




# Genotyping-in-thousands by sequencing (GT-seq) of noninvasive faecal and degraded samples: A new panel to enable ongoing monitoring of Canadian polar bear populations

Kristen M. Hayward<sup>1</sup>  | Rute B.G. Clemente-Carvalho<sup>2</sup> | Evelyn L. Jensen<sup>3</sup>  | Peter V.C. de Groot<sup>1</sup> | Marsha Branigan<sup>4</sup> | Markus Dyck<sup>5</sup> | Christina Tschirter<sup>1</sup> | Zhengxin Sun<sup>1</sup> | Stephen C. Loughheed<sup>1</sup> 

<sup>1</sup>Department of Biology, Queen's University, Kingston, Ontario, Canada

<sup>2</sup>Hakai Institute, Campbell River, British Columbia, Canada

<sup>3</sup>School of Natural and Environmental Sciences, Newcastle University, Newcastle, UK

<sup>4</sup>Department of Environment and Natural Resources, Government of the Northwest Territories, Inuvik, Northwest Territories, Canada

<sup>5</sup>Department of Environment, Government of Nunavut, Igloolik, Nunavut, Canada

## Correspondence

Kristen M. Hayward and Stephen C. Loughheed, Department of Biology, Queen's University, Kingston, ON, Canada.

Emails: [k.hayward@queensu.ca](mailto:k.hayward@queensu.ca) and [steve.loughheed@queensu.ca](mailto:steve.loughheed@queensu.ca)

## Funding information

Genome Canada, Grant/Award Number: OGI-123

## Abstract

Genetic monitoring using noninvasive 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 noninvasive samples (e.g., scat) is often challenged by poor DNA quality and contamination by nontarget 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 noninvasively 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 noninvasively collected faecal samples. We demonstrate (1) successful GT-seq genotyping of DNA from a range of sample sources, including successful genotyping (>50% loci) of 62.9% of noninvasively collected faecal 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 spatial-genetic patterns similar to previous polar bear studies but at lesser cost per sample and through use of noninvasively collected samples, indicating the potential of this approach for population monitoring. This GT-seq panel provides the foundation for a noninvasive toolkit for polar bear monitoring and can contribute to community-based programmes – a framework which may serve as a model for wildlife conservation and management for species worldwide.

## KEYWORDS

faecal, GT-seq, monitoring, noninvasive, polar bear

Rute B.G. Clemente-Carvalho and Evelyn L. Jensen contributed equally to this work.

This paper is dedicated to our friend, colleague, and collaborator Markus Dyck, who died in a helicopter accident while doing fieldwork. His love for the Canadian Arctic and polar bear research was unmatched. We will miss him.

This is an open access article under the terms of the [Creative Commons Attribution - NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *Molecular Ecology Resources* published by John Wiley & Sons Ltd.

## 1 | 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 et al., 2018; Hamilton & Derocher, 2019; Laidre et al., 2015). Population monitoring is especially urgent for species that are impacted by rapid climate change, are harvested or poached, or reside in habitats that have been heavily altered by human activity (Durner et al., 2018; Laidre et al., 2015; Robinson et al., 2014). Such monitoring can be challenging for species that have large territories, 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 (Van Coeverden de Groot et al., 2013; Solberg et al., 2006; Stapleton et al., 2014). 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 et al., 2002).

Genetic monitoring using noninvasive samples, such as scat, hair, feathers, or skin, affords an alternative that can mitigate some of the challenges of traditional monitoring. Noninvasive genetic monitoring can be deployed on a large scale and with greater frequency, potentially enabling larger sample sizes and improved temporal monitoring. Other benefits include ease of collection, reduced disturbance of study species, potentially decreased spatial or temporal biases, and diminished physical risk to collectors (Carroll et al., 2018; Morin et al., 2018; Steyer et al., 2016; Waits & Paetkau, 2005). Noninvasive genetic monitoring can provide robust and repeatable data for individual identification (including sex), movement, and population trends (e.g. Aziz et al., 2017; Quinn et al., 2019; Schmidt et al., 2020; Schultz et al., 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 et al., 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 et al., 2012), and contaminant loads (Lundin et al., 2015, 2016).

Despite the purported advantages of noninvasive genetic monitoring, it poses challenges that have limited its widespread implementation. DNA from noninvasive samples may be degraded due to environmental exposure (Bourgeois et al., 2019; Poinar et al., 1996; Schultz et al., 2018) and contaminated by nontarget species (Carroll et al., 2018; Taberlet et al., 1999). Often less than 5% of the total DNA that scat contains is host DNA, with most DNA coming from pathogens, parasites, commensal bacteria, prey, and off-target species (Han et al., 2019; Perry et al., 2010; Snyder-Mackler et al., 2016). Due to low quality and quantity of host DNA, accurate quantification and genotyping of noninvasive samples using next-generation

sequencing (NGS) methods have often proved difficult, as they require high DNA concentration and quality (Graham et al., 2015; Maroso et al., 2018; Taberlet et al., 1999). Such reduced genotyping accuracy due to contaminated and degraded DNA may increase processing efforts and costs, and complicate inferences from collected data.

Multiple methods that have been developed to improve NGS of low-quality samples (e.g., scat, hair, or archaeological samples) 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 et al., 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 noninvasive 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 et al., 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 noninvasively 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 et al., 2015). GT-seq library preparation can be performed with standard laboratory equipment and shows decreased genotyping error and genotyping costs relative to other NGS-based genotyping methods, including TaqMan assays (Campbell et al., 2015). Because GT-seq uses suites of multiplexed, species-specific primers, it may mitigate some of the challenges presented by exogenous DNA and degraded host DNA in noninvasive faecal samples. Combined with new Illumina platforms like NovaSeq, GT-seq costs can be decreased further by running up to thousands of individuals on a single lane. 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 noninvasive 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 noninvasively 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 based on noninvasively 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 et al., 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 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 et al., 2017; York et al., 2016). Two-thirds (13) of the 19 polar bear subpopulations are fully or partially found in Canada (Figure 1), 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 six subpopulations had population estimates <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 (Fontúrbel et al., 2018; Hamilton & Derocher, 2019; Hunter et al., 2010; Rode et al., 2014), there is an urgent need for new monitoring approaches.

Noninvasive scat surveys may enable more direct community participation, and provide a cost-effective complement to traditional polar bear monitoring methods. Noninvasive scat samples could be obtained through community-level monitoring programmes, in which Inuit hunters are remunerated for field sampling efforts and surveys are repeated regularly to better track the trajectory of polar bear subpopulations. Noninvasive 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 et al., 2019). We have also established that sufficient 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 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

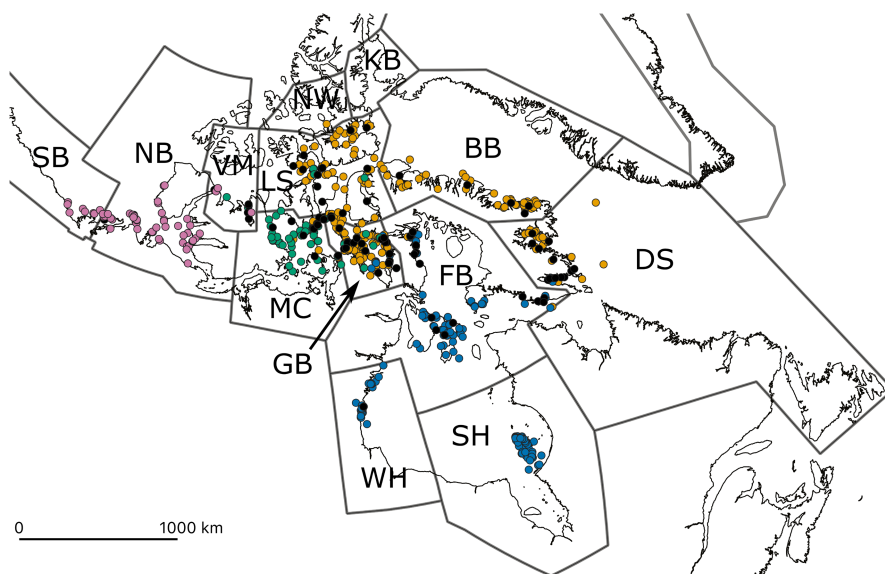
communities and are predicted to have real socioeconomic benefits for the communities involved.

In this study, we test the practicality of using GT-seq for SNP genotyping of noninvasively collected faecal samples from polar bears and apply GT-seq to population monitoring, using degraded samples that heretofore could not be genotyped with NGS-based methods, to expand our understanding of Canadian polar bear genetic structure. We show that our optimized GT-seq panel can be used to distinguish among individuals (including noninvasively collected samples), and to characterize genetic structure of the Canadian polar bear population.

## 2 | MATERIALS AND METHODS

### 2.1 | 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). The distribution of these samples is depicted in Figure 1 and Table 1 of Jensen et al. (2020). We filtered this data set 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 (Table S1 presents the full filtering workflow). This resulted in 442 retained SNP loci. To allow for sex identification, we added two SNPs known to be sex-linked in polar bears (Pagès et al., 2009). We validated our GT-seq panel of 442 SNPs and designed primers as described in Supporting Information: GT-seq Panel Development and Validation. To ensure our potential panel could distinguish individuals with high reliability, validation included calculating probability of identity (PID) and probability of identity (siblings) for an increasing number of loci from our ddRADseq data. We could reliably identify individuals with as few as



**FIGURE 1** Map of the Canadian Arctic showing currently recognized subpopulations that are fully or partially in Canada (solid line polygons). Subpopulation abbreviations are the same as in Table 2. Points correspond to sampled individuals, 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

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

Sample type	<i>n</i>	% individuals >50% data	% individuals >50% data after removal of samples qPCR = 0 ng/μl	Average % missing data by locus (range)
Set muscle (MS)	101	96.2	–	2.2 (0.31–45.3)
Colon faeces (CF)	69	88.5	90.1	2.1 (0.31–23.3)
Biopsy (BP)	134	95.7	–	3.1 (0.31–39.1)
Harvest muscle (HV)	38	97.4	–	1.1 (0.31–2.5)
Field faeces (FF)	23	30.6	62.9	14.9 (0.31–50.0)

*Note:* Set muscle, tissue from corresponding muscle and colon sets; Colon faeces, faeces removed from the colon of corresponding muscle and colon sets; Biopsy, biopsy tissue sample; Harvest muscle, tissue from annual polar bear harvest; Field faeces, noninvasively collected scat from the field.

34 SNPs, but included a panel of >300 SNPs to achieve both individual identification and sufficient resolution for population genetic analyses.

## 2.2 | Sample collection

Our work draws on multiple sample sources (Figure S1), 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 of 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 faecal genotyping error by comparing genotypes from set tissue (MS; *n* = 108) and faeces from the colon (CF; *n* = 78), with expectation that these estimates will be conservative. Noninvasive, field-collected faecal samples (FF; *n* = 72) were also located and collected by Inuit hunters under wildlife research permit WL-2018-006 to SCL. All sample types were stored in –20°C or –80°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°C or –80°C and CF and FF subsamples were stored without ethanol at –20°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 withing subpopulations ranging from 1999 (M'Clintock Channel) to 2018 (Southern Hudson Bay; Figure 1). For nine subpopulations, we had at least 10 sampled individuals (range: 10–95). However, there were only four 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).

## 2.3 | DNA extraction

Whole genomic DNA was extracted from faecal samples using the QIAamp Fast DNA Stool Mini Kit (Qiagen) according to the 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 Nucleic Acid Staining Solution (iNtRON Biotechnology) to assess quality and quantified using a Nanodrop ND\_1000 spectrophotometer (Nanodrop Technologies Inc.). We used a polar bear-specific qPCR assay targeting the *F2* gene to quantify the amount of polar bear DNA in both CF and FF samples (Hayward et al., 2020). To gauge the value of running faecal samples through this qPCR assay as a screening tool, we devised a small double-blind experiment in which we randomly divided FF samples into two subsets: 1. 8 samples for which we assayed DNA quantity before GT-seq library construction and sequencing; and 2. 89 samples (+21 duplicates) for which we assessed DNA quantity only after sequencing had already been performed (see Supporting Information: qPCR Experiment for qPCR experiment methods and results).

## 2.4 | GT-seq genotyping and genotype calling methods

A full description of GT-seq panel optimization can be found in Supporting Information: Panel Optimization. Using optimized

GT-seq conditions and primers for our final GT-seq panel, we prepared two libraries for faecal samples (CF, FF) and another three for tissue samples (HV, BP, MS). The faecal libraries contained all 150 faecal samples, as well as the 21 FF duplicates to assess subsample variation in genotyping error. We included 284 tissue samples in the tissue libraries and two technical replicates (BP) to assess genotyping error for GT-seq. Genotyping error was measured conservatively as percent discordance between technical replicates. Within our total of 457 samples, we included 65 paired sets of MS and CF samples collected from the same individuals allowing us to compare GT-seq genotyping error between 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 data set, as we considered samples with >50% data (i.e. genotypes for at least half of panel loci) to have been successfully genotyped. We considered a sample of 161 SNPs (50%) genotyped to be a “success” because it enabled sufficient resolution for both individual identification and analysis of population structure, although far fewer SNPs can be used to reliably distinguish among individuals (Supporting Information Results: GT-seq panel validation). To assess the validity of the sex identities provided by our two GT-seq SNPs, we compared the GT-seq determined polar bear sex to hunter-provided sex 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 Supporting 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 Qnd was used to call genotypes for our original ddRADseq data set (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 of either six (data set BCF-6) or 10 (data set 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 data set with the least missing data (BCF-6) was used to calculate genotyping error and percent missing data, and for all subsequent population genetic analyses.

## 2.5 | 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 data set called and filtered with our BCF-6 workflow. This data set comprised individuals genotyped at 322 autosomal SNPs using GT-seq, as well as individuals previously genotyped for these same SNPs 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. For individuals that had been included in both data sets, only GT-seq replicates were retained. Our final GT-seq+ddRADseq data set contained 642 individuals genotyped at 322 autosomal loci and 2 sex-linked loci.

For the combined GT-seq + ddRADseq data set, we estimated observed and expected heterozygosity ( $H_o/H_e$ ) and inbreeding coefficients ( $G_{is}$ ) for each subpopulation, as implemented in GENODIVE v3.04 (Meirmans, 2020). Using the RELATED package in R (Pew et al., 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 data set to test the ability of multiple relatedness estimators to distinguish among relatedness categories. We also calculated and plotted pairwise relatedness for all 642 individuals in our combined data set (Figure S4). To assess population substructure, we used Bayesian clustering analysis as implemented in STRUCTURE 2.3.4 (Pritchard et al., 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 burnin period of 100,000 MCMC. The most likely value of  $K$  was identified using the DeltaK method of Evanno et al. (2005) in STRUCTURE HARVESTER (Earl & vonHoldt, 2012). To complement the STRUCTURE analysis, we also used discriminant analysis of principal components (DAPC; Jombart et al., 2010), implemented in adegenet in R (Jombart & Ahmed, 2011) to evaluate the number of genetic clusters in the data. The best-fit value of  $K$  was tested using the find.clusters function, retaining all PCs, and the Bayesian information criterion (BIC). The chosen value of  $K$  was selected based on the lowest BIC value, and a DAPC plot was generated based on this clustering, retaining sufficient PCs to capture 80% of the variance. Principal component analysis (PCA) 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).



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

Subpopulation	<i>n</i>	<i>H<sub>o</sub></i>	<i>H<sub>E</sub></i>	<i>G<sub>IS</sub></i>	<i>p</i> ( <i>G<sub>IS</sub></i> >0)	Self-assignment	Main genetic cluster
Baffin Bay (BB)	50	0.45	0.45	−0.006	.23	0	Arctic Archipelago
Davis Strait (DS)	40	0.45	0.45	0.004	.311	0.01	Arctic Archipelago
Foxe Basin (FB)	82	0.45	0.44	−0.025	.001	0.52	Hudson Complex
Gulf of Boothia (GB)	110	0.45	0.45	0.003	.282	0.73	Arctic Archipelago
Lancaster Sound (LS)	83	0.45	0.45	−0.009	.083	0.03	Arctic Archipelago
M'Clintock Channel (MC)	80	0.45	0.44	−0.023	.001	0.56	M'Clintock Channel
Northern Beaufort Sea (NB)	55	0.44	0.43	−0.019	.01	0.04	Polar Basin
Norwegian Bay (NW)	1	–	–	–	–	–	Arctic Archipelago
Southern Beaufort Sea (SB)	12	0.45	0.42	−0.070	.001	0	Polar Basin
Southern Hudson Bay (SH)	80	0.44	0.43	−0.017	.003	0.65	Hudson Complex
Viscount Melville Sound (VM)	11	0.45	0.44	−0.018	.169	0	M'Clintock Channel/ Polar Basin
Western Hudson Bay (WH)	38	0.44	0.43	−0.017	.037	0	Hudson Complex

Abbreviations: *G<sub>IS</sub>*, inbreeding coefficient; *H<sub>E</sub>*, expected heterozygosity; *H<sub>o</sub>*, observed heterozygosity; *n*, sample size.

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 data set. The remaining 25% of individuals were used to test the rate of assignment, which was then averaged across 30 iterations.

### 3 | RESULTS

#### 3.1 | Calling workflow comparisons

Our final GT-seq data set, called using the published GT-seq pipeline (Campbell et al., 2015), included 325 autosomal SNPs and two sex-linked markers. An additional three loci were removed from the BCFtools workflow data sets after filtering for minimum depth (depth 6, 10), leaving each with 322 autosomal loci and two sex-linked markers. We removed the same three loci from our GT-seq pipeline data set 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% (*SD* = ±5.3%) on average. Based on these results and the potential for easy comparison with existing ddRADseq data, we chose to use the data set generated from

the BCF-6 calling workflow to assess genotyping error and analyse population structure.

#### 3.2 | Genotyping success

After removing individuals with >50% missing data, a total of 365 samples, including three duplicate samples, remained in our GT-seq data set. 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/μl polar bear DNA), percent individuals for which we had sufficient SNPs to estimate genetic identity (>34 SNPs) was 80.0% and percent individuals “successfully genotyped” as per our cutoff (>50% loci genotyped) was 62.9%. See Supporting 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 giving a conservative 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 prescreening, 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 some cases with large differences in genotyping success between FF duplicates. For example, one FF subsample was genotyped at only eight 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%. Fifty-six of the 65 muscle-colon

faeces sets had both sample types from the same individual successfully genotyped, giving a mean genotyping error estimate of 6.8% for faecal 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 faeces.

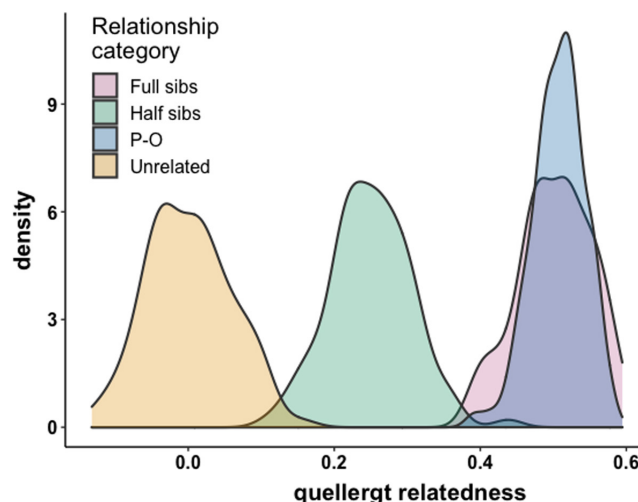
We had a total of 293 bear samples for which hunters provided a sex identity. Of these, 78.6% were genotyped successfully at one or both sex-linked loci. There were 18 instances where a sample had only a genotype at one of the sex-linked loci, and only one case of allelic dropout. For samples with genotypes 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.

### 3.3 | Population structure

Based on the final combined data set of 642 individuals (GT-seq+ddRADseq data at 322 GT-seq loci), all subpopulations display similar levels of genetic diversity (Table 2). Expected heterozygosity ranged from 0.42 to 0.45 and  $G_{IS}$  values were mostly negative, ranging from -0.006 to 0.004.

We discovered 13 true recaptures in our combined data set (relatedness >0.80), which were removed from all subsequent analysis. Recaptures spanned sample types, with pairs existing between 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 (nonlethal) was later detected using a CF sample (lethal). Based on comparison of multiple relatedness estimators, we determined that the *queller* correlation coefficient was best for our combined data set (correlation between observed and expected values = 0.97; Queller & Goodnight, 1989). Results from simulations with the *queller* coefficient yielded pairwise relatedness density plots with marked separation for full siblings, half siblings and parent-offspring pairs, and unrelated individuals (Figure 2). Although relatedness among individuals in our data set was mostly unknown prior to analysis, we can see >50 empirical relatedness values of approximately 0.5 (Figure S4), as would be expected for parent-offspring or sibling pairs.

Our STRUCTURE and DAPC analyses revealed similar patterns in population genetic structure. STRUCTURE analysis suggested an optimal value of  $K = 4$ , with  $\ln \text{Pr}(X|K)$  plateauing around  $K = 4$  (Figure 3d) and  $\Delta K$  greatest at  $K = 4$  (Figure 3e). For the find.clusters DAPC analysis, the lowest BIC scores occurred at  $K = 3$  and  $K = 4$  (Figure 3c). 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 3a,b, Figure 1). Barplots showing  $K = 3$  and  $K = 5$  are provided in Figure S6.

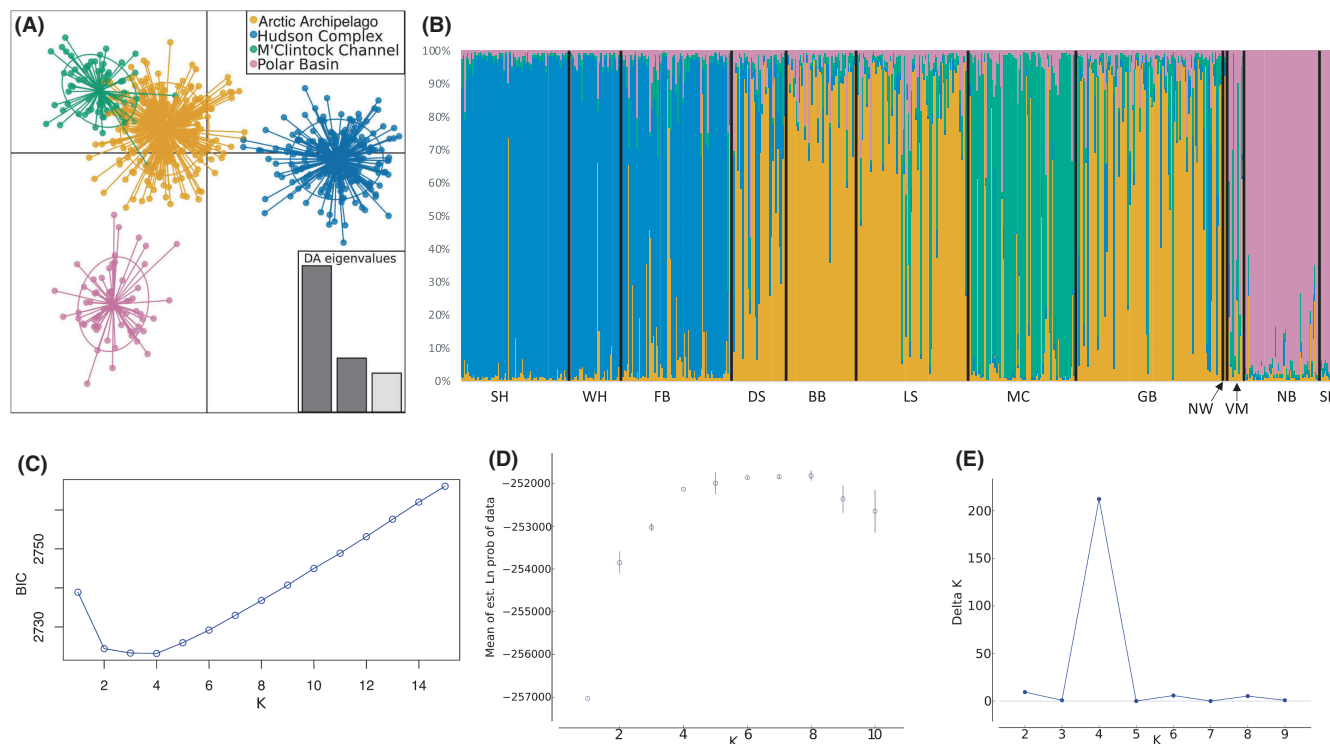


**FIGURE 2** Density plot of pairwise *queller* relatedness values generated from simulations (100 per relatedness category) based on allele frequencies from our combined GT-seq+ddRADseq data set of 642 individuals genotyped at 322 autosomal loci. Each colour represents a different relatedness category (Full sibs, full siblings; Half sibs, half siblings; P-O, parent-offspring; Unrelated, unrelated pair)

Despite 90 individuals having values of less than 0.7 membership to a particular genetic cluster (black dots in Figure 1), most of the 642 individuals in our combined ddRADseq and GT-seq data set 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. 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).

## 4 | DISCUSSION

Noninvasive samples (e.g., faeces) are viable sources of DNA for new genetic methods of population monitoring, which may help mitigate limitations associated with traditional approaches. Here, we developed a final GT-seq assay of 324 SNPs optimized to genotype polar bears based on DNA from noninvasively collected faecal samples. We demonstrated (1) successful GT-seq genotyping of DNA from field-collected polar bear faeces, and (2) the practicality of GT-seq for distinguishing individuals, assessing relatedness,



**FIGURE 3** Results of two clustering analyses performed with combined GT-seq+ddRADseq data set 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

**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 data set of 642 bears genotyped at 322 autosomal loci

Genetic cluster	<i>n</i>	Self-assignment
Arctic Archipelago	230	0.98
Polar Basin	66	0.35
Hudson Complex	188	1.00
M'Clintock Channel	68	0.83

Note: *n*, sample size.

and expanding our understanding of genetic population structure and diversity.

#### 4.1 | Individual identification

We determined that all existing unrelated bears could be distinguished with as few as 34 SNPs (Figure S2), and that 80% of field-collected scat with detectable polar bear DNA could be assigned a genetic identity (i.e.,  $>34$  SNPs genotyped). Given the current global polar bear population estimate of 23,315 (Hamilton & Derocher,

2019), our panel of 322 autosomal and two sex-linked SNPs is sufficient to reliably discern bears among most relationship categories (e.g., siblings, unrelated) with high certainty (Figure 2). Given this ability to differentiate individuals based on both faecal samples and tissue, our GT-seq assay may be particularly useful for genetic mark-recapture studies using data from traditional sources and noninvasive samples (e.g., scat, urine, hair snags).

#### 4.2 | Comparing workflows, error rates, and genotyping success

GT-seq uses a reduced panel of SNPs to genotype samples with degraded DNA, such as scat or archived tissue samples, and as long as there are loci in common, GT-seq data can be combined with data from other methods (e.g., ddRADseq). However, differences in genotyping calling methods between GT-seq and RADseq data sets may contribute to discordance among data sets (Schmidt et al., 2020). To address this, 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 for our ddRADseq data (Jensen et al., 2020). As genotypic discordance between the two calling workflows was low (1.1%) and levels of missing data were moderate (21.3%–25.4%), we suggest that workflow choice is



primarily a matter of preference. Regardless, we recommend using consistent variant calling protocols when working with both GT-seq and ddRADseq data.

GT-seq genotyping success was high across all sample types. In particular, field-collected scat samples with qPCR-detectable polar bear DNA were genotyped successfully (using our criterion of >50% SNPs genotyped) for 62.9% of individuals and demonstrated an average of 14.9% missing data. Based on MS:CF comparisons and FF sample replicates not screened before sequencing, we estimated genotyping error for field scat to be around 10%. 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 nontarget species. Thus, to minimize effort and money expended on faecal samples unlikely to yield GT-seq genotypes, we evaluated a qPCR screening assay (Hayward et al., 2020). The inclusion of a qPCR screening step substantially increased the percentage of field faecal samples successfully genotyped using GT-seq from 30.6% to 62.9%, with 22 of the 23 (95.6%) field faecal samples that passed this step being successfully genotyped. In one case, duplicate subsamples from the same field scat exhibited a large difference in genotyping success (1 failure, 1 success), suggesting that samples can be heterogeneous in the amount of host DNA. Thus, we recommend using repeated subsampling for scat samples for which genotyping is critical, and using a sample quality screening step such as our qPCR protocol for any future research making use of GT-seq and noninvasive samples. Additional testing of DNA extraction protocols for scat may also help to improve DNA recovery.

#### 4.3 | Assessment of population structure

By combining our new GT-seq data with ddRADseq data (Jensen et al., 2020), we were able to reassess Canadian polar bear population structure using a larger data set (642 vs. 358 individuals from Jensen et al., 2020) and make use of samples that were too degraded for genotyping with ddRADseq (e.g., faecal, degraded biopsy). Diversity metrics estimated using these combined data, including new genotypes from noninvasive samples, do not vary markedly among subpopulations and resemble those of previous studies based on SNPs, mtDNA, or microsatellites (e.g. Jensen et al., 2020; Malenfant et al., 2016), although our estimates may be higher due to selection of high diversity SNPs during panel design. Despite many subpopulations having only minor differences in genetic diversity, high assignment rates for a few subpopulations (e.g., GB, SH, MC) may have implications for conservation and management planning.

Sample clustering patterns and geographic distributions based on combined GT-seq and ddRADseq data were also largely consistent with genetic groups called the "Hudson Complex," "Arctic Archipelago," and "Polar Basin," which have been previously described using microsatellite and SNP data (Jensen et al., 2020; Malenfant et al., 2015, 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 1 and 3), a pattern not evident from analysis of ddRADseq data alone (Jensen et al., 2020). Rather than M'Clintock Channel comprising a major genetic cluster, it seems more likely that this new grouping reflects subtle genetic differentiation within the Arctic Archipelago (see Malenfant et al., 2016, 2020).

Considering the mean sample collection year for M'Clintock Channel of 1999 and the rapidness of environmental changes in the Canadian Arctic, our data may not represent contemporary genetic structuring of the Canadian polar bear population. Further, our data set contains samples collected between 1998 and 2018 (a 20-year span), which presents a potential temporal confound that is not uncommon in studies investigating polar bear population structure (e.g. Malenfant et al., 2016; Paetkau et al., 1995, 1999). Regardless, we show here that our GT-seq panel provides sufficient power to adequately capture known patterns that will be relevant to northern and federal governments for polar bear management. Importantly, we can detect these patterns through use of noninvasive and degraded samples, and GT-seq genotyping can be done with greater cost-efficiency than other genotyping methods irrespective of the number of samples available to be assayed at a time (Campbell et al., 2015). Thus, GT-seq presents an opportunity to iteratively assess the genetic structure of Canada's polar bear populations – a flexible, cost-effective means to continuously update our understanding as regular sampling is performed and governments shift towards non-invasive methods of population monitoring.

#### 4.4 | Other applications of GT-seq and considerations

Further considerations during GT-seq panel design and optimization may be required for other species and applications. For example, certain characteristics of a species may influence the number of markers required to discern individuals and populations, including mating system and linkage patterns within a genome. In the future, for polar bears we may wish to optimize a panel for assessment of adaptation in real time by including SNPs that are flagged as potentially under selection, or perhaps a panel to distinguish between bear species (e.g., grizzly and polar bears), similar to the application of GT-seq to invasive brown (*Rattus norvegicus*) and black rats (*R. rattus*; Sjodin et al., 2020).

Although we intended to include at least 30 individuals from each Canadian subpopulation in the ddRADseq data set (Jensen et al., 2020) that was used to develop our GT-seq panel, some subpopulations had few samples available. Ascertainment bias may thus be of concern as our GT-seq panel may not fully represent all Canadian subpopulations, nor was it designed with the inclusion of data from other Arctic nations. Thus, application to the global polar bear population may require some panel modifications. Establishment of baseline global polar bear population structure through GT-seq of noninvasive samples will be especially

important for examining trends in dispersal patterns, diversity and structure, and behaviour in the context of a rapidly changing climate.

## 5 | CONCLUSIONS

The need for monitoring using noninvasive genetic methods in the midst of rapid environmental changes, as well as the desire 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 monitor polar bears noninvasively. We demonstrate that we can use noninvasively collected samples and GT-seq to distinguish among individuals and to quantify population structure at levels comparable to other methods of genetic monitoring (e.g. Malenfant et al., 2016; Paetkau et al., 1995, 1999), but with greater efficiency even at small batch sizes (Campbell et al., 2015). The GT-seq assay that we developed here provides a foundation for new community-based programmes that can use noninvasive methods to improve temporal monitoring of polar bear populations and directly inform conservation efforts and government policy. The envisioned programmes will incorporate the perspectives of Indigenous communities throughout the planning and monitoring processes, and provide both social and economic benefits to them. This GT-seq protocol is also intended to serve as the basis 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 can be provided to communities and territorial governments annually. 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 can be adapted and applied as a model for inclusive wildlife monitoring worldwide.

## ACKNOWLEDGEMENTS

This study would not have been possible without the support and collaboration of the Inuvialuit Game Council, the Gjoa Haven and Coral Harbour Hunters and Trappers Associations, 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 laboratory work and have provided unwavering support for the project: M. Navarrete Bedolla, D. Loughheed, 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 Loughheed Laboratory. 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 programme. K. Hayward's Master's thesis is funded by an NSERC Alexander Graham Bell Canada Graduate Scholarship.

## CONFLICT OF INTEREST

The authors declare no conflicts of interest.

## AUTHOR CONTRIBUTIONS

Kristen M. Hayward participated in study design, conducted laboratory work, performed bioinformatic processing and analysis, and drafted the manuscript. Evelyn L. Jensen participated in study design, genomic data collection, bioinformatic processing, and analysis. Rute B.G. Clemente-Carvalho participated in genomic data collection and was crucial for panel optimization. Marsha Branigan, Markus Dyck, and Peter V.C. de Groot managed access to samples and provided key insights to polar bear ecology and management implications. Zhengxin Sun performed laboratory work and Christina Tschirter aided with analysis. Stephen C. Loughheed conceived and coordinated study design and helped draft the initial manuscript. All authors contributed to manuscript revisions.

## DATA AVAILABILITY STATEMENT

SNP calling pipelines (GT-seq and BCFTOOLS workflows) and analysis scripts with clear annotation and troubleshooting tips are all available on GITHUB ([https://github.com/kristenmhayward/GT-seq\\_2021](https://github.com/kristenmhayward/GT-seq_2021)). R scripts include how to perform discriminant analysis of principal components, relatedness analysis, self-assignment tests, and the logistic regression used to explore the relationship between GT-seq genotyping success and qPCR-detected DNA quantity. Also included are mainparams, extraparams, and job lists for STRUCTURE analysis, and full BCFTOOLS and GT-seq pipeline text files with clear instruction. The raw sequencing data are also available on GITHUB in vcf and genepop formats ([https://github.com/kristenmhayward/GT-seq\\_2021](https://github.com/kristenmhayward/GT-seq_2021)), and are owned by the Governments of the Northwest Territories and Nunavut and the communities that provided the samples.

## ORCID

Kristen M. Hayward  <https://orcid.org/0000-0002-2466-4587>

Evelyn L. Jensen  <https://orcid.org/0000-0002-1819-3117>

Stephen C. Loughheed  <https://orcid.org/0000-0003-1466-3668>

## 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. <https://doi.org/10.1093/nar/25.22.4692>
- 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., Shamsuddoha, M., Islam, M. A., & 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. <https://doi.org/10.1016/j.gecco.2017.09.002>
- Bergner, L. M., Orton, R. J., da Silva Filipe, A., Shaw, A. E., Becker, D. J., Tello, C., Biek, R., & Streicker, D. G. (2019). Using noninvasive metagenomics to characterize viral communities from wildlife. *Molecular Ecology Resources*, 19, 128–143. <https://doi.org/10.1111/1755-0998.12946>

- Blåhø, 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. <https://doi.org/10.1371/journal.pone.0197364>
- Bourgeois, S., Kaden, J., Senn, H., Bunnefeld, N., Jeffery, K. J., Akomo-Okoue, E. F., Ogden, R., & McEwing, R. (2019). Improving cost-efficiency of faecal genotyping: New tools for elephant species. *PLoS One*, 14, e0210811. <https://doi.org/10.1371/journal.pone.0210811>
- 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. <https://doi.org/10.1111/1755-0998.12357>
- 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. <https://doi.org/10.1111/eva.12600>
- Chen, K.-Y., Marshall, E. A., Sovic, M. G., Fries, A. C., Gibbs, H. L., & Ludsins, 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., & Frère, 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., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, 27, 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- 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. <https://doi.org/10.1007/s12686-011-9548-7>
- 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. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- 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. <https://doi.org/10.3996/112014-JFWM-080>
- Fontúrbel, F. E., Lara, A., Lobos, D., & Little, C. (2018). The cascade impacts of climate change could threaten key ecological interactions. *Ecosphere*, 9, e02485. <https://doi.org/10.1002/ecs2.2485>
- 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. <https://doi.org/10.1111/j.1365-294X.1997.tb00140.x>
- 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. <https://doi.org/10.1007/s12686-018-1017-0>
- Government of Canada (2018). *Maps of subpopulations of polar bears and protected areas*. Retrieved from <https://www.canada.ca/en/envir>
- onment-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. <https://doi.org/10.1111/acv.12439>
- Han, S., Guan, Y., Dou, H., Yang, H., Yao, M., Ge, J., & Feng, L. (2019). Comparison of the faecal 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., Coeverden, V., de Groot, P., & Jensen, E. L. (2020). A real-time PCR assay to accurately quantify polar bear DNA in faecal extracts. *PeerJ*, 8, e8884. <https://doi.org/10.7717/peerj.8884>
- Hess, J. E., Campbell, N. R., Docker, M. F., Baker, C., Jackson, A., Lampman, R., McIlraith, B., Moser, M. L., Statler, D. P., Young, W. P., Wildbill, A. J., & 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. <https://doi.org/10.1111/1755-0998.12283>
- 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. <https://doi.org/10.1890/09-1641.1>
- 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. <https://doi.org/10.1007/s00300-012-1284-2>
- Jensen, E. L., Tschirter, C., Groot, P. V. C., Hayward, K. M., Branigan, M., Dyck, M., Clemente-Carvalho, R. B. G., & Lougheed, S. C. (2020). Canadian polar bear population structure using genome-wide markers. *Ecology and Evolution*, 10, 3706–3714. <https://doi.org/10.1002/ece3.6159>
- Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27, 3070–3071. <https://doi.org/10.1093/bioinformatics/btr521>
- 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. <https://doi.org/10.1186/1471-2156-11-94>
- Kleinman-Ruiz, D., Martínez-Cruz, B., Soriano, L., Lucena-Perez, M., Cruz, F., Villanueva, B., Fernández, J., & Godoy, J. A. (2017). Novel efficient genome-wide SNP panels for the conservation of the highly endangered Iberian lynx. *BMC Genomics*, 18, 556. <https://doi.org/10.1186/s12864-017-3946-5>
- Kraus, R. H. S., vonHoldt, B., Cocchiaro, B., Harms, V., Bayerl, H., Kühn, R., Förster, D. W., Fickel, J., Roos, C., & 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. <https://doi.org/10.1111/1755-0998.12307>
- Laidre, K. L., Stern, H., Kovacs, K. M., Lowry, L., Moore, S. E., Regehr, E. V., Ferguson, S. H., Wiig, Ø., Boveng, P., Angliss, R. P., Born, E. W., Litovka, D., Quakenbush, L., Lydersen, C., Vongraven, D., & Ugarte, F. (2015). Arctic marine mammal population status, sea ice habitat loss, and conservation recommendations for the 21st century. *Conservation Biology*, 29, 724–737. <https://doi.org/10.1111/cobi.12474>
- 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. <https://doi.org/10.1093/bioinformatics/btr509>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Liu, S., Lorenzen, E. D., Fumagalli, M., Li, B. O., Harris, K., Xiong, Z., Zhou, L., Korneliusen, T. S., Somel, M., Babbitt, C., Wray, G., Li, J., He, W., Wang, Z., Fu, W., Xiang, X., Morgan, C. C., Doherty, A., O'Connell, M. J., ... Wang, J. (2014). Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. *Cell*, 157, 785–794. <https://doi.org/10.1016/j.cell.2014.03.054>
- Lundin, J. I., Dills, R. L., Ylitalo, G. M., Hanson, M. B., Emmons, C. K., Schorr, G. S., Ahmad, J., Hempelmann, J. A., Parsons, K. M., & 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. <https://doi.org/10.1007/s00244-015-0218-8>
- 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. <https://doi.org/10.1016/j.envpol.2015.07.035>
- 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., Cullingham, C., Coltman, D., Richardson, E., Dyck, M., Lunn, N., Obbard, M., Pongracz, J., Atkinson, S., Sahanatien, V., Laidre, K., Born, E., Wiig, Ø., & Davis, C. (2020). Population genomics reveals historical divergence and local adaptation in polar bears. *Authora*. <https://doi.org/10.22541/au.158136875.51417995>
- 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. <https://doi.org/10.1371/journal.pone.0148967>
- 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. <https://doi.org/10.1111/1755-0998.12998>
- Meirmans, P. G. (2020). GENODIVE version 3.0: Easy-to-use software for the analysis of genetic data of diploids and polyploids. *Molecular Ecology Resources*, 20, 1126–1131.
- Morden, C.-J.-C., Weladji, R. B., Ropstad, E., Dahl, E., Holand, Ø., Mastrotonaco, G., & Nieminen, M. (2011). Faecal hormones as a non-invasive population monitoring method for reindeer. *The Journal of Wildlife Management*, 75, 1426–1435. <https://doi.org/10.1002/jwmg.185>
- Morin, D. J., Waits, L. P., McNitt, D. C., & Kelly, M. J. (2018). Efficient single-survey estimation of carnivore density using faecal DNA and spatial capture-recapture: A bobcat case study. *Population Ecology*, 60, 197–209. <https://doi.org/10.1007/s10144-018-0606-9>
- 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. <https://doi.org/10.1111/2041-210X.13173>
- 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. <https://doi.org/10.1111/2041-210X.13271>
- 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. <https://doi.org/10.1134/S1062359018090145>
- Paetkau, D., Amstrup, S. C., Born, E. W., Calvert, W., Derocher, A. E., Garner, G. W., Messier, F., Stirling, I., Taylor, M. K., Wiig, O., & Strobeck, C. (1999). Genetic structure of the world's polar bear populations. *Molecular Ecology*, 8, 1571–1584. <https://doi.org/10.1046/j.1365-294x.1999.00733.x>
- Paetkau, D., Calvert, W., Stirling, I., & Strobeck, C. (1995). Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology*, 4, 347–354. <https://doi.org/10.1111/j.1365-294X.1995.tb00227.x>
- Pagès, 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. <https://doi.org/10.1007/s10592-008-9650-x>
- 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. <https://doi.org/10.1111/j.1365-294X.2010.04888.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, 557–561. <https://doi.org/10.1111/1755-0998.12323>
- Poinar, H. N., Höss, M., Bada, J. L., & Pääbo, S. (1996). Amino acid racemization and the preservation of ancient DNA. *Science*, 272, 864–866. <https://doi.org/10.1126/science.272.5263.864>
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959. <https://doi.org/10.1093/genetics/155.2.945>
- Queller, D. C., & Goodnight, K. F. (1989). Estimating relatedness using genetic markers. *Evolution*, 43, 258–275. <https://doi.org/10.1111/j.1558-5646.1989.tb04226.x>
- 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. <https://doi.org/10.1111/gcb.12339>
- 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. <https://doi.org/10.1111/1755-0998.13090>
- Schultz, A. J., Cristescu, R. H., Littleford-Colquhoun, B. L., Jaccoud, D., & Frère, C. H. (2018). Fresh is best: Accurate SNP genotyping from koala scats. *Ecology and Evolution*, 8, 3139–3151. <https://doi.org/10.1002/ece3.3765>
- 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. <https://doi.org/10.1371/journal.pone.0234694>
- Snyder-Mackler, N., Majoros, W. H., Yuan, M. L., Shaver, A. O., Gordon, J. B., Kopp, G. H., Schlebusch, S. A., Wall, J. D., Alberts, S. C., Mukherjee, S., Zhou, X., & Tung, J. (2016). Efficient genome-wide sequencing and low-coverage pedigree analysis from noninvasively collected samples. *Genetics*, 203, 699–714. <https://doi.org/10.1534/genetics.116.187492>
- 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. <https://doi.org/10.1016/j.biocon.2005.09.025>
- 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. <https://doi.org/10.1016/j.biocon.2013.12.040>
- Steyer, K., Kraus, R. H. S., Mölich, T., Anders, O., Cocchiara, B., Frosch, C., Geib, A., Götz, M., Herrmann, M., Hupe, K., Kohnen, A., Krüger, M., Müller, F., Pir, J. B., Reiners, T. E., Roch, S., Schade, U., Schiefenhövel, P., Siemund, M., ... Nowak, C. (2016). Large-scale genetic census of an elusive carnivore, the European wildcat (*Felis s. silvestris*). *Conservation Genetics*, 17, 1183–1199. <https://doi.org/10.1007/s10592-016-0853-2>
- Taberlet, P., Waits, L. P., & Luikart, G. (1999). Noninvasive genetic sampling: Look before you leap. *Trends in Ecology & Evolution*, 14, 323–327. [https://doi.org/10.1016/S0169-5347\(99\)01637-7](https://doi.org/10.1016/S0169-5347(99)01637-7)
- 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. <https://doi.org/10.1111/j.1748-7692.2009.00302.x>
- Van Coeverden de Groot, P. V. C., Wong, P. B. Y., Harris, C., Dyck, M. G., Kamookak, L., Pagès, M., Michaux, J., & 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. <https://doi.org/10.1002/wsb.283>
- von Thaden, A., Cocchiara, B., Jarausch, A., Jüngling, H., Karamanlidis, A. A., Tiesmeyer, A., Nowak, C., & Muñoz-Fuentes, V. (2017). Assessing SNP genotyping of noninvasively collected wildlife samples using microfluidic arrays. *Scientific Reports*, 7, 10768. <https://doi.org/10.1038/s41598-017-10647-w>
- 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. <https://doi.org/10.1111/j.1469-1795.2011.00503.x>
- 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. <https://doi.org/10.1016/j.anaerobe.2019.03.013>
- 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. <https://doi.org/10.1017/S0032247417000031>
- 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. <https://doi.org/10.1002/ece3.2030>

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Hayward, K. M., Clemente-Carvalho, R. B. G., Jensen, E. L., de Groot, P. V. C., Branigan, M., Dyck, M., Tschritter, C., Sun, Z., & Loughheed, S. C. (2022). Genotyping-in-thousands by sequencing (GT-seq) of noninvasive faecal and degraded samples: A new panel to enable ongoing monitoring of Canadian polar bear populations. *Molecular Ecology Resources*, 22, 1906–1918. <https://doi.org/10.1111/1755-0998.13583>