Title
Genomics of the Spotted Owl (Strix occidentalis) and Barred Owl (Strix varia) in Western North America

Author
Hanna, Zachary Ryan

Publication Date
2017

Peer reviewed|Thesis/dissertation

Genomics of the Spotted Owl (*Strix occidentalis*) and Barred Owl (*Strix varia*) in Western North America

By

Zachary Ryan Hanna

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Integrative Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Associate Professor Rauri C.K. Bowie, Chair
Professor Jimmy A. McGuire
Dr. John P. Dumbacher
Assistant Professor Ian J. Wang

Fall 2017

Abstract

Genomics of the Spotted Owl (*Strix occidentalis*) and Barred Owl (*Strix varia*) in Western North America

by

Zachary Ryan Hanna

Doctor of Philosophy in Integrative Biology

University of California, Berkeley

Associate Professor Rauri C.K. Bowie, Chair

We report here the assembly of a northern spotted owl (*Strix occidentalis caurina*) genome. We generated Illumina paired-end sequence data at 90X coverage using nine libraries with insert lengths ranging from approximately 250 - 9,600 nucleotides and read lengths from 100-375 nucleotides. The genome assembly is comprised of 8,108 scaffolds totaling 1.26×10^9 nucleotides in length with an N50 length of 3.98×10^6 nucleotides. We calculated the genome-wide fixation index (F_{ST}) of *S. o. caurina* with the closely related barred owl (*S. varia*) as 0.819. We examined nineteen genes that encode proteins with light-dependent functions in our genome assembly as well as in that of the barn owl (*Tyto alba*). We present genomic evidence for loss of three of these in *S. o. caurina* and four in *T. alba*. We suggest that most light-associated gene functions have been maintained in owls and their loss has not proceeded to the same extent as in other dim-light-adapted vertebrates.

We describe here the successful assembly of the complete mitochondrial genomes of the northern spotted owl (*Strix occidentalis caurina*) and the barred owl (*S. varia*). We utilized sequence data from two sequencing methodologies, Illumina paired-end sequence data with insert lengths ranging from approximately 250 nucleotides (nt) to 9,600 nt and read lengths from 100-375 nt and Sanger-derived sequences. We employed multiple assemblers and alignment methods to generate the final assemblies. The circular genomes of *S. o. caurina* and *S. varia* are comprised of 19,948 nt and 18,975 nt, respectively. Both code for two rRNAs, twenty-two tRNAs, and thirteen polypeptides. They both have duplicated control region sequences with complex repeat structures. We were not able to assemble the control regions solely using Illumina paired-end sequence data. By fully spanning the control regions, Sanger-derived sequences enabled accurate and complete assembly of these mitochondrial genomes. These are the first complete mitochondrial genome sequences of owls (Aves: Strigiformes) possessing duplicated control regions. We searched the nuclear genome of *S. o. caurina* for copies of mitochondrial genes and found at least nine separate stretches of nuclear copies of gene sequences originating in the mitochondrial genome (*Numts*). The *Numts* ranged from 226-19,522 nt in length and included copies of all mitochondrial genes except $tRNA^{Pro}$, *ND6*, and $tRNA^{Glu}$. *Strix occidentalis caurina* and *S. varia* exhibited an average of 10.74% (8.68% uncorrected p-distance) divergence across the non-tRNA mitochondrial genes.

We analyzed low-coverage, whole-genome sequence data from fifty-one barred owl (*Strix varia*) and spotted owl (*S. occidentalis*) individuals to investigate recent introgression between these two species in western North America. Although we obtained genomic confirmation that these species can hybridize and backcross, we found no evidence of widespread introgression. Plumage characteristics of western *S. varia* that suggested admixture with *S. occidentalis* appeared unrelated to *S. occidentalis* ancestry and may instead reflect local selection.

To Rachel and my parents, without whom this would not exist.

Table of Contents

Acknowledgements

Chapter 1

We thank WildCare, San Rafael for graciously providing us with blood samples from Sequoia. We thank the Cincinnati Museum Center for providing a barred owl (*Strix varia*) tissue sample. We extend special thanks to J. Graham Ruby for his assistance with initial project design and assembly methodology. We thank Anna Sellas for assistance with lab work as well as Ke Bi and Stefan Prost for helpful methodological discussions. We generated genetic data at the Center for Comparative Genomics, California Academy of Sciences. Genewiz, Inc. (U.S.A.) constructed and sequenced the Nextera (Illumina, U.S.A.) large-insert mate-pair libraries. Research performed under University of California, Berkeley Institutional Animal Care and Use Committee approval number R317.

This work was supported by funding from Michael and Katalina Simon [to John P. Dumbacher]; the Louise Kellogg Fund, Museum of Vertebrate Zoology, University of California, Berkeley [to Zachary R. Hanna]; the National Science Foundation Graduate Research Fellowship [DGE 1106400 to Zachary R. Hanna]; the Howard Hughes Medical Institute [to Joseph L. DeRisi]; and the National Science Foundation Postdoctoral Research Fellowship [DBI 1523943 to Christopher A. Emerling]. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. This work used data produced by the Vincent J. Coates Genomics Sequencing Laboratory at the University of California, Berkeley, supported by National Institutes of Health S10 Instrumentation Grants [S10 RR029668, S10 RR027303]. We published this work in *Genome Biology and Evolution* (DOI: 10.1093/gbe/evx158).

Chapter 2

We thank WildCare, San Rafael for graciously providing us with blood samples from Sequoia. We thank the Cincinnati Museum Center for providing a barred owl (*Strix varia*) tissue sample. We thank Laura Wilkinson for assistance with laboratory work. We generated genetic sequence data at the Center for Comparative Genomics, California Academy of Sciences. Genewiz, Inc. constructed and sequenced the Nextera (Illumina) large-insert mate-pair libraries. The Genomics Core Facility at the Icahn Institute and Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai constructed and sequenced the Pacific Biosciences library.

Funds provided by Michael and Katalina Simon (to John P. Dumbacher); the Louise Kellogg Fund, Museum of Vertebrate Zoology, University of California, Berkeley (to Zachary R. Hanna); and the National Science Foundation Graduate Research Fellowship (DGE 1106400 to Zachary R. Hanna) made this work possible. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. This work used data produced by the Vincent J. Coates Genomics Sequencing Laboratory at the University of California, Berkeley, supported by NIH S10 Instrumentation Grants (S10 RR029668, S10 RR027303). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We published this work in *PeerJ* (DOI: 10.7717/peerj.3901).

Chapter 3

For access to specimens and genetic samples, we thank Lowell Diller; J. Mark Higley and Aaron Pole of the Hoopa Valley Indian Reservation Tribal Forestry department; Susan Haig, Tom Mullins, and Mark Miller of the USGS Forest and Rangeland Ecosystem Science Center;

Mary Estes and Morgan of the Chintimini Wildlife Center, Corvallis; Jami Ostby-Marsh and Oroville of the West Valley Outdoor Learning Center, Spokane Valley; Melanie Piazza and WildCare, San Rafael; Sharon Birks and the Burke Museum; Herman L. Mays, Jr., Jane MacKnight, Lauren Hancock, and the Cincinnati Museum Center; Irby Lovette and the Cornell University Museum of Vertebrates; Maureen Flannery, Laura Wilkinson, and the California Academy of Sciences; and Carla Cicero, Theresa Barclay, Shelby Medina, Elizabeth Wommack, and the Museum of Vertebrate Zoology. We thank Anna Sellas for assistance with laboratory work. We generated genomic libraries at the Center for Comparative Genomics, California Academy of Sciences.

Funding from Michael and Katalina Simon [to John P. Dumbacher]; the Louise Kellogg Fund, Museum of Vertebrate Zoology, University of California, Berkeley [to Zachary R. Hanna]; the National Science Foundation Graduate Research Fellowship [DGE 1106400 to Zachary R. Hanna]; and the University of California President's Research Catalyst Award to the UC Conservation Genomics Consortium [to Jeffrey D. Wall] made this research possible. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Chapter 1

Northern spotted owl (*Strix occidentalis caurina*) genome: divergence with the barred owl (*Strix varia*) and characterization of light-associated genes

Zachary R. Hanna,[1,2,3,4] James B. Henderson,[3,4] Jeffrey D. Wall,[1,3,4,5] Christopher A. Emerling,[1,2] Jérôme Fuchs,[3,6] Charles Runckel,[7,8,9] David P. Mindell,[1] Rauri C. K. Bowie,[1,2] Joseph L. DeRisi,[7,8] John P. Dumbacher[3,4]

[1] Museum of Vertebrate Zoology, University of California, Berkeley, Berkeley, California, United States of America
[2] Department of Integrative Biology, University of California, Berkeley, Berkeley, California, United States of America
[3] Department of Ornithology & Mammalogy, California Academy of Sciences, San Francisco, California, United States of America
[4] Center for Comparative Genomics, California Academy of Sciences, San Francisco, California, United States of America
[5] Institute for Human Genetics, University of California San Francisco, San Francisco, California, United States of America
[6] UMR 7205 Institut de Systématique, Evolution, Biodiversité, CNRS, MNHN, UPMC, EPHE, Sorbonne Universités, Département Systématique et Evolution, Muséum National d'Histoire Naturelle, Paris, France
[7] Department of Biochemistry and Biophysics, University of California San Francisco, San Francisco, California, United States of America
[8] Howard Hughes Medical Institute, Bethesda, Maryland, United States of America
[9] Runckel & Associates, Portland, Oregon, United States of America.

Introduction

The spotted owl (*Strix occidentalis*) is a large, charismatic inhabitant of dense forests whose range extends along the Pacific coast of North America from southwestern British Columbia to southern California and eastward into the southwest desert states and Mexico. The northern spotted owl subspecies, *S. o. caurina*, inhabits the Pacific Northwest portion of the *S. occidentalis* range from British Columbia south along the west coast to the Golden Gate strait, California. The U.S. Fish and Wildlife Service listed *S. o. caurina* as "threatened" under the Endangered Species Act (ESA) in 1990 (Thomas et al., 1990) and the owl has been the subject of much ecological research and economic tension. Since its listing under the ESA, populations have continued to decline (Forsman et al., 1996, 2011; Dugger et al., 2015; Davis et al., 2016) despite the increased level of protection. Although it is not considered a "model species" by most researchers, there is a considerable amount of demographic and ecological data available for this species (Courtney et al., 2004), especially in comparison with other owls, which tend to be less studied than diurnal birds.

Spotted owl conservation efforts often focus on genetic challenges, including those relating to small population sizes and inbreeding, relationships to other population segments, and potential interbreeding with congeners (Barrowclough, Gutierrez & Groth, 1999; Haig et al., 2001, 2004, Barrowclough et al., 2005, 2011). A complete genome assembly could provide many useful tools for conservation geneticists, including independent estimates of effective population size (N_e), tools for identifying and developing genetic markers such as single nucleotide polymorphisms and microsatellites, and data that can provide direct and relatively accurate measures of interbreeding.

The congeneric barred owl (*S. varia*), formerly native to North America east of the Rocky Mountains (Mazur & James, 2000), has invaded western North America in the last 50-75 years and, from British Columbia to southern California, has become broadly sympatric with the spotted owl in the last 50 years (Taylor & Forsman, 1976; Livezey, 2009a,b) and likely poses a threat to the survival of the northern spotted owl (Forsman et al., 2011; Wiens, Anthony & Forsman, 2014; Dugger et al., 2015; Diller et al., 2016). In addition to competing for western forest habitat, barred and spotted owls interact at the genetic level as they can hybridize and successfully backcross (Haig et al., 2004; Kelly & Forsman, 2004; Funk et al., 2007). Much of our motivation to assemble the northern spotted owl genome was to provide a resource to aid those studying the genetics of this owl and related taxa. Thus, we included analyses of the genome of a barred owl from eastern North America as a baseline comparison to the spotted owl. We compared genome-derived estimates of N_e from both species and calculated F_{ST} between them.

Access to high-coverage, relatively complete genomes also allows researchers to address questions that, without this resource, are inaccessible or difficult to answer. For example, previous work has suggested that owls have evolved an atypical avian visual system with high numbers of dim-light-adaptive rod photoreceptors (Fite, 1973; Bowmaker & Martin, 1978) and a diminished capacity for color vision (Bowmaker & Martin, 1978; Wu et al., 2016). Whole genome sequencing can establish what mutation(s) or genomic rearrangements resulted in their reduced color vision and, with multiple genomes, one may test whether such mutations are lineage-specific or inherited from a common ancestor. The genome assembly of the barn owl (*Tyto alba*; Aves: Tytonidae) was available and allowed us to test owl-lineage-based hypotheses, but it was one of the lower-coverage, less complete of the available avian genome assemblies (Zhang et al., 2014d). A complete spotted owl genome, in addition to providing whole genome

data for a representative of Strigidae, the other of the two families of owls, could also enable a definitive search for genes involved in nocturnal visual adaptations and a better understanding of the processes of mutation that lead to such adaptations.

Material and Methods

Genome sample

We collected blood from a captive adult northern spotted owl (*S. o. caurina*) at WildCare rehabilitation facility in San Rafael, California. The captive owl, named Sequoia and referred to as such hereafter, patient card # 849, was admitted to WildCare on 5 June 2005 as an abandoned nestling found in Larkspur, Marin County, California (CAS:ORN:98821; Table 1). We chose to sequence the genome of this individual as *S. occidentalis* is known to hybridize with *S. varia* (Haig et al., 2004; Kelly & Forsman, 2004; Funk et al., 2007) and we wanted to ensure that we were sequencing the genome of a non-hybrid, non-introgressed individual. The first Marin County *S. varia* detections occurred in 2003 and researchers estimated a population size of only three individuals by 2005 (Jennings et al., 2011). First generation hybrid individuals are phenotypically diagnosable with intermediate plumage characteristics (Hamer et al., 1994). Thus, if Sequoia had any *S. varia* genetic material, it would likely have been a first generation hybrid and easily diagnosable as such. No plumage or behavioral features, such as vocalizations, suggested that it was a hybrid individual.

DNA Isolation

For genomic DNA libraries that required very high molecular weight DNA, we isolated DNA by using the precipitation method provided by the Gentra Puregene Kit (Qiagen, Hilden, Germany) and following the manufacturer's protocol. We also isolated DNA using a column-based method, the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany), and used this DNA for those libraries where very high molecular weight was not essential. We assessed the quality and concentration of all isolated DNA using a Nanodrop 2000c spectrophotometer (Thermo Fisher Scientific, U.S.A.), 2100 BioAnalyzer (Agilent Technologies, Santa Clara, California), Qubit™ 2.0 Flurometer (Invitrogen, Carlsbad, California), and by running the DNA on a 1% agarose gel. We determined that the resulting DNA from both methods had high molecular weight with most of the DNA comprising fragments greater than 50,000 nucleotides (nt) in length.

Illumina data

We obtained paired-end Illumina data from nine whole-genome libraries constructed using a variety of methods with a range of average insert lengths from 247 - 9,615 nt. In our library construction we utilized a range of DNA shearing methods including enzyme-based, ultrasonication, and hydrodynamic forces using a Hydroshear DNA Shearing Device (GeneMachines, Ann Arbor, Michigan). We amplified all but one of the libraries using polymerase chain reaction (PCR) and sequenced them with read lengths from 100 - 375 nt (Table S1; Supplementary Article section 1.1-1.8).

Trimming, merging, error-correction

We trimmed the Nextera mate-pair data using the software NxTrim version 0.2.3-alpha (O'Connell, 2014; O'Connell et al., 2015) (Supplementary Article section 1.9.1) in order to classify reads of mate pair libraries as true mate pair reads, paired-end reads, or singleton reads. We then removed adapters and low quality bases separately for the resulting mate-pair sequences, paired-end sequences, and singleton sequences using Trimmomatic version 0.32 (Bolger, Lohse & Usadel, 2014) (Supplementary Article section 1.9.2). We also used

Trimmomatic to remove adapters from all non-mate-pair libraries (Supplementary Article section 1.10.1). In order to test how various trimming methods affected the assembly outcome, we trimmed to different thresholds for some of our preliminary assemblies by changing the Trimmomatic version 0.32 (Bolger, Lohse & Usadel, 2014) average quality score parameters. We did not apply the error-correction process to reads trimmed to a stringent quality threshold. For some preliminary assemblies, we performed adapter and quality trimming, but did not merge overlapping paired-end reads (Supplementary Article section 1.13). However, since substantial portions of the paired-end reads from all of the libraries, except the Nextera700bp library, were overlapping, for the sequences that we used to generate our final assembly we joined overlapping paired reads using BBMap version 34.00 (Bushnell, 2014) (Supplementary Article section 1.10.2). We then performed quality trimming on the non-mate-pair library data using Trimmomatic version 0.32 (Bolger, Lohse & Usadel, 2014) (Supplementary Article section 1.10.3). Since we trimmed using the relatively lenient threshold of trimming the read when the average quality over 4 base pairs dropped below phred 17, we next used the k-mer-based error corrector in the SOAPdenovo2 toolkit, SOAPec version 2.01 (Luo et al., 2012), to correct sequence errors (Supplementary Article section 1.11). For any read that became unpaired due to the loss of the paired read we separately subjected it to the same adapter, quality trimming, and error-correcting steps as the reads that remained paired (Supplementary Article section 1.12).

Genome size

In order to estimate the *S. occidentalis* nuclear genome size from our Illumina data, we ran Preqc (Simpson, 2014) with the paired-end sequences from the Nextera700bp dataset (Supplementary Article section 1.14).

Assembly

We assembled the *S. occidentalis* genome using SOAPdenovo2 version 2.04 (Luo et al., 2012). In order to determine the optimal assembly parameter options, we performed numerous trial runs experimenting with different k-mer values and parameters. We utilized the insert size estimated in the output of trial assemblies to refine our estimation of the insert sizes for our libraries and used these refined values as input into subsequent assembly configuration files (Table S1). After optimizing the SOAPdenovo2 assembly options, we generated fourteen further preliminary assemblies to test how using differently filtered versions and subsets of our Illumina sequence data affected the assembly outcome. We examined how the assembly was affected by trimming our data to multiple quality thresholds, using or not applying error correction, not merging or merging our overlapping paired-end data, assembling with different k-mers, using or not using singleton data, and dropping certain libraries (Table S2). We used dupchk (Henderson & Hanna, 2016b) to check for sequence duplication in each sequenced library and found an elevated level of duplication in the Hydroshear library data, so we excluded all sequences from this library from several assemblies (Supplementary Article section 1.15).

Preliminary assembly assessment

In order to compare our preliminary assemblies, we removed contiguous sequences (contigs) or scaffolds less than or equal to 300 nt with the intent of removing any unassembled reads from the assembly. We calculated the contig and scaffold N50 as well as the number of scaffolds in various length classes using scafN50 (Henderson & Hanna, 2016c). We calculated the total length of the assembly, the percentage of "N" characters in the assembly that represent sequence gaps between contiguous sequences joined by paired-end or mate-pair data (% N's), and the total number of scaffolds using scafSeqContigInfo (Henderson & Hanna, 2016a). We were conservative in the calculation of these metrics and separated scaffolds into contigs at each

N in the sequence. We then used CEGMA version 2.5 (Parra, Bradnam & Korf, 2007) to annotate a set of highly conserved eukaryotic genes in our assembly and thereby obtain an assessment of the quality and completeness of each assembly (Supplementary Article section 1.16).

We found it useful to assess the genome assembly's continuity and completeness at each stage of the assembly process. We searched for conserved eukaryotic genes using CEGMA to evaluate our earlier assemblies. However, at this time, one of the CEGMA tool authors recommends that researchers use BUSCO in place of CEGMA (Bradnam, 2015). Since we used CEGMA to evaluate our earliest assemblies, we continued to use CEGMA for continuity. We ran BUSCO on our final assembly and the results suggested similar completeness as those of CEGMA.

Determination of final assembly

We examined multiple statistics in choosing our final assembly. We valued high contig and scaffold N50 values, low % N's in the sequence, a low total number of scaffolds, larger numbers of scaffolds longer than 1 mega nucleotide (Mnt), and completeness as reflected in the number of conserved genes found by the CEGMA pipeline. We decided that the assembly that had the best statistics across these categories was assembly 4 (Table 2) and proceeded forward with this assembly.

We filled gaps in the assembly using the gap closing tool in the SOAPdenovo2 toolkit, GapCloser version 1.12-r6 (Luo et al., 2012). The gap-closed assembly contained many sequences under 1,000 nt in length, a substantial portion of which appeared to be unassembled reads. We used ScaffSplitN50s (Henderson & Hanna, 2016d) to compare statistics describing the continuity of the assembly after removing contigs or scaffolds of lengths 300, 500, and 1,000 nt as well as when using N blocks of lengths 1, 5, 10, 15, 20, and 25 to separate contigs within scaffolds. We decided to remove all contigs and scaffolds less than 1,000 nt for downstream analyses and will refer to the resulting assembly as "StrOccCau_0.2" hereafter (Supplementary Article section 1.18).

Final assembly statistics

We calculated basic statistics on StrOccCau_0.2 using the "assemblathon_stats.pl" script, which was used for comparison of the Assemblathon 2 genome assemblies (Bradnam et al., 2013). We used both CEGMA version 2.5 (Parra, Bradnam & Korf, 2007) and BUSCO version 1.1b1 (Simão et al., 2015a,b) to annotate sets of highly conserved eukaryotic genes and thereby assess the assembly's completeness (Supplementary Article section 1.19). We also calculated basic statistics and ran CEGMA as described above for other available avian genomes, including the barn owl (*Tyto alba*) (Zhang et al., 2014a,e), downy woodpecker (*Picoides pubescens*) (Zhang et al., 2014b,e), zebra finch (*Taeniopygia guttata*) [GenBank assembly accession GCA_000151805.2; (Warren et al., 2010)], bald eagle (*Haliaeetus leucocephalus*) (Warren et al., 2014; Zhang et al., 2014e), golden eagle (*Aquila chrysaetos*) [GenBank assembly accession GCA_000766835.1; (Warren, Bussche & Minx, 2014)], chimney swift (*Chaetura pelagica*) (Zhang et al., 2014c,e), and chicken (*Gallus gallus*) [GenBank assembly accession GCA_000002315.3; (Warren et al., 2017)].

Contamination assessment

To assess whether any assembled contigs were derived from contaminant non-vertebrate organisms, we performed a local alignment of all sequences in StrOccCau_0.2 to a copy of the NCBI nucleotide database "nt" (Clark et al., 2016; NCBI Resource Coordinators, 2016) using NCBI's BLAST+ version 2.3.0 tool BLASTN (Altschul et al., 1997; Camacho et al., 2009)). We

searched for non-vertebrate hits in the top aligned sequences using a local copy of the NCBI taxonomy database (ftp://ftp.ncbi.nlm.nih.gov/pub/taxonomy; (Clark et al., 2016; NCBI Resource Coordinators, 2016) and GItaxidIsVert (Henderson & Hanna, 2016e). We re-examined those sequences where any of the 5 output alignments was an alignment to a non-vertebrate using the web version of NCBI's BLAST+ version 2.4.0 tool BLASTN (Altschul et al., 1997; Camacho et al., 2009)). We used bioawk version 1.0 (Li, 2013b) to remove contaminant scaffolds from the assembly and will refer to the resulting assembly version hereafter as "StrOccCau_1.0". We calculated basic statistics on StrOccCau_1.0 using the "assemblathon_stats.pl" script (Bradnam et al., 2013) (Supplementary Article section 1.20). We confirmed that no conserved eukaryotic genes were present in the contaminant scaffolds.

Mitochondrial genome identification

We searched for any contigs or scaffolds that were assemblies of the mitochondrial genome, rather than the nuclear genome by aligning a mitochondrial genome assembly of the brown wood owl (*S. leptogrammica*) [GenBank Accession KC953095.1; (Liu, Zhou & Gu, 2014)] to StrOccCau_1.0 using NCBI's BLAST+ version 2.4.0 tool BLASTN (Altschul et al., 1997; Camacho et al., 2009). We searched for long alignments to scaffolds with lengths not greatly exceeding 16,500 nt, the approximate size of the mitochondrial genomes of other owl (Aves: Strigiformes) species (Harrison et al., 2004; Mahmood et al., 2014; Liu, Zhou & Gu, 2014; Hengjiu et al., 2016). We extracted the scaffold corresponding to the mitochondrial genome assembly using bioawk version 1.0 (Li, 2013b) and annotated it using the MITOS WebServer version 806 (Bernt et al., 2013) (Supplementary Article section 1.21). We will refer to the mitochondrial and nuclear genome assemblies hereafter as StrOccCau_1.0_mito and StrOccCau_1.0_nuc, respectively.

Sex identification

In order to determine the sex of the *S. o. caurina* individual that supplied the genetic sample for this genome assembly, we aligned nucleotide sequences of *S. varia* chromo-helicase-DNA binding protein-W (*CHD1W*) (GenBank Accession KF425687.1) and chromo-helicase-DNA binding protein-Z (*CHD1Z*) (GenBank Accession KF412792.1) to StrOccCau_1.0 using NCBI's BLAST+ version 2.4.0 tool BLASTN (Altschul et al., 1997; Camacho et al., 2009). We extracted the scaffolds that aligned to the *CHD1W* and *CHD1Z* sequences using bioawk version 1.0 (Li, 2013b) and then used Geneious version 9.1.4 (Kearse et al., 2012; Biomatters, 2016b) to predict the length of a PCR product resulting from amplification of this region with primers 2550F and 2718R (Fridolfsson & Ellegren, 1999) (Supplementary Article section 1.22).

Repeat annotation

We ran our genome through two separate series of repeat masking steps. The purpose of the first series was to produce a masked genome without masking of low complexity regions or simple repeats, which we could then use for downstream annotation steps. The purpose of the second series was to obtain an accurate assessment of the total repeat content of the genome, including low complexity regions and simple repeats. We first performed a homology-based repeat annotation of the genome assembly using RepeatMasker version 4.0.5 (Smit, Hubley & Green, 2013) and the repeat databases of the DFAM library version 1.3 (Wheeler et al., 2013) and the Repbase-derived RepeatMasker libraries version 20140131 (Jurka, 1998, 2000; Jurka et al., 2005; Bao, Kojima & Kohany, 2015) without masking low complexity regions or simple repeats. We next performed a *de novo* modeling of the repeat elements in the genome using RepeatModeler version 1.0.8 (Smit & Hubley, 2015)) in order to create a database of repetitive regions in our genome assembly. We then further masked the genome by running RepeatMasker

using the homology-based repeat-masked genome as input and the repeat database created by our RepeatModeler run and again not masking low complexity regions or simple repeats. The output was a twice-masked genome, hereafter "StrOccCau_1.0_masked". Finally, we repeated the above steps to perform a separate homology-based and *de novo* masking of the genome with RepeatMasker runs that included masking of low complexity regions and simple repeats in order to obtain an accurate estimate of the total repeat content of the genome (Supplementary Article section 1.23).

Gene annotation

In order to annotate genes in the repeat-masked assembly, StrOccCau_1.0_masked, we followed the MAKER version 2.31.8 (Cantarel et al., 2008) pipeline as described in Campbell et al. (2014). As input for protein homology evidence, we provided MAKER the redundant protein set previously used to annotate forty-eight avian genomes (Zhang et al., 2014e). We used the genes found in our CEGMA run to train the gene prediction tool, Semi-HMM-based Nucleic Acid Parser or SNAP version 2006-07-28 (Korf, 2004). As we independently performed repeat masking, we ran MAKER without further repeat masking. We combined all of the output gene annotations using the MAKER accessory scripts “fasta_merge” and “gff3_merge” (Supplementary Article section 1.24).

We assigned putative gene functions to the MAKER annotations by comparing the output MAKER protein fasta file to the Swiss-Prot UniProt release 2016_04 (Consortium, 2015) database using NCBI’s BLAST 2.2.31+ tool “blastp” (Altschul et al., 1997; Camacho et al., 2009). In order to identify proteins with known functional domains, we ran InterProScan version 5.18-57.0 (Jones et al., 2014) on the protein sequences generated by MAKER. We then filtered transcripts with an Annotation Edit Distance (AED) less than 1 and/or a match to a domain in Pfam, a database of protein families (Finn et al., 2016), using the script “quality_filter.pl” supplied in MAKER version 3.00.0 (Cantarel et al., 2008). We compared the unfiltered and filtered GFF3 files by analyzing the AED values for all annotations using the script “AED_cdf_generator.pl” supplied in MAKER version 3.00.0 (Cantarel et al., 2008) and graphed the distribution of values using Matplotlib pyplot (Hunter, 2007) (Figure S1). Finally, we used GenomeTools version 1.5.1 (Gremme, Steinbiss & Kurtz, 2013) to calculate annotation summary statistics, including distributions of gene lengths, exon lengths, number of exons per gene, coding DNA sequence (CDS) lengths (measured in amino acids), and intron lengths (Supplementary Article section 1.24) and graphed these using Matplotlib pyplot (Hunter, 2007) (Figures S2, S3, S4, S5, and S6).

Alignment

We aligned the filtered versions of all sequences from all libraries to StrOccCau_1.0_masked using the Burrows-Wheeler aligner, BWA-MEM version 0.7.12-r1044 (Li, 2013a), and then merged, sorted, and marked duplicate reads using Picard version 1.104 (http://broadinstitute.github.io/picard; accessed 2016 Oct 1). We then assessed the genome coverage, duplication level, and other statistics of each aligned sequence library using Picard version 1.141 (http://broadinstitute.github.io/picard; accessed 2016 Oct 1) (Supplementary Article section 1.25). In order to obtain an estimate of the insert size of the mate pair libraries independent of the N-gaps in the scaffold sequences, we divided scaffolds into contigs at 25 or more N’s, aligned the mate pair libraries to this set of contigs using BWA-MEM version 0.7.12-r1044 (Li, 2013a), and then calculated insert sizes from these alignments (Supplementary Article section 1.25).

Microsatellite analysis

We searched the repeat-masked and unmasked versions of our assembly for all microsatellite primers that have been designed from sequencing of the Mexican spotted owl (*S. o. lucida*) (Thode et al., 2002) as well as additional primers that were designed from sequences obtained from other strigid (Aves: Strigidae) species (Isaksson & Tegelström, 2002; Hsu et al., 2003, 2006; Koopman, Schable & Glenn, 2004; Proudfoot, Honeycutt & Douglas Slack, 2005), but which have been used in population-level studies of *S. occidentalis* (Funk et al., 2008, 2010) and/or have been found to be useful in genetically determining F_1 and F_2 *S. occidentalis* x *S. varia* hybrids (Funk et al., 2007). We searched the assembly for 16 pairs of microsatellite primer sequences using NCBI's BLAST+ version 2.4.0 tool BLASTN (Altschul et al., 1997; Camacho et al., 2009) (Supplementary Article section 1.26).

Barred Owl divergence

In order to assess the genome-wide divergence of *S. occidentalis* and *S. varia*, we extracted genomic DNA from preserved tissue of a *S. varia* collected in Hamilton County, Ohio [CNHM<USA-OH>:ORNITH:B41533; hereafter referred to as "CMCB41533"; Table 1] using a DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany). We prepared a whole-genome library with an average insert length of 466 nt using a Nextera DNA Sample Preparation Kit (Illumina, San Diego, California) and obtained 150 nt paired-end sequence data. We performed adapter and quality trimming of the sequence data using Trimmomatic version 0.32 (Bolger, Lohse & Usadel, 2014). We aligned the trimmed sequences to StrOccCau_1.0_masked using BWA-MEM version 0.7.12-r1044 (Li, 2013a) and then merged the alignments, sorted the alignments, and marked duplicate sequences using Picard version 1.104 (http://broadinstitute.github.io/picard; accessed 2016 Oct 1). We then calculated alignment statistics using Picard version 1.141 (http://broadinstitute.github.io/picard; accessed 2016 Oct 1). We used Genome Analysis Toolkit (GATK) version 3.4-46 UnifiedGenotyper (McKenna et al., 2010; DePristo et al., 2011; Van der Auwera et al., 2013) to call variants using the *S. occidentalis* (Sequoia) and *S. varia* (CMCB41533) BWA-MEM-aligned, sorted, duplicate-marked bam files as simultaneous inputs (Supplementary Article section 1.27).

We then filtered the variants to exclude indels, sites of low genotyping quality, sites where the reference individual had a homozygous alternative allele genotype, and sites with coverage greater than the mean coverage plus five times the standard deviation, as suggested by the GATK documentation (https://software.broadinstitute.org/gatk/guide/article?id=3225; accessed 2016 Oct 1). We used GNU cut version 8.21 (Ihnat, MacKenzie & Meyering, 2013) and GNU Awk (GAWK) version 4.0.1 (Free Software Foundation, 2012) to calculate H_w, the mean number of nucleotide differences within *S. o. caurina* and *S. varia*, as well as H_b, the number of nucleotide differences between the two species, and then used these to estimate the fixation index (F_{ST}) (Hudson, Slatkin & Maddison, 1992), a measure of population differentiation (Supplementary Article section 1.27). We then used an implementation of the pairwise sequentially Markovian coalescent model, PSMC version 0.6.5-r67 (Li & Durbin, 2011; Li, 2015), with 100 rounds of bootstrapping to estimate the effective population size (N_e) through time for both *S. o. caurina* and *S. varia* (Supplementary Article section 1.28).

Light-associated gene analyses

We searched our *S. o. caurina* StrOccCau_1.0 assembly and the *T. alba* genome assembly (GenBank Accession GCA_000687205.1) for the presence of functional orthologs in nineteen genes that encode proteins with light-associated functions. These genes encode five visual pigment proteins (*LWS* [long wavelength-sensitive opsin], *SWS1* [short wavelength-sensitive 1 opsin], *SWS2* [short wavelength-sensitive 2 opsin], *Rh1* [rod opsin], *Rh2* [rod-like

cone opsin]) (Davies, Collin & Hunt, 2012)); ten non-visual photopigment proteins (*Opn3* [panopsin/encephalopsin], *Opn4m* [mammal-like melanopsin], *Opn4x* [*Xenopus*-like melanopsin], *Opn5* [neuropsin], *Opn5L1* [neuropsin-like 1], *Opn5L2* [neuropsin-like 2], *OpnP* [pinopsin], *RRH* [peropsin], *RGR* [retinal G protein-coupled receptor], *OpnVA* [vertebrate ancient opsin]) (Okano, Yoshizawa & Fukada, 1994; Shen et al., 1994; Soni & Foster, 1997; Sun et al., 1997; Blackshaw & Snyder, 1999; Halford et al., 2001; Tarttelin et al., 2003; Bellingham et al., 2006; Tomonari et al., 2008); three enzymes involved in protection from UV radiation (*EEVS*-like, *MT-Ox*, *pOPC1* [photolyase]) (Kato et al., 1994; Osborn et al., 2015); and an enzyme involved in synthesizing red ketocarotenoid pigments (*CYP2J19* [carotenoid ketolase]) (Lopes et al., 2016; Mundy et al., 2016). We queried the genome assemblies of *S. o. caurina* and *T. alba* utilizing *in silico* probes that encompassed the exons, introns and 5' and 3' flanking sequences of the above genes (see Table S3 for details on the probe sequences). We imported the *S. o. caurina* genome assembly into Geneious version 9.1.6 (Kearse et al., 2012; Biomatters, 2016a) and used the included version of the NCBI BLAST+ BLASTn tool (Zhang et al., 2000) to search for the probes in our assembly. We used the web version of NCBI BLAST+ version 2.5.0 (Zhang et al., 2000) to align the probes against the *T. alba* genome assembly sequences in the NCBI Whole-Genome-Shotgun (WGS) contigs database. After recovering matches with our BLAST searches, we used the Geneious version 9.1.6 implementation of the MUSCLE aligner (Edgar, 2004) to align the BLAST results to the probe sequences. We then used Geneious version 9.1.6 to manually adjust the alignments and examine the owl sequences for inactivating mutations, such as premature stop codons, frame shift indels (insertions/deletions), and splice site mutations. When BLAST searches were unsuccessful, we performed BLAST searches against the discarded < 1000 nt contig set. In cases of further negative results, we used synteny data from Ensembl (version 86; (Yates et al., 2016) to search for evidence of whole gene deletion (Supplementary Article section 1.29 and Table S3). Specifically, we identified genes flanking the gene of interest in other vertebrate taxa with available contiguous genomic sequence through the relevant region, and used BLAST as noted above to align the reference sequences for these flanking genes to the genome assemblies of *S. o. caurina* and *T. alba*. If both flanking genes occurred on the same contig/scaffold and the intergenic sequence was not composed of missing data (N's), this provided evidence that the gene of interest had been deleted from the genome. In order to provide further evidence of gene deletion, we used the web version of NCBI BLAST+ version 2.5.0 blastn tool (Zhang et al., 2000) to align the assembly sequence intervening the flanking genes to available sequence data in the NCBI nucleotide database "nt" (Clark et al., 2016; NCBI Resource Coordinators, 2016) to search for remnant sequences of untranslated gene regions.

In instances where we discovered evidence of potentially inactivating mutations in light-associated genes of one or both owl species, we performed dN/dS ratio (ω) analyses to test whether the owl orthologs displayed evidence of relaxation of the strength of natural selection. We obtained additional ortholog sequences for the following non-owl avian species using the web version of the NCBI BLAST+ version 2.5.0 blastn tool (Zhang et al., 2000) with the discontiguous megablast option to search the NCBI nucleotide database "nt" (Clark et al., 2016; NCBI Resource Coordinators, 2016): *Aquila chrysaetos*, turkey vulture (*Cathartes aura*), speckled mousebird (*Colius striatus*), cuckoo roller (*Leptosomus discolor*), bar-tailed trogon (*Apaloderma vittatum*), rhinoceros hornbill (*Buceros rhinoceros*), downy woodpecker (*Picoides pubescens*), and the northern carmine bee-eater (*Merops nubicus*) (see Table S9 for sequence information). After aligning the owl gene sequences with the outgroup taxa using MUSCLE

(Edgar, 2004) in Geneious version 9.1.6, we adjusted the alignments manually and removed all stop codons as well as any codon positions with questionable homology. We then modeled the evolution of the genes of interest using the codeml program from the PAML version 4.8 package (Yang, 2007) assuming the Prum et al. (2015) phylogeny and two separate codon frequency models (F1X4 and F3X4). We created nested models and tested for statistically significant differences in model fits using likelihood ratio tests [parameters included model = 0 (one ratio) or 2 (nested models), fix_omega = 0, NSsites = 0, see Tables S10 and S11 for additional information]. Most models implemented branch tests, which assumed that ω differed across branches on the phylogeny, but was equal across a gene. We estimated the foreground ω on the *Tyto* branch for *OpnP*, the *Strix* and *Tyto* branches for *CYP2J19* and *Rh2*, and the crown (*Strix* + *Tyto*) and stem Strigiformes branches for *Opn4m*. The background ω for each gene consisted of the remaining branches. In a few instances, we implemented branch-sites tests, which assumed differences in ω across the phylogeny while allowing for different ω values across different portions of a gene [parameters included model = 2, fix_omega = 1 (null) or 0 (alternative), omega = 1, NSsites = 2].

We additionally used the NCBI BLAST+ version 2.5.0 blastn tool (Zhang et al., 2000) with the discontiguous megablast option to align a reference *Opn4m* sequence to fifteen avian retinal transcriptomes, which included six owl species (Wu et al., 2016) in NCBI's Sequence Read Archive (SRA) (Leinonen, Sugawara & Shumway, 2011; NCBI Resource Coordinators, 2016) (see Supplementary Article section 1.29 for additional transcriptome information). We imported the short reads that aligned into Geneious version 9.1.6 (Kearse et al., 2012; Biomatters, 2016a) and mapped them to the reference sequence using the Geneious "map to reference" function and trying both the "medium sensitivity / fast" and "low sensitivity / fastest" settings.

Results

Contamination assessment

Our search for non-vertebrate sequences in our assembly suggested that our assembly was only very minimally contaminated with non-vertebrate sequences. For only nine out of the 8,113 final assembly scaffolds, one of the five top alignments to the NCBI nucleotide database (Clark et al., 2016; NCBI Resource Coordinators, 2016) was an alignment to a non-vertebrate sequence. Four of these scaffolds were short, ranging from 1,182 - 2,304 nt, and aligned to *Escherichia coli* sequence data. We removed these four scaffolds from the assembly. We kept the other five scaffolds in the assembly. The highest BLAST bit-score for scaffold-1085 was for an alignment to the telomere region of a human genome with 81% identity across 53% of the scaffold. The highest BLAST bit-scores for scaffold-1155 were for alignments to endogenous retrovirus regions of several vertebrate genomes. Three scaffolds (scaffold-2014, scaffold-2160, scaffold-3069) were longer scaffolds that aligned to vertebrate genome sequences with only small sequence portions that aligned to non-vertebrate sequence data; we did not feel this justified removing them from the assembly.

Mitochondrial genome identification

We identified scaffold-3674 as an assembly of the mitochondrial genome as it had a 14,649 nt alignment with 89.1% similarity to the *S. leptogrammica* mitochondrial genome. This length was the majority of the 21,628 nt scaffold-3674. After subtracting a block of 3,984 N's present in the scaffold, the length of scaffold-3674 is similar to that of other avian mitochondrial genomes (Mindell et al., 1997, 1999; Mindell, Sorenson & Dimcheff, 1998; Guan, Xu & Smith,

2016; Zhang et al., 2016). We were able to annotate all of the standard avian mitochondrial genes, except *ND6* and $tRNA^{Pro}$, which suggests that this assembly of the mitochondrial genome could be improved.

Genome size

Our *k*-mer-based estimation with Preqc yielded an estimated genome length of 1.29 giga nucleotides (Gnt). This type of estimation generally underestimates the true genome size as it collapses *k*-mers from highly repetitive regions. The total length of all sequences in our gap-closed assembly was 1.88 Gnt, but this length included all singleton sequences (many of which were unassembled reads) and N-filled gaps. After removing all contigs and scaffolds less than 1,000 nt, the combined total length of all scaffolds was 1.26 Gnt.

Assembly statistics

Gap-closing improved the assembly continuity and completeness metrics (Table 3, Table 4). Removing shorter length contigs / scaffolds improved the post gap closing assembly metrics at both the contig and the scaffold level. The unfiltered assembly had a scaffold N50 length of 1.836 Mnt and a contig N50 length of 81,400 nt. Removing contigs / scaffolds less than 300 nt increased the scaffold and contig N50 lengths over 2X to 3.916 Mnt and 168,721 nt, respectively, and generated the greatest relative increase in the other continuity metrics of any of the filtering options that we tried (Table S4). The highest scaffold and contig N50 lengths (3.983 Mnt and 171,882 nt, respectively) and the best other continuity metrics resulted from removing all contigs and scaffolds less than 1,000 nt, but this came at the slight expense of the completeness of the genome (Table S4, Table 3, and Table 4). Our gap-closed genome included complete sequences of 228 and at least partial sequences of 236 of the 248 CEGMA orthologs. We only lost one of these when we removed contigs and scaffolds less than 1,000 nt and retained 228 complete and 235 partial CEGMA orthologs in the filtered assembly (Table 4). Except for the percentage of duplicated orthologs, which was notably higher as measured by the CEGMA analysis versus the BUSCO analysis, the results of the CEGMA and BUSCO analyses closely agreed. Both found at least partial sequences of over 90% of the conserved orthologs (235 / 248 = 94.76% CEGMA and 2,815 / 3,023 = 93.12% BUSCO orthologs) under scrutiny in the final assembly (Table 4). Our final assembly contained 8,113 scaffolds and/or contigs with a scaffold N50 of 3.98 Mnt. The longest scaffold was 15.75 Mnt. The GC content was 41.31%. The N content was 1.10%.

The contig-level continuity statistics improved substantially when we allowed for longer blocks of intervening N's before demarcating separate contigs (Table S4). Relative to delineating contigs at every N (contig N50 of 51,301 nt), allowing up to 5 N's before demarcating a separate contig yielded an over 3X increase in the contig N50 of 155,200 nt. This was the greatest relative increase that we saw in the contig N50 length out of all the intervening N lengths that we tried, (Table S4). Allowing up to 25 N's before demarcating a separate contig resulted in the highest contig N50 (171.88 kilo nucleotides (knt); Table S4). In both continuity and completeness, our assembly compares favorably with those of the other avian genomes for which we calculated equivalent metrics (Table 5).

Sex identification

We determined from our assembly that the sequence came from the genome of a female *S. o. caurina*. The lengths of the *CHD1* markers on the sex chromosomes were 634 nt and 1,058 nt on scaffold-806 and scaffold-4429, respectively. These lengths are in the size range of those amplified from *S. nebulosa* samples by previous researchers (600-650 nt and 1,200 nt for

CHD1Z and *CHD1W*, respectively) (Fridolfsson & Ellegren, 1999) and suggest that scaffold-806 and scaffold-4429 are sequences from the Z and W chromosomes, respectively.

Repeat annotation

The repeat annotation and masking of the genome examined 3,754,965 individual sequences totaling 1,882,109,172 nt. The homology-based repeat annotation resulted in GC content estimation of 44.15% and masked 21.02% of the assembly as repetitive. Repeat masking using a *de novo* model of the repeat elements estimated that an additional 0.55% of the assembly was repetitive. Due to the fact that some of the annotated repetitive elements overlapped, the following repeat category percentage values do not exactly sum to the 21.57% total genome repeat content. Interspersed repeat elements including retroelements, DNA elements (DNA transposons with no RNA intermediate), and unclassified elements comprised 9.31% of the assembly; of these, retroelements were the most common, constituting 8.96% of the assembly (Table 6). Non-interspersed repeat elements including small RNA elements, satellites, simple repeats, and low complexity repeats comprised 12.33% of the assembly; of these, satellites were the most common, constituting 9.88% of the assembly.

Gene annotation

The MAKER pipeline succeeded in annotating all contigs and scaffolds except one, scaffold-1363, which is 555,526 nt long and failed the annotation pipeline for an unknown reason. The MAKER pipeline's implementation of AUGUSTUS version 3.2.1 (Keller et al., 2011; Stanke, 2015)predicted 19,692 proteins and transcripts *ab initio*. After quality filtering, we retained 16,718 annotated proteins and transcripts, 5,062 of which were non-overlapping *ab initio* predictions of proteins and transcripts.

Annotated gene sequence lengths ranged from 51 nt to 282,544 nt with a median length of 9,187.50 nt (Figure S2). Coding sequence lengths varied from 51 nt to 66,303 nt with a median length of 1,137 nt (Figure S3). Exon lengths extended to a maximum of 14,832 nt with a median length of 130 nt (Figure S4). Intron lengths ranged from 45 nt to 57,529 nt with a median length of 910 nt (Figure S5). The number of exons per gene ranged from 1 exon to 142 exons with a median number of six exons per gene (Figure S6).

Alignment

The assembly contained 1,142,612,682 non-N bases used in the calculation of the library alignment statistics. After all filters, the total mean coverage for the paired and unpaired data from all of the sequenced libraries aligned to the repeat-masked genome was 60.43X. The MP11kb mate-pair library had the highest proportion of duplicate bases (60.1%) and the PCR-free library noPCR550bp had the lowest (0.3%) (Table 7).

Insert sizes of mate pair libraries determined by mapping quality-filtered reads back to the genome assembly gave lower inserts than were expected based on bioanalyzer traces. Whereas the bioanalyzer traces gave evidence that the MP4kb, MP7kb, and MP11kb libraries had insert lengths of approximately 4.2 knt, 7.1 knt, and 10.7 knt, respectively, the results from mapping to the whole genome assembly suggested that the insert lengths were instead 3.3 knt, 5.9 knt, and 9.6 knt, respectively. We hypothesized that this difference may have been due to the number of N's added during scaffolding, we also mapped the sequences from these libraries to the assembly with all scaffolds decomposed into their constituent contigs. This yielded average insert sizes of 3.3 knt, 6.0 knt, and 10.0 knt, which suggest some potential for improving N gap lengths, but that the N stretches in the scaffolds are good approximations of the lengths of missing, intervening sequences.

Microsatellite analysis

We found 15 out of the 16 pairs of microsatellite primers for which we searched in the genome assembly (Table 8). We found loci 4E10, 4E10.2, and Oe149 on scaffold-11. The distance from the forward 4E10.2 primer to the forward 4E10 primer is 12,172 nt in our assembly, which confirms the characterization of the loci 4E10 and 4E10.2 as linked within 40 kb by the original authors who described these loci using sequences obtained from the same cosmid (Thode et al., 2002). The reverse 4E10 primer is 717,153 nt distant from the forward Oe149 primer. The remaining primer pairs aligned to separate assembly scaffolds (Table 8).

Barred Owl divergence

We estimated the nuclear genome-wide nucleotide diversity (H_w) of *S. o. caurina* as 2.008×10^{-4} and that of *S. varia* as 2.352×10^{-3}. We estimated the genome-wide nucleotide diversity between *S. o. caurina* and *S. varia* (H_b) as 7.042×10^{-3} and calculated an F_{ST} of 0.819.

PSMC analysis

Our pairwise sequentially Markovian coalescent (PSMC) model analyses suggested that the N_e of both *S. o. caurina* and *S. varia* was substantially higher in the past and has been in decline since approximately 100,000 or 80,000 years before present, respectively (figure 1). The estimated peak N_e of *S. o. caurina* was more than an order of magnitude lower than that of *S. varia* (approximately 20,000 and 250,000 for *S. o. caurina* and *S. varia*, respectively). The most recent estimate that the PSMC analysis provided for the N_e of *S. o. caurina* was also more than an order of magnitude lower than that of *S. varia* (approximately 4,000 and 50,000 for *S. o. caurina* and *S. varia*, respectively).

Light-associated gene analyses

Seven of the nineteen genes encoding proteins with light-associated functions that we examined displayed evidence of inactivation or whole gene deletion in one or both owl species [Table S3; (Hanna et al., 2017)]. We found no BLAST alignments of *SWS1* to either the *S. o. caurina* or the *T. alba* assembly. However, the genes flanking *SWS1* in zebra finch (*Taeniopygia guttata*) and human (*Homo sapiens*), *FLNC* (Filamin-C) and *CALU* (Calumenin) (Ensembl version 86; (Yates et al., 2016), are both present in the *S. o. caurina* genome assembly, but they are located on different scaffolds. Without increased genomic continuity, it is difficult to discern whether chromosomal rearrangement has occurred or whether this is a case of simple gene deletion. Recent searches in crocodilian (Crocodilia) genomes similarly found *FLNC* and *CALU* on separate contigs with *SWS1* missing from the assemblies (Emerling, 2017a), which suggests that this may be a problematic region to assemble. NCBI's Eukaryotic Genome Annotation (EGA) pipeline did not find *FLNC* and *CALU* in the *T. alba* genome assembly (NCBI *Tyto alba* Annotation Release 100; NCBI Accession GCF_000687205.1), but the absence of these genes in the assembly may be due to low assembly quality (Zhang et al., 2014d).

SWS2 and *LWS* are adjacent on the same chromosome in the Carolina anole (*Anolis carolinensis*) and African clawed frog (*Xenopus laevis*) genome assemblies and are flanked by *MECP2* (methyl-CpG binding protein 2) in *A. carolinensis* and *X. laevis*, *AVPR2* (arginine vasopressin receptor 2) in *X. laevis*, and *TEX28* (testis expressed 28) in *A. carolinensis* (Ensembl version 86; (Yates et al., 2016). We did not obtain BLAST alignments to *SWS2* or *LWS* for the *T. alba* assembly and NCBI's EGA pipeline did not find *MECP2*, *AVPR2*, or *TEX28* (NCBI *Tyto alba* Annotation Release 100; NCBI Accession GCF_000687205.1), which suggests that this portion of the genome, like the *SWS1* region, may be challenging to assemble. While we found *SWS2* and *LWS* in our *S. o. caurina* assembly, we only obtained partial coding sequences with elevated GC content of 66.9% and 68.0%, respectively. Our *S. o. caurina* assembly contained a partial *SWS2* exon 1 sequence as well as complete exon 2 and 3 sequences with all three exons

found on two separate scaffolds (scaffold-4153 and scaffold-7110). The sequences of these exons on the two scaffolds were 100% identical except for one difference in exon 3. Given the high sequence similarity and the recovery of the same portions of the *SWS2* coding region, these duplicate sequences are likely an artifact of the assembly process and do not indicate gene duplication.

SWS2, *LWS*, *Rh1*, and *Rh2* in *S. o. caurina* and *Rh1* in *T. alba* showed no evidence of potentially inactivating mutations. However, *Rh2* in *T. alba* displayed a 29 nt deletion in exon 1, single premature stop codons in both exon 2 and exon 3, and a 2 nt deletion in exon 4. Our modeling of the sequence evolution of *Rh2* in *S. o. caurina* and *T. alba* yielded evidence that selection has become relaxed in *T. alba* ($\omega = 0.22\text{-}0.37$; $p < 0.00001$) relative to other avian taxa ($\omega = 0.03\text{-}0.06$), which is consistent with pseudogenization of this gene. A branch test of *S. o. caurina* also displayed evidence of relaxed selection on *Rh2* with an elevated ω (0.16-0.21; $p < 0.05$) relative to the background. Our branch-sites test evaluated whether there was indication of positive selection across a subset of sites, but it did not yield any evidence that the elevated ω was due to adaptive evolution. We did find nine missense mutations in *S. o. caurina* that were not found in any of the non-owl avian species, but none of these were at known conserved sites (Carleton, Spady & Cote, 2005), which suggests that they have not resulted in a loss of function.

We were unable to recover *OpnP* in our *S. o. caurina* assembly, but together on the same scaffold we found the genes that flank *OpnP* in the chicken (*Gallus gallus*) and the collared flycatcher (*Ficedula albicollis*) genome assemblies, *TEX14* (testis expressed sequence 14) in *G. gallus* and *DOC2B* (double C2 domain beta) in *G. gallus* and *F. albicollis* [Ensembl version 86; (Yates et al., 2016)]. Our BLAST of the sequence intervening *TEX14* and *DOC2B* in our *S. o. caurina* assembly revealed similarity (8% query coverage, 82% identity) with the 5' untranslated region of *G. gallus OpnP*. Together, these provide strong evidence of whole gene deletion of *OpnP* in *S. o. caurina*. *OpnP* in *T. alba* is a pseudogene with numerous inactivating mutations, including the following: a start codon mutation (ACA), 13 nt deletion, 2 nt insertion, and 1 nt deletion in exon 1, a 1 nt deletion in exon 2, a 21 nt deletion of the intron 3-exon 4 boundary, a 7 nt deletion and 2 nt deletion in exon 4, and a 1 nt deletion in exon 5. We assembled sequences from outgroup taxa and confirmed that these mutations are unique to *T. alba*. Our dN/dS ratio analyses strongly suggested relaxed selection on the *T. alba* branch ($\omega = 0.51\text{-}0.7$; $p < 0.00001$) compared to purifying selection on the background branches ($\omega = 0.11\text{-}0.18$).

Opn4m displays evidence of inactivation in both *S. o. caurina* and *T. alba*, with both species sharing a 4 nt deletion in exon 8. Additionally, *S. o. caurina* has a premature stop codon in exon 8 and *T. alba* possesses a splice donor mutation (GT to AT) in intron 11. Comparisons with outgroup taxa confirmed that these mutations were unique to owls, but also demonstrated that other bird species have putative inactivating mutations in this gene, including the golden eagle (*Aquila chrysaetos*) with a premature stop codon in exon 9; speckled mousebird (*Colius striatus*) with a 1 nt deletion in exon 9, splice donor mutation in intron 9 (GT to TT), and premature stop codon exon 11; cuckoo roller (*Leptosomus discolor*) with a splice donor mutation in intron 10 (GT to GA); and rhinoceros hornbill (*Buceros rhinoceros*) with a start codon mutation (ATG to CTG). We performed dN/dS ratio analyses after removing all exons that contained putative inactivating mutations. The results indicated that the average ω for the crown owl branches is elevated ($\omega = 0.45$; $p < 0.01$) relative to the background ($\omega = 0.19$), which does not meet the expectation of neutral evolution predicted if the shared 4 nt deletion led to a loss of function of *Opn4m*. Branch-sites tests yielded evidence of positive selection on some portions of the gene for both owl branches, but this signal was not a significantly better fit than the null. Our

BLAST of an *Opn4m* sequence to fifteen bird retinal mRNA short read databases, which included data from six owl species, yielded alignments to all fifteen transcriptomes. Further investigation of these sequences in Geneious revealed evidence of different isoforms of *Opn4m*. When we used lower sensitivity alignment settings, the assemblies of mapped sequences generally terminated after exon 8 (the exon with the 4 nt deletion), suggesting that this is an abundant transcript isoform. However, using higher sensitivity alignment settings generated assemblies of multiple transcripts with distinct sequences at some of the exon-intron boundaries.

Finally, *CYP2J19* displays evidence of inactivation in both owl species. *S. o. caurina* has a 1 nt insertion and 2 nt deletion in exon 9. As Emerling (2017b) described, the *T. alba* assembly contains a premature stop codon in each of exons 1, 5, and 6 as well as a 5 nt deletion in exon 3. Both the *S. o. caurina* ($w = 0.33$-0.34; $p < 0.05$) and *T. alba* ($w = 0.68$-0.72; $p < 0.0001$) branches have elevated dN/dS ratios compared to the background (0.15-0.16), which is consistent with the hypothesis that these mutations have led to a loss of function of *CYP2J19*.

Discussion

Genome characterization

Direct comparison of assembly metrics between our *S. o. caurina* assembly and seven other avian genome assemblies, including the avian model organisms chicken (*Gallus gallus*) and zebra finch (*Taeniopygia guttata*), revealed that the *S. o. caurina* assembly is in the top tier of genomes in both continuity and completeness (Table 5). Only the golden eagle (*Aquila chrysaetos*), zebra finch, and chicken genomes had better continuity statistics as measured by scaffold and contig N50s. We compared the relative completeness of the assemblies by searching for a set of 248 conserved eukaryotic genes (CEGs) using CEGMA. Of the assemblies that we compared, we found the highest number of complete conserved gene sequences in our *S. o. caurina* assembly (228 complete CEGs), surprisingly surpassing even the chicken genome (226 complete CEGs). In terms of at least partially complete sequences of conserved genes, our *S. o. caurina* assembly contained only two fewer than the chicken genome (235 vs. 237 partial CEGs). Our assembly is both more complete and more contiguous than that of *Tyto alba*, the only other owl assembly currently available (*S. o. caurina* vs. *Tyto alba* assembly statistics include 235 vs. 198 conserved eukaryotic genes at least partially present, scaffold N50 of approximately 4.0×10^6 nucleotides vs. approximately 5.2×10^4 nucleotides, and contig N50 of approximately 1.7×10^5 nucleotides vs. approximately 1.9×10^4 nucleotides).

The number of annotated genes and the percentage of interspersed repeat elements in our *S. o. caurina* assembly are similar to those seen in other avian genomes (Zhang et al., 2014d). The number of annotated genes in our assembly (16,718 genes) was very similar to the number in the high quality chicken and zebra finch genomes (16,516 and 17,471 genes, respectively) (Zhang et al., 2014d). These values were at the upper end of the range seen in the analysis of the gene annotations of 48 avian genomes (13,454 - 17,471 genes) (Zhang et al., 2014d). Similar to the number of annotated genes, the percentage of interspersed repeat elements in our *S. o. caurina* assembly (9.31%) closely matched the percentage found in the chicken and zebra finch genomes (9.82% and 9.68%, respectively) (Zhang et al., 2014d). These values fell at the higher end of the range seen in the analysis of 48 avian genomes (4.11 - 9.82%) if one excludes the downy woodpecker (*Picoides pubescens*) outlier (22.15%) (Zhang et al., 2014d).

Our searches for conserved eukaryotic genes with both our CEGMA and BUSCO analyses revealed that our *S. o. caurina* assembly lacks only 5-7% of conserved orthologs, which is similar to the 4.4% we observed to be absent in the assembly of the chicken genome. Genome

size data estimated from flow cytometry measurement of red blood cells exist for two *S. occidentalis* congeners. The nuclear genome lengths of the tawny owl (*S. aluco*) and the great grey owl (*S. nebulosa*) are approximately 1.56 Gnt (De Vita et al., 1994; Doležel et al., 2003) and 1.61 Gnt (Doležel et al., 2003; Vinogradov, 2005), respectively, which average to 1.59 Gnt. As compared with this average, the shorter total length of our scaffolded *S. o. caurina* assembly (approximately 1.26 Gnt) suggests that 21% of the full genome sequence length of *S. o. caurina* remains unrepresented in this assembly. This is similar to the approximately 17.8% unrepresented sequence in the 1.19 Gnt golden eagle genome, assuming a genome size of approximately 1.45 Gnt (Nakamura et al., 1990; Doležel et al., 2003). The unrepresented sequence may consist largely of difficult-to-assemble repetitive content (Yamada, Nishida-Umehara & Matsuda, 2004; Wicker et al., 2005). These data illustrate that the *S. o caurina* assembly is comparable to the top tier of avian genomes assembled to date, but, as with all avian genomes, there is still improvement to be made.

Previous work on *Strix* karyotypes suggests that *S. occidentalis* likely has a typical avian karyotype of 2n=80-82 (Renzoni & Vegni-Talluri, 1966; Hammar, 1970; Belterman & Boer, 1984; Rebholz et al., 1993). Assuming 1n = 41 chromosomes, the 8,100 scaffolds in our assembly yield approximately 198 scaffolds per chromosome. However, this number may not be a very meaningful estimate of the number of sequence blocks per chromosome as *Strix* shares with other birds the feature of possessing chromosomes in a wide range of sizes with the majority of the karyotype (approximately 35 of the 41 chromosomes) comprised of microchromosomes and just 6 macrochromosomes (Rebholz et al., 1993).

The SOAPdenovo2 version 2.04 (Luo et al., 2012) assembler does not remove short sequences, which were mostly unincorporated reads. We removed all contigs and scaffolds less than 1,000 nt for our final assembly and used the resulting assembly in downstream analyses. We felt that removal of these small sequences was warranted as sequences shorter than 1,000 nt are unlikely to be useful in assessing synteny or gene structure. Some commonly used assemblers, such as ALLPATHS-LG, do not output contigs / scaffolds less than 1,000 nt (Gnerre et al., 2011). Indeed, the authors of the ALLPATHS-LG description removed contigs / scaffolds less than 1,000 nt in the comparisons of their assembler's functionality with other genome assemblers (Gnerre et al., 2011). Removal of these short sequences post assembly allowed us to better compare across assemblies and to effectively analyze what was actually assembled.

Our CEGMA results suggest that we lost minimal genome information (only 1 out of 248 conserved orthologs examined) by removing assembly contigs / scaffolds less than 1,000 nt. This validated our decision to remove these short sequences and confirmed that it was likely not worth the increase in processing time to retain these small genome fragments in downstream analyses. Additionally, larger genome assembly fragments have greater structural information.

In order to calculate the contig N50 statistic, scaffolds must be decomposed into constituent contigs. We explored how the criteria for splitting scaffolds into contigs affected assembly statistics. As one might expect, allowing longer blocks of N's before breaking a scaffold into contigs resulted in better continuity statistic values. We chose to allow up to 25 N's before separating contigs in our final assembly metric calculations as this was the default used in the "assemblathon_stats.pl" script used for calculating assembly statistics of the Assemblathon 2 genome assemblies (Bradnam et al., 2013). Indeed, even though the "assemblathon_stats.pl" script allowed the user to set a flag to change the number of N's that would separate contigs, our examination of the code revealed that the 25 N's was actually hard-coded into the script and overrode any value set by the user.

We found that our assemblies had better continuity metrics when we did not include all of our available short read data in the assembly. Of particular benefit was the exclusion of the Hydroshear dataset, which displayed a high level of sequence duplication. This suggests that checking libraries for evidence of elevated levels of duplication prior to an assembly could be beneficial.

We found that all of the microsatellite primer pairs previously used for *S. occidentalis* genetic studies (Funk et al., 2007, 2008, 2010) mapped at reasonable distances from each other and predicted PCR products in normal microsatellite size ranges. We found no evidence of linkage except for three primer pairs that mapped to the same scaffold. The other eleven primer sets that we were able to align to the assembly mapped to separate scaffolds. A chromosome-level genomic sequence assembly would help further evaluate the independence of these loci.

Genome-wide divergence of spotted owl and barred owl

As *S. o. caurina* and *S. varia* are separate species, we expected a high genome-wide F_{ST} estimate, but our estimate is elevated even relative to values calculated for other congeneric bird species pairs (Toews et al., 2016). It is difficult to interpret this value, however, as the genome-wide nucleotide diversity within *S. varia* is approximately ten-fold greater than that of *S. o. caurina*. We hypothesize that a difference in N_e for the two species is likely the largest contributor to this difference in nucleotide diversity, especially as the Marin *S. o. caurina* population of which our *S. o. caurina* genome is a sample is known to be an isolated population of this extinction-threatened species (Barrowclough et al., 2005). Following from the ten-fold difference in nucleotide diversity of the two species' genomes, our PSMC analyses suggested that the N_e of *S. varia* was consistently approximately an order of magnitude greater than that of *S. o. caurina* over the past 100,000 years. The PSMC analyses also suggested that the N_e of both *S. o. caurina* and *S. varia* has been in decline over the past tens of millennia, but we caution that precise timing of the past maximum N_e for both species and its subsequent decline is highly dependent on the values chosen for the substitution rate and generation time, which likely require further optimization for these *Strix* species and for owls in general.

Light-associated gene analyses

We have provided genomic evidence of inactivation and deletion of genes with light-associated functions in two species of predominantly nocturnal owls. Ancestral birds likely possessed tetrachromatic color vision (Borges et al., 2015) characterized by four cone photoreceptor opsin pigments with distinct spectral sensitivities, but it appears that owls have a reduced capacity to discriminate colors. Our genomic data for the color vision system in owls are largely consistent with the results of a retinal microspectrophotometry study (Bowmaker & Martin, 1978), retinal transcriptome analyses (Wu et al., 2016), and a recent genomic study of avian visual opsins (Borges et al., 2015). Specifically, the absence of *SWS1*, which absorbs light in the violet/ultraviolet (Davies, Collin & Hunt, 2012), in both *S. o. caurina* and *T. alba* is corroborated by the absence of a violet/ultraviolet-sensitive photopigment in *S. aluco* (Bowmaker & Martin, 1978), the lack of *SWS1* retinal mRNA transcripts in a tytonid and species from all three of the strigid subfamilies (Wu et al., 2016), and a genomic analysis of *T. alba* that also failed to find *SWS1* in the genome assembly (Borges et al., 2015). In our *S. o. caurina* assembly we were able to locate, albeit on separate scaffolds, the genes that flank *SWS1* in other avian taxa, but not *SWS1* itself. More data is needed to confirm whether there are *SWS1* remnants in the *S. o. caurina* and *T. alba* genomes and their absence in the current assemblies is simply due to assembly incompleteness or errors. However, together the data accumulated to date strongly indicate that owls lack *SWS1*, potentially since their most recent common ancestor,

leading to a reduced capacity for color discrimination. The loss of *SWS1* is highly unusual in Aves (Borges et al., 2015). Other than in owls, it has only been inferred to have been lost in the nocturnal North Island brown kiwi (*Apteryx mantelli*) (Le Duc et al., 2015). In contrast, it has occurred repeatedly in nocturnal, subterranean, and marine mammals (Jacobs, 2013; Emerling et al., 2015) as well as in the crocodilians, a lineage believed to have undergone an extensive period of nocturnal adaptation (Walls, 1942; Emerling, 2017a).

The inactivation of *Rh2* in *T. alba* was previously suggested (Borges et al., 2015) and we confirmed this result with the two premature stop codons and two frameshift indels we found in the gene sequence. Additionally, there is evidence that the retinal transcriptome of a congener, *T. longimembris*, does not include *Rh2* transcripts (Wu et al., 2016). The intact copy of *Rh2* in our *S. o. caurina* genome, the transcription of this gene in multiple strigid species (Wu et al., 2016), and the expression of a cone pigment consistent with the Rh2 protein in *S. aluco* (Bowmaker & Martin, 1978) all support the hypothesis that *Rh2* was lost uniquely in the tytonid lineage and not across Strigiformes (Wu et al., 2016). Among avian species, *Rh2* is also inactivated in the kiwi *Apteryx mantelli* (Le Duc et al., 2015) as well as in the Adélie (*Pygoscelis adeliae*) and emperor penguins (*Aptenodytes forsteri*) (Li et al., 2014; Borges et al., 2015), two marine predators that frequently feed at great depths under dim-light conditions. A third penguin species, the Humboldt penguin (*Spheniscus humboldti*) lacks cones with a peak absorbance typical of *Rh2* (Bowmaker & Martin, 1985). The loss of *Rh2* occurred in several other vertebrate groups that are thought to have experienced long periods of inhabiting dim-light environments, including stem Mammalia (Walls, 1942; Davies et al., 2007; Gerkema et al., 2013), Crocodilia (Emerling, 2017a), and snakes (Reptilia: Serpentes) (Castoe et al., 2013; Vonk et al., 2013; Simões et al., 2015; Emerling, 2017c).

The apparent absence of *SWS2* and *LWS* in *T. alba* is likely due to the assembly being incomplete. These genes are in tandem in *A. carolinensis* and *X. laevis*, but the avian assemblies in Ensembl version 86 (Yates et al., 2016) contain *SWS2* and *LWS* on separate small contigs and not adjacent to other genes. This is consistent with our recovery of only partial *SWS2* and *LWS* in *S. o. caurina* and previous difficulties in assembling full *SWS2* and *LWS* sequences in dozens of other avian genomes (Le Duc et al., 2015; Borges et al., 2015), which may be attributable to the high GC content of these genes (Borges et al., 2015). Researchers recovered intact *SWS2* and *LWS* mRNAs in the retinal transcriptomes of five strigid and one tytonid species (Wu et al., 2016) and have demonstrated that the tawny owl (*Strix aluco*) expresses photoreceptor pigments with peak absorptions consistent with *SWS2* and *LWS* (Bowmaker & Martin, 1978), suggesting that *SWS2* and *LWS* are likely retained in owls.

Together, the confluence of data from genomics, transcriptomics, and retinal microspectrophotometry suggests that *SWS1* was likely lost in stem Strigiformes, which resulted in a reduction in the degree of color vision from tetrachromacy to trichomacy by the time of the last common ancestor of owls. *Rh2* became subsequently inactivated in Tytonidae, resulting in further reduced capacity for color discrimination (dichromacy) in this family. Owls, kiwis, and penguins represent the few known avian taxa that deviated from the ancestral avian state of tetrachromatic color vision, likely as a result of an increased dependence on highly sensitive rod photoreceptors for foraging in low light conditions.

The inactivation (*T. alba*) or deletion (*S. o. caurina*) of the gene encoding pinopsin (*OpnP*) may have resulted in the loss of direct photosensitivity of the pineal gland in owls. Pinopsin is expressed in the pineal gland of birds (Okano, Yoshizawa & Fukada, 1994) and likely regulates the daily rhythms of melatonin synthesis. Owls have a relatively small and

simple pineal with little response to differences in luminance (Taniguchi et al., 1993), which suggests that, similar to mammals, the gland may receive photic input indirectly from the eyes (Falcón et al., 2009). *OpnP* is also inactivated in the penguins *Pygoscelis adeliae* and *Aptenodytes forsteri* (Li et al., 2014), but it otherwise appears intact across Aves (Borges et al., 2015). Notably, the loss of pinopsin has also occurred in the historically dim-light-environment-inhabiting Mammalia, Crocodilia, and Serpentes (Walls, 1942; Gerkema et al., 2013; Emerling, 2017a,c). Crocodilians appear to lack a pineal gland entirely (Roth et al., 1980), whereas mammals have a pineal gland that has moved from a more superficial to a deeper position in the brain (Falcón et al., 2009), presumably resulting in a loss of photosensitivity. Together these data suggest that the loss of direct photosensitivity of the pineal gland is a common theme in amniotes (Tetrapoda: Amniota) that experience minimal exposure to light.

Although we found several putative inactivating mutations in *Opn4m*, these are unlikely to have led to complete loss of function. The shared 4 nt mutation in *T. alba* and *S. o. caurina* suggests that it was inherited from the common ancestor of Strigiformes. If this mutation disrupted the function of *Opn4m* in the common ancestor of Strigiformes, then this gene sequence should have been evolving neutrally in all of the descendant lineages. However, Strigidae and Tytonidae split approximately 45 million years ago (Prum et al., 2015) yet each ortholog has only accumulated a single additional putative inactivating mutation, both of which are downstream of exon 8. Our dN/dS ratio analyses of crown owl branches yielded an ω less than 1 ($\omega = 0.45$), which is consistent with the hypothesis that *Opn4m* remains functional in owls. Furthermore, we were able to assemble *Opn4m* from the retinal mRNA sequences from six additional owls (five strigid and one tytonid), which indicates that *Opn4m* is still being transcribed in the eyes of those species. We found evidence of multiple *Opn4m* isoforms in the avian retinal transcriptome sequences and the genomic sequences of several other avian taxa possessed putative inactivating mutations. These potentially inactivating mutations were almost all distributed on or after exon 8. Notably, when we used the lowest sensitivity setting of the Geneious aligner to map *Opn4m* BLAST hits from the avian retinal transcriptomes, we primarily obtained assembled sequences that terminated after exon 8. Previous work has found multiple *Opn4m* isoforms in vertebrates (Verra et al., 2011; Hughes et al., 2012). Our results suggest loss of some of these isoforms in owls and other birds. *Opn4m* is involved in entraining circadian rhythms in mammals via the pineal gland, in part, as well as in regulating pupil diameter (Hankins, Peirson & Foster, 2008). Given the diminished importance of the pineal gland in owls, alteration of the circadian function of *Opn4m* is a possibility.

CYP2J19 has recently been implicated as the carotenoid ketolase responsible for synthesizing red carotenoids in birds (Lopes et al., 2016; Mundy et al., 2016; Emerling, 2017b). Carotenoids, in addition to being involved in pigmentation of avian skin and feathers, are located in oil droplets anterior to the photosensitive outer segments of cone photoreceptors. These oil droplets fine-tune color vision by absorbing shorter wavelengths and reducing spectral overlap between cone visual pigments (Vorobyev, 2003). However, these droplets also reduce the number of photons that reach cone photoreceptors and, therefore, may be less beneficial under dim-light conditions. Among owls, *Strix aluco*, *Athene noctua* (little owl), and *Asio flammeus* (short-eared owl) are known to possess red cone oil droplets, whereas *Strix uralensis* (Ural owl), *Bubo scandiacus* (snowy owl), and *Tyto alba* lack them (Erhard, 1924; Yew, Woo & Meyer, 1977; Bowmaker & Martin, 1978; Gondo & Ando, 1995). In *Strix aluco*, the red oil droplets are limited to less than 1% of the cone photoreceptor population (Bowmaker & Martin, 1978), which is an extremely low proportion compared to other avian species (Bowmaker, 1980; Partridge,

1989). Additionally, there is recent evidence that *CYP2J19* is inactivated in *T. alba*, is transcribed as a pseudogene in the retinal transcriptome of *Asio otus* (long-eared owl), and is transcribed at low levels in five other owl species as compared to the level observed in diurnal outgroup avian taxa (Emerling, 2017b). Among non-owl Aves, the absence of red cone oil droplets has only been reported in two penguin species, *Spheniscus humboldti* (Bowmaker & Martin, 1985) and *Aptenodytes patagonicus* (Gondo & Ando, 1995). Among non-owls, *CYP2J19* is inactivated in the penguins *Pygoscelis adeliae* and *Aptenodytes forsteri* as wells as in the kiwi *Apteryx mantelli* (Emerling, 2017b), which all forage under dim-light conditions. The *CYP2J19* pseudogene reported here for *Strix occidentalis caurina* provides further evidence that owls have repeatedly been losing red carotenoid oil droplets in parallel, potentially to maximize retinal sensitivity in their predominantly nocturnal niche.

Perhaps what is most notable about the loss of light-associated genes in Strigiformes is not the fact that it has occurred, but that it has not ensued to the same extent as in other historically dim-light-adapted vertebrates. Of the nineteen genes we examined, all but one (*CYP2J19*) were likely present in the common ancestor of amniotes (Gerkema et al., 2013; Osborn et al., 2015; Twyman et al., 2016). Excluding *CYP2J19*, mammals lost nine (Mammalia: Marsupialia and Monotremata) to ten of these genes (Mammalia: Placentalia) during a hypothesized nocturnal or mesopic bottleneck (Walls, 1942; Heesy & Hall, 2010; Davies, Collin & Hunt, 2012; Gerkema et al., 2013) and crocodilians lost seven during a similarly hypothesized period of dim-light adaptation (Walls, 1942; Emerling, 2017a). Among squamates (Reptilia: Squamata), snakes lost seven of these genes during a putative nocturnal and/or fossorial period early in their history, whereas the largely nocturnal geckos lost six (Walls, 1942; Emerling, 2017c). As for owls, tytonids have lost three of the light-associated genes we examined (*SWS1*, *Rh2*, *OpnP*), whereas strigids have lost only two (*SWS1*, *OpnP*).

Conclusions

We report the first genome of a member of Strigidae, the largest family of owls. We anticipate that this draft whole genome assembly will be useful to those studying the genetics, demography, and conservation of the spotted owl and related taxa. It will be of particular use in genetic identification of hybrid spotted / barred owls (*S. occcidentalis* x *varia*) and in ascertaining the frequency of hybridization between these two species in the forests of western North America. The phylogenetic position of owls within Neoaves is at the base of a large clade containing mousebirds (Coliiformes), cuckoo-rollers (Leptosomiformes), trogons (Trongoniformes), hornbills (Bucerotiformes), woodpeckers (Piciformes), and kingfishers (Coraciiformes) (Jarvis et al., 2014; Prum et al., 2015). This placement of owls suggests that our spotted owl genome assembly will be useful in genomic studies that span a substantial component of avian morphologic diversity and life history strategies.

Despite potentially more than 45 million years of dim-light specialization in Strigiformes, owls have retained a diverse array of non-visual opsin pigments and mechanisms to protect against ultraviolet photo-oxidative damage. Although tytonids have a reduced color vision capacity that is similar to ancestral mammals, crocodilians, and snakes, strigids have retained trichromatic color vision akin to that of humans. Many light-associated gene functions have been maintained in owls, perhaps enabling activities during daylight, a time when most owls are presumed to be generally inactive. It appears that what many consider the quintessential nocturnal birds are not as independent of light as are other nocturnal or crepuscular amniote lineages.

Data deposition

Specimen Sequoia blood sample deposited as CAS:ORN:98821, California Academy of Sciences, San Francisco, California, United States of America. This Whole Genome Shotgun sequencing project has been deposited at DDBJ/ENA/GenBank under the accession NIFN00000000. The version described in this paper is version NIFN01000000. *Strix occidentalis caurina* raw genomic DNA sequences obtained from CAS:ORN:98821 are available from NCBI Sequence Read Archive (SRA) (SRA run accessions SRR4011595, SRR4011596, SRR4011597, SRR4011614, SRR4011615, SRR4011616, SRR4011617, SRR4011618, SRR4011619, and SRR4011620). *Strix varia* raw genomic DNA sequences obtained from CNHM<USA-OH>:ORNITH:B41533 are available from NCBI Sequence Read Archive (SRA) (SRA run accessions SRR5428115, SRR5428116, and SRR5428117). Program ScaffSplitN50s deposited at Zenodo <http://doi.org/10.5281/zenodo.163683> and available at <https://github.com/calacademy-research/ScaffSplitN50s>. Program dupchk deposited at Zenodo <http://doi.org/10.5281/zenodo.163722> and available at <https://github.com/calacademy-research/dupchk>. Program GItaxidIsVert deposited at Zenodo <http://doi.org/10.5281/zenodo.163737> and available at <https://github.com/calacademy-research/GItaxidIsVert>. Program scafSeqContigInfo deposited at Zenodo <http://doi.org/10.5281/zenodo.163748> and available at <https://github.com/calacademy-research/scafSeqContigInfo>. Program scafN50 deposited at Zenodo <http://doi.org/10.5281/zenodo.163739> and available at <https://github.com/calacademy-research/scafN50>. Additional scripts deposited as NSO-genome-scripts at Zenodo <http://doi.org/10.5281/zenodo.805012> and available at <https://github.com/calacademy-research/NSO-genome-scripts>. Gene and repeat annotation files, the raw variant call file, alignments of light-associated gene orthologs as well as assemblies of transcriptome sequences deposited at Zenodo <http://doi.org/10.5281/zenodo.822859>.

Tables

Table 1. Specimen data.

Information regarding the *Strix occidentalis caurina* and *Strix varia* individuals from which we obtained genomic sequences for this study including the county, state, country, and date of collection for each specimen as well as the specimen code and institution where each specimen is archived.

Specimen	County	State	Country	Date	Specimen institution
CAS:ORN:98821	Marin County	California	U.S.A.	26 Jun 2005	California Academy of Sciences
CNHM<USA-OH>:ORNITH:B41533	Hamilton County	Ohio	U.S.A.	29 Nov 2010	Cincinnati Museum Center

Table 2. Metrics of preliminary assemblies.

Various continuity and completeness summary statistics for our preliminary assemblies. We removed contigs / scaffolds < 300 nt in order to remove unassembled reads from the assemblies before calculating these statistics. We defined contigs with the very restrictive parameter that each N split a scaffold into a separate contig. "Partial CEGs found by CEGMA" refers to the number of gene sequences found by CEGMA in the assembly in at least partial completeness out of 248 total conserved eukaryotic genes. An asterisk and bolded font mark the preliminary assembly that we chose to use as the basis for the final assembly.

Assembly	contig N50 (nt)	scaffold N50 (nt)	Total length of assembly (Gnt)	Ns (%)	Total number of scaffolds	Number of scaffolds > 1 Mnt in length	Partial CEGs found by CEGMA	Complete CEGs found by CEGMA
1	9,499	3,869,235	1.275	4.77	51,843	292	231	205
2	12,096	3,522,724	1.274	4.40	48,264	295	233	205
3	10,425	4,007,375	1.272	4.88	47,075	0	226	200
4*	**13,983**	**3,919,460**	**1.275**	**4.26**	**47,900**	**303**	**235**	**221**
5	10,315	4,164,870	1.272	4.45	46,146	287	232	206
6	9,142	3,780,867	1.275	4.86	51,615	296	230	202
7	9,802	3,478,271	1.274	4.42	54,240	327	233	209
8	12,650	3,665,028	1.271	4.18	43,092	313	231	204
9	12,006	3,587,241	1.271	4.66	44,939	307	226	201
10	12,487	3,586,666	1.271	4.26	44,345	314	232	204
11	14,651	3,917,141	1.276	4.26	50,636	293	234	217
12	14,627	3,728,521	1.276	4.28	50,349	305	234	219
13	14,672	3,917,121	1.276	4.26	50,129	293	234	217
14	13,967	3,431,044	1.3	4.50	127,384	318	238	218

Table 3. Final assembly metrics.

Assembly (contaminant and mitochondrial sequences removed) metrics before gap-closing, after gap-closing, and after both gap-closing and removal of all contigs and scaffolds less than 1000 nt in length. Strings of 25 or more N's broke scaffolds into contigs.

Assembly version	No gap-closing, no scaffolds or contigs removed	Gap-closed, no scaffolds or contigs removed	Gap-closed, scaffolds and contigs <1000 nt removed
Number of scaffolds	3,754,960	3,754,960	8,108
Total size of scaffolds	1,884,397,264 nt	1,882,081,621 nt	1,255,541,132 nt
Longest scaffold	15,783,852 nt	15,750,186 nt	15,750,186 nt
Shortest scaffold	128 nt	128 nt	1,000 nt
Number of scaffolds > 1K nt	8,112 (0.2%)	8,095 (0.2%)	8,095 (99.8%)
Number of scaffolds > 10K nt	1,754 (0.0%)	1,746 (0.0%)	1,746 (21.5%)
Number of scaffolds > 100K nt	661 (0.0%)	661 (0.0%)	661 (8.2%)
Number of scaffolds > 1M nt	303 (0.0%)	303 (0.0%)	303 (3.7%)
Number of scaffolds > 10M nt	9 (0.0%)	9 (0.0%)	9 (0.1%)
Mean scaffold size	502 nt	501 nt	154,852 nt
Median scaffold size	150 nt	150 nt	1,904 nt
N50 scaffold length (L50 scaffold count)	1,843,286 nt (209)	1,836,279 nt (209)	3,983,020 nt (92)
N60 scaffold length (L60 scaffold count)	622,124 nt (370)	619,581 nt (371)	3,012,707 nt (129)
N70 scaffold length (L70 scaffold count)	255 nt (216,251)	255 nt (218,976)	2,162,240 nt (178)
N80 scaffold length (L80 scaffold count)	174 nt (1,110,583)	174 nt (1,113,245)	1,545,070 nt (246)
N90 scaffold length (L90 scaffold count)	143 nt (2,336,958)	143 nt (2,338,577)	618,731 nt (372)
scaffold %GC	42.81%	43.82%	41.31%
scaffold %N	2.89%	0.74%	1.10%
Percentage of assembly in scaffolded contigs	66.4%	65.7%	98.5%
Percentage of assembly in unscaffolded contigs	33.6%	34.3%	1.5%
Average number of contigs per scaffold	1.0	1.0	3.4
Average length of break (>25 Ns) between contigs in scaffold	311	703	716
Number of contigs	3,929,029	3,774,552	27,252
Number of contigs in scaffolds	179,939	22,372	21,478
Number of contigs not in scaffolds	3,749,090	3,752,180	5,774
Total size of contigs	1,830,109,624 nt	1,868,296,631 nt	1,241,823,123 nt
Longest contig	186,255 nt	1,259,046 nt	1,259,046 nt
Shortest contig	5 nt	128 nt	130 nt
Number of contigs > 1K nt	123,891 (3.2%)	23,915 (0.6%)	23,915 (87.8%)
Number of contigs > 10K nt	37,347 (1.0%)	12,373 (0.3%)	12,373 (45.4%)
Number of contigs > 100K nt	58 (0.0%)	3,909 (0.1%)	3,909 (14.3%)
Number of contigs > 1M nt	0 (0.0%)	8 (0.0%)	8 (0.0%)
Mean contig size	466 nt	495 nt	45,568 nt
Median contig size	150 nt	150 nt	6,702 nt
N50 contig length (L50 contig count)	7,855 nt (46,856)	81,400 nt (4,678)	171,882 nt (2,057)
N60 contig length (L60 contig count)	3,275 nt (81,600)	33,521 nt (8,121)	134,419 nt (2,876)
N70 contig length (L70 contig count)	254 nt (448,715)	255 nt (254,729)	98,604 nt (3,955)
N80 contig length (L80 contig count)	170 nt (1,346,255)	173 nt (1,148,692)	66,668 nt (5,484)
N90 contig length (L90 contig count)	142 nt (2,548,877)	142 nt (2,367,845)	34,559 nt (8,023)

Table 4. Summary of conserved ortholog searches.

Comparison of the number of conserved orthologous genes found in the final assembly (gap-closed, contigs / scaffolds < 1000 nt removed) using the CEGMA and BUSCO tools. In order to illustrate the effect of gap-closing and removal of small fragments on assembly completeness metrics, also included are the results of CEGMA gene searches conducted on two draft versions of the final assembly where we either did not perform gap-closing and removed contigs / scaffolds < 300 nt or performed gap-closing and did not remove any small contigs / scaffolds.

Assembly	**Draft. No gap-closing, contigs / scaffolds < 300 nt removed**	**Draft. Gap-closed, no removal of small contigs / scaffolds.**	**Final. Gap-closed, contigs / scaffolds < 1000 nt removed.**	**Final. Gap-closed, contigs / scaffolds < 1000 nt removed.**
Method	**CEGMA**	**CEGMA**	**CEGMA**	**BUSCO**
Total conserved orthologs examined	248	248	248	3023
Complete orthologs (% of total)	221 (89.11%)	228 (91.94%)	228 (91.94%)	2605 (86.17%)
At least partial orthologs (% of total)	235 (94.76%)	236 (95.16%)	235 (94.76%)	2815 (93.12%)
Duplicated orthologs (% of total)	92 (37.10%)	83 (33.47%)	99 (39.92%)	46 (1.52%)
Missing orthologs	13 (5.24%)	12 (4.84%)	13 (5.24)%	208 (6.88%)

Table 5. Comparative statistics of avian genomes.

Comparative statistics of our *Strix occidentalis caurina* assembly with those of a selection of other avian genome assemblies.

Species	Common name	Scaffold N50 (nt)	# scaffolds / contigs	Contig N50 (nt)	Length (Gnt)	Ns (%)	Complete CEGs (% of 248)	Partial CEGs (% of 248)
Strix occidentalis caurina	Northern Spotted Owl	3,983,020	8,108	171,882	1.26	1.10	228 (91.94%)	235 (94.76%)
Tyto alba	Barn Owl	51,873	166,092	19,113	1.14	1.02	144 (58.06%)	198 (79.84%)
Picoides pubescens	Downy Woodpecker	2,086,781	85,828	29,578	1.17	3.72	196 (79.03%)	216 (87.10%)
Taeniopygia guttata	Zebra Finch	62,374,962	37,095	38,644	1.23	0.75	192 (77.42%)	214 (86.29%)
Haliaeetus leucocephalus	Bald Eagle	669,725	346,419	10,218	1.26	3.97	217 (87.50%)	240 (96.77%)
Aquila chrysaetos	Golden Eagle	9,230,743	1,141	215,151	1.19	1.07	226 (91.13%)	238 (95.97%)
Chaetura pelagica	Chimney swift	3,839,435	60,234	33,918	1.13	4.02	191 (77.02%)	222 (89.52%)
Gallus gallus	Chicken	82,310,166	23,474	2,905,620	1.23	0.96	226 (91.13%)	237 (95.56%)

Table 6. Repetitive element summary.

Summary of the repeat elements found during two rounds of repeat masking (homology-based followed by denovo-model-based masking). Depending on the type of repeat element, we provide information at different category summary levels. We use the "Type level" column headings to organize these categories.

Type level 1	Type level 2	Type level 3	Type level 4	Number of elements	Element total length (nt)	Assembly portion (%)
Total interspersed repeats					175,287,790	9.31
	Total retroelements			727,006	168,672,903	8.96
	Retroelement	SINE		40,360	4,770,020	0.25
	Retroelement	SINE	ALU	53	6,194	0.00
	Retroelement	SINE	MIR	15,510	1,558,420	0.08
	Retroelement	Penelope		169	35,110	0.00
	Retroelement	Total LINEs		486,310	115,604,290	6.14
	Retroelement	LINE	LINE1	622	58,117	0.00
	Retroelement	LINE	LINE2	3,116	317,864	0.02
	Retroelement	LINE	L3/CR1	28,122	5,153,289	0.27
	Retroelement	LINE	CRE/SLACS	0	0	0.00
	Retroelement	LINE	L2/CR1/Rex	452,030	109,807,316	5.83
	Retroelement	LINE	R1/LOA/Jockey	0	0	0.00
	Retroelement	LINE	R2/R4/NeSL	131	44,590	0.00
	Retroelement	LINE	RTE/Bov-B	15	3,492	0.00
	Retroelement	LINE	L1/CIN4	98	23,441	0.00
	Retroelement	Total LTR elements		200,336	48,298,593	2.57
	Retroelement	LTR	BEL/Pao	0	0	0.00
	Retroelement	LTR	ERV_classI	983	122,219	0.01
	Retroelement	LTR	ERV_classII	400	54,854	0.00
	Retroelement	LTR	ERVL	436	91,660	0.00
	Retroelement	LTR	ERVL-MaLRs	51	4,838	0.00
	Retroelement	LTR	Gypsy/DIRS1	111	14,921	0.00
	Retroelement	LTR	Retroviral	197,967	47,947,799	2.55
	Retroelement	LTR	Ty1/Copia	0	0	0.00
	Total DNA elements			37,526	5,628,486	0.30
	DNA element		En-Spm	0	0	0.00
	DNA element		hAT-Charlie	418	28,220	0.00
	DNA element		hobo-Activator	4,235	719,417	0.04
	DNA element		MuDR-IS905	0	0	0.00
	DNA element		PiggyBac	0	0	0.00
	DNA element		Tc1-IS630-Pogo	806	141,663	0.01
	DNA element		TcMar-Tigger	528	39,074	0.00
	DNA element		Tourist/Harbinger	9,255	958,360	0.05
	DNA element		Other (Mirage, P-element, Transib)	0	0	0.00
	Rolling-circles			0	0	0.00
	Unclassified interspersed repeats			6,225	986,401	0.05
Total non-interspersed repeats				1,907,394	232,038,709	12.33
	Small RNA			12,051	1,645,166	0.09
	Satellites			1,261,021	185,995,538	9.88
	Simple repeats			564,508	40,568,395	2.16
	Low complexity repeats			69,814	3,829,610	0.20

Table 7. Library alignment statistics.

Alignment statistics for all Sequoia (*Strix occidentalis caurina*) libraries and the CMCB41533 (*Strix varia*) library calculated using Picard's CollectWgsMetrics.

Library	Mean paired and unpaired read genome coverage post filtering (X)	Standard deviation of paired and unpaired read genome coverage post filtering (X)	Fraction of aligned bases from unpaired reads	Total fraction of filtered aligned bases	Fraction aligned bases filtered due to mapping quality < 20	Fraction aligned bases filtered as duplicates	Fraction aligned bases filtered as low quality with Q < 20	Fraction aligned bases filtered as second observation from overlapping reads	Fraction aligned bases filtered from regions already with > 1000X coverage
Nextera350bp lane 1	4.369	5.484	0.048	0.533	0.060	0.444	0.004	0.023	1.52E-03
Nextera350bp lane 2	11.162	8.960	0.039	0.559	0.056	0.480	0.005	0.017	1.43E-03
Hydroshear	1.093	2.784	0.004	0.549	0.033	0.429	0.005	0.081	2.03E-03
Nextera550bp lane 1	2.741	3.708	0.393	0.096	0.034	0.038	0.011	0.011	1.05E-03
Nextera550bp lane 2	5.790	5.435	0.327	0.126	0.032	0.066	0.019	0.008	1.26E-03
Nextera700bp	23.357	14.710	0.041	0.216	0.046	0.126	0.009	0.032	3.64E-03
noPCR550bp	3.244	2.661	0.241	0.059	0.013	0.003	0.014	0.029	4.32E-04
PCR900bp	1.978	1.894	0.073	0.052	0.012	0.024	0.014	0.001	3.34E-04
MP4kb	2.528	2.745	0.300	0.361	0.048	0.306	0.002	0.004	5.36E-04
MP7kb	2.528	2.734	0.256	0.449	0.045	0.397	0.002	0.004	4.53E-04
MP11kb	1.641	2.205	0.168	0.652	0.046	0.601	0.001	0.004	2.56E-04
CMCB41533	15.552	12.254	0.030	0.341	0.299	0.037	2.37E-04	2.59E-03	2.50E-03

Table 8. Genomic locations of selected microsatellite loci.

Locations of commonly used microsatellite loci in our draft genome assembly. We searched for all of the primer pairs used in several *Strix occidentalis* population genetics studies as well all of those designed for use in *S. o. lucida* (Thode et al., 2002). The "Primer" column designates the forward or reverse primer with "F" or "R", respectively. The "Reference" column gives the citation of the publication that originally described each primer pair. The "Comment" column gives the citation(s) of the publication(s) in which a primer pair has been used for population-level study of *S. occidentalis* or and/or study of *S. occidentalis* x *S. varia* hybrids. "Length alignment" refers to the length of the BLASTN (Altschul et al., 1997; Camacho et al., 2009) alignment. The "Microsatellite length" refers to the inferred length of the microsatellite PCR product based on the length of the primers and their mapping positions in the genome assembly.

Locus	Primer	Reference	Usage comments	Length primer	Length alignment	Mismatches	Genome scaffold	Genome start	Genome end	Microsatellite length (nt)
13D8	F	(Thode et al., 2002)	population genetics (Funk et al., 2008, 2010)	22	22	0	scaffold88	4,241,040	4,241,019	187
13D8	R			21	21	0	scaffold88	4,240,854	4,240,874	
15A6	F	(Thode et al., 2002)	population genetics (Funk et al., 2008, 2010)	21	21	0	scaffold233	2,208,703	2,208,723	148
15A6	R			19	16	0	scaffold233	2,208,847	2,208,832	
1C6	F	(Thode et al., 2002)	None	20	20	0	scaffold178	2,550,734	2,550,753	110
1C6	R			20	20	0	scaffold178	2,550,843	2,550,824	
4E10	F	(Thode et al., 2002)	None	22	22	0	scaffold11	768,391	768,371	230
4E10	R			22	22	0	scaffold11	768,162	768,183	
4E10.2	F	(Thode et al., 2002)	population genetics (Funk et al., 2008, 2010)	18	18	0	scaffold11	780,562	780,579	226
4E10.2	R			18	18	0	scaffold11	780,787	780,770	
6H8	F	(Thode et al., 2002)	population genetics (Funk et al., 2008, 2010)	21	21	0	scaffold103	3,773,885	3,773,865	93
6H8	R			16	16	0	scaffold103	3,773,793	3,773,808	
8G11	F	(Thode et al., 2002)	None	18	-	-	-	-	-	-
8G11	R			17	-	-	-	-	-	
Bb126	F	(Isaksson & Tegelström, 2002)	hybrid diagnostic (Funk et al., 2007)	20	20	0	scaffold219	2,548,147	2,548,166	185
Bb126	R			24	24	0	scaffold219	2,548,331	2,548,308	
BOOW18	F	(Koopman, Schable & Glenn, 2004)	hybrid diagnostic (Funk et al., 2007)	19	19	1	scaffold244	648,444	648,426	205
BOOW18	R			20	20	1	scaffold244	648,240	648,259	
FEPO5	F	(Proudfoot, Honeycutt & Douglas Slack, 2005)	population genetics (Funk et al., 2008, 2010)	22	22	0	scaffold138	720,315	720,336	270
FEPO5	R			25	25	2	scaffold138	720,584	720,560	
Oe045	F	(Hsu et al., 2003)	hybrid diagnostic (Funk et al., 2007)	23	23	2	scaffold173	3,777,655	3,777,677	127
Oe045	R			19	19	0	scaffold173	3,777,781	3,777,763	
Oe053	F	(Hsu et al., 2003)	population genetics (Funk et al., 2008, 2010)	23	23	1	scaffold136	299,240	299,262	218
Oe053	R			22	22	1	scaffold136	299,457	299,436	
Oe128	F	(Hsu et al., 2003)	hybrid diagnostic (Funk et al., 2007), population genetics (Funk et al., 2008, 2010)	27	27	0	scaffold722	802,232	802,206	319
Oe128	R			24	24	0	scaffold722	801,914	801,937	
Oe129	F	(Hsu et al., 2006)	population genetics (Funk et al., 2008, 2010)	24	21	2	scaffold529	3,066,759	3,066,739	266
Oe129	R			24	24	1	scaffold529	3,066,497	3,066,520	
Oe149	F	(Hsu et al., 2006)	population genetics (Funk et al., 2008, 2010)	21	21	1	scaffold11	51,010	50,990	258
Oe149	R			20	20	0	scaffold11	50,753	50,772	
Oe3-7	F	(Hsu et al., 2003)	population genetics (Funk et al., 2008, 2010)	20	19	1	scaffold35	572,329	572,347	129
Oe3-7	R			23	23	0	scaffold35	572,456	572,434	

Figures

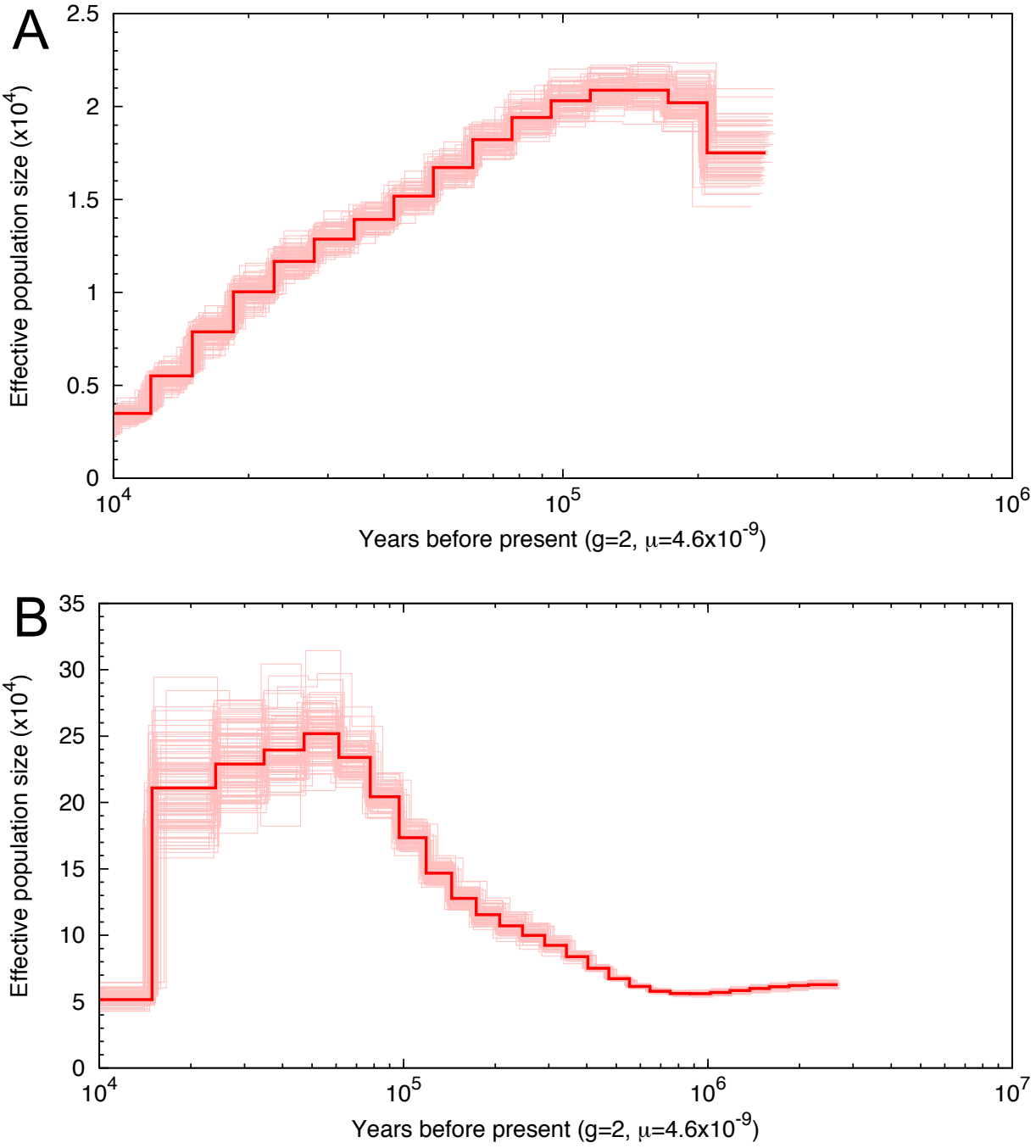

Figure 1. Demographic history of *Strix occidentalis caurina* and *Strix varia* with bootstrap replicates.

Panel A depicts the demographic history estimated for *Strix occidentalis caurina*. Panel B depicts the demographic history estimated for *Strix varia*.

References

Altschul SF., Madden TL., Schäffer AA., Zhang J., Zhang Z., Miller W., Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25:3389–3402. DOI: 10.1093/nar/25.17.3389.

Bao W., Kojima KK., Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* 6:1–6. DOI: 10.1186/s13100-015-0041-9.

Barrowclough GF., Groth JG., Mertz LA., Gutiérrez RJ. 2005. Genetic structure, introgression, and a narrow hybrid zone between northern and California spotted owls (*Strix occidentalis*). *Molecular Ecology* 14:1109–1120. DOI: 10.1111/j.1365-294X.2005.02465.x.

Barrowclough GF., Gutierrez RJ., Groth JG. 1999. Phylogeography of Spotted Owl (*Strix occidentalis*) Populations Based on Mitochondrial DNA Sequences: Gene Flow, Genetic Structure, and a Novel Biogeographic Pattern. *Evolution* 53:919–931. DOI: 10.2307/2640731.

Barrowclough GF., Gutiérrez RJ., Groth JG., Lai JE., Rock DF. 2011. The Hybrid Zone between Northern and California Spotted Owls in the Cascade—Sierran Suture Zone. *The Condor* 113:581–589. DOI: 10.1525/cond.2011.100203.

Bellingham J., Chaurasia SS., Melyan Z., Liu C., Cameron MA., Tarttelin EE., Iuvone PM., Hankins MW., Tosini G., Lucas RJ. 2006. Evolution of Melanopsin Photoreceptors: Discovery and Characterization of a New Melanopsin in Nonmammalian Vertebrates. *PLOS Biol* 4:e254. DOI: 10.1371/journal.pbio.0040254.

Belterman RHR., Boer LEMD. 1984. A karyological study of 55 species of birds, including karyotypes of 39 species new to cytology. *Genetica* 65:39–82. DOI: 10.1007/BF00056765.

Bernt M., Donath A., Jühling F., Externbrink F., Florentz C., Fritzsch G., Pütz J., Middendorf M., Stadler PF. 2013. MITOS: Improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution* 69:313–319. DOI: 10.1016/j.ympev.2012.08.023.

Biomatters 2016a. Geneious. Version 9.1.6. [Accessed 2016 Oct 1]. Available from: http://www.geneious.com.

Biomatters 2016b. Geneious. Version 9.1.4. [Accessed 2016 Oct 1]. Available from: http://www.geneious.com.

Blackshaw S., Snyder SH. 1999. Encephalopsin: A Novel Mammalian Extraretinal Opsin Discretely Localized in the Brain. *The Journal of Neuroscience* 19:3681–3690.

Bolger AM., Lohse M., Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. DOI: 10.1093/bioinformatics/btu170.

Borges R., Khan I., Johnson WE., Gilbert MT., Zhang G., Jarvis ED., O'Brien SJ., Antunes A. 2015. Gene loss, adaptive evolution and the co-evolution of plumage coloration genes with opsins in birds. *BMC Genomics* 16:751. DOI: 10.1186/s12864-015-1924-3.

Bowmaker JK. 1980. Colour vision in birds and the role of oil droplets. *Trends in Neurosciences* 3:196–199. DOI: 10.1016/0166-2236(80)90072-7.

Bowmaker JK., Martin GR. 1978. Visual pigments and colour vision in a nocturnal bird, *Strix aluco* (tawny owl). *Vision Research* 18:1125–1130. DOI: 10.1016/0042-6989(78)90095-0.

Bowmaker JK., Martin GR. 1985. Visual pigments and oil droplets in the penguin,*Spheniscus humboldti*. *Journal of Comparative Physiology A* 156:71–77. DOI: 10.1007/BF00610668.

Bradnam K. 2015. Goodbye CEGMA, hello BUSCO! *ACGT*. [Accessed 2016 Jun 24]. Available from: http://www.acgt.me/blog/2015/5/18/goodbye-cegma-hello-busco.

Bradnam KR., Fass JN., Alexandrov A., Baranay P., Bechner M., Birol I., Boisvert S., Chapman JA., Chapuis G., Chikhi R., Chitsaz H., Chou W-C., Corbeil J., Del Fabbro C., Docking TR., Durbin R., Earl D., Emrich S., Fedotov P., Fonseca NA., Ganapathy G., Gibbs RA., Gnerre S., Godzaridis E., Goldstein S., Haimel M., Hall G., Haussler D., Hiatt JB., Ho IY., Howard J., Hunt M., Jackman SD., Jaffe DB., Jarvis ED., Jiang H., Kazakov S., Kersey PJ., Kitzman JO., Knight JR., Koren S., Lam T-W., Lavenier D., Laviolette F., Li Y., Li Z., Liu B., Liu Y., Luo R., MacCallum I., MacManes MD., Maillet N., Melnikov S., Naquin D., Ning Z., Otto TD., Paten B., Paulo OS., Phillippy AM., Pina-Martins F., Place M., Przybylski D., Qin X., Qu C., Ribeiro FJ., Richards S., Rokhsar DS., Ruby JG., Scalabrin S., Schatz MC., Schwartz DC., Sergushichev A., Sharpe T., Shaw TI., Shendure J., Shi Y., Simpson JT., Song H., Tsarev F., Vezzi F., Vicedomini R., Vieira BM., Wang J., Worley KC., Yin S., Yiu S-M., Yuan J., Zhang G., Zhang H., Zhou S., Korf IF. 2013. Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. *GigaScience* 2:10. DOI: 10.1186/2047-217X-2-10.

Bushnell B. 2014. BBMap. Version 34.00. [Accessed 2016 Oct 1]. Available from: http://sourceforge.net/projects/bbmap.

Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. DOI: 10.1186/1471-2105-10-421.

Campbell MS., Holt C., Moore B., Yandell M. 2014. Genome Annotation and Curation Using MAKER and MAKER-P. *Current Protocols in Bioinformatics* 48:4.11.1-4.11.39. DOI: 10.1002/0471250953.bi0411s48.

Cantarel BL., Korf I., Robb SMC., Parra G., Ross E., Moore B., Holt C., Alvarado AS., Yandell M. 2008. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research* 18:188–196. DOI: 10.1101/gr.6743907.

Carleton KL., Spady TC., Cote RH. 2005. Rod and Cone Opsin Families Differ in Spectral Tuning Domains but Not Signal Transducing Domains as Judged by Saturated Evolutionary Trace Analysis. *Journal of Molecular Evolution* 61:75–89. DOI: 10.1007/s00239-004-0289-z.

Castoe TA., Koning APJ de., Hall KT., Card DC., Schield DR., Fujita MK., Ruggiero RP., Degner JF., Daza JM., Gu W., Reyes-Velasco J., Shaney KJ., Castoe JM., Fox SE., Poole AW., Polanco D., Dobry J., Vandewege MW., Li Q., Schott RK., Kapusta A., Minx P., Feschotte C., Uetz P., Ray DA., Hoffmann FG., Bogden R., Smith EN., Chang BSW., Vonk FJ., Casewell NR., Henkel CV., Richardson MK., Mackessy SP., Bronikowsi AM., Yandell M., Warren WC., Secor SM., Pollock DD. 2013. The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. *Proceedings of the National Academy of Sciences*:201314475. DOI: 10.1073/pnas.1314475110.

Clark K., Karsch-Mizrachi I., Lipman DJ., Ostell J., Sayers EW. 2016. GenBank. *Nucleic Acids Research* 44:D67–D72. DOI: 10.1093/nar/gkv1276.

Consortium TU. 2015. UniProt: a hub for protein information. *Nucleic Acids Research* 43:D204–D212. DOI: 10.1093/nar/gku989.

Courtney SP., Blakesley JA., Bigley RE., Cody ML., Dumbacher JP., Fleischer RC., Franklin AB., Franklin JF., Gutiérrez RJ., Marzluff JM., Sztukowski L. 2004. *Scientific evaluation of the status of the Northern Spotted Owl*. Sustainable Ecosystems Institute: Portland, OR.

Davies WL., Carvalho LS., Cowing JA., Beazley LD., Hunt DM., Arrese CA. 2007. Visual pigments of the platypus: A novel route to mammalian colour vision. *Current Biology* 17:R161–R163. DOI: 10.1016/j.cub.2007.01.037.

Davies WIL., Collin SP., Hunt DM. 2012. Molecular ecology and adaptation of visual photopigments in craniates. *Molecular Ecology* 21:3121–3158. DOI: 10.1111/j.1365-294X.2012.05617.x.

Davis RJ., Hollen B., Hobson J., Gower JE., Keenum D. 2016. Northwest Forest Plan—the first 20 years (1994-2013): status and trends of northern spotted owl habitats. [Accessed 2016 Oct 7]. Available from: http://www.treesearch.fs.fed.us/pubs/50567.

De Vita R., Cavallo D., Eleuteri P., Dell'Omo G. 1994. Evaluation of interspecific DNA content variations and sex identification in Falconiformes and Strigiformes by flow cytometric analysis. *Cytometry* 16:346–350. DOI: 10.1002/cyto.990160409.

DePristo MA., Banks E., Poplin R., Garimella KV., Maguire JR., Hartl C., Philippakis AA., del Angel G., Rivas MA., Hanna M., McKenna A., Fennell TJ., Kernytsky AM., Sivachenko AY., Cibulskis K., Gabriel SB., Altshuler D., Daly MJ. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* 43:491–498. DOI: 10.1038/ng.806.

Diller LV., Hamm KA., Early DA., Lamphear DW., Dugger KM., Yackulic CB., Schwarz CJ., Carlson PC., McDonald TL. 2016. Demographic response of northern spotted owls to barred owl removal. *The Journal of Wildlife Management* 80:691–707. DOI: 10.1002/jwmg.1046.

Doležel J., Bartoš J., Voglmayr H., Greilhuber J. 2003. Letter to the editor. *Cytometry Part A* 51A:127–128. DOI: 10.1002/cyto.a.10013.

Dugger KM., Forsman ED., Franklin AB., Davis RJ., White GC., Schwarz CJ., Burnham KP., Nichols JD., Hines JE., Yackulic CB., Doherty PF., Bailey L., Clark DA., Ackers SH., Andrews LS., Augustine B., Biswell BL., Blakesley J., Carlson PC., Clement MJ., Diller LV., Glenn EM., Green A., Gremel SA., Herter DR., Higley JM., Hobson J., Horn RB., Huyvaert KP., McCafferty C., McDonald T., McDonnell K., Olson GS., Reid JA., Rockweit J., Ruiz V., Saenz J., Sovern SG. 2015. The effects of habitat, climate, and Barred Owls on long-term demography of Northern Spotted Owls. *The Condor* 118:57–116. DOI: 10.1650/CONDOR-15-24.1.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32:1792–1797. DOI: 10.1093/nar/gkh340.

Emerling CA. 2017a. Archelosaurian Color Vision, Parietal Eye Loss, and the Crocodylian Nocturnal Bottleneck. *Molecular Biology and Evolution* 34:666–676. DOI: 10.1093/molbev/msw265.

Emerling CA. 2017b. Independent pseudogenization of CYP2J19 in penguins, owls and kiwis implicates gene in red carotenoid synthesis. *bioRxiv*:130468. DOI: 10.1101/130468.

Emerling CA. 2017c. Genomic regression of claw keratin, taste receptor and light-associated genes provides insights into biology and evolutionary origins of snakes. *Molecular Phylogenetics and Evolution* 115:40–49. DOI: 10.1016/j.ympev.2017.07.014.

Emerling CA., Huynh HT., Nguyen MA., Meredith RW., Springer MS. 2015. Spectral shifts of mammalian ultraviolet-sensitive pigments (short wavelength-sensitive opsin 1) are associated with eye length and photic niche evolution. *Proc. R. Soc. B* 282:20151817. DOI: 10.1098/rspb.2015.1817.

Erhard H. 1924. Messende Untersuchungen über den Farbensinn der Vögel. *Zool. Jahrb., Allg. Zool. Physiol* 41:489–552.

Falcón J., Besseau L., Fuentès M., Sauzet S., Magnanou E., Boeuf G. 2009. Structural and Functional Evolution of the Pineal Melatonin System in Vertebrates. *Annals of the New York Academy of Sciences* 1163:101–111. DOI: 10.1111/j.1749-6632.2009.04435.x.

Finn RD., Coggill P., Eberhardt RY., Eddy SR., Mistry J., Mitchell AL., Potter SC., Punta M., Qureshi M., Sangrador-Vegas A., Salazar GA., Tate J., Bateman A. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research* 44:D279–D285. DOI: 10.1093/nar/gkv1344.

Fite KV. 1973. Anatomical and behavioral correlates of visual acuity in the great horned owl. *Vision Research* 13:219. DOI: 10.1016/0042-6989(73)90101-6.

Forsman ED., Anthony RG., Dugger KM., Glenn EM., Franklin AB., White GC., Schwarz CJ., Burnham KP., Anderson DR., Nichols JD., Hines JE., Lint JB., Davis RJ., Ackers SH., Andrews LS., Biswell BL., Carlson PC., Diller LV., Gremel SA., Herter DR., Higley JM., Horn RB., Reid JA., Rockweit J., Schaberl JP., Snetsinger TJ., Sovern SG. 2011. *Population Demography of Northern Spotted Owls: Published for the Cooper Ornithological Society*. Berkeley and Los Angeles, California: University of California Press.

Forsman ED., DeStefano S., Raphael MG., Gutiérrez RJ. (eds.) 1996. *Demography of the Northern Spotted Owl*. Lawrence, Kansas: Allen Press, Inc.

Free Software Foundation 2012. GNU Awk . Version 4.0.1. [Accessed 2016 Oct 1]. Available from: https://www.gnu.org/software/gawk.

Fridolfsson A-K., Ellegren H. 1999. A Simple and Universal Method for Molecular Sexing of Non-Ratite Birds. *Journal of Avian Biology* 30:116–121. DOI: 10.2307/3677252.

Funk WC., Forsman ED., Johnson M., Mullins TD., Haig SM. 2010. Evidence for recent population bottlenecks in northern spotted owls (Strix occidentalis caurina). *Conservation Genetics* 11:1013–1021. DOI: 10.1007/s10592-009-9946-5.

Funk WC., Forsman ED., Mullins TD., Haig SM. 2008. Introgression and dispersal among spotted owl (*Strix occidentalis*) subspecies. *Evolutionary Applications* 1:161–171. DOI: 10.1111/j.1752-4571.2007.00002.x.

Funk WC., Mullins TD., Forsman ED., Haig SM. 2007. Microsatellite loci for distinguishing spotted owls (*Strix occidentalis*), barred owls (*Strix varia*), and their hybrids. *Molecular Ecology Notes* 7:284–286. DOI: 10.1111/j.1471-8286.2006.01581.x.

Gerkema MP., Davies WIL., Foster RG., Menaker M., Hut RA. 2013. The nocturnal bottleneck and the evolution of activity patterns in mammals. *Proceedings of the Royal Society B: Biological Sciences* 280. DOI: 10.1098/rspb.2013.0508.

Gnerre S., MacCallum I., Przybylski D., Ribeiro FJ., Burton JN., Walker BJ., Sharpe T., Hall G., Shea TP., Sykes S., Berlin AM., Aird D., Costello M., Daza R., Williams L., Nicol R., Gnirke A., Nusbaum C., Lander ES., Jaffe DB. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of the National Academy of Sciences* 108:1513–1518. DOI: 10.1073/pnas.1017351108.

Gondo M., Ando H. 1995. Comparative and histophysiological study of oil droplets in the avian retina. *The Kobe Journal of Medical Sciences* 41:127–139.

Gremme G., Steinbiss S., Kurtz S. 2013. GenomeTools: A Comprehensive Software Library for Efficient Processing of Structured Genome Annotations. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 10:645–656. DOI: 10.1109/TCBB.2013.68.

Guan X., Xu J., Smith EJ. 2016. The complete mitochondrial genome sequence of the budgerigar, *Melopsittacus undulatus*. *Mitochondrial DNA Part A* 27:401–402. DOI: 10.3109/19401736.2014.898277.

Haig SM., Mullins TD., Forsman ED., Trail PW., Wennerberg L. 2004. Genetic identification of spotted owls, barred owls, and their hybrids: legal implications of hybrid identity. *Conservation Biology* 18:1347–1357.

Haig SM., Wagner RS., Forsman ED., Mullins TD. 2001. Geographic variation and genetic structure in Spotted Owls. *Conservation Genetics* 2:25–40. DOI: 10.1023/A:1011561101460.

Halford S., Freedman Melanie S., Bellingham J., Inglis SL., Poopalasundaram S., Soni BG., Foster RG., Hunt DM. 2001. Characterization of a Novel Human Opsin Gene with Wide Tissue Expression and Identification of Embedded and Flanking Genes on Chromosome 1q43. *Genomics* 72:203–208. DOI: 10.1006/geno.2001.6469.

Hamer TE., Forsman ED., Fuchs AD., Walters ML. 1994. Hybridization between Barred and Spotted Owls. *The Auk* 111:487–492. DOI: 10.2307/4088616.

Hammar B. 1970. The karyotypes of thirty-one birds. *Hereditas* 65:29–58. DOI: 10.1111/j.1601-5223.1970.tb02306.x.

Hankins MW., Peirson SN., Foster RG. 2008. Melanopsin: an exciting photopigment. *Trends in Neurosciences* 31:27–36. DOI: 10.1016/j.tins.2007.11.002.

Hanna ZR., Henderson JB., Wall JD., Emerling CA., Fuchs J., Runckel C., Mindell DP., Bowie RCK., DeRisi JL., Dumbacher JP. 2017. Supplemental dataset for Northern Spotted Owl (*Strix occidentalis caurina*) genome assembly version 1.0. *Zenodo*. DOI: 10.5281/zenodo.822859.

Harrison GL (Abby)., McLenachan PA., Phillips MJ., Slack KE., Cooper A., Penny D. 2004. Four New Avian Mitochondrial Genomes Help Get to Basic Evolutionary Questions in the Late Cretaceous. *Molecular Biology and Evolution* 21:974–983. DOI: 10.1093/molbev/msh065.

Heesy CP., Hall MI. 2010. The Nocturnal Bottleneck and the Evolution of Mammalian Vision. *Brain, Behavior and Evolution* 75:195–203. DOI: 10.1159/000314278.

Henderson JB., Hanna ZR. 2016a. scafSeqContigInfo. Version 1.0.0. *Zenodo*. DOI: 10.5281/zenodo.163748.

Henderson JB., Hanna ZR. 2016b. dupchk. Version 1.0.0. *Zenodo*. DOI: 10.5281/zenodo.163722.

Henderson JB., Hanna ZR. 2016c. scafN50. Version 1.0.0. *Zenodo*. DOI: 10.5281/zenodo.163739.

Henderson JB., Hanna ZR. 2016d. ScaffSplitN50s. Version 1.0.0. *Zenodo*. DOI: 10.5281/zenodo.163683.

Henderson JB., Hanna ZR. 2016e. GItaxidIsVert. Version 1.0.0. *Zenodo*. DOI: 10.5281/zenodo.163737.

Hengjiu T., Jianwei J., Shi Y., Zhiming Z., Laghari MY., Narejo NT., Lashari P. 2016. Complete mitochondrial genome of Eagle Owl (*Bubo bubo*, Strigiformes; Strigidae) from China. *Mitochondrial DNA Part A* 27:1455–1456. DOI: 10.3109/19401736.2014.953090.

Hsu Y-C., Li S-H., Lin Y-S., Severinghaus LL. 2006. Microsatellite Loci from Lanyu Scops Owl (*Otus elegans botelensis*) and their Cross-Species Application in Four Species of Strigidae. *Conservation Genetics* 7:161–165. DOI: 10.1007/s10592-005-5477-x.

Hsu Y-C., Severinghaus LL., Lin Y-S., Li S-H. 2003. Isolation and characterization of microsatellite DNA markers from the Lanyu scops owl (*Otus elegans botelensis*). *Molecular Ecology Notes* 3:595–597. DOI: 10.1046/j.1471-8286.2003.00523.x.

Hudson RR., Slatkin M., Maddison WP. 1992. Estimation of Levels of Gene Flow from DNA Sequence Data. *Genetics* 132:583–589.

Hughes S., Welsh L., Katti C., González-Menéndez I., Turton M., Halford S., Sekaran S., Peirson SN., Hankins MW., Foster RG. 2012. Differential Expression of Melanopsin Isoforms Opn4L and Opn4S during Postnatal Development of the Mouse Retina. *PLOS ONE* 7:e34531. DOI: 10.1371/journal.pone.0034531.

Hunter JD. 2007. Matplotlib: A 2D graphics environment. *Computing In Science & Engineering* 9:90–95. DOI: 10.1109/MCSE.2007.55.

Ihnat DM., MacKenzie D., Meyering J. 2013. cut (GNU coreutils). Version 8.21. [Accessed 2016 Oct 1]. Available from: http://www.gnu.org/software/coreutils/coreutils.html.

Isaksson M., Tegelström H. 2002. Characterization of polymorphic microsatellite markers in a captive population of the eagle owl (Bubo bubo) used for supportive breeding. *Molecular Ecology Notes* 2:91–93. DOI: 10.1046/j.1471-8286.2002.00156.x.

Jacobs GH. 2013. Losses of functional opsin genes, short-wavelength cone photopigments, and color vision—A significant trend in the evolution of mammalian vision. *Visual Neuroscience* 30:39–53. DOI: 10.1017/S0952523812000429.

Jarvis ED., Mirarab S., Aberer AJ., Li B., Houde P., Li C., Ho SYW., Faircloth BC., Nabholz B., Howard JT., Suh A., Weber CC., Fonseca RR da., Li J., Zhang F., Li H., Zhou L., Narula N., Liu L., Ganapathy G., Boussau B., Bayzid MS., Zavidovych V., Subramanian S., Gabaldón T., Capella-Gutiérrez S., Huerta-Cepas J., Rekepalli B., Munch K., Schierup M., Lindow B., Warren WC., Ray D., Green RE., Bruford MW., Zhan X., Dixon A., Li S., Li N., Huang Y., Derryberry EP., Bertelsen MF., Sheldon FH., Brumfield RT., Mello CV., Lovell PV., Wirthlin M., Schneider MPC., Prosdocimi F., Samaniego JA., Velazquez AMV., Alfaro-Núñez A., Campos PF., Petersen B., Sicheritz-Ponten T., Pas A., Bailey T., Scofield P., Bunce M., Lambert DM., Zhou Q., Perelman P., Driskell AC., Shapiro B., Xiong Z., Zeng Y., Liu S., Li Z., Liu B., Wu K., Xiao J., Yinqi X., Zheng Q., Zhang Y., Yang H., Wang J., Smeds L., Rheindt FE., Braun M., Fjeldsa J., Orlando L., Barker FK., Jønsson KA., Johnson W., Koepfli K-P., O'Brien S., Haussler D., Ryder OA., Rahbek C., Willerslev E., Graves GR., Glenn TC., McCormack J., Burt D., Ellegren H., Alström P., Edwards SV., Stamatakis A., Mindell DP., Cracraft J., Braun EL., Warnow T., Jun W., Gilbert MTP., Zhang G. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 346:1320–1331. DOI: 10.1126/science.1253451.

Jennings S., Cormier RL., Gardali T., Press D., Merkle WW. 2011. Status and Distribution of the Barred Owl in Marin County, California. *Western Birds* 42:103–110.

Jones P., Binns D., Chang H-Y., Fraser M., Li W., McAnulla C., McWilliam H., Maslen J., Mitchell A., Nuka G., Pesseat S., Quinn AF., Sangrador-Vegas A., Scheremetjew M.,

Yong S-Y., Lopez R., Hunter S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30:1236–1240. DOI: 10.1093/bioinformatics/btu031.
Jurka J. 1998. Repeats in genomic DNA: mining and meaning. *Current Opinion in Structural Biology* 8:333–337. DOI: 10.1016/S0959-440X(98)80067-5.
Jurka J. 2000. Repbase Update: a database and an electronic journal of repetitive elements. *Trends in Genetics* 16:418–420. DOI: 10.1016/S0168-9525(00)02093-X.
Jurka J., Kapitonov VV., Pavlicek A., Klonowski P., Kohany O., Walichiewicz J. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research* 110:462–467. DOI: 10.1159/000084979.
Kato T., Todo T., Ayaki H., Ishizaki K., Morita T., Mitra S., Ikenaga M. 1994. Cloning of a marsupial DNA photolyase gene and the lack of related nucleotide sequences in placental mammals. *Nucleic Acids Research* 22:4119–4124.
Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A., Markowitz S., Duran C., Thierer T., Ashton B., Meintjes P., Drummond A. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649. DOI: 10.1093/bioinformatics/bts199.
Keller O., Kollmar M., Stanke M., Waack S. 2011. A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics* 27:757–763. DOI: 10.1093/bioinformatics/btr010.
Kelly EG., Forsman ED. 2004. Recent Records of Hybridization Between Barred Owls (*Strix varia*) and Northern Spotted Owls (*S. occidentalis caurina*). *The Auk* 121:806–810. DOI: 10.1642/0004-8038(2004)121[0806:RROHBB]2.0.CO;2.
Koopman ME., Schable NA., Glenn TC. 2004. Development and optimization of microsatellite DNA primers for boreal owls (*Aegolius funereus*). *Molecular Ecology Notes* 4:376–378. DOI: 10.1111/j.1471-8286.2004.00658.x.
Korf I. 2004. Gene finding in novel genomes. *BMC Bioinformatics* 5:1–9. DOI: 10.1186/1471-2105-5-59.
Le Duc D., Renaud G., Krishnan A., Almén MS., Huynen L., Prohaska SJ., Ongyerth M., Bitarello BD., Schiöth HB., Hofreiter M., Stadler PF., Prüfer K., Lambert D., Kelso J., Schöneberg T. 2015. Kiwi genome provides insights into evolution of a nocturnal lifestyle. *Genome Biology* 16:1–15. DOI: 10.1186/s13059-015-0711-4.
Leinonen R., Sugawara H., Shumway M. 2011. The Sequence Read Archive. *Nucleic Acids Research* 39:D19–D21. DOI: 10.1093/nar/gkq1019.
Li H. 2013a. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv:1303.3997 Q-Bio. [Accessed 2016 Feb 16]. Available from: http://arxiv.org/abs/1303.3997.
Li H. 2013b. bioawk. Version 1.0. [Accessed 2016 Oct 1]. Available from: https://github.com/lh3/bioawk.
Li H. 2015. PSMC. Version 0.6.5-r67. [Accessed 2016 Oct 1]. Available from: https://github.com/lh3/psmc.
Li H., Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature* 475:493–496. DOI: 10.1038/nature10231.
Li C., Zhang Y., Li J., Kong L., Hu H., Pan H., Xu L., Deng Y., Li Q., Jin L., Yu H., Chen Y., Liu B., Yang L., Liu S., Zhang Y., Lang Y., Xia J., He W., Shi Q., Subramanian S., Millar CD., Meader S., Rands CM., Fujita MK., Greenwold MJ., Castoe TA., Pollock

DD., Gu W., Nam K., Ellegren H., Ho SY., Burt DW., Ponting CP., Jarvis ED., Gilbert MTP., Yang H., Wang J., Lambert DM., Wang J., Zhang G. 2014. Two Antarctic penguin genomes reveal insights into their evolutionary history and molecular changes related to the Antarctic environment. *GigaScience* 3:1–15. DOI: 10.1186/2047-217X-3-27.

Liu G., Zhou L., Gu C. 2014. The complete mitochondrial genome of Brown wood owl *Strix leptogrammica* (Strigiformes: Strigidae). *Mitochondrial DNA* 25:370–371. DOI: 10.3109/19401736.2013.803540.

Livezey KB. 2009a. Range Expansion of Barred Owls, Part I: Chronology and Distribution. *The American Midland Naturalist* 161:49–56. DOI: 10.1674/0003-0031-161.1.49.

Livezey KB. 2009b. Range Expansion of Barred Owls, Part II: Facilitating Ecological Changes. *The American Midland Naturalist* 161:323–349. DOI: 10.1674/0003-0031-161.2.323.

Lopes RJ., Johnson JD., Toomey MB., Ferreira MS., Araujo PM., Melo-Ferreira J., Andersson L., Hill GE., Corbo JC., Carneiro M. 2016. Genetic Basis for Red Coloration in Birds. *Current Biology* 26:1427–1434. DOI: 10.1016/j.cub.2016.03.076.

Luo R., Liu B., Xie Y., Li Z., Huang W., Yuan J., He G., Chen Y., Pan Q., Liu Y., Tang J., Wu G., Zhang H., Shi Y., Liu Y., Yu C., Wang B., Lu Y., Han C., Cheung DW., Yiu S-M., Peng S., Xiaoqian Z., Liu G., Liao X., Li Y., Yang H., Wang J., Lam T-W., Wang J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* 1:18. DOI: 10.1186/2047-217X-1-18.

Mahmood MT., McLenachan PA., Gibb GC., Penny D. 2014. Phylogenetic Position of Avian Nocturnal and Diurnal Raptors. *Genome Biology and Evolution* 6:326–332. DOI: 10.1093/gbe/evu016.

Mazur KM., James PC. 2000. Barred Owl (*Strix varia*). *The Birds of North America Online* (A. Poole, Ed.) Ithaca: Cornell Lab of Ornithology. [Accessed 2016 Oct 1]. Retrieved from the Birds of North America Online: https://birdsna.org/Species-Account/bna/species/brdowl. DOI: 10.2173/bna.508.

McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernytsky A., Garimella K., Altshuler D., Gabriel S., Daly M., DePristo MA. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20:1297–1303. DOI: 10.1101/gr.107524.110.

Mindell DP., Sorenson MD., Dimcheff DE. 1998. Multiple independent origins of mitochondrial gene order in birds. *Proceedings of the National Academy of Sciences* 95:10693–10697.

Mindell DP., Sorenson MD., Dimcheff DE., Hasegawa M., Ast JC., Yuri T. 1999. Interordinal Relationships of Birds and Other Reptiles Based on Whole Mitochondrial Genomes. *Systematic Biology* 48:138–152. DOI: 10.1080/106351599260490.

Mindell DP., Sorenson MD., Huddleston CJ., Miranda HC., Knight A., Sawchuk SJ., Yuri T. 1997. Phylogenetic Relationships among and within Select Avian Orders Based on Mitochondrial DNA. In: *Avian Molecular Evolution and Systematics*. San Diego, California: Academic Press, 211–245.

Mundy NI., Stapley J., Bennison C., Tucker R., Twyman H., Kim K-W., Burke T., Birkhead TR., Andersson S., Slate J. 2016. Red Carotenoid Coloration in the Zebra Finch Is Controlled by a Cytochrome P450 Gene Cluster. *Current Biology* 26:1435–1440. DOI: 10.1016/j.cub.2016.04.047.

Nakamura D., Tiersch TR., Douglass M., Chandler RW. 1990. Rapid identification of sex in birds by flow cytometry. *Cytogenetics and Cell Genetics* 53:201–205.

NCBI Resource Coordinators 2016. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research* 44:D7–D19. DOI: 10.1093/nar/gkv1290.
O'Connell J. 2014. NxTrim. Version 0.2.3-alpha. [Accessed 2016 Oct 1]. Available from: https://github.com/sequencing/NxTrim.
O'Connell J., Schulz-Trieglaff O., Carlson E., Hims MM., Gormley NA., Cox AJ. 2015. NxTrim: optimized trimming of Illumina mate pair reads. *Bioinformatics* 31:2035–2037. DOI: 10.1093/bioinformatics/btv057.
Okano T., Yoshizawa T., Fukada Y. 1994. Pinopsin is a chicken pineal photoreceptive molecule. *Nature* 372:94–97. DOI: 10.1038/372094a0.
Osborn AR., Almabruk KH., Holzwarth G., Asamizu S., LaDu J., Kean KM., Karplus PA., Tanguay RL., Bakalinsky AT., Mahmud T. 2015. De novo synthesis of a sunscreen compound in vertebrates. *eLife* 4:e05919. DOI: 10.7554/eLife.05919.
Parra G., Bradnam K., Korf I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23:1061–1067. DOI: 10.1093/bioinformatics/btm071.
Partridge JC. 1989. The visual ecology of avian cone oil droplets. *Journal of Comparative Physiology A* 165:415–426. DOI: 10.1007/BF00619360.
Proudfoot G., Honeycutt R., Douglas Slack R. 2005. Development and characterization of microsatellite DNA primers for ferruginous pygmy-owls (*Glaucidium brasilianum*). *Molecular Ecology Notes* 5:90–92. DOI: 10.1111/j.1471-8286.2004.00842.x.
Prum RO., Berv JS., Dornburg A., Field DJ., Townsend JP., Lemmon EM., Lemmon AR. 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* 526:569–573. DOI: 10.1038/nature15697.
Rebholz WER., Boer LEMD., Sasaki M., Belterman RHR., Nishida-Umehara C. 1993. The Chromosomal Phylogeny of Owls (Strigiformes) and New Karyotypes of Seven Species. *Cytologia* 58:403–416. DOI: 10.1508/cytologia.58.403.
Renzoni A., Vegni-Talluri M. 1966. The karyograms of some Falconiformes and Strigiformes. *Chromosoma* 20:133–150. DOI: 10.1007/BF00335203.
Roth JJ., Gern WA., Roth EC., Ralph CL., Jacobson E. 1980. Nonpineal melatonin in the alligator (*Alligator mississippiensis*). *Science* 210:548–550. DOI: 10.1126/science.7423204.
Shen D., Jiang M., Hao W., Tao L., Salazar M., Fong HK. 1994. A human opsin-related gene that encodes a retinaldehyde-binding protein. *Biochemistry* 33:13117–13125.
Simão FA., Waterhouse RM., Ioannidis P., Kriventseva EV., Zdobnov EM. 2015a. BUSCO. Version 1.1b1. [Accessed 2016 Oct 1]. Available from: http://busco.ezlab.org.
Simão FA., Waterhouse RM., Ioannidis P., Kriventseva EV., Zdobnov EM. 2015b. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212. DOI: 10.1093/bioinformatics/btv351.
Simões BF., Sampaio FL., Jared C., Antoniazzi MM., Loew ER., Bowmaker JK., Rodriguez A., Hart NS., Hunt DM., Partridge JC., Gower DJ. 2015. Visual system evolution and the nature of the ancestral snake. *Journal of Evolutionary Biology* 28:1309–1320. DOI: 10.1111/jeb.12663.
Simpson JT. 2014. Exploring genome characteristics and sequence quality without a reference. *Bioinformatics* 30:1228–1235. DOI: 10.1093/bioinformatics/btu023.
Smit AFA., Hubley R. 2015. RepeatModeler Open-1.0. [Accessed 2016 Oct 1]. Available from: http://www.repeatmasker.org.

Smit A., Hubley R., Green P. 2013. RepeatMasker Open-4.0. [Accessed 2016 Oct 1]. Available from: http://www.repeatmasker.org.

Soni BG., Foster RG. 1997. A novel and ancient vertebrate opsin. *FEBS Letters* 406:279–283. DOI: 10.1016/S0014-5793(97)00287-1.

Stanke M. 2015. AUGUSTUS. Version 3.2.1. [Accessed 2016 Oct 1]. Available from: http://bioinf.uni-greifswald.de/augustus.

Sun H., Gilbert DJ., Copeland NG., Jenkins NA., Nathans J. 1997. Peropsin, a novel visual pigment-like protein located in the apical microvilli of the retinal pigment epithelium. *Proceedings of the National Academy of Sciences of the United States of America* 94:9893–9898.

Taniguchi M., Murakami N., Nakamura H., Nasu T., Shinohara S., Etoh T. 1993. Melatonin release from pineal cells of diurnal and nocturnal birds. *Brain Research* 620:297–300. DOI: 10.1016/0006-8993(93)90169-N.

Tarttelin EE., Bellingham J., Bibb LC., Foster RG., Hankins MW., Gregory-Evans K., Gregory-Evans CY., Wells DJ., Lucas RJ. 2003. Expression of opsin genes early in ocular development of humans and mice. *Experimental Eye Research* 76:393–396. DOI: 10.1016/S0014-4835(02)00300-7.

Taylor AL., Forsman ED. 1976. Recent range extensions of the Barred Owl in western North America, including the first records for Oregon. *Condor* 78:560–561.

Thode AB., Maltbie M., Hansen LA., Green LD., Longmire JL. 2002. Microsatellite markers for the Mexican spotted owl (*Strix occidentalis lucida*). *Molecular Ecology Notes* 2:446–448. DOI: 10.1046/j.1471-8286.2002.00267.x.

Thomas JW., Forsman ED., Lint JB., Meslow EC., Noon BR., Verner J. 1990. A conservation strategy for the northern spotted owl: report of the Interagency Scientific Committee to address the conservation of the northern spotted owl. USDA Forest Service; USDI Bureau of Land Management, Fish and Wildlife Service, National Park Service: Portland, Oregon.

Toews DPL., Taylor SA., Vallender R., Brelsford A., Butcher BG., Messer PW., Lovette IJ. 2016. Plumage Genes and Little Else Distinguish the Genomes of Hybridizing Warblers. *Current Biology* 26:2313–2318. DOI: 10.1016/j.cub.2016.06.034.

Tomonari S., Migita K., Takagi A., Noji S., Ohuchi H. 2008. Expression patterns of the opsin 5–related genes in the developing chicken retina. *Developmental Dynamics* 237:1910–1922. DOI: 10.1002/dvdy.21611.

Twyman H., Valenzuela N., Literman R., Andersson S., Mundy NI. 2016. Seeing red to being red: conserved genetic mechanism for red cone oil droplets and co-option for red coloration in birds and turtles. *Proc. R. Soc. B* 283:20161208. DOI: 10.1098/rspb.2016.1208.

Van der Auwera GA., Carneiro MO., Hartl C., Poplin R., del Angel G., Levy-Moonshine A., Jordan T., Shakir K., Roazen D., Thibault J., Banks E., Garimella KV., Altshuler D., Gabriel S., DePristo MA. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics* 11:11.10.1-11.10.33. DOI: 10.1002/0471250953.bi1110s43.

Verra DM., Contín MA., Hicks D., Guido ME. 2011. Early Onset and Differential Temporospatial Expression of Melanopsin Isoforms in the Developing Chicken Retina. *Investigative Ophthalmology & Visual Science* 52:5111–5120. DOI: 10.1167/iovs.11-75301.

Vinogradov AE. 2005. Genome size and chromatin condensation in vertebrates. *Chromosoma* 113:362–369. DOI: 10.1007/s00412-004-0323-3.

Vonk FJ., Casewell NR., Henkel CV., Heimberg AM., Jansen HJ., McCleary RJR., Kerkkamp HME., Vos RA., Guerreiro I., Calvete JJ., Wüster W., Woods AE., Logan JM., Harrison RA., Castoe TA., Koning APJ de., Pollock DD., Yandell M., Calderon D., Renjifo C., Currier RB., Salgado D., Pla D., Sanz L., Hyder AS., Ribeiro JMC., Arntzen JW., Thillart GEEJM van den., Boetzer M., Pirovano W., Dirks RP., Spaink HP., Duboule D., McGlinn E., Kini RM., Richardson MK. 2013. The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. *Proceedings of the National Academy of Sciences* 110:20651–20656. DOI: 10.1073/pnas.1314702110.

Vorobyev M. 2003. Coloured oil droplets enhance colour discrimination. *Proceedings of the Royal Society of London B: Biological Sciences* 270:1255–1261. DOI: 10.1098/rspb.2003.2381.

Walls GL. 1942. *The vertebrate eye and its adaptive radiation.* Bloomfield Hills, Michigan: Cranbrook Institute of Science. DOI: 10.5962/bhl.title.7369.

Warren W., Bussche RAVD., Minx P. 2014. Direct Submission. *Aquila chrysaetos canadensis* isolate GSEH35GE, whole genome shotgun sequencing project submitted (09-OCT-2014) to NCBI accession JRUM00000000.1. The Genome Institute, Washington University School of Medicine, St. Louis, Missouri, U.S.A.

Warren WC., Clayton DF., Ellegren H., Arnold AP., Hillier LW., Künstner A., Searle S., White S., Vilella AJ., Fairley S., Heger A., Kong L., Ponting CP., Jarvis ED., Mello CV., Minx P., Lovell P., Velho TAF., Ferris M., Balakrishnan CN., Sinha S., Blatti C., London SE., Li Y., Lin Y-C., George J., Sweedler J., Southey B., Gunaratne P., Watson M., Nam K., Backström N., Smeds L., Nabholz B., Itoh Y., Whitney O., Pfenning AR., Howard J., Völker M., Skinner BM., Griffin DK., Ye L., McLaren WM., Flicek P., Quesada V., Velasco G., Lopez-Otin C., Puente XS., Olender T., Lancet D., Smit AFA., Hubley R., Konkel MK., Walker JA., Batzer MA., Gu W., Pollock DD., Chen L., Cheng Z., Eichler EE., Stapley J., Slate J., Ekblom R., Birkhead T., Burke T., Burt D., Scharff C., Adam I., Richard H., Sultan M., Soldatov A., Lehrach H., Edwards SV., Yang S-P., Li X., Graves T., Fulton L., Nelson J., Chinwalla A., Hou S., Mardis ER., Wilson RK. 2010. The genome of a songbird. *Nature* 464:757–762. DOI: 10.1038/nature08819.

Warren WC., Hillier LW., Tomlinson C., Minx P., Kremitzki M., Graves T., Markovic C., Bouk N., Pruitt KD., Thibaud-Nissen F., Schneider V., Mansour TA., Brown CT., Zimin A., Hawken R., Abrahamsen M., Pyrkosz AB., Morisson M., Fillon V., Vignal A., Chow W., Howe K., Fulton JE., Miller MM., Lovell P., Mello CV., Wirthlin M., Mason AS., Kuo R., Burt DW., Dodgson JB., Cheng HH. 2017. A New Chicken Genome Assembly Provides Insight into Avian Genome Structure. *G3: Genes, Genomes, Genetics* 7:109–117. DOI: 10.1534/g3.116.035923.

Warren W., Jarvis ED., Wilson RK., Howard JT., Gilbert MP., Zhang G., The Avian Genome Consortium 2014. Genomic data of the Bald Eagle (*Haliaeetus leucocephalus*). *GigaScience Database*. DOI: 10.5524/101040.

Wheeler TJ., Clements J., Eddy SR., Hubley R., Jones TA., Jurka J., Smit AFA., Finn RD. 2013. Dfam: a database of repetitive DNA based on profile hidden Markov models. *Nucleic Acids Research* 41:D70–D82. DOI: 10.1093/nar/gks1265.

Wicker T., Robertson JS., Schulze SR., Feltus FA., Magrini V., Morrison JA., Mardis ER., Wilson RK., Peterson DG., Paterson AH., Ivarie R. 2005. The repetitive landscape of the chicken genome. *Genome Research* 15:126–136. DOI: 10.1101/gr.2438004.

Wiens JD., Anthony RG., Forsman ED. 2014. Competitive interactions and resource partitioning between northern spotted owls and barred owls in western Oregon. *Wildlife Monographs* 185:1–50. DOI: 10.1002/wmon.1009.

Wu Y., Hadly EA., Teng W., Hao Y., Liang W., Liu Y., Wang H. 2016. Retinal transcriptome sequencing sheds light on the adaptation to nocturnal and diurnal lifestyles in raptors. *Scientific Reports* 6:33578. DOI: 10.1038/srep33578.

Yamada K., Nishida-Umehara C., Matsuda Y. 2004. A new family of satellite DNA sequences as a major component of centromeric heterochromatin in owls (Strigiformes). *Chromosoma* 112:277–287. DOI: 10.1007/s00412-003-0267-z.

Yang Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution* 24:1586–1591. DOI: 10.1093/molbev/msm088.

Yates A., Akanni W., Amode MR., Barrell D., Billis K., Carvalho-Silva D., Cummins C., Clapham P., Fitzgerald S., Gil L., Girón CG., Gordon L., Hourlier T., Hunt SE., Janacek SH., Johnson N., Juettemann T., Keenan S., Lavidas I., Martin FJ., Maurel T., McLaren W., Murphy DN., Nag R., Nuhn M., Parker A., Patricio M., Pignatelli M., Rahtz M., Riat HS., Sheppard D., Taylor K., Thormann A., Vullo A., Wilder SP., Zadissa A., Birney E., Harrow J., Muffato M., Perry E., Ruffier M., Spudich G., Trevanion SJ., Cunningham F., Aken BL., Zerbino DR., Flicek P. 2016. Ensembl 2016. *Nucleic Acids Research* 44:D710–D716. DOI: 10.1093/nar/gkv1157.

Yew DT., Woo HH., Meyer DB. 1977. Further studies on the morphology of the owl's retina. *Cells Tissues Organs* 99:166–168.

Zhang G., Li B., Gilbert M., Jarvis E., The Avian Genome Consortium, Wang J. 2014a. Genomic data of the Barn owl (*Tyto alba*). *GigaScience Database*. DOI: 10.5524/101039.

Zhang G., Li B., Gilbert M., Jarvis E., The Avian Genome Consortium, Wang J. 2014b. Genomic data of the Downy Woodpecker (*Picoides pubescens*). *GigaScience Database*. DOI: 10.5524/101012.

Zhang G., Li B., Gilbert M., Jarvis E., The Avian Genome Consortium, Wang J. 2014c. Genomic data of the Chimney Swift (*Chaetura pelagica*). *GigaScience Database*. DOI: 10.5524/101005.

Zhang G., Li B., Li C., Gilbert M., Jarvis E., Wang J., The Avian Genome Consortium 2014d. Comparative genomic data of the Avian Phylogenomics Project. *GigaScience* 3:1–8. DOI: 10.1186/2047-217X-3-26.

Zhang G., Li C., Li Q., Li B., Larkin DM., Lee C., Storz JF., Antunes A., Greenwold MJ., Meredith RW., Ödeen A., Cui J., Zhou Q., Xu L., Pan H., Wang Z., Jin L., Zhang P., Hu H., Yang W., Hu J., Xiao J., Yang Z., Liu Y., Xie Q., Yu H., Lian J., Wen P., Zhang F., Li H., Zeng Y., Xiong Z., Liu S., Zhou L., Huang Z., An N., Wang J., Zheng Q., Xiong Y., Wang G., Wang B., Wang J., Fan Y., Fonseca RR da., Alfaro-Núñez A., Schubert M., Orlando L., Mourier T., Howard JT., Ganapathy G., Pfenning A., Whitney O., Rivas MV., Hara E., Smith J., Farré M., Narayan J., Slavov G., Romanov MN., Borges R., Machado JP., Khan I., Springer MS., Gatesy J., Hoffmann FG., Opazo JC., Håstad O., Sawyer RH., Kim H., Kim K-W., Kim HJ., Cho S., Li N., Huang Y., Bruford MW., Zhan X., Dixon A., Bertelsen MF., Derryberry E., Warren W., Wilson RK., Li S., Ray DA., Green RE., O'Brien SJ., Griffin D., Johnson WE., Haussler D., Ryder OA., Willerslev E.,

Graves GR., Alström P., Fjeldså J., Mindell DP., Edwards SV., Braun EL., Rahbek C., Burt DW., Houde P., Zhang Y., Yang H., Wang J., Jarvis ED., Gilbert MTP., Wang J., Ye C., Liang S., Yan Z., Zepeda ML., Campos PF., Velazquez AMV., Samaniego JA., Avila-Arcos M., Martin MD., Barnett R., Ribeiro AM., Mello CV., Lovell PV., Almeida D., Maldonado E., Pereira J., Sunagar K., Philip S., Dominguez-Bello MG., Bunce M., Lambert D., Brumfield RT., Sheldon FH., Holmes EC., Gardner PP., Steeves TE., Stadler PF., Burge SW., Lyons E., Smith J., McCarthy F., Pitel F., Rhoads D., Froman DP. 2014e. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* 346:1311–1320. DOI: 10.1126/science.1251385.

Zhang Z., Schwartz S., Wagner L., Miller W. 2000. A Greedy Algorithm for Aligning DNA Sequences. *Journal of Computational Biology* 7:203–214. DOI: 10.1089/10665270050081478.

Zhang Y., Song T., Pan T., Sun X., Sun Z., Qian L., Zhang B. 2016. Complete sequence and gene organization of the mitochondrial genome of *Asio flammeus* (Strigiformes, strigidae). *Mitochondrial DNA Part A* 27:2665–2667. DOI: 10.3109/19401736.2015.1043538.

Chapter 1 Supplementary Article

1 Supplementary Material and Methods

1.1 Nextera350nt library

1.1.1 We intended this library to be a Nextera-sheared library with a small insert size. We isolated DNA using a Gentra Puregene Kit (Qiagen, Hilden, Germany) following the protocol entitled "Protocol: DNA Purification from Tissue Using the Gentra Puregene Tissue Kit" (Qiagen, Hilden, Germany). We used 50 ng of the DNA to prepare a genomic library using a Nextera DNA Sample Prep Kit (Illumina-compatible) (Epicentre, Madison, Wisconsin). After tagmentation, we cleaned the reaction with a DNA Clean & Concentrator -5 kit (Zymo Research, Irvine, California). We amplified the reaction for 5 cycles of PCR using a Nextera DNA Sample Prep Kit (Illumina-compatible) (Epicentre, Madison, Wisconsin) and the Nextera PCR Enzyme (Epicentre, Madison, Wisconsin). We then cleaned the reaction with a DNA Clean & Concentrator -5 kit (Zymo Research, Irvine, California). We used a LabChip XT DNA 750 Assay Kit on a LabChip XT (PerkinElmer, Waltham, Massachusetts) automated nucleic acid fractionation system to select library fragments in the size range of 375-600 nt, which, after subtracting the 141 nt of adapters, corresponds to an average fragment size of 346.5 nt. We performed a final PCR using 5 µL Klentaq LA 10X Buffer with MgCl (Sigma-Aldrich, St. Louis, Missouri), 1 µL 12.5 µM dNTPs, 1 µL each of two Illumina-adapter-compatible primers at 10 µM, 1 µL KlenTaq LA DNA Polymerase Mix (Sigma-Aldrich, St. Louis, Missouri), 5 µL library off of LabChip, and water to make a 50 µL reaction volume. We ran the PCR at 94°C for 2 min; then 5 cycles of denaturation at 94°C for 30 s, annealing at 58°C for 30 s, and extension at 72°C for 3 min; and we performed a final extension at 72°C for 5 min. We removed the PCR products after the final extension and and then cleaned them using a DNA Clean & Concentrator -5 kit (Zymo Research, Irvine, California). We obtained one lane of 100 nt paired-end data using a TruSeq PE Cluster Kit v2-cBot-HS kit and a TruSeq SBS v2-HS kit on a HiSeq 2000 (Illumina, San Diego, California) and a second lane of 100 nt paired-end data using a TruSeq PE Cluster Kit v3-cBot-HS kit and a TruSeq SBS v3-HS kit on a HiSeq 2000 (Illumina, San Diego, California).

1.2 Nextera700nt library

1.2.1 We attempted to construct a Nextera-sheared library with a moderate insert size. We isolated DNA using a Gentra Puregene Kit (Qiagen, Hilden, Germany) and used 50 ng to prepare a genomic library using a Nextera DNA Sample Preparation Kit (Illumina, San Diego, California). After tagmentation, we cleaned the reaction with a DNA Clean & Concentrator -5 kit (Zymo Research, Irvine, California). We amplified the reaction for 5 cycles of PCR using a KAPA Library Amplification kit (KAPA Biosystems, Wilmington, Massachusetts) and then cleaned the reaction with a DNA Clean & Concentrator -5 kit (Zymo Research, Irvine, California). We used a BluePippin (Sage Science, Beverly, Massachusetts) to select library fragments in the size range of 734-934 nt, which, after subtracting the 134 nt of adapters, corresponded to selecting an average insert size of 700 nt. We performed a real-time PCR (rtPCR) using a KAPA Real-Time Library Amplification Kit (KAPA Biosystems, Wilmington, Massachusetts) on a CFX96 Touch Real-Time PCR Detection System (Bio-Rad, Hercules, California) to amplify the library. We amplified the library with 6 cycles PCR and then cleaned the PCR products with a

DNA Clean & Concentrator -5 kit (Zymo Research, Irvine, California). We lastly assessed the library fragment size distribution with a 2100 BioAnalyzer (Agilent Technologies, Santa Clara, California) and the concentration of double-stranded DNA material with a Qubit 2.0 Flurometer (Invitrogen, Carlsbad, California). We obtained one lane of 150 nt paired-end data sequenced on a HiSeq 2500 (Illumina, San Diego, California) in rapid mode.

1.3 Nextera550nt library

1.3.1 We aimed to construct a Nextera-sheared library with overlapping reads, which could be merged into long fragments. We isolated DNA using a Gentra Puregene Kit (Qiagen, Hilden, Germany) and used 50 ng to prepare a genomic library using a Nextera DNA Sample Preparation Kit (Illumina, San Diego, California). After tagmentation, we cleaned the reaction with a DNA Clean & Concentrator -5 kit (Zymo Research, Irvine, California). We amplified the reaction for 5 cycles of PCR using a KAPA Library Amplification kit (KAPA Biosystems, Wilmington, Massachusetts) and then cleaned the reaction with a DNA Clean & Concentrator -5 kit (Zymo Research, Irvine, California). We then used a BluePippin (Sage Science, Beverly, Massachusetts) to select library fragments in the size range of 634-709 nt, which, after subtracting the 134 nt of adapters, corresponded to selecting an average insert size of 537.5 nt. We assessed the library fragment size distribution with a 2100 BioAnalyzer (Agilent Technologies, Santa Clara, California). We cleaned the size-selected product with 0.6X Agencourt AMPure XP (Beckman Coulter, Brea, California) magnetic beads to remove adapter dimer of approximately 250 nt in size. We then performed a real-time PCR (rtPCR) using a KAPA Real-Time Library Amplification Kit (KAPA Biosystems, Wilmington, Massachusetts) on a CFX96 Touch Real-Time PCR Detection System (Bio-Rad, Hercules, California) to amplify the library. We amplified the library with 8 cycles PCR and then cleaned the PCR products with a DNA Clean & Concentrator -5 kit (Zymo Research, Irvine, California). We lastly assessed the library fragment size distribution with a 2100 BioAnalyzer (Agilent Technologies, Santa Clara, California) and the concentration of double-stranded DNA material with a Qubit 2.0 Flurometer (Invitrogen, Carlsbad, California). We obtained one lane of 300 nt paired-end data sequenced using a MiSeq Reagent Kit v3 on a MiSeq (Illumina, San Diego, California). We obtained a second lane of 375 nt read 1 and 225 nt read 2 for a total of 600 nt of paired-end read data sequenced using a MiSeq Reagent Kit v3 on a MiSeq (Illumina, San Diego, California).

1.4 noPCR550nt library

1.4.1 We extracted genomic DNA from blood using a DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany). We sheared 4,460 ng genomic DNA in 130 µL in a microTUBE AFA Fiber Pre-Slit Snap-Cap tube (Covaris, Woburn, Massachusetts) using a M220 focused-ultrasonicator (Covaris, Woburn, Massachusetts) targeting 550 nt as the center of the fragment distribution. We used peak incident power 50 W, 20% duty factor, 200 cycles per burst, and 45 s treatment time at 20°C. We then removed small fragments and concentrated the sheared material using a DNA Clean & Concentrator -5 kit (Zymo Research, Irvine, California). We next constructed a genomic library by using a TruSeq DNA PCR-Free Library kit (Illumina, San Diego, California) and following the manufacturer's protocol, including the use of bead-based size selection to remove large and small DNA fragments in succession to target a mean fragment size of 550 nt. We

assessed the concentration of double-stranded DNA material in the final library with a Qubit 2.0 Flurometer (Invitrogen, Carlsbad, California).

1.5 900ntPCR library

1.5.1 We extracted genomic DNA from blood using a DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany). We sheared 4,580 ng genomic DNA in 130 µL in a microTUBE AFA Fiber Pre-Slit Snap-Cap tube (Covaris, Woburn, Massachusetts) using a M220 focused-ultrasonicator (Covaris, Woburn, Massachusetts) targeting 900 nt as the center of the fragment distribution. We used peak incident power 50 W, 5% duty factor, 200 cycles per burst, and 70 s treatment time at 20°C. We then removed small fragments and concentrated the sheared material using a DNA Clean & Concentrator -5 kit (Zymo Research, Irvine, California). We next constructed a genomic library by using a TruSeq DNA PCR-Free Library kit (Illumina, San Diego, California) and following the manufacturer's protocol, except that we only performed a bead-based size selection to remove small fragments and not large fragments. We used a 0.45X bead to sample ratio in order to eliminate fragments smaller than approximately 700 nt. Following A-tailing and prior to adapter ligation, we took 10% of the sample (by volume) and separated it from the noPCR aliquot for use in a PCR-amplified library. We ligated adapters to these two aliquots separately and cleaned the finished ligations with a DNA Clean & Concentrator -5 kit (Zymo Research, Irvine, California). We then only went forward with the aliquot for use in a PCR-amplified library. We used a BluePippin (Sage Science, Beverly, Massachusetts) to select library fragments in the size range of 800-1100 nt, which, after subtracting the 121 nt of adapters, corresponded to selecting an average insert size of 829 nt. We next cleaned the eluted material with a DNA Clean & Concentrator -5 kit (Zymo Research, Irvine, California) and then performed real-time PCR (rtPCR) using a KAPA Real-Time Library Amplification Kit (KAPA Biosystems, Wilmington, Massachusetts) on a CFX96 Touch Real-Time PCR Detection System (Bio-Rad, Hercules, California) to amplify the library. We amplified the library with 11 cycles PCR and then cleaned the PCR products with 1X Agencourt AMPure XP (Beckman Coulter, Brea, California) magnetic beads. We lastly assessed the library fragment size distribution with a 2100 BioAnalyzer (Agilent Technologies, Santa Clara, California) and the concentration of double-stranded DNA material with a Qubit 2.0 Flurometer (Invitrogen, Carlsbad, California).

1.6 Hydroshear library

1.6.1 We isolated DNA using a Gentra Puregene Kit (Qiagen, Hilden, Germany) and used a Hydroshear DNA Shearing Device (GeneMachines, Ann Arbor, Michigan) to shear 25 µg in DNA in 100 µL volume with 30 cycles of shearing using speed code 3. We checked the sheared DNA on a 1% agarose gel and saw that fragments had been sheared between 400-1000 nt. We additionally mechanically sheared the DNA by performing 15 passes through a 28 gauge x 1/2 inch needle attached to a 1 cc U-100 Insulin Syringe (Becton, Dickinson and Company). We performed end-repair using 4266 ng sheared DNA in an End-It DNA End-Repair Kit (Epicentre, Madison, Wisconsin). We incubated the reaction at room temperature for 45 minutes and then inactivated the enzyme by heating to 72°C for 10 minutes followed by cleaning with a DNA Clean & Concentrator -5 kit (Zymo Research, Irvine, California). We then added 3' A tails in a reaction with 2 µL 10X NEBuffer 2, 0.5 µL 100 mM dATP (Invitrogen, Carlsbad, California), 1 µL Klenow Fragment (3'→5' exo-) (New England Biolabs, Ipswich, Massachusetts), and 16.5 µL

cleaned end-repaired product. We incubated for 45 min at 37°C and then 20 min at 75°C to inactivate the enzyme. We cleaned the reaction with a DNA Clean & Concentrator -5 kit (Zymo Research, Irvine, California). We then ligated Illumina-compatible adapters using 1 µL 10X Fast-Link Ligation Buffer (Epicentre, Madison, Wisconsin), 1 µL 10 mM ATP (Epicentre, Madison, Wisconsin), 5 µL of end-repaired DNA (0.7835 µg), 2 µL of annealed Illumina-compatible adapters at 10 µM (Integrated DNA Technologies), and 1 µL Fast-Link DNA Ligase (Epicentre, Madison, Wisconsin) for 10 µL total reaction volume. We incubated the ligation reaction overnight at 16°C and then used 1.5X Agencourt AMPure XP (Beckman Coulter, Brea, California) magnetic beads to clean the ligase reaction and remove any extra adapters. We performed a PCR using 10 µL Klentaq LA 10X Buffer with MgCl (Sigma-Aldrich, St. Louis, Missouri), 2 µL 12.5 µM dNTPs, 2 µL each of two Illumina-adapter-compatible primers at 10 µM, 2 µL KlenTaq LA DNA Polymerase Mix (Sigma-Aldrich, St. Louis, Missouri), half of the cleaned ligase reaction in 10 µL, and water to make a 100 µL reaction volume. We ran the PCR in two 50 µL aliquots at 94°C for 5 min; then 2 cycles of denaturation at 94°C for 30 s, annealing at 55°C for 30 s, and extension at 68°C for 3 min; and we performed a final extension at 68°C for 5 min. We removed the PCR products after the final extension and and then cleaned them using a DNA Clean & Concentrator -5 kit (Zymo Research, Irvine, California). We used a LabChip XT DNA 750 Assay Kit on a LabChip XT (PerkinElmer, Waltham, Massachusetts) automated nucleic acid fractionation system to select library fragments in the size range of 600-700 nt. We performed a final PCR using 5 µL Klentaq LA 10X Buffer with MgCl (Sigma-Aldrich, St. Louis, Missouri), 1 µL 12.5 µM dNTPs, 1 µL each of two Illumina-adapter-compatible primers at 10 µM, 1 µL KlenTaq LA DNA Polymerase Mix (Sigma-Aldrich, St. Louis, Missouri), 5 µL library off of LabChip, and water to make a 50 µL reaction volume. We ran the PCR at 94°C for 2 min; then 17 cycles of denaturation at 94°C for 30 s, annealing at 58°C for 30 s, and extension at 72°C for 3 min; and we performed a final extension at 72°C for 5 min. We removed the PCR products after the final extension and and then cleaned them using a DNA Clean & Concentrator -5 kit (Zymo Research, Irvine, California). We next assessed the library fragment size distribution with a 2100 BioAnalyzer (Agilent Technologies, Santa Clara, California) and the concentration of double-stranded DNA material with a Qubit 2.0 Flurometer (Invitrogen, Carlsbad, California).

1.7 noPCR550nt, 900ntPCR, and Hydroshear libraries

1.7.1 We pooled the barcoded noPCR550nt, 900ntPCR, and Hydroshear libraries equimolarly and we obtained 350 nt read 1 and 250 nt read 2 for a total of 600 nt of paired-end read data from one lane (approximately ⅓ of one lane per library) using a 600-cycle MiSeq Reagent Kit v3 on a MiSeq (Illumina, San Diego, California).

1.8 MP4kb, MP7kb, and MP11kb libraries

1.8.1 We constructed and sequenced three large-insert mate-pair libraries. We isolated DNA using a Gentra Puregene Kit (Qiagen, Hilden, Germany) and sent 41.3 µg to GENEWIZ (www.genewiz.com). We requested barcoded mate-pair libraries with insert sizes of 4 kb, 6 kb, and 11 kb constructed using the Nextera Mate Pair Sample Preparation Kit (Illumina, San Diego, California). GENEWIZ followed the procedure detailed in the Nextera Mate Pair Sample Preparation Guide (Illumina, Part # 15035209 Rev. C, January 2013). Traces obtained using a 2100 BioAnalyzer (Agilent Technologies, Santa Clara, California) showed the centers of the distributions of the sheared fragmentsthat went into

the circularization step of the three mate-pair libraries as 4.2 kb, 7.1 kb, and 10.7 kb. GENEWIZ pooled the three libraries equimolarly and we obtained one lane (approximately ⅓ of one lane per library) of 100 nt paired-end data sequenced on a HiSeq 2000 (Illumina, San Diego, California).

1.9 Trimming - long-insert mate-pair data

1.9.1 We trimmed the Nextera mate-pair data using NxTrim version 0.2.3-alpha (O'Connell, 2014; O'Connell et al., 2015), which required BOOST version 1.57.0 (http://www.boost.org). When running NxTrim, we used the "--preserve-mp" flag to prefer mate pair reads as output even if paired-end reads would be longer. NxTrim utilizes the position of the junction identifier sequence in Nextera mate-pair data to classify reads of mate pair libraries as true mate pair reads, paired-end reads, or singleton reads.

1.9.2 We trimmed adapters and low quality bases separately for the resulting mate-pair data, paired-end reads, and singleton reads using Trimmomatic version 0.32 (Bolger, Lohse & Usadel, 2014). We trimmed adapters using options "ILLUMINACLIP:<fasta of Illumina adapter sequences >:2:30:10". We removed low quality bases from the beginning and end of the reads using the following options: LEADING:3 TRAILING:3 to remove bases below Phred 3. We trimmed off low quality sequence portions using: SLIDINGWINDOW:4:17, which trimmed the read when the average quality over 4 basepairs dropped below Phred 17. Finally, we trimmed reads less than 36 basepairs in length using "MINLEN:36".

1.10 Trimming - short-insert paired-end data

1.10.1 We first trimmed adapters from all non-mate-pair libraries using Trimmomatic version 0.32 (Bolger, Lohse & Usadel, 2014). We used the ILLUMINACLIP function with the following options: <fasta of Illumina adapter sequences >:2:30:10.

1.10.2 Since substantial portions of the paired-end reads from all of the libraries, except the Nextera700nt library were overlapping, we joined overlapping paired reads using the BBMerge tool in the BBMap tool suite version 34.00 (Bushnell, 2014). We merged overlapping reads using the options "minoverlapinsert=110 mininsert=110 strict=t" for the datasets Nextera350nt lane 1 and Nextera350nt lane 2, We used the options "minoverlapinsert=400 mininsert=400 strict=t" for the datasets Nextera550nt lane 1, Nextera550nt lane 2, noPCR550nt, and PCR900nt, which had longer read lengths.

1.10.3 We then performed quality trimming using Trimmomatic version 0.32 (Bolger, Lohse & Usadel, 2014). We removed low quality bases from the beginning and end of the reads using the options "LEADING:3 TRAILING:3" to remove bases below Phred 3. We trimmed off low quality sequence portions using "SLIDINGWINDOW:4:17", which trimmed the read when the average quality over 4 basepairs dropped below Phred 17. Finally, we trimmed reads less than 36 basepairs in length using "MINLEN:36".

1.11 Error-correction

1.11.1 Since we trimmed using a moderately low quality threshold, we used the k-mer-based error corrector in the SOAPdenovo2 toolkit, SOAPec version 2.01 (Luo et al., 2012), to correct sequence errors. We first used the KmerFreq_HA tool to create a k-mer frequency spectrum with default options except "-k 27 -L 600", which indicate that we used a k-mer size of 27 for creating the frequency spectrum and the maximum read length was 600 nt. We then used the Corrector_HA tool along with the k-mer frequency spectrum that we created to correct all of our trimmed reads using default options except "-k 27 -r 36",

which indicate that we used a k-mer size of 27 for the error correction and kept trimmed reads as short as 36 nt.

1.12 Single-end data

1.12.1 In each stage of the trimming, merging, and error-correction process, some reads previously paired became unpaired due to the loss of their paired read in a trimming step. We handled the single-end reads separate from the paired reads and subjected them to the same adapter, quality trimming, and error-correcting steps as the reads that remained paired. We used all of these single read sets in the final assembly.

1.13 Read processing variation for some preliminary assemblies

1.13.1 For a trim level of an average Phred 7 or 28, the only difference from the methodology described above was that we trimmed off low quality sequence portions using Trimmomatic with the parameter "SLIDINGWINDOW:4:7" or "SLIDINGWINDOW:4:28", respectively.

1.13.2 We did not apply the error-correction process to reads trimmed to an average Phred 28.

1.13.3 For some preliminary assemblies, we did not merge overlapping paired-end reads. This entailed leaving out the BBMerge step described above, but still performing adapter and quality trimming as noted.

1.14 Genome size

1.14.1 Genome size data estimated from flow cytometry measurement of red blood cells exist for two *S. occidentalis* congeners of, *S. aluco* and *S. nebulosa*. *Strix aluco* has a C-value of 1.59 pg (De Vita et al., 1994), which is approximately 1.56 Gnt (Doležel et al., 2003). *Strix nebulosa* has a C-value of 1.65 pg (Vinogradov, 2005), which is approximately 1.61 Gnt (Doležel et al., 2003).

1.14.2 We ran Preqc (Simpson, 2014), a module within SGA version 0.10.14 (Simpson & Durbin, 2010, 2016), which used Google SparseHash library version 2.0.2 (google-sparsehash@googlegroups.com, 2012), zlib version 1.2.8 (Gailly & Adler, 2013) and BamTools version 2.4.0 (Barnett et al., 2011, 2015) requiring CMake version 3.2.3 (Hoffman & Martin, 2003; Kitware, 2015), and on the 150 nt paired-end reads from the Nextera700nt dataset to estimate the genome size. Preqc estimated the genome size by sampling 20,000 reads and counting the frequency of *k*-mers of length 31 nt while applying a correction for sequencing errors.

1.15 Assembly

1.15.1 We used SOAPdenovo2 version 2.04 (Luo et al., 2012) to assemble the genome. We performed numerous trial runs experimenting with different k-mer values and parameters. We utilized the insert size estimated in the output of initial, trial assemblies to refine our estimation of the insert sizes for our libraries and used these refined values as input into subsequent assembly configuration files (Table S1). We settled on using the default parameters other than the options "SOAPdenovo-127mer all -N 1500000000 -K 23 -m 127 -k 65 -d 1 -R -F". These options indicate that we used the 127 k-mer version of the assembler and ran the assembly using multiple k-mer sizes starting at 23 and ending with a maximum of 127, we gave an estimated genome size of 1.5 Gnt, we allowed reads as small as 65 nt to map to contigs during scaffolding, we ignored singleton k-mers, we tried to resolve repeats with reads, and we attempted to fill gaps in scaffolds.

1.15.2 In our configuration files for all of the preliminary assemblies, we used the default minimum alignment lengths between a read and contig (32 for paired-end reads, 35 for

mate-pair reads) and the default minimum pair number cutoffs (3 for paired-end reads, 5 for mate-pair reads).

1.15.3 We used dupchk (Henderson & Hanna, 2016b), which utilized the first and last 21 nt of each read as a read fingerprint, to check for sequence duplication in each sequenced library.

1.16 Preliminary assembly assessment

1.16.1 In order to compare our preliminary assemblies, we removed contigs / scaffolds <= 300 nt in order to remove any unassembled reads from the assembly. We calculated the contig and scaffold N50 as well as the number of scaffolds in various length classes using scafN50 (Henderson & Hanna, 2016c). We calculated the total length of the assembly, the % Ns, and the total number of scaffolds using scafSeqContigInfo (Henderson & Hanna, 2016a). We were conservative and separated scaffolds into contigs at each N in the sequence, which is the default option for scafSeqContigInfo (Henderson & Hanna, 2016a).

1.16.2 We then used CEGMA version 2.5 (Parra, Bradnam & Korf, 2007), which required GeneWise from the Wise2 version 2.2.3-rc7 package (Birney, Clamp & Durbin, 2004; Birney, 2008), HMMER version 3.0 (http://hmmer.org), geneid version 1.4.4 (Guigó, 1998; Blanco et al., 2011), and NCBI's BLAST+ version 2.2.25 (Altschul et al., 1997; Camacho et al., 2009), to annotate a set of highly conserved eukaryotic genes in our assembly and thereby obtain an assessment of the quality and completeness of each assembly. In order to install CEGMA's GeneWise dependency, we followed the source code modification recommendations documented by Markus Grohme (<http://korflab.ucdavis.edu/datasets/cegma/ubuntu_instructions_1.txt>) and the Homebrew Science GeneWise formula (<https://github.com/Homebrew/homebrew-science/blob/master/genewise.rb>).

1.17 Determination of final assembly

1.17.1 We examined multiple statistics in choosing our final assembly. We valued high contig and scaffold N50 values, low % Ns in the sequence, a low total number of scaffolds, larger numbers of long scaffolds, and completeness as reflected in the number of conserved genes found by the CEGMA pipeline. We decided that the assembly that had the best statistics across these categories was assembly 4 (Table 2) and we went forward with this assembly as our final assembly.

1.18 Gap closing

1.18.1 We found that using the "-F" flag to fill gaps using the SOAPdenovo2 version 2.04 (Luo et al., 2012) *de novo* assembler was ineffective at gap filling during the assembly. We then filled gaps using the gap closing tool in the SOAPdenovo2 toolkit, GapCloser version 1.12-r6 (Luo et al., 2012), with the default options other than "-l 600" to specify that our longest read length was 600 nt. The program output a warning stating that the maximum supported read length was 155 nt and that it would use that setting for the analysis. We assumed that the program just used the first 155 nt of reads with a total length exceeding 155 nt.

1.18.2 The gap-closed assembly contained many contigs and/or scaffolds under 1000 nt in length, a substantial portion of which appeared to be unassembled reads. We used ScaffSplitN50s (Henderson & Hanna, 2016d) to compare the continuity statistics resulting after removing contigs / scaffolds of lengths 300, 500, and 1,000 nt as well as when using N blocks of lengths 1, 5, 10, 15, 20, and 25 to separate contigs within

scaffolds. Based on these results, we removed all contigs and scaffolds less than 1000 nt for downstream analyses.

1.19 Final assembly stats

1.19.1 We used CEGMA version 2.5 (Parra, Bradnam & Korf, 2007), which required GeneWise from the Wise2 version 2.2.3-rc7 package (Birney, Clamp & Durbin, 2004; Birney, 2008), HMMER version 3.0 (http://hmmer.org), geneid version 1.4.4 (Guigó, 1998; Blanco et al., 2011), and NCBI's BLAST+ version 2.3.0 (Altschul et al., 1997; Camacho et al., 2009), to annotate a set of highly conserved eukaryotic genes in our assembly and thereby obtain an assessment of the quality and completeness of the assembly. We ran CEGMA with default parameters other than specifying "--vrt" to optimize the searches for a vertebrate genome.

1.19.2 We used BUSCO version 1.1b1 (Simão et al., 2015a,b), which used NCBI's BLAST+ version 2.2.28 (Altschul et al., 1997; Camacho et al., 2009), HMMER version 3.1b2 (http://hmmer.org), and AUGUSTUS version 3.2.1 (Keller et al., 2011; Stanke, 2015) to assess the assembly quality by searching for conserved orthologs. We ran BUSCO with default genome mode parameters other than specifying "vertebrata" as the evolutionary lineage with the option "-l" and using "-sp chicken" to employ the AUGUSTUS parameters optimized for the chicken genome.

1.20 Contamination assessment

1.20.1 We performed a local alignment of all scaffolds in NSO-wgs-v0 to a copy the NCBI nucleotide database (nt) that we downloaded on 24 June 2016 (Clark et al., 2016; NCBI Resource Coordinators, 2016) using NCBI's BLAST+ version 2.3.0 tool BLASTN (Altschul et al., 1997; Camacho et al., 2009) with default parameters other than "-outfmt 10 -num_alignments 5 -max_hsps 1". We used these parameters to limit to 5 the maximum number of alignments to unique subjects output and to limit to 1 the number of outputted alignments per subject. This allowed us to examine the top 5 alignments to different subject sequences and ascertain whether those subject sequences were obtained from vertebrate or non-vertebrate organisms.

1.20.2 In order to parse the taxonomy of the subject sequences in the alignment output, we obtained the a local copy of the NCBI taxonomy database using NCBI's BLAST+ version 2.3.0 script, update_blastdb.pl with the parameters "--passive --timeout 300 --force --verbose taxdb". We also downloaded the files taxdump.tar.gz and gi_taxid_nucl.dmp.gz from NCBI (ftp://ftp.ncbi.nlm.nih.gov/pub/taxonomy) (Clark et al., 2016; NCBI Resource Coordinators, 2016). We then used GItaxidIsVert (Henderson & Hanna, 2016e) with default options other than using the parameter "-n" to filter the alignment output for non-vertebrate alignments.

1.20.3 We used the web version of NCBI's BLAST+ version 2.4.0 tool BLASTN (Altschul et al., 1997; Camacho et al., 2009) with default parameters.

1.21 Mitochondrial genome identification

1.21.1 We searched NSO-wgs-v1 (not repeat-masked, all contigs / scaffolds < 1,000 nt removed, contaminant scaffolds removed) for any of the contigs / scaffolds that were assemblies of the mitochondrial genome, rather than the nuclear genome using NCBI's BLAST+ version 2.4.0 tool BLASTN (Altschul et al., 1997; Camacho et al., 2009) with default parameters other than "-outfmt 6".

1.21.2 We annotated the scaffold using the MITOS WebServer version 806 (Bernt et al., 2013) and specifying "genetic code = 02 - Vertebrate" with default settings otherwise.

1.22 Sex identification

1.22.1 We searched NSO-wgs-v1 for matches to *S. varia CHD1W* and *CHD1Z* nucelotide sequences using NCBI's BLAST+ version 2.4.0 tool BLASTN (Altschul et al., 1997; Camacho et al., 2009) with default parameters other than "-outfmt 6".

1.22.2 We used the Geneious version 9.1.4 aligner through the "map to reference" function (Kearse et al., 2012; Biomatters, 2016b) with default options to align primers 2550F and 2718R (Fridolfsson & Ellegren, 1999)) to the scaffolds and then extract the region bounded by the aligned primers.

1.23 Repeat annotation

1.23.1 We performed a homology-based repeat annotation of the genome assembly using RepeatMasker version 4.0.5 (Smit, Hubley & Green, 2013), which employs the repeat databases of the DFAM library version 1.3 (Wheeler et al., 2013) and the Repbase-derived RepeatMasker libraries version 20140131 (Jurka, 1998, 2000; Jurka et al., 2005; Bao, Kojima & Kohany, 2015). Our installation of the RepeatMasker tool utilized NCBI's BLAST+ version 2.2.30 (Altschul et al., 1997; Camacho et al., 2009) and RMBlast version 2.2.28 (Smit, Hubley & National Center for Biotechnology Information, 2015) sequence search engines as well as the tandem repeats finder (TRF) version 4.0.7b (Benson, 1999, p. 199). We ran RepeatMasker with default options other than parameters "-gccalc -nolow -species aves". The purpose of this run was to produce a masked genome without masking of low complexity regions or simple repeats, which we could then use for downstream annotation steps.

1.23.2 We performed a *de novo* modeling of the repeat elements in the genome using RepeatModeler version 1.0.8 (Smit & Hubley, 2015), which uses two *de novo* repeat finders, RECON version 1.08 (Bao & Eddy, 2002) and RepeatScout version 1.0.5 (Price, Jones & Pevzner, 2005), as well as the tandem repeats finder (TRF) version 4.0.7b (Benson, 1999), the RMBlast version 2.2.28 (Smit, Hubley & National Center for Biotechnology Information, 2015) sequence search engine, and RepeatMasker version 4.0.5 {Smit et al., 2015} with Repbase-derived RepeatMasker libraries version 20140131 (Jurka, 1998, 2000; Jurka et al., 2005; Bao, Kojima & Kohany, 2015). We built a sequence database from our genome and ran RepeatModeler with default options.

1.23.3 We further masked the genome by running RepeatMasker again with the masked genome as input, using the repeat database created by our RepeatModeler run, and with default options other than parameters "-gccalc -nolow".

1.23.4 We performed homology-based repeat masking using RepeatMasker as above with default options other than parameters "-gccalc -species aves". We then performed a second run of RepeatMasker using the repeat database created by our RepeatModeler run with the masked genome as input and using default options other than parameters "-gccalc -nolow". Our output was a second twice-masked genome with masked low complexity regions and simple repeats.

1.24 Gene annotation

1.24.1 We used the MAKER accessory script, cegma2zff, to convert the GFF file output from our CEGMA run on the GapClosed assembly into ZFF format to use in training of the gene prediction tool Semi-HMM-based Nucleic Acid Parser (SNAP) version 2006-07-28 (Korf, 2004). We used the fathom tool of the SNAP package with the parameters "-categorize 1000", followed by fathom with the parameters "-export 1000", then the forge element of the SNAP package, then the hmm-assembler.pl script from the SNAP package

to convert the ZFF files to an HMM file, which was then the newly trained gene finder that we provided SNAP in the MAKER configuration file (Campbell et al., 2014).

1.24.2 We ran MAKER using NCBI's BLAST version 2.2.31+ (Altschul et al., 1997; Camacho et al., 2009); the sequence comparison tool, exonerate version 2.2.0 (Slater & Birney, 2005) with glib version 2.46.2; and the gene prediction tool, AUGUSTUS version 3.2.1 (Keller et al., 2011) for which we specified "chicken" for the gene prediction species model. We employed default parameters for all BLAST and exonerate statistics thresholds and default parameters for all other MAKER configuration options. We used Open MPI version 1.10.2 (Gabriel et al., 2004) to run MAKER on 62 cores for 50.62 hours.

1.24.3 We combined the annotations for all of the genes using the MAKER accessory scripts "fasta_merge" and "gff3_merge" with default options.

1.24.4 We assigned putative gene functions to the MAKER annotations by first obtaining the Uniprot manually annotated and non-redundant protein sequence database Swiss-Prot UniProt release 2016_04 (Consortium, 2015) on 2016 April 25 and indexing it using NCBI's BLAST version 2.2.31+ (Altschul et al., 1997; Camacho et al., 2009) tool "makeblastdb" with default parameters other than the options "-input_type fasta -dbtype prot". We then compared the combined MAKER protein fasta file to the Swiss-Prot UniProt database using the BLAST 2.2.31+ tool "blastp" with default parameters other than the options "-evalue .000001 -outfmt 6 -num_alignments 1 -seg yes -soft_masking true -lcase_masking -max_hsps 1". We then used the MAKER accessory script "maker_functional_gff" to add the protein homology data to the combined MAKER GFF3 file and the MAKER accessory script "maker_functional_fasta" to add the protein homology data to the combined MAKER protein and transcript fasta files.

1.24.5 In order to identify proteins with known functional domains, we ran InterProScan version 5.18-57.0 (Jones et al., 2014) with options "-appl PfamA -iprlookup -goterms -f tsv", which limited searches to Pfam, a database of protein family domains, on the protein sequences generated by MAKER. We then used the MAKER accessory script "ipr_update_gff" to update the MAKER-generated GFF3 file with the results of the InterProScan run and add information on protein family domain matches.

1.24.6 We then filtered transcripts with an Annotation Edit Distance (AED) less than 1 and/or a match to a Pfam domain using the option "-s" in the script "quality_filter.pl" supplied in MAKER version 3.00.0 (Cantarel et al., 2008).

1.24.7 We used the "stat" tool of GenomeTools version 1.5.1 (Gremme, Steinbiss & Kurtz, 2013) to calculate annotation summary statistics, including distributions of gene lengths, exon lengths, number of exons per gene, and coding DNA sequence (CDS) lengths (measured in amino acids). We also used the "stat" tool of GenomeTools with the options "-addintrons" and "-intronlengthdistri" to infer intron lengths within the annotated gene boundaries and calculate the distribution of intron lengths.

1.25 Alignment

1.25.1 We aligned each set of reads to NSO-wgs-v1-masked using bwa version 0.7.12-r1044 (Li, 2013a) with default options other than parameters "bwa mem -M". We separately aligned paired-end and unpaired reads. For alignment of the paired-end data, we set the insert size to be equal to our estimates from our initial assemblies. We set the parameter "-w" to be equal to twice the standard deviation of the insert size we estimated from our initial assemblies.

1.25.2 We merged the paired-end and unpaired read alignments using the Picard version 1.104 function MergeSamFiles (http://broadinstitute.github.io/picard) and sorted them using the Picard version 1.104 function SortSam (http://broadinstitute.github.io/picard), employing default settings for both tools. We next marked duplicate reads (both PCR and optical) using the Picard version 1.104 function MarkDuplicates (http://broadinstitute.github.io/picard), employing default settings.

1.25.3 We assessed the genome coverage, duplication level, and other statistics of each read set based on the read alignments. We used the Picard version 1.141 function CollectWgsMetrics (http://broadinstitute.github.io/picard) with the bam file output by MarkDuplicates as the input file, employing default settings, except setting COUNT_UNPAIRED=True to include coverage contributed by unpaired reads when calculating the alignment statistics. The default CollectWgsMetrics settings included setting the minimum mapping quality for a read to contribute coverage as 20 and the minimum base quality for a base to contribute coverage as 20. We also ran CollectWgsMetrics with the default settings and COUNT_UNPAIRED=False to obtain the portion of the total aligned reads contributed by unpaired reads.

1.25.4 In order to obtain an estimate of the insert size of the mate pair libraries independent of the N-gaps in the scaffold sequences, we divided the scaffolds into contigs at 25 or more N's using make-contig-ref.sh from NSO-genome-scripts version 1.0.0 (Hanna & Henderson, 2017) with bioawk version 1.0 (Li, 2013b), GNU Awk (GAWK) version 4.0.1 (Free Software Foundation, 2012), and GNU fold version 8.21 (MacKenzie, 2013). We then aligned the mate pair libraries to this set of contigs using bwa version 0.7.10-r789 (Li, 2013a) with default options other than parameters "bwa mem -M". For alignment of the paired-end data, we set the insert size to be equal to our estimates from our initial assemblies. We set the parameter "-w" to be equal to twice the standard deviation of the insert size we estimated from our initial assemblies. We calculated the insert sizes for each of the three mate pair libraries from these alignments using calcInsertLen.sh from NSO-genome-scripts version 1.0.0 (Hanna & Henderson, 2017) with bioawk version 1.0 (Li, 2013b).

1.26 Microsatellite analysis

1.26.1 We searched the assembly for 16 pairs of microsatellite primer sequences using NCBI's BLAST+ version 2.4.0 tool BLASTN (Altschul et al., 1997; Camacho et al., 2009) with default parameters other than "-outfmt 6 –word_size 7".

1.27 Barred owl divergence

1.27.1 We used 50 ng genomic DNA to prepare a whole-genome library using a Nextera DNA Sample Preparation Kit (Illumina, San Diego, California). After tagmentation, we cleaned the reaction with a DNA Clean & Concentrator -5 kit (Zymo Research, Irvine, California). We amplified the reaction with 5 cycles of PCR using a KAPA Library Amplification kit (KAPA Biosystems, Wilmington, Massachusetts) and then cleaned the reaction with a DNA Clean & Concentrator -5 kit (Zymo Research, Irvine, California). We used a BluePippin (Sage Science, Beverly, Massachusetts) to select library fragments in the size range of 500-700 nt, which, after subtracting the 134 nt of adapters, corresponded to selecting an average insert size of 466 nt. We cleaned the BluePippin products with 0.6X Agencourt AMPure XP (Beckman Coulter, Brea, California) magnetic beads and then performed a real-time PCR (rtPCR) using a KAPA Real-Time Library Amplification Kit (KAPA Biosystems, Wilmington, Massachusetts) on a CFX96

Touch Real-Time PCR Detection System (Bio-Rad, Hercules, California) to amplify the library with 8 cycles PCR. We then cleaned the PCR products with a DNA Clean & Concentrator -5 kit (Zymo Research, Irvine, California). We lastly assessed the library fragment size distribution with a 2100 BioAnalyzer (Agilent Technologies, Santa Clara, California) and the concentration of double-stranded DNA material with a Qubit 2.0 Flurometer (Invitrogen, Carlsbad, California). We combined this library with others and sequenced it on two successive runs of 150 nt paired-end sequencing using a 2-lane flow cell on a HiSeq 2500 (Illumina, San Diego, California) in rapid mode. On the first run, we obtained sequencing data from a portion of each of the two flow cell lanes. On the second run, we obtained data from a portion of one of the two flow cell lanes. We combined all of the data from the two runs for the downstream steps.

1.27.2 We performed adapter and quality trimming of the sequence data using Trimmomatic version 0.32 (Bolger, Lohse & Usadel, 2014). We used the following options: "ILLUMINACLIP:<fasta of Illumina adapter sequences>:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:28 MINLEN:36".

1.27.3 We aligned trimmed paired and unpaired reads to NSO-wgs-v1-masked using bwa mem version 0.7.12-r1044 (Li, 2013a) with default options other than parameters "bwa mem -M". We separately aligned paired-end and unpaired reads. For alignment of the paired-end reads, we set the insert size to be equal to the size estimate of the final library given by the 2100 BioAnalyzer (Agilent Technologies, Santa Clara, California) minus the length of the adapters, which gave an insert size of 466 nt. Additionally, for the alignment of the paired-end reads we set the parameter "-w", the maximum insert size, equal to 1000.

1.27.4 We merged the paired-end and unpaired sequence alignments using the Picard version 1.104 function MergeSamFiles (http://broadinstitute.github.io/picard) and sorted them using the Picard version 1.104 function SortSam (http://broadinstitute.github.io/picard), employing default settings for both tools. We next marked duplicate sequences (both PCR and optical) using the Picard version 1.104 function MarkDuplicates (http://broadinstitute.github.io/picard), employing default settings.

1.27.5 We calculated various alignment statistics using the Picard version 1.141 function CollectWgsMetrics (http://broadinstitute.github.io/picard) with the bam file output by MarkDuplicates as input and employing default settings except setting COUNT_UNPAIRED=True in order to include coverage contributed by unpaired reads in the calculation of the statistics on the aligned reads. The default CollectWgsMetrics settings include setting the minimum mapping quality for a read to contribute coverage as 20 and the minimum base quality for a base to contribute coverage as 20. We also ran CollectWgsMetrics with the default settings and COUNT_UNPAIRED=False to obtain the portion of the total aligned reads contributed by unpaired reads.

1.27.6 We used Genome Analysis Toolkit (GATK) version 3.4-46 UnifiedGenotyper (McKenna et al., 2010; DePristo et al., 2011; Van der Auwera et al., 2013) to call SNPs using the *S. occidentalis* (Sequoia) and *S. varia* (CMCB41533) bwa-aligned, sorted, duplicate-marked bam files as simultaneous inputs and employing default options other than setting "--output_mode EMIT_ALL_SITES".

1.27.7 We first filtered the variant file using the following GNU Awk (GAWK) version 4.0.1 (Free Software Foundation, 2012) command: “awk 'NF==11 && substr($1, 1, 2) != "##" && $6>=50 && $1 != "#CHROM" && $1 != "C7961234" && $1 != "C7963448" &&

$1 != "C7970814" && $1 != "C8091874" && $1 != "scaffold3674"' | awk '$4=="A" || $4=="C" || $4=="G" || $4=="T"' | awk '$5=="A" || $5=="C" || $5=="G" || $5=="T"' > filtered1.vcf". This removed lines without 11 fields, header lines, variant sites where the Phred-scaled probability that a polymorphism exists was < 50, contaminant scaffolds, the mitochondrial genome scaffold, indels, and non-polymorphic sites.

1.27.8 We then calculated the unfiltered allele depth (the number of reads that supported an allele) summed across all of the alleles at each of the remaining variant sites using the following GNU cut version 8.21 (Ihnat, MacKenzie & Meyering, 2013) and GNU Awk (GAWK) version 4.0.1 (Free Software Foundation, 2012) command: "cat filtered1.vcf | cut -f10,11 | awk 'BEGIN {cov} {split($1,a,":"); split(a[2],acov,","); split($2,b,":"); split(b[2],bcov,","); totcov = acov[1]+acov[2]+bcov[1]+bcov[2]; print totcov}' > vcf-coverage.out". We then graphed these depths and calculated the mean and standard deviation (σ) of the distribution using vcf-coverage-calc.py from NSO-genome-scripts version 1.0.0 (Hanna & Henderson, 2017) with Python version 2.7.12 (Python Software Foundation, 2016), matplotlib version 1.5.1 (Hunter, 2007; Matplotlib Development Team, 2016), and NumPy version 1.11.1 (NumPy Developers, 2016).

1.27.9 When calculating the nucleotide diversity both within and between samples (H_w and H_b), we removed variants where the unfiltered allele depth summed across all of the alleles was greater than 5σ greater than the mean depth, variants without information for both samples, and variants where the *S. o. caurina* genotype was homozygous for the non-reference allele. We used calc-pi-exclude-onlySPOW.sh and calc-pi-exclude-onlyBADO.sh from NSO-genome-scripts version 1.0.0 (Hanna & Henderson, 2017) with GNU cut version 8.21 (Ihnat, MacKenzie & Meyering, 2013) and GNU Awk (GAWK) version 4.0.1 (Free Software Foundation, 2012) to calculate the H_w for *S. o. caurina* and *S. varia*, respectively. We used calc-pi-exclude.sh from NSO-genome-scripts version 1.0.0 (Hanna & Henderson, 2017) with GNU Awk (GAWK) version 4.0.1 (Free Software Foundation, 2012) to calculate H_b for *S. o. caurina* and *S. varia*. In order to report H_w and H_b in terms of the number of nucleotide differences per site within the sample, we divided the output from the scripts above by the number of ACGT characters in NSO-wgs-v1-nuc (the whole-genome assembly without the contaminant or mitochondrial scaffolds), which we obtained using "assemblathon-stats-ex.pl" from NSO-genome-scripts (Bradnam et al., 2013; Hanna & Henderson, 2017).

1.27.10 We averaged the values of H_w for *S. o. caurina* and *S. varia* and then used this average along with H_b in equation 3 from a study by Hudson, Slatkin & Maddison (1992) in order to estimate F_{ST} between *S. o. caurina* and *S. varia*.

1.28 PSMC analysis

1.28.1 In order to prepare our data for input into an analysis using an implementation of the pairwise sequentially Markovian coalescent model, PSMC version 0.6.5-r67 (Li & Durbin, 2011; Li, 2015), we used Samtools version 1.3.1 with HTSlib 1.3.1 (Li et al., 2009, 2016b), bcftools version 1.3.1 (Li et al., 2016a), and the vcfutils.pl script from bcftools to call variants with the command "samtools mpileup -C50 –uf reference-genome.fa alignment-file.bam | bcftools call -c - | vcfutils.pl vcf2fq -d minimum-read-depth -D maximum-read-depth | gzip >variants.fq.gz". As per the recommendation of the PSMC documentation (https://github.com/lh3/psmc), we used a third of the average read depth as the minimum read depth (-d) and twice the average read depth as the maximum read depth (-D) (-d 20 -D 126 and -d5 -D 33 for *S. o. caurina* and *S. varia*, respectively).

We determined the average read depth using Samtools version 1.3.1 with HTSlib 1.3.1 (Li et al., 2009, 2016b) and GNU Awk (GAWK) version 4.0.1 (Free Software Foundation, 2012) with the command "samtools depth alignment-file.bam | awk '{sum += $3} END {print sum / NR}'".

1.28.2 After variant calling, we used the PSMC script "fq2psmcfa" next with the command "fq2psmcfa -q20 variants.fq.gz >variants.psmcfa". We then ran PSMC with the command "psmc -N25 -t15 -r5 -p "4+25*2+4+6" -o variants.psmc variants.psmcfa". We next ran the PSMC scripts "psmc2history.pl" and "history2ms.pl" with the command "psmc2history.pl variants.psmc | history2ms.pl > variants.psmc_ms-cmd.sh".

1.28.3 We ran 100 rounds of bootstraping by first splitting long reference sequences into shorter lengths in the variants.psmcfa file using the PSMC script "splitfa" with the command "splitfa variants.psmcfa >variants-split.psmcfa" and then running PSMC with the command "parallel -j25 'psmc -N25 -t15 -r5 -b -p "4+25*2+4+6" -o variants-split-round-{}.psmc variants-split.psmcfa' :::: <(seq 100)".

1.28.4 We graphed the output of our PSMC run and rounds of bootstrapping by first combining using GNU cat version 8.21 (Granlund & Stallman, 2013) with the command "cat variants.psmc variants-split-round-*.psmc >variants-combined.psmc". We then plotted the output using the PSMC script "psmc_plot.pl" with the command "psmc_plot.pl -u 4.6e-09 -g 2 variants-combined-plot variants-combined.psmc". We used 2 years as the generation time (-g option for psmc_plot.pl) for both *S. o. caurina* and *S. varia* (Gutiérrez, Franklin & Lahaye, 1995; Mazur & James, 2000) although *S. o. caurina* may breed in its first year (Hamer et al., 1994) and some researchers have estimated the generation time *S. o. caurina* as 10 years (Noon & Biles, 1990; USDA Forest Service, 1992). We used 4.6×10^{-9} mutations per site per generation (Smeds, Qvarnström & Ellegren, 2016) as the mutation rate (-u option for psmc_plot.pl).

1.29 Light-associated gene analyses

1.29.1 We searched in NSO-wgs-v1 for regions orthologous to probes for 19 genes that encode proteins with light-associated functions using Geneious version 9.1.6 (Kearse et al., 2012; Biomatters, 2016a) and the included version of the NCBI BLAST+ BLASTn tool (Zhang et al., 2000) with default options. On 1-10 November, 2016, we used the web version of NCBI BLAST+ version 2.5.0 (Zhang et al., 2000) with discontiguous megablast options to align the probes against sequences in the NCBI Whole-Genome-Shotgun (WGS) contigs database limited by specifying the organism *T. alba* (taxid:56313).

1.29.2 When BLAST searches were unsuccessful, we used synteny data from Ensembl (version 86; (Yates et al., 2016) to search for evidence of whole gene deletion. We identified genes flanking the gene of interest in related taxa, and subsequently used BLAST to align the reference sequences for these genes against the *S. o. caurina* and *T. alba* genome assemblies. We imported the *S. o. caurina* genome assembly into Geneious version 9.1.6 (Kearse et al., 2012; Biomatters, 2016a) and used the included version of the NCBI BLAST+ BLASTn tool (Zhang et al., 2000) to search for the flanking genes in our assembly. We used the web version of NCBI BLAST+ version 2.5.0 (Zhang et al., 2000) to align the flanking genes against *T. alba* sequences in the NCBI Whole-Genome-Shotgun (WGS) contigs database.

1.29.3 We used the NCBI BLAST+ version 2.5.0 blastn tool (Zhang et al., 2000) with the discontiguous megablast option to align a reference *Opn4m* sequence to fifteen avian retinal transcriptomes in NCBI's Sequence Read Archive (SRA) (Leinonen, Sugawara &

Shumway, 2011; NCBI Resource Coordinators, 2016) including the pied harrier (*Circus melanoleucos*) (SRA accession SRR3203217), long-eared owl (*Asio otus*) (SRA accession SRR3203220), eastern grass owl (*Tyto longimembris*) (SRA accession SRR3203222), hoopoe (*Upupa epops*) (SRA accession SRR3203224), Eurasian eagle-owl (*Bubo bubo*) (SRA accession SRR3203225), black-winged kite (*Elanus caeruleus*) (SRA accession SRR3203227), Eurasian scops owl (*Otus scops*) (SRA accession SRR3203230), common kestrel (*Falco tinnunculus*) (SRA accession SRR3203231), grey-faced buzzard (*Butastur indicus*) (SRA accession SRR3203233), besra (*Accipiter virgatus*) (SRA accession SRR3203234), cinereous vulture (*Aegypius monachus*) (SRA accession SRR3203236), Eurasian hobby (*Falco subbuteo*) (SRA accession SRR3203238), grey-headed woodpecker (*Picus canus*) (SRA accession SRR3203240), little owl (*Athene noctua*) (SRA accession SRR3203242), Indian scops owl (*Otus bakkamoena*) (SRA accession SRR3203243) (Wu et al., 2016).

2 Supplementary Results and Discussion

2.1 Scaffold numbering

2.1.1 When referring to specific scaffolds in the results and discussion sections, we have inserted a dash ("-") between the word "scaffold" and the scaffold number for legibility. These dashes are not present in any of the assembly data files. Thus, "scaffold-1085" referenced in the manuscript will appear as "scaffold1085" in the assembly and other associated files.

3 Supplementary Tables

Table S1. Sequence data collected for use in genome assembly.

We here provide information on the insert size, fragmentation method, amplification, sequencing length, and raw data quantity for all libraries sequenced for this genome assembly. We have numbered the libraries and refer to these numbers in other sections of this manuscript.

Library number	Library name	Average insert size (nt)	Insert size standard deviation (nt)	Library Fragmentation method	PCR amplification used (Yes / No)	Paired-end read lengths forward / reverse (nt)	Raw reads passing onboard Illumina quality filter coverage of 1.5 Gnt genome (1X-fold coverage)
1	Nextera350nt lane 1	247	118	Nextera	Yes	100 / 100	9.80
2	Nextera350nt lane 2	247	118	Nextera	Yes	100 / 100	26.44
3	Hydroshear	500	52	Hydroshear	Yes	350 / 250	2.55
4	Nextera550nt lane 1	560	25	Nextera	Yes	300 / 300	3.65
5	Nextera550nt lane 2	560	25	Nextera	Yes	375 / 225	8.90
6	Nextera700nt	566	194	Nextera	Yes	150 / 150	31.14
7	noPCR550nt	619	132	Covaris	No	350 / 250	3.50
8	PCR900nt	687	58	Covaris	Yes	350 / 250	2.04
9	MP4kb	3,316	213	Nextera Mate Pair	Yes	100 / 100	7.84
10	MP7kb	5,904	537	Nextera Mate Pair	Yes	100 / 100	8.48
11	MP11kb	9,615	1930	Nextera Mate Pair	Yes	100 / 100	8.19

Table S2. Preliminary assembly parameters.

We here report the parameters used in our preliminary assemblies using SOAPdenovo2. "Trim level" indicates the average Phred score to which we trimmed using Trimmomatic. A higher Phred score indicates a more restrictive trimming. “Error correction” refers to whether we performed error correction on the input reads for the assembly. We provide information on how we specified that the assembler use the paired-end and unpaired data for each assembly. For a given assembly, we note which libraries provided data and in which portions of the assembly process that data was used. For a given portion of the assembly process, we give the numbers of the utilized libraries followed, in parentheses, by the rank given to each library in the assembly configuration file. Please refer to Table S1 for information about the libraries to which the numbers refer. An asterisk is next to the preliminary assembly that we chose to use as the basis for the final assembly.

Assembly	Trim level	Error correction	Assembly notes	Unpaired data - only contig	Paired-end data - only scaffold	Paired-end data - both contig and scaffold	Unpaired data - only gap closure
1	28	No	N/A	1-11 (6)	9 (3), 10 (4), 11 (5)	1-3 (1), 6 (2), 7 (1), 8 (2)	None
2	28	No	N/A	1-2 (6), 4-11 (6)	9 (3), 10 (4), 11 (5)	1-2 (1), 6 (2), 7 (1), 8 (2)	None
3	28		Only reads merged with BBMerge used as unpaired data	1-5 (6), 7-8 (6)	9 (3), 10 (4), 11 (5)	1-2 (1), 6 (2), 7 (1), 8 (2)	None
4*	17	Yes	N/A	1-2 (6), 4-11 (6)	9 (3), 10 (4), 11 (5)	1-2 (1), 6 (2), 7 (1), 8 (2)	None
5	28	No	No merging of paired-end reads performed		9 (3), 10 (4), 11 (5)	1-5 (1), 6 (2), 7 (1), 8 (2)	None
6	28	No	N/A	1-11 (6)	9 (3), 10 (4), 11 (5)	1-3 (1), 6 (2), 7 (1)	None
7	28	No	N/A	1-11 (6)	9 (3), 10 (4), 11 (5)	1 (1), 2 (1), 3 (1), 4 (1), 5 (1), 6 (2), 7 (1), 8 (2)	None
8	28	No	No merging of paired-end reads performed.		9 (3), 10 (4), 11 (5)	1 (1), 2 (1), 4 (1), 5 (1), 6 (2), 7 (1), 8 (2)	None
9	28	No	Only reads merged with BBMerge used as unpaired data	1-2 (6), 4-5 (6), 7-8 (6)	9 (3), 10 (4), 11 (5)	1 (1), 2 (1), 6 (2), 7 (1), 8 (2)	None
10	28	No	Only reads merged with BBMerge used as unpaired data	1-2 (6), 4-5 (6), 7-8 (6)	9 (3), 10 (4), 11 (5)	1-2 (1), 4-5 (1), 6 (2), 7 (1), 8 (2)	None
11	17	Yes	Only reads merged with BBMerge used as unpaired data, library 3 excluded.	1-2 (6), 4-5 (6), 7-8 (6)	9 (3), 10 (4), 11 (5)	1-2 (1), 4-5 (1), 6 (2), 7 (1), 8 (2)	None
12	17	Yes	All unpaired reads used, library 3 excluded.	1-2 (6) 4-5 (6), 6-11 (6)	9 (3), 10 (4), 11 (5)	1-2 (1), 4-5 (1), 6 (2), 7 (1), 8 (2)	None
13	17	Yes	Reads merged with BBMerged used for contig assembly, other unpaired reads used only for gap closure.	1-2 (6), 4-5 (6), 7-8 (6)	9 (3), 10 (4), 11 (5)	1-2 (1), 2 (1), 4-5 (1), 6 (2), 7 (1), 8 (2)	1-2 (7), 4-11 (7)
14	7	Yes	N/A	1-2 (6), 4-11 (6)	9 (3), 10 (4), 11 (5)	1-2 (1), 4-5 (1), 6 (2), 7 (1), 8 (2)	None

Table S3. Light-associated gene searches information.

This table provides details on the reference sequences used for and the results of our searches for light-associated genes in the genome assemblies of *Strix occidentalis caurina* and *Tyto alba*. "Stop" indicates the presence of a premature stop codon. "Del" indicates a frameshift deletion. "Ins" indicates a frameshift insertion.

Gene	**Reference Sequence**	***Strix occidentalis* Sequence**	***Tyto alba* Sequence**
SWS1	GenBank: AH007798 *Columba livia*	**No BLAST results**	**No BLAST results**
SWS1 notes	**Synteny:** *Taeniopygia guttata* and *Homo sapiens,* 5' end *FLNC* (REV), 3' end *CALU* (REV); *Anolis carolinensis*, 3' end *CALU* (REV)	*FLNC:* scaffold-4221 *CALU:* scaffold-15	No gene predictions for *FLNC* or *CALU* in *Tyto*
SWS2	GenBank: AH007799 *Columba livia*	scaffold-4153 & scaffold-7110: Functional	**No BLAST results**
SWS2 notes	**Synteny:** *Anolis carolinensis* and *Xenopus laevis* 5' end *MECP2* (REV), 3' end *LWS;* avian contigs in Ensembl are very short and do not include flanking genes	Only exons 1 (partial), 2 and 3 recovered; partial exon 1 flanked by N's, and exon 3 is towards the end of the scaffold; 2 different scaffolds; 100% identical except 1-nt diff in exon 3, nonsynonymous	*MECP2* and *LWS* not predicted in *Tyto*
Rh1	GenBank: AH007730 *Columba livia*	scaffold-133: Functional	JJRD01003728, JJRD01003729: Functional
Rh2	GenBank: AH007731 *Columba livia*	scaffold-1932: Functional	**JJRD01131248, JJRD01131249: Pseudogene (exon 1: 29-nt del; exon 2: stop; exon 3: stop; exon 4: 2-nt del)**
LWS	GenBank: AH007800 *Columba livia*	scaffold-6263: Functional	**No BLAST results**
LWS notes	**Synteny:** *Anolis carolinensis* 5' end *SWS2,* 3' end TEX28 (REV); *Xenopus laevis* 5' end *SWS2*, 3' end *AVPR2*; avian contigs in Ensembl are very short and do not include flanking genes	Only exons 2, partial 5 and 6; 3-5 are N's, no hits for exon 1	No gene predictions for *SWS2, AVPR2* or *TEX28*
OpnP	GenBank: U15762, WGS: AADN03007691 *Gallus gallus*	**No BLAST results; After BLASTing intergenic region, has hit with *Gallus gallus* genomic pinopsin, non-coding region 5' of cds is retained**	**JJRD01162372, JJRD01162373: Pseudogene (exon 1: start codon mutation ACA, 13-nt del, 2-nt ins, 1-nt del, exon 2: 1-nt del; intron 3-exon 4 boundary: 21 nt-del; exon 4: 7-nt del, 2-nt del; exon 5: 1-nt del)**
OpnP notes	**Synteny:** *Gallus gallus DOC2B,* 5' end, 3' end *TEX14* (REV); *Ficedula albicollis DOC2B*, 5' end	DOC2B: scaffold-86 TEX14: scaffold-86	
OpnVA	GenBank: EF055883, WGS: AADN03005037 *Gallus gallus*	Scaffold205: Functional	JJRD01088850, JJRD01088852, JJRD01106859, JJRD01168068: Functional
Opn4x	GenBank: NM_204625, WGS: AADN03004364 *Gallus gallus*	scaffold-147: Functional	JJRD01038044: Functional

Gene	**Reference Sequence**	***Strix occidentalis* Sequence**	***Tyto alba* Sequence**
Opn4m	GenBank: AY882944, WGS: AADN04000143 *Gallus gallus*	**scaffold-219: Pseudogene? (exon 8: stop, 4-nt del)**	**JJRD01098086, JJRD01098087: Pseudogene? (exon 8: 4-nt del; intron 11: splice donor mutation GT to AT)**
Opn3	GenBank: XM_426139, WGS: AADN04000318 *Gallus gallus*	scaffold-728: Functional	JJRD01072701: Functional (No BLAST results for exon 1)
Opn5	GenBank: NM_001130743 WGS: AADN04000287 *Gallus gallus*	scaffold-546: Functional	JJRD01001581, JJRD01133804: Functional
Opn5L1	GenBank: NM_001310056, WGS: AADN04000228 *Gallus gallus*	scaffold-6: Functional	JJRD01004196: Functional
Opn5L2	GenBank: NM_001162892, WGS: AADN04000287 *Gallus gallus*	scaffold-722: Functional	JJRD01082691: Functional
RRH	GenBank: NM_001079759, WGS: AADN04000018 *Gallus gallus*	scaffold-22: Functional	JJRD01123735: Functional
RGR	GenBank: NM_001031216, WGS: AADN04000143 *Gallus gallus*	scaffold-219: Functional	JJRD01065549: Functional
EEVS-like	GenBank: XM_013180282, WGS: AOGC01018216 *Anser cygnoides*	scaffold-133: Functional	JJRD01160345: Functional
MT-Ox	GenBank: XM_015293238, WGS: AADN04000009 *Gallus gallus*	scaffold-133: Functional	JJRD01160345, JJRD01160346, JJRD01160347, JJRD01160348: Functional
Photolyase	GenBank: XM_422729, WGS: AADN04000078 *Gallus gallus*	scaffold-742: Functional	JJRD01136093, JJRD01136094: Functional
CYP2J19	GenBank: XM_422553, WGS: AADN04000032 *Gallus gallus*	**scaffold-313: Pseudogene? (exon 9: 1-nt ins, 2-nt del)**	**JJRD01034859: Pseudogene (exon 1: stop; exon 3: 5-nt del; exon 5: stop; exon 6: stop)**

Table S4. Assembly metrics with a range of cutoffs.

These are statistics on the final (post gap closing) assembly that display the consequence of choosing various cutoffs for minimum scaffold length and the number of N's that separate a contig. We have marked the line with the cutoffs and statistics that correspond to the final chosen assembly version with an asterisk.

Scaffold minimum length (nt)	Scaffold N50 (nt)	Scaffold L50	Number of Scaffolds	Total sequence length (nt)	Number of N's to split contigs	Contig N50	Contig L50	Number of contigs	Total sequence length (nt)
1000*	3,983,020	92	8,113	1,255,568,683	25	171,882	2,057	27,258	1,241,846,690
1000	-	-	-	-	20	167,327	2,112	27,729	1,241,836,309
1000	-	-	-	-	15	163,476	2,166	28,200	1,241,828,287
1000	-	-	-	-	10	159,062	2,233	28,719	1,241,822,133
1000	-	-	-	-	5	155,200	2,286	29,229	1,241,818,593
1000	-	-	-	-	1	51,301	7,054	65,092	1,241,782,051
500	3,937,821	93	17,952	1,262,291,236	25	170,589	2,076	37,544	1,248,502,317
500	-	-	-	-	20	166,062	2,132	38,023	1,248,491,764
500	-	-	-	-	15	162,595	2,186	38,504	1,248,483,572
500	-	-	-	-	10	158,193	2,254	39,038	1,248,477,239
500	-	-	-	-	5	153,747	2,308	39,562	1,248,473,599
500	-	-	-	-	1	50,930	7,119	76,379	1,248,436,081
300	3,915,799	95	48356	1,273,290,518	25	168,721	2,109	67,949	1,259,501,544
300	-	-	-	-	20	164,817	2,166	68,428	1,259,490,991
300	-	-	-	-	15	161,269	2,220	68,909	1,259,482,799
300	-	-	-	-	10	156,434	2,289	69,443	1,259,476,466
300	-	-	-	-	5	152,072	2,344	69,967	1,259,472,826
300	-	-	-	-	1	50,425	7,228	106,823	1,259,435,266
None	1,836,279	209	3,754,965	1,882,109,172	25	81,400	4,678	3,774,558	1,868,320,198
None	-	-	-	-	20	79,089	4,800	3,775,037	1,868,309,645
None	-	-	-	-	15	77,624	4,921	3,775,518	1,868,301,453
None	-	-	-	-	10	76,045	5,061	3,776,052	1,868,295,120
None	-	-	-	-	5	73,935	5,180	3,776,576	1,868,291,480
None	-	-	-	-	1	25,761	15,609	3,813,432	1,868,253,920

Table S5. Final SOAPdenovo2 parameters.

This table lists the SOAPdenovo2 parameters that we specified for each library to generate the final assembly.

Library	Paired or unpaired reads	Configuration file insert size (nt)	Used in contig or scaffold building	Assembly usage rank	Pair number cutoff	Mapping length (nt)
Nextera350nt lane 1	paired	247	both	1	3	32
Nextera350nt lane 2	paired	247	both	1	3	32
Nextera700nt	paired	566	both	2	3	32
noPCR550nt	paired	619	both	1	3	32
PCR900nt	paired	687	both	2	3	32
MP4kb	paired	3,316	scaffold	3	5	35
MP7kb	paired	5,904	scaffold	4	5	35
MP11kb	paired	9,615	scaffold	5	5	35
Nextera350nt lane 1	unpaired	N/A	contig	6	3	32
Nextera350nt lane 2	unpaired	N/A	contig	6	3	32
Nextera550nt lane 1	unpaired	N/A	contig	6	3	32
Nextera550nt lane 2	unpaired	N/A	contig	6	3	32
Nextera700nt	unpaired	N/A	contig	6	3	32
noPCR550nt	unpaired	N/A	contig	6	3	32
PCR900nt	unpaired	N/A	contig	6	3	32
MP4kb	unpaired	N/A	contig	6	3	32
MP7kb	unpaired	N/A	contig	6	3	32
MP11kb	unpaired	N/A	contig	6	3	32

Table S6. Full assembly metrics.

Listed here are metrics on the full assembly (no contaminate or mitochondrial sequences removed) before gap-closing, after gap-closing, and after gap-closing and removal of all contigs and scaffolds less than 1000 nt in length. Strings of 25 or more N's broke scaffolds into contigs.

Assembly version	No gap-closing, scaffolds and contigs <1000 nt removed	Gap-closed, no scaffolds or contigs removed	Gap-closed, scaffolds and contigs <1000 nt removed
Number of scaffolds	3,754,965	3,754,965	8,113
Total size of scaffolds	1,884,424,465 nt	1,882,109,172 nt	1,255,568,683 nt
Longest scaffold	15,783,852 nt	15,750,186 nt	15,750,186 nt
Shortest scaffold	128 nt	128 nt	1,000 nt
Number of scaffolds > 1K nt	8,117 (0.2%)	8,100 (0.2%)	8,100 (99.8%)
Number of scaffolds > 10K nt	1,755 (0.0%)	1,747 (0.0%)	1,747 (21.5%)
Number of scaffolds > 100K nt	661 (0.0%)	661 (0.0%)	661 (8.1%)
Number of scaffolds > 1M nt	303 (0.0%)	303 (0.0%)	303 (3.7%)
Number of scaffolds > 10M nt	9 (0.0%)	9 (0.0%)	9 (0.1%)
Mean scaffold size	502 nt	501 nt	154,760 nt
Median scaffold size	150 nt	150 nt	1,903 nt
N50 scaffold length (L50 scaffold count)	1,843,286 nt (209)	1,836,279 nt (209)	3,983,020 nt (92)
N60 scaffold length (L60 scaffold count)	622,124 nt (370)	619,581 nt (371)	3,012,707 nt (129)
N70 scaffold length (L70 scaffold count)	255 nt (216,224)	255 nt (218,948)	2,142,451 nt (178)
N80 scaffold length (L80 scaffold count)	174 nt (1,110,557)	174 nt (1,113,218)	1,545,070 nt (246)
N90 scaffold length (L90 scaffold count)	143 nt (2,336,944)	143 nt (2,338,563)	618,731 nt (372)
scaffold %GC	42.81%	43.82%	41.31%
scaffold %N	2.89%	0.74%	1.10%
Percentage of assembly in scaffolded contigs	66.4%	65.7%	98.5%
Percentage of assembly in unscaffolded contigs	33.6%	34.3%	1.5%
Average number of contigs per scaffold	1.0	1.0	3.4
Average length of break (>25 Ns) between contigs in scaffold	311	703	716
Number of contigs	3,929,051	3,774,558	27,258
Number of contigs in scaffolds	179,957	22,374	21,480
Number of contigs not in scaffolds	3,749,094	3,752,184	5,778
Total size of contigs	1,830,129,061 nt	1,868,320,198 nt	1,241,846,690 nt
Longest contig	186,255 nt	1,259,046 nt	1,259,046 nt
Shortest contig	5 nt	128 nt	130 nt
Number of contigs > 1K nt	123,899 (3.2%)	23,921 (0.6%)	23,921 (87.8%)
Number of contigs > 10K nt	37,347 (1.0%)	12,374 (0.3%)	12,374 (45.4%)
Number of contigs > 100K nt	58 (0.0%)	3,909 (0.1%)	3,909 (14.3%)
Number of contigs > 1M nt	0 (0.0%)	8 (0.0%)	8 (0.0%)
Mean contig size	466 nt	495 nt	45,559 nt
Median contig size	150 nt	150 nt	6,696 nt
N50 contig length (L50 contig count)	7,855 nt (46,857)	81,400 nt (4,678)	171,882 nt (2,057)
N60 contig length (L60 contig count)	3,275 nt (81,604)	33521 nt (8,121)	134,419 nt (2,876)
N70 contig length (L70 contig count)	254 nt (448,713)	255 nt (254,707)	98,599 nt (3,956)
N80 contig length (L80 contig count)	170 nt (1,346,253)	173 nt (1,148,670)	66,629 nt (5,485)
N90 contig length (L90 contig count)	142 nt (2,548,885)	142 nt (2,367,834)	34,559 nt (8,023)

Table S7. Statistics from after quality-filtering MAKER annotations.

This is a table of annotation summary statistics resulting from quality-filtering our MAKER pipeline annotation output.

	Values post -s filter
parsed genome node DAGs	745,622
sequence regions	8,112 (total length: 1,255,013,157 nt)
multi-features	15,712
genes	16,718
protein-coding genes	16,718
mRNAs	16,718
protein-coding mRNAs	16,718
exons	146,689
CDSs	146,217

Table S8. Mitochondrial genome assembly gene annotations.

This is a table of the gene annotations of the assembly of a partial mitochondrial genome represented by scaffold-3674. The coordinates are 1-based.

Gene	Scaffold	Start position	End position	Direction
tRNA^{Thr}	scaffold3674	231	299	-
Cytb	scaffold3674	307	1431	-
ND5	scaffold3674	1463	3268	-
tRNA^{Leu1}	scaffold3674	3269	3339	-
tRNA^{Ser1}	scaffold3674	3342	3407	-
tRNA^{His}	scaffold3674	3410	3479	-
ND4L	scaffold3674	3490	4857	-
ND4L	scaffold3674	4854	5147	-
tRNA^{Arg}	scaffold3674	5149	5218	-
ND3 b	scaffold3674	5224	5397	-
ND3 a	scaffold3674	5399	5572	-
tRNA^{Gly}	scaffold3674	5573	5641	-
COIII	scaffold3674	5643	6425	-
ATP6	scaffold3674	6431	7108	-
ATP8	scaffold3674	7105	7266	-
tRNA^{Lys}	scaffold3674	7268	7338	-
COII	scaffold3674	7357	8031	-
tRNA^{Asp}	scaffold3674	8034	8102	-
tRNA^{Ser2}	scaffold3674	8106	8177	+
COI	scaffold3674	8178	9710	-
tRNA^{Tyr}	scaffold3674	9721	9791	+
tRNA^{Cys}	scaffold3674	9792	9860	+
tRNA^{Asn}	scaffold3674	9863	9936	+
tRNA^{Ala}	scaffold3674	9938	10006	+
tRNA^{Trp}	scaffold3674	10008	10083	-
ND2	scaffold3674	10094	11122	-
tRNA^{Met}	scaffold3674	11123	11191	-
tRNA^{Gln}	scaffold3674	11191	11261	+
tRNA^{Ile}	scaffold3674	11273	11344	-
ND1	scaffold3674	11352	12299	-
tRNA^{Leu2}	scaffold3674	12314	12387	-
16S	scaffold3674	12387	13982	-
tRNA^{Val}	scaffold3674	13983	14054	-
12S	scaffold3674	14054	15041	-
tRNA^{Phe}	scaffold3674	15041	15108	-
tRNA^{Glu}	scaffold3674	21542	21614	+

Table S9. Information on searches for light-associated genes in non-owl genome assemblies.

This table provides information on the results of our searches for a subset of the light-associated genes in several non-owl avian genome assemblies. "Stop" indicates the presence of a premature stop codon. "Del" indicates a frameshift deletion. For these searches we employed the same reference sequences used in the owl genome searches, detailed in Table S3.

	Rh2	***OpnP***	***Opn4m***	***CYP2J19***
Reference Sequence	GenBank: AH007731 *Columba livia*	GenBank: U15762, WGS: AADN03007691 *Gallus gallus*	GenBank: AY882944, WGS: AADN04000143 *Gallus gallus*	GenBank: XM_422553, WGS: AADN04000032 *Gallus gallus*
***Aquila chrysaetos* Sequence**	JRUM01011001	JRUM01006324	**JRUM01004396: Pseudogene? (exon 9: stop)**	JRUM01002169
***Cathartes aura* Sequence**	JMFT01083953	JMFT01020150, JMFT01020151, JMFT01020152, JMFT01020153	JMFT01012857, JMFT01012858, JMFT01012859	JMFT01168756
***Colius striatus* Sequence**	JJRP01038063, JJRP01092220	JJRP01068983	**JJRP01099016, JJRP01099018, JJRP01099019: Pseudogene? (exon 9: 1-bp del; intron 9: splice donor mutation GT to TT; exon 11: stop)**	JJRP01092926
***Leptosomus discolor* Sequence**	JJRK01095962, JJRK01095963	JJRK01016598, JJRK01016599	**JJRK01001211, JJRK01001212, JJRK01001213: Pseudogene? (intron 10: splice donor mutation GT to GA)**	JJRK01096026
***Apaloderma vittatum* Sequence**	JMFV01047445, JMFV01047446	JMFV01046166, JMFV01046167	JMFV01094831	JMFV01067118, JMFV01102670, JMFV01104326, JMFV01105382
***Buceros rhinoceros* Sequence**	JMFK01024225	JMFK01144445, JMFK01144446, JMFK01144447, JMFK01144448,	**JMFK01158949, JMFK01158950, JMFK01158951, JMFK01158952: Pseudogene? (exon 1: start codon mutation CTG)**	JMFK01006414, JMFK01073748
***Picoides pubescens* Sequence**	JJRU01080411, JJRU01080413	JJRU01064065	JJRU01054812	JJRU01010544, JJRU01010545
***Merops nubicus* Sequence**	JJRJ01051189	JJRJ01058175	JJRJ01007844	JJRJ01011917, JJRJ01033855

Table S10. Details of branch tests.

This table gives the details of the branch tests performed to test for evidence of changes in selection pressure on the owl branches. "BG" indicates the background branches, "lnL" denotes the log likelihood of the model, "LRT" denotes the value of the likelihood ratio test (given by 2 times the difference in the likelihoods of the models), and "cf" denotes the codon frequency model used to calculate the equilibrium codon frequencies with "cf 1" indicating that we used the average nucleotide frequencies and "cf 2" indicating that we used the average nucleotide frequencies at each of the 3 codon positions. "Model" corresponds to the number of ω values employed among branches with one ω value assumed for all branches under model "0", two ω values used under model "1", and 3 ω values used with model "2". "*Tyto*" and "*Strix*" indicate whether the value pertains to sequence in the *Tyto alba* or *Strix occidentalis caurina* genome assembly, respectively. For model comparisons, bold font indicates significant difference ($p < 0.05$) between models.

Gene	Model	BG ω	*Tyto* ω	*Strix* ω	Stem Owl ω	lnL	Models compared	LRT
***CYP2J19* (cf 1)**	0	0.206				-5045.714		
	1	0.173	0.719			-5029.495	**1 vs. 0**	32.437
	2	0.164	0.719	0.336		-5027.178	**2 vs. 1**	4.633
***CYP2J19* (cf 2)**	0	0.194				-5050.277		
	1	0.163	0.681			-5034.027	**1 vs. 0**	32.499
	2	0.154	0.680	0.333		-5031.418	**2 vs. 1**	5.219
***OPN4M* (cf 1)**	0	0.214				-3345.378		
	1	0.192	0.448	0.448	0.895	-3341.951	1 vs. 0	6.854
***OPN4M* (cf 2)**	0	0.213				-3350.019		
	1	0.190	0.452	0.452	0.864	-3346.487	**1 vs. 0**	7.066
***OPNP* (cf 1)**	0	0.234				-3937.560		
	1	0.180	0.695			-3918.377	**1 vs. 0**	38.446
***OPNP* (cf 2)**	0	0.152				-3892.939		
	1	0.114	0.508			-3870.379	**1 vs. 0**	45.121
***RH2* (cf 1)**	0	0.079				-3155.354		
	1	0.057	0.367			-3139.086	**1 vs. 0**	32.536
	2	0.052	0.358	0.208		-3136.501	**2 vs. 1**	5.170
***RH2* (cf 2)**	0	0.043				-3054.733		
	1	0.031	0.219			-3037.200	**1 vs. 0**	35.065
	2	0.028	0.205	0.158		-3033.836	**2 vs. 1**	6.728

Table S11. Details of branch-site tests.

This table provides details of the tests performed using branch-site models implemented in the phylogenetic analysis by maximum likelihood (PAML) package to detect positive selection affecting certain sites on the owl lineages. "*Tyto*" and "*Strix*" indicate whether the values pertain to sequence in the *Tyto alba* or *Strix occidentalis caurina* genome assembly, respectively. "BG" indicates the background branches, "FG" denotes the foreground branch, "lnL" denotes the log likelihood of the model, "LRT" denotes the value of the likelihood ratio test (given by 2 times the difference in the likelihoods of the models), and "cf" denotes the codon frequency model used to calculate the equilibrium codon frequencies with "cf 1" indicating that we used the average nucleotide frequencies and "cf 2" indicating that we used the average nucleotide frequencies at each of the 3 codon positions. "Site class" indicates the ω category with "0" indicating sites under purifying selection, "1" sites under relaxed selection, "2a" sites that are under positive selection on the foreground branch and under purifying selection on the background branches, and "2b" indicating positive selection on the foreground branch and relaxed selection on the background branches. "Proportion" indicates the proportion of sites in a given class. "Model" denotes either the positive selection model ("Positive") or the null model ("Null").

Gene	Taxon	Site class	Proportion	BG ω	FG ω	Model	lnL	LRT
***OPN4M* (cf 1)**	*Strix*	0	0.778	0.047	0.047			
		1	0.184	1	1			
		2a	0.031	0.047	4.291			
		2b	0.007	1	4.291			
						Positive	-3305.681	
						Null	-3305.984	-0.605
	Tyto	0	0.773	0.046	0.046			
		1	0.190	1	1			
		2a	0.030	0.046	1.660			
		2b	0.007	1	1.660			
						Positive	-3306.308	
						Null	-3306.325	-0.033
***OPN4M* (cf 2)**	*Strix*	0	0.773	0.047	0.047			
		1	0.182	1	1			
		2a	0.036	0.047	4.051			
		2b	0.009	1	4.051			
						Positive	-3310.564	
						Null	-3310.887	-0.646
	Tyto	0	0.788	0.050	0.050			
		1	0.189	1	1			
		2a	0.019	0.050	2.072			
		2b	0.004	1	2.072			
						Positive	-3311.582	
						Null	-3311.605	-0.047

4 Supplementary Figures

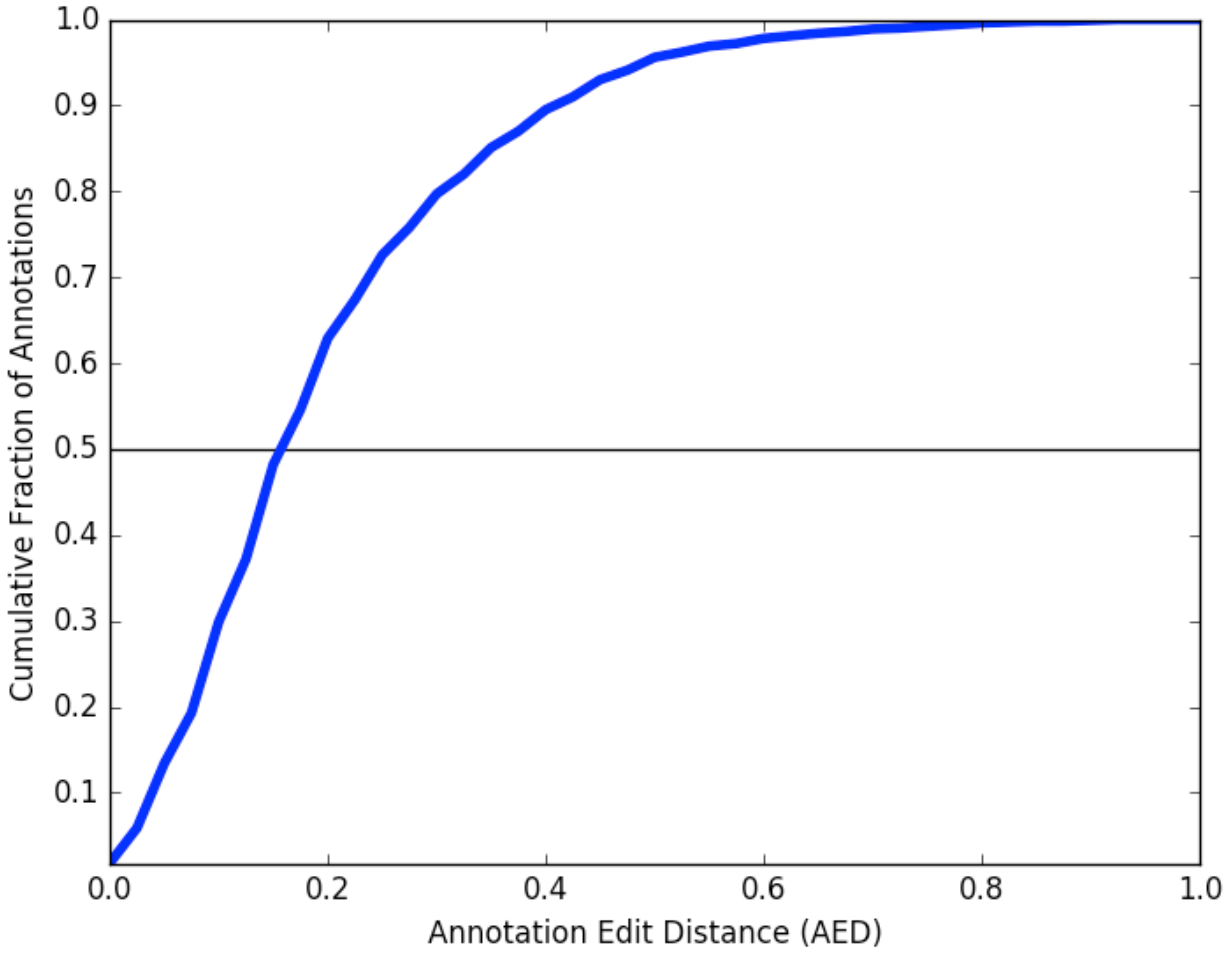

Figure S1. Cumulative distribution of annotation edit distances of MAKER-generated annotations.

This is a graph of the cumulative distribution of annotation edit distances (AED) of the annotations generated by MAKER. Included here are all of the annotations in the MAKER final output. We have drawn a horizontal line denoting 50% of the annotations. After quality filtering, the cumulative distribution appeared identical.

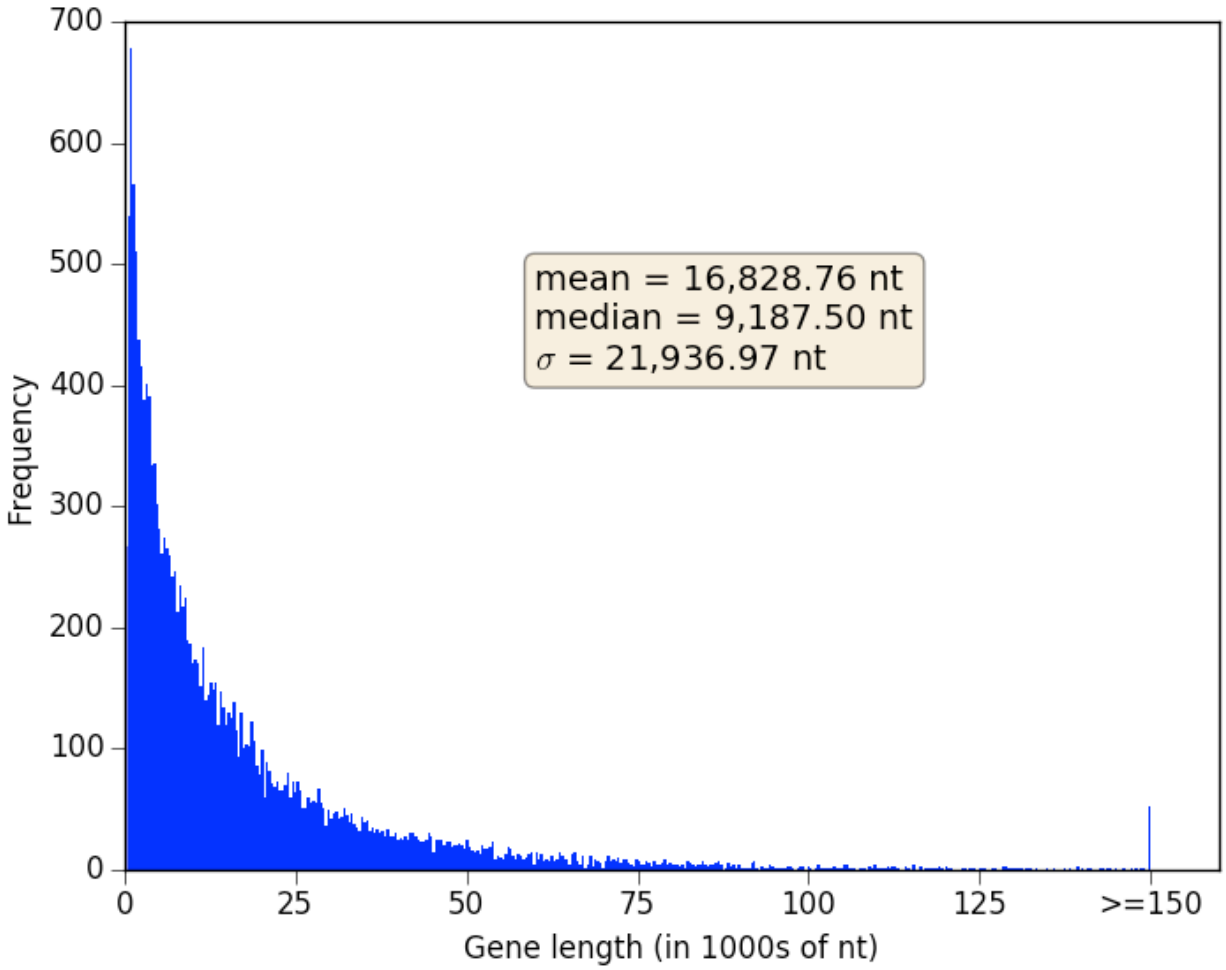

Figure S2. Histogram of the lengths of genes annotated by MAKER.

This is a histogram of the distribution of the lengths of genes annotated by MAKER. We included all of the gene annotations in the MAKER final output. We grouped the values into 400 frequency bins, one of these including all genes greater than or equal to 150,000 nt in length. We have provided the mean, median, and standard deviation of the gene lengths in a text box.

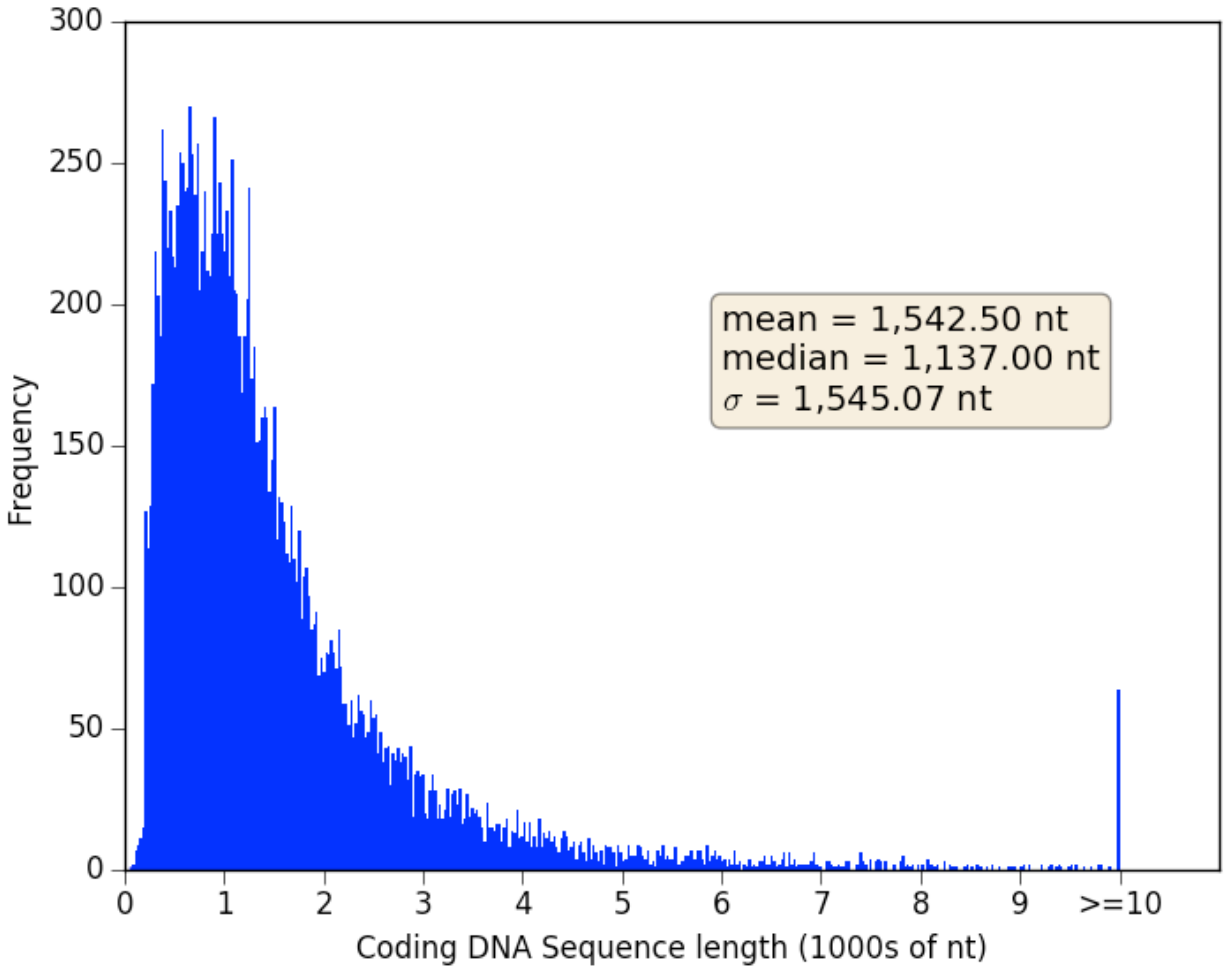

Figure S3. Histogram of the coding DNA sequence length in genes annotated by MAKER.

This is a histogram of the lengths of coding DNA sequences in genes annotated by MAKER. We included all of the gene annotations in the MAKER final output. We grouped the values into 400 frequency bins, one of these including all coding DNA sequences greater than or equal to 10,000 nt in length. We have provided the mean, median, and standard deviation of the lengths in a text box.

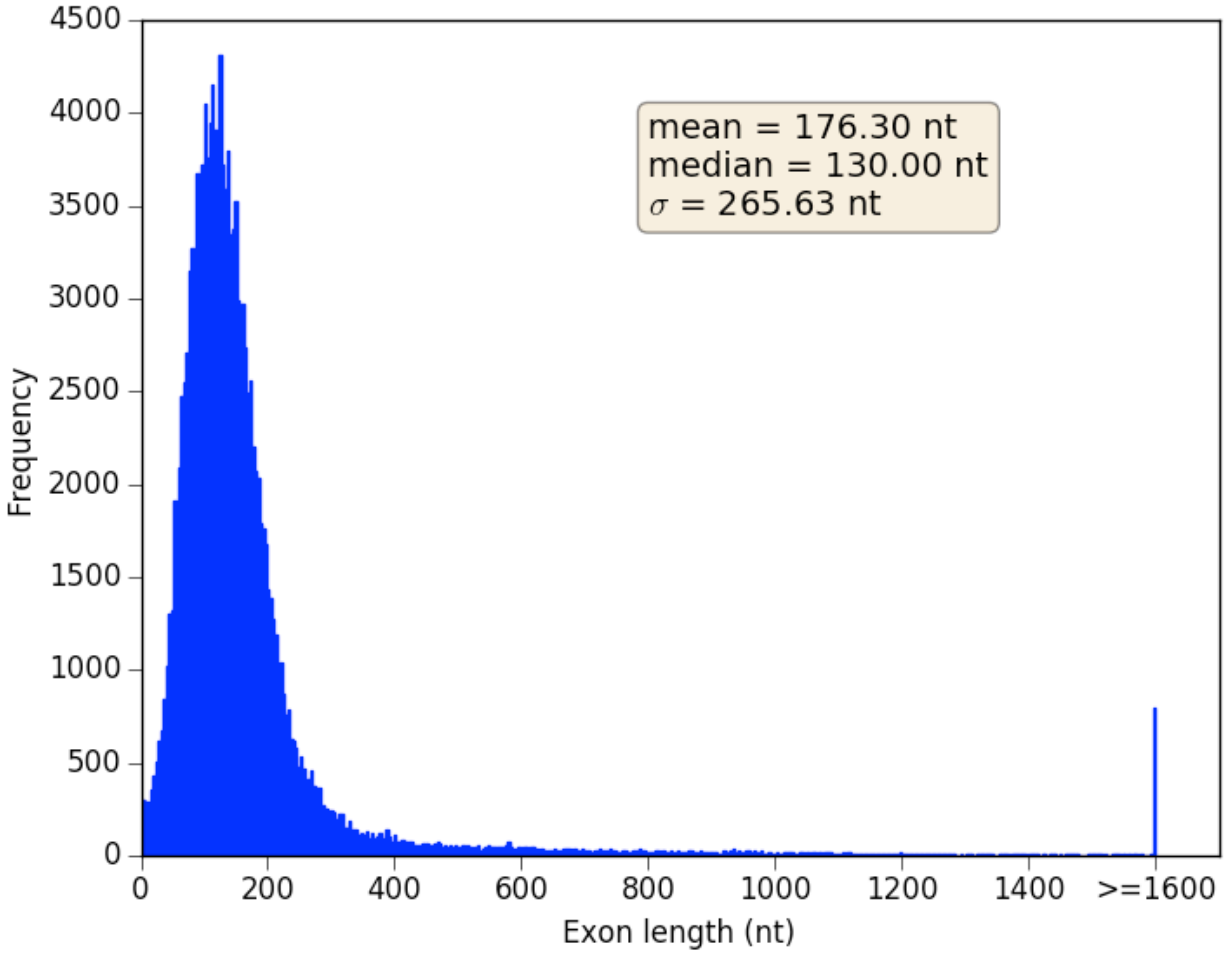

Figure S4. Histogram of the lengths of exons in genes annotated by MAKER.

This is a histogram of the lengths of exons in genes annotated by MAKER. We included the exons from all of the gene annotations in the MAKER final output. We grouped the values into 400 frequency bins, one of these including all exons greater than or equal to 1,600 nt in length. We have provided the mean, median, and standard deviation of the exon lengths in a text box.

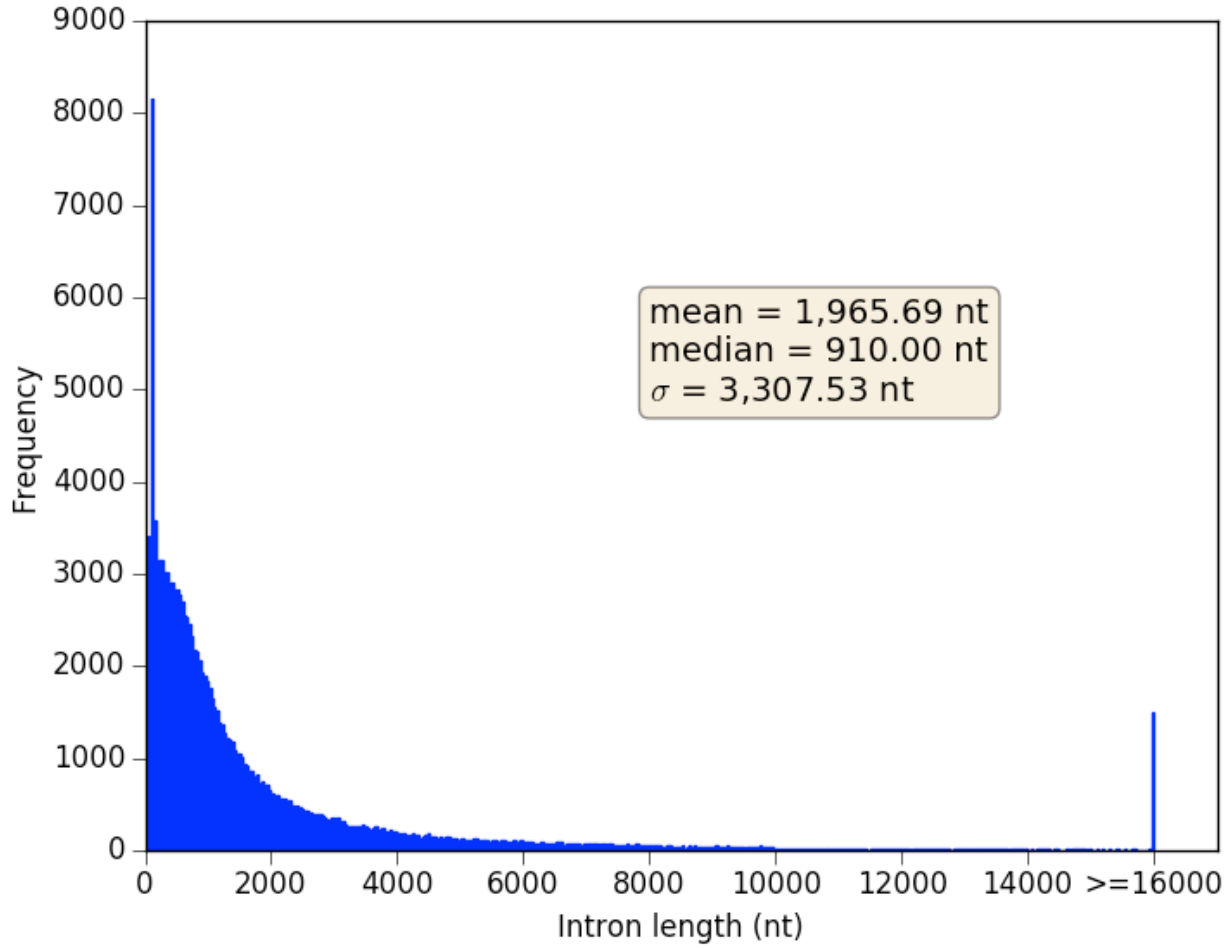

Figure S5. Histogram of the lengths of introns in genes annotated by MAKER.

This is a histogram of the lengths of introns in genes annotated by MAKER. We included the introns from all of the gene annotations in the MAKER final output. We grouped the values into 400 frequency bins, one of these including all introns greater than or equal to 16,000 nt in length. We have provided the mean, median, and standard deviation of the intron lengths in a text box.

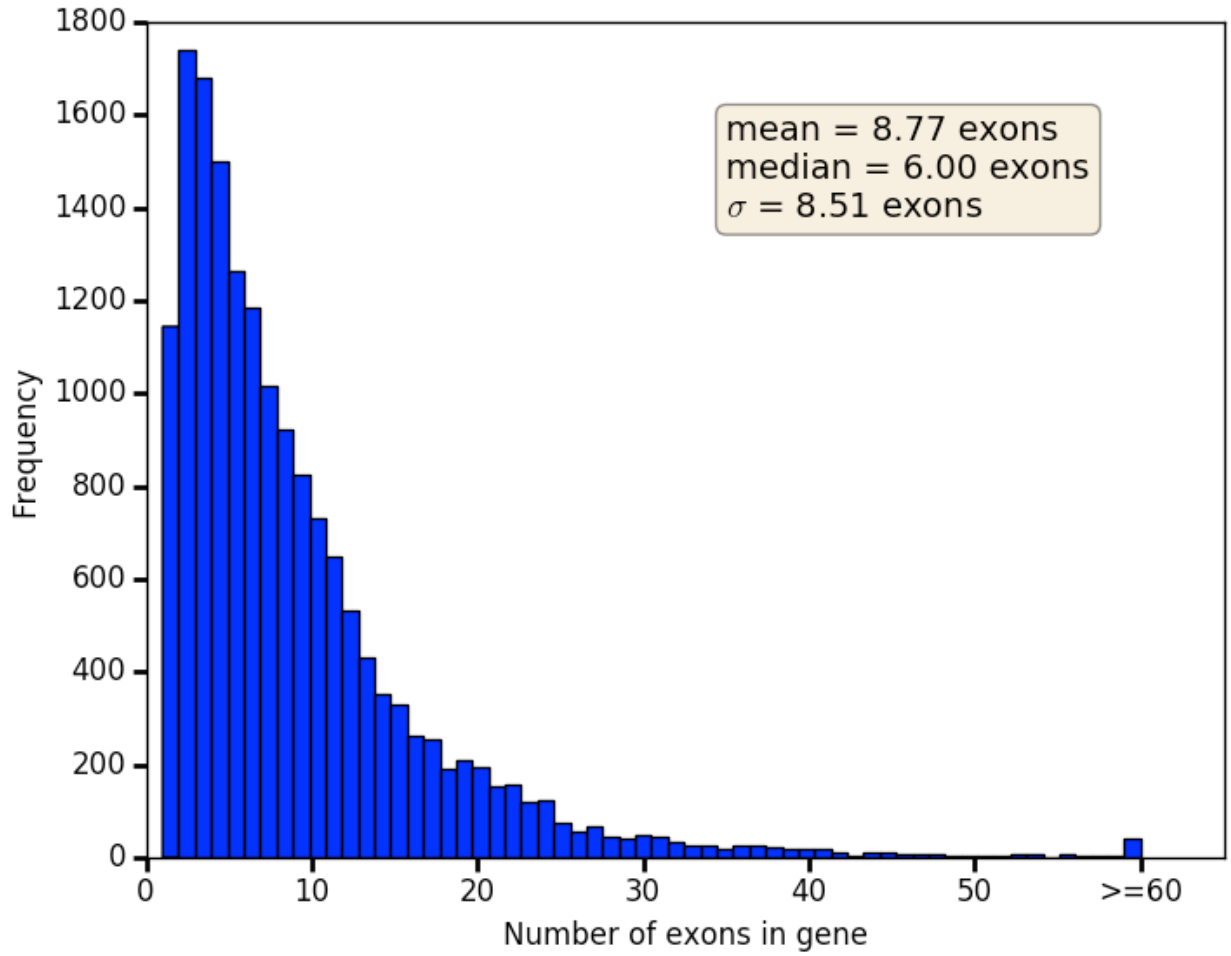

Figure S6. Histogram of the number of exons in genes annotated by MAKER.

This is a histogram of the number of exons in genes annotated by MAKER. We included the exons from all of the gene annotations in the MAKER final output. We grouped the values into 60 frequency bins, one of these including all genes with greater than or equal to 60 exons. We have provided the mean, median, and standard deviation of the number of exons per gene in a text box.

5 Supplementary References

Altschul SF., Madden TL., Schäffer AA., Zhang J., Zhang Z., Miller W., Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25:3389–3402. DOI: 10.1093/nar/25.17.3389.

Bao Z., Eddy SR. 2002. Automated De Novo Identification of Repeat Sequence Families in Sequenced Genomes. *Genome Research* 12:1269–1276. DOI: 10.1101/gr.88502.

Bao W., Kojima KK., Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* 6:1–6. DOI: 10.1186/s13100-015-0041-9.

Barnett D., Garrison E., Marth G., Strömberg M. 2015. BamTools. Version 2.4.0. [Accessed 2016 Oct 1]. Available from: https://github.com/pezmaster31/bamtools.

Barnett DW., Garrison EK., Quinlan AR., Strömberg MP., Marth GT. 2011. BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* 27:1691–1692. DOI: 10.1093/bioinformatics/btr174.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research* 27:573–580. DOI: 10.1093/nar/27.2.573.

Bernt M., Donath A., Jühling F., Externbrink F., Florentz C., Fritzsch G., Pütz J., Middendorf M., Stadler PF. 2013. MITOS: Improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution* 69:313–319. DOI: 10.1016/j.ympev.2012.08.023.

Biomatters 2016a. Geneious. Version 9.1.6. [Accessed 2016 Oct 1]. Available from: http://www.geneious.com.

Biomatters 2016b. Geneious. Version 9.1.4. [Accessed 2016 Oct 1]. Available from: http://www.geneious.com.

Birney E. 2008. Wise2. Version 2.2.3-rc7. [Accessed 2016 Oct 1]. Available from: http://korflab.ucdavis.edu/datasets/cegma/#SCT4.

Birney E., Clamp M., Durbin R. 2004. GeneWise and Genomewise. *Genome Research* 14:988–995. DOI: 10.1101/gr.1865504.

Blanco E., Alioto T., Guigó R., Parra G., Camara F., Abril JF., Burset M., Messequer X. 2011. geneid. Version 1.1.4. [Accessed 2016 Oct 1]. Available from: http://genome.crg.es/software/geneid.

Bolger AM., Lohse M., Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. DOI: 10.1093/bioinformatics/btu170.

Bradnam KR., Fass JN., Alexandrov A., Baranay P., Bechner M., Birol I., Boisvert S., Chapman JA., Chapuis G., Chikhi R., Chitsaz H., Chou W-C., Corbeil J., Del Fabbro C., Docking TR., Durbin R., Earl D., Emrich S., Fedotov P., Fonseca NA., Ganapathy G., Gibbs RA., Gnerre S., Godzaridis E., Goldstein S., Haimel M., Hall G., Haussler D., Hiatt JB., Ho IY., Howard J., Hunt M., Jackman SD., Jaffe DB., Jarvis ED., Jiang H., Kazakov S., Kersey PJ., Kitzman JO., Knight JR., Koren S., Lam T-W., Lavenier D., Laviolette F., Li Y., Li Z., Liu B., Liu Y., Luo R., MacCallum I., MacManes MD., Maillet N., Melnikov S., Naquin D., Ning Z., Otto TD., Paten B., Paulo OS., Phillippy AM., Pina-Martins F., Place M., Przybylski D., Qin X., Qu C., Ribeiro FJ., Richards S., Rokhsar DS., Ruby JG., Scalabrin S., Schatz MC., Schwartz DC., Sergushichev A., Sharpe T., Shaw TI., Shendure J., Shi Y., Simpson JT., Song H., Tsarev F., Vezzi F., Vicedomini R., Vieira BM., Wang J., Worley KC., Yin S., Yiu S-M., Yuan J., Zhang G., Zhang H., Zhou S.,

Korf IF. 2013. Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. *GigaScience* 2:10. DOI: 10.1186/2047-217X-2-10.
Bushnell B. 2014. BBMap. Version 34.00. [Accessed 2016 Oct 1]. Available from: http://sourceforge.net/projects/bbmap.
Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. DOI: 10.1186/1471-2105-10-421.
Campbell MS., Holt C., Moore B., Yandell M. 2014. Genome Annotation and Curation Using MAKER and MAKER-P. *Current Protocols in Bioinformatics* 48:4.11.1-4.11.39. DOI: 10.1002/0471250953.bi0411s48.
Cantarel BL., Korf I., Robb SMC., Parra G., Ross E., Moore B., Holt C., Alvarado AS., Yandell M. 2008. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research* 18:188–196. DOI: 10.1101/gr.6743907.
Clark K., Karsch-Mizrachi I., Lipman DJ., Ostell J., Sayers EW. 2016. GenBank. *Nucleic Acids Research* 44:D67–D72. DOI: 10.1093/nar/gkv1276.
Consortium TU. 2015. UniProt: a hub for protein information. *Nucleic Acids Research* 43:D204–D212. DOI: 10.1093/nar/gku989.
De Vita R., Cavallo D., Eleuteri P., Dell'Omo G. 1994. Evaluation of interspecific DNA content variations and sex identification in Falconiformes and Strigiformes by flow cytometric analysis. *Cytometry* 16:346–350. DOI: 10.1002/cyto.990160409.
DePristo MA., Banks E., Poplin R., Garimella KV., Maguire JR., Hartl C., Philippakis AA., del Angel G., Rivas MA., Hanna M., McKenna A., Fennell TJ., Kernytsky AM., Sivachenko AY., Cibulskis K., Gabriel SB., Altshuler D., Daly MJ. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* 43:491–498. DOI: 10.1038/ng.806.
Doležel J., Bartoš J., Voglmayr H., Greilhuber J. 2003. Letter to the editor. *Cytometry Part A* 51A:127–128. DOI: 10.1002/cyto.a.10013.
Free Software Foundation 2012. GNU Awk . Version 4.0.1. [Accessed 2016 Oct 1]. Available from: https://www.gnu.org/software/gawk.
Fridolfsson A-K., Ellegren H. 1999. A Simple and Universal Method for Molecular Sexing of Non-Ratite Birds. *Journal of Avian Biology* 30:116–121. DOI: 10.2307/3677252.
Gabriel E., Fagg GE., Bosilca G., Angskun T., Dongarra JJ., Squyres JM., Sahay V., Kambadur P., Barrett B., Lumsdaine A., Castain RH., Daniel DJ., Graham RL., Woodall TS. 2004. Open MPI: Goals, Concept, and Design of a Next Generation MPI Implementation. In: *Proceedings, 11th European PVM/MPI Users' Group Meeting*. Budapest, Hungary, 97–104.
Gailly J., Adler M. 2013. zlib. Version 1.2.8. [Accessed 2016 Oct 1]. Available from: http://www.zlib.net.
google-sparsehash@googlegroups.com 2012. Google SparseHash. Version 2.0.2. [Accessed 2016 Oct 1]. Available from: https://github.com/sparsehash/sparsehash.
Granlund T., Stallman RM. 2013. cat (GNU coreutils). Version 8.21. [Accessed 2016 Oct 1]. Available from: http://www.gnu.org/software/coreutils/coreutils.html.
Gremme G., Steinbiss S., Kurtz S. 2013. GenomeTools: A Comprehensive Software Library for Efficient Processing of Structured Genome Annotations. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 10:645–656. DOI: 10.1109/TCBB.2013.68.

Guigó R. 1998. Assembling Genes from Predicted Exons in Linear Time with Dynamic Programming. *Journal of Computational Biology* 5:681–702. DOI: 10.1089/cmb.1998.5.681.
Gutiérrez RJ., Franklin AB., Lahaye WS. 1995. Spotted Owl (*Strix occidentalis*). *The Birds of North America Online* (A. Poole, Ed.). Ithaca: Cornell Lab of Ornithology. [Accessed 2016 Oct 1]. Retrieved from the Birds of North America Online: https://birdsna.org/Species-Account/bna/species/spoowl. DOI: 10.2173/bna.179.
Hamer TE., Forsman ED., Fuchs AD., Walters ML. 1994. Hybridization between Barred and Spotted Owls. *The Auk* 111:487–492. DOI: 10.2307/4088616.
Hanna ZR., Henderson JB. 2017. NSO-genome-scripts. Version 1.0.0. *Zenodo*. DOI: 10.5281/zenodo.805012.
Henderson JB., Hanna ZR. 2016a. scafSeqContigInfo. Version 1.0.0. *Zenodo*. DOI: 10.5281/zenodo.163748.
Henderson JB., Hanna ZR. 2016b. dupchk. Version 1.0.0. *Zenodo*. DOI: 10.5281/zenodo.163722.
Henderson JB., Hanna ZR. 2016c. scafN50. Version 1.0.0. *Zenodo*. DOI: 10.5281/zenodo.163739.
Henderson JB., Hanna ZR. 2016d. ScaffSplitN50s. Version 1.0.0. *Zenodo*. DOI: 10.5281/zenodo.163683.
Henderson JB., Hanna ZR. 2016e. GItaxidIsVert. Version 1.0.0. *Zenodo*. DOI: 10.5281/zenodo.163737.
Hoffman W., Martin K. 2003. The CMake Build Manager. [Accessed 2016 Oct 1]. Available from: http://www.drdobbs.com/cpp/the-cmake-build-manager/184405251. *Dr. Dobb's*.
Hudson RR., Slatkin M., Maddison WP. 1992. Estimation of Levels of Gene Flow from DNA Sequence Data. *Genetics* 132:583–589.
Hunter JD. 2007. Matplotlib: A 2D graphics environment. *Computing In Science & Engineering* 9:90–95. DOI: 10.1109/MCSE.2007.55.
Ihnat DM., MacKenzie D., Meyering J. 2013. cut (GNU coreutils). Version 8.21. [Accessed 2016 Oct 1]. Available from: http://www.gnu.org/software/coreutils/coreutils.html.
Jones P., Binns D., Chang H-Y., Fraser M., Li W., McAnulla C., McWilliam H., Maslen J., Mitchell A., Nuka G., Pesseat S., Quinn AF., Sangrador-Vegas A., Scheremetjew M., Yong S-Y., Lopez R., Hunter S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30:1236–1240. DOI: 10.1093/bioinformatics/btu031.
Jurka J. 1998. Repeats in genomic DNA: mining and meaning. *Current Opinion in Structural Biology* 8:333–337. DOI: 10.1016/S0959-440X(98)80067-5.
Jurka J. 2000. Repbase Update: a database and an electronic journal of repetitive elements. *Trends in Genetics* 16:418–420. DOI: 10.1016/S0168-9525(00)02093-X.
Jurka J., Kapitonov VV., Pavlicek A., Klonowski P., Kohany O., Walichiewicz J. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research* 110:462–467. DOI: 10.1159/000084979.
Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A., Markowitz S., Duran C., Thierer T., Ashton B., Meintjes P., Drummond A. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649. DOI: 10.1093/bioinformatics/bts199.

Keller O., Kollmar M., Stanke M., Waack S. 2011. A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics* 27:757–763. DOI: 10.1093/bioinformatics/btr010.

Kitware 2015. CMake. Version 3.2.3. [Accessed 2016 Oct 1]. Available from: https://cmake.org.

Korf I. 2004. Gene finding in novel genomes. *BMC Bioinformatics* 5:1–9. DOI: 10.1186/1471-2105-5-59.

Leinonen R., Sugawara H., Shumway M. 2011. The Sequence Read Archive. *Nucleic Acids Research* 39:D19–D21. DOI: 10.1093/nar/gkq1019.

Li H. 2013a. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv:1303.3997 Q-Bio. [Accessed 2016 Feb 16]. Available from: http://arxiv.org/abs/1303.3997.

Li H. 2013b. bioawk. Version 1.0. [Accessed 2016 Oct 1]. Available from: https://github.com/lh3/bioawk.

Li H. 2015. PSMC. Version 0.6.5-r67. [Accessed 2016 Oct 1]. Available from: https://github.com/lh3/psmc.

Li H., Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature* 475:493–496. DOI: 10.1038/nature10231.

Li H., Handsaker B., Danecek P., McCarthy S., Marshall J. 2016a. bcftools. Version 1.3.1. [Accessed 2016 Oct 1]. Available from: https://github.com/samtools/bcftools.

Li H., Handsaker B., Marshall J., Danecek P. 2016b. Samtools. Version 1.3.1 with HTSlib 1.3.1. [Accessed 2016 Oct 1]. Available from: http://www.htslib.org.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R., Subgroup 1000 Genome Project Data Processing. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. DOI: 10.1093/bioinformatics/btp352.

Luo R., Liu B., Xie Y., Li Z., Huang W., Yuan J., He G., Chen Y., Pan Q., Liu Y., Tang J., Wu G., Zhang H., Shi Y., Liu Y., Yu C., Wang B., Lu Y., Han C., Cheung DW., Yiu S-M., Peng S., Xiaoqian Z., Liu G., Liao X., Li Y., Yang H., Wang J., Lam T-W., Wang J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* 1:18. DOI: 10.1186/2047-217X-1-18.

MacKenzie D. 2013. fold (GNU coreutils). Version 8.21. Available at <http://www.gnu.org/software/coreutils/coreutils.html>.

Matplotlib Development Team 2016. matplotlib. Version 1.5.1. [Accessed 2017 Sep 29]. Available from: https://github.com/matplotlib/matplotlib. DOI: 10.5281/zenodo.44579.

Mazur KM., James PC. 2000. Barred Owl (*Strix varia*). *The Birds of North America Online* (A. Poole, Ed.) Ithaca: Cornell Lab of Ornithology. [Accessed 2016 Oct 1]. Retrieved from the Birds of North America Online: https://birdsna.org/Species-Account/bna/species/brdowl. DOI: 10.2173/bna.508.

McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernytsky A., Garimella K., Altshuler D., Gabriel S., Daly M., DePristo MA. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20:1297–1303. DOI: 10.1101/gr.107524.110.

NCBI Resource Coordinators 2016. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research* 44:D7–D19. DOI: 10.1093/nar/gkv1290.

Noon BR., Biles CM. 1990. Mathematical Demography of Spotted Owls in the Pacific Northwest. *The Journal of Wildlife Management* 54:18–27. DOI: 10.2307/3808895.

NumPy Developers 2016. NumPy. Version 1.11.1. [Accessed 2017 Sep 29]. Available from: http://www.numpy.org.
O'Connell J. 2014. NxTrim. Version 0.2.3-alpha. [Accessed 2016 Oct 1]. Available from: https://github.com/sequencing/NxTrim.
O'Connell J., Schulz-Trieglaff O., Carlson E., Hims MM., Gormley NA., Cox AJ. 2015. NxTrim: optimized trimming of Illumina mate pair reads. *Bioinformatics* 31:2035–2037. DOI: 10.1093/bioinformatics/btv057.
Parra G., Bradnam K., Korf I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23:1061–1067. DOI: 10.1093/bioinformatics/btm071.
Price AL., Jones NC., Pevzner PA. 2005. De novo identification of repeat families in large genomes. *Bioinformatics* 21:i351–i358. DOI: 10.1093/bioinformatics/bti1018.
Python Software Foundation 2016. Python. Version 2.7.12. [Accessed 2016 Oct 1]. Available from: https://www.python.org.
Simão FA., Waterhouse RM., Ioannidis P., Kriventseva EV., Zdobnov EM. 2015a. BUSCO. Version 1.1b1. [Accessed 2016 Oct 1]. Available from: http://busco.ezlab.org.
Simão FA., Waterhouse RM., Ioannidis P., Kriventseva EV., Zdobnov EM. 2015b. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212. DOI: 10.1093/bioinformatics/btv351.
Simpson JT. 2014. Exploring genome characteristics and sequence quality without a reference. *Bioinformatics* 30:1228–1235. DOI: 10.1093/bioinformatics/btu023.
Simpson JT., Durbin R. 2010. Efficient construction of an assembly string graph using the FM-index. *Bioinformatics* 26:i367–i373. DOI: 10.1093/bioinformatics/btq217.
Simpson JT., Durbin R. 2016. SGA - String Graph Assembler. Version 0.10.14. [Accessed 2016 Oct 1]. Available from: https://github.com/jts/sga.
Slater GSC., Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6:1–11. DOI: 10.1186/1471-2105-6-31.
Smeds L., Qvarnström A., Ellegren H. 2016. Direct estimate of the rate of germline mutation in a bird. *Genome Research* 26:1211–1218. DOI: 10.1101/gr.204669.116.
Smit AFA., Hubley R. 2015. RepeatModeler Open-1.0. [Accessed 2016 Oct 1]. Available from: http://www.repeatmasker.org.
Smit A., Hubley R., Green P. 2013. RepeatMasker Open-4.0. [Accessed 2016 Oct 1]. Available from: http://www.repeatmasker.org.
Smit A., Hubley R., National Center for Biotechnology Information 2015. RMBlast. [Accessed 2016 Oct 1]. Available from: http://www.repeatmasker.org/RMBlast.html.
Stanke M. 2015. AUGUSTUS. Version 3.2.1. [Accessed 2016 Oct 1]. Available from: http://bioinf.uni-greifswald.de/augustus.
USDA Forest Service 1992. Final Environmental Impact Statement on Management for the Northern Spotted Owl in the National Forests. USDA Forest Service, National Forest System: Portland, Oregon. 2 vol.
Van der Auwera GA., Carneiro MO., Hartl C., Poplin R., del Angel G., Levy-Moonshine A., Jordan T., Shakir K., Roazen D., Thibault J., Banks E., Garimella KV., Altshuler D., Gabriel S., DePristo MA. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics* 11:11.10.1-11.10.33. DOI: 10.1002/0471250953.bi1110s43.

Vinogradov AE. 2005. Genome size and chromatin condensation in vertebrates. *Chromosoma* 113:362–369. DOI: 10.1007/s00412-004-0323-3.

Wheeler TJ., Clements J., Eddy SR., Hubley R., Jones TA., Jurka J., Smit AFA., Finn RD. 2013. Dfam: a database of repetitive DNA based on profile hidden Markov models. *Nucleic Acids Research* 41:D70–D82. DOI: 10.1093/nar/gks1265.

Wu Y., Hadly EA., Teng W., Hao Y., Liang W., Liu Y., Wang H. 2016. Retinal transcriptome sequencing sheds light on the adaptation to nocturnal and diurnal lifestyles in raptors. *Scientific Reports* 6:33578. DOI: 10.1038/srep33578.

Yates A., Akanni W., Amode MR., Barrell D., Billis K., Carvalho-Silva D., Cummins C., Clapham P., Fitzgerald S., Gil L., Girón CG., Gordon L., Hourlier T., Hunt SE., Janacek SH., Johnson N., Juettemann T., Keenan S., Lavidas I., Martin FJ., Maurel T., McLaren W., Murphy DN., Nag R., Nuhn M., Parker A., Patricio M., Pignatelli M., Rahtz M., Riat HS., Sheppard D., Taylor K., Thormann A., Vullo A., Wilder SP., Zadissa A., Birney E., Harrow J., Muffato M., Perry E., Ruffier M., Spudich G., Trevanion SJ., Cunningham F., Aken BL., Zerbino DR., Flicek P. 2016. Ensembl 2016. *Nucleic Acids Research* 44:D710–D716. DOI: 10.1093/nar/gkv1157.

Zhang Z., Schwartz S., Wagner L., Miller W. 2000. A Greedy Algorithm for Aligning DNA Sequences. *Journal of Computational Biology* 7:203–214. DOI: 10.1089/10665270050081478.

Chapter 2

Complete mitochondrial genome sequences of the northern spotted owl (*Strix occidentalis caurina*) and the barred owl (*Strix varia*; Aves: Strigiformes: Strigidae) confirm the presence of a duplicated control region

Zachary R. Hanna,[1,2,3,4] James B. Henderson,[3,4] Anna B. Sellas,[4,5] Jérôme Fuchs,[3,6] Rauri C. K. Bowie,[1,2] John P. Dumbacher[3,4]

[1] Museum of Vertebrate Zoology, University of California, Berkeley, Berkeley, California, United States of America
[2] Department of Integrative Biology, University of California, Berkeley, Berkeley, California, United States of America
[3] Department of Ornithology & Mammalogy, California Academy of Sciences, San Francisco, California, United States of America
[4] Center for Comparative Genomics, California Academy of Sciences, San Francisco, California, United States of America
[5] Chan Zuckerberg Biohub, San Francisco, California, United States of America
[6] UMR 7205 Institut de Systématique, Evolution, Biodiversité, CNRS, MNHN, UPMC, EPHE, Sorbonne Universités, Muséum National d'Histoire Naturelle, Paris, France

Introduction

The chicken (*Gallus gallus*) was the first avian species with a complete mitochondrial genome assembly (Desjardins & Morais, 1990). Subsequently, researchers assembled the mitochondrial genomes of members of the Paleognathae (e.g., ostriches, emus, kiwis) and other members of the Galloanserae (ducks, chicken-like birds) and recovered the same gene order found in the mitochondrial genome of the chicken, which led to the conclusion that the mitochondrial genome of the chicken is representative of the ancestral avian gene order (Desjardins & Morais, 1990; Mindell, Sorenson & Dimcheff, 1998a; Haddrath & Baker, 2001; Gibb et al., 2007). Almost a decade after publication of the chicken mitochondrial genome, Mindell, Sorenson & Dimcheff (1998a) described an alternative or, to use their terminology, "novel" avian gene order from that of the chicken, which included a different positioning of $tRNA^{Pro}$, *ND6*, and $tRNA^{Glu}$ relative to the control region sequence as well as an additional noncoding segment that they hypothesized was a degraded copy of the control region. A few years later, researchers first described the presence of an intact, duplicated control region in the mitochondrial genomes of *Amazona* parrots (Eberhard, Wright & Bermingham, 2001) and the common buzzard *Buteo buteo* (Haring et al., 2001).

Mindell, Sorenson & Dimcheff (1998a) detected their novel avian gene order in the mitochondrial genomes of taxa in multiple avian orders that spanned a significant portion of Neoaves, but did not detect it in the single owl species that they studied, *Otus asio* (Mindell, Sorenson & Dimcheff, 1998a). However, further investigation of owl (Strigiformes) mitochondrial genomes has revealed several surprises.

First, at least three wood owl species [*Strix aluco*, *S. uralensis* (Brito, 2005), and *S. varia* (Barrowclough et al., 2011)] contain the novel mitochondrial gene order of Mindell, Sorenson & Dimcheff (1998a) as well as duplicate control regions. The use of a primer in $tRNA^{Thr}$ to amplify a fragment of the control region suggests that the novel gene order is present in two additional wood owl species, *S. occidentalis* (Barrowclough, Gutierrez & Groth, 1999) and *S. nebulosa* (Hull et al., 2010). However, the novel gene order was not reported as present in the mitochondrial genome of *S. leptogrammica* (Liu, Zhou & Gu, 2014).

Second, some species of eagle-owls (genus *Bubo*) have a large control region (up to ~ 3,800 nucleotides) relative to *Strix*, their putative sister genus (Fuchs et al., 2008; Wink et al., 2009), largely due to a tandem repeat structure in the 3' end of the control region (Omote et al., 2013). Such control region tandem repeat blocks appear to be widespread in Strigidae (Xiao et al., 2006; Omote et al., 2013). These results suggest that the structures of owl mitochondrial genomes are surprisingly dynamic and in need of further investigation, particularly for species of conservation concern for which portions of the control region are used in population genetic studies (Barrowclough, Gutierrez & Groth, 1999; Haig et al., 2004; Hull et al., 2010, 2014).

We here provide the complete mitochondrial genome sequence of both a northern spotted owl (*Strix occidentalis caurina*) and barred owl (*S. varia*). The spotted owl (*S. occidentalis*) is a large and charismatic denizen of dense forests whose range includes the Pacific coast of North America from southwestern British Columbia to southern California and extends eastward into the deserts of the Southwestern United States and southward to central Mexico. The range of the northern spotted owl (*S. o. caurina*) subspecies includes the Pacific Northwest portion of the *S. occidentalis* range from British Columbia south to the Golden Gate strait, California. The U.S. Fish and Wildlife Service has listed *S. o. caurina* as "threatened" under the Endangered Species Act since 1990.

The barred owl (*S. varia*), formerly native east of the Rocky Mountains (Mazur & James, 2000), has extended its range into the western U.S. in the last 50-100 years and, from British Columbia to southern California, has become broadly sympatric with the northern spotted owl in the last 50 years. Barred and spotted owls hybridize and successfully backcross (Haig et al., 2004; Kelly & Forsman, 2004; Funk et al., 2007). Mitochondrial DNA sequencing has served as a valuable tool in ascertaining the maternal lineage of western birds, especially in potential hybrids (Zink, 1994; Haig et al., 2004; Barrowclough et al., 2005; Ruegg, 2008; Krosby & Rohwer, 2009; Williford et al., 2014).

Population-level studies of the genetics of *S. occidentalis* and *S. varia* have mainly used two mitochondrial markers, a partial control region sequence (Barrowclough, Gutierrez & Groth, 1999; Haig et al., 2004; Barrowclough et al., 2005) and *cytochrome b* (*cyt b*) (Haig et al., 2004), although a phylogeographic study of *S. varia* also utilized portions of *ND6* and *COIII* (Barrowclough et al., 2011). The sequences of the complete genomes of the mitochondria of these two species will aid researchers in utilizing additional mitochondrial markers in population genetic studies of these owls.

It is well known that mitochondrial genes can transfer to the nuclear genome; such regions of the nuclear genome are sometimes called *Numts* (Lopez et al., 1994; Sorenson & Quinn, 1998). As a high-quality nuclear genome of *S. o. caurina* is available (Hanna et al., 2017), we were able to explore the incidence of *Numts* within the nucleus and investigate which mitochondrial genes have most often transferred. Furthermore, by assessing divergence between mitochondrial genes and their descendent *Numts*, we ascertained the likelihood of them posing problems for phylogenetic and other types of studies.

Methods

Strix occidentalis mitochondrial genome assembly

We sourced *Strix occidentalis caurina* DNA from a blood sample collected by a veterinarian from a captive adult female *S. o. caurina* at WildCare rehabilitation facility in San Rafael, California. Found as an abandoned nestling in Larkspur, Marin County, California, WildCare admitted the captive owl as patient card # 849 on 5 June 2005 and named her Sequoia (sample preserved as CAS:ORN:98821; Table 1).

To assemble the *S. o. caurina* mitochondrial genome, we used paired-end Illumina sequence data from nine different genomic libraries constructed, sequenced, and processed as described in Hanna et al. (2017). The raw sequences from sample CAS:ORN:98821 (Table 1) are available from the NCBI Sequence Read Archive (SRA) (SRA run accessions SRR4011595, SRR4011596, SRR4011597, SRR4011614, SRR4011615, SRR4011616, SRR4011617, SRR4011618, SRR4011619, and SRR4011620). For our initial assembly, we used BLATq version 1.02 (Henderson & Hanna, 2016a), which was a modification of BLAT version 35 (Kent, 2002, 2012), to find Illumina reads that aligned to the *Ninox novaeseelandiae* mitochondrial genome (GenBank Accession AY309457.1) (Harrison et al., 2004) (Supplementary Materials (SM) 1.1.1) and extracted those matching reads using excerptByIds version 1.0.2 (Henderson & Hanna, 2016b) (SM 1.1.2). We then used SOAPdenovo2 version 2.04 (Luo et al., 2012) to assemble those sequences (SM 1.1.3).

We used the web version of the NCBI BLAST+ version 2.2.29 tool BLASTN (Altschul et al., 1990; Zhang et al., 2000; Morgulis et al., 2008; Camacho et al., 2009) to search the NCBI nucleotide collection (Johnson et al., 2008; Boratyn et al., 2013; Benson et al., 2015; NCBI Resource Coordinators, 2015) (NCBI-nt) to assess the completeness of the resulting

assembled continuous sequences (contigs) by aligning them to available mitochondrial genome sequences (SM 1.1.4). We confirmed that we had assembled a contig with the genes for *tRNA*^{Phe} through *cyt b* to *tRNA*^{Thr} that was approximately 18,000 nucleotides (nt) in length, but lacked the complete control region sequence. We used GNU Grep version 2.16 (Free Software Foundation, 2014) to search the Illumina reads for matches to the assembled sequence of *tRNA*^{Phe} or *tRNA*^{Thr} (SM 1.1.5). We found three reads that spanned *tRNA*^{Phe} and combined them using the Geneious version 9.1.4 *de novo* assembler (Kearse et al., 2012; Biomatters, 2016) (SM 1.1.6). We then extended this assembled contig using a targeted assembly approach with the software PRICE version 1.2 (Ruby, Bellare & DeRisi, 2013; Ruby, 2014) (SM 1.1.7). This PRICE run produced an improved and lengthened assembly after 31 cycles, but the assembly still lacked the complete control region sequence.

We used BLATq version 1.0.2 to align Illumina sequences to the assembly output by PRICE (SM 1.1.8) and extracted aligned reads using excerptByIds version 1.0.2 (SM 1.1.9). We then performed another PRICE assembly with the same initial contig as before, but with the extracted additional Illumina sequence data (SM 1.1.10). This run produced an assembly of one contig of length 18,489 nt after 26 cycles.

We annotated this PRICE assembly using the MITOS WebServer version 605 (Bernt et al., 2013) (SM 1.1.11), which confirmed that this assembly contained the genes for *tRNA*^{Phe} through *cyt b* to *tRNA*^{Thr} followed by control region 1 (CR1), *tRNA*^{Pro}, ND6, *tRNA*^{Glu}, and control region 2 (CR2). We searched for repetitive regions using Tandem Repeats Finder version 4.07b (Benson, 1999, 2012) (SM 1.1.12).

In order to confirm the assemblies of both CR1 and CR2 with longer sequences that could span the repetitive sections of these regions, we designed primers to gene sequences outside of CR1 and CR2 and used Sanger sequencing to obtain verifying sequences across them. We successfully amplified CR2 using a polymerase chain reaction (PCR) with primers 17589F and 41R (Table 2), which primed in *tRNA*^{Glu} and *tRNA*^{Phe}, respectively. We then sequenced both ends of the PCR-amplified fragment using BigDye terminator chemistry (Applied Biosystems, Foster City, California) on an ABI 3130xl automated sequencer (Applied Biosystems, Foster City, California; SM 1.2.1). We also used primer 17572F, which primed in *tRNA*^{Glu}, and primer 41R (Table 2) to successfully PCR-amplify a slightly longer fragment than above, which also included all of CR2, and then sequenced across the repetitive section of CR2 using internal primers 18327F and 19911R (Table 2), which primed outside of the repetitive region (SM 1.2.2).

We edited the sequences using Geneious version 9.1.4 (Kearse et al., 2012; Biomatters, 2016) and then used the Geneious mapper to align the sequences to the 19,946 nt preliminary mitochondrial genome assembly (SM 1.2.3). These Sanger-derived sequences confirmed that there were nine complete repetitions of a 78 nt motif in CR2 and extended the assembly length to 19,948 nt.

Similarly, we confirmed the CR1 sequence with Sanger-derived sequence data by first PCR-amplifying CR1 with primers cytb-F1 and 17122R (Table 2), which primed in *cyt b* and *ND6*, respectively (SM 1.2.4). We visualized the PCR products on a 1% agarose gel, which revealed two PCR products approximately 2,250 and 3,500 nt in length. We re-ran the PCR and gel visualization to confirm this result, which was consistent. We then excised each band from a 1% low melting point agarose gel, performed gel purification using a Zymoclean Gel DNA Recovery Kit (Zymo Research, Irvine, California), and sequenced the purified fragments using the original external primers as well as the internal primers CR1-F1, CR1-F1-RC, CR1-R2, CR1-R2-RC, and N1 (Barrowclough, Gutierrez & Groth, 1999) (Table 2) with BigDye terminator

chemistry on an ABI 3130xl automated sequencer. We edited the sequences using Geneious version 9.1.4 and then used the Geneious *de novo* assembler and mapper to assemble the sequences and then align them to the 19,948 nt preliminary mitochondrial genome assembly. We were able to assemble the entirety of the smaller PCR product, but we were unable to completely assemble the CR1 repetitive region in the larger PCR product. Thus, our mitochondrial genome assembly contains the CR1 sequence obtained from the smaller PCR product. The assembly length was then 19,889 nt as the Sanger-confirmed CR1 sequence contained a shorter repetitive region than we assembled with the shorter Illumina sequences. The length of the CR1 repetitive region in the Illumina-sequence-only assembly was also different from the length we expected in the larger PCR product.

In order to use all of the available Illumina sequence data to verify our mitochondrial genome assembly, we took the draft whole genome assembly of *S. o. caurina* (Hanna et al., 2017) and replaced scaffold-3674, which was the incomplete assembly of the mitochondrial genome, with the 19,889 nt mitochondrial genome assembly from our targeted assembly methodology (SM 1.3.1).

We aligned all filtered Illumina sequences to this new draft reference genome using bwa version 0.7.13-r1126 (Li, 2013a) and then merged, sorted, and marked duplicate reads using Picard version 2.2.4 (http://broadinstitute.github.io/picard) (SM 1.3.2). We filtered the alignment file to only retain alignments to the preliminary targeted mitochondrial genome assembly using Samtools version 1.3 with HTSlib 1.3.1 (Li et al., 2009, 2015). We then used Samtools and GNU Awk (GAWK) version 4.0.1 (Free Software Foundation, 2012) to filter the alignments (SM 1.3.3-1.3.4). We next visualized the alignment across the reference sequence in Geneious version 9.1.4 to confirm that the sequence evidence matched our assembly (SM 1.3.5).

We annotated the final assembly using the MITOS WebServer version 806 (Bernt et al., 2013) (SM 1.4.1) followed by manual inspection of the coding loci and comparison with predicted open reading frames and sequences from *Gallus gallus* [GenBank Accessions NC_001323 (Desjardins & Morais, 1990) and AB086102.1 (Wada et al., 2004)] in Geneious version 9.1.4. We annotated the repetitive regions using the web version of Tandem Repeats Finder version 4.09 (Benson, 1999, 2016) (SM 1.4.2). We used bioawk version 1.0 (Li, 2013b) and GAWK version 4.0.1 to find goose hairpin sequences in CR1 and CR2 (SM 1.4.3). We compared the sequences of the annotated genes in our final mitochondrial genome assembly with those of the incomplete mitochondrial genome assembly that was output as a byproduct of the *S. o. caurina* whole nuclear genome assembly (Hanna et al., 2017) in order to evaluate the efficacy of the nuclear genome assembler in assembling mitochondrial genes. We aligned all of the nucleotide sequences of the genes in the final mitochondrial genome against a database of the scaffold-3674 gene nucleotide sequences using NCBI BLAST+ version 2.4.0 tool BLASTN (Altschul et al., 1990; Zhang et al., 2000; Morgulis et al., 2008; Camacho et al., 2009) (SM 1.4.4).

In order to visualize the binding sites of the primers that we developed to PCR-amplify CR1 and CR2 as well as the primers used by Barrowclough, Gutierrez & Groth (1999) to PCR-amplify a portion of CR1 we used Geneious version 9.1.4 (SM 1.4.5). We assessed the similarity of CR1 and CR2 by performing a multiple alignment using the Geneious version 9.1.4 implementation of MUSCLE version 3.8.425 (Edgar, 2004) (SM 1.4.6). In order to assess whether published control region sequences of related species are more similar to CR1 or CR2, we used the web version of NCBI's BLAST+ version 2.5.0 tool BLASTN (Altschul et al., 1990; Zhang et al., 2000; Morgulis et al., 2008; Camacho et al., 2009) to search NCBI-nt for sequences

similar to CR1 and CR2 (SM 1.4.7). As a result of these searches, we aligned the primers used by Omote et al. (2013) to PCR-amplify the control region in *Strix uralensis* in their study to our final *S. o. caurina* assembly using Geneious version 9.1.4 (SM 1.4.8).

Nuclear pseudogenation of Strix occidentalis mitochondrial genes

In order to examine the incidence of genetic transfer from mitochondria to the nucleus, we examined the draft nuclear genome assembly for evidence of nuclear pseudogenes or nuclear copies of mitochondrial genes (*Numts*) (Lopez et al., 1994), in the *S. o. caurina* draft nuclear genome assembly (Hanna et al., 2017). We aligned the final *S. o. caurina* mitochondrial genome assembly to the draft nuclear genome assembly using the NCBI BLAST+ version 2.4.0 tool BLASTN (SM 1.5.1) using the default threshold Expect value (E-value) of 10. We then used GAWK version 4.0.1 to remove all alignments to scaffold-3674, which was the assembly of the mitochondrial genome in the draft nuclear genome assembly. We visually inspected the results to insure that all alignments were of reasonable length and that all E-values were < 0.0001 (De Wit et al., 2012). Indeed, all alignments exceeded 100 nt and all E-values were $< 1 \times 10^{-25}$. We next used GAWK version 4.0.1 to reformat the BLAST output into a Browser Extensible Data (BED) formatted file (SM 1.5.3). In order to determine the mitochondrial genes spanned by each *Numt*, we used BEDTools version 2.26.0 (Quinlan & Hall, 2010) to produce a BED file of the intersection of the BED-formatted BLAST output with the BED file output from the MITOS annotation of the final mitochondrial genome assembly (SM 1.5.4).

Strix varia mitochondrial genome assembly

In order to assess the divergence between *S. occidentalis* and *S. varia* across all genes of the mitochondrial genome, we constructed a complete *S. varia* mitochondrial genome assembly. We did this by utilizing available whole-genome Illumina data from two *S. varia* individuals collected outside of the zone of contact of *S. varia* and *S. o. caurina* (Haig et al., 2004). The main set of *S. varia* whole-genome Illumina data originated from sequencing of a tissue sample collected in Hamilton County, Ohio, United States of America (CNHM<USA-OH>:ORNITH:B41533; Table 1), hereafter "CMCB41533". The paired-end Illumina sequence data was from a genomic library constructed, sequenced, and the data processed as described in Hanna et al. (2017). The raw sequences are available from NCBI (SRA run accessions SRR5428115, SRR5428116, and SRR5428117).

The second *S. varia* individual was from Marion County, Indiana, United States of America (CAS:ORN:95964; Table 1), hereafter "CAS95964". Sequence data from this individual informed the assembly process, but none of these data are included in the final *S. varia* mitochondrial genome assembly (SM 1.6.1). The raw sequences are available from NCBI (SRA run accession SRR6026668). We performed adapter and quality trimming of these sequence data using Trimmomatic version 0.30 (Bolger, Lohse & Usadel, 2014) (SM 1.6.2). For use in only the SOAPdenovo2 assembly, we trimmed the sequences using a different set of parameters and performed error-correction of the sequences using SOAPec version 2.01 (Luo et al., 2012) (SM 1.6.3).

We constructed the complete *S. varia* mitochondrial genome of sample CMCB41533 by building a succession of assemblies that contributed information to the final assembly from which we extracted the gene sequences. We used partial mitochondrial assemblies of sample CAS95964 to inform the assembly process, but, as we had more sequence data for sample CMCB41533, we chose to only produce a final genome assembly for this sample to compare with that of *S. o. caurina*.

We used two contigs (ContigInput1 and ContigInput2) as the starting material for our final CMCB41533 *S. varia* assembly. In order to generate ContigInput1, we used bwa version 0.7.13-r1126 to align all of the trimmed CMCB41533 paired read 1 and 2 sequences to a reference sequence that included the draft *S. o. caurina* whole nuclear genome along with our final mitochondrial genome assembly (SM 1.9.1). We then merged the paired-end and unpaired read alignments, sorted the reads, and marked duplicate reads using Picard version 2.2.4 (SM 1.9.2).

We filtered the alignment file to only retain alignments to the final mitochondrial genome assembly using Samtools version 1.3 with HTSlib 1.3.1 (Li et al., 2009, 2015). We then used Samtools and GAWK version 4.0.1 to filter out duplicate reads, low quality alignments, secondary alignments, and alignments where both reads of a pair did not align to the mitochondrial assembly (SM 1.9.2-1.9.3). We next visualized the alignment across the reference sequence in Geneious version 9.1.4 and generated a consensus sequence from the alignment (SM 1.9.4). We extracted three sequences from this consensus sequence based on the *S. o. caurina* mitochondrial genome annotations and then used these extracted sequences as three separate seed contigs in an assembly using PRICE version 1.2 (SM 1.9.5). This run produced one contig (ContigInput1) of length 9,690 nt after 16 cycles.

The series of assemblies that resulted in ContigInput2, an input to our final *S. varia* assembly, involved first using SOAPdenovo2 version 2.04 to assemble all of the trimmed, error-corrected CAS95964 sequences (SM 1.10.1). We extended the output 15,019 nt contig using PRICE version 1.2 (SM 1.10.2). After seven cycles, this run produced an assembly of one contig of length 16,652 nt, which included the sequence for $tRNA^{Phe}$ through $tRNA^{Thr}$ and part of CR1. We used this CAS95964 contig to seed a more complete assembly using PRICE version 1.2 with the larger CMCB41533 Illumina sequence dataset (SM 1.11.1). After four cycles, this assembly produced one contig of length 17,073 nt, which we will refer to as "ContigInput2" below.

We performed a final assembly using PRICE version 1.2 and the 9,690 nt ContigInput1 and the 17,073 nt ContigInput2 as the initial contigs (SM 1.12.1). After two cycles, this assembly produced one contig of length 19,589 nt. We then used Sanger sequencing to confirm the sequences of CR1 and CR2.

We PCR-amplified CR1 with primers cytb-F1 and 17122R (Table 2), which primed in *cyt b* and *ND6*, respectively (SM 1.12.2). We then sequenced the fragment using the original external primers as well as the internal primers CR1-F1, CR1-F1-RC, CR1-R2, CR1-R2-RC, and N1 (Barrowclough, Gutierrez & Groth, 1999) (Table 2). We PCR-amplified CR2 with primers ND6-ext1F and 12S-ext1R (Table 2), which primed in *ND6* and *12S*, respectively (SM 1.12.3). We then sequenced the PCR-amplified fragment using the original external primers as well as the internal primers final-CR2F, 18547F, 19088R, and 19088R-RC. We performed all sequencing using BigDye terminator chemistry (Applied Biosystems, Foster City, California) on an ABI 3130xl automated sequencer (Applied Biosystems, Foster City, California).

We edited the sequences using Geneious version 9.1.4 and then used the Geneious *de novo* assembler and mapper to assemble the sequences and then align them to the 19,589 nt preliminary mitochondrial genome assembly. These Sanger-derived sequences confirmed that the preliminary assembly was inaccurate in the control regions and reduced the total length to a final size of 18,975 nt. We annotated the assembly using the MITOS WebServer version 605 (SM 1.12.4) followed by manual inspection of the coding loci and comparison with predicted open reading frames and sequences from *Gallus gallus* [GenBank Accessions NC_001323 (Desjardins & Morais, 1990) and AB086102.1 (Wada et al., 2004)] in Geneious version 9.1.4.

We annotated the repetitive regions using the web version of Tandem Repeats Finder version 4.09 (Benson, 1999, 2016) (SM 1.4.2). We used bioawk version 1.0 (Li, 2013b) and GAWK version 4.0.1 to find goose hairpin sequences in CR1 and CR2 (SM 1.4.3).

Comparison of Strix occidentalis and Strix varia mitochondrial genes

In order to compare mitochondrial gene sequences of *S. occidentalis* and *S. varia*, we extracted the nucleotide sequence for all non-tRNA genes (stop codons excluded) from our final *S. o. caurina* and *S. varia* assemblies. We aligned them using MAFFT version 7.305b (Katoh et al., 2002; Katoh & Standley, 2013; Katoh, 2016) (SM 1.13.1). We verified the alignments by eye and then used trimAl version 1.4.rev15 (Capella-Gutiérrez, Silla-Martínez & Gabaldón, 2009; Capella-Gutiérrez & Gabaldón, 2013) to convert the alignments to MEGA format (Kumar, Tamura & Nei, 1994; Kumar, Stecher & Tamura, 2016) (SM 1.13.2). We then used MEGA version 7.0.18 (Kumar, Stecher & Tamura, 2016) to calculate the p-distance (SM 1.13.3) and the corrected pairwise distance (Tamura & Nei, 1993) (SM 1.13.4) between *S. o. caurina* and *S. varia* for each gene. We calculated a weighted average pairwise distance across all of the genes (SM 1.13.5).

Avian mitochondrial gene order comparisons

We downloaded the mitochondrial genome sequences of *Gallus gallus* (GenBank Accession NC_001323.1) (Desjardins & Morais, 1990), *Melopsittacus undulatus* (GenBank Accession NC_009134.1) (Guan, Xu & Smith, 2016), *Falco peregrinus* (GenBank Accession NC_000878.1) (Mindell et al., 1997; Mindell, Sorenson & Dimcheff, 1998a; Mindell et al., 1999), *Bubo bubo* (GenBank Accession AB918148.1) (Hengjiu et al., 2016), *Ninox novaeseelandiae* (GenBank Accession AY309457.1) (Harrison et al., 2004), *Tyto alba* (GenBank Accession EU410491.1) (Pratt et al., 2009), *Strix leptogrammica* (GenBank Accession KC953095.1) (Liu, Zhou & Gu, 2014), *Glaucidium brodiei* (GenBank Accession KP684122.1) (Sun et al., 2016), and *Asio flammeus* (GenBank Accession KP889214.1) (Zhang et al., 2016), which were all submitted as complete genomes apart from *Tyto alba*, which was submitted as a partial genome. The *Gallus gallus* mitochondrion represented the ancestral avian order (Desjardins & Morais, 1990; Mindell, Sorenson & Dimcheff, 1998a; Haddrath & Baker, 2001; Gibb et al., 2007). The mitochondrial gene order of *Falco peregrinus* was illustrative of the novel gene order first described by Mindell, Sorenson & Dimcheff (1998a) with a remnant CR2 (Gibb et al., 2007) while the mitochondrial gene order of *Melopsittacus undulatus* exemplified an intact, duplicated control region first described in Psittaciformes by Eberhard, Wright & Bermingham (2001). We visualized the mitochondrial genome sequences and the accompanying annotations using Geneious version 9.1.4. For a coarse assessment of gene similarity, we next used the Geneious version 9.1.4 implementation of MUSCLE version 3.8.425 in order to align all of the owl (Aves: Strigiformes) mitochondrial genomes as well as to align the *S. leptogrammica* mitochondrial genome with our *S. o. caurina* and *S. varia* assemblies.

Results

We deposited the complete mitochondrial genome sequences of *Strix occidentalis caurina* sample CAS:ORN:98821 and *Strix varia* sample CNHM<USA-OH>:ORNITH:B41533 as NCBI GenBank Accessions MF431746 and MF431745, respectively. The lengths of the final *S. o. caurina* and *S. varia* mitochondrial genome assemblies were 19,889 nt and 18,975 nt, respectively. As for all typical avian mitochondrial genomes, they are circular and code for 2 rRNAs, 22 tRNAs, and 13 polypeptides (Figure SM1 and Figure SM2). The annotations produced by MITOS identified a 1 nt gap that split *ND3*, which is consistent with the

untranslated nucleotide and translational frameshift seen in *ND3* in some other bird species (Mindell, Sorenson & Dimcheff, 1998b), including owls (Strigiformes) (Fuchs et al., 2008).

Both the *S. o. caurina* and *S. varia* mitochondrial genomes contain a duplicated control region (Figure 1). In both genomes, CR1 and CR2 each include a C-rich sequence near the 5' end, the goose hairpin (Quinn & Wilson, 1993), which is identical across the two species and across CR1 and CR2. The *S. o. caurina* CR1 contains a 70 nt motif repeated 6.8 times near the 3' end while CR2 includes two sets of tandem repeats near the 3' end of the region, a 70 nt motif repeated 4 times followed by 9.5 repetitions of a 78 nt motif (Table 3).

The *S. o. caurina* CR1 and CR2 share a conserved central block of 1,222 nt with only two mismatches between CR1 and CR2 (Figure 2). This conserved block includes 202 nt of the 5' portion of the repetitive regions. The *S. varia* CR1 and CR2 share a conserved 1,041 nt central sequence stretch containing five mismatches. In CR1, this conserved block begins in the 3' 57 nt of the CR1 repetitive region, but in CR2 it does not extend into the repetitive region. The 5' and 3' regions surrounding the conserved central blocks of the control regions in both *S. o. caurina* and *S. varia* are more divergent from each other.

We obtained an alignment (88.37% identity) of 1,429 nt from the 5' ends of the *S. o. caurina* and *S. varia* CR1 sequences, but it included fifteen gaps (Figure 3). In contrast, the more 3' repetitive sections of the *S. o. caurina* and *S. varia* CR1 sequences yielded an uninformative alignment with numerous, long gap regions. Similarly to CR1, the 5' ends of the *S. o. caurina* and *S. varia* CR2 sequences aligned well (90.62% identity), yielding a 1,300 nt alignment that included four gaps. However, the alignment of the 3' region of the CR2 sequences was uninformative with numerous, long gaps due to conflicts between the 78 nt motif repetitive regions of the two CR2 sequences. We found no evidence of mitochondrial pseudogenes in the control region sequences of either *S. o. caurina* or *S. varia*.

Across all of the 35 genes that were present in the previous, incomplete *S. o. caurina* assembly that was produced as a byproduct of the assembly of the *S. o. caurina* whole nuclear genome (Hanna et al., 2017), we only found one mismatch with our complete assembly, which occurred between the two *ND1* sequences. This assembly improves upon the previous version by providing the complete sequences of *ND6*, $tRNA^{Pro}$, and the two control regions.

The *S. o. caurina* CR1 is 2,021 nt in length and the *S. varia* CR1 is 1,686 nt long. In both species, the 5' end of CR1 borders $tRNA^{Thr}$ and the 3' end is adjacent to $tRNA^{Pro}$, then *ND6*, and then $tRNA^{Glu}$ (Figure 1). The initial 1,104 nt of the *S. o. caurina* CR1 are identical to a *S. o. caurina* partial control region sequence (GenBank Accession AY833630.1) (Barrowclough et al., 2005). All of the top 100 matches of the BLASTN searches of the *S. o. caurina* CR1 to NCBI-nt were to either *S. occidentalis* or *S. varia* control region sequences deposited by other researchers, as we expected.

CR2 follows $tRNA^{Glu}$ and is 2,319 nt in length in *S. o. caurina* and 1,719 nt long in *S. varia*. The initial 549 nt of the *S. o. caurina* CR2 matches the beginning of the D-loop sequence of an annotated complete genome of a *Bubo bubo* mitochondrion (GenBank Accession AB918148.1) (Hengjiu et al., 2016). One of the top 100 matches of the BLASTN searches of the *S. o. caurina* CR1 to NCBI-nt, which had the highest total score (2,177) and query coverage (96% versus 36-41% for the other matches) of the top 100 matches, was to a *S. uralensis* control region sequence (GenBank Accession AB743794.1) (Omote et al., 2013). The majority of the primers used by Omote et al. (2013) to PCR-amplify the control region in *S. uralensis* align within and around the *S. o. caurina* CR2. Four of the control-region-specific primers align to the middle of CR2 in our *S. o. caurina* sequence, which is identical to the middle of the *S. o. caurina*

CR1 sequence. Perhaps most crucially, the primer L16728 aligns in the forward direction in *tRNA*^{Glu} such that it would amplify CR2, if present in the species.

As we mentioned in the methodology, our PCR-amplification of the *S. o. caurina* CR1 using primers that spanned from *cyt b* to *ND6* yielded two products approximately 2,250 and 3,500 nt in length (Figure SM3). The sequences of these fragments were identical in the *cyt b* and *ND6* portions as well as in the adjacent CR1 sections except when they entered the repetitive region at the 3' end of CR1. We were only able to obtain sequence spanning the entirety of this repetitive region in the 2,250 nt fragment. This was largely due to the fact that the 3,500 nt fragment, in addition to the 70 nt motif repetitive section observed in the sequence of the 2,250 nt fragment, contained another repetitive region on the *tRNA*^{Pro} side of the 70 nt motif region with at least 13.1 copies of a 67 nt motif. We did not find any copies of *tRNA*^{Pro} or *ND6* in the *S. o. caurina* nuclear genome, but we did find nuclear copies of *cyt b* and *tRNA*^{Thr}. With *ND6* absent from the nuclear genome, PCR-amplification using primers in *cyt b* and *ND6* should only generate mitochondrial genome fragments. Additionally, the *cyt b* and *tRNA*^{Thr} sequence in both the 2,250 nt and 3,500 nt fragments did not match the nuclear genome copies of these genes. In summary, we believe that both of these fragments were mitochondrial in origin and this evidence suggests that at least two different versions of the mitochondrial genome were present in this *S. o. caurina* individual.

The annotations of the mitochondrial genome sequences of the owls (Aves: Strigiformes) *Tyto alba*, *Ninox novaeseelandiae*, *Strix leptogrammica*, *Glaucidium brodiei*, and *Asio flammeus* indicate that those owls all share the same mitochondrial gene order as *Gallus gallus*, the ancestral avian mitochondrial gene order (Desjardins & Morais, 1990; Mindell, Sorenson & Dimcheff, 1998a; Haddrath & Baker, 2001) (Figure 1). Our alignment of the *S. leptogrammica* mitochondrial genome to the mitochondrial genomes of other owls, including our *S. o. caurina* and *S. varia* assemblies, resulted in a poor, gap-filled alignment of the genes from the second half of the *S. leptogrammica cyt b* sequence through *ND6* to *tRNA*^{Phe}. We could not obtain a reasonable alignment of the last 210 nt of the *S. leptogrammica* D-loop adjacent to the *tRNA*^{Phe} sequence to our *S. o. caurina* and *S. varia* assemblies or to the mitochondrial genomes of any of the other Strigiformes whose sequences we examined. Additionally, alignment of the *S. leptogrammica* mitochondrial genome with our *S. o. caurina* and *S. varia* assemblies yielded an *ND5* alignment with seven gaps and numerous mismatches (85.60% and 84.82% identity to *S. o. caurina* and *S. varia*, respectively). Together, these results suggest that the *S. leptogrammica* sequence potentially contains significant errors in the sequences of the genes from *ND5* through *ND6* to *tRNA*^{Phe}.

We found 29,520 nt of *Numt* sequences in the draft *S. o. caurina* nuclear genome assembly spanning nine *Numts* (Table 4). The *Numts* ranged in length from 226-19,522 nt and had an average length of 3,280 nt. The *Numts* provided evidence of nuclear copies of all mitochondrial genes, except *tRNA*^{Pro}, *ND6*, and *tRNA*^{Glu}, the three genes between CR1 and CR2. *Numt* #9 (Table 4) aligns to both CR1 and CR2 with the alignments extending into the conserved block shared by the control regions. The portion of genome scaffold-294 aligned to CR2 for this *Numt* is 519 nt, whereas the length aligned to CR1 is 592 nt. As we could not be sure of which control region was incorporated into the nuclear genome, we have provided information for both alignments and derived the length of the *Numt* from the alignment to CR1 (Table 4).

Strix occidentalis caurina and *S. varia* display an average of 10.74% (8.68% uncorrected p-distance) base substitutions per site across the 2 rRNA genes and 13 polypeptide genes (the

non-tRNA mitochondrial genes) (Table 5). The lowest number of base substitutions per site occurs within *16S* and the highest within *ATP8* (Table 5).

Discussion

Sequences of most mitochondrial genes can often be recovered from high-throughput short-read sequencing data if genome complexity is not too great. Algorithms using short-read data have more difficulty assembling low-complexity or repetitive regions due to an inability to span these regions. Thus, assembly of complete mitochondrial genome sequences can be more difficult when such genomes include regions of low-complexity. The sequence of the avian control region can both contain blocks of tandem repeats (Omote et al., 2013) and be duplicated (Eberhard, Wright & Bermingham, 2001; Haring et al., 2001). Moreover, the presence of multiple controls regions that are similar or identical, which has been observed in snakes (Kumazawa et al., 1996), can cause problems with assembly. In such situations, additional types of sequencing data that complement short-read data may be necessary in order to obtain an accurate and complete assembly of the mitochondrial genome. This proved to be the case in our study where the longer Sanger-derived sequence data were crucial in obtaining the complete sequence of the lengthy, repeat-rich control regions in *S. o. caurina* and *S. varia*. Although Brito (2005) and Barrowclough et al. (2011) inferred the presence of a duplicated control region structure in the mitochondrial genomes of at least three wood owl species, *Strix aluco*, *S. uralensis*, and *S. varia*, they did not sequence complete mitochondrial genomes. They likely deduced that a duplication was present from the appearance of multiple bands on agarose gels resulting from PCR-amplification of portions of the mitochondrial control region. Here we describe the first complete genome sequences of the mitochondrion of an owl (Aves: Strigiformes) with a duplicate control region.

The mitochondrial genomes of *S. o. caurina* and *S. varia* exhibit the novel avian gene order first described by Mindell, Sorenson & Dimcheff (1998a) for several bird orders, but not reported by them as present in the owl *Otus asio*. As mentioned above, this duplicated control region structure and novel gene order has previously been reported in the mitochondrial genome of *S. varia* (Barrowclough et al., 2011) and the congeners *S. aluco* and *S. uralensis* (Brito, 2005). The novel gene order was previously implied for *S. occidentalis* by the placement of primer N1 in $tRNA^{Thr}$ by Barrowclough, Gutierrez & Groth (1999) to PCR-amplify the control region (CR1) fragment used in their study. Hull et al. (2010) also used the Barrowclough, Gutierrez & Groth (1999) N1 primer to PCR-amplify the control region in their study of *S. nebulosa*, so we can infer that the *S. nebulosa* mitochondrion also possesses the Mindell, Sorenson & Dimcheff (1998a) novel gene order. Notably, this mitochondrial gene order was not reported as present in *S. leptogrammica* (Liu, Zhou & Gu, 2014). However, our alignments of this mitochondrial genome to our *S. o. caurina* and *S. varia* sequences as well as the sequences of other owl mitochondrial genomes indicated problems with the *S. leptogrammica* sequence from *cyt b* through *ND6* to $tRNA^{Phe}$. If we then leave aside the *S. leptogrammica* sequence, available evidence suggests that the novel gene order and duplicate control region structure is present across the genus *Strix*.

The primers developed by Barrowclough, Gutierrez & Groth (1999) to PCR-amplify a fragment of the control region (CR1) in *S. occidentalis* have been used extensively in additional genetic studies of owl species (Haig et al., 2004; Brito, 2005; Marthinsen et al., 2009; Hull et al., 2010; Barrowclough et al., 2011; Hausknecht et al., 2014). The Barrowclough, Gutierrez & Groth (1999) control region primers D16 (the most 3' of their primers) and D20 (more 5' relative

to primer D16) align to a region conserved between CR1 and CR2, although the length of the distance from the binding site of primer N1 in *tRNA^{Thr}* to the CR2 sites of primers D16 and D20 (3,742 nt and 3,392 nt, respectively, in our *S. o. caurina* assembly) likely reduces the probability of this second primer binding site causing problems in the PCR-amplification of the CR1 fragment.

The second control region appears intact, not degraded as found in some other avian taxa (Mindell, Sorenson & Dimcheff, 1998a). This gene order corresponds to the "Type D Duplicate CR gene order" of Gibb et al. (2007) and the "Duplicate CR gene order I" of Eberhard & Wright (2016). The goose-hairpin structure is typically found near the beginning of the control region in avian mitochondria (Marshall & Baker, 1997; Randi & Lucchini, 1998; Bensch & H, 2000) and, in agreement with what we found, it appears in the beginning of the intact, duplicated control region sequences in the genomes of *Amazona* (Eberhard, Wright & Bermingham, 2001) and additional parrot mitochondria (Eberhard & Wright, 2016).

The lengths of the *S. o. caurina* CR1 and CR2 (2,021 nt and 2,319 nt, respectively) and of the *S. varia* CR1 and CR2 (1,686 nt and 1,719 nt , respectively) are all shorter than the length reported for the control regions of some species in the sister genus of owls, *Bubo* (Wink et al., 2009), which have lengths up to approximately 3,800 nt due to tandem repeats in the 3' end of the control region (Omote et al., 2013). Similar tandem repeat blocks occur in the control regions of several other owl species in the family Strigidae (Xiao et al., 2006; Omote et al., 2013). The length of the tandem repeat motif unit is 78 nt in the 3' end of the control region sequences of *Bubo blakistoni*, *Bubo virginianus*, *Strix uralensis* (Omote et al., 2013), and *Strix aluco* (Xiao et al., 2006); 78 nt is also the length of the motif in the longest tandem repeat block in both the *S. o. caurina* and *S. varia* CR2 (Table 3).

As we previously mentioned, both *S. uralensis* and *S. aluco* exhibit a duplicated control region structure in their mitochondrial genomes (Brito, 2005). Neither Omote et al. (2013) nor Xiao et al. (2006) report the presence of a duplicated control region structure in either *S. uralensis* or *S. aluco*, respectively, in their discussions of the repetitive content of the control region sequences of these two species. It is not overtly clear from their methodologies which control region they sequenced. The precise primer combinations used for the PCR-amplification and sequencing of the control region of the *Bubo* species and *S. uralensis* are not provided by Omote et al. (2013), but mapping the primer sequences used by the researchers to our *S. o. caurina* genome suggests that, if the structure of the *S. uralensis* mitochondrial genome shares that of *S. o. caurina*, they likely sequenced CR2 in at least *S. uralensis* and in the *Bubo* species if a CR2 was present. We are unsure how placement of the primers in *cyt b* and *12S*, as reported in the methodology of Xiao et al. (2006) could PCR-amplify a single control region sequence for *S. aluco*, given the duplicated control region structure (Brito, 2005).

The duplicated control region structure is unreported in Strigiformes outside of *Strix*, but we infer that it is also likely present in *Bubo* due to the positioning of primers used for PCR-amplification of the control region in previous studies (Marthinsen et al., 2009; Omote et al., 2013). If also present in *Bubo*, then the duplicate control region structure appears to have arisen in the common ancestor of *Strix* and *Bubo*, but a proper phylogenetic test of this hypothesis with increased taxon sampling is warranted. Further work on the structure of control region sequences in *Strix* and related taxa is needed to elucidate the pattern of evolution of this region across the Strigidae phylogeny.

Although inconclusive and warranting further investigation, our evidence for two versions of the 3' repetitive region of CR1 suggests that mitochondrial heteroplasmy is present in

this *S. o. caurina* individual. Mitochondrial heteroplasmy due to tandem repeat variability in the control region has been shown to occur in other bird species (Berg, Moum & Johansen, 1995; Mundy, Winchell & Woodruff, 1996). Previous work has suggested that the most likely mechanism by which the gain and loss of such tandem repeat elements occurs in the mitochondrial control region is that the repetitive region forms a stable, single-stranded secondary structure and there is slippage during replication (Levinson & Gutman, 1987; Wilkinson & Chapman, 1991; Fumagalli et al., 1996; Faber & Stepien, 1998). Greater numbers of repeats may improve the stability of the secondary structure (Faber & Stepien, 1998). Utilizing sequence from the 3' region of CR1 for population genetic study of *S. o. caurina* is not likely to be useful due to the variability (in terms of the number of copies of the tandem repeat motifs in this region) that is potentially present within a single individual.

The 29,520 nt of *Numt* sequence in the draft *S. o. caurina* nuclear genome assembly is more than triple the 8,869 nt of *Numt* sequence found in a *Gallus gallus* draft nuclear genome assembly (Pereira & Baker, 2004). The 3,280 nt average *Numt* size exceeds the average size in all of the eukaryotic genomes examined by Richly & Leister (2004). There are markedly fewer control region *Numts* in the *S. o. caurina* draft genome assembly than found in a *Gallus gallus* draft genome assembly (Pereira & Baker, 2004). We only found one control region *Numt* (Table 4). Indeed the longest *Numt*, *Numt* #1, extends through almost the entire mitochondrial genome sequence including from $tRNA^{Phe}$ through $tRNA^{Thr}$, immediately adjacent to, but ending at CR1. The percentage identity of the nuclear pseudogenes with the true mitochondrial genes ranges from 77.5-87.81%, so care must be taken to insure that *Numts* are not PCR-amplified in place of mitochondrial gene sequences. As the control region is the mitochondrial segment that has been used most often in studies of the population genetics of *Strix* species (Barrowclough, Gutierrez & Groth, 1999; Haig et al., 2004; Barrowclough et al., 2005, 2011; Brito, 2005; Hull et al., 2010), it is encouraging that only one, short *Numt* included CR1 or CR2 (Table 4). As long as researchers PCR-amplify sequences that span beyond the 592 nt *Numt* #9, they should have confidence in amplifying the true mitochondrial control regions.

The average pairwise sequence divergence between *S. occidentalis* and *S. varia* has been previously reported as 13.9% for a 524 nt section of CR1 (Haig et al., 2004). This exceeds the weighted average of 10.74% (8.68% uncorrected p-distance) that we calculated across the non-tRNA mitochondrial genes (Table 5), which is unsurprising as the control region is known to be rapidly evolving in birds (Quinn & Wilson, 1993). However, the pairwise sequence divergence between *S. occidentalis* and *S. varia* appears higher in *ND3*, *ND4L*, *ND6*, and *ATP8* (Table 5) than in the CR1 portion. Hull et al. (2010) found uncorrected *p*-distances of 13.73-13.93% for *ND2* and 14.58-14.81% for the control region between *S. nebulosa* and *S. occidentalis*. Wink & Heidrich (1999) calculated uncorrected *p*-distances of 8.15-11.72% for *cyt b* sequence pairwise comparisons of six *Strix* species (*S. aluco*, *S. butleri*, *S. nebulosa*, *S. rufipes*, *S. uralensis*, and *S. woodfordii*). Of these six *Strix* species, only *S. aluco* x *S. uralensis* are known to hybridize (McCarthy, 2006), albeit in captivity (Scherzinger, 1982), and their *cyt b* pairwise divergence was 8.15%, the lowest of those calculated by Wink & Heidrich (1999) between *Strix* species. Our *cyt b* uncorrected *p*-distance value between *S. o. caurin*a and *S. varia* was 9.21%, which is also on the lower end of the range of Wink & Heidrich (1999) *Strix* interspecific divergences. Overall, however, high levels of interspecific pairwise divergence of mitochondrial DNA seem to be typical for the genus *Strix*, even for species able to hybridize. We anticipate that these whole mitochondrial genome resources will be useful to those with an interest in developing new mitochondrial markers to study the genetics of *S. o. caurina*, *S. varia*, and related taxa.

Data deposition

Complete mitochondrial genome sequence of *Strix occidentalis caurina* CAS:ORN:98821 deposited as NCBI GenBank Accession MF431746. Complete mitochondrial genome sequence of *Strix varia* CNHM<USA-OH>:ORNITH:B41533 deposited as NCBI GenBank Accession MF431745. The raw sequences for *Strix varia* sample CAS:ORN:95964 are available from NCBI (SRA run accession SRR6026668). Script "BLATq" deposited at Zenodo DOI: 10.5281/zenodo.61136 available at https://github.com/calacademy-research/blatq. Script "excerptByIDs" deposited at Zenodo DOI: 10.5281/zenodo.61134 and available at https://github.com/calacademy-research/excerptByIDs.

Tables

Table 1. *Strix* specimen data.

We here provide further information regarding the datasets that archive the *Strix* specimens to which we refer throughout the manuscript.

Specimen	Data publisher	Date accessed	Link to dataset
CAS:ORN:95964	CAS Ornithology (ORN), California Academy of Sciences, San Francisco, California, United States of America	2016 Aug 15	http://ipt.calacademy.org:8080/ipt/resource.do?r=orn
CAS:ORN:98821	CAS Ornithology (ORN), California Academy of Sciences, San Francisco, California, United States of America	2016 Aug 15	http://ipt.calacademy.org:8080/ipt/resource.do?r=orn
CNHM<USA-OH>:ORNITH:B41533	CMC ORNI-T, Museum of Natural History & Science, Cincinnati Museum Center, Cincinnati, Ohio, United States of America	2017 Sep 3	https://www.idigbio.org/portal/records/57d299f0-2cc7-44f7-aa5f-3c2ea175e757

Table 2. Sequence of primers used in Sanger sequencing of control regions.

These are the sequences of all of the primers that we used to amplify control regions 1 and 2 in order to confirm the final sequence of these regions in the mitochondrial genome assemblies.

Primer name	Relevant region	Species used on	External or Internal	Primer sequence (5' → 3')	Source
cytb-F1	CR1	*S. o. caurina*, *S. varia*	External	ATCCTCATTCTCTTCCCCGT	This study
17122R	CR1	*S. o. caurina*, *S. varia*	External	GGTGGGGGTTATTATTAACTTT	This study
CR1-F1	CR1	*S. o. caurina*, *S. varia*	Internal	CTCSASCAAATCCCAAGTTT	This study
CR1-F1-RC	CR1	*S. o. caurina*, *S. varia*	Internal	AAACTTGGGATTTGSTSGAG	This study
CR1-R2	CR1	*S. o. caurina*, *S. varia*	Internal	GGAGGGCGAGAATAGTTGRT	This study
CR1-R2-RC	CR1	*S. o. caurina*, *S. varia*	Internal	AYCAACTATTCTCGCCCTCC	This study
N1	CR1	*S. o. caurina*	Internal	AACATTGGTCTTGTAAACCAA	Barrowclough et al. 1999
41R	CR2	*S. o. caurina*	External	GCATCTTCAGTGCCATGCTT	This study
17572F	CR2	*S. o. caurina*	External	ATTATCCAAGGTCTGCGGCC	This study
17589F	CR2	*S. o. caurina*	Internal	GCCTGAAAAACCGCCGTTAA	This study
18327F	CR2	*S. o. caurina*	Internal	CACTTTTGCGCCTCTGGTTC	This study
19911R	CR2	*S. o. caurina*	Internal	AGAGAGGCTCTGATTGCTTG	This study
ND6-ext1F	CR2	*S. varia*	External	ACAACCCCATAATAYGGCGA	This study
12S-ext1R	CR2	*S. varia*	External	GGTAGATGGGCATTTACACT	This study
final-CR2F	CR2	*S. varia*	Internal	TCAAACCAAACGATCGAGAA	This study
18547F	CR2	*S. varia*	Internal	CTCACGTGAAATCAGCAACC	This study
19088R	CR2	*S. varia*	Internal	ATTCAACTAAAATTCGTTACAAATCTT	This study
19088R-RC	CR2	*S. varia*	Internal	AAGATTTGTAACGAATTTTAGTTGAAT	This study

Table 3. Tandem Repeat Annotations

This summarizes the repetitive regions of the northern spotted owl (*Strix occidentalis caurina* or *S. o. caurina*) and barred owl (*S. varia*) mitochondrial genomes annotated by Tandem Repeats Finder. "Period size" refers to the size of the repeated motif. "Copy number" refers to the number of copies of the repeat in the region. "Consensus size" is the length of the consensus sequence summarizing all copies of the repeat, which may or may not be different from the period size. "Percent matches" refers to the percentage of nucleotides that match between adjacent copies of the repeat. "Percent indels" refers to the percentage of indels between adjacent copies of the repeat. We present the percent composition of each of the four nucleotides in the repetitive region. We have included the genomic regions that intersect each repetitive span in the "Region" column. "CR1" and "CR2" refer to control region 1 and control region 2, respectively.

Taxon	Coordinates (nt)	Region	Period Size (nt)	Copy Number	Consensus Size (nt)	Percent Matches (%)	Percent Indels (%)	A (%)	C (%)	G (%)	T (%)
S. o. caurina	10,267-10,309	*ND4*	18	2.3	19	84	4	25	46	0	27
S. o. caurina	15,066-15,162	CR1	22	4.3	22	70	7	37	27	6	28
S. o. caurina	15,169-15,311	CR1	67	2.1	67	83	8	40	31	6	21
S. o. caurina	16,243-16,715	CR1	70	6.8	70	98	1	39	21	4	33
S. o. caurina	16,245-16,715	CR1	139	3.4	139	99	0	39	22	4	33
S. o. caurina	16,403-16,515	CR1	37	3.2	37	61	27	40	23	3	32
S. o. caurina	17,679-17,795	CR2	44	2.6	45	87	4	43	30	4	21
S. o. caurina	17,719-17,795	CR2	22	3.5	22	89	0	45	31	3	19
S. o. caurina	18,798-19,076	CR2	70	4.0	70	99	0	39	21	4	34
S. o. caurina	18,800-19,076	CR2	139	2.0	139	100	0	39	21	4	34
S. o. caurina	18,958-19,070	CR2	37	3.2	37	61	27	40	23	3	32
S. o. caurina	19,110-19,853	CR2	78	9.5	78	99	0	41	15	15	27
S. varia	15,126-15,209	CR1	22	3.8	22	82	4	36	27	4	30
S. varia	15,193-15,340	CR1	67	2.2	68	83	1	37	32	8	22
S. varia	17,384-17,482	CR2	22	4.4	23	87	5	41	34	5	19
S. varia	18,548-18,951	CR2	78	5.2	77	93	2	40	17	15	26

Table 4. Mitochondrion-derived nuclear pseudogenes (*Numts*) identified in the *Strix occidentalis caurina* nuclear genome sequence and statistics of the results of BLASTN searches.

We indicate the mitochondrial genes that a *Numt* spans in the "Genes included" column. If a *Numt* spans more than two genes, we indicate the first and last genes that it spans as well as a gene in the middle of the *Numt* in order to indicate the direction that the *Numt* extends. The *Numt* additionally spans all of the intervening genes in such cases. "Start mtDNA" and "End mtDNA" indicate the mitochondrial genome assembly sequence positions and "Start Scaffold" and "End Scaffold" denote the nuclear genome assembly contig/scaffold sequence positions in the alignments of the mitochondrial genome assembly to the nuclear genome assembly. "% ID" indicates the percentage of identical matches in an alignment. "E-value" is the Expect value. "Bit score" is a log-scaled version of the alignment score. We characterized some of the *Numts* by examining more than one alignment and concluding that a *Numt* spanned across those individual alignments.

Numt #	Genes included	Start mtDNA	End mtDNA	Nuclear Genome Scaffold	Start Scaffold	End Scaffold	Orien-tation	% ID	E-value	Bit score	Length alignment (nt)	Length Numt (nt)
1	*$tRNA^{Phe}$ - 12S - 16S*	1	2,225	scaffold478	47,666	49,858	+	79.92	0.0	1,565	2,261	19,522
	16S	2,367	2,645	scaffold478	49,871	50,143	+	87.81	2.16e-84	322	279	-
	16S - ND2 - $tRNA^{Asn}$	2,706	5,223	scaffold478	50,161	52,680	+	80.66	0.0	1,921	2,549	-
	$tRNA^{Asn}$ - COI - $tRNA^{Ser}$2	5,219	6,932	scaffold478	57,635	59,328	+	83.22	0.0	1,552	1,716	-
	$tRNA^{Ser}$2 - $tRNA^{Asp}$ - COII	6,988	7,103	scaffold478	59,382	59,496	+	87.18	1.41e-26	130	117	-
	ATP6 - ND4 - ND5	8,382	13,249	scaffold478	59,498	64,306	+	80.59	0.0	3,672	4,893	-
	cyt b	14,047	14,733	scaffold478	44,785	45,459	+	82.82	1.92e-169	604	687	-
	$tRNA^{Thr}$	14,729	14,878	scaffold478	46,066	46,222	+	82.80	1.09e-27	134	157	-
2	*16S*	1,682	2,603	scaffold215	5,517,239	5,518,161	-	81.97	0.0	773	932	923
3	*$tRNA^{Ser}$2 - ATP8 - ND3 a*	6,989	9,584	scaffold215	5,513,222	5,515,749	-	79.01	0.0	1,690	2615	2,528
4	*16S - $tRNA^{Leu}$2*	2,290	2,788	scaffold632	1,548,886	1,549,372	+	77.50	6.14e-70	274	511	487
5	*ND1 - $tRNA^{Gln}$ - ND2*	2,810	4,646	scaffold167	11,322,764	11,324,590	+	80.54	0.0	1,400	1,840	2,732
	ND2 - $tRNA^{Asn}$ - COI	4,692	5,597	scaffold167	11,324,598	11,325,495	+	83.68	0.0	846	907	-
6	*$tRNA^{Glu}$ - ND2 - COI*	3,851	5,526	scaffold1500	35,914	37,582	-	84.21	0.0	1,620	1,678	1,669
7	*ND2 - $tRNA^{Asn}$ - $tRNA^{Tyr}$*	4,500	5,348	scaffold173	750,945	751,785	-	81.40	0.0	680	855	841
8	*ND5*	12,082	12,310	scaffold143	586,822	587,047	+	81.30	3.83e-42	182	230	226
9	*CR1*	15,026	15,640	scaffold294	2,356,468	2,357,059	-	83.07	9.17e-148	532	620	592
	CR2	17,677	18,195	scaffold294	2,356,468	2,356,986	-	80.87	9.70e-108	399	528	-

Table 5. Divergence of *Strix occidentalis caurina* and *Strix varia* at all protein-coding genes

This provides the number of base substitutions per site for all mitochondrial protein-coding genes and rRNAs between the mitochondrial sequences of *Strix occidentalis occidentalis* and *S. varia*. P-distance refers to an uncorrected pairwise distance while TN93 refers to the pairwise distance corrected by the Tamura-Nei 1993 model (Tamura et al., 1993).

Gene	Number of sites in alignment (nt)	p-distance	Distance with TN93 model
12S	984	5.79%	6.61%
16S	1,589	5.48%	6.14%
ATP6	681	9.10%	11.07%
ATP8	165	14.55%	20.81%
COI	1,548	7.88%	9.31%
COII	681	9.10%	11.23%
COIII	783	7.54%	8.89%
cyt_b	1,140	9.21%	11.35%
ND1	957	10.66%	13.46%
ND2	1,038	9.34%	11.60%
*ND3*_a	174	10.92%	13.96%
*ND3*_b	174	11.49%	14.86%
ND4	1,377	10.31%	13.12%
ND4L	294	11.22%	14.29%
ND5	1,818	9.19%	11.29%
ND6	516	9.69%	14.71%

Figures

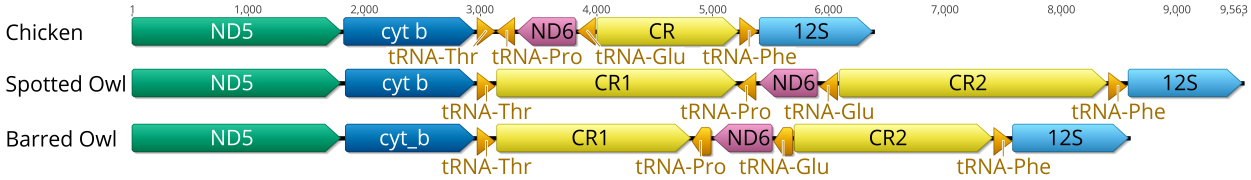

Figure 1. Ancestral avian mitochondrial gene order surrounding the control region compared with that of *Strix occidentalis caurina* and *Strix varia*.

The Chicken panel displays the gene order of *Gallus gallus*, which is the presumed ancestral avian gene order. The Spotted Owl panel depicts the gene order of *Strix occidentalis caurina* and the Barred Owl panel depicts the gene order of *Strix varia*. All rRNAs, tRNAs, and protein-coding genes outside of the displayed region exhibit the same order in all of these mitochondrial genomes. "CR" denotes the control region with "CR1" and "CR2" referring to control regions 1 and 2, respectively. We added 100 nucleotides to each of the tRNAs to improve visualization. Apart from the tRNAs, the annotations are to scale relative to each other with the numbers at the top of the figure denoting nucleotides. The order of the genes outside of the region depicted in this figure is the same in the chicken, spotted owl, and barred owl.

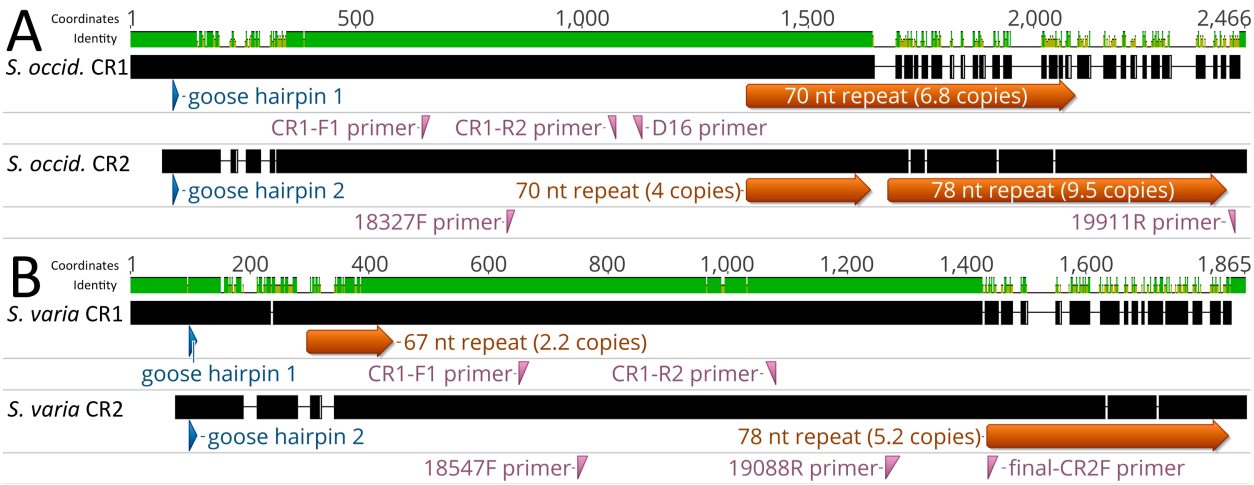

Figure 2. Alignment of control regions 1 and 2 within *Strix occidentalis caurina* and *Strix varia*. Panel A depicts an alignment of the *Strix occidentalis caurina* control regions 1 and 2. Panel B displays an alignment of the *Strix varia* control regions 1 and 2. The numerical coordinates at the top of each panel correspond to the coordinates of the alignment. Black rectangles for each control region denote continuous sequence, whereas intervening horizontal lines denote gaps in the alignment. The sequence identity rectangle is green at full height when there is agreement between the sequences, yellow at less than full height when the sequences disagree, and flat in gap regions. The location of the goose hairpin sequence in each control region is annotated in blue. The alignment locations of the primers we developed to amplify control regions 1 and 2 as well as the D16 primer used by Barrowclough et al. (1999) to amplify a portion of control region 1 are annotated in reddish purple.

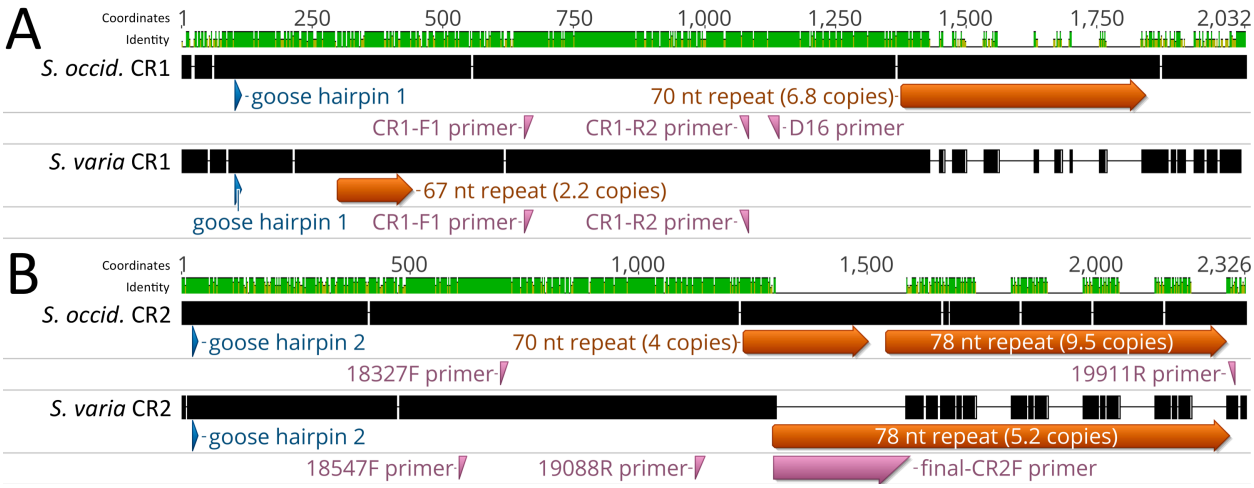

Figure 3. Alignment of *Strix occidentalis caurina* control regions 1 and 2 with those of *Strix varia*.

Panel A depicts an alignment of the *Strix occidentalis caurina* control region 1 with that of *Strix varia*. Panel B displays an alignment of the *Strix occidentalis caurina* control region 2 with that of *Strix varia*. The numerical coordinates at the top of each panel correspond to the coordinates of the alignment. Black rectangles for each control region denote continuous sequence, whereas intervening horizontal lines denote gaps in the alignment. The sequence identity rectangle is green at full height when there is agreement between the sequences, yellow at less than full height when the sequences disagree, and flat in gap regions. The location of the goose hairpin sequence in each control region is annotated in blue. The alignment locations of the primers we developed to amplify control regions 1 and 2 as well as the D16 primer used by Barrowclough et al. (1999) to amplify a portion of control region 1 are annotated in reddish purple. The annotation of primer final-CR2F is elongated as it is situated across a gap region in the alignment.

References

Altschul SF., Gish W., Miller W., Myers EW., Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215:403–410. DOI: 10.1016/S0022-2836(05)80360-2.

Barrowclough GF., Groth JG., Mertz LA., Gutiérrez RJ. 2005. Genetic structure, introgression, and a narrow hybrid zone between northern and California spotted owls (*Strix occidentalis*). *Molecular Ecology* 14:1109–1120. DOI: 10.1111/j.1365-294X.2005.02465.x.

Barrowclough GF., Groth JG., Odom KJ., Lai JE. 2011. Phylogeography of the Barred Owl (*Strix varia*): Species Limits, Multiple Refugia, and Range Expansion. *The Auk* 128:696–706. DOI: 10.1525/auk.2011.11057.

Barrowclough GF., Gutierrez RJ., Groth JG. 1999. Phylogeography of Spotted Owl (*Strix occidentalis*) Populations Based on Mitochondrial DNA Sequences: Gene Flow, Genetic Structure, and a Novel Biogeographic Pattern. *Evolution* 53:919–931. DOI: 10.2307/2640731.

Bensch S., H A. 2000. Mitochondrial Genomic Rearrangements in Songbirds. *Molecular Biology and Evolution* 17:107–113.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research* 27:573–580. DOI: 10.1093/nar/27.2.573.

Benson G. 2012. Tandem Repeats Finder. Version 4.07b. [Accessed 2016 Oct 1]. Available from: https://tandem.bu.edu/trf/trf.html.

Benson G. 2016. Tandem Repeats Finder. Version 4.09. [Accessed 2016 Oct 1]. Available from: https://tandem.bu.edu/trf/trf.html.

Benson DA., Clark K., Karsch-Mizrachi I., Lipman DJ., Ostell J., Sayers EW. 2015. GenBank. *Nucleic Acids Research* 43:D30–D35. DOI: 10.1093/nar/gku1216.

Berg T., Moum T., Johansen S. 1995. Variable numbers of simple tandem repeats make birds of the order Ciconiiformes heteroplasmic in their mitochondrial genomes. *Current Genetics* 27:257–262. DOI: 10.1007/BF00326158.

Bernt M., Donath A., Jühling F., Externbrink F., Florentz C., Fritzsch G., Pütz J., Middendorf M., Stadler PF. 2013. MITOS: Improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution* 69:313–319. DOI: 10.1016/j.ympev.2012.08.023.

Biomatters 2016. Geneious. Version 9.1.4. [Accessed 2016 Oct 1]. Available from: http://www.geneious.com.

Bolger AM., Lohse M., Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. DOI: 10.1093/bioinformatics/btu170.

Boratyn GM., Camacho C., Cooper PS., Coulouris G., Fong A., Ma N., Madden TL., Matten WT., McGinnis SD., Merezhuk Y., Raytselis Y., Sayers EW., Tao T., Ye J., Zaretskaya I. 2013. BLAST: a more efficient report with usability improvements. *Nucleic Acids Research* 41:W29–W33. DOI: 10.1093/nar/gkt282.

Brito PH. 2005. The influence of Pleistocene glacial refugia on tawny owl genetic diversity and phylogeography in western Europe. *Molecular Ecology* 14:3077–3094. DOI: 10.1111/j.1365-294X.2005.02663.x.

Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. DOI: 10.1186/1471-2105-10-421.

Capella-Gutiérrez S., Gabaldón T. 2013. trimAl. Version 1.4.rev15. [Accessed 2016 Oct 1]. Available from: https://github.com/scapella/trimal.
Capella-Gutiérrez S., Silla-Martínez JM., Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973. DOI: 10.1093/bioinformatics/btp348.
De Wit P., Pespeni MH., Ladner JT., Barshis DJ., Seneca F., Jaris H., Therkildsen NO., Morikawa M., Palumbi SR. 2012. The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. *Molecular Ecology Resources* 12:1058–1067. DOI: 10.1111/1755-0998.12003.
Desjardins P., Morais R. 1990. Sequence and gene organization of the chicken mitochondrial genome. *Journal of Molecular Biology* 212:599–634. DOI: 10.1016/0022-2836(90)90225-B.
Eberhard JR., Wright TF. 2016. Rearrangement and evolution of mitochondrial genomes in parrots. *Molecular Phylogenetics and Evolution* 94, Part A:34–46. DOI: 10.1016/j.ympev.2015.08.011.
Eberhard JR., Wright TF., Bermingham E. 2001. Duplication and Concerted Evolution of the Mitochondrial Control Region in the Parrot Genus *Amazona*. *Molecular Biology and Evolution* 18:1330–1342.
Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32:1792–1797. DOI: 10.1093/nar/gkh340.
Faber JE., Stepien CA. 1998. Tandemly Repeated Sequences in the Mitochondrial DNA Control Region and Phylogeography of the Pike-Perches *Stizostedion*. *Molecular Phylogenetics and Evolution* 10:310–322. DOI: 10.1006/mpev.1998.0530.
Free Software Foundation 2012. GNU Awk . Version 4.0.1. [Accessed 2016 Oct 1]. Available from: https://www.gnu.org/software/gawk.
Free Software Foundation 2014. GNU Grep. Version 2.16. [Accessed 2016 Oct 1]. Available from: https://www.gnu.org/software/grep.
Fuchs J., Pons J-M., Goodman SM., Bretagnolle V., Melo M., Bowie RC., Currie D., Safford R., Virani MZ., Thomsett S., Hija A., Cruaud C., Pasquet E. 2008. Tracing the colonization history of the Indian Ocean scops-owls (Strigiformes: *Otus*) with further insight into the spatio-temporal origin of the Malagasy avifauna. *BMC Evolutionary Biology* 8:197. DOI: 10.1186/1471-2148-8-197.
Fumagalli L., Taberlet P., Favre L., Hausser J. 1996. Origin and evolution of homologous repeated sequences in the mitochondrial DNA control region of shrews. *Molecular Biology and Evolution* 13:31–46.
Funk WC., Mullins TD., Forsman ED., Haig SM. 2007. Microsatellite loci for distinguishing spotted owls (*Strix occidentalis*), barred owls (*Strix varia*), and their hybrids. *Molecular Ecology Notes* 7:284–286. DOI: 10.1111/j.1471-8286.2006.01581.x.
Gibb GC., Kardailsky O., Kimball RT., Braun EL., Penny D. 2007. Mitochondrial Genomes and Avian Phylogeny: Complex Characters and Resolvability without Explosive Radiations. *Molecular Biology and Evolution* 24:269–280. DOI: 10.1093/molbev/msl158.
Guan X., Xu J., Smith EJ. 2016. The complete mitochondrial genome sequence of the budgerigar, *Melopsittacus undulatus*. *Mitochondrial DNA Part A* 27:401–402. DOI: 10.3109/19401736.2014.898277.
Haddrath O., Baker AJ. 2001. Complete mitochondrial DNA geonome sequences of extinct birds: ratite phylogenetics and the vicariance biogeography hypothesis. *Proceedings of*

the Royal Society of London B: Biological Sciences 268:939–945. DOI: 10.1098/rspb.2001.1587.
Haig SM., Mullins TD., Forsman ED., Trail PW., Wennerberg L. 2004. Genetic identification of spotted owls, barred owls, and their hybrids: legal implications of hybrid identity. *Conservation Biology* 18:1347–1357.
Hanna ZR., Henderson JB., Wall JD., Emerling CA., Fuchs J., Runckel C., Mindell DP., Bowie RCK., DeRisi JL., Dumbacher JP. 2017. Northern Spotted Owl (*Strix occidentalis caurina*) Genome: Divergence with the Barred Owl (*Strix varia*) and Characterization of Light-Associated Genes. *Genome Biology and Evolution* 9:2522–2545. DOI: 10.1093/gbe/evx158.
Haring E., Kruckenhauser L., Gamauf A., Riesing MJ., Pinsker W. 2001. The Complete Sequence of the Mitochondrial Genome of *Buteo buteo* (Aves, Accipitridae) Indicates an Early Split in the Phylogeny of Raptors. *Molecular Biology and Evolution* 18:1892–1904.
Harrison GL (Abby)., McLenachan PA., Phillips MJ., Slack KE., Cooper A., Penny D. 2004. Four New Avian Mitochondrial Genomes Help Get to Basic Evolutionary Questions in the Late Cretaceous. *Molecular Biology and Evolution* 21:974–983. DOI: 10.1093/molbev/msh065.
Hausknecht R., Jacobs S., Müller J., Zink R., Frey H., Solheim R., Vrezec A., Kristin A., Mihok J., Kergalve I., Saurola P., Kuehn R. 2014. Phylogeographic analysis and genetic cluster recognition for the conservation of Ural Owls (*Strix uralensis*) in Europe. *Journal of Ornithology* 155:121–134. DOI: 10.1007/s10336-013-0994-8.
Henderson JB., Hanna ZR. 2016a. BLATq. Version 1.0.2. *Zenodo*. DOI: 10.5281/zenodo.61136.
Henderson JB., Hanna ZR. 2016b. excerptByIDs. Version 1.0.2. *Zenodo*. DOI: 10.5281/zenodo.61134.
Hengjiu T., Jianwei J., Shi Y., Zhiming Z., Laghari MY., Narejo NT., Lashari P. 2016. Complete mitochondrial genome of Eagle Owl (*Bubo bubo*, Strigiformes; Strigidae) from China. *Mitochondrial DNA Part A* 27:1455–1456. DOI: 10.3109/19401736.2014.953090.
Hull JM., Englis A., Medley JR., Jepsen EP., Duncan JR., Ernest HB., Keane JJ. 2014. A New Subspecies of Great Gray Owl (*Strix nebulosa*) in the Sierra Nevada of California, U.S.A. *Journal of Raptor Research* 48:68–77. DOI: 10.3356/JRR-13-35.1.
Hull JM., Keane JJ., Savage WK., Godwin SA., Shafer JA., Jepsen EP., Gerhardt R., Stermer C., Ernest HB. 2010. Range-wide genetic differentiation among North American great gray owls (*Strix nebulosa*) reveals a distinct lineage restricted to the Sierra Nevada, California. *Molecular Phylogenetics and Evolution* 56:212–221. DOI: 10.1016/j.ympev.2010.02.027.
Johnson M., Zaretskaya I., Raytselis Y., Merezhuk Y., McGinnis S., Madden TL. 2008. NCBI BLAST: a better web interface. *Nucleic Acids Research* 36:W5–W9. DOI: 10.1093/nar/gkn201.
Katoh K. 2016. MAFFT: a multiple sequence alignment program. Version 7.305b. [Accessed 2016 Oct 1]. Available from: http://mafft.cbrc.jp/alignment/software.
Katoh K., Misawa K., Kuma K., Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* 30:3059–3066. DOI: 10.1093/nar/gkf436.
Katoh K., Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* 30:772–780. DOI: 10.1093/molbev/mst010.

Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A., Markowitz S., Duran C., Thierer T., Ashton B., Meintjes P., Drummond A. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649. DOI: 10.1093/bioinformatics/bts199.

Kelly EG., Forsman ED. 2004. Recent Records of Hybridization Between Barred Owls (*Strix varia*) and Northern Spotted Owls (*S. occidentalis caurina*). *The Auk* 121:806–810. DOI: 10.1642/0004-8038(2004)121[0806:RROHBB]2.0.CO;2.

Kent WJ. 2002. BLAT—The BLAST-Like Alignment Tool. *Genome Research* 12:656–664. DOI: 10.1101/gr.229202.

Kent WJ. 2012. BLAT—The BLAST-Like Alignment Tool. Version 35. [Accessed 2016 Oct 1]. Available from: https://users.soe.ucsc.edu/~kent.

Krosby M., Rohwer S. 2009. A 2000 km genetic wake yields evidence for northern glacial refugia and hybrid zone movement in a pair of songbirds. *Proceedings of the Royal Society of London B: Biological Sciences* 276:615–621. DOI: 10.1098/rspb.2008.1310.

Kumar S., Stecher G., Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* 33:1870–1874. DOI: 10.1093/molbev/msw054.

Kumar S., Tamura K., Nei M. 1994. MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. *Computer applications in the biosciences : CABIOS* 10:189–191. DOI: 10.1093/bioinformatics/10.2.189.

Kumazawa Y., Ota H., Nishida M., Ozawa T. 1996. Gene rearrangements in snake mitochondrial genomes: highly concerted evolution of control-region-like sequences duplicated and inserted into a tRNA gene cluster. *Molecular Biology and Evolution* 13:1242–1254. DOI: 10.1093/oxfordjournals.molbev.a025690.

Levinson G., Gutman GA. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution* 4:203–221.

Li H. 2013a. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv:1303.3997 Q-Bio. [Accessed 2016 Feb 16]. Available from: http://arxiv.org/abs/1303.3997.

Li H. 2013b. bioawk. Version 1.0. [Accessed 2016 Oct 1]. Available from: https://github.com/lh3/bioawk.

Li H., Handsaker B., Marshall J., Danecek P. 2015. Samtools. Version 1.3 with HTSlib 1.3.1. [Accessed 2016 Oct 1]. Available from: http://www.htslib.org.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R., Subgroup 1000 Genome Project Data Processing. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. DOI: 10.1093/bioinformatics/btp352.

Liu G., Zhou L., Gu C. 2014. The complete mitochondrial genome of Brown wood owl *Strix leptogrammica* (Strigiformes: Strigidae). *Mitochondrial DNA* 25:370–371. DOI: 10.3109/19401736.2013.803540.

Lopez JV., Yuhki N., Masuda R., Modi W., O'Brien SJ. 1994. Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. *Journal of Molecular Evolution* 39:174–190.

Luo R., Liu B., Xie Y., Li Z., Huang W., Yuan J., He G., Chen Y., Pan Q., Liu Y., Tang J., Wu G., Zhang H., Shi Y., Liu Y., Yu C., Wang B., Lu Y., Han C., Cheung DW., Yiu S-M.,

Peng S., Xiaoqian Z., Liu G., Liao X., Li Y., Yang H., Wang J., Lam T-W., Wang J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* 1:18. DOI: 10.1186/2047-217X-1-18.

Marshall HD., Baker AJ. 1997. Structural conservation and variation in the mitochondrial control region of fringilline finches (*Fringilla* spp.) and the greenfinch (*Carduelis chloris*). *Molecular Biology and Evolution* 14:173–184.

Marthinsen G., Wennerberg L., Solheim R., Lifjeld JT. 2009. No phylogeographic structure in the circumpolar snowy owl (*Bubo scandiacus*). *Conservation Genetics* 10:923–933. DOI: 10.1007/s10592-008-9581-6.

Mazur KM., James PC. 2000. Barred Owl (*Strix varia*). *The Birds of North America Online* (A. Poole, Ed.) Ithaca: Cornell Lab of Ornithology. [Accessed 2016 Oct 1]. Retrieved from the Birds of North America Online: https://birdsna.org/Species-Account/bna/species/brdowl. DOI: 10.2173/bna.508.

McCarthy EM. 2006. *Handbook of avian hybrids of the world*. New York, New York: Oxford University Press.

Mindell DP., Sorenson MD., Dimcheff DE. 1998a. Multiple independent origins of mitochondrial gene order in birds. *Proceedings of the National Academy of Sciences* 95:10693–10697.

Mindell DP., Sorenson MD., Dimcheff DE. 1998b. An extra nucleotide is not translated in mitochondrial ND3 of some birds and turtles. *Molecular Biology and Evolution* 15:1568–1571.

Mindell DP., Sorenson MD., Dimcheff DE., Hasegawa M., Ast JC., Yuri T. 1999. Interordinal Relationships of Birds and Other Reptiles Based on Whole Mitochondrial Genomes. *Systematic Biology* 48:138–152. DOI: 10.1080/106351599260490.

Mindell DP., Sorenson MD., Huddleston CJ., Miranda HC., Knight A., Sawchuk SJ., Yuri T. 1997. Phylogenetic Relationships among and within Select Avian Orders Based on Mitochondrial DNA. In: *Avian Molecular Evolution and Systematics*. San Diego, California: Academic Press, 211–245.

Morgulis A., Coulouris G., Raytselis Y., Madden TL., Agarwala R., Schäffer AA. 2008. Database indexing for production MegaBLAST searches. *Bioinformatics* 24:1757–1764. DOI: 10.1093/bioinformatics/btn322.

Mundy NI., Winchell CS., Woodruff DS. 1996. Tandem Repeats and Heteroplasmy in the Mitochondrial DNA Control Region of the Loggerhead Shrike (*Lanius ludovicianus*). *Journal of Heredity* 87:21–26.

NCBI Resource Coordinators 2015. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research* 43:D6–D17. DOI: 10.1093/nar/gku1130.

Omote K., Nishida C., Dick MH., Masuda R. 2013. Limited phylogenetic distribution of a long tandem-repeat cluster in the mitochondrial control region in *Bubo* (Aves, Strigidae) and cluster variation in Blakiston's fish owl (*Bubo blakistoni*). *Molecular Phylogenetics and Evolution* 66:889–897. DOI: 10.1016/j.ympev.2012.11.015.

Pereira SL., Baker AJ. 2004. Low number of mitochondrial pseudogenes in the chicken (*Gallus gallus*) nuclear genome: implications for molecular inference of population history and phylogenetics. *BMC Evolutionary Biology* 4:17. DOI: 10.1186/1471-2148-4-17.

Pratt RC., Gibb GC., Morgan-Richards M., Phillips MJ., Hendy MD., Penny D. 2009. Toward Resolving Deep Neoaves Phylogeny: Data, Signal Enhancement, and Priors. *Molecular Biology and Evolution* 26:313–326. DOI: 10.1093/molbev/msn248.

Quinlan AR., Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842. DOI: 10.1093/bioinformatics/btq033.
Quinn TW., Wilson AC. 1993. Sequence evolution in and around the mitochondrial control region in birds. *Journal of Molecular Evolution* 37:417–425. DOI: 10.1007/BF00178871.
Randi E., Lucchini V. 1998. Organization and evolution of the mitochondrial DNA control region in the avian genus *Alectoris*. *Journal of Molecular Evolution* 47:449–462.
Richly E., Leister D. 2004. NUMTs in Sequenced Eukaryotic Genomes. *Molecular Biology and Evolution* 21:1081–1084. DOI: 10.1093/molbev/msh110.
Ruby JG. 2014. PRICE. Version 1.2. [Accessed 2016 Oct 1]. Available from: http://derisilab.ucsf.edu/software/price.
Ruby JG., Bellare P., DeRisi JL. 2013. PRICE: Software for the Targeted Assembly of Components of (Meta) Genomic Sequence Data. *G3: Genes|Genomes|Genetics* 3:865–880. DOI: 10.1534/g3.113.005967.
Ruegg K. 2008. Genetic, Morphological, and Ecological Characterization of a Hybrid Zone That Spans a Migratory Divide. *Evolution* 62:452–466. DOI: 10.1111/j.1558-5646.2007.00263.x.
Scherzinger W. 1982. Beobachtungen an WaldkauzHabichtskauz-Hybriden (*Strix aluco* × *Strix uralensis*). *Zoologische Gärten* 53:133–148.
Sorenson MD., Quinn TW. 1998. Numts: A Challenge for Avian Systematics and Population Biology. *The Auk* 115:214–221. DOI: 10.2307/4089130.
Sun X., Zhou W., Sun Z., Qian L., Zhang Y., Pan T., Zhang B. 2016. The complete mitochondrial genome of *Glaucidium brodiei* (Strigiformes: Strigidae). *Mitochondrial DNA Part A* 27:2508–2509. DOI: 10.3109/19401736.2015.1036252.
Tamura K., Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* 10:512–526.
Wada Y., Yamada Y., Nishibori M., Yasue H. 2004. Complete Nucleotide Sequence of Mitochondrial Genome in Silkie Fowl (*Gallus gallus var. domesticus*). *The Journal of Poultry Science* 41:76–82. DOI: 10.2141/jpsa.41.76.
Wilkinson GS., Chapman AM. 1991. Length and sequence variation in evening bat D-loop mtDNA. *Genetics* 128:607–617.
Williford D., DeYoung RW., Honeycutt RL., Brennan LA., Hernández F., Heffelfinger JR., Harveson LA. 2014. Phylogeography of the Gambel's Quail (*Callipepla gambelii*) of western North America. *The Wilson Journal of Ornithology* 126:218–235. DOI: 10.1676/13-111.1.
Wink M., El-Sayed A-A., Sauer-Gürth H., Gonzalez J. 2009. Molecular Phylogeny of Owls (Strigiformes) Inferred from DNA Sequences of the Mitochondrial Cytochrome *b* and the Nuclear *RAG-1* gene. *Ardea* 97:581–591. DOI: 10.5253/078.097.0425.
Wink M., Heidrich P. 1999. Molecular Evolution and Systematics of the Owls (Strigiformes). In: *Owls - A Guide to the Owls of the World*. Sussex, United Kingdom: Pica Press, 39–57.
Xiao B., Ma F., Sun Y., Li Q-W. 2006. Comparative Analysis of Complete Mitochondrial DNA Control Region of Four Species of Strigiformes. *Acta Genetica Sinica* 33:965–974. DOI: 10.1016/S0379-4172(06)60131-5.
Zhang Z., Schwartz S., Wagner L., Miller W. 2000. A Greedy Algorithm for Aligning DNA Sequences. *Journal of Computational Biology* 7:203–214. DOI: 10.1089/10665270050081478.

Zhang Y., Song T., Pan T., Sun X., Sun Z., Qian L., Zhang B. 2016. Complete sequence and gene organization of the mitochondrial genome of *Asio flammeus* (Strigiformes, strigidae). *Mitochondrial DNA Part A* 27:2665–2667. DOI: 10.3109/19401736.2015.1043538.

Zink RM. 1994. The Geography of Mitochondrial DNA Variation, Population Structure, Hybridization, and Species Limits in the Fox Sparrow (*Passerella iliaca*). *Evolution* 48:96–111. DOI: 10.2307/2410006.

Chapter 2 Supplementary Article

1 Supplementary Methods

1.1 Initial S. o. caurina assembly

1.1.1 BLATq version 1.0.2 (Henderson & Hanna, 2016a) used to align our 150 bp read 1 sequences from the Nextera700bp library (Hanna et al., 2017) to the *Ninox novaeseelandiae* mitochondrial genome (GenBank accession AY309457) using default BLAT parameters other than "-stepSize=5 -repMatch=100000 -out=blast8".

1.1.2 excerptByIds version 1.0.2 (Henderson & Hanna, 2016b) used, to extract the pairs of reads that had BLATq hits to the *Ninox novaeseelandiae* mitochondrial genome.

1.1.3 SOAPdenovo2 version 2.04 (Luo et al., 2012) used with default settings except "SOAPdenovo-127mer all -m 127 -R". In our configuration file for this assembly we used the default minimum alignment length between a read and contig (32 for paired-end reads) and the default minimum pair number cutoff (3 for paired-end reads) and set the reads to be used for the assembly of both contigs and scaffolds.

1.1.4 Web version of the NCBI BLAST+ version 2.2.29 tool BLASTn (Altschul et al., 1990; Zhang et al., 2000; Morgulis et al., 2008; Camacho et al., 2009) with default parameters to search the NCBI nucleotide collection (Johnson et al., 2008; Boratyn et al., 2013; NCBI Resource Coordinators, 2015; Benson et al., 2015).

1.1.5 GNU Grep version 2.16 (Free Software Foundation, 2014) used to search the trimmed and merged reads from lane 1 and 2 of the Nextera550bp library for reads that matched the assembled sequence of $tRNA^{Phe}$ or $tRNA^{Thr}$.

1.1.6 We found 3 reads that spanned $tRNA^{Phe}$ (1 of which was a merged read pair (Hanna et al., 2017)) and combined them using the Geneious version 9.1.4 (Kearse et al., 2012; Biomatters, 2016) *de novo* assembler. The resulting contig contained spanned from the control region through $tRNA^{Phe}$ and into *12S*.

1.1.7 PRICE version 1.2 (Ruby, Bellare & DeRisi, 2013; Ruby, 2014) used with the assembled contig spanning $tRNA^{Phe}$ as the initial contig and the parameters "-spf <merged_reads> 300 600 -icf <initial_contig.fasta> 1 1 5 -mol 25 -mpi 85 -MPI 80 -nc 60 -lenf 40 5 -lenf 90 10 -a 10 -target 85 1 1 1 -maxHp 25 -o <output.fa> -o <output.priceq>". The "merged_reads" parameter referred to the merged overlapping sequences from lane 1 of the Nextera550bp library.

1.1.8 BLATq version 1.0.2 (Henderson & Hanna, 2016a) to align our 150 bp read 1 sequences from the Nextera700bp library to the assembly output by PRICE using default settings other than "-stepSize=5 -repMatch=100000 -out=blast8".

1.1.9 excerptByIds version 1.0.2 (Henderson & Hanna, 2016b) to extract the read 2 of the paired-end sequences corresponding to the aligned read 1 sequences.

1.1.10 PRICE assembly with the same initial contig as before, but with more sequence data, including the merged overlapping sequences from both lane 1 and 2 of the Nextera550bp library and the matching paired-end sequences from the Nextera700bp library. We used the default PRICE assembly settings with the following exceptions: "-fp <Nextera700bp read 1 matches> <Nextera700bp read 2 matches> 809 -spf <Nextera550bp lane 1 merged reads> 300 600 -spf <Nextera550bp lane 2 merged reads> 300 600 -icf <initial_contig.fasta> 1 1 5 -mol 25 -mpi 85 -MPI 80 -nc 60 -lenf 40 5 -lenf 90 10 -a 10 -target 85 1 1 1 -maxHp 25 -o <output.fa> -o <output.priceq>".

1.1.11 MITOS WebServer version 605 (Bernt et al., 2013) specifying “genetic code = 02 - Vertebrate”.

1.1.12 Tandem Repeats Finder version 4.07b (Benson, 1999, 2012) used with default options.

1.2 S. o. caurina Sanger sequencing assembly confirmation

1.2.1 We isolated genomic DNA using a DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany). Polymerase chain reaction (PCR) conditions for primers 17589F and 41R included an initial denaturation at 94°C for 1 min; then 30 cycles at 94°C for 30 s, 60°C for 30 s, and 72°C for 2 min; and a final extension at 72°C for 7 min. We then sequenced both ends of the PCR-amplified fragment using BigDye terminator chemistry (Applied Biosystems, Foster City, California) on an ABI 3130xl automated sequencer (Applied Biosystems, Foster City, California).

1.2.2 Polymerase chain reaction (PCR) conditions for primers 17572F and 41R were the same as for primers 17589F and 41R (initial denaturation at 94°C for 1 min; then 30 cycles at 94°C for 30 s, 60°C for 30 s, and 72°C for 2 min; and a final extension at 72°C for 7 min). We sequenced the PCR products using BigDye terminator chemistry (Applied Biosystems, Foster City, California) on an ABI 3130xl automated sequencer (Applied Biosystems, Foster City, California).

1.2.3 We used the Geneious mapper through the “map to reference” function with default options other than sensitivity set to “highest sensitivity / slow” to align the edited Sanger-derived sequences to the 19,946 nt preliminary mitochondrial genome assembly.

1.2.4 Polymerase chain reaction (PCR) conditions for primers cytb-F1 and 17122R included an initial denaturation at 94°C for 3 min; then 35 cycles at 94°C for 30 s, 50°C for 30 s, and 72°C for 2 min; and a final extension at 72°C for 10 min.

1.3 S. o. caurina final assembly

1.3.1 We removed scaffold-3674 from the draft whole nuclear genome assembly (Hanna et al., 2017) using the “filterbyname.sh” tool in the BBMap tool suite version 36.02 (Bushnell, 2016) and replaced it with the 19,948 nt mitochondrial genome assembly from our targeted assembly methodology. When referring to specific scaffolds here and in the manuscript, we have inserted a dash (“-”) between the word “scaffold” and the scaffold number for legibility. These dashes are not present in any of the assembly data files. Thus, “scaffold-3674” referenced in the manuscript will appear as “scaffold3674” in the assembly and other associated files.

1.3.2 We aligned all filtered Illumina sequences to this new draft reference genome using bwa version 0.7.13-r1126 (Li, 2013a) using default options except parameters "bwa mem -M". We separately aligned paired-end and unpaired reads. We then merged the paired-end and unpaired read alignments using the Picard version 2.2.4 (http://broadinstitute.github.io/picard) function MergeSamFiles and sorted them using the Picard function SortSam, employing default settings for both tools. We next marked duplicate reads (both PCR and optical) using the Picard function MarkDuplicates, employing default settings.

1.3.3 We then filtered out duplicate reads, low quality alignments, secondary alignments, and alignments where both reads of a pair did not align to the mitochondrial assembly. We used Samtools to filter out duplicate reads marked by Picard using the Samtools parameters “-F 0x400”. We next used Samtools to filter out alignments of quality less than 10 with the parameters “-q 10”. We then filtered out secondary alignments with the Samtools parameters “-F 0x100”. We then used Samtools with GNU Awk (GAWK)

version 4.0.1 (Free Software Foundation, 2012) to filter out paired reads where one of the reads mapped to a different contig/scaffold than the mitochondrial genome using parameters “samtools view -h <input.bam> | awk '$7 == "=" || $7 == "*" || $1 ~ "^@"' | samtools view -Sb - > <output.bam>”.

1.3.4 We used Samtools version 1.3 with HTSlib 1.3.1 (Li et al., 2009, 2015) with GAWK version 4.0.1 (Free Software Foundation, 2012) to filter out all aligned sequences less than 300 nt using parameters “samtools view -h <input.bam> | awk '$1 ~ "^@"' || length($10) >= 300’ | samtools view -Sb - > <output.bam>”.

1.3.5 We visually inspected all sites where there was lower coverage and any hint of disagreement between reads and, except in the case of CR1 and CR2 where we relied on the Sanger-derived sequence data, decided in favor of majority evidence, which matched our preliminary assembly at all sites, providing confirmation of our assembly methodology.

1.4 S. o. caurina final annotation

1.4.1 MITOS WebServer version 806 (Bernt et al., 2013) specifying “genetic code = 02 - Vertebrate”.

1.4.2 Web version of Tandem Repeats Finder version 4.09 (Benson, 1999, 2016) employing default options.

1.4.3 We used bioawk version 1.0 (Li, 2013b) and GAWK version 4.0.1 (Free Software Foundation, 2012) to find goose hairpin sequences by searching for 7 C’s followed by 1 to 3 D nucleotides (A, G, or T) followed by 7 C’s in a FASTA format file (Pearson & Lipman, 1988) with each control region input as a separate entry with the command “bioawk -c fastx '{print $seq}' $1 | awk '{pos=match($0, /CCCCCCC[AGT]{1,3}CCCCCCC/);if(pos){print pos}}'”.

1.4.4 We used the NCBI BLAST+ version 2.4.0 (Altschul et al., 1990; Zhang et al., 2000; Morgulis et al., 2008; Camacho et al., 2009) tool “makeblastdb” with options “-parse_seqids -dbtype nucl” to create a database of the scaffold-3674 gene sequences and then the tool “blastn” with default options except “-outfmt 6” to align the targeted assembly gene sequences against this database.

1.4.5 We aligned the primers we developed to amplify control region 2 and the N1 primer used by Barrowclough, Gutierrez & Groth (1999) to amplify a portion of control region 1 to the final assembly using Geneious version 9.1.4 mapper through the “map to reference” function (Kearse et al., 2012; Biomatters, 2016) with default options other than using the sensitivity set at “highest sensitivity / slow”. We determined the position in the final assembly of the D16 primer used by Barrowclough, Gutierrez & Groth (1999) to amplify a portion of control region 1 by using the Geneious version 9.1.4 *de novo* assembler (Kearse et al., 2012; Biomatters, 2016) with default parameters other than setting sensitivity at “highest sensitivity / slow” to assemble control region 1 with the D16 primer.

1.4.6 We performed a multiple alignment of control regions 1 and 2 using the Geneious version 9.1.4 (Kearse et al., 2012; Biomatters, 2016) implementation of the MUSCLE version 3.8.425 (Edgar, 2004) aligner with default options.

1.4.7 We used the web version of NCBI’s BLAST+ Version 2.5.0 tool BLASTN (Altschul et al., 1990; Zhang et al., 2000; Morgulis et al., 2008; Camacho et al., 2009) with default parameters to search the NCBI nucleotide collection (Johnson et al., 2008; Boratyn et al., 2013; NCBI Resource Coordinators, 2015; Benson et al., 2015) for sequences similar to

control regions 1 and 2 in order to assess whether published control region sequences of related species are more similar to control region 1 or 2.

1.4.8 As a result of these searches, we aligned the primers used by Omote et al. (2013) to amplify the control region in *Strix uralensis* to the final assembly using Geneious version 9.1.4 mapper with the "map to reference" function (Kearse et al., 2012; Biomatters, 2016) with default options other than using the Geneious mapper with sensitivity set at "highest sensitivity / slow".

1.5 Nuclear pseudogenation of S. o. caurina mitochondrial genes

1.5.1 We first used the NCBI BLAST+ version 2.4.0 tool BLASTN (Altschul et al., 1990; Zhang et al., 2000; Morgulis et al., 2008; Camacho et al., 2009) to align the final *S. o. caurina* mitochondiral genome assembly to the draft nuclear genome assembly (Hanna et al., 2017) using default parameters other than "-outfmt 6".

1.5.2 We used GAWK version 4.0.1 (Free Software Foundation, 2012) to remove all alignments to scaffold3674, which was the assembly of the mitochondrial genome in the draft nuclear genome assembly.

1.5.3 We next used GAWK version 4.0.1 (Free Software Foundation, 2012) to reformat the BLAST output into a BED (Browser Extensible Data) formatted file (http://genome.ucsc.edu/FAQ/FAQformat#format1) with the parameters "cat <filtered_BLAST_file> | awk 'BEGIN {OFS = "\t"} {print "Strix_Occidentalis",$7,$8,$2,$9,$10,$12}' | awk 'BEGIN {OFS = $"\t"} {if ($3<$2) print $1,$3,$2,$4,$5,$6,7; else print $0}'".

1.5.4 We used BEDTools version 2.26.0 tool "intersect" (Quinlan & Hall, 2010) to produce a BED file of the intersection of the BED-formatted BLAST output with the BED file output from the MITOS annotation of the final mitochondrial genome assembly using the parameters "-a <BED-formatted BLAST output file> -b <MITOS annotation BED file> -wo". We then used the output of the intersection to determine the mitochondrial genes spanned by each *Numt*.

1.6 Strix varia sample CAS95964

1.6.1 We extracted genomic DNA using a DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany). We used 50 ng genomic DNA to prepare a whole-genome library using a Nextera DNA Sample Preparation Kit (Illumina, San Diego, California). After tagmentation, we cleaned the reaction with a DNA Clean & Concentrator -5 kit (Zymo Research, Irvine, California). We amplified the reaction with 5 cycles of PCR using a KAPA Library Amplification kit (KAPA Biosystems, Wilmington, Massachusetts) and then cleaned the reaction with a DNA Clean & Concentrator -5 kit (Zymo Research, Irvine, California). We used Dye-Free, 1.5% agarose, 250-1,500 base pair (bp) cassette on a BluePippin (Sage Science, Beverly, Massachusetts) to select library fragments in the size range of 534-634 bp, which, after subtracting the 134 bp of adapters, corresponded to selecting an average insert size of 450 bp. We next performed a real-time PCR (rtPCR) using a KAPA Real-Time Library Amplification Kit (KAPA Biosystems, Wilmington, Massachusetts) on a CFX96 Touch Real-Time PCR Detection System (Bio-Rad, Hercules, California) to further amplify the library with 9 cycles PCR. We then cleaned the PCR products with a DNA Clean & Concentrator -5 kit (Zymo Research, Irvine, California). We lastly assessed the library fragment size distribution with a 2100 BioAnalyzer (Agilent Technologies, Santa Clara, California) and the concentration of double-stranded DNA material with a Qubit 2.0 Flurometer (Invitrogen, Carlsbad,

California). Due to the presence of small peaks in the BioAnalyzer trace, we further cleaned the library using 0.6X Agencourt AMPure XP (Beckman Coulter, Brea, California) magnetic beads. We obtained approximately one lane of 100 bp paired-end data sequenced in an indexed pool using a 2-lane flow cell with a HiSeq PE Rapid Cluster Kit and a 200 cycle HiSeq Rapid SBS Kit v1 on a HiSeq 2500 (Illumina, San Diego, California). The raw sequences are available upon request.

1.6.2 We performed adapter and quality trimming of the sequence data using Trimmomatic version 0.30 (Bolger, Lohse & Usadel, 2014) with the following options: "ILLUMINACLIP:<FASTA format file of Illumina adapter sequences>:2:30:7 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36".

1.6.3 For use in only the SOAPdenovo2 assembly, we trimmed the sequences using a different set of parameters and performed error-correction of the sequences. We performed adapter and quality trimming using Trimmomatic version 0.30 (Bolger, Lohse & Usadel, 2014) with the following options: "ILLUMINACLIP:<FASTA format file of Illumina adapter sequences>:2:30:7 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36 HEADCROP:12". We merged into a single file all of the single reads that resulted after their pair was dropped in the trimming process. Then we used the k-mer-based error corrector in the SOAPdenovo2 toolkit, SOAPec version 2.01 (Luo et al., 2012), to correct sequence errors. We first used the KmerFreq_HA tool to create a kmer frequency spectrum with default options except "-k 27", which indicate that we used a kmer size of 27 for creating the spectrum. We then used the Corrector_HA tool along with the kmer frequency spectrum that we created to correct all of our trimmed reads using default options except "-k 27 -r 36", which indicate that we used a kmer size of 27 for the error correction and kept trimmed reads as short as 36 bp.

1.7 Strix varia sample CMCB41533

1.7.1 We obtained tissue from a *S. varia* collected in Hamilton County, Ohio, United States of America (CNHM<USA-OH>:ORNITH:B41533; Table 1), hereafter "CMCB41533", which is well outside of the zone of contact of *S. varia* and *S. occidentalis caurina* (Haig et al., 2004). We obtained paired-end Illumina sequence data from a genomic library constructed, sequenced, and the data processed as described in (Hanna et al., 2017). The raw sequences are available from NCBI (SRA run accessions SRR5428115, SRR5428116, and SRR5428117).

1.8 Strix varia mitochondrial genome assembly

1.8.1 We generated the mitochondrial genome assembly of the *S. varia* sample CMCB41533 by building a succession of assemblies that contributed information to the final assembly from which we extracted the gene sequences. We used assemblies of sample CAS95964 to inform the process, but, as we had more sequence data for sample CMCB41533, we chose to only produce a final assembly for this sample.

1.9 Assembly of Strix varia ContigInput1

1.9.1 We used bwa version 0.7.13-r1126 (Li, 2013a) with default options other than parameters "bwa mem -M". We separately aligned paired-end and unpaired reads.

1.9.2 We merged the paired-end and unpaired read alignments using the Picard version 2.2.4 (http://broadinstitute.github.io/picard) function MergeSamFiles and sorted them using the Picard function SortSam, employing default settings for both tools. We next marked duplicate reads (both PCR and optical) using the Picard function MarkDuplicates, employing default settings.

1.9.3 We filtered the alignment file for only alignments to the final mitochondrial genome assembly using Samtools version 1.3 with HTSlib 1.3.1 (Li et al., 2009, 2015). We then used Samtools to filter out duplicate reads marked by Picard using the Samtools parameters "-F 0x400". We next used Samtools to filter out alignments of quality less than 10 with the parameters "-q 10". We then filtered out secondary alignments with the Samtools parameters "-F 0x100". We then used Samtools with GNU Awk (GAWK) version 4.0.1 (Free Software Foundation, 2012) to filter out paired reads where one of the reads mapped to a different contig/scaffold than the mitochondrial genome using parameters "samtools view -h <input.bam> | awk '$7 == "=" || $7 == "*" || $1 ~ "^@"' | samtools view -Sb - > <output.bam>".

1.9.4 We visualized the alignment across the reference sequence in Geneious version 9.1.4 (Kearse et al., 2012; Biomatters, 2016). We used Geneious to generate a consensus sequence with default parameters for the alignment to the mitochondrial genome.

1.9.5 We extracted 3 sequences from this consensus sequence based on the *S. o. caurina* annotations to give a 142 nt fragment spanning from nucleotide 5 of $tRNA^{Phe}$ part way into *12S*; a longer, 844 nt fragment spanning from nucleotide 4 of $tRNA^{Phe}$ part way into *12S*; and a 1,042 nt fragment spanning from nucleotide 142 of *cyt b* part way into $tRNA^{Thr}$. We then used these extracted sequences as three separate seed contigs in an assembly using PRICE version 1.2 (Ruby, Bellare & DeRisi, 2013; Ruby, 2014). We used the trimmed CMCB41533 paired read 1 and 2 sequences as the sequence data for this run. Our PRICE assembly parameters were the defaults other than the following "-fp <read 1 paired sequences> < read 2 paired sequences> 466 -icf <initial_contig.fasta> 1 1 5 -mol 30 -mpi 90 -MPI 85 -nc 60 -a 8 -target 85 1 1 1 -maxHp 25 -o <output.fa> -o <output.priceq>".

1.10 Assembly of ContigInput2 - CAS95964

1.10.1 We first used SOAPdenovo2 version 2.04 (Luo et al., 2012) to assemble all of the trimmed, error-corrected CAS95964 sequences, employing default parameters other than the options "SOAPdenovo all -m 63 -R". In our configuration file for this assembly we used the default minimum alignment length between a read and contig (32) and the default minimum pair number cutoff (3) for both the paired-end and single-end reads. We set the paired-end reads to be used for the assembly of both contigs and scaffolds and the single-end reads for use only in contig assembly. We input both the paired-end and the single-end reads with a rank of 1 and set the average insert size as 446 bp. This produced a contig of length 15,019 nt.

1.10.2 We extended the contig using PRICE version 1.2 (Ruby, Bellare & DeRisi, 2013; Ruby, 2014). We employed the trimmed (and not error-corrected) CAS95964 paired read 1 and 2 sequences as the sequence data for this run. We used the 15,019 nt contig output from the SOAPdenovo2 run above as the initial contig. Our PRICE assembly parameters were the defaults other than the following "-fp <read 1 paired sequences> <read 2 paired sequences> 446 -icf <initial_contig.fasta> 1 1 5 -mol 25 -mpi 85 -MPI 80 -nc 60 -lenf 40 5 -lenf 90 10 -a 10 -target 85 1 1 1 -maxHp 25 -o <output.fa> -o <output.priceq>".

1.11 Assembly of ContigInput2 - CMCB41533 PRICE

1.11.1 We employed the trimmed CMCB41533 paired read 1 and 2 sequences as the sequence data for this run. We used the 16,652 nt contig output from the CAS95964 run above as the initial contig. Our PRICE version 1.2 (Ruby, Bellare & DeRisi, 2013; Ruby, 2014) assembly parameters were the defaults other than the following "-fp <read 1 paired

sequences> < read 2 paired sequences> 350 -icf <initial_contig.fasta> 1 1 5 -mol 25 -mpi 85 -MPI 80 -nc 60 -lenf 40 5 -lenf 90 10 -a 10 -target 85 1 1 1 -maxHp 25 -o <output.fa> -o <output.priceq>".

1.12 Final Strix varia assembly

1.12.1 We used PRICE version 1.2 (Ruby, Bellare & DeRisi, 2013; Ruby, 2014) the 9,690 nt ContigInput1 contig output from cycle 16 of the initial CMCB41533 PRICE run and the 17,073 nt ContigInput2 as the initial contigs. We used the CMCB41533 paired read 1 and 2 sequences as well as the unpaired read 1 sequences that lost their mate as a result of quality trimming for the sequence data input for this assembly. Our PRICE assembly parameters were the defaults other than the following "-fp <read 1 paired sequences> < read 2 paired sequences> 400 -spf <read 1 unpaired> 110 200 -icf <initial_contig.fasta> 1 1 5 -mol 25 -mpi 85 -MPI 80 -nc 60 -lenf 40 5 -lenf 90 10 -a 24 -target 85 1 1 1 -maxHp 25 -o <output.fa> -o <output.priceq>".

1.12.2 PCR conditions for primers cytb-F1 and 17122R included an initial denaturation at 94°C for 3 min; then 35 cycles at 94°C for 30 s, 53°C for 30 s, and 72°C for 2 min; and a final extension at 72°C for 10 min.

1.12.3 PCR conditions for primers ND6-ext1F and 12S-ext1R included an initial denaturation at 94°C for 3 min; then 35 cycles at 94°C for 30 s, 54°C for 30 s, and 72°C for 2 min; and a final extension at 72°C for 10 min.

1.12.4 We annotated the PRICE assembly using the MITOS WebServer version 605 (Bernt et al., 2013) specifying "genetic code = 02 - Vertebrate".

1.13 Comparison of Strix occidentalis and Strix varia mitochondrial genes

1.13.1 MAFFT version 7.305b (Katoh et al., 2002; Katoh & Standley, 2013; Katoh, 2016) used with the default options other than parameters "--auto --clustalout".

1.13.2 trimAl version 1.4.rev15 (Capella-Gutiérrez, Silla-Martínez & Gabaldón, 2009; Capella-Gutiérrez & Gabaldón, 2013) with default options other than "-mega" used to convert the alignments to MEGA format (Kumar, Tamura & Nei, 1994; Kumar, Stecher & Tamura, 2016).

1.13.3 MEGA version 7.0.18 (Kumar, Stecher & Tamura, 2016) used to calculate the p-distance (an uncorrected pairwise distance that is the proportion of nucleotide sites at which two sequences are different obtained by dividing the number of differences by the total number of nucleotide sites) between *S. occidentalis caurina* and *S. varia* for each gene with all alignment positions with gaps or missing data removed from the analysis.

1.13.4 MEGA version 7.0.18 (Kumar, Stecher & Tamura, 2016) used to calculate for each gene the pairwise distance corrected by the Tamura-Nei model (TN93) of DNA sequence evolution (Tamura & Nei, 1993) with rate variation among sites modeled using a gamma distribution with shape parameter = 1, differences in the composition bias of sequences considered in the comparisons (Tamura & Kumar, 2002), and with all alignment positions with gaps or missing data removed from the analysis.

1.13.5 We weighted each distance by the length of the gene alignment from which it was derived as a proportion of the total alignment length across all gene alignments and calculated a weighted average pairwise distance across all of the genes.

2 Supplementary Figures

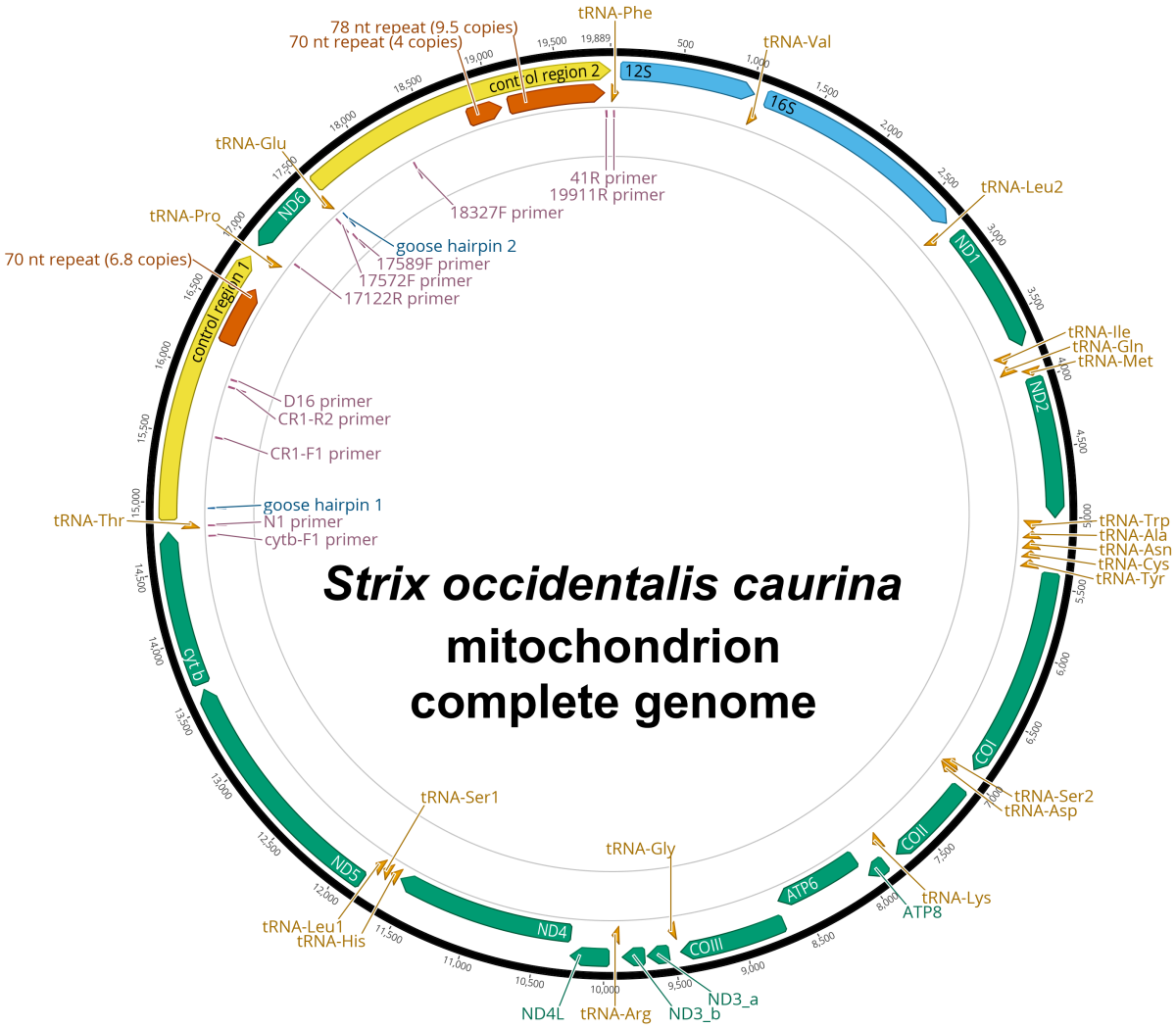

Figure S1. Complete genome of the *Strix occidentalis caurina* mitochondrion.

This is a graphical representation of the annotated complete genome of the northern spotted owl (*Strix occidentalis caurina*) mitochondrion. We have color-coded the various annotations, including genes for rRNA in sky blue, tRNA genes in orange, and all other genes in bluish green. The control regions are in yellow and the goose hairpin for each control region is depicted in blue. The locations of the primers we developed to amplify control regions 1 and 2 as well as the N1 and D16 primers used by Barrowclough et al. (1999) to amplify a portion of control region 1 are in reddish purple. The reverse complement versions of primers used ("-RC" versions) are not shown. Regions with repetitive motifs are in vermillion. The base numbers around the perimeter of the figure are in nucleotides. We used Geneious version 9.1.4 (Kearse et al., 2012; Biomatters, 2016) to construct this figure.

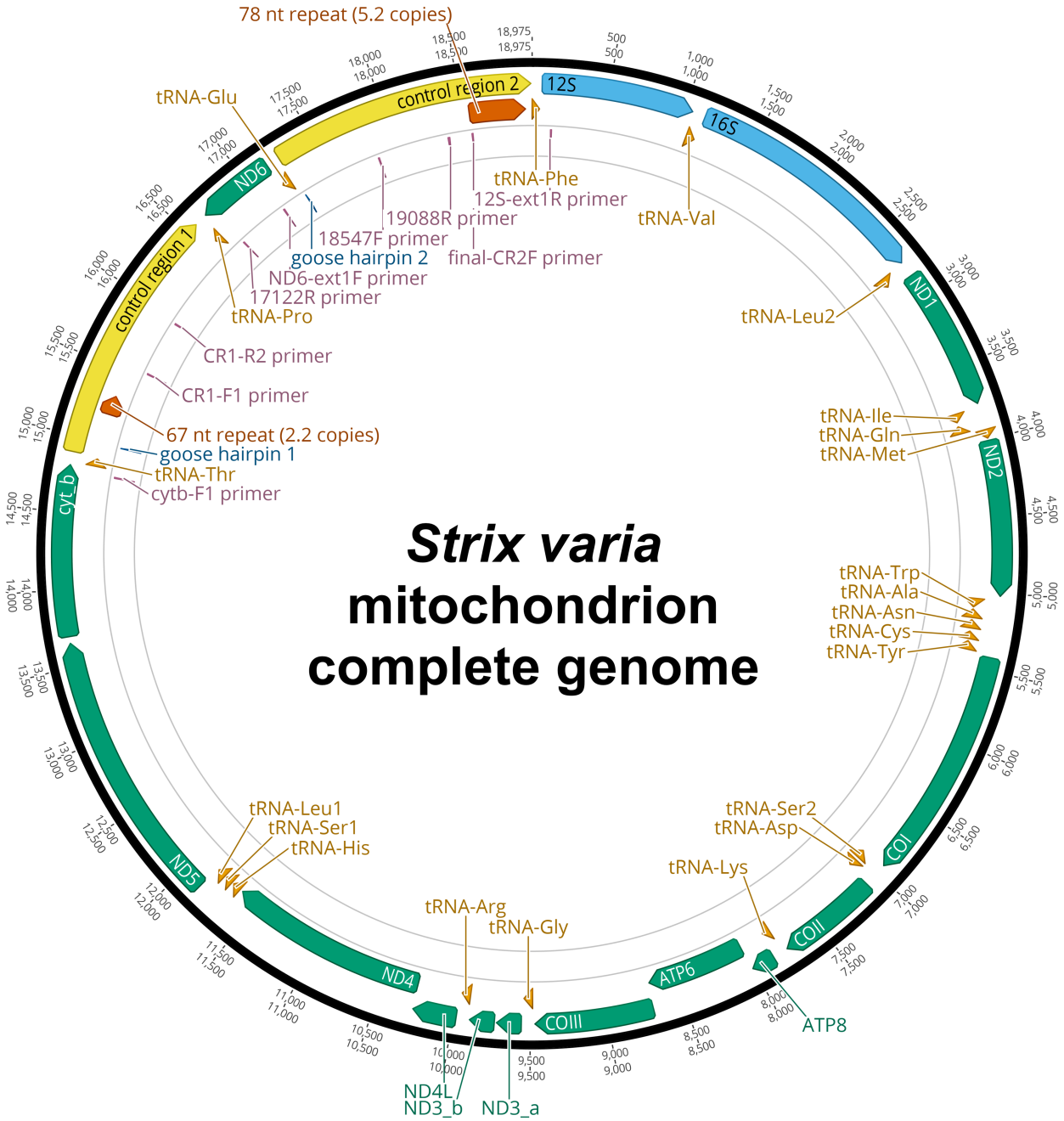

Figure S2. Complete genome of the *Strix varia* mitochondrion.

This is a graphical representation of the annotated complete genome of the barred owl (*Strix varia*) mitochondrion. We have color-coded the various annotations, including genes for rRNA in sky blue, tRNA genes in orange, and all other genes in bluish green. The control regions are in yellow and the goose hairpin for each control region is depicted in blue. The locations of the primers we developed to amplify control regions 1 and 2 are in reddish purple. The reverse complement versions of primers used ("-RC" versions) are not shown. Regions with repetitive motifs are in vermillion. The base numbers around the perimeter of the figure are in nucleotides. We used Geneious version 9.1.4 (Kearse et al., 2012; Biomatters, 2016) to construct this figure.

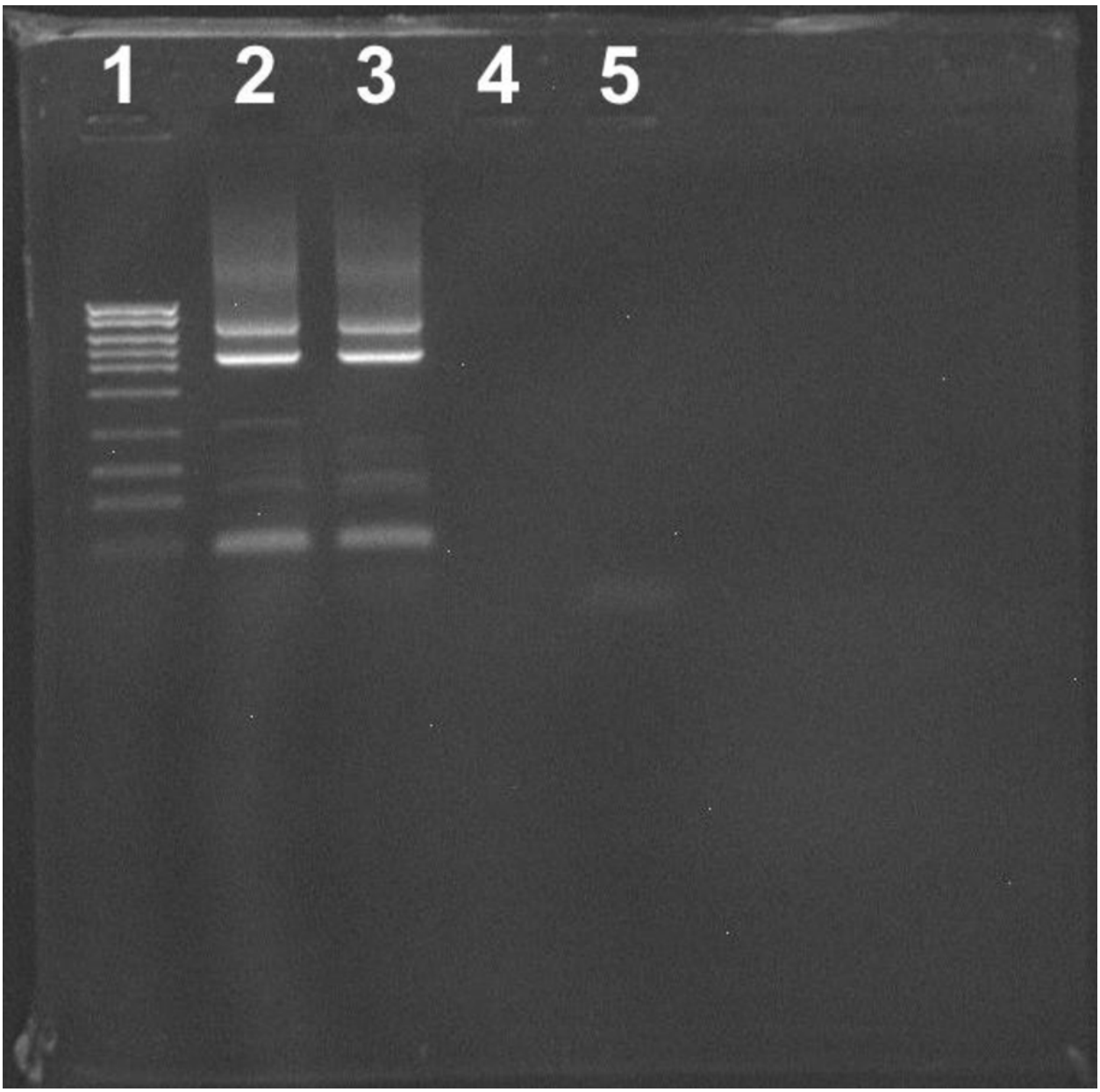

Figure S3. *Strix occidentalis caurina* CR1 PCR-amplification products.

This photograph of an agarose gel displays the lengths of the two products of PCR-amplification of the *S. o. caurina* CR1 using primers cyb-F1 and 17122R. In lane 1 we loaded Fisher BioReagents exACTGene DNA Ladder (Cat. No. BP2576100; Fisher Scientific) the ten bands of which were of lengths 5,000; 4,000; 3,000; 2,500; 2,000; 1,500; 1,000; 700; 500; and 300 nt. Lanes 2 and 3 contained independent PCR replicates of CR1 amplification products. Lane 4 was blank. In lane 5 we loaded the negative control for the PCR.

3 Supplementary References

Altschul SF., Gish W., Miller W., Myers EW., Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215:403–410. DOI: 10.1016/S0022-2836(05)80360-2.

Barrowclough GF., Gutierrez RJ., Groth JG. 1999. Phylogeography of Spotted Owl (*Strix occidentalis*) Populations Based on Mitochondrial DNA Sequences: Gene Flow, Genetic Structure, and a Novel Biogeographic Pattern. *Evolution* 53:919–931. DOI: 10.2307/2640731.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research* 27:573–580. DOI: 10.1093/nar/27.2.573.

Benson G. 2012. Tandem Repeats Finder. Version 4.07b. [Accessed 2016 Oct 1]. Available from: https://tandem.bu.edu/trf/trf.html.

Benson G. 2016. Tandem Repeats Finder. Version 4.09. [Accessed 2016 Oct 1]. Available from: https://tandem.bu.edu/trf/trf.html.

Benson DA., Clark K., Karsch-Mizrachi I., Lipman DJ., Ostell J., Sayers EW. 2015. GenBank. *Nucleic Acids Research* 43:D30–D35. DOI: 10.1093/nar/gku1216.

Bernt M., Donath A., Jühling F., Externbrink F., Florentz C., Fritzsch G., Pütz J., Middendorf M., Stadler PF. 2013. MITOS: Improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution* 69:313–319. DOI: 10.1016/j.ympev.2012.08.023.

Biomatters 2016. Geneious. Version 9.1.4. [Accessed 2016 Oct 1]. Available from: http://www.geneious.com.

Bolger AM., Lohse M., Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. DOI: 10.1093/bioinformatics/btu170.

Boratyn GM., Camacho C., Cooper PS., Coulouris G., Fong A., Ma N., Madden TL., Matten WT., McGinnis SD., Merezhuk Y., Raytselis Y., Sayers EW., Tao T., Ye J., Zaretskaya I. 2013. BLAST: a more efficient report with usability improvements. *Nucleic Acids Research* 41:W29–W33. DOI: 10.1093/nar/gkt282.

Bushnell B. 2016. BBMap. Version 36.02. [Accessed 2016 Oct 1]. Available from: http://sourceforge.net/projects/bbmap.

Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. DOI: 10.1186/1471-2105-10-421.

Capella-Gutiérrez S., Gabaldón T. 2013. trimAl. Version 1.4.rev15. [Accessed 2016 Oct 1]. Available from: https://github.com/scapella/trimal.

Capella-Gutiérrez S., Silla-Martínez JM., Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973. DOI: 10.1093/bioinformatics/btp348.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32:1792–1797. DOI: 10.1093/nar/gkh340.

Free Software Foundation 2012. GNU Awk . Version 4.0.1. [Accessed 2016 Oct 1]. Available from: https://www.gnu.org/software/gawk.

Free Software Foundation 2014. GNU Grep. Version 2.16. [Accessed 2016 Oct 1]. Available from: https://www.gnu.org/software/grep.

Haig SM., Mullins TD., Forsman ED., Trail PW., Wennerberg L. 2004. Genetic identification of spotted owls, barred owls, and their hybrids: legal implications of hybrid identity. *Conservation Biology* 18:1347–1357.

Hanna ZR., Henderson JB., Wall JD., Emerling CA., Fuchs J., Runckel C., Mindell DP., Bowie RCK., DeRisi JL., Dumbacher JP. 2017. Northern Spotted Owl (*Strix occidentalis caurina*) Genome: Divergence with the Barred Owl (*Strix varia*) and Characterization of Light-Associated Genes. *Genome Biology and Evolution* 9:2522–2545. DOI: 10.1093/gbe/evx158.

Henderson JB., Hanna ZR. 2016a. BLATq. Version 1.0.2. *Zenodo*. DOI: 10.5281/zenodo.61136.

Henderson JB., Hanna ZR. 2016b. excerptByIDs. Version 1.0.2. *Zenodo*. DOI: 10.5281/zenodo.61134.

Johnson M., Zaretskaya I., Raytselis Y., Merezhuk Y., McGinnis S., Madden TL. 2008. NCBI BLAST: a better web interface. *Nucleic Acids Research* 36:W5–W9. DOI: 10.1093/nar/gkn201.

Katoh K. 2016. MAFFT: a multiple sequence alignment program. Version 7.305b. [Accessed 2016 Oct 1]. Available from: http://mafft.cbrc.jp/alignment/software.

Katoh K., Misawa K., Kuma K., Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* 30:3059–3066. DOI: 10.1093/nar/gkf436.

Katoh K., Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* 30:772–780. DOI: 10.1093/molbev/mst010.

Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A., Markowitz S., Duran C., Thierer T., Ashton B., Meintjes P., Drummond A. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649. DOI: 10.1093/bioinformatics/bts199.

Kumar S., Stecher G., Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* 33:1870–1874. DOI: 10.1093/molbev/msw054.

Kumar S., Tamura K., Nei M. 1994. MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. *Computer applications in the biosciences : CABIOS* 10:189–191. DOI: 10.1093/bioinformatics/10.2.189.

Li H. 2013a. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv:1303.3997 Q-Bio. [Accessed 2016 Feb 16]. Available from: http://arxiv.org/abs/1303.3997.

Li H. 2013b. bioawk. Version 1.0. [Accessed 2016 Oct 1]. Available from: https://github.com/lh3/bioawk.

Li H., Handsaker B., Marshall J., Danecek P. 2015. Samtools. Version 1.3 with HTSlib 1.3.1. [Accessed 2016 Oct 1]. Available from: http://www.htslib.org.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R., Subgroup 1000 Genome Project Data Processing. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. DOI: 10.1093/bioinformatics/btp352.

Luo R., Liu B., Xie Y., Li Z., Huang W., Yuan J., He G., Chen Y., Pan Q., Liu Y., Tang J., Wu G., Zhang H., Shi Y., Liu Y., Yu C., Wang B., Lu Y., Han C., Cheung DW., Yiu S-M.,

Peng S., Xiaoqian Z., Liu G., Liao X., Li Y., Yang H., Wang J., Lam T-W., Wang J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* 1:18. DOI: 10.1186/2047-217X-1-18.

Morgulis A., Coulouris G., Raytselis Y., Madden TL., Agarwala R., Schäffer AA. 2008. Database indexing for production MegaBLAST searches. *Bioinformatics* 24:1757–1764. DOI: 10.1093/bioinformatics/btn322.

NCBI Resource Coordinators 2015. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research* 43:D6–D17. DOI: 10.1093/nar/gku1130.

Omote K., Nishida C., Dick MH., Masuda R. 2013. Limited phylogenetic distribution of a long tandem-repeat cluster in the mitochondrial control region in *Bubo* (Aves, Strigidae) and cluster variation in Blakiston's fish owl (*Bubo blakistoni*). *Molecular Phylogenetics and Evolution* 66:889–897. DOI: 10.1016/j.ympev.2012.11.015.

Pearson WR., Lipman DJ. 1988. Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences* 85:2444–2448.

Quinlan AR., Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842. DOI: 10.1093/bioinformatics/btq033.

Ruby JG. 2014. PRICE. Version 1.2. [Accessed 2016 Oct 1]. Available from: http://derisilab.ucsf.edu/software/price.

Ruby JG., Bellare P., DeRisi JL. 2013. PRICE: Software for the Targeted Assembly of Components of (Meta) Genomic Sequence Data. *G3: Genes|Genomes|Genetics* 3:865–880. DOI: 10.1534/g3.113.005967.

Tamura K., Kumar S. 2002. Evolutionary Distance Estimation Under Heterogeneous Substitution Pattern Among Lineages. *Molecular Biology and Evolution* 19:1727–1736.

Tamura K., Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* 10:512–526.

Zhang Z., Schwartz S., Wagner L., Miller W. 2000. A Greedy Algorithm for Aligning DNA Sequences. *Journal of Computational Biology* 7:203–214. DOI: 10.1089/10665270050081478.

Chapter 3

Whole-genome sequences confirm lack of widespread introgression between the spotted owl (*Strix occidentalis*) and barred owl (*Strix varia*) (Aves: Strigiformes: Strigidae) in western North America

Introduction

We now know of several examples from the past century where humans have introduced non-native vertebrate species into the native range of a closely related species in western North America and generated moving hybrid swarms. In California, genes of the non-native barred tiger salamander (*Ambystoma tigrinum*) are spreading into the range of the California tiger salamander (*A. californiense*) (Fitzpatrick et al., 2009, 2010). In the Flathead River system of Montana and British Columbia, the non-native rainbow trout (*Oncorhynchus mykiss*) is rapidly hybridizing with the native westslope cutthroat trout (*O. clarkii lewisi*) (Muhlfeld et al., 2014). In addition to these intentional introductions of non-native species, with changing global climate and the documented movement of species ranges, many species are invading novel geographic regions (Parmesan et al., 1999; Parmesan, 2006) and establishing broad contact with related taxa (Rieseberg et al., 2007).

The spotted owl (*Strix occidentalis*) is a large wood owl inhabitant of western North American forests. In 1990, the U.S. Fish and Wildlife Service listed the northern spotted owl (*S. o. caurina*) as "threatened" under the Endangered Species Act (ESA) (Thomas et al., 1990). Due to continuing population declines (Dugger et al., 2015; Davis et al., 2016), the northern spotted owl remains protected. When initially listed under the ESA, habitat loss was regarded as the primary threat to the northern spotted owl (Forsman, Meslow & Wight, 1984; Anderson & Burnham, 1992). Recent research has confirmed a second major threat to the persistence of the spotted owl - the invasion and expansion of the congeneric barred owl (*S. varia*) into western North American forests (Dugger et al., 2015; Diller et al., 2016). Formerly, the barred owl (*S. varia*) only inhabited areas east of the Rocky Mountains (Mazur & James, 2000), but, in the last 50-100 years, it has expanded its range to western North America (Dark, Gutiérrez & Gould, 1998; Livezey, 2009a,b). At present, sympatric populations of spotted and barred owls exist from British Columbia to southern California (Taylor & Forsman, 1976; Haig et al., 2004; Livezey, 2009a).

Spotted and barred owls are genetically divergent, being on average 13.9% divergent in the mitochondrial control region (Haig et al., 2004), 10.74% divergent in the non-tRNA mitochondrial genes (Hanna et al., 2017a), and 7.04×10^{-3} (~ 0.7%) divergent across the nuclear genome (Hanna et al., 2017c). Barred and spotted owls can hybridize and backcross (Haig et al., 2004; Kelly & Forsman, 2004; Funk et al., 2007), with heterospecific matings and F_1 hybrids more commonly reported where barred owls are rare and spotted owls common (Kelly & Forsman, 2004). Early studies reported at least sixteen F_2 individuals in Washington and Oregon (Kelly & Forsman, 2004; Funk et al., 2007). In museum collections of western barred owl specimens, there is striking morphological variation. Eastern Klamath Mountain, California birds have overall darker plumage, more spotting on the belly, and are smaller than barred owls from the Coast Range (Figures S1 and S2), suggesting either local selection for this phenotype or possible introgression of spotted owl genes into the eastern Klamath Mountain population.

This is a natural, dynamic, ongoing invasion with secondary contact that can be studied across space in real time. With hybridization of these owl species, there is a potential for a loss of

biodiversity in western North America due to either replacement of the spotted owl by the barred owl or collapse of the boundaries of the two species (Huxel, 1999). Here we present an analysis of fifty-one low-coverage, whole-genome sequences from barred and spotted owls sampled across their contact zone in western North America and identify the extent of introgression between these species.

Methods

Samples

We obtained fifty-one samples from museum collections that included eleven *Strix occidentalis* samples (two samples predated contact with *S. varia*), thirty-eight *S. varia* samples (including five from eastern North America), and two samples identified by other researchers as probable hybrid *S. varia* x *occidentalis* individuals (Tables S1, S2, and S3). We mapped the samples using QGIS version 2.18.2 (Quantum GIS Development Team, 2017) with raster and vector files from Natural Earth (http://www.naturalearthdata.com; accessed 2017 Oct 1) (Supplementary Materials (SM) section 1.1).

Sequence data

We utilized whole genome sequencing data from a previous study (Hanna et al., 2017c) for our reference pre-contact *Strix occidentalis* and eastern *S. varia* samples (NCBI Sequence Read Archive (SRA) run accessions SRR4011595, SRR4011596, SRR4011597, SRR4011614, SRR4011615, SRR4011616, SRR4011617, SRR4011618, SRR4011619, and SRR4011620 for *S. occidentalis* sample CAS:ORN:98821; SRR5428115, SRR5428116, and SRR5428117 for *S. varia* sample CNHM<USA-OH>:ORNITH:B41533, hereafter referred to as CNHMB41533). We prepared whole genome libraries for fifty-one additional (i.e. non-reference) *Strix* samples using a Nextera DNA Sample Preparation Kit (Illumina) and obtained paired-end sequences from a HiSeq 2500 (Illumina) (SM section 1.2) resulting in coverage ranging from 0.02-6.41X after filtering. The raw sequences are available from the NCBI SRA (see run accessions in Table S1).

Alignment and filtering

For the sequence data of the reference samples *Strix occidentalis* CAS:ORN:98821 and *S. varia* CNHMB41533, which Hanna et al. (2017c) generated for their study, we followed the processing methods described in Hanna et al. (2017c) and used the data here in its final, processed form. For all other samples we used Trimmomatic version 0.32 (Bolger, Lohse & Usadel, 2014) to remove adapter sequences and perform quality trimming of all of the low-coverage, short-read data (SM section 1.3). We used BWA-MEM version 0.7.12-r1044 (Li, 2013) to align the processed reference and low-coverage sequences to the repeat-masked *S. o. caurina* genome "StrOccCau_1.0_nuc_masked" (Hanna et al., 2017b,c). We merged the alignments, sorted the alignments, and marked duplicate sequences using Picard version 1.104 (http://broadinstitute.github.io/picard; accessed 2017 Oct 1) (SM section 1.4.1-1.4.2). We filtered the alignment files to only retain alignments of high quality using the Genome Analysis Toolkit (GATK) version 3.4-46 PrintReads tool (McKenna et al., 2010; DePristo et al., 2011; Van der Auwera et al., 2013) (SM section 1.4.3).

Variant calling and filtering

We called variants using the GATK version 3.4-46 UnifiedGenotyper tool (McKenna et al., 2010; DePristo et al., 2011; Van der Auwera et al., 2013) with the alignment files for all samples included as simultaneous inputs (SM section 1.5.1). We used the vcf_qual_filter.sh script from SPOW-BDOW-introgression-scripts version 1.1.0 (Hanna, Henderson & Wall, 2017)

to exclude indels, and low genotyping quality sites while retaining only diallelic sites where CAS:ORN:98821 (the source of the StrOccCau_1.0_nuc_masked reference genome) was homozygous for the reference allele and CNHMB41533, the *Strix varia* reference sample, was homozygous for the alternative allele (SM section 1.6.1). Of the remaining variable sites, we excluded those with excessively high coverage [greater than the mean plus five times the standard deviation (σ), as recommended by the GATK documentation (https://software.broadinstitute.org/gatk/guide/article?id=3225; accessed 2017 Oct 1)] (SM section 1.6.2). We used DP_means_std_dev.sh from SPOW-BDOW-introgression-scripts version 1.1.0 to calculate the mean and standard deviation (σ) of the depth of coverage for each sample across the final set of variant sites.

Ancestry and diversity analyses

For each sample at each variant site, we calculated a percentage spotted owl ancestry, which was the percentage of the coverage that supported the CAS:ORN:98821 (the *Strix occidentalis* reference sequence) allele. We calculated the average and standard deviation of the spotted owl ancestry of each sample across all variant sites (SM section 1.6.3). We tested for significant differences between the average spotted owl ancestries in populations using Welch's *t*-test (Welch, 1947) as the populations had unequal numbers of samples and then applied a Bonferroni adjustment (Dunn, 1961) when we evaluated significance (SM section 1.6.4).

We estimated the probabilities of observing an introgressed region greater than 50,000 nt, 100,000 nt, or 150,000 nt in length if *Strix varia* and *S. occidentalis* hybridized in 1945, approximately the earliest date of their potential contact (Livezey, 2009a), using the formula from Racimo et al. (2015). For the recombination rate, we used 1.5 centimorgans/million nucleotides (cM/Mnt), which Backström et al. (2010) estimated for the zebra finch (*Taeniopygia guttata*). For the number of generations since the earliest potential date of hybridization, we assumed a generation time of two years (Gutiérrez, Franklin & Lahaye, 1995; Mazur & James, 2000) even though *S. o. caurina* is able to breed in its first year and others have used ten years as the generation time for *S. o. caurina* (Noon & Biles, 1990; USDA Forest Service, 1992). With that generation time, approximately thirty-five generations have potentially elapsed since the two species first contacted in 1945 and 2014, the date of our most recent sample.

In order to probe for further for evidence of introgression in the samples that did not appear as hybrids from their genome-wide average spotted owl ancestry, we attempted to identify regions that were outliers from the genome-wide ancestry average by conducting a sliding window analysis. We examined adjacent windows of 50,000 nucleotides (nt) where a sample had data for at least ten variant sites within that window and calculated the average spotted owl ancestry for the window. We assumed that, if a region was introgressed from the other species, the average should be close to 0.5. Thus, in samples with an average genome-wide ancestry close to 0, we called a window an outlier if the average spotted owl ancestry was >= 0.4. Inversely, in samples with an average genome-wide ancestry close to 1, we called a window an outlier if the average spotted owl ancestry was <= 0.6 (SM sections 1.6.5-1.6.6).

In order to estimate the genome-wide diversity harbored by *Strix varia* and *S. occidentalis* populations, we considered all diallelic variant sites (not just those fixed between our *S. varia* and *S. occidentalis* references) and calculated H_w, the number of nucleotide differences within populations, and H_b, the number of nucleotide differences between populations using the countFstPi script from SPOW-BADO-introgression-scripts (Hanna, Henderson & Wall, 2017). We also used countFstPi to calculate the fixation index (F_{ST}) (Hudson,

Slatkin & Maddison, 1992) in order to estimate the differentiation of *S. varia* and *S. occidentalis* populations (SM section 1.6.7).

Results

Our fifty-one low-coverage, whole-genome sequences included *Strix varia* and *S. occidentalis* individuals sampled both outside and across the contact zone of these two species in western North America (Figure 1). After filtering, our final set of variable sites fixed between our *S. varia* and *S. occidentalis* reference individuals included 5,816,692 sites. Except for the two putative hybrid samples that we had included as a test of our methodology, the genome-wide average spotted owl ancestry for all samples was close to either 0 or 1, indicating that they were either pure *S. varia* or *S. occidentalis*, respectively (Figure 2 and Table S4). A genome-wide average spotted owl ancestry of 0.538 confirmed the F_1 hybrid (*S. varia* x *occidentalis*) identity of a sample from Humboldt County, California. We calculated a spotted owl ancestry of 0.359 for the second hybrid sample from Benton County, Oregon, which suggested that this individual was likely a F_2 hybrid (F_1 x *S. varia* backcross).

The mean genome-wide spotted owl ancestry of the Siskiyou County *Strix varia* population was 0.0696 whereas the mean was 0.0699 for the rest of the western *S. varia* (Table S5). There was no significant difference in spotted owl ancestry between these two populations (Table S6). When we combined all *S. varia* from western North America together (0.0698 mean spotted owl ancestry) and compared their spotted owl ancestry with that of the eastern *S. varia* (0.0676 mean spotted owl ancestry), we found no significant difference in ancestry between the western and eastern *S. varia* after applying a Bonferroni adjustment (Tables S5 and S6). There was also no significant difference in spotted owl ancestry between *S. occidentalis* individuals sampled from populations not in contact with *S. varia* and those from populations already in contact with *S. varia* (mean ancestries of 0.9930 and 0.9952, respectively) (Tables S5 and S6).

The average spotted owl ancestry in the *Strix varia* samples ranged from approximately 6.55-7.28% greater than the 0% value at which our methodology set the reference *S. varia* (Table S4). The *S. occidentalis* samples ranged from approximately 0.43-0.94% less than the 100% value for the reference *S. occidentalis*. The standard deviation in the *S. varia* samples was consistently more than two times greater than the standard deviation in the *S. occidentalis* samples.

Based upon an estimate of thirty-five generations as the maximum number of generations since contact of *Strix varia* and *S. occidentalis* (Gutiérrez, Franklin & Lahaye, 1995; Mazur & James, 2000; Livezey, 2009a) and the recombination rate of *Taeniopygia guttata* (Backström et al., 2010), we estimated that the probability of observing a track > 50,000 nt resulting from hybridization during the initial contact of *S. varia* and *S. occidentalis* was 97.41%, the probability of observing a track > 100,000 nt was 94.89%, and the probability of observing a track > 150,000 nt was 92.43%.

Of the forty-nine samples for which we conducted an outlier window analysis, we detected outlier windows in thirty-nine samples (79.6%). Across all samples, we detected 316 outlier windows of length 50,000 nt, forty-one of length 100,000 nt, and only three of length 150,000 nt and none exceeded this length (Figure 3). In all samples the outlier windows represented < 1.01% of the analyzed windows. For thirty-six of the thirty-nine samples with outliers, the number of outlier windows was < 0.08% of the analyzed windows. There were three samples for which the outlier windows represented between 0.1% and 1.01% of the analyzed windows. However, the increased proportion of outlier windows in these samples appeared to be

related to exceptionally low sequence coverage as these three *Strix varia* samples had the lowest coverage (0.036-0.118X) and, consequently, the fewest number of analyzed windows of any of the samples in which we detected outlier windows (Figure S3). A *S. occidentalis* sample with 0.017X coverage was the only sample with lower coverage than those three, but our analyses did not recover any outlier windows for it.

We found little evidence of differentiation between the Siskiyou *Strix varia* and the other western *S. varia*, recovering a low F_{ST} (0.008) and very similar levels of nucleotide diversity in the two populations (Table 1). Similar levels of nucleotide diversity also exist in the *S. varia* populations from western and eastern North America. We additionally estimated a low F_{ST} value (0.051) between western and eastern *S. varia*, which suggests a low level of differentiation between these populations. *Strix occidentalis* populations pre and post-contact with *S. varia* exhibited similar levels of nucleotide diversity and appeared weakly differentiated (F_{ST} = 0.022). We estimated approximately 14X greater nucleotide diversity in *S. varia* than *S. occidentalis* and a high level of divergence (F_{ST} = 0.833) between the species.

Discussion

Our genome-wide average spotted owl ancestry analysis suggested that, apart from the F_1 and F_2 hybrid individuals, all of the other individuals appeared to be pure *Strix occidentalis* or pure *S. varia*. *Strix varia* individuals from the West did not differ in spotted owl ancestry from those from eastern North America, providing no evidence of introgression of *S. occidentalis* genomic material into *S. varia*. *Strix occidentalis* individuals sampled from populations not in contact with *S. varia* did not differ from those sampled from populations already in contact with *S. varia*, providing no evidence of introgression of *S. varia* genomic material into *S. occidentalis*.

We hypothesize that the reason that the genome-wide average spotted owl ancestry values for the *Strix varia* samples deviated more from that of the reference *S. varia* than did the *S. occidentalis* samples from their reference was due to the greater amount of genetic variation within *S. varia* (Hanna et al., 2017c). The sites fixed between our reference *S. varia* and *S. occidentalis* samples do not appear to be fixed across *S. varia* and *S. occidentalis*. Further high-coverage sequencing of whole-genomes for both species would enable better identification of fixed genetic differences between the two species.

Our sliding window analyses confirmed the results of the genome-wide ancestry averages. Although we detected windows of spotted owl ancestry in 80% of our samples that were outliers from the genome-wide spotted owl ancestry average for those samples, the outliers represented a small proportion of the total analyzed windows and we calculated a > 92% probability of observing windows longer than these if hybridization had taken place at any point over the last 50-70 years since *Strix varia* and *S. occidentalis* first came into contact in western North America (Taylor & Forsman, 1976; Livezey, 2009a). We viewed our estimated maximum number of generations since contact as a conservative estimate in terms of erring on the side of overestimating the number of generations that the two species have been in contact. Even with this conservative estimate, there is a > 97% probability of introgressed regions being larger than the 50,000 nt windows that we used to check for potential introgression and a > 92% probability of the introgressed regions being larger than the 150,000 nt length of the longest outlier window that we detected with our sliding window analysis. Of the samples with outlier windows, we detected the highest proportion of outliers in the three samples with the lowest sequence coverage. We hypothesize that the outlier window proportions in these samples would be reduced with higher coverage sequencing data.

Hybridization of *Strix varia* and *S. occidentalis* has previously been investigated using a set of four microsatellite (Funk et al., 2007) and fourteen amplified fragment-length polymorphism (Haig et al., 2004) markers, which the authors found useful for diagnosing F_1 and F_2 hybrids (Haig et al., 2004; Funk et al., 2007). These sets of markers are not as effective in identifying introgressed genomic regions resulting from hybridization that took place more than two generations in the past. Our whole-genome approach allowed us to sample more than five million sites spread throughout the entire genome. With this quantity of markers, we are confident that we should have been able to detect any introgression that has taken place over the last 50-70 years that *S. varia* and *S. occidentalis* have been in contact in western North America (Taylor & Forsman, 1976; Livezey, 2009a).

When a taxon expands its range into that of a closely related species with which it readily hybridizes, theory predicts asymmetric introgression into the invading taxon and that the greatest proportion of introgressed genomic material will occur in populations on the leading edge of the invasion (Currat et al., 2008; Excoffier, Foll & Petit, 2009). Coupled with geographic history of *Strix varia*'s expansion into western North America and through *S. occidentalis* populations from British Columbia into southern California (Taylor & Forsman, 1976; Haig et al., 2004; Livezey, 2009a), this theoretical prediction suggests that we might expect *S. varia* individuals in the southern portion of the zone of sympatry of *S. varia* and *S. occidentalis* to exhibit the highest proportion of introgressed genomic material. It was with this prediction in mind that we focused our sampling on California *S. varia* populations (Figure 1). In addition, we sampled a western *S. varia* population in Siskiyou County, California that was morphologically anomalous. It is notable that we found no evidence of admixture even though we sampled the populations we expected to contain the highest degree of introgressed genomic material. Range expansion simulations suggest that we should predict asymmetric introgression into *S. varia* even when there is a rate of hybridization of < 2% (Currat & Excoffier, 2011). Coupled with these predictions, our findings suggest that, although hybridization between *S. varia* and *S. occidentalis* occurs, it has either been vanishingly rare on the edge of the *S. varia* expansion wave or some other process is occurring to counter the expected widespread introgression of *S. occidentalis* genetic material into *S. varia*.

Our estimation of a more than 10X greater nucleotide diversity in *Strix varia* than in *S. occidentalis* and of a high F_{ST} between the species closely matched the results of an analysis of whole-genome sequences of one individual per species done by Hanna et al. (2017c), which is unsurprising as we included the high coverage data from those two individuals along with our low-coverage whole-genome sequences in this study. Our estimation of similar levels of nucleotide diversity in the Siskiyou *S. varia* population compared with the population including the other western *S. varia* samples is consistent with our finding no difference in spotted owl ancestry between these populations. Similarly, *S. occidentalis* populations pre and post-contact with *S. varia* exhibited similar levels of nucleotide diversity, appeared weakly differentiated, and did not differ in spotted owl ancestry.

More surprising, however, was our estimation of similar levels of nucleotide diversity in western and eastern North American *Strix varia* populations. We expected the western *S. varia* populations to harbor reduced genetic diversity compared to the eastern *S. varia* after having been subjected to successive founder effects and corresponding reductions in nucleotide diversity (Austerlitz et al., 1997). The similarity of the western *S. varia* nucleotide diversity to that of the eastern population suggests that the *S. varia* range expansion has proceeded with minimal to no reduction in genetic diversity.

Simulations have suggested that long-distance dispersal by individuals of a species undergoing a range expansion can inhibit a loss of genetic diversity in the newly formed populations on the edge of the range (Ray & Excoffier, 2010). Engler et al. (2016) suggested this as an explanation for the retention of genetic diversity in an Old World warbler, *Hippolais polyglotta*, experiencing a range expansion. Recent simulations have also suggested that long-distance dispersal in an invading taxon can counteract introgression of local genetic material into the invader by inhibiting the surfing of introgressed genetic regions (Amorim et al., 2017). Livezey (2009b) reported the mean dispersal distance of *Strix varia* as 41.3 km, but mentioned that some individuals have dispersed as far as 488.1 km. Even if long-distance dispersal has only been occurring at low levels during the *S. varia* range expansion, then, in addition to explaining the lack of reduction in genetic diversity in western *S. varia*, this could also account for the lack of large-scale introgression of *S. occidentalis* genetic material into western *S. varia* populations (Ray & Excoffier, 2010; Amorim et al., 2017). Long-distance dispersal would have been especially capable of countering introgression of *S. occidentalis* material if non-introgressed *S. varia* have been dispersing to the front of the expansion wave (Amorim et al., 2017). Long-distance dispersal may also simply be a component of a high rate of intraspecific gene flow in western *S. varia*, which could be inhibiting the loss of *S. varia* genetic diversity and countering introgression of *S. occidentalis* genetic material (Ray, Currat & Excoffier, 2003; Currat et al., 2008; Petit & Excoffier, 2009).

Although our results provide genomic confirmation that hybridization and backcrossing does occur, we found no evidence of widespread admixture between *Strix varia* and *S. occidentalis* in western North America. The distinctive plumage *S. varia* collected in Siskiyou County, California does not appear to be a result of hybridization with *S. occidentalis*. We propose that plumage patterning that seems to have characters intermediate between *S. varia* and *S. occidentalis* may not be indicative of hybridization. Previous investigators have issued similar cautionary statements after their genetic studies of hybridization in these taxa (Haig et al., 2004; Funk et al., 2007). Despite quantification being needed, the plumages of western *S. varia* appear to be quite variable. The lack of spotted owl ancestry in these oddly plumaged western *S. varia* suggests that some western *S. varia* may be undergoing local selection, which has affected plumage and size, but further study is needed. Although our low-coverage data suggest that the divergence between western of eastern North American *S. varia* populations is low, higher-coverage sequence data will enable us to obtain a better grasp of the genomics of the *S. varia* expansion and aid us in understanding the genetic distinctness (or lack thereof) of western barred owls. Coupled with demographic studies (Kelly, Forsman & Anthony, 2003; Dugger et al., 2015; Diller et al., 2016), our results indicate that the expansion of *S. varia* into the range of *S. occidentalis* in western North America is largely following a pattern of pure replacement, rather than inducing extinction through hybridization and introgression (Rhymer & Simberloff, 1996). It seems unlikely that even introgressed remnants of the *S. occidentalis* genome will remain in areas in contact with *S. varia* if *S. occidentalis* is not able to persist.

Tables

Table 1. Nucleotide diversity and fixation index statistics calculated for various population comparisons. The π_{Within} statistic signifies the average number of pairwise differences between two individuals sampled from the same population. The $\pi_{Between}$ statistic denotes the average number of pairwise differences between two individuals sampled from different populations (Populations 1 and 2). "Pop 1" and "Pop 2" refer to Population 1 or 2 from columns 1 and 2, respectively. The "All Western Barred Owls" population is a combination of the "Western Barred Owls" and "Siskiyou Barred Owls" populations. The "Spotted Owls (pre-contact)" and "Spotted Owls (post)" populations indicate *Strix occidentalis* from populations not in contact or in contact with *S. varia*, respectively.

Population 1	Population 2	π_{Within} Pop 1	π_{Within} Pop 2	$\pi_{Between}$	F_{ST}
Western Barred Owls	Siskiyou Barred Owls	2.097E-03	2.068E-03	2.100E-03	0.008
Western Barred Owls	Eastern Barred Owls	2.119E-03	2.228E-03	2.291E-03	0.051
Siskiyou Barred Owls	Eastern Barred Owls	2.066E-03	2.203E-03	2.259E-03	0.055
All Western Barred Owls	Eastern Barred Owls	2.128E-03	2.242E-03	2.301E-03	0.051
All Barred Owls	All Spotted Owls	2.202E-03	1.572E-04	7.052E-03	0.833
Spotted Owls (pre-contact)	Spotted Owls (post)	1.073E-04	9.998E-05	1.060E-04	0.022

Figures

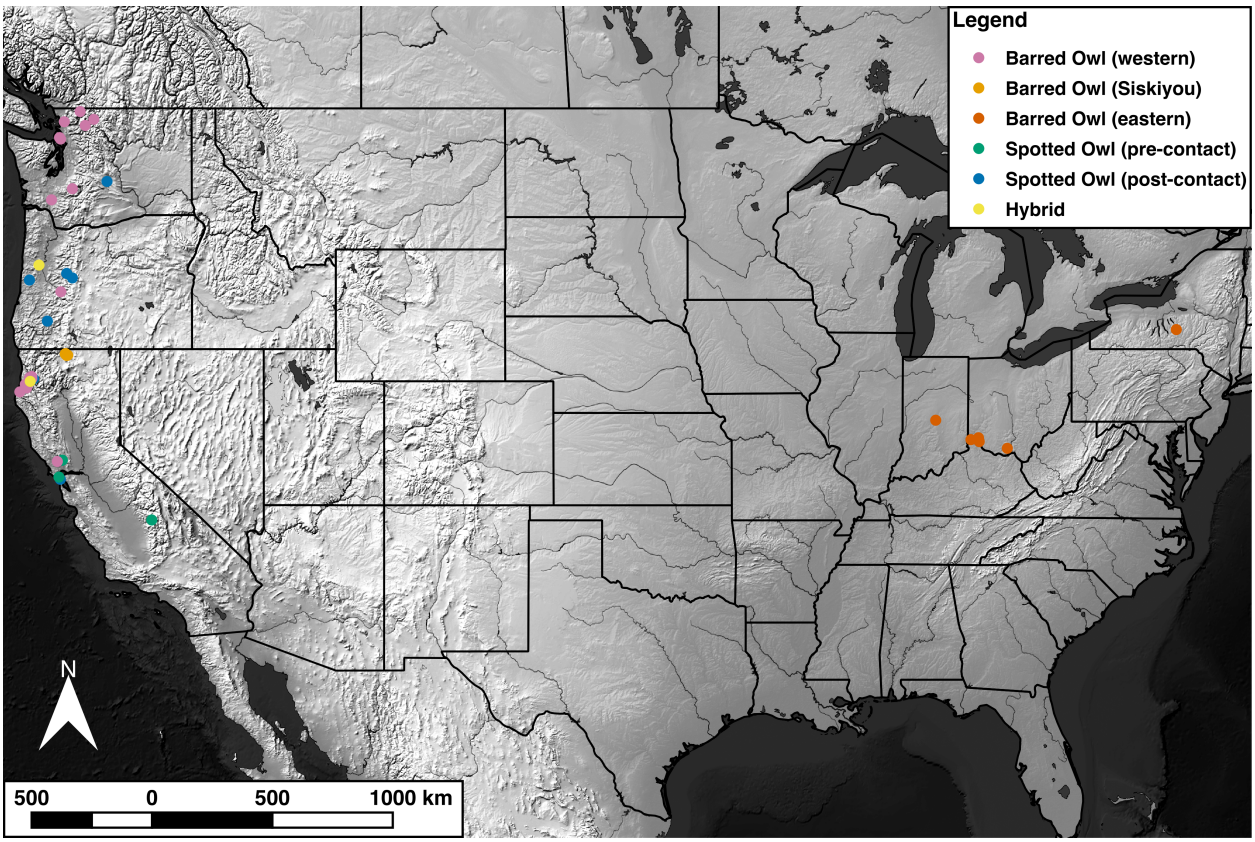

Figure 1. Map of *Strix* samples.

This map displays the sampling locations of all of the *Strix* specimens included in this study.

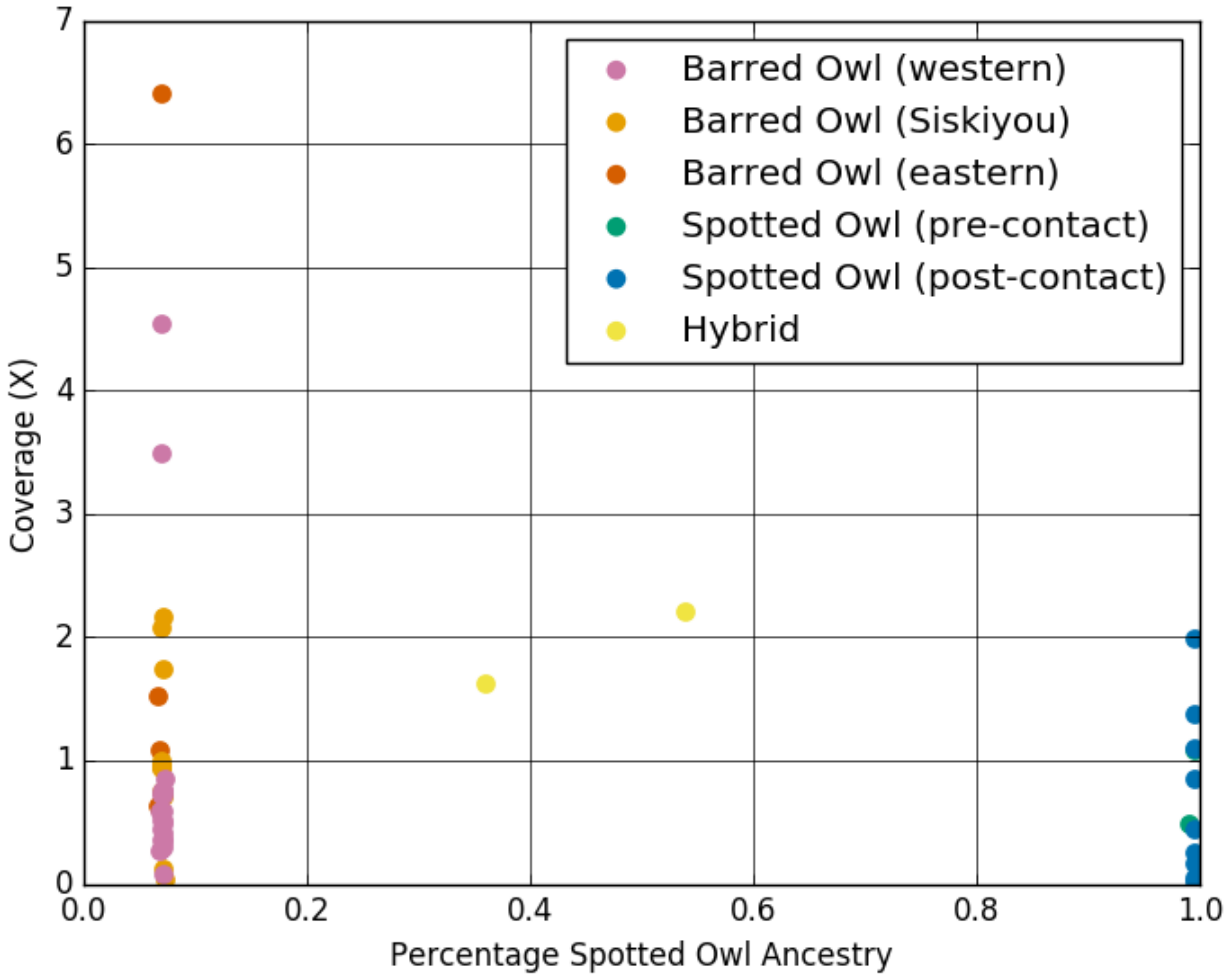

Figure 2. Plot of coverage versus genome-wide average spotted owl ancestry.

The average spotted owl ancestry of all of the samples for which we collected low-coverage, whole-genome sequence data. We plotted DNA sequence coverage on the y-axis to display that the average percentage of spotted owl (*Strix occidentalis*) ancestry was independent of the amount of coverage for a given sample.

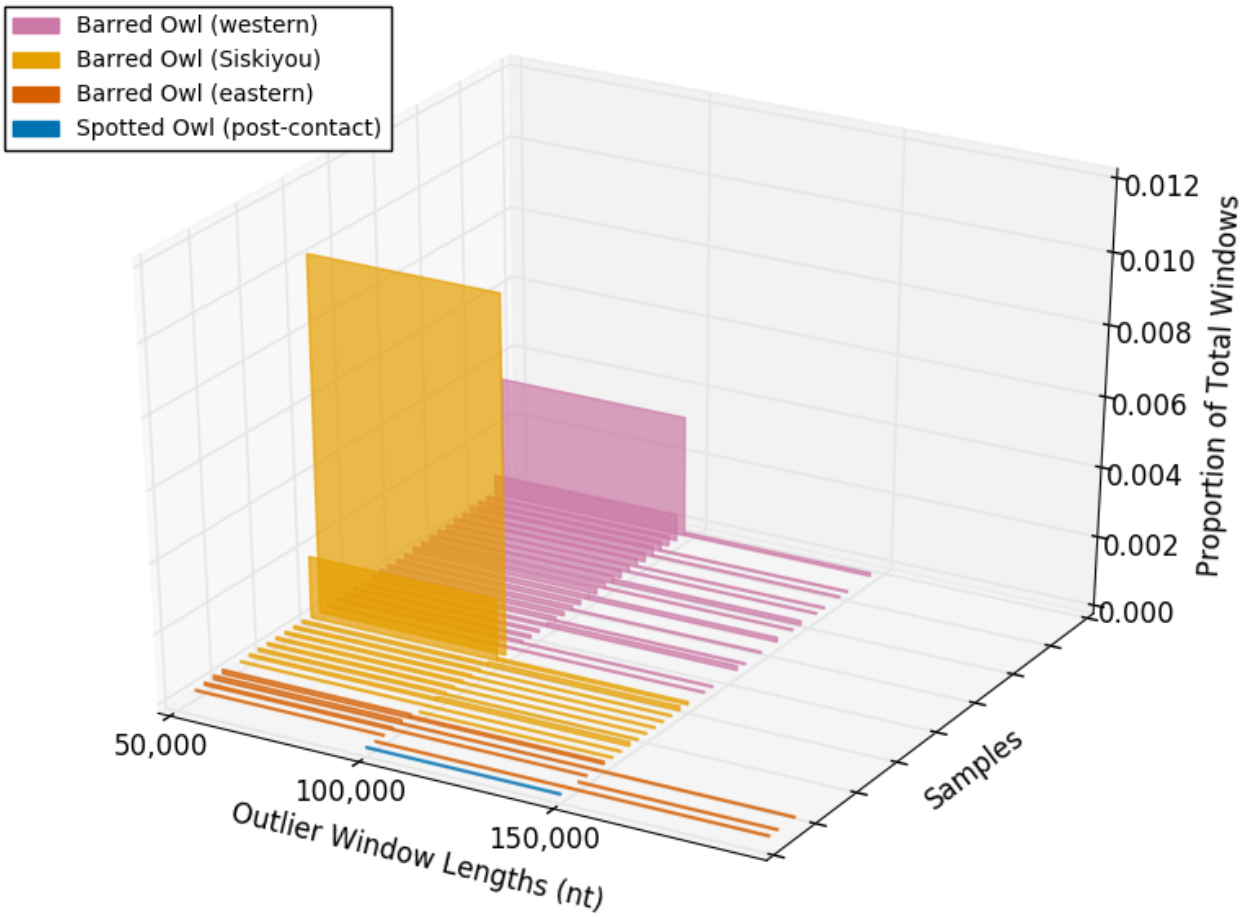

Figure 3. Plot of outlier window proportion versus outlier window length for each sample. The x-axis plots the lengths of outlier windows in increments of 50,000 nucleotides (nt). The y-axis displays the number of outlier windows of a given length as a proportion of all analyzed windows. The z-axis separates individual samples, which we grouped by population.

References

Amorim CEG., Hofer T., Ray N., Foll M., Ruiz-Linares A., Excoffier L. 2017. Long-distance dispersal suppresses introgression of local alleles during range expansions. *Heredity* 118:135–142. DOI: 10.1038/hdy.2016.68.

Anderson DR., Burnham KP. 1992. Demographic analysis of Northern Spotted Owl populations. In Final Draft Recovery Plan for the Northern Spotted Owl, Volume 2. USDI Fish and Wildlife Service, Region 1, Portland, OR, USA. p. 66–75.

Austerlitz F., Jung-Muller B., Godelle B., Gouyon P-H. 1997. Evolution of Coalescence Times, Genetic Diversity and Structure during Colonization. *Theoretical Population Biology* 51:148–164. DOI: 10.1006/tpbi.1997.1302.

Backström N., Forstmeier W., Schielzeth H., Mellenius H., Nam K., Bolund E., Webster MT., Öst T., Schneider M., Kempenaers B., Ellegren H. 2010. The recombination landscape of the zebra finch *Taeniopygia guttata* genome. *Genome Research* 20:485–495. DOI: 10.1101/gr.101410.109.

Bolger AM., Lohse M., Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. DOI: 10.1093/bioinformatics/btu170.

Currat M., Excoffier L. 2011. Strong reproductive isolation between humans and Neanderthals inferred from observed patterns of introgression. *Proceedings of the National Academy of Sciences* 108:15129–15134. DOI: 10.1073/pnas.1107450108.

Currat M., Ruedi M., Petit RJ., Excoffier L. 2008. THE HIDDEN SIDE OF INVASIONS: MASSIVE INTROGRESSION BY LOCAL GENES. *Evolution* 62:1908–1920. DOI: 10.1111/j.1558-5646.2008.00413.x.

Dark SJ., Gutiérrez RJ., Gould GI Jr. 1998. The Barred Owl (*Strix varia*) Invasion in California. *The Auk* 115:50–56. DOI: 10.2307/4089110.

Davis RJ., Hollen B., Hobson J., Gower JE., Keenum D. 2016. Northwest Forest Plan—the first 20 years (1994-2013): status and trends of northern spotted owl habitats. [Accessed 2016 Oct 7]. Available from: http://www.treesearch.fs.fed.us/pubs/50567.

DePristo MA., Banks E., Poplin R., Garimella KV., Maguire JR., Hartl C., Philippakis AA., del Angel G., Rivas MA., Hanna M., McKenna A., Fennell TJ., Kernytsky AM., Sivachenko AY., Cibulskis K., Gabriel SB., Altshuler D., Daly MJ. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* 43:491–498. DOI: 10.1038/ng.806.

Diller LV., Hamm KA., Early DA., Lamphear DW., Dugger KM., Yackulic CB., Schwarz CJ., Carlson PC., McDonald TL. 2016. Demographic response of northern spotted owls to barred owl removal. *The Journal of Wildlife Management* 80:691–707. DOI: 10.1002/jwmg.1046.

Dugger KM., Forsman ED., Franklin AB., Davis RJ., White GC., Schwarz CJ., Burnham KP., Nichols JD., Hines JE., Yackulic CB., Doherty PF., Bailey L., Clark DA., Ackers SH., Andrews LS., Augustine B., Biswell BL., Blakesley J., Carlson PC., Clement MJ., Diller LV., Glenn EM., Green A., Gremel SA., Herter DR., Higley JM., Hobson J., Horn RB., Huyvaert KP., McCafferty C., McDonald T., McDonnell K., Olson GS., Reid JA., Rockweit J., Ruiz V., Saenz J., Sovern SG. 2015. The effects of habitat, climate, and Barred Owls on long-term demography of Northern Spotted Owls. *The Condor* 118:57–116. DOI: 10.1650/CONDOR-15-24.1.

Dunn OJ. 1961. Multiple Comparisons among Means. *Journal of the American Statistical Association* 56:52–64. DOI: 10.1080/01621459.1961.10482090.

Engler JO., Secondi J., Dawson DA., Elle O., Hochkirch A. 2016. Range expansion and retraction along a moving contact zone has no effect on the genetic diversity of two passerine birds. *Ecography* 39:884–893. DOI: 10.1111/ecog.01520.

Excoffier L., Foll M., Petit RJ. 2009. Genetic Consequences of Range Expansions. *Annual Review of Ecology, Evolution, and Systematics* 40:481–501. DOI: 10.1146/annurev.ecolsys.39.110707.173414.

Fitzpatrick BM., Johnson JR., Kump DK., Shaffer HB., Smith JJ., Voss SR. 2009. Rapid fixation of non-native alleles revealed by genome-wide SNP analysis of hybrid tiger salamanders. *BMC Evolutionary Biology* 9:176. DOI: 10.1186/1471-2148-9-176.

Fitzpatrick BM., Johnson JR., Kump DK., Smith JJ., Voss SR., Shaffer HB. 2010. Rapid spread of invasive genes into a threatened native species. *Proceedings of the National Academy of Sciences* 107:3606–3610. DOI: 10.1073/pnas.0911802107.

Forsman ED., Meslow EC., Wight HM. 1984. Distribution and Biology of the Spotted Owl in Oregon. *Wildlife Monographs*:3–64.

Funk WC., Mullins TD., Forsman ED., Haig SM. 2007. Microsatellite loci for distinguishing spotted owls (*Strix occidentalis*), barred owls (*Strix varia*), and their hybrids. *Molecular Ecology Notes* 7:284–286. DOI: 10.1111/j.1471-8286.2006.01581.x.

Gutiérrez RJ., Franklin AB., Lahaye WS. 1995. Spotted Owl (*Strix occidentalis*). *The Birds of North America Online* (A. Poole, Ed.). Ithaca: Cornell Lab of Ornithology. [Accessed 2016 Oct 1]. Retrieved from the Birds of North America Online: https://birdsna.org/Species-Account/bna/species/spoowl. DOI: 10.2173/bna.179.

Haig SM., Mullins TD., Forsman ED., Trail PW., Wennerberg L. 2004. Genetic identification of spotted owls, barred owls, and their hybrids: legal implications of hybrid identity. *Conservation Biology* 18:1347–1357.

Hanna ZR., Henderson JB., Sellas AB., Fuchs J., Bowie RCK., Dumbacher JP. 2017a. Complete mitochondrial genome sequences of the northern spotted owl (*Strix occidentalis caurina*) and the barred owl (*Strix varia*; Aves: Strigiformes: Strigidae) confirm the presence of a duplicated control region. *PeerJ* 5:e3901. DOI: 10.7717/peerj.3901.

Hanna ZR., Henderson JB., Wall JD. 2017. SPOW-BDOW-introgression-scripts. Version 1.1.0. *Zenodo*. DOI: 10.5281/zenodo.1065074.

Hanna ZR., Henderson JB., Wall JD., Emerling CA., Fuchs J., Runckel C., Mindell DP., Bowie RCK., DeRisi JL., Dumbacher JP. 2017b. Supplemental dataset for Northern Spotted Owl (*Strix occidentalis caurina*) genome assembly version 1.0. *Zenodo*. DOI: 10.5281/zenodo.822859.

Hanna ZR., Henderson JB., Wall JD., Emerling CA., Fuchs J., Runckel C., Mindell DP., Bowie RCK., DeRisi JL., Dumbacher JP. 2017c. Northern Spotted Owl (*Strix occidentalis caurina*) Genome: Divergence with the Barred Owl (*Strix varia*) and Characterization of Light-Associated Genes. *Genome Biology and Evolution* 9:2522–2545. DOI: 10.1093/gbe/evx158.

Hudson RR., Slatkin M., Maddison WP. 1992. Estimation of Levels of Gene Flow from DNA Sequence Data. *Genetics* 132:583–589.

Huxel GR. 1999. Rapid displacement of native species by invasive species: effects of hybridization. *Biological Conservation* 89:143–152. DOI: 10.1016/S0006-3207(98)00153-0.

Kelly EG., Forsman ED. 2004. Recent Records of Hybridization Between Barred Owls (*Strix varia*) and Northern Spotted Owls (*S. occidentalis caurina*). *The Auk* 121:806–810. DOI: 10.1642/0004-8038(2004)121[0806:RROHBB]2.0.CO;2.

Kelly EG., Forsman ED., Anthony RG. 2003. Are barred owls displacing spotted owls? *The Condor* 105:45–53. DOI: 10.1650/0010-5422(2003)105[45:ABODSO]2.0.CO;2.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv:1303.3997 Q-Bio. [Accessed 2016 Feb 16]. Available from: http://arxiv.org/abs/1303.3997.

Livezey KB. 2009a. Range Expansion of Barred Owls, Part I: Chronology and Distribution. *The American Midland Naturalist* 161:49–56. DOI: 10.1674/0003-0031-161.1.49.

Livezey KB. 2009b. Range Expansion of Barred Owls, Part II: Facilitating Ecological Changes. *The American Midland Naturalist* 161:323–349. DOI: 10.1674/0003-0031-161.2.323.

Mazur KM., James PC. 2000. Barred Owl (*Strix varia*). *The Birds of North America Online* (A. Poole, Ed.) Ithaca: Cornell Lab of Ornithology. [Accessed 2016 Oct 1]. Retrieved from the Birds of North America Online: https://birdsna.org/Species-Account/bna/species/brdowl. DOI: 10.2173/bna.508.

McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernytsky A., Garimella K., Altshuler D., Gabriel S., Daly M., DePristo MA. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20:1297–1303. DOI: 10.1101/gr.107524.110.

Muhlfeld CC., Kovach RP., Jones LA., Al-Chokhachy R., Boyer MC., Leary RF., Lowe WH., Luikart G., Allendorf FW. 2014. Invasive hybridization in a threatened species is accelerated by climate change. *Nature Climate Change* 4:620–624. DOI: 10.1038/nclimate2252.

Noon BR., Biles CM. 1990. Mathematical Demography of Spotted Owls in the Pacific Northwest. *The Journal of Wildlife Management* 54:18–27. DOI: 10.2307/3808895.

Parmesan C. 2006. Ecological and Evolutionary Responses to Recent Climate Change. *Annual Review of Ecology, Evolution, and Systematics* 37:637–669.

Parmesan C., Ryrholm N., Stefanescu C., Hill JK., Thomas CD., Descimon H., Huntley B., Kaila L., Kullberg J., Tammaru T., Tennent WJ., Thomas JA., Warren M. 1999. Poleward shifts in geographical ranges of butterfly species associated with regional warming. *Nature* 399:579–583. DOI: 10.1038/21181.

Petit RJ., Excoffier L. 2009. Gene flow and species delimitation. *Trends in Ecology & Evolution* 24:386–393. DOI: 10.1016/j.tree.2009.02.011.

Quantum GIS Development Team 2017. Quantum GIS Geographic Information System. Open Source Geospatial Foundation Project. [Accessed 2017 Sep 16]. Available from: http://qgis.org.

Racimo F., Sankararaman S., Nielsen R., Huerta-Sánchez E. 2015. Evidence for archaic adaptive introgression in humans. *Nature Reviews Genetics* 16:359–371. DOI: 10.1038/nrg3936.

Ray N., Currat M., Excoffier L. 2003. Intra-Deme Molecular Diversity in Spatially Expanding Populations. *Molecular Biology and Evolution* 20:76–86. DOI: 10.1093/molbev/msg009.

Ray N., Excoffier L. 2010. A first step towards inferring levels of long-distance dispersal during past expansions. *Molecular Ecology Resources* 10:902–914. DOI: 10.1111/j.1755-0998.2010.02881.x.

Rhymer JM., Simberloff D. 1996. Extinction by Hybridization and Introgression. *Annual Review of Ecology and Systematics* 27:83–109. DOI: 10.1146/annurev.ecolsys.27.1.83.

Rieseberg LH., Kim S-C., Randell RA., Whitney KD., Gross BL., Lexer C., Clay K. 2007. Hybridization and the colonization of novel habitats by annual sunflowers. *Genetica* 129:149–165. DOI: 10.1007/s10709-006-9011-y.

Taylor AL., Forsman ED. 1976. Recent range extensions of the Barred Owl in western North America, including the first records for Oregon. *Condor* 78:560–561.

Thomas JW., Forsman ED., Lint JB., Meslow EC., Noon BR., Verner J. 1990. A conservation strategy for the northern spotted owl: report of the Interagency Scientific Committee to address the conservation of the northern spotted owl. USDA Forest Service; USDI Bureau of Land Management, Fish and Wildlife Service, National Park Service: Portland, Oregon.

USDA Forest Service 1992. Final Environmental Impact Statement on Management for the Northern Spotted Owl in the National Forests. USDA Forest Service, National Forest System: Portland, Oregon. 2 vol.

Van der Auwera GA., Carneiro MO., Hartl C., Poplin R., del Angel G., Levy-Moonshine A., Jordan T., Shakir K., Roazen D., Thibault J., Banks E., Garimella KV., Altshuler D., Gabriel S., DePristo MA. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics* 11:11.10.1-11.10.33. DOI: 10.1002/0471250953.bi1110s43.

Welch BL. 1947. The Generalization of "Student's" Problem When Several Different Population Variances Are Involved. *Biometrika* 34:28–35. DOI: 10.1093/biomet/34.1-2.28.

Chapter 3 Supplementary Article

1 Supplementary Material and Methods

1.1 Mapping

1.1.1 We created a map of the samples in QGIS version 2.18.2 (Quantum GIS Development Team, 2017) using the high resolution (21,600 x 10,800 pixels), 1:10 million-scale Gray Earth with Shaded Relief, Hypsography, Ocean Bottom, and Drainages version 2.1.0 raster file from Natural Earth (http://www.naturalearthdata.com; accessed 2017 Oct 1) as the base map layer. We overlaid this with the 1:50 million-scale Admin 1 - States, Provinces boundaries version 3.0.0 vector file from Natural Earth and then plotted the coordinates of our samples.

1.2 Sequence data

1.2.1 We prepared libraries and sequenced the samples in two indexed pools, which we will refer to as "Sample Set 1" and "Sample Set 2" (see Table S1 for the samples included in each sample set).

1.2.2 **Sample Set 1 - *Strix varia* sample CAS:ORN:95964.** We extracted genomic DNA from CAS:ORN:95964 using a DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany). We used 50 ng genomic DNA to prepare a whole-genome library using a Nextera DNA Sample Preparation Kit (Illumina, San Diego, California). After tagmentation, we cleaned the reaction with a DNA Clean & Concentrator -5 kit (Zymo Research, Irvine, California). We amplified the reaction with five cycles of PCR using a KAPA Library Amplification kit (KAPA Biosystems, Wilmington, Massachusetts) and then cleaned the reaction with a DNA Clean & Concentrator -5 kit (Zymo Research, Irvine, California). We used Dye-Free, 1.5% agarose, 250-1,500 base pair (bp) cassette on a BluePippin (Sage Science, Beverly, Massachusetts) to select library fragments in the size range of 534-634 bp, which, after subtracting the 134 bp of adapters, corresponded to selecting an average insert size of 450 bp. We next performed a real-time PCR (rtPCR) using a KAPA Real-Time Library Amplification Kit (KAPA Biosystems, Wilmington, Massachusetts) on a CFX96 Touch Real-Time PCR Detection System (Bio-Rad, Hercules, California) to further amplify the library with nine cycles PCR. We then cleaned the PCR products with a DNA Clean & Concentrator -5 kit (Zymo Research, Irvine, California). We assessed the library fragment size distribution with a 2100 BioAnalyzer (Agilent Technologies, Santa Clara, California) and the concentration of double-stranded DNA material with a Qubit 2.0 Flurometer (Invitrogen, Carlsbad, California). Due to the presence of small peaks in the BioAnalyzer trace, we further cleaned the library using 0.6X Agencourt AMPure XP (Beckman Coulter, Brea, California) magnetic beads. We then reassessed the concentration of double-stranded DNA material with a Qubit 2.0 Flurometer (Invitrogen, Carlsbad, California) and the library fragment size distribution with a 2100 BioAnalyzer (Agilent Technologies, Santa Clara, California), which revealed that the average fragment size of the pool was 583 nucleotides (nt).

1.2.3 **Sample Set 1 - sixteen additional samples.** For each sample, we used 10 ng genomic DNA to prepare a whole-genome library using a Nextera DNA Sample Preparation Kit (Illumina, San Diego, California). After tagmentation, we cleaned the reaction with a DNA Clean & Concentrator -5 kit (Zymo Research, Irvine, California). We amplified the reaction and added Illumina indexed adapters with five cycles of PCR using a KAPA

Library Amplification kit (KAPA Biosystems, Wilmington, Massachusetts) and then cleaned the reaction with a DNA Clean & Concentrator -5 kit (Zymo Research, Irvine, California). We assessed the concentration of double-stranded DNA material with a Qubit 2.0 Flurometer (Invitrogen, Carlsbad, California), combined the library into an equimolar sixteen-sample pool, and then concentrated the pool with a DNA Clean & Concentrator -5 kit (Zymo Research, Irvine, California). We used a BluePippin (Sage Science, Beverly, Massachusetts) to select library fragments in the size range of 550-750 nt, which, after subtracting the 134 nt of adapters, corresponded to selecting an average insert size of 516 nt. We then performed a real-time PCR (rtPCR) using a KAPA Real-Time Library Amplification Kit (KAPA Biosystems, Wilmington, Massachusetts) on a CFX96 Touch Real-Time PCR Detection System (Bio-Rad, Hercules, California) to amplify the pool with nine cycles of PCR. We then cleaned the PCR products with a DNA Clean & Concentrator -5 kit (Zymo Research, Irvine, California). We assessed the concentration of double-stranded DNA material with a Qubit 2.0 Flurometer (Invitrogen, Carlsbad, California) and the library fragment size distribution with a 2100 BioAnalyzer (Agilent Technologies, Santa Clara, California), which revealed that the average fragment size of the pool was 607 nt.

1.2.4 **Sample Set 1 - final pool.** We pooled the CAS:ORN:95964 library in an equimolar ratio with the equimolar pool of the sixteen other libraries. We then sequenced the final pool across both lanes of a two-lane flow cell with a HiSeq PE Rapid Cluster Kit and a 200 cycle HiSeq Rapid SBS Kit v1 on a HiSeq 2500 (Illumina, San Diego, California). The raw sequences are available from the NCBI Sequence Read Archive (SRA) in the run accessions indicated in Table S1.

1.2.5 **Sample Set 2.** For each sample, we used 10 ng genomic DNA to prepare a whole-genome library using a Nextera DNA Sample Preparation Kit (Illumina, San Diego, California). After tagmentation, we cleaned the reaction with a DNA Clean & Concentrator -5 kit (Zymo Research, Irvine, California). We amplified the reaction and added Illumina indexed adapters with five cycles of PCR using a KAPA Library Amplification kit (KAPA Biosystems, Wilmington, Massachusetts) and then cleaned the reaction with a DNA Clean & Concentrator -5 kit (Zymo Research, Irvine, California). We assessed the concentration of double-stranded DNA material with a Qubit 2.0 Flurometer (Invitrogen, Carlsbad, California), combined the indexed library into a thirty-six-sample pool, and then concentrated the pool with a DNA Clean & Concentrator -5 kit (Zymo Research, Irvine, California). We used a BluePippin (Sage Science, Beverly, Massachusetts) to select fragments in the size range of 500-700 nt, which, after subtracting the 134 nt of adapters, corresponded to selecting an average insert size of 466 nt. We cleaned the BluePippin products with 0.6X Agencourt AMPure XP (Beckman Coulter, Brea, California) magnetic beads and then performed a real-time PCR (rtPCR) using a KAPA Real-Time Library Amplification Kit (KAPA Biosystems, Wilmington, Massachusetts) on a CFX96 Touch Real-Time PCR Detection System (Bio-Rad, Hercules, California) to amplify the pool with eight cycles of PCR. We then cleaned the PCR products with a DNA Clean & Concentrator -5 kit (Zymo Research, Irvine, California). We assessed the concentration of double-stranded DNA material with a Qubit 2.0 Flurometer (Invitrogen, Carlsbad, California) and the fragment size distribution with a 2100 BioAnalyzer (Agilent Technologies, Santa Clara, California), which indicated that the average fragment size of the pool was 579 nt. We sequenced the pool

on two successive runs of 150 nt paired-end sequencing using a two-lane flow cell with a HiSeq PE Rapid Cluster Kit and a 300 cycle HiSeq Rapid SBS Kit v1 on a HiSeq 2500 (Illumina, San Diego, California) in rapid mode. We obtained sequencing data from each of the two flow cell lanes on the first run. We obtained data from a portion of one of the two flow cell lanes on the second run. The raw sequences from the Sample Set 2 samples are available from the NCBI SRA in the run accessions indicated in Table S1.

1.3 Sequence data processing

1.3.1 We performed adapter and quality trimming of the low-coverage sequence data using Trimmomatic version 0.32 (Bolger, Lohse & Usadel, 2014) with the following options: "ILLUMINACLIP:<fasta of Illumina adapter sequences>:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:28 MINLEN:36".

1.4 Alignment and filtering

1.4.1 For all samples, in order to align trimmed paired and unpaired reads to "StrOccCau_1.0_nuc_masked" (Hanna et al., 2017a,b) we used bwa mem version 0.7.12-r1044 (Li, 2013) with default options other than parameters "bwa mem -M". We separately aligned paired-end and unpaired reads. For alignment of the paired-end reads, we set the insert size to be equal to the size estimate of the final library given by the 2100 BioAnalyzer (Agilent Technologies, Santa Clara, California) minus the length of the adapters (insert sizes for CAS:ORN:98821 and CNHM<USA-OH>:ORNITH:B41533 obtained from Hanna et al. (2017a,b), 446 nt insert size used for CAS:ORN:95964, 481 nt insert size for the rest of Sample Set 1, and 466 nt insert size for Sample Set 2). Additionally, for the alignment of the paired-end reads we set the parameter "-w", the maximum insert size, equal to 1000.

1.4.2 We used the Picard version 1.104 function MergeSamFiles (http://broadinstitute.github.io/picard) with default settings to merge the paired-end and unpaired sequence alignments and then used the Picard version 1.104 function SortSam (http://broadinstitute.github.io/picard) with default settings to sort the alignments. We used the Picard version 1.104 function MarkDuplicates (http://broadinstitute.github.io/picard) with default settings to mark duplicate sequences (both PCR and optical).

1.4.3 We used the Genome Analysis Toolkit (GATK) version 3.4-46 PrintReads tool (McKenna et al., 2010; DePristo et al., 2011; Van der Auwera et al., 2013) with the parameters "--read_filter BadCigar --read_filter BadMate --read_filter UnmappedRead --read_filter NotPrimaryAlignment --read_filter FailsVendorQualityCheck --read_filter DuplicateRead --read_filter MappingQualityUnavailable" to filter the bam files to only retain the high quality alignments. Alignments with these flags are ignored by the GATK SNP discovery tools, but since reads with these flags can still contribute to the DepthPerAlleleBySample (depth of coverage of each allele per sample) field in the variant call format files output by the GATK tools (GATK Dev Team, 2017) and we intended to use this field downstream for purposes where this extra coverage may be misleading, we performed this filtering of the bam files.

1.5 SNP calling

1.5.1 We used the GATK version 3.4-46 UnifiedGenotyper tool (McKenna et al., 2010; DePristo et al., 2011; Van der Auwera et al., 2013) to call SNPs using all of the filtered bam files as simultaneous inputs and employing default options other than setting "--output_mode EMIT_ALL_SITES".

1.5.2 There is a sample included in the variant call format (vcf) file output by UnifiedGenotyper for which we do not report any results. We initially included this sample that was provided by another research group as a potential hybrid to test. Our results suggested that the sample was not what it was originally purported to be. It was clear that it was a *Strix occidentalis* sample, but, as we did not know the geographic origin or date of collection of the sample, we decided to drop it from our analyses. We provide this information as explanation for its presence in the vcf file.

1.6 Filtering and spotted owl ancestry analyses

1.6.1 We used vcf_qual_filter.sh from SPOW-BDOW-introgression-scripts version 1.1.0 (Hanna, Henderson & Wall, 2017) to retain only biallelic sites sites where CAS:ORN:98821 (the source of the StrOccCau_1.0_nuc_masked reference genome) was homozygous for the reference allele and CNHMB41533, the *S. varia* reference sample, was homozygous for the alternative allele. We used this script also to exclude indels. We only retained sites where the Phred-scaled probability that a polymorphism exists was >50 and the Phred-scaled genotype quality was >=30 for both CAS:ORN:98821 and CNHMB41533. We required that CNHMB41533 had zero reads that supported the CAS:ORN:98821 allele at each retained variant site. We also required that CAS:ORN:98821 had zero reads that supported the CNHMB41533 allele and >=10 reads in support of the CAS:ORN:98821 allele at each retained variant site.

1.6.2 We used dp_cov_script.sh from SPOW-BDOW-introgression-scripts version 1.1.0 (Hanna, Henderson & Wall, 2017) to calculate the mean and standard deviation of the total coverage at the remaining sites. We then used vcf_dp_filter.sh from SPOW-BDOW-introgression-scripts version 1.1.0 to remove those with coverage in excess of the mean + 5σ (we only kept sites with coverage <301 X).

1.6.3 We used AD_pct.sh from SPOW-BDOW-introgression-scripts version 1.1.0 to calculate the spotted owl ancestry for each sample at each variant site. We then used compute_ad_mean_stdev.sh from SPOW-BDOW-introgression-scripts version 1.1.0 to calculate the mean and standard deviation (σ) of the spotted owl ancestry across all variant sites.

1.6.4 We tested for significant difference in the spotted owl ancestry of the individuals in four sets of populations using Welch's *t*-test (Welch, 1947) and applied a Bonferroni adjustment (Dunn, 1961) to the p-value cut-off to correct for multiple comparisons. We performed Welch's *t*-test using the Welch_ttest.py script from SPOW-BDOW-introgression-scripts version 1.1.0. We first tested for significant difference in the spotted owl ancestry of the Siskiyou barred owls versus the rest of the western barred owls. Based on the result of this test, we then grouped the Siskiyou barred owls with the other western barred owls into a population including all western barred owls. Similarly, we tested for significant difference between the pre and post-contact spotted owl samples before grouping all of the spotted owls together into a combined spotted owl population. We then tested for significant difference between all western barred owls and the eastern barred owls. Finally, we grouped all of the barred owl samples together and tested for significant difference in spotted owl ancestry between the barred and spotted owls.

1.6.5 We used AD_pct_ex.sh from SPOW-BDOW-introgression-scripts version 1.1.0 to calculate the spotted owl ancestry for each sample at each variant site and also return the number of reads for the sample the site. We then used ext_fmt_sliding_window_reads.sh from SPOW-BDOW-introgression-scripts version 1.1.0 with default parameters to

conduct a sliding window analysis with 50,000 nt adjacent windows and calculate the average spotted owl ancestry in each window.

1.6.6 We used outlier_window_detection.py from SPOW-BDOW-introgression-scripts version 1.1.0 to detect and graph outlier windows. For each sample, we only considered windows where that sample had data for at least ten sites in that window. Outlier windows were those that had an average spotted owl ancestry >= 0.4 in samples with an average genome-wide ancestry close to 0. In samples with an average genome-wide ancestry close to 1, outliers were windows with an average spotted owl ancestry <= 0.6. The outlier_window_detection.py script also merged adjacent outlier windows.

1.6.7 In order to create the input file for calculation of π and F_{ST} statistics, we filtered the raw variant file using vcf_qual_filter_pi.sh from SPOW-BDOW-introgression-scripts version 1.1.0 to retain only biallelic sites sites where CAS:ORN:98821 (the source of the StrOccCau_1.0_nuc_masked reference genome) was not homozygous for the alternate allele. We used this script also to exclude indels, only retain sites where the Phred-scaled probability that a polymorphism exists was >50, and to filter out high coverage sites so as to only retain sites with coverage <301 X. We then used the script ad_pi_no_coords.sh from SPOW-BDOW-introgression-scripts version 1.1.0 to output the number of reads in the AD field that supported either the reference or alternative allele at each site. We then used this file as input to the script countFstPi from SPOW-BDOW-introgression-scripts version 1.1.0 to calculate π_{Within}, $\pi_{Between}$, and F_{ST} for various population comparisons. We used the "Category" column from Table S4 for population groupings. We included the reference genome source sample (CAS:ORN:98821) in the "Spotted Owl (pre-contact)" population. In order to arrive at our final π_{Within} and $\pi_{Between}$ values, we divided the output of the script by the number of A, C, G, and T characters in the reference genome sequence.

2 Supplementary Tables

Table S1. DNA sequence data details for each sample.

The "Specimen Identifier" column provides the voucher specimen codes. The "Other Sample Identifier" column provides an abbreviated sample code. Column "Sample Set" refers to the round of sequencing that produced the sequence data for a given sample. The main and supplemental methodology sections provide details of the production of these two sets of sequence data. Column "SRA ACCN" provides NCBI Sequence Read Archive (SRA) run accessions in which the raw sequences for each sample are archived.

Voucher Specimen Identifier	Other Sample Identifier	Sample Set	SRA ACCN
CAS:ORN:98821	Sequoia	N/A	SRR4011595, SRR4011596, SRR4011597, SRR4011614, SRR4011615, SRR4011616, SRR4011617, SRR4011618, SRR4011619, SRR4011620,
CNHM<USA-OH>:ORNITH:B41533	CMCB41533	N/A	SRR5428115, SRR5428116, SRR5428117
CAS:ORN:87569	CAS87569	1	
CAS:ORN:92982	ASG007	1	
CAS:ORN:95475	MK994	1	
CAS:ORN:95789	JMR920	1	
CAS:ORN:95790	ASG037	1	
CAS:ORN:95964	MEF457	1	
CAS:ORN:97181	MK1020	1	
CNHM<USA-OH>:ORNITH:B40819	CMCB40819	1	
CNHM<USA-OH>:ORNITH:B40824	CMC40824	1	
CNHM<USA-OH>:ORNITH:B41566	CMCB41566	1	
CUMV:Bird:51478	CU51478	1	
MVZ:Bird:189508	ZRH455	1	
UWBM:Bird:62061	UWBM62061	1	
UWBM:Bird:76815	UWBM76815	1	
UWBM:Bird:91379	UWBM91379	1	
UWBM:Bird:91382	UWBM91382	1	
UWBM:Bird:91408	UWBM91408	1	
CAS:ORN:92979	MK968	2	
CAS:ORN:92980	MK987	2	
CAS:ORN:92981	MEF404	2	
CAS:ORN:95476	MK998	2	
CAS:ORN:95477	ASG017	2	
CAS:ORN:97049	LCW491	2	
CAS:ORN:97052	LCW443	2	
CAS:ORN:97174	MEF432	2	
CAS:ORN:97175	MK1012	2	
CAS:ORN:97176	JPD386	2	
CAS:ORN:97177	MEF435	2	
CAS:ORN:97201	LCW405	2	
CAS:ORN:97815	Hoopa20005	2	
CAS:ORN:97816	Hoopa20018	2	
CAS:ORN:97818	Hoopa20011	2	
CAS:ORN:97819	Hoopa20019	2	
CAS:ORN:97820	Hoopa20017	2	
CAS:ORN:97822	Hoopa20014	2	
CAS:ORN:98171	ZRH962	2	
CAS:ORN:98198	ZRH602	2	
CAS:ORN:99315	ZRH604	2	
CAS:ORN:99320	ZRH607	2	
CAS:ORN:99423	NSO138799040	2	
CAS:ORN:99425	NSO168709365	2	
UWBM:Bird:53433	UWBM53433	2	
UWBM:Bird:65055	UWBM65055	2	
UWBM:Bird:67015	UWBM67015	2	
UWBM:Bird:74078	UWBM74078	2	
UWBM:Bird:79007	UWBM79007	2	
UWBM:Bird:79049	UWBM79049	2	
UWBM:Bird:79141	UWBM79141	2	
UWBM:Bird:91380	UWBM91380	2	
UWBM:Bird:91392	UWBM91392	2	
UWBM:Bird:91393	UWBM91393	2	

Table S2. Specimen institution data.

This table provides information regarding the collections that archive the *Strix* specimens utilized in this study.

Specimen Identifier	Specimen Collection
CAS:ORN:87569	CAS Ornithology (ORN), California Academy of Sciences, San Francisco, California, U.S.A.
CAS:ORN:92979	CAS Ornithology (ORN), California Academy of Sciences, San Francisco, California, U.S.A.
CAS:ORN:92980	CAS Ornithology (ORN), California Academy of Sciences, San Francisco, California, U.S.A.
CAS:ORN:92981	CAS Ornithology (ORN), California Academy of Sciences, San Francisco, California, U.S.A.
CAS:ORN:92982	CAS Ornithology (ORN), California Academy of Sciences, San Francisco, California, U.S.A.
CAS:ORN:95475	CAS Ornithology (ORN), California Academy of Sciences, San Francisco, California, U.S.A.
CAS:ORN:95476	CAS Ornithology (ORN), California Academy of Sciences, San Francisco, California, U.S.A.
CAS:ORN:95477	CAS Ornithology (ORN), California Academy of Sciences, San Francisco, California, U.S.A.
CAS:ORN:95789	CAS Ornithology (ORN), California Academy of Sciences, San Francisco, California, U.S.A.
CAS:ORN:95790	CAS Ornithology (ORN), California Academy of Sciences, San Francisco, California, U.S.A.
CAS:ORN:95964	CAS Ornithology (ORN), California Academy of Sciences, San Francisco, California, U.S.A.
CAS:ORN:97049	CAS Ornithology (ORN), California Academy of Sciences, San Francisco, California, U.S.A.
CAS:ORN:97052	CAS Ornithology (ORN), California Academy of Sciences, San Francisco, California, U.S.A.
CAS:ORN:97174	CAS Ornithology (ORN), California Academy of Sciences, San Francisco, California, U.S.A.
CAS:ORN:97175	CAS Ornithology (ORN), California Academy of Sciences, San Francisco, California, U.S.A.
CAS:ORN:97176	CAS Ornithology (ORN), California Academy of Sciences, San Francisco, California, U.S.A.
CAS:ORN:97177	CAS Ornithology (ORN), California Academy of Sciences, San Francisco, California, U.S.A.
CAS:ORN:97181	CAS Ornithology (ORN), California Academy of Sciences, San Francisco, California, U.S.A.
CAS:ORN:97201	CAS Ornithology (ORN), California Academy of Sciences, San Francisco, California, U.S.A.
CAS:ORN:97815	CAS Ornithology (ORN), California Academy of Sciences, San Francisco, California, U.S.A.
CAS:ORN:97816	CAS Ornithology (ORN), California Academy of Sciences, San Francisco, California, U.S.A.
CAS:ORN:97818	CAS Ornithology (ORN), California Academy of Sciences, San Francisco, California, U.S.A.
CAS:ORN:97819	CAS Ornithology (ORN), California Academy of Sciences, San Francisco, California, U.S.A.
CAS:ORN:97820	CAS Ornithology (ORN), California Academy of Sciences, San Francisco, California, U.S.A.
CAS:ORN:97822	CAS Ornithology (ORN), California Academy of Sciences, San Francisco, California, U.S.A.
CAS:ORN:98171	CAS Ornithology (ORN), California Academy of Sciences, San Francisco, California, U.S.A.
CAS:ORN:98198	CAS Ornithology (ORN), California Academy of Sciences, San Francisco, California, U.S.A.
CAS:ORN:98821	CAS Ornithology (ORN), California Academy of Sciences, San Francisco, California, U.S.A.
CAS:ORN:99315	CAS Ornithology (ORN), California Academy of Sciences, San Francisco, California, U.S.A.
CAS:ORN:99320	CAS Ornithology (ORN), California Academy of Sciences, San Francisco, California, U.S.A.
CAS:ORN:99423	CAS Ornithology (ORN), California Academy of Sciences, San Francisco, California, U.S.A.
CAS:ORN:99425	CAS Ornithology (ORN), California Academy of Sciences, San Francisco, California, U.S.A.
CNHM<USA-OH>:ORNITH:B40819	Museum of Natural History & Science, Cincinnati Museum Center, Cincinnati, Ohio, U.S.A.
CNHM<USA-OH>:ORNITH:B40824	Museum of Natural History & Science, Cincinnati Museum Center, Cincinnati, Ohio, U.S.A.
CNHM<USA-OH>:ORNITH:B41533	Museum of Natural History & Science, Cincinnati Museum Center, Cincinnati, Ohio, U.S.A.
CNHM<USA-OH>:ORNITH:B41566	Museum of Natural History & Science, Cincinnati Museum Center, Cincinnati, Ohio, U.S.A.
CUMV:Bird:51478	Bird Collection, Cornell University Museum of Vertebrates, Ithaca, New York, U.S.A.
MVZ:Bird:189508	Bird Collection, Museum of Vertebrate Zoology, University of California, Berkeley, California, U.S.A.
UWBM:Bird:53433	Ornithology Collection, Burke Museum, University of Washington, Seattle, Washington, U.S.A.
UWBM:Bird:62061	Ornithology Collection, Burke Museum, University of Washington, Seattle, Washington, U.S.A.
UWBM:Bird:65055	Ornithology Collection, Burke Museum, University of Washington, Seattle, Washington, U.S.A.
UWBM:Bird:67015	Ornithology Collection, Burke Museum, University of Washington, Seattle, Washington, U.S.A.
UWBM:Bird:74078	Ornithology Collection, Burke Museum, University of Washington, Seattle, Washington, U.S.A.
UWBM:Bird:76815	Ornithology Collection, Burke Museum, University of Washington, Seattle, Washington, U.S.A.
UWBM:Bird:79007	Ornithology Collection, Burke Museum, University of Washington, Seattle, Washington, U.S.A.
UWBM:Bird:79049	Ornithology Collection, Burke Museum, University of Washington, Seattle, Washington, U.S.A.
UWBM:Bird:79141	Ornithology Collection, Burke Museum, University of Washington, Seattle, Washington, U.S.A.
UWBM:Bird:91379	Ornithology Collection, Burke Museum, University of Washington, Seattle, Washington, U.S.A.
UWBM:Bird:91380	Ornithology Collection, Burke Museum, University of Washington, Seattle, Washington, U.S.A.
UWBM:Bird:91382	Ornithology Collection, Burke Museum, University of Washington, Seattle, Washington, U.S.A.
UWBM:Bird:91392	Ornithology Collection, Burke Museum, University of Washington, Seattle, Washington, U.S.A.
UWBM:Bird:91393	Ornithology Collection, Burke Museum, University of Washington, Seattle, Washington, U.S.A.
UWBM:Bird:91408	Ornithology Collection, Burke Museum, University of Washington, Seattle, Washington, U.S.A.

Table S3. Additional specimen data.

We here provide additional data for each sample, including the taxonomic identification, the county and state of the collection locality, and the date of collection. The column "Pre or Post Contact" documents whether, based upon the date of collection, a sample's population was in contact with the other species.

Specimen Identifier	Genus	species	subspecies	County	State	Collection Date	Pre or Post Contact
CAS:ORN:99423	*Strix*	*occidentalis*	*caurina*	Humboldt	California	1-Jul-2011	Post
CAS:ORN:99425	*Strix*	*occidentalis*	*caurina*	Humboldt	California	28-Jul-2011	Post
CAS:ORN:98821	*Strix*	*occidentalis*	*caurina*	Marin	California	25-Jun-2005	Pre
MVZ:Bird:189508	*Strix*	*occidentalis*	*caurina*	Marin	California	9-Dec-2012	Post
CAS:ORN:87569	*Strix*	*occidentalis*	*caurina*	Napa	California	1998	Pre
UWBM:Bird:91380	*Strix*	*occidentalis*	*caurina*	Douglas	Oregon	4-Jun-2008	Post
UWBM:Bird:53433	*Strix*	*occidentalis*	*caurina*	Lane	Oregon	18-Jul-1995	Post
UWBM:Bird:91392	*Strix*	*occidentalis*	*caurina*	Lane	Oregon	9-Jun-2008	Post
UWBM:Bird:91379	*Strix*	*occidentalis*	*caurina*	Linn	Oregon	11-Jun-2008	Post
UWBM:Bird:91393	*Strix*	*occidentalis*	*caurina*	Linn	Oregon	12-Jun-2008	Post
UWBM:Bird:91408	*Strix*	*occidentalis*	*caurina*	Kittitas	Washington	28-Dec-2010	Post
UWBM:Bird:62061	*Strix*	*occidentalis*	*occidentalis*	Tulare	California	Jun-1994	Pre
CAS:ORN:95475	*Strix*	*varia*	*varia*	Humboldt	California	24-May-2006	Post
CAS:ORN:95476	*Strix*	*varia*	*varia*	Humboldt	California	25-May-2006	Post
CAS:ORN:95477	*Strix*	*varia*	*varia*	Humboldt	California	29-May-2006	Post
CAS:ORN:97049	*Strix*	*varia*	*varia*	Humboldt	California	3-Dec-2009	Post
CAS:ORN:97052	*Strix*	*varia*	*varia*	Humboldt	California	23-Aug-2010	Post
CAS:ORN:97201	*Strix*	*varia*	*varia*	Humboldt	California	5-Aug-2009	Post
CAS:ORN:97815	*Strix*	*varia*	*varia*	Humboldt	California	16-Oct-2013	Post
CAS:ORN:97816	*Strix*	*varia*	*varia*	Humboldt	California	30-Oct-2013	Post
CAS:ORN:97818	*Strix*	*varia*	*varia*	Humboldt	California	23-Oct-2013	Post
CAS:ORN:97819	*Strix*	*varia*	*varia*	Humboldt	California	30-Oct-2013	Post
CAS:ORN:97820	*Strix*	*varia*	*varia*	Humboldt	California	30-Oct-2013	Post
CAS:ORN:97822	*Strix*	*varia*	*varia*	Humboldt	California	30-Oct-2013	Post
CAS:ORN:92979	*Strix*	*varia*	*varia*	Siskiyou	California	28-Jul-2005	Post
CAS:ORN:92980	*Strix*	*varia*	*varia*	Siskiyou	California	3-Aug-2005	Post
CAS:ORN:92981	*Strix*	*varia*	*varia*	Siskiyou	California	29-Jul-2005	Post
CAS:ORN:92982	*Strix*	*varia*	*varia*	Siskiyou	California	29-Jul-2005	Post
CAS:ORN:95789	*Strix*	*varia*	*varia*	Siskiyou	California	29-Aug-2006	Post
CAS:ORN:95790	*Strix*	*varia*	*varia*	Siskiyou	California	28-Aug-2006	Post
CAS:ORN:97174	*Strix*	*varia*	*varia*	Siskiyou	California	3-Aug-2006	Post
CAS:ORN:97175	*Strix*	*varia*	*varia*	Siskiyou	California	10-Aug-2006	Post
CAS:ORN:97176	*Strix*	*varia*	*varia*	Siskiyou	California	10-Aug-2006	Post
CAS:ORN:97177	*Strix*	*varia*	*varia*	Siskiyou	California	3-Aug-2006	Post
CAS:ORN:97181	*Strix*	*varia*	*varia*	Siskiyou	California	23-Jul-2006	Post
CAS:ORN:98198	*Strix*	*varia*	*varia*	Sonoma	California	20-Nov-2013	Post
CAS:ORN:95964	*Strix*	*varia*	*varia*	Marion	Indiana	18-Dec-2002	Pre
CUMV:Bird:51478	*Strix*	*varia*	*varia*	Cortland	New York	9-Apr-2005	Pre
CNHM<USA-OH>:ORNITH:B40824	*Strix*	*varia*	*varia*	Hamilton	Ohio	27-Jan-2008	Pre
CNHM<USA-OH>:ORNITH:B41566	*Strix*	*varia*	*varia*	Hamilton	Ohio	27-Nov-2009	Pre
CNHM<USA-OH>:ORNITH:B41533	*Strix*	*varia*	*varia*	Hamilton	Ohio	29-Nov-2010	Pre
CNHM<USA-OH>:ORNITH:B40819	*Strix*	*varia*	*varia*	Scioto	Ohio	29-Mar-2008	Pre
UWBM:Bird:91382	*Strix*	*varia*	*varia*	Lane	Oregon	22-Sep-2004	Post
UWBM:Bird:67015	*Strix*	*varia*	*varia*	Island	Washington	2-Aug-1997	Post
UWBM:Bird:79141	*Strix*	*varia*	*varia*	Island	Washington	13-Nov-1996	Post
CAS:ORN:99315	*Strix*	*varia*	*varia*	Lewis	Washington	22-Jul-2011	Post
UWBM:Bird:76815	*Strix*	*varia*	*varia*	Pierce	Washington	8-Oct-1990	Post
UWBM:Bird:74078	*Strix*	*varia*	*varia*	Skagit	Washington	14-Nov-2000	Post
UWBM:Bird:79049	*Strix*	*varia*	*varia*	Skagit	Washington	21-Jul-1993	Post
UWBM:Bird:65055	*Strix*	*varia*	*varia*	Whatcom	Washington	6-Jul-1994	Post
UWBM:Bird:79007	*Strix*	*varia*	*varia*	Whatcom	Washington	16-Jul-1990	Post
CAS:ORN:98171	*Strix*	*varia* x *occidentalis*		Humboldt	California	26-Mar-2014	Post
CAS:ORN:99320	*Strix*	*varia* x *occidentalis*		Benton	Oregon	24-Dec-2005	Post

Table S4. Population assignment, ancestry, and site coverage values for each sample.

The "Category" column provides the population into which we grouped each sample. The spotted owl ancestry values are averages of the ancestry at each variant site across all sites with data for an individual. The site coverage is the average sequence coverage across all sites examined. "SD" stands for "standard deviation".

Specimen Identifier	Category	Mean Spotted Owl Ancestry	SD Spotted Owl Ancestry	Mean Site Coverage (X)	SD Site Coverage (X)
CNHM<USA-OH>:ORNITH:B41533	Reference Barred Owl	0.000	0.000	15.549	5.811
CNHM<USA-OH>:ORNITH:B40824	Eastern Barred Owl	0.066	0.243	0.627	0.827
CNHM<USA-OH>:ORNITH:B41566	Eastern Barred Owl	0.066	0.239	1.521	1.353
UWBM:Bird:91382	Western Barred Owl	0.067	0.248	0.263	0.551
CNHM<USA-OH>:ORNITH:B40819	Eastern Barred Owl	0.068	0.243	1.091	1.116
CAS:ORN:95475	Western Barred Owl	0.068	0.248	0.597	0.812
CAS:ORN:97181	Siskiyou Barred Owl	0.068	0.245	0.976	1.048
CAS:ORN:92982	Siskiyou Barred Owl	0.069	0.247	0.746	0.913
CUMV:Bird:51478	Eastern Barred Owl	0.069	0.245	0.961	1.042
CAS:ORN:95789	Siskiyou Barred Owl	0.069	0.246	0.948	1.031
UWBM:Bird:67015	Western Barred Owl	0.069	0.247	0.723	0.897
CAS:ORN:95790	Siskiyou Barred Owl	0.069	0.246	0.997	1.057
UWBM:Bird:76815	Western Barred Owl	0.069	0.248	0.754	0.904
CAS:ORN:97816	Western Barred Owl	0.069	0.251	0.354	0.608
CAS:ORN:97175	Siskiyou Barred Owl	0.069	0.241	2.072	1.606
CAS:ORN:97815	Western Barred Owl	0.069	0.251	0.444	0.681
UWBM:Bird:74078	Western Barred Owl	0.070	0.250	0.511	0.735
CAS:ORN:95964	Eastern Barred Owl	0.070	0.227	6.407	3.134
CAS:ORN:98198	Western Barred Owl	0.070	0.235	3.492	2.205
CAS:ORN:99315	Western Barred Owl	0.070	0.234	4.546	2.817
CAS:ORN:97174	Siskiyou Barred Owl	0.070	0.247	0.945	1.016
CAS:ORN:97819	Western Barred Owl	0.070	0.252	0.278	0.536
CAS:ORN:92980	Siskiyou Barred Owl	0.070	0.250	0.713	0.881
UWBM:Bird:79007	Western Barred Owl	0.070	0.255	0.081	0.286
CAS:ORN:92981	Siskiyou Barred Owl	0.070	0.243	1.745	1.409
CAS:ORN:97177	Siskiyou Barred Owl	0.070	0.241	2.167	1.599
CAS:ORN:97820	Western Barred Owl	0.070	0.252	0.421	0.662
CAS:ORN:92979	Siskiyou Barred Owl	0.070	0.255	0.118	0.349
CAS:ORN:97822	Western Barred Owl	0.070	0.252	0.515	0.736
UWBM:Bird:79141	Western Barred Owl	0.070	0.250	0.759	0.903
UWBM:Bird:79049	Western Barred Owl	0.070	0.251	0.590	0.798
CAS:ORN:97049	Western Barred Owl	0.071	0.250	0.722	0.875
CAS:ORN:97052	Western Barred Owl	0.071	0.252	0.483	0.711
UWBM:Bird:65055	Western Barred Owl	0.071	0.253	0.377	0.625
CAS:ORN:97818	Western Barred Owl	0.071	0.254	0.297	0.553
CAS:ORN:97201	Western Barred Owl	0.071	0.250	0.752	0.900
CAS:ORN:95477	Western Barred Owl	0.071	0.253	0.345	0.599
CAS:ORN:97176	Siskiyou Barred Owl	0.072	0.257	0.036	0.190
CAS:ORN:95476	Western Barred Owl	0.073	0.252	0.854	0.985
CAS:ORN:99320	Hybrid	0.359	0.428	1.624	1.395
CAS:ORN:98171	Hybrid	0.538	0.377	2.209	1.615
UWBM:Bird:62061	Spotted Owl (pre-contact)	0.991	0.095	0.495	0.732
CAS:ORN:99423	Spotted Owl (post-contact)	0.995	0.069	1.380	1.228
CAS:ORN:99425	Spotted Owl (post-contact)	0.995	0.067	1.985	1.492
UWBM:Bird:91380	Spotted Owl (post-contact)	0.995	0.071	0.017	0.131
UWBM:Bird:91393	Spotted Owl (post-contact)	0.995	0.070	0.261	0.522
UWBM:Bird:91392	Spotted Owl (post-contact)	0.995	0.069	0.044	0.211
UWBM:Bird:53433	Spotted Owl (post-contact)	0.995	0.069	0.168	0.418
CAS:ORN:87569	Spotted Owl (pre-contact)	0.995	0.066	1.080	1.104
MVZ:Bird:189508	Spotted Owl (post-contact)	0.996	0.064	0.850	0.965
UWBM:Bird:91379	Spotted Owl (post-contact)	0.996	0.065	0.449	0.697
UWBM:Bird:91408	Spotted Owl (post-contact)	0.996	0.063	1.107	1.133

CAS:ORN:98821	Reference Spotted Owl	1.000	0.000	60.815	15.939

Table S5. Mean and standard deviation spotted owl ancestry by population.

We provide the mean and standard deviation of spotted owl ancestry for each population. The "All Western Barred Owls" population was a superset of the Siskiyou and Western Barred Owl populations. The "All Barred Owls" population is a combination of all of the *Strix varia* samples and the "All Spotted Owls" population is a combination of all of the *S. occidentalis* samples. "SD" stands for "standard deviation".

Population	Mean Spotted Owl Ancestry	SD Spotted Owl Ancestry
Siskiyou Barred Owls	6.964E-02	8.968E-04
Western Barred Owls	6.992E-02	1.072E-03
All Western Barred Owls	6.983E-02	1.026E-03
Eastern Barred Owls	6.757E-02	1.475E-03
All Barred Owls	6.953E-02	1.334E-03
Spotted Owls (pre-contact)	9.930E-01	2.360E-03
Spotted Owls (post-contact)	9.952E-01	3.902E-04
All Spotted Owls	9.948E-01	1.363E-03

Table S6. Tests of significant difference in spotted owl ancestry.

We here provide the t-values from multiple Welch's *t*-tests conducted for comparisons of spotted owl ancestry among populations. The "All Western Barred Owls" population was a superset of the Siskiyou and Western Barred Owl populations. An asterisk (*) and bold font indicate those tests with $p<0.0125$, which is the significance cut-off after applying the Bonferroni correction. The "All Barred Owls" population is a combination of all of the *Strix varia* samples and the "All Spotted Owls" population is a combination of all of the *S. occidentalis* samples.

Populations Compared	**t-value**	**p-value**
Siskiyou Barred Owls vs. Western Barred Owls	-0.771	0.449
All Western Barred Owls vs. Eastern Barred Owls	2.968	0.036
Spotted Owls (pre-contact) vs. Spotted Owls (post-contact)	-0.931	0.522
All Barred Owls vs. All Spotted Owls	**-1913.494***	3.008 E-43

3 Supplementary Figures

Figure S1. Eastern barred owl, Siskiyou County barred owl, and northern spotted owl plumage comparison.

This image displays the darker ventral plumage of a *Strix varia* collected in Siskiyou County, California compared with that of typical *S. varia* and *S. occidentalis caurina* individuals.

On the left is the ventral plumage of a *Strix varia* from eastern North America. In the center is a *S. varia* from Siskiyou County, California. On the right is a *S. occidentalis caurina* from northern California.

Figure S2. Image of barred owls from Siskiyou County.

This image displays the ventral plumage of three *Strix varia* collected in Siskiyou County, California. Owl A is specimen CAS:ORN:92981. Owl B is CAS:ORN:92979. Owl C is CAS:ORN:97181.

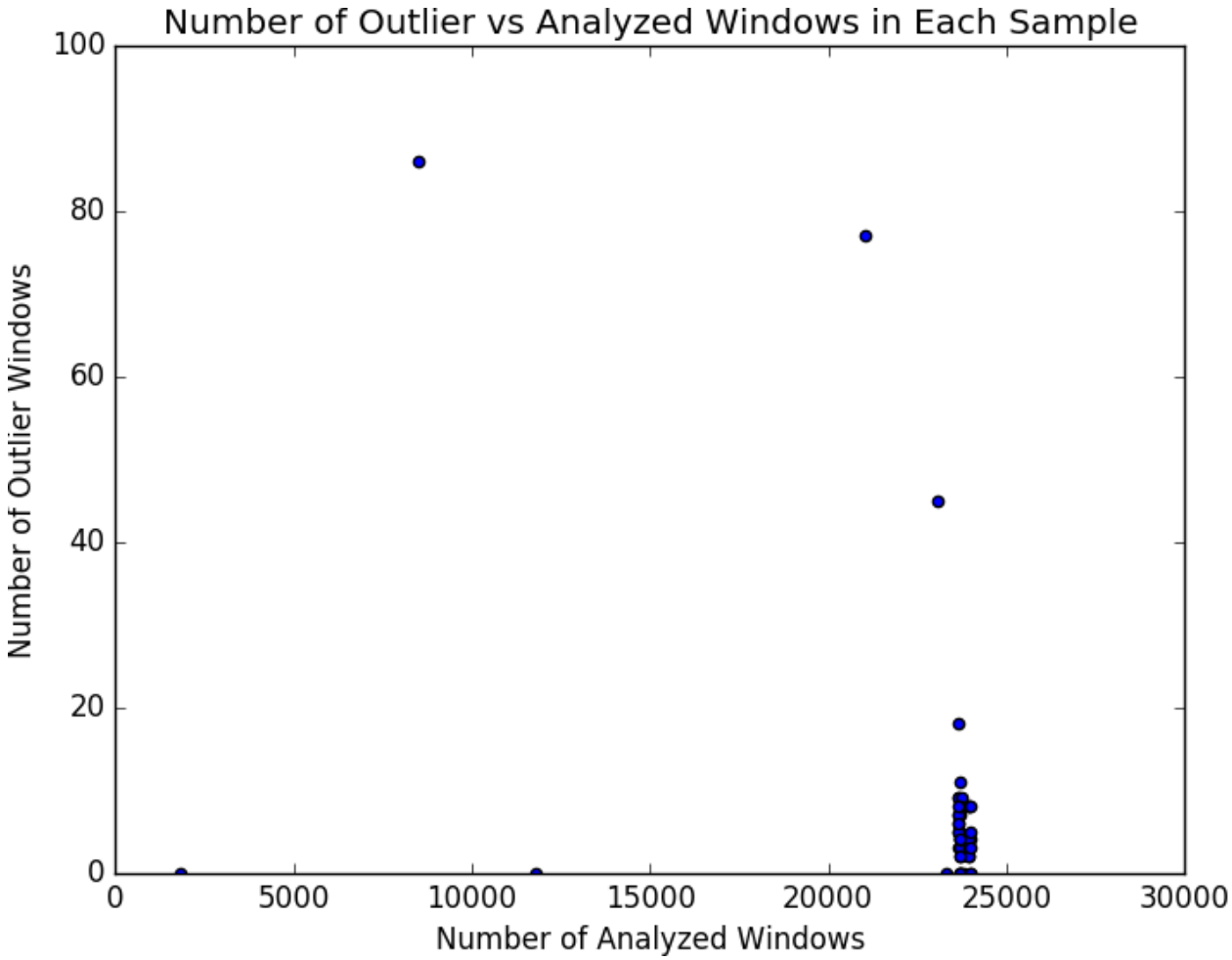

Figure S3. Plot of Number of outlier windows versus analyzed windows.

The number of spotted owl ancestry windows of length ≥50,000 nt that were outliers relative to the genome-wide average ancestry for those samples is on the y-axis. The x-axis represents the total number of windows analyzed for each sample. We required the presence of data for at least ten variant sites in order to analyze a window for a given sample. Samples with lower sequence coverage tended to have fewer windows that could be analyzed.

4 Supplementary References

Bolger AM., Lohse M., Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. DOI: 10.1093/bioinformatics/btu170.

DePristo MA., Banks E., Poplin R., Garimella KV., Maguire JR., Hartl C., Philippakis AA., del Angel G., Rivas MA., Hanna M., McKenna A., Fennell TJ., Kernytsky AM., Sivachenko AY., Cibulskis K., Gabriel SB., Altshuler D., Daly MJ. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* 43:491–498. DOI: 10.1038/ng.806.

Dunn OJ. 1961. Multiple Comparisons among Means. *Journal of the American Statistical Association* 56:52–64. DOI: 10.1080/01621459.1961.10482090.

GATK Dev Team 2017.GATK Tool Documentation. [Accessed 2017 Oct 3]. Available from: https://software.broadinstitute.org/gatk/documentation/

Hanna ZR., Henderson JB., Wall JD. 2017. SPOW-BDOW-introgression-scripts. Version 1.1.0. *Zenodo*. DOI: 10.5281/zenodo.1065074.

Hanna ZR., Henderson JB., Wall JD., Emerling CA., Fuchs J., Runckel C., Mindell DP., Bowie RCK., DeRisi JL., Dumbacher JP. 2017a. Supplemental dataset for Northern Spotted Owl (*Strix occidentalis caurina*) genome assembly version 1.0. *Zenodo*. DOI: 10.5281/zenodo.822859.

Hanna ZR., Henderson JB., Wall JD., Emerling CA., Fuchs J., Runckel C., Mindell DP., Bowie RCK., DeRisi JL., Dumbacher JP. 2017b. Northern Spotted Owl (*Strix occidentalis caurina*) Genome: Divergence with the Barred Owl (*Strix varia*) and Characterization of Light-Associated Genes. *Genome Biology and Evolution* 9:2522–2545. DOI: 10.1093/gbe/evx158.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv:1303.3997 Q-Bio. [Accessed 2016 Feb 16]. Available from: http://arxiv.org/abs/1303.3997.

McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernytsky A., Garimella K., Altshuler D., Gabriel S., Daly M., DePristo MA. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20:1297–1303. DOI: 10.1101/gr.107524.110.

Quantum GIS Development Team 2017. Quantum GIS Geographic Information System. Open Source Geospatial Foundation Project. [Accessed 2017 Sep 16]. Available from: http://qgis.org.

Van der Auwera GA., Carneiro MO., Hartl C., Poplin R., del Angel G., Levy-Moonshine A., Jordan T., Shakir K., Roazen D., Thibault J., Banks E., Garimella KV., Altshuler D., Gabriel S., DePristo MA. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics* 11:11.10.1-11.10.33. DOI: 10.1002/0471250953.bi1110s43.

Welch BL. 1947. The Generalization of "Student's" Problem When Several Different Population Variances Are Involved. *Biometrika* 34:28–35. DOI: 10.1093/biomet/34.1-2.28.