Conservation Genetics Resources
https://doi.org/10.1007/s12686-019-01082-2

METHODS AND RESOURCES ARTICLE

Improving our conservation genetic toolkit: ddRAD-seq for SNPs in snow leopards

Safia Janjua[1,2] · Jeffrey L. Peters[1] · Byron Weckworth[3] · Fakhar I. Abbas[2] · Volker Bahn[1] · Orjan Johansson[4,5] · Thomas P. Rooney[1]

Received: 29 November 2018 / Accepted: 6 January 2019

Abstract
Snow leopards (*Panthera uncia*) are an enigmatic, high-altitude species whose challenging habitat, low population densities and patchy distribution have presented challenges for scientists studying its biology, population structure, and genetics. Molecular scatology brings a new hope for conservation efforts by providing valuable insights about snow leopards, including their distribution, population densities, connectivity, habitat use, and population structure for assigning conservation units. However, traditional amplification of microsatellites from non-invasive sources of DNA are accompanied by significant genotyping errors due to low DNA yield and poor quality. These errors can lead to incorrect inferences in the number of individuals and estimates of genetic diversity. Next generation technologies have revolutionized the depth of information we can get from a species' genome. Here we used double digest restriction-site associated DNA sequencing (ddRAD-seq), a well-established technique for studying non-model organisms, to develop a reference sequence library for snow leopards using blood samples from five Mongolian individuals. Our final data set reveals 4504 loci with a median size range of 221 bp. We identified 697 SNPs and low nucleotide diversity (0.00032) within these loci. However, the probability that two random individuals will share identical genotypes is about 10^{-168}. We developed probes for DNA capture using this sequence library which can now be used for genotyping individuals from scat samples. Genetic data from ddRAD-seq will be invaluable for conducting population and landscape scale studies that can inform snow leopard conservation strategies.

Keywords Snow leopard · ddRAD-seq · Next generation sequencing · SNP discovery

Introduction

Despite being a high profile and charismatic carnivore, information on snow leopard ecology, population structure, and genetics lags behind most other large carnivores (Caragiulo et al. 2016 and Fox and Chundawat 2016). This lack of information is largely because of the animal's cryptic nature and remote habitat. Therefore, it is difficult to obtain sufficient data on numbers, locations of peripheral and core populations, and areas where the snow leopard populations are in decline. Such information is critical for the conservation of this apex predator.

Efforts are underway to determine the status of the species across its extensive (approximately 1.6 million km^2) but highly fragmented range (McCarthy et al. 2016). One such effort is the genomic analysis of DNA, extracted from non-invasively collected samples. This can benefit conservation efforts by providing information about population densities, connectivity, source/sink dynamics, and habitat use, among others, as well as how to define distinct conservation units upon which to focus specific conservation efforts.

Several studies have genotyped non-invasively collected snow leopard DNA samples (Waits et al. 2007; Janečka et al. 2008; Karmacharya et al. 2011; Aryal et al. 2014). However, a common problem associated with these studies is that the genetic markers used were either not snow leopard specific or had to be modified to improve their specificity.

✉ Safia Janjua
janjua.2@wright.edu

1 Department of Biological Sciences, Wright State University, Dayton, USA

2 Bioresource Research Centre, Islamabad, Pakistan

3 PANTHERA, New York, USA

4 Grimso Wildlife Research Station, Swedish University of Agricultural Sciences, Uppsala, Sweden

5 Snow Leopard Trust, Seattle, USA

Published online: 13 February 2019

The methods used are also challenged by genotyping errors (McKelvey and Schwartz 2004) due to the inherent low yield and poor quality of DNA from such samples. These errors can lead to incorrect inferences, including the misidentification of individuals (Waits and Paetkau 2005), which in turn can lead to incorrect estimates of population size and patterns of genetic diversity. Janecka et al. (2017) attempted to overcome these limitations by designing 33 snow leopard-specific microsatellite markers and using them to evaluate snow leopard populations across its range. This is the most detailed genetic study of this cat to date. However, it still has conventional limitations associated with microsatellites, such as amplification failure, allele dropouts, and the appearance of false alleles. Hence, the issues remained unresolved.

Today, cutting-edge technologies, like next generation sequencing (NGS), have revolutionized the depth of information we can get from a species' genome by providing sequence information from thousands of loci. Thus, NGS enables scientists to supplement and further refine existing research by thoroughly analyzing the genomes of organisms to better evaluate evolutionary patterns and signatures that can be beneficial for conservation efforts. Additionally, NGS amplifies short stretches of DNA, making it well-suited for analyzing non-invasive samples, where collected DNA tends to already be fragmented (Waits and Paetkau 2005).

Double digest restriction-site associated DNA sequencing (ddRAD-seq) is a well-established NGS technique used to study non-model organisms (Peterson et al. 2012). This method involves pseudorandom sampling of whole genomes of organisms (Miller et al. 2007; Baird et al. 2008; Lavretsky et al. 2015) and subsequent discovery of SNPs from sequenced genomes (Peterson et al. 2012). We developed ddRAD-seq libraries for snow leopard and identified a SNP panel for studying their population genetics. Given the need for high quality snow leopard genetic data and the limitations associated with existing methods, the primary objectives of this study were:

i. Genome-wide SNP discovery in snow leopards;
ii. Determine the utility of this SNP panel for identifying individuals; and
iii. Compare the resolution between SNPs and microsatellites.

Methods

Sampling and DNA extraction

We used five blood samples of wild Mongolian snow leopards, archived at the American Museum of Natural History, which were collected as part of another project (Johansson et al. 2013; sample IDs used in this study are: M1, M9, M10, F8, F9; details of samples can be found in the supplementary material). DNA from blood samples was extracted by using a DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA). The DNA extractions were quantified using a NanoDrop (Thermo Scientific).

ddRAD-seq library

The extracted DNA was used to generate ddRAD-seq libraries following DaCosta and Sorenson (2014). Briefly, 0.5–1 µg of genomic DNA was digested using 20 U of the restriction enzymes *Eco*RI and *Sbf*I, producing fragments with sticky ends. Uniquely barcoded adapters were ligated to the digestion products. DNA in the size range of 300–450 bp was excised from a 2% low melt agarose gel and purified using a MinElute Gel Extraction Kit (Qiagen, Valencia, CA). This size-selected DNA was then amplified using standard PCR with Phusion high fidelity DNA polymerase (Thermo Scientific, Pittsburgh, PA), and the products were purified using magnetic SPRI beads (Company). We used real-time PCR to quantify PCR products using an Illumina library quantification kit (KAPA Biosystems, Wilmington, MA) and an ABI 7900HT SDS (Applied Biosystems, Foster City, CA). Equimolar concentrations of libraries with unique barcodes were pooled and sequenced (single-end, 150 bp reads) on an Illumina HiSeq 2000 at Tufts University.

Data processing

DaCosta and Sorenson (2014) computational pipeline was used to process the raw data obtained from Illumina sequence reads (Python scripts available at http://github.com/BU-RAD-seq/ddRAD-seqPipeline). For each individual, identical reads were collapsed while retaining read counts and the highest quality score for each position. Individual reads that were > 10% divergent from all others (using the UCLUST function in USEARCH v. 5; Edgar 2010) and/or those with an average Phred score < 20 were removed. We then clustered the filtered reads from all individuals into putative loci using UCLUST (-id setting of 0.85). The highest quality read from each cluster (i.e., putative locus) was localized in the tiger (*Panthera tigris*) genome (GCA_000464555.1 PanTig1.0), using BLASTN v. 2 (Altschul et al. 1990). Clusters that did not generate a BLAST hit were carried through the pipeline as anonymous loci. After combining clusters that had the same BLAST hits, we aligned reads from each cluster using MUSCLE v. 3 (Edgar 2004).

The final alignments were used to genotype individuals at each locus using the RADGenotypes.py script. Individuals were scored as homozygous at a locus if ≥ 93% of reads were consistent with a single haplotype across polymorphic sites, and heterozygous if a second haplotype was represented by

at least 29% of reads (DaCosta and Sorenson 2014). We also scored individuals as heterozygous if a second allele was represented by as few as 20% of reads, but only if the second allele was known from other individuals in the sample. Individual genotypes that did not meet either of these criteria, or had evidence of more than two haplotypes, were flagged (0.001% of all genotypes in the final data set); for these samples, we retained only the allele represented by the majority of reads and scored the second allele as missing data. Similarly, the second allele was scored as missing for apparently homozygous genotypes based on 1–5 reads, which were considered "low depth" (1.1% of all genotypes). We retained for analysis all loci that had complete genotypes for at least four of our five individuals.

Estimates of genetic diversity

Standard estimates of genetic diversity were calculated using the R package—PopGenome (Pfeifer et al. 2014) and Structure 2.2.3 (Pritchard et al. 2000). Observed heterozygosity (H_{obs}), expected heterozygosity (H_e), nucleotide diversity, and inbreeding coefficient (F_{IS}) were calculated. Probability of Identity (PID), which is the probability that two random individuals will have identical genotypes, was calculated using Paetkau and Strobeck (1994) equation: $PID = \sum P_i^4 + \sum \left(2P_iP_j\right)^2$, where *Pi* is the frequency of the *i*th allele and *Pj* is the frequency of *j*th allele.

Comparison with microsatellites

The individual snow leopards used in this study were also genotyped by Caragiulo et al. (unpublished) using 12 microsatellite loci (Caragiulo et al. 2015). We calculated H_{obs}, H_e, F_{IS}, and PID values for the microsatellite data as we did for ddRAD-seq loci for comparisons.

Results and discussion

Given significant knowledge gaps on the genetics of snow leopards, and the importance of this information for conservation, we used ddRAD-seq for a high resolution and low cost development of a reference sequence library. This method is widely becoming an important component of ecological and evolutionary studies (Andrews et al. 2018; Ba et al. 2017; Peters et al. 2016), especially for organisms like snow leopards where little is known about their genome. We obtained an average of ~4,000,000 high quality sequence reads per individual. After assembling these reads, we retained 4504 loci that were recovered from a minimum of 80% of individuals and contained no flagged genotypes. Among these reads median fragment size was 221 bp. The final data set comprised 511 loci with one or more polymorphic sites plus 3993 constant loci and a total of 7,428,063 aligned nucleotides and 697 SNPs. Overall nucleotide diversity in the five Mongolian samples was 0.00032 with nucleotide diversity ranging from 0.00081 to 0.07935 among the variable loci.

To evaluate the utility of these SNPs, we estimated different descriptive statistics important in population genetics, comparing them between traditional microsatellites and our SNP panel. The average expected heterozygosity (He) was 0.042, which was slightly lower than the observed heterozygosity (Ho) of 0.047 for ddRAD-sEq. However, He and Ho did not differ significantly ($p = 0.07$). In contrast, He was slightly higher than Ho for microsatellites but the difference is statically insignificant ($p = 0.341$) (Table 1). Fixation index (F) for both microsatellites and ddRAD-seq was negative. However, these values were calculated for only five samples from Mongolia and are not necessarily representative of the total genetic diversity within the population.

To evaluate the power of ddRAD-seq loci for individual identification, we calculated PID. This is the probability that two individuals drawn at random from a population will have the same genotype at multiple loci (Waits et al. 2001, Valière 2002). The PID is widely used to assess the statistical confidence for individual identification, for non-invasive sampling (Reed et al. 1997; Kohn et al. 1999; Mills et al. 2000; Waits and Leberg 2000). It is therefore useful for estimating the number of individuals with higher confidence (Ernest et al. 2000). The value for PID for ddRAD-seq is very low (i.e. 1.55×10^{-168}) compared to the value for microsatellite loci (i.e. 2.35×10^{-7}), Therefore the power of ddRAD-seq loci to identify individuals is over 100 orders of magnitude higher than that of microsatellites.

As the role of genomic methods (e.g. SNPs) has evolved in conservation practice, the transition from, and juxtaposition with, traditional conservation genetics methods (Ouburg et al. 2010) has brought to light advantages and challenges (Allendorf et al. 2010). For snow leopards, the advantages are particularly critical. Microsatellites are notoriously difficult to standardize across different labs. In the case of non-invasive samples, there is the added challenge of genotyping

Table 1 The mean genetic diversity estimated in five snow leopards genotyped using 12 microsatellites and ddRAD-seq loci

	Microsatellites	ddRAD-seq loci
N	4.333	4.973
Na	2.917	1.118
Ho	0.536	0.047
He	0.544	0.042
F	− 0.023	− 0.113

N average sample size, *Na* average number of different alleles, *Ho* observed heterozygosity, *He* expected heterozygosity, *F* fixation index

error due to allelic drop-out and false alleles. Specifically in snow leopards, microsatellites appear to have generally low variability (Caragiulo et al. 2016). These challenges limit the utility of microsatellites as a universal maker to use across labs and research groups. In addition, an often overlooked obstacle is that given the species' legal status and the political sensitivities in many of its range countries, it is sometimes impossible to transport DNA samples across borders, further precluding work being done in a single, standardized lab. Instead, lab work must occur within the country where the samples were collected. A hypervariable SNP panel for snow leopards, such as done here, circumvents all of the challenges of microsatellites. The next step is to design probes from our ddRAD-seq libraries for a targeted DNA capture method that is well suited for non-invasive samples (Perry et al. 2010). This work is ongoing and will provide a crucial new tool for conducting conservation genetic research on this imperiled species.

Funding The funding was provided by Panthera.

References

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. Nat Rev Genet 11:697–709

Altschul S, Gish W, Miller W, Myers E, Lipman D (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Andrews KR, Adams JR, Cassirer EF, Plowright RK, Gardner C, Dwire M, Hohenlohe PA, Waits LP (2018) A bioinformatic pipeline for identifying informative SNP panels for parentage assignment from RAD seq data. Mol Ecol Resour. https://doi.org/10.1111/1755-0998.12910

Aryal A, Brunton D, Ji W, Karmacharya D, McCarthy T, Bencini R, Raubenheimer D (2014) Multipronged strategy including genetic analysis for assessing conservation options for the snow leopard in the central Himalaya. J Mammal 95(4):871–881

Ba H, Jia B, Wang G, Yang Y, Kedem G, Li C (2017) Genome-wide SNP discovery and analysis of genetic diversity in farmed sika deer (*Cervus nippon*) in northeast China using double-digest restriction site-associated DNA sequencing. G3: Genes Genomes Genet 7(9):3169–3176

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Johnson EA (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3(10):e3376

Caragiulo A, Kang Y, Rabinowitz S, Dias-Freedman I, Loss S, Zhou XW, Bao WD, Amato G (2015) Presence of the endangered amur tiger Panthera tigris altaica in Jilin Province, China, detected using non-invasive genetic techniques. Oryx 49(4):632–635

Caragiulo A, Amato G, Weckworth B (2016) Conservation genetics of snow leopards. In: Nyhus PJ, McCarthy T, Mallon D (eds) Snow leopards—biodiversity of the world: conservation from genes to landscapes. Elsevier Inc., London, pp 368–371

DaCosta JM, Sorenson MD (2014) Amplification biases and consistent recovery of loci in a double-digest RAD-seq protocol. PLoS ONE 9(9):e106713

Edgar R (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461

Ernest H, Penedo M, May B, Syvanen M, Boyce W (2000) Molecular tracking of mountain lions in the Yosemite Valley region in California: genetic analysis using microsatellites and faecal DNA. Mol Ecol 9:433–442

Fox JL, Chundawat RS (2016) What is a Snow Leopard? Behavior and Ecology. In: Nyhus PJ, McCarthy T, Mallon D (eds) Snow Leopards – Biodiversity of the world: conservation from genes to landscapes. Elsevier Inc., London, pp 13–21

Janečka JE, Jackson R, Yuquang Z, Diqiang L, Munkhtsog B, Buckley-Beason V, Murphy WJ (2008) Population monitoring of snow leopards using noninvasive collection of scat samples: a pilot study. Anim Conserv 11(5):401–411

Janecka JE, Zhang Y, Li D, Munkhtsog B, Bayaraa M, Galsandorj N, Wangchuk TR, Karmacharya D, Li J, Lu Z, Uulu KZ, Gaur A, Kumar S, Kumar K, Hussain S, Muhammad G, Jevit M, Hacker C, Burger P, Wultsch C, Janecka MJ, Helgen K, Murphy WJ, Jackson R (2017) Range-wide snow leopard phylogeography supports three subspecies. J Hered 108(6):597–607

Johansson Ö, Malmsten J, Mishra C, Lkhagvajav P, McCarthy T (2013) Reversible immobilization of free-ranging snow leopards (*Panthera uncia*) with a combination of medetomidine and tiletamine-zolazepam. J Wildl Dis 49(2):338–346

Karmacharya DB, Thapa K, Shrestha R, Dhakal M, Janecka JE (2011) Noninvasive genetic population survey of snow leopards (Panthera uncia) in Kangchenjunga conservation area, Shey Phoksundo National Park and surrounding buffer zones of Nepal. BMC Res Notes 4:516

Kohn MH, York EC, Kamradt DA, Haught G, Sauvajot RM, Wayne RK (1999) Estimating population size by genotyping faeces. Proc R Soc Lond Ser B 266:1–7

Lavretsky P, Dacosta JM, Hernández-Baños BE, Engilis A, Sorenson MD, Peters JL (2015) Speciation genomics and a role for the Z chromosome in the early stages of divergence between Mexican ducks and mallards. Mol Ecol 24(21):5364–5378

McCarthy T, Mallon D, Sanderson EW, Zahler P, Fisher K (2016) Biogeography and status overview. In: Nyhus PJ, McCarthy T, Mallon D (eds) Snow leopards. Biodiversity of the world: conservation from genes to landscapes. Elsevier, London, pp 23–42

McKelvey KS, Schwartz MK (2004) Genetic errors associated with population estimation using non-invasive molecular tagging: problems and new solutions. J Wildl Manag 68(3):439–448

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res 17:240–248

Mills LS, Citta JJ, Lair KP, Schwartz MK, Tallmon DA (2000) Estimating animal abundance using noninvasive DNA sampling: promise and pitfalls. Ecol Appl 10(1):283–294

Ouborg NJ, Pertoldi C, Loeschcke V, Bijlsma RK, Hedrick PW (2010) Conservation genetics in transition to conservation genomics. Trends Genet 26:177–187

Paetkau D, Strobeck C (1994) Microsatellite analysis of genetic variation in black bear populations. Mol Ecol 3:489–495

Paetkau D, Waits LP, Clarkson PL, Craighead L, Vyse E, Ward R, Strobeck C (1998) Variation in genetic diversity across the range of North American brown bears. Conserv Biol 12:418–429

Perry GH, Marioni JC, Melsted P, Gilad Y (2010) Genomic-scale capture and sequencing of endogenous DNA from feces. Mol Ecol 19(24):5332–5344

Peters JL, Lavretsky P, DaCosta JM, Bielefeld RR, Feddersen JC, Sorenson MD (2016) Population genomic data delineate conservation units in mottled ducks (*Anas fulvigula*). Biol Conserv 203:272–281

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model

species. PLoS ONE 7(5):e37135. https://doi.org/10.1371/journal.pone.0037135
Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ (2014) PopGenome: an efficient Swiss army knife for population genomic analyses in R. Mol Biol Evol 31(7):1929–1936
Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959
Reed JZ, Tollit D, Thompson P, Amos W (1997) Molecular scatology: the use of molecular genetic analysis to assign species, sex, and individual identity to seal faeces. Mol Ecol 6:225–234
Valière N (2002) GIMLET: a computer program for analysing genetic individual identification data. Mol Ecol Notes 2:377–379
Waits J, Leberg P (2000) Biases associated with population estimation using molecular tagging. Anim Conserv 3:191–199
Waits LP, Paetkau D (2005) Noninvasive genetic sampling tools for wildlife biologists: a review of applications and recommendations for accurate data collection. J Wildl Manag 69(4):1419–1433
Waits LP, Luikart G, Taberlet P (2001) Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. Mol ecol 10(1):249–256
Waits LP, BUCKLEY-BEASON VA, Johnson WE, Onorato D, McCarthy TOM (2007) A select panel of polymorphic microsatellite loci for individual identification of snow leopards (Panthera uncia). Mol Ecol Notes 7(2):311–314