter, C., V.C. de Groot, P., Hayward, K.M., Branigan, M., Dyck, M., ... Lougheed, S.C. (2020). Canadian polar bear population structure using genome-wide markers. *Ecology and Evolution* , **10** , 3706-3714.
- Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, **27** , 3070-3071.
- Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics* , **11** , 94.
- Kleinman-Ruiz, D., Martinez-Cruz, B., Soriano, L., Lucena-Perez, M., Cruz, F., Villanueva, B., ... Godoy, J.A. (2017). Novel efficient genome-wide SNP panels for the conservation of the highly endangered Iberian lynx. *BMC Genomics* , **18** , 556.
- Kraus, R.H.S., vonHoldt, B., Cocchiararo, B., Harms, V., Bayerl, H., Kühn, R., ... Nowak, C. (2015). A single-nucleotide polymorphism-based approach for rapid and cost-effective genetic wolf monitoring in Europe based on noninvasively collected samples. *Molecular Ecology Resources* , **15** , 295–305.
- Laidre, K.L., Stern, H., Kovacs, K.M., Lowry, L., Moore, S.E., Regehr, E.V., ... Ugarte, F. (2015). Arctic marine mammal population status, sea ice habitat loss, and conservation recommendations for the 21st century. *Conservation Biology* , **29** , 724–737.
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* , **27** , 2987-2993.
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler

transform. *Bioinformatics* , **25** , 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* , **25** , 2078-2079.

Liu, S., Lorenzen, E.D., Fumagalli, M., Li, B., Harris, K., Xiong, Z., . . . Wang, J. (2014). Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. *Cell* , **157** , 785-794.

Lundin, J.I., Dills, R.L., Ylitalo, G.M., Hanson, M.B., Emmons, C.K., Schorr, G.S., . . . Wasser, S.K. (2016). Persistent organic pollutant determination in killer whale scat samples: Optimization of a gas chromatography/mass spectrometry method and application to field samples. *Archives of Environmental Contamination and Toxicology* , **70** , 9-19. Lundin, J.I., Riffell, J.A., & Wasser, S.K. (2015). Polycyclic aromatic hydrocarbons in caribou, moose, and wolf scat samples from three areas of the Alberta oil sands. *Environmental Pollution* , **206** , 527-534.

Malenfant, R.M., Coltman, D.W., & Davis, C.S. (2015). Design of a 9K Illumina BeadChip for polar bears (*Ursus maritimus* ) from RAD and transcriptome sequencing. *Molecular Ecology Resources* , **15** , 587-600.

Malenfant, R.M., Davis, C.S., Cullingham, C.I., & Coltman, D.W. (2016). Circumpolar genetic structure and recent gene flow of polar bears: A reanalysis. *PLoS ONE* , **11** , e0148967.

Malenfant, R., Cullingham, C., Coltman, D., Richardson, E., Dyck, M., Lunn, N., . . . Davis, C. (in press). Population genomics reveals historical divergence and local adaptation in polar bears. *Authorea* , submitted to *Molecular Ecology* .

Maroso, F., Hillen, J.E.J., Pardo, B.G., Gkagkavouzis, K., Coscia, I., Hermida, M., . . . AquaTrace Consortium. (2018). Performance and precision of double digest RAD (ddRAD) genotyping in large multiplexed datasets of marine fish species. *Marine Genomics* , **39** , 64-72.

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

Meirmans, P.G., & Van Tienderen, P.H. (2004). GENOTYPE and GENODIVE: Two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes* , **4** , 792-794.

Morden, C.-J.C., Weladji, R.B., Ropstad, E., Dahl, E., Holand, Ø., Mastro Monaco, G., & Nieminen, M. (2011). Fecal hormones as a non-invasive population monitoring method for reindeer. *The Journal of Wildlife Management* , **75** , 1426-1435.

- Morin, D.J., Waits, L.P., McNitt, D.C., & Kelly, M.J. (2018). Efficient single-survey estimation of carnivore density using fecal DNA and spatial capture-recapture: A bobcat case study. *Population Ecology* , **60** , 197-209.
- Natesh, M., Taylor, R.W., Truelove, N.K., Hadly, E.A., Palumbi, S.R., Petrov, D.A., & Ramakrishnan, U. (2019). Empowering conservation practice with efficient and economical genotyping from poor quality samples. *Methods in Ecology and Evolution* , **10** , 853-859.
- Nelms, S.E., Parry, H.E., Bennett, K.A., Galloway, T.S., Godley, B.J., Santillo, D., & Lindeque, P.K. (2019). What goes in, must come out: Combining scat-based molecular diet analysis and quantification of ingested microplastics in a marine top predator. *Methods in Ecology and Evolution*, **10**, 1712-1722.
- Ogurtsov, S.S. (2018). The diet of the brown bear (*Ursus arctos* ) in the Central Forest Nature Reserve (West-European Russia), based on scat analysis data. *Biology Bulletin* , **45** , 1039-1054.
- Paetkau, D., Amstrup, S.C., Born, E.W., Calvert, W., Derocher, A.E., Garner, G.W., ... Strobeck, C. (1999). Genetic structure of the world's polar bear populations. *Molecular Ecology* , **8** , 1571–1584.
- Paetkau, D., Calvert, W., Stirling, I., & Strobeck, C. (1995). Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology* , **4** , 347-354.
- Pages, M., Maudet, C., Bellemain, E., Taberlet, P., Hughes, S., & Hänni, C. (2009). A system for sex determination from degraded DNA: A useful tool for palaeogenetics and conservation genetics of ursids. *Conservation Genetics* , **10** , 897-907.
- Perry, G.H., Marioni, J.C., Melsted, P., & Gilad, Y. (2010). Genomic-scale capture and sequencing of endogenous DNA from feces. *Molecular Ecology* , **19** , 5332-5344.
- 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** , 557-561.
- Poinar, H.N., Höss, M., Bada, J.L., & Pääbo, S. (1996). Amino acid racemization and the preservation of ancient DNA. *Science* , **272** , 864–866.
- Pritchard, J.K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* , **155** , 945– 959.
- Queller, D.C., & Goodnight, K.F. (1989). Estimating relatedness using genetic markers. *Evolution* , **43** , 258–275.
- Quinn, C.B., Alden, P.B., & Sacks, B.N. (2019). Noninvasive sampling reveals short-term genetic rescue in an insular red fox population. *Journal of Heredity* , **110** , 559-576.
- Robinson, R.A., Morrison, C.A., & Baillie, S.R. (2014). Integrating demographic data: Towards

a framework for monitoring wildlife populations at large spatial scales. *Methods in Ecology and Evolution*, **5**, 1361-1372.

Rode, K.D., Regehr, E.V., Douglas, D.C., Durner, G., Derocher, A.E., Thiemann, G.W., & Budge, S.M. (2014). Variation in the response of an Arctic top predator experiencing habitat loss: Feeding and reproductive ecology of two polar bear populations.

*Global Change*

*Biology* , **20** , 76–88.

Schmidt, D.A., Campbell, N.R., Govindarajulu, P., Larsen, K.W., Russello, M.A. (2020).

Genotyping-in-thousands by sequencing (GT-seq) panel development and application to minimally invasive DNA samples to support studies in molecular ecology. *Molecular Ecology Resources* , **20** , 114-124.

Schultz, A.J., Cristescu, R.H., Littleford-Colquhoun, B.L., Jaccoud, D., & Frere, C.H. (2018). Fresh is best: Accurate SNP genotyping from koala scats. *Ecology and Evolution* , **8** , 3139– 3151.

Sjodin, B.M.F., Irvine, R.L., & Russello, M.A. (2020). RapidRat: Development, validation and application of a genotyping-by-sequencing panel for rapid biosecurity and invasive species management. *PLoS ONE* , **15** , e0234694.

Snyder-Mackler, N., Marjoros, W.H., Yuan, M.L., Shaver, A.O., Gordon, J.B., Kopp, G.H., ... Tung, J. (2016). Efficient genome-wide sequencing and low-coverage pedigree analysis from noninvasively collected samples. *Genetics* ,**203** , 699-714.

Solberg, K.H., Bellemain, E., Drageset, O.M., Taberlet, P., & Swenson, J.E. (2006). An evaluation of field and non-invasive genetic methods to estimate brown bear (*Ursus arctos* ) population size. *Biological Conservation* ,**128** , 158–168.

Stapleton, S., Atkinson, S., Hedman, D., & Garshelis, D. (2014). Revisiting Western Hudson Bay: Using aerial surveys to update polar bear abundance in a sentinel population. *Biological Conservation* , **170** , 38–47.

Steyer, K., Kraus, R.H.S, Mölich, T., Anders, O., Cocchiararo, B., Frosch, C., ... Nowak, C. (2016). Large-scale genetic census of an elusive carnivore, the European wildcat (*Felis s. silvestris* ). *Conservation Genetics* ,**17** ,1183–1199.

Taberlet, P., Waits, L.P., & Luikart, G. (1999). Noninvasive genetic sampling: Look before you leap. *Trends in Ecology & Evolution* , **14** , 323–327.

Taylor, M.K., Laake, J., McLoughlin, P.D., Cluff, H.D., & Messier, F. (2009). Demography and population viability of polar bears in the Gulf of Boothia, Nunavut.

*Marine Mammal*

*Science* , **25** , 778–796.

van Coeverden De Groot, P., Wong, P.B.Y., Harris, C., Dyck, M.G., Kamookak, L., Pages, M., ... Boag, P.T. (2013). Toward a non-invasive Inuit polar bear survey: Genetic data from polar bear hair snags. *Wildlife Society Bulletin* , **37** , 394-401.

von Thaden, A., Cocchiararo, B., Jarausch, A., Jüngling, H., Karamanlidis, A.A., Tiesmeyer, A., ... Munoz-Fuentes, V. (2017). Assessing SNP genotyping of noninvasively collected wildlife samples using microfluidic arrays. *Scientific Reports* , **7** , 10768.

Vynne, C., Baker, M.R., Breuer, Z.K., & Wasser, S.K. (2012). Factors influencing degradation of DNA and hormones in maned wolf scat. *Animal Conservation* , **15** , 184-194.

Waits, L.P., & Paetkau, D. (2005). Noninvasive genetic sampling tools for wildlife biologists: A review of applications and recommendations for accurate data collection.

*The Journal*

*of Wildlife Management* , **69** ,1419-1433.

Weese, J.S., Salgado-Bierman, F., Rupnik, M., Smith, D.A., & van Coeverden de Groot, P. (2019). *Clostridium (Clostridioides ) difficile* shedding by polar bears (*Ursus maritimus* ) in the Canadian Arctic. *Anaerobe* , **57** , 35-38.

Wong, P.B.Y., Dyck, M.G., & Murphy, R.W. (2017). Inuit perspectives of polar bear research: Lessons for community-based collaborations. *Polar Record* , **53** , 257-270.

York, J., Dowsley, M., Cornwell, A., Kuc, M., & Taylor, M. (2016). Demographic and traditional knowledge perspectives on the current status of Canadian polar bear subpopulations. *Ecology and Evolution* , **6** , 2897-2924.

#### Data Accessibility and Benefit-Sharing Statement

Upon acceptance, we will upload raw genotyping reads, analysis scripts, and sequencing primers to an online repository. The sequencing data are currently owned by the Governments of the Northwest Territories and Nunavut. We will seek approval for these data to be published.

#### Author Contributions

KMH participated in study design, conducted laboratory work, performed bioinformatic processing and analysis, and drafted the manuscript. ELJ participated in study design, genomic data collection, bioinformatic processing, and analysis. RCC participated in genomic data collection and was crucial for panel optimization. MB, MD, and PVC DG managed access to samples and provided key insights to polar bear ecology and management implications. ZS performed laboratory work and CT aided with analysis. SCL conceived and coordinated study design and helped draft the initial manuscript. All authors contributed to manuscript revisions.