Supplemental Document 1

Genomic DNA was converted into nextRAD genotyping-by-sequencing libraries (SNPsaurus, LLC) as in Russello et al (ref: Russello MA, Waterhouse MD, Etter PD, Johnson EA. (2015) From promise to practice: pairing non-invasive sampling with genomics in conservation. PeerJ 3:e1106 https://dx.doi.org/10.7717/peerj.1106).

Genomic DNA was first fragmented with Nextera reagent (Illumina, Inc), which also ligates short adapter sequences to the ends of the fragments. The Nextera reaction was scaled for fragmenting 7 ng of genomic DNA, although 17.5 ng of genomic DNA was used for input to compensate for the amount of degraded DNA in the samples and to increase fragment sizes. Fragmented DNA was then amplified for 26 cycles at 73 degrees, with one of the primers matching the adapter and extending 9 nucleotides into the genomic DNA with the selective sequence GTGTAGAGG. Thus, only fragments starting with a sequence that can be hybridized by the selective sequence of the primer will be efficiently amplified. The nextRAD libraries were sequenced on a HiSeq 4000 with one lane of 150 bp reads (University of Oregon). We acknowledge that sampling of regions with high GC content can be under-represented due to non-random sequencing from Illumina sequencers.

Genotyping analysis used custom scripts (SNPsaurus, LLC) that trimmed reads to 100 bp, removed Nextera adapter, and trimmed the end of the read if bases were below a quality score of 10 using bbduk (BBMap tools, <http://sourceforge.net/projects/bbmap/>, bash bbmap/bbduk.sh in=$file out=$outfile ktrim=r k=17 hdist=1 mink=8 ref=bbmap/resources/nextera.fa.gz minlen=100 ow=t qtrim=r trimq=10).

Next, a de novo reference was created by collecting 10 million reads in total, evenly from the samples, excluding reads that had counts fewer than 10 (likely to be a low frequency sequencing error) or more than 1,000 (likely to be a repeat sequence error). To account for potential paralogs, SNPs with more than 2 alleles in a sample are excluded. The remaining loci were then aligned to each other to identify allelic loci and collapse allelic haplotypes to a single representative. All reads were mapped to the reference with an alignment identity threshold of 95% using bbmap (BBMap tools). Genotype calling was done using Samtools and bcftools (Li, 2011; Li et al., 2009; samtools mpileup -gu -Q 10 -t DP,DPR -f ref.fasta -b samples.txt | bcftools call -cv - > genotypes.vcf). The vcf was filtered to remove alleles with a population frequency of less than 3%. Loci were removed that were heterozygous in all samples or had more than 2 alleles in a sample (suggesting collapsed paralogs). The absence of artifacts was checked by counting SNPs at each read nucleotide position and determining that SNP number did not increase with reduced base quality at the end of the read.

Additional filtering using VCFTOOLS (Danecek et al., 2011) removed loci that had less than 10x coverage, minor alleles with a frequency of less than 0.05 and any sites with >20% missing data. After previous filtering was completed, the dataset was thinned to keep only one SNP per fragment to reduce linkage between loci. All remaining loci were evaluated for Hardy-Weinberg equilibrium (HWE) so that loci were removed if more than seven populations were out of HWE at p = 0.01. Any individual with >20% missing data was excluded from analyses.