

Articles

A New Panel of SNP Markers for the Individual Identification of North American Pumas

Robert R. Fitak,* Ashwin Naidu, Ron W. Thompson, Melanie Culver

R.R. Fitak

Department of Biology, Duke University, Durham, North Carolina 27708

A. Naidu

School of Natural Resources and the Environment, University of Arizona, Tucson, Arizona 85721

R.W. Thompson

Primero Conservation, Box 16106, Portal, Arizona 85632

M. Culver

Arizona Cooperative Fish and Wildlife Research Unit, U.S. Geological Survey, School of Natural Resources and the Environment, University of Arizona, Tucson, Arizona 85721

Abstract

Pumas *Puma concolor* are one of the most studied terrestrial carnivores because of their widespread distribution, substantial ecological impacts, and conflicts with humans. Over the past decade, managing pumas has involved extensive efforts including the use of genetic methods. Microsatellites have been the most commonly used genetic markers; however, technical artifacts and little overlap of frequently used loci render large-scale comparison of puma genetic data across studies challenging. Therefore, a panel of genetic markers that can produce consistent genotypes across studies without the need for extensive calibrations is essential for range-wide genetic management of puma populations. Here, we describe the development of PumaPlex, a high-throughput assay to genotype 25 single nucleotide polymorphisms in pumas. We validated PumaPlex in 748 North American pumas *Puma concolor cougar*, and demonstrated its ability to generate reproducible genotypes and accurately identify individuals. Furthermore, in a test using fecal deoxyribonucleic acid (DNA) samples, we found that PumaPlex produced significantly more genotypes with fewer errors than 12 microsatellite loci, 8 of which are commonly used. Our results demonstrate that PumaPlex is a valuable tool for the genetic monitoring and management of North American puma populations. Given the analytical simplicity, reproducibility, and high-throughput capability of single nucleotide polymorphisms, PumaPlex provides a standard panel of markers that promotes the comparison of genotypes across studies and independent of the genotyping technology used.

Keywords: conservation; cougar; fecal DNA; mountain lion; next-generation sequencing; noninvasive genetics; PumaPlex

Received: November 17, 2014; Accepted November 13, 2015; Published Online Early: November 2015; Published: June 2016

Citation: Fitak RR, Naidu A, Thompson RW, Culver M. 2015. A new panel of SNP markers for the individual identification of North American pumas. *Journal of Fish and Wildlife Management* 7(1):13–27; e1944-687X. doi: 10.3996/112014-JFWM-080

Copyright: All material appearing in the *Journal of Fish and Wildlife Management* is in the public domain and may be reproduced or copied without permission unless specifically noted with the copyright symbol ©. Citation of the source, as given above, is requested.

The findings and conclusions in this article are those of the author(s) and do not necessarily represent the views of the U.S. Fish and Wildlife Service.

* Corresponding author: robert.fitak@duke.edu

Introduction

Pumas *Puma concolor* are the most widespread terrestrial carnivore in the western hemisphere, ranging from Alaska and northern Canada to the southernmost extent of Argentina and Chile (Sunquist and Sunquist 2002). Regarded as one of the most iconic and well-studied species in the Americas, pumas often attract large amounts of attention as a game species (Logan and Sweanor 2001), for preying upon (at times endangered) wildlife (Turner et al. 1992; Wehausen 1996; Hayes et al. 2000; Schaefer et al. 2000), and for conflicts with humans (Beier 1991; Torres et al. 1996). They are frequently the center of conservation programs targeting entire landscapes (Weber and Rabinowitz 1996) because the removal of pumas, like other large carnivores, can have substantial impacts on the dynamics of both plant and animal communities (Logan and Irwin 1985; Turner et al. 1992; Schmitz et al. 2000; Ripple et al. 2014). In response to the significant ecological impacts of pumas, many wildlife agencies are actively researching and monitoring puma populations to make informed decisions regarding their conservation and management (Culver and Schwartz 2011).

As part of effective wildlife conservation and management programs, genetic studies are often initiated to investigate many aspects of a species' ecology (Frankham 2003). In pumas, genetic studies have concentrated either on the examination of gene flow on continental (Culver et al. 2000) and local scales (e.g., Loxterman [2011]; Andreassen et al. [2012]; Balkenhol et al. [2014]), or on the identification of individuals (Haag et al. 2009; Miotto et al. 2011; Naidu et al. 2011). Microsatellites have been the genetic marker of choice, and numerous loci derived from either domestic cats *Felis catus* (Menotti-Raymond and O'Brien 1995; Menotti-Raymond et al. 1997, 1999) or directly from pumas (Kurushima et al. 2006; Rodzen et al. 2007) have been developed and used in many studies (Figure 1). However, microsatellite allele sizes can vary among studies as a result of differences in the *Taq* polymerase, fluorescent dye, genotyping instrument, and analytical software used (Hahn et al. 2001; Vignal et al. 2002). This size variation renders comparing microsatellite genotypes across laboratories challenging and depositing genotypes into databases difficult without calibration of positive control samples. Furthermore, the loci analyzed to date vary across studies, often with little overlap. For example, across 17 genetic studies of western North American pumas published since 2000, 56 total different microsatellite loci were used (Figure 1). Many loci were used once ($n = 25$), whereas 9 loci were used in at least half the studies and no locus was shared across all studies. As a result of the technical difficulties in comparing microsatellite genotypes and little overlap of loci, genetic data from pumas have often been obtained from smaller geographic regions and rarely integrated into studies across larger landscapes (see references in Figure 1). Because pumas are capable of traversing long distances ($>1,000$ km; e.g., Thompson and Jenks 2005; Stoner et al. 2008), a consistent panel of genetic markers that can be analyzed in a high-

throughput fashion and compared across laboratories without the need for calibrating genotypes will facilitate the long-term and range-wide genetic monitoring of North American puma populations.

Recently, another genetic marker, called a single nucleotide polymorphism (SNP), has become increasingly popular in population genetic studies. A SNP is a single base in the DNA sequence that contains multiple variants (alleles) segregating in the population. There are several features of SNPs that make them effective markers for population genetic studies (see reviews by Vignal et al. 2002; Morin et al. 2004; Garvin et al. 2010). Most importantly, SNPs are numerous, broadly distributed across various genomic regions (coding, noncoding, organellar), and amenable to high-throughput technologies (Garvin et al. 2010). Alleles are encoded simply as bases (i.e., A, C, T, G) rather than amplicon lengths, and are thus independent of the genotyping technology used. Because SNP loci are generally biallelic and have a well-defined mutation model (Garvin et al. 2010), the mathematical descriptions of many parameters using them are simpler than multiallelic loci (e.g., microsatellites). However, the biallelic nature of SNPs often reduces the information content per locus. As a result, more SNPs (2–6 \times) are theoretically required to match the resolution of microsatellites (Morin et al. 2004), although several empirical studies have demonstrated that $<1\text{--}3\times$ as many SNPs can perform equally as well (Seddon et al. 2005; Ryyänen et al. 2007; Coates et al. 2009; Fernández et al. 2013). This number can vary, however, depending on a variety of factors, including the parameter in question, per-locus information content, levels of genetic variation, absolute number of markers, and sample size. Furthermore, because SNPs can be genotyped using short fragments of DNA (often <100 base-pairs [bp]), it has been suggested that they may outperform microsatellites in amplification success from noninvasively collected samples (Morin et al. 2004; Seddon et al. 2005; Morin and McCarthy 2007; Ogden 2011).

To date, the use of SNPs in wildlife conservation and management programs remains relatively uncommon and has often been limited to taxa closely related to species with available genome sequences (e.g., Pertoldi et al. 2010; Cronin et al. 2014; Kraus et al. 2015). There currently exists no set of SNPs specific for use in pumas, but advances in next-generation sequencing technologies have made the process of SNP discovery much more tractable in nonmodel species (Garvin et al. 2010). Here we use Roche 454 pyrosequencing of cDNA prepared from whole blood to assemble the first transcriptome of pumas. We identify SNPs from the transcriptome assembly and use them to develop a high-throughput, 25-SNP genotyping assay for North American pumas *Puma concolor cougar*. We demonstrate the utility of our assay in 748 samples collected throughout their range and its ability to accurately identify individuals. Furthermore, we compare genotyping success of our SNP assay with that of 12 microsatellites, 8 of which are commonly used in studies of North American pumas, in 46 fecal DNA samples. Finally, we discuss the relevance

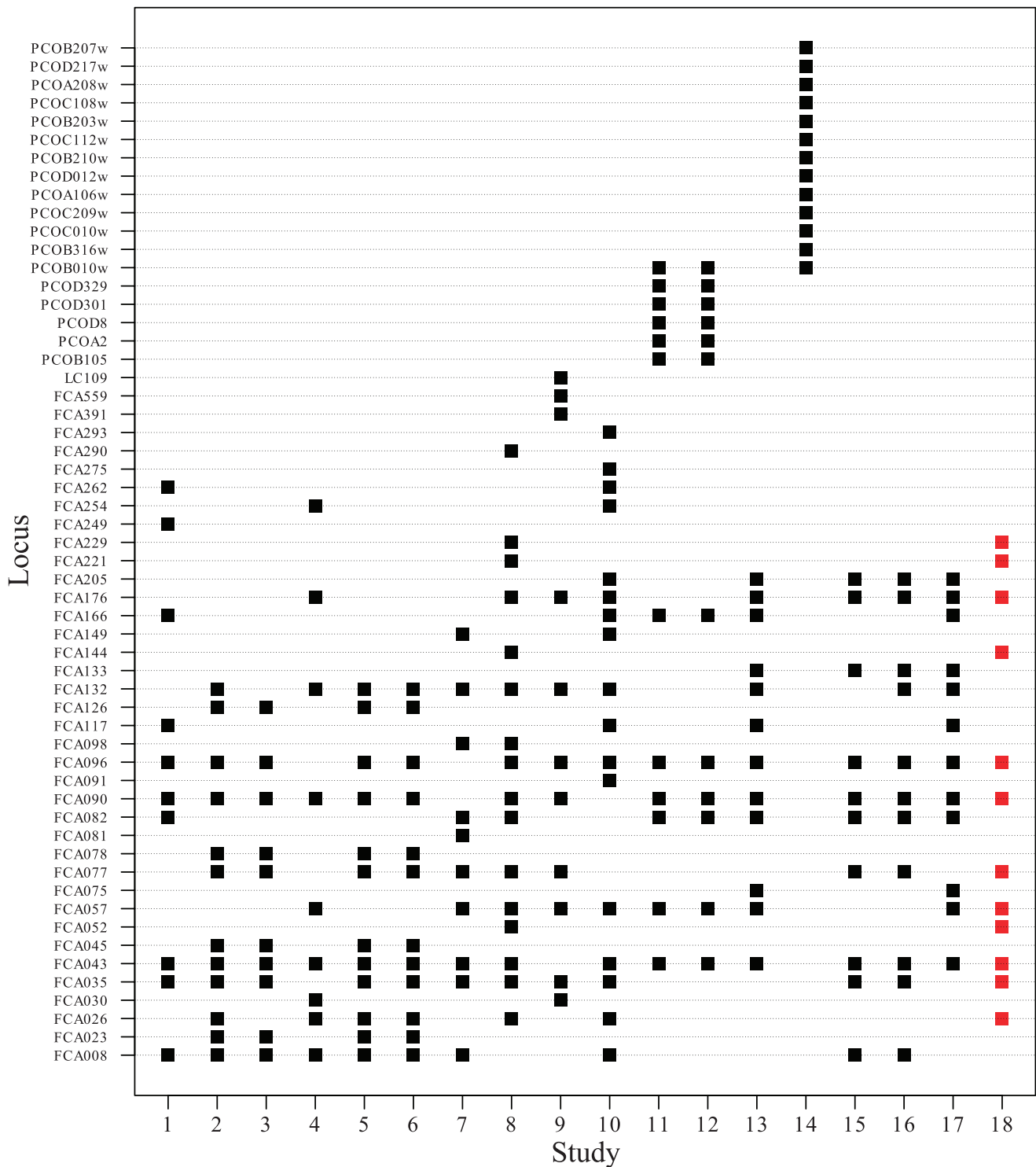


Figure 1. The microsatellite loci used in 17 genetic studies of western North American pumas *Puma concolor cougar* between 2000 and 2014 (studies 1–17, black squares) and this study (study 18, red squares). The 17 studies are as follows: 1—Culver et al. (2000), 2—Ernest et al. (2000), 3—Walker et al. (2000), 4—Sinclair et al. (2001), 5—Ernest et al. (2002), 6—Ernest et al. (2003), 7—Anderson et al. (2004), 8—McRae et al. (2005), 9—Biek et al. (2006), 10—Loxterman (2011), 11—Naidu et al. (2011), 12—Nicholson et al. (2011), 13—Onorato et al. (2011), 14—Andreasen et al. (2012), 15—Holbrook et al. (2012b), 16—Holbrook et al. (2012a), 17—Balkenhol et al. (2014).

of our assay as an effective tool for monitoring and managing individual pumas.

Methods

Transcriptome assembly and annotation

We collected whole blood samples from 12 pumas of wild origin via saphenous venipuncture. All pumas were from different locations in Arizona, with the exception of one puma from Bosque del Apache National Wildlife Refuge, New Mexico. Local agency personnel captured wild individual pumas under their live-capture permits. For each individual we filled two ribonucleic acid (RNA) Protect Animal Blood Tubes (Qiagen, Germantown, MD) with approximately 500 μ L of whole blood each and immediately placed the tubes in liquid nitrogen, or on ice for <24 h and then in liquid nitrogen, until RNA extraction was performed. Remaining whole blood was stored at -20°C for DNA extraction. We extracted total RNA using the RNA Protect Animal Blood System (Qiagen) according to the manufacturer's recommendations, and combined the two extracted RNA samples per individual for the final sample.

We quantified total RNA in each sample using a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA) and combined the 12 RNA samples in equimolar ratios to obtain a pooled RNA sample. We synthesized cDNA using the SMARTer Pico polymerase chain reaction (PCR) cDNA Synthesis Kit (Clontech Laboratories, Mountain View, CA) following the manufacturer's recommendations. This kit targets polyadenylated RNA and minimizes contamination from ribosomal and other small RNAs. To avoid biases in cDNA library construction and amplification, we prepared six independent cDNA libraries as described above, quantified each on a 2100 Bioanalyzer, and pooled them in equimolar concentrations for the final cDNA library. We sequenced the cDNA library on half of a picotiter plate on a GS-FLX Titanium platform (454 Life Sciences, Branford, CT) at the University of Arizona Genetics Core (<http://uagc.arl.arizona.edu/>).

We assembled all the sequences into contigs using GSASSEMBLER (454 Life Sciences), employing parameters to screen reads against NCBI's vector-contaminant database, to trim the reads of the SMARTer oligonucleotide primer used in cDNA library preparation, and to remove contigs <100 bp in length. The GSASSEMBLER software also quality-trimmed reads according to default criteria (3' ends trimmed to approx. $Q \geq 20$). We functionally annotated the contigs using FASTANNOTATOR (Chen et al. 2012), which integrates BLAST2GO (Conesa et al. 2005), PRIAM (Claudel-Renard et al. 2003), and RPSBLAST (<http://blast.ncbi.nlm.nih.gov>) to assign Gene Ontology (GO) terms, protein domains, and Enzyme Commission numbers to contigs.

PumaPlex development and genotyping

Next, we cleaned the raw sequence reads by removing the cDNA library primer, trimming the 3' ends of each

read to a minimum phred-scaled base quality score of 20, and filtering reads shorter than 100 bp using CUTADAPT v1.0 (Martin 2011). We mapped the cleaned reads to the assembly using BWA v0.6.2 (Li and Durbin 2009). We identified and scored polymorphisms using FREEBAYES v0.9.4 (Garrison and Marth 2012) and filtered the resulting polymorphisms to include only biallelic SNPs with a phred-scaled quality score ≥ 20 and a minimum allele count ≥ 2 . We performed a reciprocal BLAST (<http://blast.ncbi.nlm.nih.gov>) search to exclude potential false-positive SNPs as a result of the alignment of sequence paralogs. Briefly, we selected a fragment spanning approximately 100 bp on either side of each SNP and performed a nucleotide BLAST search against the RefSeq database. We omitted a SNP from further analyses if its fragment had multiple matches within the same genome with an e-value $< 1 \times 10^{-30}$ and similarity $> 90\%$. Finally, we removed all SNPs whose sequence had a best match to mitochondrial or ribosomal RNA genes.

In addition to the 12 samples described above, we collected 520 tissue samples from hunter-harvested pumas as part of a monitoring program by the Arizona Game and Fish Department (AZ; $n = 532$; Figure 2). We also used tissue samples previously collected by Culver et al. (2000), which represent a majority of the range of *P. c. cougar* (NAM; $n = 216$; Figure 2). We extracted DNA using a BioSprint 96 automated system (Qiagen) with Dynabeads magnetic bead technology (Invitrogen, Grand Island, NY) and quantified samples using the Quant-iT PicoGreen reagent (Invitrogen).

Using the candidate SNPs identified above, we designed four primer multiplexes for simultaneous analyses of 28, 29, 36, and 36 ($n = 129$ total) SNPs using the Assay Design v3.1 (Sequenom, San Diego, CA). This software designs PCR primers and a single, internal "extend primer" for each locus. Next, the software combines the loci into a multiplex that minimizes interactions between oligonucleotides and maximizes their amplification and genotyping potential following the instrument's requirements. We screened each multiplex for polymorphic SNPs in a subset of DNA samples ($n = 191$) on the MassARRAY system (Sequenom) at the University of Arizona Genetics Core. The 191 samples included the 12 pumas used to construct the cDNA library and 179 of the AZ pumas. We developed a final multiplex, PumaPlex 1.0 (hereafter referred to as PumaPlex), from the polymorphic SNPs using the Assay Design v3.1 as described above. Using PumaPlex, we genotyped the remaining samples ($n = 748$ total). All PCR reagents and cycling conditions for the PumaPlex runs can be found in Table S1 (*Supplemental Material*). We replicated 24 samples twice within and between the different Sequenom analyses and included negative controls for each assay. The replicate with the least number of called genotypes was removed from subsequent analyses. We annotated the SNPs on PumaPlex using a nucleotide BLAST search and assigned GO terms



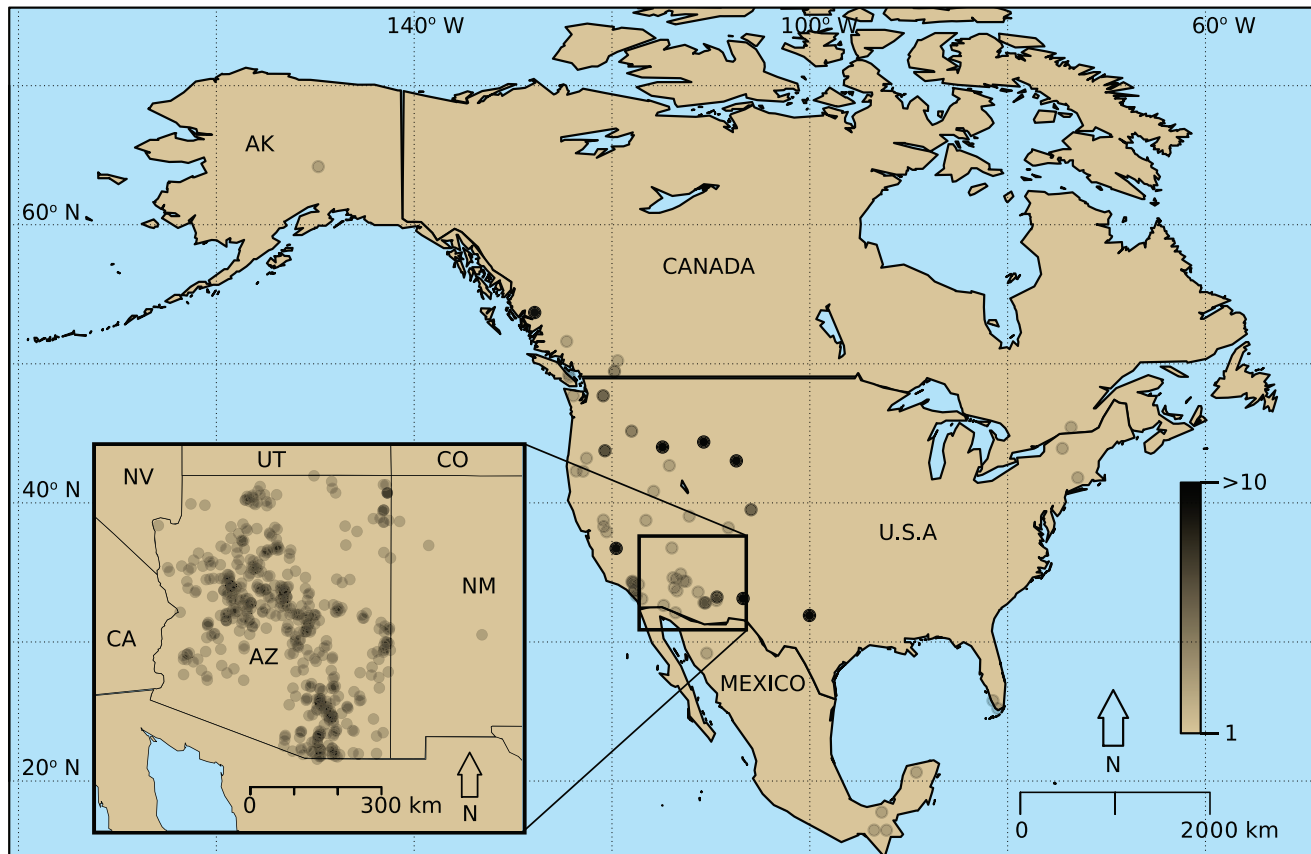


Figure 2. Map of the locations of the pumas *Puma concolor cougar* genotyped in this study. The larger image represents those samples previously collected by Culver et al. (2000; NAM) from 1983 to 1995, whereas the inset represents those collected during this study (AZ) from 2007 to 2013. The sampling density for both images is indicated by increasing opacity of black points according to the scale given.

using the genomic resources available for the domestic cat (www.ensembl.org).

To further validate genotypes, we selected 8 SNP loci at random for verification using traditional PCR and Sanger sequencing in the 12 original puma samples. For each SNP we designed primers using PRIMER3 (Rozen and Skaletsky 2000) to target a fragment between 150 and 250 bp in length and to include ≥ 50 bp of flanking sequence on either side of the SNP. We amplified each SNP using approximately 20 ng of template DNA, 0.5 μ M each of forward and reverse primer described above, 1 \times PCR buffer (USB, Cleveland, OH), 1.5 mM MgCl₂, 0.25 mM each dNTP, 0.05% BSA, and 0.5 U *Taq* DNA polymerase (USB) in a final volume of 20 μ L. Cycling conditions consisted of an initial melting step of 5 min at 95°C, then 40 cycles of 95°C for 45 s, 52–62°C for 45 s, and 72°C for 1 min, with a final extension step at 72°C for 7 min. We purified all PCR products using the ExoSAP-IT PCR Cleanup Kit (USB) according to manufacturer's specifications. The PCR products were sequenced on an ABI3730 DNA Sequencer (Applied Biosystems, Foster City, CA) at the University of Arizona Genetics Core in both forward and reverse directions. We visually inspected the sequences for errors and assembled them using Sequencher v5.1 (Gene Codes, Ann Arbor, MI). We manually identified

genotypes at each SNP locus after visual inspection of the chromatograms.

Using the results from PumaPlex, we removed individuals missing $>20\%$ of genotypes ($n = 21$; see Results). For each locus we calculated summary statistics in GENALEX v6.5 (Peakall and Smouse 2012), including allele frequencies, observed heterozygosity, and expected heterozygosity. Using only AZ pumas, we tested each locus for significant deviations from Hardy–Weinberg equilibrium with a chi-square test implemented in GENALEX and screened for the presence of null alleles following the procedure described by Girard (2011) using the median F_{IS} (calculated in GENALEX) and 100,000 simulations. This method is effective at detecting locus-specific deviations from neutrality caused by null alleles in both panmictic and inbred populations using any codominant marker. We calculated the observed, theoretical, and sibling probability of identity (P_{ID}) according to Waits et al. (2001) also in GENALEX.

Genotyping comparison in feces

We genotyped 46 DNA samples extracted from feces previously collected in southwestern AZ and identified as puma in origin (Naidu et al. 2011, 2014a), using both

Table 1. Summary statistics of the sequence reads before (Raw Reads) and after (Cleaned Reads) quality trimming and of the assembled transcriptome (Assembly) prepared from a pooled cDNA library of 12 pumas *Puma concolor couguar* collected in Arizona and New Mexico, USA, between 2009 and 2010. bp is base-pair.

Statistic	Raw reads	Cleaned reads	Assembly
No. of sequences	358,049	268,744	2,109
Total no. of bases	182,476,147 bp	92,153,346 bp	1,180,385 bp
Mean length (SD ^a)	509.6 (135.2) bp	342.9 (106.0) bp	560.0 (386.8) bp
GC content	50.6%	53.3%	47.9%
N50 ^b	522 bp	395 bp	671 bp
N50 count ^c	150,362	105,010	554

^a standard deviation.

^b N50 = 50% of bases are found in sequences (or contigs) of this length or longer.

^c The no. of sequences (or contigs) greater than or equal to the N50.

PumaPlex and 12 microsatellite loci: FCA026, FCA035, FCA043, FCA052, FCA057, FCA077, FCA090, FCA096, FCA144, FCA176, FCA221, FCA229 (Menotti-Raymond et al. 1999). The loci, eight of which are commonly used in studies of North American pumas, were selected specifically for their high polymorphism and previous use in pumas from the same region (McRae et al. 2005; Naidu et al. 2011; Figure 1). We followed PCR conditions for microsatellite loci as reported by Menotti-Raymond et al. (1999) and fragment analysis as described in Naidu et al. (2011). We replicated genotypes three times for both marker sets. We calculated the overall proportion of called genotypes as the number of called genotypes divided by the product of the total number of reactions performed ($3 \times 25 \times 46 = 3,450$ for SNPs and $3 \times 12 \times 46 = 1,656$ for microsatellites). We then constructed consensus genotypes ($25 \times 46 = 1,150$ for SNPs and $12 \times 46 = 552$ for microsatellites) and estimated the frequencies of false alleles (genotyping errors between replicates) and allelic dropout in GIMLET v1.3.3 (Valière 2002) using two degrees of stringency: 1) at least two replicates must produce a genotype at a given locus, and 2) all three replicates must produce a genotype at a given locus. We compared results between the SNP and microsatellite data sets using a Wilcoxon signed-rank test in R v3.0.0 (www.r-project.org).

Data archiving and accessibility

In fulfillment of data archiving guidelines (Wenburg 2011; Whitlock 2011), we deposited raw sequence data in GenBank (Data A1, *Archived Material*, <http://www.ncbi.nlm.nih.gov/sra/SRX633288>) and made the transcriptome assembly, the raw SNPs identified, and all individual genotypes accessible in Data A2 (*Archived Material*, <http://doi.pangaea.de/10.1594/PANGAEA.835154>). We have also made the first SNP data from PumaPlex available as a spatially intuitive genetic database for pumas called the Puma Genetic Database (Naidu et al. 2014b), hosted on Environmental Systems Research Institute's ArcGIS Online through the University of Arizona (<http://www.arcgis.com/home/item.html?id=4d9e04e504bb453691fbff736df49b3b>). This online database links each individual sample with SNP genotypes and a variety of metadata including collection

date, sample type, Global Positioning System location, etc., and is displayed as an interactive map. Users can filter for criteria of interest and subsequently download the respective information.

Results

Transcriptome assembly and annotation

Our cDNA library sequencing produced 358,049 sequence reads with a mean length of 509.6 (SD = 135.2) bp (Table 1). We assembled the reads into 2,109 contigs with an N50 of 671 bp (N50 indicates 50% of all bases were found in contigs of this length or larger; Table 1). We assigned 12,231 GO terms to 1,244 contigs (59%) and were unable to assign a functional annotation to 865 (41%) of contigs (Table S2, *Supplemental Material*). The GO term assigned most often in the category of Biological Process was “oxygen transport” ($n = 70$), and the other commonly assigned terms included a variety of housekeeping and immune-related processes (Figure S1, *Supplemental Material*). Genes that localize to the “cytosol” or are “integral to the plasma membrane” were most often represented in the category of Cellular Component (Figure S2, *Supplemental Material*), and genes involved in “protein binding” were the most frequent Molecular Function (Figure S3, *Supplemental Material*). We ascribed 73 Enzyme Commission codes to 64 contigs and annotated 707 contigs with at least one protein domain (Table S2, *Supplemental Material*).

PumaPlex development and genotyping

After cleaning and quality control, 268,744 (75.1%) sequence reads and approximately 92 million bp (50.5%) remained (Table 1). We were able to accurately map 34,675 (12.9%) of the cleaned reads to our transcriptome assembly. We identified 434 candidate SNPs that passed our quality criteria. We screened 129 of the candidate SNPs using four different Sequenom multiplexes and verified 30 (23.3%) as polymorphic. We were able to incorporate 25 of the verified SNPs into PumaPlex (the information necessary to construct PumaPlex using Sequenom or other methods can be found in Table S3 and Data S1, *Supplemental Material*). We annotated each of the 25 SNPs using either the domestic cat *Felis catus* or

Table 2. Description and annotation of the 25 single nucleotide polymorphisms (SNPs) on PumaPlex. These 25 SNPs were developed from a transcriptome sequenced from a pooled cDNA library of 12 pumas *Puma concolor couguar* collected in Arizona and New Mexico, USA, between 2009 and 2010. All SNPs were further validated by genotyping 748 pumas collected in North America from 1983 to 2013. Accession numbers and organisms represent the most similar sequence found in GenBank (<http://ncbi.nlm.nih.gov>). For a list of all the contigs or the 25 sequences represented on PumaPlex, see data available in Data A2 (Archived Material, <http://doi.pangaea.de/10.1594/PANGAEA.835154>) or Data S1 (Supplemental Material), respectively.

SNP name	Contig	Position	Alleles	Gene name	Accession no.	Organism	SNP type ^a (codon position)
PP01	Contig01593	378	A/G	FAS	XM_007079290	<i>Panthera tigris</i>	3' UTR
PP02	Contig01979	320	C/T	ZYX	XM_006929460	<i>Felis catus</i>	3' UTR
PP03	Contig00248	320	C/T	LOC101098507	XM_006941899	<i>Felis catus</i>	Silent (3 rd)
PP04	Contig00368	499	A/T	PPBP	XM_007086942	<i>Panthera tigris</i>	3'-UTR
PP05	Contig00515	508	A/G	TKT	XM_006928861	<i>Felis catus</i>	Silent (3 rd)
PP06	Contig00520	925	A/G	AP2M1	XM_007083632	<i>Panthera tigris</i>	3'-UTR
PP07	Contig00540	285	C/T	MSRB1	XM_007098707	<i>Panthera tigris</i>	3' UTR
PP08	Contig00891	1,341	C/T	VIM	XM_003988131	<i>Felis catus</i>	Silent (3 rd)
PP09	Contig00907	709	A/G	TAL1	XM_006934722	<i>Felis catus</i>	3' UTR
PP10	Contig00915	1,070	C/T	SDCBP	XM_007084815	<i>Panthera tigris</i>	3' UTR
PP11	Contig00918	673	A/G	PRDX6	XM_007084027	<i>Panthera tigris</i>	3' UTR
PP12	Contig00956	289	A/G	PDLIM1	XM_006938044	<i>Felis catus</i>	3' UTR
PP13	Contig00960	839	A/G	FAM213A	XM_007095603	<i>Panthera tigris</i>	3' UTR
PP14	Contig01001	561	C/T	MARCH2	XM_003981780	<i>Felis catus</i>	Silent (3 rd)
PP15	Contig01008	815	A/C	ARHGDI1	XM_007082762	<i>Panthera tigris</i>	3' UTR
PP16	Contig01059	462	C/T	IFI16	XM_007084298	<i>Panthera tigris</i>	Missense (2 nd), V/A
PP17	Contig01090	445	A/C	MGAT1	XM_007083048	<i>Panthera tigris</i>	3' UTR
PP18	Contig01122	392	C/T	LOC101085837	XM_003995040	<i>Felis catus</i>	Silent (3 rd)
PP19	Contig01160	468	A/G	FHL1	XM_004000937	<i>Felis catus</i>	3' UTR
PP20	Contig01166	434	C/T	PSMB8	XM_007093253	<i>Panthera tigris</i>	Silent (3 rd)
PP21	Contig01239	372	C/T	PTPRC	XM_007096096	<i>Panthera tigris</i>	Silent (3 rd)
PP22	Contig01365	537	A/G	COPE	XM_007088371	<i>Panthera tigris</i>	3' UTR
PP23	Contig01403	257	A/G	CCDC109B	XM_006930919	<i>Felis catus</i>	Missense (1 st), M/V
PP24	Contig01405	397	A/G	LAPTM5	XM_007092383	<i>Panthera tigris</i>	3' UTR
PP25	Contig00453	772	C/T	ITGB3	XM_003997035	<i>Felis catus</i>	3' UTR

^a For missense SNPs, the two amino acids are shown using their International Union of Pure and Applied Chemistry abbreviation; 3' UTR = untranslated region downstream of the stop codon.

tiger *Panthera tigris* genomes, of which 16 SNPs were found in 3' untranslated regions, 7 were coding but silent (synonymous) changes, and only 2 were missense (nonsynonymous) changes (Table 2). A functional description of each gene represented on PumaPlex and their associated GO terms is available in Table S4 (Supplemental Material).

We genotyped 772 total samples (748 pumas + 24 replicated samples) and our overall genotyping rate was 95.2%. Across 24 replicated samples, we found 100% agreement between called genotypes. We omitted 14 AZ and 7 NAM individuals missing genotypes at >20% of loci. Our final genotyping rate was 97.4% in AZ and 95.3% in NAM. In 12 individuals tested at 8 loci, we found 100% agreement between Sequenom- and Sanger-called genotypes. In AZ, minimum allele frequency ranged between 0.069 and 0.490, observed heterozygosity between 0.066 and 0.534, and expected heterozygosity between 0.128 and 0.500 (Table 3). Only three loci (PP05, PP08, and PP19) showed significant deviations from Hardy-Weinberg equilibrium after Bonferroni correction ($P < 0.05$; Table 3). Using the median F_{IS} of 0.061 we observed no evidence for the presence of null alleles (Figure 3). The theoretical P_{ID} in AZ and NAM was 7.7×10^{-9} and 1.6×10^{-7} , respectively, and no genotype

matches (observed P_{ID}) occurred when >17 or >22 loci, respectively, were included (Figure 4). Sibling P_{ID} was approximately one order of magnitude greater than the observed P_{ID} in both AZ and NAM at 7.1×10^{-5} and 3.6×10^{-4} , respectively (Figure 4).

Genotyping comparison in feces

We genotyped 46 fecal DNA samples using both PumaPlex and 12 microsatellite loci. The overall proportion of called genotypes was significantly greater for SNPs (59.8%) than microsatellites (39.9%; Wilcoxon signed-rank test, $v = 802.5$, $P = 7.1 \times 10^{-5}$; Figure 5a). Using a cutoff of at least two called genotypes per locus, SNPs produced significantly more consensus genotypes (Wilcoxon signed-rank test, $v = 709$, $P = 3.1 \times 10^{-4}$; Figure 5a) and fewer false alleles (Wilcoxon signed-rank test, $v = 64$, $P = 0.025$; Figure 5b) than microsatellites. The frequency of allelic dropout was also less but not statistically significant (Wilcoxon signed-rank test, $v = 29.5$, $P = 0.123$; Figure 5b). After applying a more stringent cutoff of at least three genotypes called per locus, the proportion of genotypes successfully called was approximately 2.5 times greater for SNPs than microsatellites (Wilcoxon signed-rank test, $v = 597.5$, $P = 3.9 \times 10^{-6}$; Figure 5a). The frequency of false alleles did

Table 3. Descriptive statistics per locus for pumas *Puma concolor cougar* from Arizona (AZ) and the rest of North America (NAM) from 1983 to 2013. Sample sizes are indicated within parentheses. Abbreviations include: minimum allele frequency (*q*), observed heterozygosity (H_O), expected heterozygosity (H_E), and standard deviation (SD). The asterisks indicate loci significantly deviating from Hardy–Weinberg equilibrium after Bonferroni correction ($P < 0.05$) in AZ only.

SNP name	AZ (n = 518)			NAM (n = 209)		
	<i>q</i>	H_O	H_E	<i>q</i>	H_O	H_E
PP01	0.409	0.459	0.484	0.450	0.478	0.495
PP02	0.069	0.125	0.128	0.057	0.115	0.108
PP03	0.164	0.241	0.274	0.143	0.122	0.245
PP04	0.292	0.401	0.413	0.187	0.278	0.304
PP05	0.231	0.292	0.355*	0.349	0.389	0.454
PP06	0.346	0.440	0.452	0.122	0.187	0.214
PP07	0.253	0.363	0.378	0.098	0.187	0.177
PP08	0.383	0.365	0.473*	0.413	0.396	0.485
PP09	0.456	0.440	0.496	0.299	0.378	0.419
PP10	0.222	0.320	0.345	0.194	0.292	0.312
PP11	0.266	0.367	0.391	0.171	0.293	0.283
PP12	0.353	0.451	0.457	0.270	0.263	0.395
PP13	0.193	0.309	0.312	0.084	0.148	0.153
PP14	0.103	0.180	0.186	0.098	0.167	0.177
PP15	0.166	0.247	0.277	0.151	0.244	0.256
PP16	0.490	0.463	0.500	0.363	0.409	0.462
PP17	0.349	0.455	0.454	0.428	0.407	0.490
PP18	0.197	0.311	0.316	0.084	0.149	0.154
PP19	0.097	0.066	0.175*	0.091	0.038	0.165
PP20	0.381	0.442	0.472	0.225	0.249	0.349
PP21	0.107	0.164	0.191	0.051	0.071	0.097
PP22	0.324	0.393	0.438	0.167	0.230	0.279
PP23	0.321	0.455	0.436	0.267	0.320	0.391
PP24	0.199	0.282	0.319	0.194	0.282	0.312
PP25	0.428	0.534	0.490	0.406	0.591	0.482
Mean	0.272	0.343	0.368	0.214	0.267	0.306
SD	0.120	0.121	0.113	0.126	0.133	0.130

not differ (Wilcoxon signed-rank test, $v = 55.5$, $P = 0.88$; Figure 5b) between SNPs and microsatellites, and we observed no allelic dropout in either data set.

Discussion

Transcriptome assembly and annotation

In this study we sequenced and assembled the transcriptome of a pool of 12 pumas in order to identify candidate SNPs for use in population genetic studies. Despite the fact that mammalian peripheral blood is primarily composed of nonnucleated erythrocytes and considered to contain little transcript abundance, we were able to produce >2,000 contigs and annotate 59% of them. The most common functional annotations were consistent with those reported from human erythrocytes (Kabanova et al. 2009) and with well-known properties of blood such as oxygen transport. We also identified a variety of genes implicated in immune-related processes, suggesting that future comparison of transcriptome profiles between healthy and diseased individuals could improve our understanding of wildlife health and diseases (Burczynski and Dorner 2006). Furthermore,

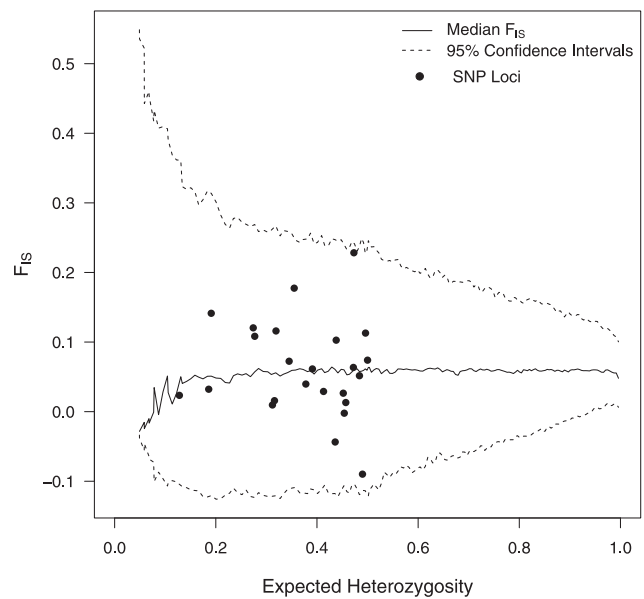


Figure 3. Mean F_{IS} (solid line) for varying levels of heterozygosity predicted after 100,000 simulations following the method of Girard (2011). The upper and lower 95% confidence intervals are shown as dashed lines. Each black circle represents the observed values for each single nucleotide polymorphism (SNP) locus (some loci overlap) from PumaPlex genotyped in pumas *Puma concolor cougar* collected from Arizona and New Mexico, USA, between 2007 and 2013. Loci above the 95% confidence interval have an excess of homozygosity, suggesting a potential for null alleles, whereas those loci below the 95% confidence interval display an excess of heterozygosity compared with the predicted distribution.

the use of our sequence reads and transcriptome assembly can greatly improve the identification and annotation of genome sequences from pumas or other closely related species (Wolfsberg and Landsman 1997; Kan et al. 2001; Denton et al. 2014).

PumaPlex development and genotyping

The PumaPlex assay is currently set up to genotype 25 SNPs per sample in either a 384- or 96-sample format on the Sequenom MassARRAY system. This technology is available at many laboratories and genetic core service centers across the country, where the workflow, from sample to results, can be completed in as little as 1 d (www.sequenom.com). This type of automated workflow is much faster than traditional microsatellite analysis, saving time and costs for wildlife practitioners. PumaPlex also requires very small amounts of DNA (approx. 10 ng/sample), thus providing an efficient and economical use of samples containing limited DNA quantity such as in noninvasive and museum-collected samples. We demonstrated that the MassARRAY system generates high call rates and reproducible genotypes, consistent with that reported elsewhere for the technology (Pusch et al. 2002; Cronin et al. 2014). Our genotypes were also congruent with those verified using traditional Sanger sequencing. Unlike other high-throughput SNP panels

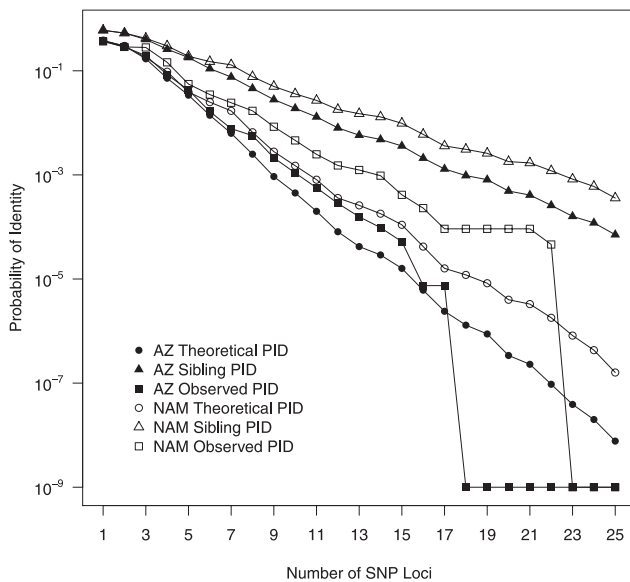


Figure 4. Theoretical (circles), sibling (triangles), and observed (squares) probability of identity (P_{ID}) for increasing single nucleotide polymorphism (SNP) locus combinations in pumas *Puma concolor couguar* collected from Arizona and New Mexico, USA (AZ; black shapes) and the rest of North America (NAM; white shapes) from 1983 to 2013 and genotyped using PumaPlex. Theoretical and observed P_{ID} are calculated according to Waits et al. (2001) using the allele frequencies from the population sampled. Observed P_{ID} is calculated as the proportion of all pairwise genotype comparisons that match at the given number of loci.

designed for wildlife species with available genome sequences (Cronin et al. 2014; Kraus et al. 2015), there were no SNPs previously known to be polymorphic in pumas. The MassARRAY system provided an effective method to screen candidate SNPs identified from small amounts of next-generation sequencing data.

With PumaPlex we were able to accurately identify individuals. Both theoretical and observed P_{ID} surpassed the recommended threshold (0.01–0.0001) proposed by Waits et al. (2001) using a minimum of 11 and 14 loci, respectively. Sethi et al. (2014) recommended ≥ 32 SNPs with a mean minimum allele frequency of 0.2 to reduce errors in mark–recapture studies, albeit less is sufficient when allelic dropout is low (< 0.6) and the number of SNPs can adequately identify individuals. The authors also demonstrated methods to minimize mark–recapture errors through the use of repeated genotyping or error-tolerant matching protocols (e.g., Creel et al. 2003). With PumaPlex, we observed no allelic dropout and low P_{ID} , demonstrating that the assay is useful for studies that require individual identification, such as estimating and monitoring population sizes, home range sizes, paternity, kinship, and forensic applications. The P_{ID} we calculated includes all 25 loci, despite 3 loci (PP05, PP08, and PP19) deviating significantly from Hardy–Weinberg proportions. Even after removal of these three loci, the P_{ID} (9.5×10^{-8}) still remained well below the recommended

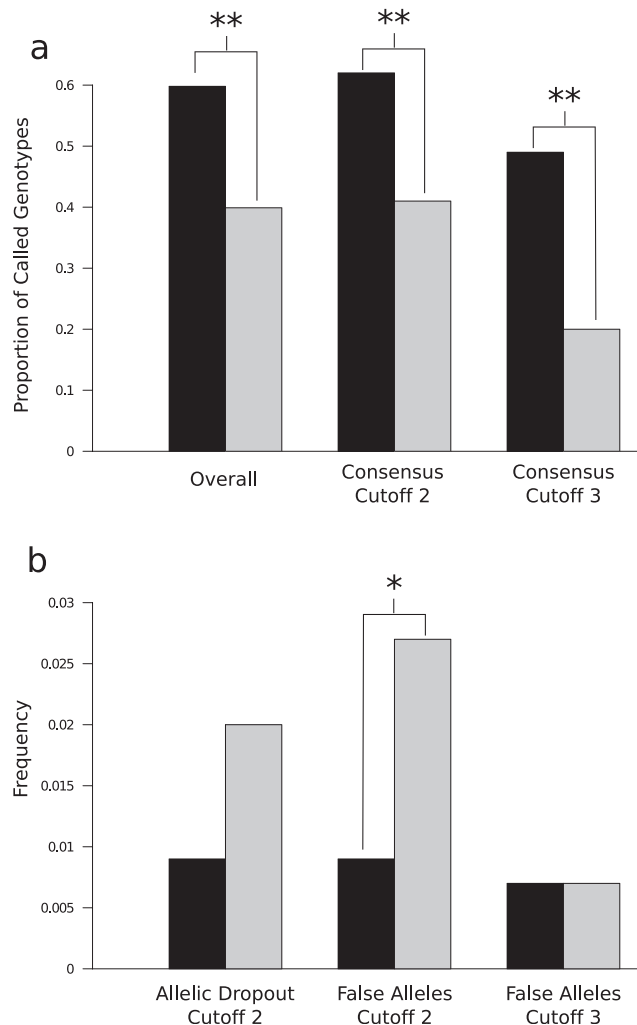


Figure 5. Comparisons of (a) the proportion of overall and consensus genotypes called, and (b) the frequency of allelic dropout and false alleles in 46 fecal DNA samples from pumas *Puma concolor couguar* collected in Arizona and Mexico from 2007 to 2013 and genotyped for 25 single nucleotide polymorphisms (SNPs) using PumaPlex (black bars) and 12 microsatellite loci (grey bars). “Cutoff 2” and “Cutoff 3” are the number of replicate genotypes required to call a consensus using GIMLET v1.3.3 (Valière 2002). Single and double asterisks indicate significance at $P < 0.05$ and $P < 0.001$, respectively, using a Wilcoxon signed-rank test.

threshold. We reported the P_{ID} across all loci to estimate the minimum possible, and we recommend PumaPlex users exclude markers that significantly deviate from the expected proportions (e.g., using tests of Hardy–Weinberg equilibrium, null alleles, or selection) prior to estimating P_{ID} . All loci, however, will provide useful information when determining sibship or excluding parents.

It is also important to understand the potential effects of ascertainment bias, or the systematic deviations from theoretical expectations between the population used for SNP discovery and the population under study. In our case, we used a small SNP discovery panel ($n = 12$) from

AZ, which may have biased loci to those with alleles at intermediate frequencies in the population (Morin et al. 2004). This is of relatively little concern when used to analyze AZ, but must be taken into account when examining other, more distantly related populations. This bias can account for the higher mean minimum allele frequency and smaller P_{ID} in AZ compared with NAM, in addition to population structure and other demographic factors. For this reason we refrained from reporting several statistics in the NAM population (i.e., Hardy–Weinberg equilibrium, test for null alleles) and provided the data solely to demonstrate polymorphism across the subspecies. We suggest that researchers who use PumaPlex are cautious when analyzing populations genetically distinct from AZ for parameters sensitive to the allele frequency spectrum, such as genetic variation, effective population size, and demographic history, or account for ascertainment bias using statistical methods described elsewhere (Wakeley et al. 2001; Nielsen and Signorovitch 2003). However, PumaPlex still provides many useful, individual-based applications less sensitive to this bias, including individual identification, paternity, and assignment tests (Morin et al. 2004).

Although we did not investigate population structure in this study, work by Morin et al. (2009) suggested that approximately 30 SNPs would only be able to detect moderate differentiation ($F_{ST} = 0.01$), whereas ≥ 80 SNPs would be necessary to characterize demographic independence ($F_{ST} < 0.005$). This argues that PumaPlex, in its current form, is better suited to monitoring and identification of individuals rather than detecting population structure, although strong structure ($F_{ST} \gg 0.01$) is likely detectable. In 22/25 (88%) of loci, we observed a heterozygosity deficit, which can often be attributed to demographic effects such as inbreeding or population structure (Wahlund effect). Technical artifacts such as null alleles may also produce this effect, yet our results suggest that null alleles are not present (Figure 3). In Arizona, McRae et al. (2005) observed differentiation between pumas collected north and south of Interstate 40, implicating the highway as a possible barrier to gene flow. Because our Arizona data set (AZ) includes the entire state, population structure is a plausible explanation for the heterozygosity deficit.

Ongoing work in our laboratory is testing additional candidate SNPs to expand the number of loci on future versions of PumaPlex. Although a majority of the SNPs on PumaPlex are noncoding, it is generally considered that variants in coding or other regulatory elements (such as 3' and 5' untranslated regions) are more likely to cause phenotypic effects and be under selection. Only three loci differed from Hardy–Weinberg proportions; therefore, the assumption of neutrality cannot be excluded for the remaining loci. However, there are special situations where strong selection can have no effect on Hardy–Weinberg proportions (see Lewontin and Cockerham 1959), and, in contrast with overdominance and underdominance, directional selection has

relatively little impact on genotype frequencies (Waples 2015). This assumption of neutrality does not differ from that of other markers, such as microsatellites, which also may be linked to selected loci or themselves be the targets of selection (reviewed by Haas and Payseur 2013). Future versions of PumaPlex will target an increasing number of intergenic and synonymous SNPs in addition to those ascertained from all six subspecies of pumas to expand its potential applications while minimizing potential impacts of ascertainment biases and selection.

Application to noninvasively collected samples

The ability to identify individuals has become popular with the expansion of population surveys using noninvasive sampling (Waits and Paetkau 2005). In pumas, noninvasive genetic studies have focused on the amplification of microsatellites from fecal DNA, although often encountering poor genotyping success (e.g., Ernest et al. 2002; Miotto et al. 2007; Naidu et al. 2011). Our genotyping rate of SNPs in tissue samples was much higher (95.2%) than from fecal samples (59.8%); this result is consistent with the harsh desert conditions in southwestern Arizona and northern Mexico, including high temperatures and extensive exposure to ultraviolet light that can cause substantial damage to DNA (Friedberg 2003; Hofreiter et al. 2015).

Several studies have proposed that SNP genotyping, which often utilizes short DNA fragments, may be more conducive than microsatellites to the conditions of DNA often encountered in noninvasively collected specimens (Morin et al. 2004; Seddon et al. 2005; Morin and McCarthy 2007; Ogden 2011). Our results support this hypothesis, because PumaPlex produced more overall and consensus genotypes with fewer false alleles than microsatellites. A similar result has also been reported from wolf *Canis lupus* scats (Fabbri et al. 2012); however, our study is the first to corroborate this using the MassARRAY system. It is possible that dissimilar PCR amplification and cycling conditions explain the difference between SNPs and microsatellites. However, this is unlikely because conditions for both marker sets have been optimized either in silico (SNPs) or through extensive empirical use in puma scats (microsatellites; Ernest et al. 2000, 2002, 2003; Miotto et al. 2007; Naidu et al. 2011). The optimized PCR conditions for the SNPs on PumaPlex are standardized across all MassArray genotyping systems, demonstrating the advantage of such an assay relative to the time and resources necessary to develop study-specific conditions often required for microsatellite analysis. Nevertheless, the reduction in genotype error is important because errors can substantially inflate population size estimates using mark-recapture methods (Waits and Leberg 2000). However, users must also take the information content of the loci into account, because microsatellites are more informative per locus than SNPs on account of the increased number of alleles (Morin et al. 2004).



Data archiving and accessibility

We created the Puma Genetic Database to have this SNP data set stored in a secure location and make it available for download in multiple standardized formats (e.g., CSV, File Geodatabase, KML, Shapefile) offered through the “Perform Analysis” tools in ArcGIS Online. This service can facilitate the exchange of data among researchers and laboratories with authorized access to the Puma Genetic Database. This database is now live at (<http://www.arcgis.com/home/item.html?id=4d9e04e504bb453691fbff736df49b3b>) and can be viewed publicly, but interested persons should contact the corresponding author to request authorization to upload and download data from the Puma Genetic Database. Using the options available in ArcGIS Online, this database facilitates easy filtering and selection of individual genotypes based on a variety of criteria (e.g., any user-defined geographic region, sex, date, sample type, country, etc.). The Puma Genetic Database will be continually updated as more genotyping is performed using PumaPlex. We encourage future users of PumaPlex to share their data and to contact the corresponding author to obtain instructions on formatting and submitting data for the database. If researchers intend to design independent genotyping assays using other techniques, the contig sequences and SNP information necessary are provided in Data A2 (*Archived Material*, <http://doi.pangaea.de/10.1594/PANGAEA.835154>) and Data S1 (*Supplemental Material*). Genotypes from these alternative genotyping technologies can also be submitted to the Puma Genetic Database.

In this study we developed PumaPlex, the first set of SNPs specific for use in pumas, and demonstrated its effectiveness in genotyping DNA extracted from both tissue and fecal material. Furthermore, we created a database to host the genetic and other information collected from pumas to promote data-sharing among researchers. Both tools are valuable resources to facilitate range-wide conservation genetic assessments and management of North American pumas.

Supplemental Material

Please note: The *Journal of Fish and Wildlife Management* is not responsible for the content or functionality of any supplemental material. Queries should be directed to the corresponding author for the article.

Data S1. Sequences of the 25 contigs containing the single nucleotide polymorphisms (SNPs) on PumaPlex. Each SNP and its two alleles are indicated within brackets (e.g., “[A/G]”). These SNPs were developed from a transcriptome sequenced from a pooled cDNA library of 12 pumas *Puma concolor couguar* collected in Arizona and New Mexico, USA, between 2009 and 2010. The file is in FASTA format and can be viewed in any text editor.

Found at DOI: <http://dx.doi.org/10.3996/112014-JFWM-080.S1> (26 KB TXT).

Table S1. The concentration and volume of polymerase chain reaction (PCR) reagents and cycling conditions for each step of the PumaPlex genotyping using the MassARRAY system (Sequenom). These PCR reaction (rxn) conditions were used to genotype 25 single-nucleotide polymorphisms in all puma *Puma concolor couguar* tissue and scat samples collected in North America from 1983 to 2013. The first step (“30 Plex, Full Enzyme”) is a traditional PCR, the second step (“SAP Reaction”) is a clean-up protocol using shrimp alkaline phosphatase (SAP), and the third step (“iPLEX Extension 30 Plex, Full Enzyme/Term”) is a single-base extension PCR using the internal, extension primer. All reactions were carried out at the University of Arizona Genetics Core using the manufacturer’s specifications and supplied reagents. The file is formatted for use in Microsoft Excel. Other abbreviations: HPLC = high-performance liquid chromatography,

Found at DOI: <http://dx.doi.org/10.3996/112014-JFWM-080.S2> (12 KB XLSX).

Table S2. The annotations for each contig of the transcriptome assembly produced by FASTANNOTATOR (Chen et al. 2012). The transcriptome was sequenced from a pooled cDNA library of 12 pumas *Puma concolor couguar* collected in Arizona and New Mexico, USA, between 2009 and 2010. The annotations include a summary of the best BLAST result to GenBank’s nonredundant database (“Top BLAST Hit,” “Hit Length,” “E-value,” “Bit Score”), associated gene ontology (GO) terms (see www.geneontology.org), Enzyme Commission (EC) numbers (“Enzyme,” see <http://enzyme.expasy.org/>), and protein domains identified (“Domain,” see <http://pfam.xfam.org/>). The dashes indicate contigs unable to be annotated. The file is in a tab-delimited, text format.

Found at DOI: <http://dx.doi.org/10.3996/112014-JFWM-080.S3> (704 KB TXT).

Table S3. The primers and information necessary to build PumaPlex on the Sequenom MassARRAY system (see Sequenom MassARRAY Assay Design v3.1 user manual; table 3 p. 70). The oligonucleotides shown here were used to amplify the 25 loci on PumaPlex in all puma *Puma concolor couguar* tissue and scat samples collected in North America from 1983 to 2013. WELL = the well number assigned to the assay; multiplexed assays are assigned the same well number, TERM = termination mix, SNP_ID = name of the input sequence, 2nd-PCRPR = secondary amplification primer (includes secondary tag), 1st-PCRPR = primary amplification primer (includes primer tag), AMP_LEN = amplicon length (in bases; includes primer tags and maximum SNP sequence length), UP_CONF = uniplex amplification score; this score indicates how well the multiplexed amplicon design meets the design criteria, MP_CONF = multiplex amplification score; this score indicates how well the multiplexed amplicon design meets the design criteria, Tm(NN) = extend primer melting temperature, calculated by Nearest Neighbor method, PcGC = percent GC content of the extend primer, PWARN = assay design warning codes (D/d = primer dimer potential between

primers of multiplexed assays; G/g = primer subsequence of contiguous G nucleotides; H/h = primer hairpin or self-dimer potential; S/s = false priming potential), UEP_DIR = direction of MassEXTEND® (F = Forward, R = Reverse), UEP_MASS = extend primer mass, UEP_SEQ = extend primer sequence, EXT1_CALL = name given to the analyte 1 mass peak in the mass spectrum, EXT1_MASS = mass of analyte 1, EXT1_SEQ = sequence of analyte 1, EXT2_CALL = name given to the analyte 2 mass peak in the mass spectrum, EXT2_MASS = mass of analyte 2, EXT2_SEQ = sequence of analyte 2. The file is formatted for use in Microsoft Excel.

Found at DOI: <http://dx.doi.org/10.3996/112014-JFWM-080.S4> (46 KB XLS).

Table S4. Description and Gene Ontology (GO) terms (see www.geneontology.org) assigned to the genes of the 25 single nucleotide polymorphisms (SNPs) on PumaPlex. These 25 SNPs were developed from a transcriptome sequenced from a pooled cDNA library of 12 pumas *Puma concolor couguar* collected in Arizona and New Mexico, USA, between 2009 and 2010. All data were retrieved from the genomic resources available for the domestic cat *Felis catus* (www.ensembl.org). The file is formatted for use in Microsoft Excel.

Found at DOI: <http://dx.doi.org/10.3996/112014-JFWM-080.S5> (30 KB XLSX).

Figure S1. The frequency of the 50 most common Biological Process Gene Ontology (GO) terms assigned to the transcriptome sequenced from a pooled cDNA library of 12 pumas *Puma concolor couguar* collected in Arizona and New Mexico, USA, between 2009 and 2010. A complete list of GO terms and detailed descriptions is maintained by the Gene Ontology Consortium (see www.geneontology.org).

Found at DOI: <http://dx.doi.org/10.3996/112014-JFWM-080.S6> (136 KB PDF).

Figure S2. The frequency of the 50 most common Cellular Component Gene Ontology (GO) terms assigned to the transcriptome sequenced from a pooled cDNA library of 12 pumas *Puma concolor couguar* collected in Arizona and New Mexico, USA, between 2009 and 2010. A complete list of GO terms and detailed descriptions is maintained by the Gene Ontology Consortium (see www.geneontology.org).

Found at DOI: <http://dx.doi.org/10.3996/112014-JFWM-080.S7> (125 KB PDF).

Figure S3. The frequency of the 50 most common Molecular Function Gene Ontology (GO) terms assigned to the transcriptome sequenced from a pooled cDNA library of 12 pumas *Puma concolor couguar* collected in Arizona and New Mexico, USA, between 2009 and 2010. A complete list of GO terms and detailed descriptions is maintained by the Gene Ontology Consortium (see www.geneontology.org).

Found at DOI: <http://dx.doi.org/10.3996/112014-JFWM-080.S8> (133 KB PDF).

Reference S1. Naidu A, Fitak R, Culver M. 2014a. Landscape genetics of mountain lions (*Puma concolor*) in southwestern Arizona. Final report to the Arizona Game and Fish Department Habitat Partnership Committee, Project number HPC-09-406, Tucson, Arizona.

Found at DOI: <http://dx.doi.org/10.3996/112014-JFWM-080.S9> (1723 KB PDF).

Reference S2. Naidu A, Fitak R, Culver M. 2014b. Data sharing for wildlife management: the puma genetic database. Final report to the Arizona Game and Fish Department Habitat Partnership Committee, Project number HPC-10-705, Tucson, Arizona.

Found at DOI: <http://dx.doi.org/10.3996/112014-JFWM-080.S10> (1020 KB PDF).

Archived Material

Please note: The Journal of Fish and Wildlife Management is not responsible for the content or functionality of any archived material. Queries should be addressed to the corresponding author for the article.

To cite this archived material, please cite both the journal article (formatting found in the Abstract section of this article) and the following recommended format for the archived material.

Data A1. We deposited the raw sequence data into the Sequence Read Archive in GenBank under accession number SRX633288, titled “A transcriptome resource for pumas *Puma concolor couguar*” (available at: <http://www.ncbi.nlm.nih.gov/sra/SRX633288>). The transcriptome was sequenced from a pooled cDNA library of 12 pumas collected in Arizona and New Mexico, USA, between 2009 and 2010.

Data A2. Fitak, R et al. (2014): PumaPlex 1.0: Data generated during the development of the SNP markers and SNP genotypes of pumas. doi:10.1594/PANGAEA.835154.

Acknowledgments

We would like to thank those who provided samples and expertise, especially S. Bender, the Navajo Nation Zoo, L. Searles of the Southwest Wildlife Conservation Center, the late Andy Salazar, and the Arizona Game and Fish Department. We thank the NSF-IGERT program in comparative genomics and the Habitat Partnership Committee grants HPC 09-406 and 10-705 for funding. We are deeply indebted to C. Ramsower at the University of Arizona Genetics Core for assistance with Sequenom genotyping. We are grateful to T. Edwards, S. Blount, C. Voirin, E. Mohandesan, the Subject Editor, and three anonymous reviewers for their useful comments on earlier drafts of this manuscript.

Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.



References

- Anderson CR, Lindzey FG, McDonald DB. 2004. Genetic structure of cougar populations across the Wyoming Basin: metapopulation or megapopulation. *Journal of Mammalogy* 85:1207–1214.
- Andreasen AM, Stewart KM, Longland WS, Beckmann JP, Forister ML. 2012. Identification of source–sink dynamics in mountain lions of the Great Basin. *Molecular Ecology* 21:5689–5701.
- Balkenhol N, Holbrook JD, Onorato D, Zager P, White C, Waits LP. 2014. A multi-method approach for analyzing hierarchical genetic structures: a case study with cougars *Puma concolor*. *Ecography* 37:552–563.
- Beier P. 1991. Cougar attacks on humans in the United States and Canada. *Wildlife Society Bulletin* 19:403–412.
- Biek R, Akamine N, Schwartz MK, Ruth TK, Murphy KM, Poss M. 2006. Genetic consequences of sex-biased dispersal in a solitary carnivore: Yellowstone cougars. *Biology Letters* 2:312–315.
- Burczynski ME, Dorner AJ. 2006. Transcriptional profiling of peripheral blood cells in clinical pharmacogenomic studies. *Pharmacogenomics* 7:187–202.
- Chen TW, Gan RC, Wu TH, Huang PJ, Lee CY, Chen YY, Chen CC, Tang P. 2012. FastAnnotator—an efficient transcript annotation web tool. *BMC Genomics* 13(Suppl 7):S9.
- Claudel-Renard C, Chevalet C, Faraut T, Kahn D. 2003. Enzyme-specific profiles for genome annotation: PRIAM. *Nucleic Acids Research* 31:6633–6639.
- Coates BS, Sumerford DV, Miller NJ, Kim KS, Sappington TW, Siegfried BD, Lewis LC. 2009. Comparative performance of single nucleotide polymorphism and microsatellite markers for population genetic analysis. *Journal of Heredity* 100:556–564.
- Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674–3676.
- Creel S, Spong G, Sands JL, Rotella J, Zeigle J, Joe L, Murphy KM, Smith D. 2003. Population size estimation in Yellowstone wolves with error-prone noninvasive microsatellite genotypes. *Molecular Ecology* 12:2003–2009.
- Cronin MA, Rincon G, Meredith RW, MacNeil MD, Islas-Trejo A, Canovas A, Medrano JF. 2014. Molecular phylogeny and SNP variation of polar bears (*Ursus maritimus*), brown bears (*U. arctos*), and black bears (*U. americanus*) derived from genome sequences. *Journal of Heredity* 105:312–323.
- Culver M, Johnson WE, Pecon-Slattery J, O'Brien SJ. 2000. Genomic ancestry of the American puma (*Puma concolor*). *Journal of Heredity* 91:186–197.
- Culver M, Schwartz MK. 2011. Conservation genetics and cougar management. Pages 91–110 in Jenks JA, editor. *Managing cougars in North America*. Logan, Utah: Jack H. Berryman Institute.
- Denton JF, Lugo-Martinez J, Tucker AE, Schrider DR, Warren WC, Hahn MW. 2014. Extensive error in the number of genes inferred from draft genome assemblies. *PLoS Computational Biology* 10:e1003998.
- Ernest HB, Boyce WM, Bleich VC, May B, Stiver SJ, Torres SG. 2003. Genetic structure of mountain lion (*Puma concolor*) populations in California. *Conservation Genetics* 4:353–366.
- Ernest HB, Penedo MCT, May BP, Syvanen M, Boyce WM. 2000. Molecular tracking of mountain lions in the Yosemite Valley region in California: genetic analysis using microsatellites and faecal DNA. *Molecular Ecology* 9:433–441.
- Ernest HB, Rubin ES, Boyce WM. 2002. Fecal DNA analysis and risk assessment of mountain lion predation of bighorn sheep. *Journal of Wildlife Management* 66:75–85.
- Fabbri E, Caniglia R, Mucci N, Thomsen HP, Krag K, Pertoldi C, Loeschcke V, Randi E. 2012. Comparison of single nucleotide polymorphisms and microsatellites in non-invasive genetic monitoring of a wolf population. *Archives of Biological Science Belgrade* 64:321–335.
- Fernández ME, Goszczynski DE, Lirón JP, Villegas-Castagnasso EE, Carino MH, Ripoli MV, Rogberg-Muñoz A, Posik DM, Peral-García P, Giovambattista G. 2013. Comparison of the effectiveness of microsatellites and SNP panels for genetic identification, traceability and assessment of parentage in an inbred Angus herd. *Genetics and Molecular Biology* 36:185–191.
- Frankham R. 2003. Genetics and conservation biology. *Comptes Rendus Biologies* 326:22–29.
- Friedberg EC. 2003. DNA damage and repair. *Nature* 421:436–440.
- Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. *ArXiv e-prints*:1207.3907.
- Garvin MR, Saitoh K, Gharrett AJ. 2010. Application of single nucleotide polymorphisms to non-model species: a technical review. *Molecular Ecology Resources* 10:915–934.
- Girard P. 2011. A robust statistical method to detect null alleles in microsatellite and SNP datasets in both panmictic and inbred populations. *Statistical Applications in Genetics and Molecular Biology* 10:1–10.
- Haag T, Santos AS, De Angelo C, Srbek-Araujo AC, Sana DA, Morato RG, Salzano FM, Eizirik E. 2009. Development and testing of an optimized method for DNA-based identification of jaguar (*Panthera onca*) and puma (*Puma concolor*) faecal samples for use in ecological and genetic studies. *Genetica* 136:505–512.
- Haas RJ, Payseur BA. 2013. Microsatellites as targets of natural selection. *Molecular Biology and Evolution* 30:285–298.
- Hahn M, Wilhelm J, Pingoud A. 2001. Influence of fluorophore dye labels on the migration behavior of polymerase chain reaction amplified short tandem repeats during denaturing capillary electrophoresis. *Electrophoresis* 22:2691–2700.

- Hayes CL, Rubin ES, Jorgensen MC, Botta RA, Boyce WM. 2000. Mountain lion predation of bighorn sheep in the Peninsular Ranges, California. *Journal of Wildlife Management* 64:954–959.
- Hofreiter M, Paijmans JLA, Goodchild H, Speller CF, Barlow A, Fortes GG, Thomas JA, Ludwig A, Collins MJ. 2015. The future of ancient DNA: technical advances and conceptual shifts. *Bioessays* 37:284–293.
- Holbrook JD, DeYoung RW, Janecka JE, Tewes ME, Honeycutt RL, Young JH. 2012a. Genetic diversity, population structure, and movements of mountain lions (*Puma concolor*) in Texas. *Journal of Mammalogy* 93:989–1000.
- Holbrook JD, DeYoung RW, Tewes ME, Young JH. 2012b. Demographic history of an elusive carnivore: using museums to inform management. *Evolutionary Applications* 5:619–628.
- Kabanova S, Kleinbongard P, Volkmer J, Andree B, Kelm M, Jax TW. 2009. Gene expression analysis of human red blood cells. *International Journal of Medical Sciences* 6:156–159.
- Kan Z, Rouchka EC, Gish WR, States DJ. 2001. Gene structure prediction and alternative splicing analysis using genomically aligned ESTs. *Genome Research* 11:889–900.
- Kraus RH, vonHoldt B, Cocchiara B, Harms V, Bayerl H, Kuhn R, Forster DW, 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.
- Kurushima JD, Collins JA, Well JA, Ernest HB. 2006. Development of 21 microsatellite loci for puma (*Puma concolor*) ecology and forensics. *Molecular Ecology Notes* 6:1260–1262.
- Lewontin RC, Cockerham CC. 1959. The goodness-of-fit test for detecting natural selection in random mating populations. *Evolution* 13:561–564.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760.
- Logan KA, Irwin LL. 1985. Mountain lion habitats in the Big Horn Mountains, Wyoming. *Wildlife Society Bulletin* 13:257–262.
- Logan KA, Sweanor LL. 2001. Desert puma: evolutionary ecology and conservation of an enduring carnivore. Washington D.C.: Island Press.
- Loxterman JL. 2011. Fine scale population genetic structure of pumas in the Intermountain West. *Conservation Genetics* 12:1049–1059.
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal* 17:10–12.
- McRae BH, Beier P, Dewald LE, Huynh LY, Keim P. 2005. Habitat barriers limit gene flow and illuminate historical events in a wide-ranging carnivore, the American puma. *Molecular Ecology* 14:1965–1977.
- Menotti-Raymond M, David VA, Stephens JC, Lyons LA, O'Brien SJ. 1997. Genetic individualization of domestic cats using feline STR loci for forensic applications. *Journal of Forensic Sciences* 42:1039–1051.
- Menotti-Raymond M, David VA, Lyons LA, Schaffer AA, Tomlin JF, Hutton MK, O'Brien SJ. 1999. A genetic linkage map of microsatellites in the domestic cat (*Felis catus*). *Genomics* 57:9–23.
- Menotti-Raymond MA, O'Brien SJ. 1995. Evolutionary conservation of ten microsatellite loci in four species of Felidae. *Journal of Heredity* 86:319–322.
- Miotto RA, Cervini M, Figueiredo MG, Begotti RA, Galetti PM. 2011. Genetic diversity and population structure of pumas (*Puma concolor*) in southeastern Brazil: implications for conservation in a human-dominated landscape. *Conservation Genetics* 12:1447–1455.
- Miotto RA, Rodrigues FP, Ciocheti G, Galetti PM. 2007. Determination of the minimum population size of pumas (*Puma concolor*) through fecal DNA analysis in two protected cerrado areas in the Brazilian Southeast. *Biotropica* 39:647–654.
- Morin PA, Luikart G, Wayne RK, the SNP Workshop Group. 2004. SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution* 19:208–216.
- Morin PA, Martien KK, Taylor BL. 2009. Assessing statistical power of SNPs for population structure and conservation studies. *Molecular Ecology Resources* 9:66–73.
- Morin PA, McCarthy M. 2007. Highly accurate SNP genotyping from historical and low-quality samples. *Molecular Ecology Notes* 7:937–946.
- Naidu A, Fitak R, Culver M. 2014a. Landscape genetics of mountain lions (*Puma concolor*) in southwestern Arizona. Final report to the Arizona Game and Fish Department Habitat Partnership Committee, Project number HPC-09-406, Tucson, Arizona (see *Supplemental Material*, Reference S1, <http://dx.doi.org/10.3996/112014-JFWM-080.S1>).
- Naidu A, Fitak R, Culver M. 2014b. Data sharing for wildlife management: the puma genetic database. Final report to the Arizona Game and Fish Department Habitat Partnership Committee, Project number HPC-10-705, Tucson, Arizona (see *Supplemental Material*, Reference S2, <http://dx.doi.org/10.3996/112014-JFWM-080.S2>).
- Naidu A, Smythe LA, Thompson RW, Culver M. 2011. Genetic analysis of scats reveals minimum number and sex of recently documented mountain lions. *Journal of Fish and Wildlife Management* 2:106–111.
- Nicholson KL, Krausman PR, Munguia-Vega A, Culver M. 2011. Spatial and temporal interactions of sympatric mountain lions in Arizona. *European Journal of Wildlife Research* 57:1151–1163.
- Nielsen R, Signorovitch J. 2003. Correcting for ascertainment biases when analyzing SNP data: applications to the estimation of linkage disequilibrium. *Theoretical Population Biology* 63:245–255.
- Ogden R. 2011. Unlocking the potential of genomic technologies for wildlife forensics. *Molecular Ecology Resources* 11:109–116.



- Onorato D, Desimone R, White C, Waits LP. 2011. Genetic assessment of paternity and relatedness in a managed population of cougars. *Journal of Wildlife Management* 75:378–384.
- Peakall R, Smouse PE. 2012. GenAEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28:2537–2539.
- Pertoldi C, Wójcik JM, Tokarska M, Kawalko A, Kristensen TN, Loeschcke V, Gregersen VR, Coltman D, Wilson GA, Randi E, Henryon M, Bendixen C. 2010. Genome variability in European and American bison detected using the BovineSNP50 BeadChip. *Conservation Genetics* 11:627–634.
- Pusch W, Wurmbach JH, Thiele H, Kostrzewa M. 2002. MALDI-TOF mass spectrometry-based SNP genotyping. *Pharmacogenomics* 3:537–548.
- Ripple WJ, Estes JA, Beschta RL, Wilmers CC, Ritchie EG, Hebblewhite M, Berger J, Elmhagen B, Letnic M, Nelson MP, Schmitz OJ, Smith DW, Wallach AD, Wirsing AJ. 2014. Status and ecological effects of the world's largest carnivores. *Science* 343:1241484–1241484.
- Rodzen JA, Banks JD, Meredith EP, Jones KC. 2007. Characterization of 37 microsatellite loci in mountain lions (*Puma concolor*) for use in forensic and population applications. *Conservation Genetics* 8:1239–1241.
- Rozen S, Skaletsky HJ. 2000. Primer3 on the WWW for general users and for biologist programmers. Pages 365–386 in Krawetz S, Misener S, editors. *Bioinformatics methods and protocols: methods in molecular biology*. Totowa, New Jersey: Humana Press.
- Ryynänen HJ, Tonteri A, Vasemägi A, Primmer CR. 2007. A comparison of biallelic markers and microsatellites for the estimation of population and conservation genetic parameters in Atlantic salmon (*Salmo salar*). *Journal of Heredity* 98:692–704.
- Schaefer RJ, Torres SG, Bleich VC. 2000. Survivorship and cause-specific mortality in sympatric populations of mountain sheep and mule deer. *California Fish and Game* 86:127–135.
- Schmitz OJ, Hamback PA, Beckerman AP. 2000. Trophic cascades in terrestrial systems: a review of the effects of carnivore removals on plants. *American Naturalist* 155:141–153.
- Seddon JM, Parker HG, Ostrander EA, Ellegren H. 2005. SNPs in ecological and conservation studies: a test in the Scandinavian wolf population. *Molecular Ecology* 14:503–511.
- Sethi SA, Cook GM, Lemons P, Wenburg J. 2014. Guidelines for MSAT and SNP panels that lead to high-quality data for genetic mark–recapture studies. *Canadian Journal of Zoology* 92:515–526.
- Sinclair EA, Swenson EL, Wolfe ML, Choate DC, Bates B, Crandall KA. 2001. Gene flow estimates in Utah's cougars imply management beyond Utah. *Animal Conservation* 4:257–264.
- Stoner DC, Weith WR, Wolfe ML, Mecham MB, Neville A. 2008. Long-distance dispersal of a female cougar in a basin and range landscape. *Journal of Wildlife Management* 72:933–939.
- Sunquist ME, Sunquist F. 2002. The essence of cats. Pages 11–13 in Sunquist ME, Sunquist F, editors. *Wild cats of the world*. Chicago: University of Chicago Press.
- Thompson DJ, Jenks JA. 2005. Long-distance dispersal by a subadult male cougar from the Black Hills, South Dakota. *Journal of Wildlife Management* 69:818–820.
- Torres SG, Mansfield TM, Foley JE, Lupo T, Brinkhaus A. 1996. Mountain lion and human activity in California: testing speculations. *Wildlife Society Bulletin* 24:451–460.
- Turner JW, Wolfe ML, Kirkpatrick JF. 1992. Seasonal mountain lion predation on a feral horse population. *Canadian Journal of Zoology* 70:929–934.
- Valière N. 2002. Gimlet: a computer program for analysing genetic individual identification data. *Molecular Ecology Notes* 2:377–379.
- Vignal A, Milan D, SanCristobal M, Eggen A. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution* 34:275.
- Waits JL, Leberg PL. 2000. Biases associated with population estimation using molecular tagging. *Animal Conservation* 3:191–199.
- Waits LP, Luikart G, Taberlet P. 2001. Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. *Molecular Ecology* 10:249–256.
- Waits LP, Paetkau D. 2005. Noninvasive genetic sampling tools for wildlife biologists: a review of applications and recommendations for accurate data collection. *Journal of Wildlife Management* 69:1419–1433.
- Wakeley J, Nielsen R, Liu-Cordero SN, Ardlie K. 2001. The discovery of single-nucleotide polymorphisms—and inferences about human demographic history. *American Journal of Human Genetics* 69:1332–1347.
- Walker CW, Harveson LA, Pittman MT, Tewes ME, Honeycutt RL. 2000. Microsatellite variation in two populations of mountain lions (*Puma concolor*) in Texas. *Southwestern Naturalist* 45:196–203.
- Waples RS. 2015. Testing for Hardy–Weinberg proportions: have we lost the plot? *Journal of Heredity* 106:1–19.
- Weber W, Rabinowitz A. 1996. A global perspective on large carnivore conservation. *Conservation Biology* 10:1046–1054.
- Wehausen JD. 1996. Effects of mountain lion predation on bighorn sheep in the Sierra Nevada and Granite Mountains of California. *Wildlife Society Bulletin* 24:471–479.
- Wenburg JK. 2011. Data archiving. *Journal of Fish and Wildlife Management* 2:1–2.
- Whitlock MC. 2011. Data archiving in ecology and evolution: best practices. *Trends in Ecology and Evolution* 26:61–65.
- Wolfsberg TG, Landsman D. 1997. A comparison of expressed sequence tags (ESTs) to human genomic sequences. *Nucleic Acids Research* 25:1626–1632.

