A Resource of Genome-Wide Single Nucleotide Polymorphisms (Snps) for the Conservation and Management of Golden Eagles

Authors: Ronald A. Van Den Bussche, Megan E. Judkins, Michael J. Montague, and Wesley C. Warren

Source: Journal of Raptor Research, 51(3) : 368-377

Published By: Raptor Research Foundation

URL: https://doi.org/10.3356/JRR-16-47.1

J. Raptor Res. 51(3):368–377

A RESOURCE OF GENOME-WIDE SINGLE NUCLEOTIDE POLYMORPHISMS (SNPS) FOR THE CONSERVATION AND MANAGEMENT OF GOLDEN EAGLES

RONALD A. VAN DEN BUSSCHE[1]
Department of Integrative Biology, 501 Life Sciences West, Oklahoma State University, Stillwater, OK 74078 U.S.A.

MEGAN E. JUDKINS
Department of Integrative Biology, 501 Life Sciences West, Oklahoma State University, Stillwater, OK 74078 U.S.A.
and
Grey Snow Eagle House, Iowa Tribe of Oklahoma, Perkins, OK 74059 U.S.A.

MICHAEL J. MONTAGUE AND WESLEY C. WARREN
McDonnell Genome Institute, Washington University, St. Louis, MO 63108 U.S.A.

ABSTRACT.—Elucidating the genetic structure and ascertaining the natal origin of Golden Eagles (*Aquila chrysaetos*) are challenging for a number of reasons, including the lack of highly reproducible, variant genetic loci. Here, we developed a new high-quality Golden Eagle genome reference to serve as a computational atlas for future genetic investigations. We then generated unique genetic resources for the Golden Eagle by performing low-coverage genomic sequencing for 32 individuals ranging from Alaska to southern New Mexico and California to Nebraska. By aligning the reads from these 32 individuals to our Golden Eagle reference genome, we detected approximately 900,000 population variants in the form of Single Nucleotide Polymorphisms (SNPs). Using linkage disequilibrium and other quality filters, we next derived a set of 30,006 SNPs that were used to cluster our samples into three genetic groups. Although additional work is needed to fully characterize these loci, we provide a high-quality Golden Eagle genome reference and a comprehensive set of genetic markers for the conservation and management of Golden Eagles. Additionally, with a more comprehensive Golden Eagle genome assembly and associated transcriptomes, it is now possible to target specific genes or other biologically relevant regions for evaluating the effects of many anthropogenic stressors on Golden Eagle survival.

KEY WORDS: *Golden Eagle*; Aquila chrysaetos; *genetic structure*; *population genomics*; *SNPs*.

POLIMORFISMOS DE NUCLEÓTIDOS INDIVIDUALES A NIVEL DE GENOMA (PNI) PARA LA CONSERVACIÓN Y EL MANEJO DE *AQUILA CHRYSAETOS*

RESUMEN.—Dilucidar la estructura genética y determinar el origen natal de *Aquila chrysaetos* representa un desafío por numerosas razones, incluyendo la falta de loci genéticos variables altamente reproducibles. Desarrollamos un nuevo genoma de referencia de elevada calidad para *A. chrysaetos* para que sirva como un atlas computacional para futuras investigaciones genéticas. A continuación, generamos fuentes genéticas únicas para esta especie realizando secuenciaciones genómicas de baja cobertura para 32 individuos distribuidos entre Alaska y el sur de Nuevo México y entre California y Nebraska. Al alinear las lecturas de estos 32 individuos con nuestro genoma de referencia para *A. chrysaetos*, detectamos aproximadamente 900,000 variantes poblacionales en la forma de Polimorfismos de Nucleótidos Individuales (PNI). Utilizamos el desequilibrio de ligamiento y otros filtros de calidad y luego derivamos un conjunto de 30,006 PNI que fueron utilizados para agrupar nuestras muestras en tres grupos genéticos. Aunque se necesita trabajo adicional para caracterizar completamente estos loci, proporcionamos un genoma de referencia y un detallado conjunto de marcadores para la conservación y la gestión de *A. chrysaetos*. Adicionalmente, con un ensamble genómico y transcriptomas asociados más detallados, es posible ahora identificar genes específicos

[1] Email address: ron.van_den_bussche@okstate.edu

u otras regiones biológicamente relevantes para evaluar los efectos de las diversas presiones antrópicas sobre la supervivencia de *A. chrysaetos*.

[Traducción del equipo editorial]

Delineating the biologically relevant boundaries within a species' range should constitute the first step in any conservation or management program (Palsbøll et al. 2007). This critical step informs wildlife managers, biologists, and policy makers of the "units" they are attempting to conserve or manage while also setting the biological and theoretical foundations for future decisions (Funk et al. 2012). For the past three decades, microsatellite loci served as the genetic marker of choice for delineating population boundaries; however, recent advances in speed and accuracy, and rapidly decreasing costs have facilitated Next Generation Sequencing (NGS) approaches for non-model organisms (Allendorf et al. 2010, Davey et al. 2011, Helyar et al. 2011). NGS facilitates the genotyping by sequencing approach that provides unprecedented resolution of segregating single nucleotide polymorphisms (SNPs) spanning the entire genome (Krück et al. 2013, Larson et al. 2014), thereby catalyzing the shift from microsatellite loci to SNPs for population genomic studies (van Bers et al. 2010, Pujolar et al. 2013, Malenfant et al. 2015). Although microsatellite loci typically possess greater allelic diversity than SNPs, the latter offer several advantages, including a high abundance, regular distribution throughout the genome (i.e., SNPs occur in both coding and noncoding regions), high reproducibility and transferability within and among laboratories, low scoring error rates, and a simple model of nucleotide evolution (Morin et al. 2004, Kraus et al. 2015). Moreover, when utilizing large numbers of SNPs (1000 or more), as few as four individuals per population provide accurate estimates of population differentiation (F_{ST}), whereas for typical microsatellite studies, the number of individuals required to obtain accurate F_{ST} estimates ranges between 25 and 30 (Hale et al. 2012).

Numerous advances have also been made over the past decade with regard to population genetic analyses. For example, a variety of statistical analyses, including BAPS (Corander et al. 2004), BayesAss+ (Wilson and Rannala 2003), GeneClass (Piry et al. 2004), GENELAND (Guillot et al. 2005a, 2005b), and STRUCTURE (Pritchard et al. 2000), make use of multilocus genotypes to determine the number of clusters from which the samples under study were collected or to assign individuals to predetermined clusters. These programs are useful for determining population structure as well as for detecting contemporary migrants; when used with large numbers of SNPs, even weak genetic structure can be revealed. For example, 586 American lobsters (*Homarus americanus*) from 17 locations were genotyped at 10,156 SNPs, revealing hierarchical genetic structure (Benestan et al. 2015). Not only could these loci separate northern lobsters from southern lobsters, but they also revealed 11 distinct populations and provided strong evidence for fine-scale genetic structuring within each region. Similarly, 494 eulachon (*Thaleichthys pacificus*) were collected from 12 sites and genotyped at 4104 SNP loci (Candy et al. 2015). Of these loci, 193 were putatively adaptive, so population differentiation was assessed using adaptive loci as well as neutral loci. Levels of population differentiation were similar based on either the 3911 neutral SNPs or the 193 putatively adaptive SNPs. However, the putatively adaptive SNPs provided greater resolution of stocks than the 3911 neutral SNPs. The authors proposed putative divergent selective pressures in the different freshwater and marine environments that acted on the regional populations of eulachon. More importantly, they posited that these adaptive differences should be given strong consideration when delineating genetic boundaries for conservation purposes. This study, as well as others that assessed Atlantic herring (*Clupea harengus*; Lamichhaney et al. 2012) and Pacific lamprey (*Entosphenus tridentatus*; Hess et al. 2013), highlights the potential for detecting adaptive variation using SNPs and represents an advantage over microsatellites. In fact, adaptive variation may be important for delineating biologically relevant units for management and conservation (Funk et al. 2012).

Delineating population boundaries of Golden Eagles (*Aquila chrysaetos*), if they exist, has proved challenging, due to Golden Eagles' capacity for long-distance movements, especially among immature individuals and during periods of nonbreeding. As Golden Eagles face increased anthropogenic pressures (e.g., wind energy development, electrocution, lead poisoning, habitat alteration and loss), it is paramount that biologically meaningful population

boundaries be determined for their proper management and conservation (U.S. Fish and Wildlife Service 2009). Until recently, Golden Eagles in North America were managed using Golden Eagle Management Units that approximate the Bird Conservation Regions established by the North American Bird Conservation Initiative (U.S. Fish and Wildlife Service 2009, 2013). The correspondence of Golden Eagle Management Units and Bird Conservation Regions was based on the estimated natal dispersal distance for Golden Eagles (Millsap et al. 2014).

In one of the first studies to use genetic loci to evaluate genetic structure in North American Golden Eagles, Doyle et al. (2016) identified 159 autosomal SNPs by comparing a previous Golden Eagle genome assembly (Doyle et al. 2013) with our unpublished genome (described herein) of Golden Eagles, each bird from an unknown natal origin. They genotyped 160 eaglets from known provenances at these 159 SNPs, a sex-linked SNP, and two mitochondrial SNPs. Their primary objective was to test the null hypothesis of panmixia, which was rejected due to the presence of three genetically defined groups: Alaska, California, and a cluster containing individuals from Arizona, Colorado, Nebraska, New Mexico, Utah, and Wyoming (Doyle et al. 2016).

Prior to the results of Doyle et al. (2016), the U.S. Fish and Wildlife Service began considering two alternative frameworks for the management of Golden Eagles: Golden Eagle Management Units as described above or Flyway Eagle Management Units which follow the administrative flyways (U.S. Fish and Wildlife Service 2016). Although Doyle et al. (2016) developed a suite of 159 autosomal SNPs, it may be necessary to examine thousands of loci to accurately define population boundaries and assign individuals to natal areas, particularly for species with high dispersal capabilities. Other studies have required greater numbers of SNPs: lobsters (Benestan et al. 2015), eulachon (Candy et al. 2015), herring (Lamichhaney et al. 2012), lamprey (Hess et al. 2013), killer whales (*Orcinus orca*; Moura et al. 2014), European wolves (*Canis lupus*; Pilot et al. 2014), polar bears (*Ursus maritimus*; Malenfant et al. 2015), House Sparrow (*Passer domesticus*; Hagen et al. 2013), European eel (*Anguilla anguilla*; Pujolar et al. 2013), rainbow trout (*Oncorhynchus mykiss*; Palti et al. 2015) and Great Tit (*Parus major*; van Bers et al. 2010). The objective of this study was to evaluate whether a larger suite of SNPs, derived from a sample with more geographic representation, would provide greater power for delineating population boundaries of Golden Eagles than those in Doyle et al. (2016). To address this objective, we sequenced the genome and transcriptomes from three tissue types of one Golden Eagle and performed low-coverage genome sequencing of 32 individual eagles. We then used these data to ascertain 30,006 SNPs for preliminary population genomic analyses and to evaluate their potential for determining population boundaries for North American Golden Eagles.

METHODS

Whole Genome Sequencing and Assembly. We extracted DNA from a blood sample from a male eagle with the unique identification number of GSEH35GE (tissue source courtesy of the Iowa Tribe of Oklahoma's Grey Snow Eagle House). Our sequencing technique followed the recommendations provided in the ALLPATHS-LG assembler (Butler et al. 2008). This model requires 45× sequence coverage of each fragment (overlapping paired reads approximately 180 bp length) from 3 kb paired end (PE) reads, as well as 5× coverage of 8 kb PE reads. All sequences were generated on the HiSeq2000 Illumina instrument (Illumina, Inc., San Diego, CA U.S.A.). Total assembled sequence coverage was ~88× (overlapping reads, 3 kb and 8 kb PE reads) using a genome size estimate of 1.2 Gb. All sequences were assembled using ALLPATHS-LG software (Butler et al. 2008). Finally, contaminating contigs, adaptors, ambiguous bases (i.e., Ns) in the sequence and all contigs <200 bp were removed prior to submission. This genome assembly referred to as Aquila_chrysaetos-1.0.2 is available for download using GenBank accession number JRUM00000000.1.

Gene Annotation. The Aquila_chrysaetos-1.0.2 assembly was annotated using the NCBI pipeline (Pruitt et al. 2014), including masking of repeats prior to *ab initio* gene predictions, for evidence-supported gene model building. An injured female Golden Eagle was brought to the Grey Snow Eagle House and a veterinarian determined that due to the nature of the injuries, the individual needed to be euthanized. Tissue samples (whole brain, muscle, liver) were obtained immediately after euthanasia and placed in RNALater and total RNA was isolated. Whole brain, muscle, and liver using the RNA Easy Plus Mini Kit with the manufacturer-supplied protocol and a double elution during the last step

(Qiagen, Redwood City, CA). From these pools of RNA, we generated tissue-specific cDNA libraries and sequenced each indexed library on a HiSeq2000 instrument. A final set of merged transcripts was generated after several filters were used to remove pseudogenes, noncoding RNA, and several other possible contaminating data sources. These RNA-Seq data were used to further improve gene model accuracy by alignment to nascent gene models to delineate boundaries of untranslated regions, as well as to identify genes not found through interspecific similarity evidence from other species. A full description of the NCBI gene annotation pipeline was previously described (Pruitt et al. 2014).

Population Sampling and Sequencing. We collaborated with field biologists permitted by the U.S. Fish and Wildlife Service to collect blood samples from nestling eagles. For the initial isolation and characterization of SNP loci, blood was collected from eaglets in Alaska ($n = 5$), Arizona ($n = 1$), British Columbia ($n = 3$), California ($n = 3$), Colorado ($n = 2$), Idaho ($n = 3$), Nebraska ($n = 3$), New Mexico ($n = 1$), Oklahoma ($n = 1$), Oregon ($n = 3$), and Wyoming ($n = 7$). Blood (0.5 ml) was placed in 5 ml of lysis buffer (Longmire et al. 1997) and shipped to our laboratory at Oklahoma State University, where we isolated whole genomic DNA following standard protocols (Longmire et al. 1997). We shipped approximately 5 μg of genomic DNA to the McDonnell Genome Institute at Washington University, St. Louis, MO, for DNA sequencing. We used a standardized input of 200 ng of DNA from each individual to construct Illumina PCR-free TruSeq Nano libraries according to the manufacturer's protocol, and sequenced each library on an Illumina HiSeq2000 instrument to generated 100 bp length paired end reads.

Mapping the Reads and SNP Detection. For each of the 32 Golden Eagles, we aligned their sequence reads to the Golden Eagle reference assembly using BWA-MEM (Li and Durbin 2009). The Golden Eagle reference assembly was downloaded from NCBI (http://www.ncbi.nlm.nih.gov/assembly/GCF_000766835.1). All reads were preprocessed to remove duplicate reads using Picard v.1.113 (The Picard toolkit. http://picard.sourceforge.net). After alignment, we identified base position differences between the reference assembly and the re-sequenced samples using the convergent outcomes of the software SAMtools (Li et al. 2009) and VarScan 2 (Koboldt et al. 2012). Parameters included a P value of 0.1, a map quality of 10, a minimum coverage of three reads, a minimum variant frequency of 0.2 and parameters for filtering by false positives, including a minimum variant read count of three reads and a bam read count minimum base quality of 15.

Population Structure Analysis. To perform a preliminary analysis of population structure based on these 32 individuals, a backfilled, multi-sample genotype file in vcf format was parsed to include the 26 longest scaffolds, representing 38.2% of the assembly. Next, the variant sites within these scaffolds were evaluated for data quality and pruned using PLINK (Purcell et al. 2007). We first performed linkage disequilibrium (LD) based SNP pruning using the method that relies on pairwise genotypic correlation. For each window of 50 SNPs, we calculated LD between each pair of SNPs in the window, and removed one of a pair of SNPs if the LD was greater than 0.2, and then shifted the window five SNPs forward to repeat the procedure. We next removed all SNPs missing in more than 30% of the Golden Eagles, as well as all SNPs with <5% minor allele frequency (MAF). We also excluded any position that failed the Hardy-Weinberg test at the default significance threshold ($P < 0.001$). Following quality-control filtering, a total of 30,006 autosomal SNPs remained. We performed model-based clustering with STRUCTURE (Pritchard et al. 2000). We ran five replicates for each of K = 1 through K = 8 using the admixture model and correlated allele frequencies (Falush et al. 2003), and each independent analysis for 50,000 iterations following a burn-in period of 50,000 iterations. We merged resulting Q output files for each replicate at each K with the LargeKGreedy method (with random input orders) using the program CLUMPP (Jakobsson and Rosenberg 2007). This allowed us to output the final proportions of each K for each Golden Eagle, and we plotted the results manually for the most optimal value of K. Results from STRUCTURE were interpreted using mean likelihood values of K and ΔK. We verified the genetic relationships among the samples using common measurements for pairwise relatedness, as implemented in VCFtools (Danecek et al. 2011).

RESULTS

Whole Genome Sequencing, Assembly, and Annotation. To facilitate the discovery of neutral and adaptive loci, we sequenced and assembled a Golden Eagle genome and the transcriptomes from brain, liver, and muscle. Our Golden Eagle genome

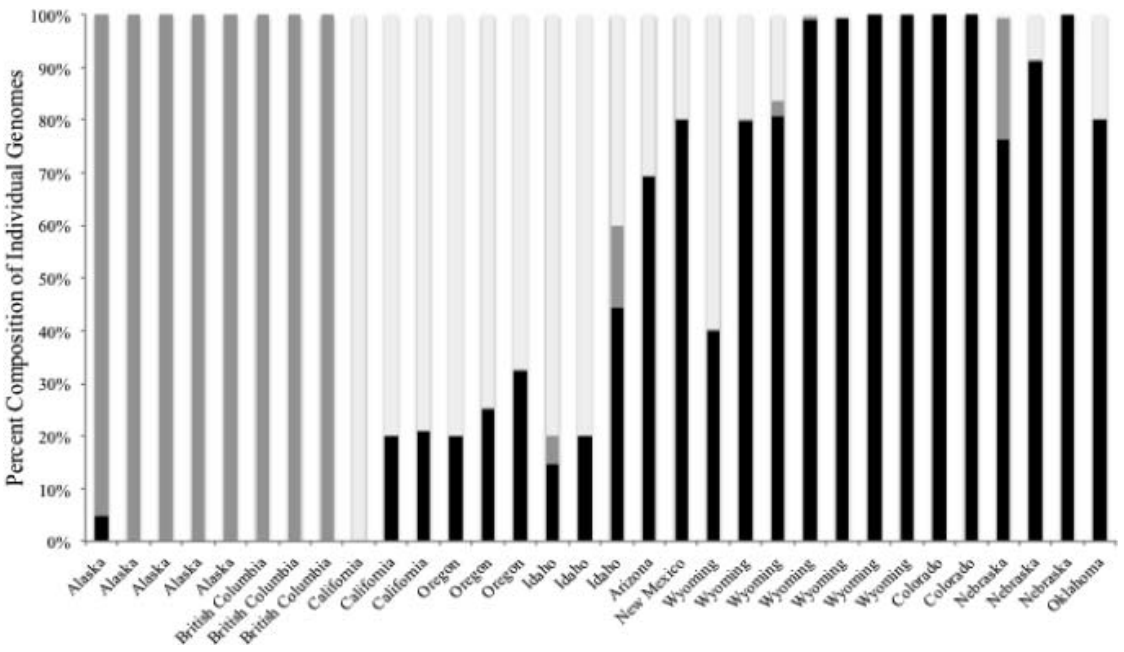

Figure 1. Results from STRUCTURE analysis showing the three genetically defined groups of Golden Eagles based on our analysis of 30,006 Single Nucleotide Polymorphisms and 32 individuals. Individual histograms represent the percentage of each individual's genome that was assigned to each of the three genetic units. The Alaska-British Columbia group is colored gray, California-Oregon-Idaho group is white, and the Arizona-Colorado-Nebraska-New Mexico-Oklahoma-Wyoming group is black.

(Aquila_chrysaetos-1.0.2) had an assembly coverage of ~88×, assembled to 1.07 Gb with N50 contig and scaffold lengths of 172 kb and 9.2 Mb, respectively, and 17,032 contigs. We comprehensively annotated the Aquila_chrysaetos-1.0.2 assembly, taking advantage of species-specific transcript evidence, i.e., RNAseq data and predicted 17,291 nuclear genes of which 15,658 are protein-coding. Our protein-coding and noncoding gene counts were similar to other published bird genomes and a comprehensive report of gene annotation was posted on the National Center for Biotechnology Information website: http://www.ncbi.nlm.nih.gov/nuccore/NW_011950869.1.

Preliminary Population Analysis. We generated an average of 53.2 million reads (100 bp) from each of 32 Golden Eagles and identified a mean of 1,849,225 variants per individual, with the average read depth at these variant sites calculated at over 5×. Variants were assessed at a total of 4,750,603 sites, which passed the thresholds using our initial variant-calling methods with SAMtools (Li et al. 2009) and VarScan 2 (Koboldt et al. 2012). The samples ranged from 4.4 to 6.0 fold in mean sequence coverage across variant sites. We identified the following numbers of high-quality variants: (1) heterozygous SNPs ranging from 207,601 to 423,749 among all samples, with a mean of 296,426 heterozygous sites per individual; (2) homozygous variant SNPs ranging from 476,321 to 621,634 among all samples, with a mean of 555,379 homozygous variant sites per individual; and (3) unique or singleton SNPs ranging from 31,832 to 59,661 among all samples, with a mean of 45,169 per individual. The mean transition to transversion ratio was 2.53, suggesting very low influence of sequencing error on SNP calling (Pujolar et al. 2013). The mean inbreeding coefficient was 0.06 (range = –0.27 to 0.37), while the mean unadjusted A_{jk} pairwise relatedness (Yang et al. 2010) was –0.03 (range = –0.01 to –0.06). An analysis of population structure detected three genetically distinct groups (K = 3, mean average ln P(X|K) = –661829.96) corresponding to (1) Alaska and British Columbia, (2) California, Oregon, and Idaho, and (3) Arizona, New Mexico, Wyoming, Colorado, Oklahoma, and Nebraska (Fig. 1, 2).

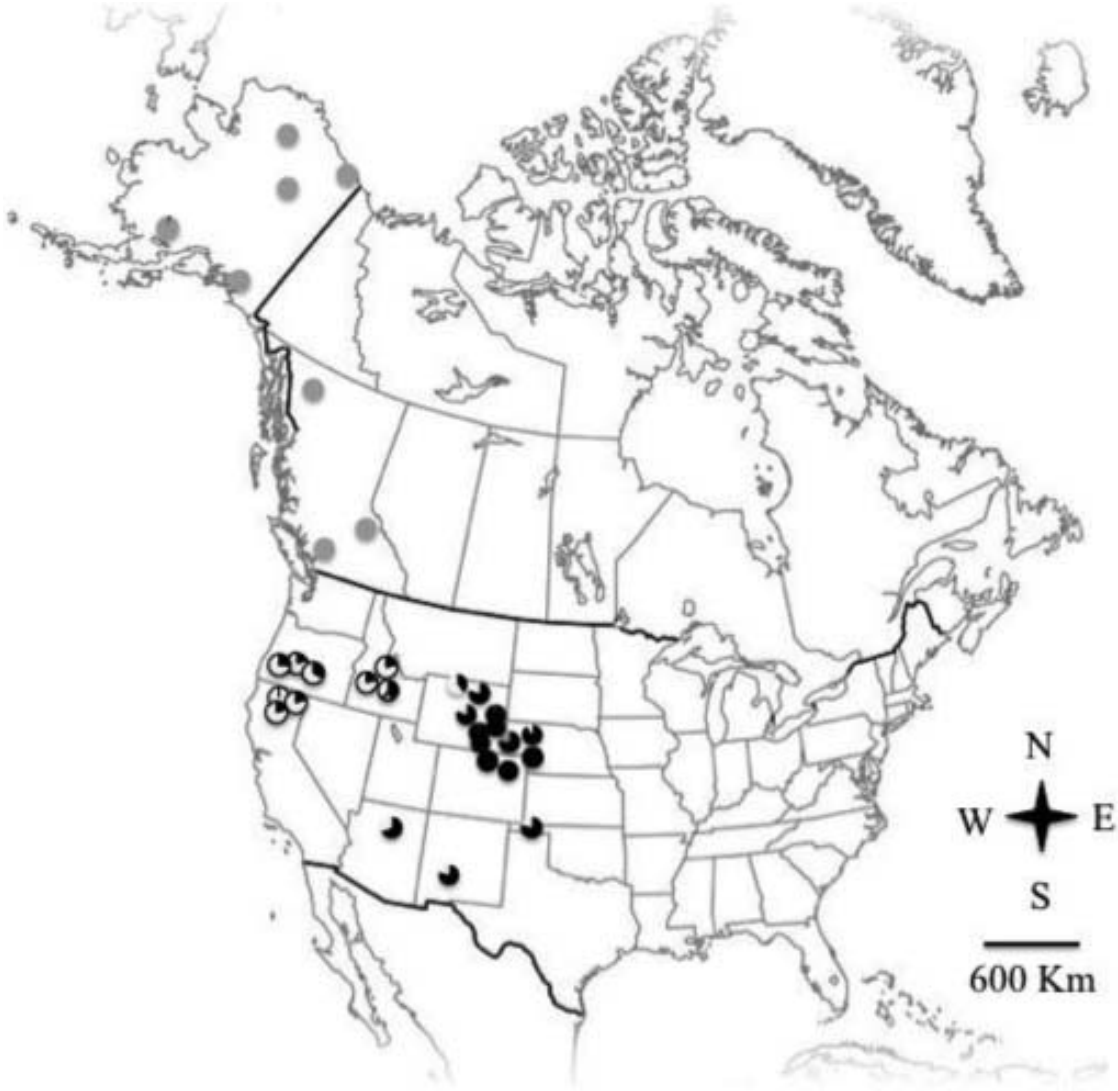

Figure 2. Spatially explicit population structure of our sample of 32 Golden Eagles. The colors in the pie charts reflects the relative posterior probability of membership of the individual's genome belonging to each of the three genetic groups identified by STRUCTURE with the Alaska-British Columbia, California-Oregon-Idaho, and Arizona-Colorado-Nebraska-New Mexico-Oklahoma-Wyoming groups represented by gray, white, and black, respectively.

Discussion

Development of Genomic Resources. To facilitate our objective of discovering a large suite of genetic loci from the Golden Eagle, we sequenced and assembled the genome of a male Golden Eagle and the transcriptomes from brain, liver, and muscle. Doyle et al. (2013) published the first genome assembly of a Golden Eagle (AquilaChrysaetos1) with an assembled size of 1.17 GB and continuity metrics of N50 contig and scaffold length estimated at 16 kb and 1.74 Mb, respectively. Our Golden Eagle genome (Aquila_chrysaetos-1.0.2) assembled to 1.07 GB with N50 contig and scaffold length of 172 kb and 9.2 Mb, respectively, represents a substantial improvement in contiguity. Importantly, the total number of assembly gaps is estimated by the total number of scaffolds minus the expected number of chromosomes. Ideally for Golden Eagles, which have

a diploid number of 62 chromosomes (Masuda et al. 1998), there would be 30 scaffolds to represent each of the haploid autosomes and single scaffolds for each sex chromosome for a female (ZW), with each scaffold defined as an ordered and oriented set of contigs (called consensus bases with no intervening gaps). Our assembly has 1142 scaffolds versus 42,881 for AquilaChrysaetos1 (Doyle et al. 2013). For future efforts to assign scaffolds to chromosomes, this clearly shows the advantage of our less fragmented assembly.

Doyle et al. (2013) were able to predict 16,571 nuclear genes using comparative evidence from closely-related bird species. Using transcriptomes from three tissues types, we were able to annotate the Aquila_chrysateos-1.0.2 assembly and predicted 17,291 nuclear genes, of which 15,658 are protein-coding. Thus, gene representation, both protein-coding and noncoding, is more comprehensive in the Aquila_chrysaetos-1.0.2 assembly for all metrics (see http://www.ncbi.nlm.nih.gov/nuccore/NW_011950869.1 for other measures).

Preliminary Population Analyses. To provide a preliminary assessment of the levels of genetic structure among the 32 Golden Eagles in this study, we performed population genomic analyses using 30,006 SNPs. STRUCTURE indicated that these 32 individuals best represent three genetically defined groups as follows (Fig. 1, 2): Alaska and British Columbia (AK-BC group), northern California, southern Oregon, and southern Idaho (CA-OR-ID group), and a group containing individuals from Arizona, Colorado, Nebraska, New Mexico, Oklahoma, and Wyoming (AZ-CO-NE-NM-OK-WY group). The most genetically homogeneous group contained five individuals from Alaska and three individuals from British Columbia, with a mean percent level of admixture calculated at 0.6%. For example, the genetic signature from one Alaskan bird shows 99.5% probability of belonging to the AK-BC group with the remaining 0.5% apportioned to AZ-CO-NE-NM-OK-WY group. A second individual also represents an admixed genome with a 99.9% assignment to the AK-BC group and the remaining 0.1% representing admixture from the AZ-CO-NE-NM-OK-WY group. The SNPs from the remaining six individuals showed no detectable admixture from the other two genetic groups (Fig. 2).

The CA-OR-ID group contains nine individuals, with the mean level of group membership calculated at 75.6% (range: 40.15–100%; Fig. 1). Thus, with one exception, individuals within this group were characterized as admixed and most of the extra-group was apportioned to the AZ-CO-NE-NM-OK-WY group. Three individuals showed evidence of admixture from all three genetic groups (Fig. 1, 2). Interestingly, the genotypes for one nestling from southern Idaho produced a highly admixed signal consisting of 44.4% AZ-CO-NE-NM-OK-WY, 15.5% AK-BC, and 40.2% CA-OR-ID. STRUCTURE assigned this individual to the AZ-CO-NE-NM-OK-WY group, but due to nest locality, we placed it within the CA-OR-ID group.

Mean percent membership within the AZ-CO-NE-NM-OK-WY group was 89.7%, which ranged from >99.9% for five individuals to a low of 69.3% (Fig. 1). Individuals within this cluster represent the most geographically widespread group (Fig. 2). Within this group, two individuals from Wyoming and Colorado as well as one individual from Nebraska showed no evidence of admixture. Moreover, two individuals from Wyoming possessed genotypes that were 99% characteristic of this group with historical influence from the AK-BC group (one individual) or the CA-OR-ID group (one individual). The remainder of the individuals possessed genotypes characteristic of this group but with genetic influence from one or more of the AK-BC and CA-OR-ID groups (Fig. 1, 2).

Although Golden Eagles generally settle for breeding within 46.4 km of their natal origin (Millsap et al. 2014), the results from our analysis suggest some low level of gene flow among Golden Eagles in the contiguous states. It may also be more common for Golden Eagles from the AK-BC group to contribute genes to populations in the lower 48 states than the reverse, resulting in asymmetric gene flow. This low level of gene flow is sufficient to reduce inbreeding and maintain genetically healthy populations. Average inbreeding coefficients (F) for individuals within the AK-BC, CA-OR-ID, and AZ-CO-NE-NM-OK-WY groups were –0.072, 0.053, and 0.130, respectively.

There are now two preliminary population genomic studies on North American Golden Eagles. Although these studies differ in number of loci (n = 159 [Doyle et al. 2016] vs. n = 30,006 [this study]) and number of samples (n = 160 [Doyle et al. 2016] vs. n = 32 [this study]), both studies detected three, relatively congruent genetically defined groups using STRUCTURE. Doyle et al. (2016) detected a northern group of 24 Golden Eagles from Alaska, whereas our corresponding group consisted of five individuals from Alaska and three from British

Columbia. The average probability of individuals being assigned to this group was 0.77 ± 0.15 (Doyle et al. 2016) and 0.99 ± 0.02 (this study). The second population cluster, as detected by Doyle et al. (2016), consisted of 117 Golden Eagles that represented a north-to-south distribution of individuals from California, whereby the average probability of individuals assignment was 0.71 ± 0.20 (Doyle et al. 2016). Likewise, our second population cluster consisted of nine individuals from northern California ($n = 3$), southern Oregon ($n = 3$) and southern Idaho ($n = 3$) with a similar average probability of group assignment (0.76 ± 0.15). The third genetically defined group, as detected by Doyle et al. (2016) consisted of 113 Golden Eagles sampled in the western states of Arizona, Colorado, Nebraska, New Mexico, Utah, and Wyoming. The average probability of individual assignment to this cluster was 0.50 ± 0.24. For our analysis, the third genetically defined group consisted of individuals from Arizona, Colorado, Nebraska, New Mexico, Oklahoma, and Wyoming. The average probability of individual assignment to our group of western states was 0.87 ± 0.13.

Implications for Future Studies. Being an apex predator, Golden Eagles are an important component of the ecosystem of western North America and therefore a concern from the conservation and management perspectives. Our higher continuity assembly will facilitate easier assignment of sequences to chromosomes, a necessary next step if we are to discover selection signatures in Golden Eagle compared to other avian apex predators. Based on our reference assembly activities, using whole genome sequencing from one male Golden Eagle, coupled with transcriptome sequencing from three tissue types, we were able to align additional low coverage (~5×) whole genome sequencing reads from 32 Golden Eagles to generate a genomic resource of 896,974 SNPs. By utilizing a subset of 30,006 SNPs from these 32 individuals we were able to show that overall, the western Golden Eagle gene pool possesses sufficient genetic variation and that this variation is partitioned into three genetic groups. Although our analyses detected three genetically defined groups, at this time these groupings should be viewed cautiously as these analyses are based on relatively uncharacterized SNPs for only 32 individuals with limited geographic sampling.

Our preliminary results, as well as results from several other studies of wide-ranging species that assayed tens of thousands of SNPs (e.g., Lamichhaney et al. 2012, Hagen et al. 2013, Benestan et al. 2015, Malenfant et al. 2015), suggest the value in further characterizing a larger suite of SNPs and developing a Golden Eagle SNP Chip. As such, the full set of population SNPs that we identified (approximately 900,000) will be thoroughly evaluated to obtain a subset of markers capable of unequivocally assigning Golden Eagles with unknown population associations to management units in a cost-effective way. For example, we could employ such markers on samples from wounded individuals to determine their natal origins. Additionally, these markers could be used on migratory individuals collected at trapping stations to better understand which populations are utilizing the various flyways. Finally, not only will these loci be important for conservation/management initiatives, but they will also be utilized to provide additional biological insight, including management of genetic diversity, effective population size, levels and direction of gene flow, inbreeding, and pairwise relatedness. Without this genetic information, such factors will be difficult to ascertain. A better understanding of these characteristics will advance our ability to protect this species in the face of climate variability and other anthropogenic stressors. The generation of such a large panel of novel SNPs represents an important step in terms of genomic resources available for this species and will facilitate additional population genomics studies in the future.

Acknowledgments

This project was funded by the Iowa Tribe of Oklahoma, including a contribution to the Iowa Tribe of Oklahoma from the Shakopee Mdewakanton Sioux Tribe in Minnesota for this work. Megan Judkins was supported by the Ford Foundation during this work. We thank Victor Roubidoux, Aviary Manager of the Iowa Tribe of Oklahoma's Grey Snow Eagle House for his continued support of our research on Bald and Golden eagles. We also thank all the biologists who provided blood samples for this study, especially Rob Domenech (Raptor View Research Institute), Garth Herring (U.S. Geological Survey), Brian Millsap (U.S. Fish and Wildlife Service), Gary Roemer (New Mexico State University), and Brian Smith (U.S. Fish and Wildlife Service). This work was conducted under U.S. Fish and Wildlife permits to the Iowa Tribe of Oklahoma's Grey Snow Eagle House (EASC MB109421) and Brian Millsap of U.S. Fish and Wildlife Service (EASC MB58285B). This article was based on a presentation given at the Golden Eagles in a Changing World Symposium held at the 2015 Raptor Research Foundation Annual Meeting in Sacramento, CA. We thank the Academy for the Environment, University of Nevada, Reno, for providing financial support for publication costs. We also thank the U.S. Fish and Wildlife Service, especially the Western

Golden Eagle Team, for their financial and logistical support of this work.

LITERATURE CITED

ALLENDORF, F.W, P.A. HOHENLOHE, AND G. LUIKART. 2010. Genomics and the future of conservation genetics. *Nature Reviews Genetics* 11:697–709.

BENESTAN, L., T. GOSSELIN, C. PERRIER, B. SAINTE-MARIE, R. ROCHETTE, AND L. BERNATCHEZ. 2015. RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (*Homarus americanus*). *Molecular Ecology* 24:3299–3315.

BUTLER, J., I. MACCALLUM, M. KLEBER, I.A. SHLYAKHTER, M.K. BELMONTE, E.S. LANDER, C. NUSBAUM, AND D.B. JAFFE. 2008. ALLPATHS: de novo assembly of whole-genome shotgun microreads. *Genome Research* 18:810–820.

CANDY, J.R., N.R. CAMPBELL, M.H. GRINNELL, T.D. BEACHAM, W.A. LARSON, AND S.R. NARUM. 2015. Population differentiation determined from putative neutral and divergent adaptive genetic markers in Eulachon (*Thaleichthys pacificus*, Osmeridae), an anadromous Pacific smelt. *Molecular Ecology Resources* 15:1421–1434.

CORANDER, J., P. WALDMANN, P. MARTTINEN, AND M.J. SILLANPÄÄ. 2004. BAPS 2: enhanced possibilities for the analysis of genetic population structure. *Bioinformatics* 20:2363–2369.

DANECEK, P., A. AUTON, G. ABECASIS, C.A. ALBERS, E. BANKS, M.A. DEPRISTO, R.E. HANDSAKER, G. LUNTER, G.T. MARTH, S.T. SHERRY, G. MCVEAN, R. DURBIN, AND 1000 GENOMES PROJECT ANALYSIS GROUP. 2011. The variant call format and VCFtools. *Bioinformatics* 27:2156–2158.

DAVEY, J.W., P.A. HOHENLOHE, P.D. ETTER, J.Q. BOONE, J.M. CATCHEN, AND M.L. BLAXTER. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* 12:499–510.

DOYLE, J.M., T.E. KATZNER, P.H. BLOOM, Y. JI, B.K. WIJAYAWARDENA, AND J.A. DEWOODY. 2013. The genome sequence of a widespread apex predator, the Golden Eagle (*Aquila chrysaetos*). *PLoS One* 9: e95599. Doi:10.1371/journal.pone.0095599.

———, ———, ———, G.W. ROEMER, J.W. CAIN, III, B. MILLSAP, C. MCINTYRE, S.A. SONSTHAGEN, N. FERNANDEZ, M. WHEELER, Z. BULUT, P.H. BLOOM, AND J.A. DEWOODY. 2016. Genetic structure and viability selection in the Golden Eagle (*Aquila chrysaetos*), a vagile raptor with a Holarctic distribution. *Conservation Genetics* 17:1–16.

FALUSH, D., M. STEPHENS, AND J.K. PRITCHARD. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164:1567–1587.

FUNK, W.C., J.K. MCKAY, P.A. HOHENLOHE, AND F.W. ALLENDORF. 2012. Harnessing genomics for delineating conservation units. *Trends in Ecology and Evolution* 27:489–496.

GUILLOT, G., A. ESTOUP, F. MORTIER, AND J.F. COSSON. 2005b. A spatial statistical model for landscape genetics. *Genetics* 170:1261–1280.

———, F. MORTIER, AND A. ESTOUP. 2005a. GENELAND: A computer package for landscape genetics. *Molecular Ecology Notes* 5:712–715.

HAGEN, I.J., A.M. BILLING, B. RONNING, S.A. PEDERSEN, H. PARN, J. SLATE, AND H. JENSEN. 2013. The easy road to genome-wide medium density SNP screening in a non-model species: development and application of a 10 K SNP-chip for the House Sparrow (*Passer domesticus*). *Molecular Ecology Resources* 13:429–439.

HALE, M.L., T.M. BURG, AND T.E. STEEVES. 2012. Sampling for microsatellite-based population genetic studies: 25–30 individuals per population is enough to accurately estimate allele frequencies. *PLoS ONE* 7:e45170.

HELYAR, S.J., J. HEMMER-HANSEN, D. BEKKEVOLD, M.I. TAYLOR, R. OGDEN, M.T. LIMBORG, A. CARIANI, G.E. MAES, E. DIOPERE, G.R. CARVALHOO, AND E.E. NIELSEN. 2011. Application of SNPs for population genetics of non-model organisms: new opportunities and challenges. *Molecular Ecology Resources* 11 Suppl 1:123–136.

HESS, J.E., N.R. CAMPBELL, D.A. CLOSE, M.F. DOCKER, AND S.R. NARUM. 2013. Population genomics of Pacific lamprey: adaptive variation in a highly dispersive species. *Molecular Ecology* 22:2898–2916.

JAKOBSSON, M. AND N.A. ROSENBERG. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23:1801–1806.

KOBOLDT, D.C., Q. ZHANG, D.E. LARSON, D. SHEN, M.D. MCLELLAN, L. LIN, C.A. MILLER, E.R. MARDIS, L. DING, AND R.K. WILSON. 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research* 22:568–576.

KRAUS, R.H.S., B. VONHOLDT, B. COCCHIARARO, V. HARMS, H. BAYERL, R. KÜHN, D.W. FORSTER, J. FICKEL, C. ROOS, AND C. NOWAK. 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.

KRÜCK, N.C., D.I. INNES, AND J.R. OVENDEN. 2013. New SNPs for population genetic analysis reveal possible cryptic speciation of eastern Australian sea mullet (*Mugil cephalus*). *Molecular Ecology Resources* 13:715–725.

LAMICHHANEY, S., Á.M. BARRIO, N. RAFATI, G. SUNDSTRÖM, C.-J. RUBIN, E.R. GILBERT, J. BERGLUND, A. WETTERBOM, L. LAIKRE, M.T. WEBSTER, M. GRABHERR, N. RYMAN, AND L. ANDERSSON. 2012. Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. *Proceedings of the National Academy of Science* 109:19345–19350.

LARSON, W.A., L.W. SEEB, M.V. EVERETT, R.K. WAPLES, W.D. TEMPLIN, AND J.E. SEEB. 2014. Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (*Oncorhynchus tshawytscha*). *Evolutionary Applications* 7:355–369.

LI, H. AND R.R. DURBIN. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.

———, B. HANDSAKER, A. WYSOKER, T. FENNELL, J. RUAN, N. HOMER, G. MARTH, G. ABECASIS, R. DURBIN, AND 1000 GENOME PROJECT DATA PROCESSING SUBGROUP. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.

LONGMIRE, J.L., M. MALTBIE, AND R.J. BAKER. 1997. Use of "lysis buffer" in DNA isolation and its implication for museum collections. *Occasional Papers, Museum of Texas Tech University* 163:1–3.

MALENFANT, R.M., D.W. COLTMAN, AND C.S. DAVIS. 2015. Design of a 9K illumine BeadChip for polar bears (*Ursus maritimus*) from RAD and transcriptome sequencing. *Molecular Ecology Resources* 15:587–600.

MASUDA, R., M. NORO, N. KUROSE, C. NISHIDA-UMEHARA, H. TAKECHI, T. YAMAZAKI, M. KOSUGE, AND M.C. YOSHIDA. 1998. Genetic characteristics of endangered Japanese Golden Eagles (*Aquila chrysaetos japonica*) based on mitochondrial DNA D-loop sequences and karyotypes. *Zoo Biology* 17:111–121.

MILLSAP B.A., A.R. HARMATA, D.W. STAHLECKER, AND D.G. MIKESIC. 2014. Natal dispersal distance of Bald and Golden eagles originating in the coterminous United States as inferred from band encounters. *Journal of Raptor Research* 48:13–23.

MORIN, P.A., G. LUIKART, AND R.K. WAYNE. 2004. The SNP Workshop Group. SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution* 19:208–216.

MOURA, A.E., J.G. KENNY, R. CHAUDHURI, M.A. HUGHES, A.J. WELCH, R.R. REISINGER, P.J. NICO DE BRUYN, M.E. DAHLHEIM, N. HALL, AND A.R. HOELZEL. 2014. Population genomics of the killer whale indicates ecotype evolution in sysmpatry involving both selection and drift. *Molecular Ecology* 23:5179–5192.

PALSBØLL, P.J., M. BÉRUBÉ, AND F.W. ALLENDORF. 2007. Identification of management units using population genetic data. *Trends in Ecology and Evolution* 22:11–16.

PALTI, Y., G. GAO, S. LIU, M.P. KENT, S. LIEN, M.R. MILLER, C.E. REXROAD, III, AND T. MOEN. 2015. The development and characterization of a 57K single nucleotide polymorphism array for rainbow trout. *Molecular Ecology Resources* 15:662–672.

PILOT, M., C. GRECO, B.M. VONHOLDT, B. JEDRZEEWSKA, E. RANDI, W. JEDRZEJEWSKI, V.E. SIDOROVICH, E.A. OSTRANDER, AND R.K. WAYNE. 2014. Genome-wide signatures of populaton bottlenecks and diversifying selection in European wolves. *Heredity* 112:428–442.

PIRY, S., A. ALAPETITE, J.-M. CORNUET, D. PAETKAU, L. BAUDOUIN, AND A. ESTOUP. 2004. GENECLASS2: a software for genetic assignment and first-generation migrant detection. *Journal of Heredity* 95:536–539.

PRITCHARD, J.K., M. STEPHENS, AND P. DONNELLY. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.

PRUITT, K.D., G.R. BROWN, S.M. HIATT, F. THIBAUD-NISSEN, A. ASTASHYN, O. ERMOLAEVA, C.M. FARRLELL, J. HART, M.J. LANDRUM, K.M. MCGARVEY, M.R. MURPHY, N.A. O'LEARY, S. PUJAR, B. RAJPUT, S.H. RANGWALA, L.D. RIDDICK, A. SHKEDA, H. SUN, P. TAMEZ, R.E. TULLY, C. WALLIN, D. WEBB, J. WEBER, W. WU, M. DICUCCIO, P. KITTS, D.R. MAGLOTT, T.D. MURPHY, AND J.M. OSTELL. 2014. RefSeq: an update on mammalian reference sequences. *Nucleic Acids Research* 42(Database issue):D756–D763.

PUJOLAR, J.M., M.W. JACOBSEN, J. FRYDENBERG, T.D. ALS, P.F. LARSEN, G.E. MAES, L. ZANE, J.B. HIAN, L. CHENG, AND M.M. HANSEN. 2013. A resource of genome-wide single-nucleotide polymorphisms generated by RAD tag sequencing in the critically endangered European eel. *Molecular Ecology Resources* 13:706–714.

PURCELL, S., B. NEALE, K. TODD-BROWN, L. THOMAS, M.A.R. FERREIRA, D. BENDER, J. MALLER, P. SKLAR, P.I.W. DE BAKKER, M.J. DALY, AND P.C. SHAM. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* 81:559–575.

U.S. FISH AND WILDLIFE SERVICE. 2009. Final environmental assessment. Proposal to permit take provided under the Bald and Golden Eagle Protection Act. U.S.D.I. Fish and Wildlife Service, Division of Migratory Bird Management, Washington, DC U.S.A.

———. 2013. Eagle Conservation Plan guidance: Module 1—land-based wind energy. Version 2. U.S.D.I. Fish and Wildlife Service, Division of Migratory Bird Management, Washington, DC U.S.A. http://www.fws.gov/migratorybirds/pdf/management/eagleconservationplanguidance.pdf (last accessed 7 April 2017).

———. 2016. Bald and Golden eagles: population demographics and estimation of sustainable take in the United States, 2016 update. U.S.D.I. Fish and Wildlife Service, Division of Migratory Bird Management, Washington, DC U.S.A.

VAN BERS, N.E.M., K. VAN OERS, H.H.D. KERSTENS, B.W. DIBBITS, R.P. CROOIJMANS, M.E. VISSER, AND M.A. GROENEN. 2010. Genome-wide SNP detection in the Great Tit *Parus major* using high throughput sequencing. *Molecular Ecology* 19 Suppl 1:89–99.

WILSON, G.A. AND B. RANNALA. 2003. Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* 163:1177–1191.

YANG, J., B. BEENYAMIN, B.P. MCEVOY, S. GORDON, A.K. HENDERS, D.R. NYHOLT, P.A. MADDEN, A.C. HEATH, N.G. MARTIN, G.W. MONTGOMERY, M.E. GODDARD, AND P.M. VISSCHER. 2010. Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics* 42:565–569.

Received 6 May 2016; accepted 2 November 2016
Associate Editor: Christopher W. Briggs