

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} and John P. Dumbacher^{3,4}

¹ Museum of Vertebrate Zoology, University of California, Berkeley, CA, United States of America

² Department of Integrative Biology, University of California, Berkeley, CA, United States of America

³ Department of Ornithology & Mammalogy, California Academy of Sciences, San Francisco, CA, United States of America

⁴ Center for Comparative Genomics, California Academy of Sciences, San Francisco, CA, United States of America

⁵ Chan Zuckerberg Biohub, San Francisco, CA, 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

ABSTRACT

We report 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.

Submitted 31 July 2017
Accepted 18 September 2017
Published 10 October 2017

Corresponding author
Zachary R. Hanna,
zachanna@berkeley.edu

Academic editor
Michael Wink

Additional Information and
Declarations can be found on
page 20

DOI 10.7717/peerj.3901

© Copyright
2017 Hanna et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects Genetics, Genomics, Molecular Biology, Zoology

Keywords Mitochondrial genome, Bird, Northern spotted owl, Barred owl, Mtgenome, mtDNA, Control region

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](#); [Hadzrath & 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 & Birmingham, 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 researchers’ 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](#); [Hull et al., 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 US 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 US 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; Crosby & 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, Mullins & Forsman, 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 descendant *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

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	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

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; Kent, 2012*), to find Illumina reads that aligned to the *Ninox novaeseelandiae* mitochondrial genome (GenBank Accession AY309457.1) (*Harrison et al., 2004*) (Supplemental Material 1.1.1) and extracted those matching reads using excerptByIds version 1.0.2 (*Henderson & Hanna, 2016b*) (Supplemental Material 1.1.2). We then used SOAPdenovo2 version 2.04 (*Luo et al., 2012*) to assemble those sequences (Supplemental Material 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 (Supplemental Material 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}* (Supplemental Material 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*) (Supplemental Material 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*) (Supplemental Material 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 (Supplemental Material 1.1.8) and extracted aligned reads using excerptByIds

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	<i>S. o. caurina</i> , <i>S. varia</i>	External	ATCCTCATTCTCTCCCCGT	This study
17122R	CR1	<i>S. o. caurina</i> , <i>S. varia</i>	External	GGTGGGGGTTATTATTAACCTT	This study
CR1-F1	CR1	<i>S. o. caurina</i> , <i>S. varia</i>	Internal	CTCSASCAAATCCCAAGTT	This study
CR1-F1-RC	CR1	<i>S. o. caurina</i> , <i>S. varia</i>	Internal	AAACTTGGGATTGGSTSGAG	This study
CR1-R2	CR1	<i>S. o. caurina</i> , <i>S. varia</i>	Internal	GGAGGGCGAGAATAGTTGRT	This study
CR1-R2-RC	CR1	<i>S. o. caurina</i> , <i>S. varia</i>	Internal	AYCAACTATTCTGCCCTCC	This study
N1	CR1	<i>S. o. caurina</i>	Internal	AACATTGGCTTGAAACCAA	Barrowclough, Gutierrez & Groth (1999)
41R	CR2	<i>S. o. caurina</i>	External	GCATCTTCAGTGCCATGCTT	This study
17572F	CR2	<i>S. o. caurina</i>	External	ATTATCCAAGGTCTGCGGCC	This study
17589F	CR2	<i>S. o. caurina</i>	Internal	GCCTGAAAAACGCCGTTAA	This study
18327F	CR2	<i>S. o. caurina</i>	Internal	CACTTTGCGCCTCTGGITC	This study
19911R	CR2	<i>S. o. caurina</i>	Internal	AGAGAGGCTCTGATTGCTTG	This study
ND6-ext1F	CR2	<i>S. varia</i>	External	ACAACCCCATAATAYGGCGA	This study
12S-ext1R	CR2	<i>S. varia</i>	External	GGTAGATGGGCATTACACT	This study
final-CR2F	CR2	<i>S. varia</i>	Internal	TCAAACCAAACGATCGAGAA	This study
18547F	CR2	<i>S. varia</i>	Internal	CTCACGTGAAATCAGCAACC	This study
19088R	CR2	<i>S. varia</i>	Internal	ATTCAACTAAAATCGTTACAAATCTT	This study
19088R-RC	CR2	<i>S. varia</i>	Internal	AAGATTGTAAACGAATTAGTTGAAT	This study

version 1.0.2 ([Supplemental Material 1.1.9](#)). We then performed another PRICE assembly with the same initial contig as before, but with the extracted additional Illumina sequence data ([Supplemental Material 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](#)) ([Supplemental Material 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](#); [Benson, 2012](#)) ([Supplemental Material 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, CA, USA) on an ABI 3130xl automated sequencer (Applied Biosystems, Foster City, CA, USA; [Supplemental Material 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 (Supplemental Material 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 (Supplemental Material 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 (Supplemental Material 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, CA, USA), 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 (Supplemental Material 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>) (Supplemental Material 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; Li *et al.*, 2015). We then used Samtools and GNU Awk (GAWK) version 4.0.1 (Free Software Foundation, 2012) to filter the alignments (Supplemental Material 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 (Supplemental Material 1.3.5).

We annotated the final assembly using the MITOS WebServer version 806 (Bernt *et al.*, 2013) (Supplemental Material 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; Benson, 2016](#)) ([Supplemental Material 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 ([Supplemental Material 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](#)) ([Supplemental Material 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 ([Supplemental Material 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](#)) ([Supplemental Material 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 ([Supplemental Material 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 ([Supplemental Material 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 ([Supplemental Material 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 Brower Extensible Data (BED) formatted file ([Supplemental Material 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 ([Supplemental Material 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 ([Supplemental Material 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](#)) ([Supplemental Material 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](#)) ([Supplemental Material 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 ([Supplemental Material 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 ([Supplemental Material 1.9.2](#)).

We filtered the alignment file to only retain alignments to the final mitochondrial genome assembly using Samtools version 1.3 with HTSlip 1.3.1 ([Li et al., 2009](#); [Li et al., 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 ([Supplemental Material 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 ([Supplemental Material 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 ([Supplemental Material 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 ([Supplemental Material 1.10.1](#)). We extended the output 15,019 nt contig using PRICE version 1.2 ([Supplemental Material 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 ([Supplemental Material 1.11.1](#)). After four cycles, this assembly produced one contig (ContigInput2) of length 17,073 nt.

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 ([Supplemental Material 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 ([Supplemental Material 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 ([Supplemental Material 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 ([Table 2](#)). We performed all sequencing using BigDye terminator chemistry (Applied Biosystems, Foster City, CA, USA) on an ABI 3130xl automated sequencer (Applied Biosystems, Foster City, CA, USA).

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 ([Supplemental Material 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; Benson, 2016*) ([Supplemental Material 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 ([Supplemental Material 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](#)) ([Supplemental Material 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](#)) ([Supplemental Material 1.13.2](#)). We then used MEGA version 7.0.18 ([Kumar, Stecher & Tamura, 2016](#)) to calculate the *p*-distance ([Supplemental Material 1.13.3](#)) and the corrected pairwise distance ([Tamura & Nei, 1993](#)) ([Supplemental Material 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 ([Supplemental Material 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,

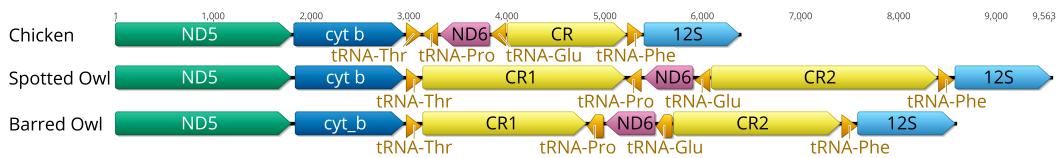


Figure 1 Ancestral avian mitochondrial gene order surrounding the control region compared with that of *Strix occidentalis caurina* and *Strix varia*. The Chicken panel displays the gene order of *Gallus gallus*, which is the presumed ancestral avian gene order. The Spotted Owl panel depicts the gene 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 orders of the genes outside of the region depicted in this figure are the same in the chicken, spotted owl, and barred owl.

respectively. As for all typical avian mitochondrial genomes, they are circular and code for two rRNAs, 22 tRNAs, and 13 polypeptides (Figs. S1 and S2). 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 (Fig. 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 four 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 (Fig. 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 (Fig. 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

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

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}* ([Fig. 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* CR2 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](#))

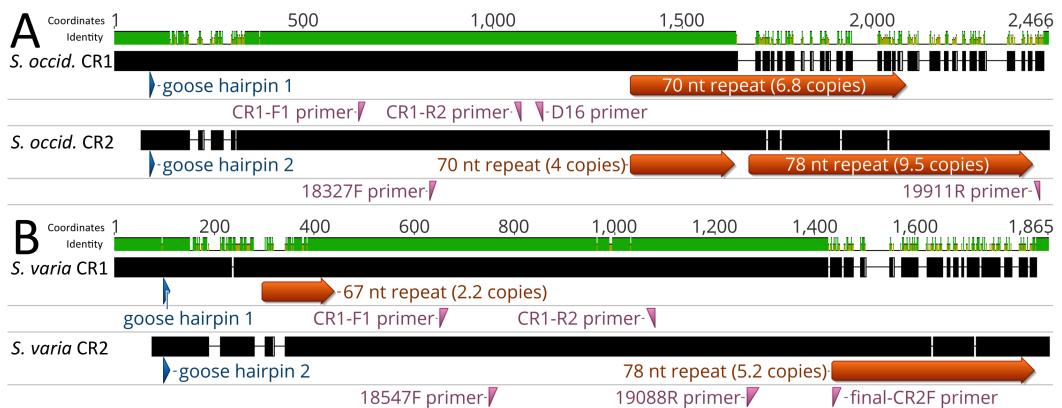


Figure 2 Alignment of control regions 1 and 2 within *Strix occidentalis caurina* and *Strix varia*. (A) depicts an alignment of the *Strix occidentalis caurina* control regions 1 and 2. (B) displays an alignment of the *Strix varia* control regions 1 and 2. The numerical coordinates at the top of each 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, Gutierrez & Groth \(1999\)](#) to amplify a portion of control region 1 are annotated in reddish purple.

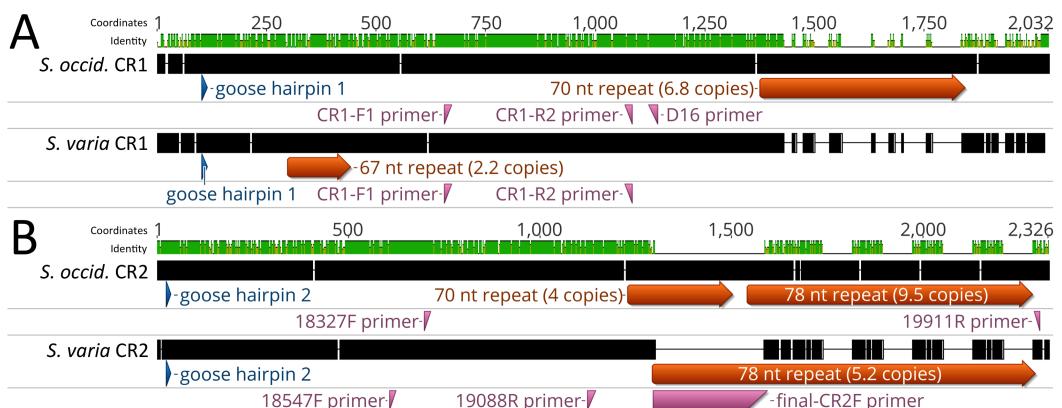


Figure 3 Alignment of *Strix occidentalis caurina* control regions 1 and 2 with those of *Strix varia*. (A) depicts an alignment of the *Strix occidentalis caurina* control region 1 with that of *Strix varia*. (B) displays an alignment of the *Strix occidentalis caurina* control region 2 with that of *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, Gutierrez & Groth \(1999\)](#) 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.

([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 ([Fig. S3](#)). 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](#)) ([Fig. 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

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.

<i>Numt</i> #	Genes included	Start mtDNA	End mtDNA	Nuclear genome scaffold	Start scaffold	End scaffold	Orientation	% ID	E-value	Bit score	Length alignment (nt)	Length <i>numt</i> (nt)
1	<i>tRNA^{Phe}</i> - 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 - <i>tRNA^{Asn}</i>	2,706	5,223	scaffold478	50,161	52,680	+	80.66	0.0	1,921	2,549	-
	<i>tRNA^{Asn}</i> - COI - <i>tRNA^{Ser}</i> 2	5,219	6,932	scaffold478	57,635	59,328	+	83.22	0.0	1,552	1,716	-
	<i>tRNA^{Ser}</i> 2 - <i>tRNA^{Asp}</i> - 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	-
	<i>cyt b</i>	14,047	14,733	scaffold478	44,785	45,459	+	82.82	1.92e-169	604	687	-
	<i>tRNA^{Thr}</i>	14,729	14,878	scaffold478	46,066	46,222	+	82.80	1.09e-27	134	157	-
	16S	1,682	2,603	scaffold215	5,517,239	5,518,161	-	81.97	0.0	773	932	923
	<i>tRNA^{Ser}</i> 2 - ATP8 - ND3_a	6,989	9,584	scaffold215	5,513,222	5,515,749	-	79.01	0.0	1,690	2,615	2,528
4	16S - <i>tRNA^{Leu}</i> 2	2,290	2,788	scaffold632	1,548,886	1,549,372	+	77.50	6.14e-70	274	511	487
5	ND1 - <i>tRNA^{Gln}</i> - ND2	2,810	4,646	scaffold167	11,322,764	11,324,590	+	80.54	0.0	1,400	1,840	2,732
	ND2 - <i>tRNA^{Asn}</i> - COI	4,692	5,597	scaffold167	11,324,598	11,325,495	+	83.68	0.0	846	907	-
6	<i>tRNA^{Glu}</i> - ND2 - COI	3,851	5,526	scaffold1500	35,914	37,582	-	84.21	0.0	1,620	1,678	1,669
7	ND2 - <i>tRNA^{Asn}</i> - <i>tRNA^{Tyr}</i>	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 & Nei, 1993).

Gene	Number of sites in alignment (nt)	<i>p</i> -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%

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 & Härlid, 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 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; Brito, 2005; Hull et al., 2010; Barrowclough et al., 2011), 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 524 nt 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. caurina* and *S. varia* was 9.21%, which is also on the lower end of the range of the 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.

ACKNOWLEDGEMENTS

We thank WildCare, San Rafael for graciously providing 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.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

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.

Grant Disclosures

The following grant information was disclosed by the authors:

Louise Kellogg Fund, Museum of Vertebrate Zoology, University of California, Berkeley.

National Science Foundation: DGE 1106400.

National Institutes of Health: S10 RR029668, S10 RR027303.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Zachary R. Hanna conceived and designed the experiments, performed the experiments, analyzed the data, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.
- James B. Henderson, Anna B. Sellas and John P. Dumbacher conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.
- Jérôme Fuchs conceived and designed the experiments, performed the experiments, analyzed the data, wrote the paper, reviewed drafts of the paper.

- Rauri C.K. Bowie conceived and designed the experiments, contributed reagents/materials/analysis tools, wrote the paper, reviewed drafts of the paper.

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

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](#)).

Data Availability

The following information was supplied regarding data availability:

BLATq deposited at Zenodo DOI: [10.5281/zenodo.61136](#) available at <https://github.com/calacademy-research/blatq>.

excerptByIDs deposited at Zenodo DOI: [10.5281/zenodo.61134](#) and available at <https://github.com/calacademy-research/excerptByIDs>.

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.3901#supplemental-information>.

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, Härlid A. 2000. Mitochondrial genomic rearrangements in songbirds. *Molecular Biology and Evolution* 17:107–113 DOI [10.1093/oxfordjournals.molbev.a026223](#).
- 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. Available at <https://tandem.bu.edu/trf/trf.html> (accessed on October 2016).
- Benson G. 2016. Tandem repeats finder. Version 4.09. Available at <https://tandem.bu.edu/trf/trf.html> (accessed on October 2016).

- 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](https://doi.org/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](https://doi.org/10.1007/BF00326158).
- Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, Pütz J, Midendorf 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](https://doi.org/10.1016/j.ympev.2012.08.023).
- Biomatters. 2016.** Geneious. Version 914. Available at <http://www.geneious.com> (accessed on 1 October 2016).
- Bolger AM, Lohse M, Usadel B. 2014.** Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**:2114–2120 DOI [10.1093/bioinformatics/btu170](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1186/1471-2105-10-421).
- Capella-Gutiérrez S, Gabaldón T. 2013.** trimAl. Version 1.4.rev15. Available at <https://github.com/scapella/trimal> (accessed on 1 October 2016).
- 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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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 DOI [10.1093/oxfordjournals.molbev.a003917](https://doi.org/10.1093/oxfordjournals.molbev.a003917).
- Edgar RC. 2004.** MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**:1792–1797 DOI [10.1093/nar/gkh340](https://doi.org/10.1093/nar/gkh340).

- Faber JE, Stepień 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](https://doi.org/10.1006/mpev.1998.0530).
- Free Software Foundation.** 2012. GNU Awk. Version 4.0.1. Available at <https://www.gnu.org/software/gawk> (accessed on 1 October 2016).
- Free Software Foundation.** 2014. GNU Grep. Version 2.16. Available at <https://www.gnu.org/software/grep> (accessed on 1 October 2016).
- 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](https://doi.org/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 DOI [10.1093/oxfordjournals.molbev.a025568](https://doi.org/10.1093/oxfordjournals.molbev.a025568).
- 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](https://doi.org/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](https://doi.org/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](https://doi.org/10.3109/19401736.2014.898277).
- Haddrath O, Baker AJ.** 2001. Complete mitochondrial DNA genome 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](https://doi.org/10.1098/rspb.2001.1587).
- Haig SM, Mullins TD, Forsman ED.** 2004. Subspecific relationships and genetic structure in the spotted owl. *Conservation Genetics* **5**:683–705 DOI [10.1007/s10592-004-1864-y](https://doi.org/10.1007/s10592-004-1864-y).
- 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 DOI [10.1111/j.1523-1739.2004.00206.x](https://doi.org/10.1111/j.1523-1739.2004.00206.x).
- 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* Epub ahead of print Aug 23 2017 DOI [10.1093/gbe/evx158](https://doi.org/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 DOI [10.1093/oxfordjournals.molbev.a003730](https://doi.org/10.1093/oxfordjournals.molbev.a003730).

- 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. *Zenodo*. Version 1.0.2 DOI 10.5281/zenodo.61136.
- Henderson JB, Hanna ZR.** 2016b. excerptByIDs. *Zenodo*. Version 1.0.2 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, USA. *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. Available at <http://mafft.cbrc.jp/alignment/software> (accessed on 1 October 2016).
- 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 Owlklks. *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](https://doi.org/10.1101/gr.229202).
- Kent WJ.** 2012. BLAT—the BLAST-like alignment tool. Version 35. Available at <https://users.soe.ucsc.edu/~kent> (accessed on 1 October 2016).
- 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 B: Biological Sciences* **276**:615–621 DOI [10.1098/rspb.2008.1310](https://doi.org/10.1098/rspb.2008.1310).
- Kumar S, Stecher G, Tamura K.** 2016. *Molecular Biology and Evolution* **33**:1870–1874 DOI [10.1093/molbev/msw054](https://doi.org/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](https://doi.org/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](https://doi.org/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 preprint. [arXiv:1303.3997](https://arxiv.org/abs/1303.3997).
- Li H.** 2013b. bioawk. Version 1.0. Available at <https://github.com/lh3/bioawk> (accessed on 01 October 2016).
- Li H, Handsaker B, Marshall J, Danecek P.** 2015. Samtools. Version 1.3 with HTSlip 1.3.1. Available at <http://www.htslip.org> (accessed on 01 October 2016).
- 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](https://doi.org/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](https://doi.org/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**:1–6 DOI [10.1186/2047-217X-1-18](https://doi.org/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 DOI [10.1093/oxfordjournals.molbev.a025750](https://doi.org/10.1093/oxfordjournals.molbev.a025750).

- 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](https://doi.org/10.1007/s10592-008-9581-6).
- Mazur KM, James PC.** 2000. Barred Owl (*Strix varia*). In: Poole A, ed. *The birds of North America online*. Ithaca: Cornell Lab of Ornithology.
- McCarthy EM.** 2006. *Handbook of avian hybrids of the world*. 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 of the United States of America* **95**:10693–10697 DOI [10.1073/pnas.95.18.10693](https://doi.org/10.1073/pnas.95.18.10693).
- 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 DOI [10.1093/oxfordjournals.molbev.a025884](https://doi.org/10.1093/oxfordjournals.molbev.a025884).
- 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](https://doi.org/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: 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](https://doi.org/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 DOI [10.1093/oxfordjournals.jhered.a022948](https://doi.org/10.1093/oxfordjournals.jhered.a022948).
- NCBI Resource Coordinators.** 2015. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research* **43**:D6–D17 DOI [10.1093/nar/gku1130](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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 DOI [10.1007/PL00006402](https://doi.org/10.1007/PL00006402).
- Richly E, Leister D.** 2004. NUMTs in sequenced eukaryotic genomes. *Molecular Biology and Evolution* **21**:1081–1084 DOI [10.1093/molbev/msh110](https://doi.org/10.1093/molbev/msh110).
- Ruby JG.** 2014. PRICE. Version 1.2. Available at <http://derisilab.ucsf.edu/software/price> (accessed on 01 October 2016).
- 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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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. *The Journal of Poultry Science* **41**:76–82 DOI [10.2141/jpsa.41.76](https://doi.org/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](https://doi.org/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](https://doi.org/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: 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](https://doi.org/10.1016/S0379-4172(06)60131-5).

- 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](https://doi.org/10.3109/19401736.2015.1043538).
- 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](https://doi.org/10.1089/10665270050081478).
- 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](https://doi.org/10.2307/2410006).

1 **Supplementary Materials for:**

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

5

6 Zachary R. Hanna,^{1,2,3,4,7} James B. Henderson,^{3,4} Anna B. Sellas,^{4,5} Jérôme Fuchs,^{3,6} Rauri C. K.
7 Bowie,^{1,2} John P. Dumbacher^{3,4}

8

9 ¹ Museum of Vertebrate Zoology, University of California, Berkeley, Berkeley, California,
10 United States of America

11 ² Department of Integrative Biology, University of California, Berkeley, Berkeley, California,
12 United States of America

13 ³ Department of Ornithology & Mammalogy, California Academy of Sciences, San Francisco,
14 California, United States of America

15 ⁴ Center for Comparative Genomics, California Academy of Sciences, San Francisco, California,
16 United States of America

17 ⁵ Chan Zuckerberg Biohub, San Francisco, California, United States of America

18 ⁶ UMR 7205 Institut de Systématique, Evolution, Biodiversité, CNRS, MNHN, UPMC, EPHE,
19 Sorbonne Universités, Muséum National d'Histoire Naturelle, Paris, France

20 ⁷ Corresponding Author:
21 Zachary Hanna
22 Email address: zachanna@berkeley.edu

- 23 **1 Supplementary Methods**
- 24 *1.1 Initial S. o. caurina assembly*
- 25 1.1.1 BLATq version 1.0.2 (Henderson & Hanna, 2016a) used to align our 150 bp read 1
26 sequences from the Nextera700bp library (Hanna et al., 2017) to the *Ninox*
27 *novaeseelandiae* mitochondrial genome (GenBank accession AY309457) using default
28 BLAT parameters other than “-stepSize=5 -repMatch=100000 -out=blast8”.
- 29 1.1.2 excerptByIds version 1.0.2 (Henderson & Hanna, 2016b) used, to extract the pairs of
30 reads that had BLATq hits to the *Ninox novaeseelandiae* mitochondrial genome.
- 31 1.1.3 SOAPdenovo2 version 2.04 (Luo et al., 2012) used with default settings except
32 “SOAPdenovo-127mer all -m 127 -R”. In our configuration file for this assembly we
33 used the default minimum alignment length between a read and contig (32 for paired-end
34 reads) and the default minimum pair number cutoff (3 for paired-end reads) and set the
35 reads to be used for the assembly of both contigs and scaffolds.
- 36 1.1.4 Web version of the NCBI BLAST+ version 2.2.29 tool BLASTn (Altschul et al., 1990;
37 Zhang et al., 2000; Morgulis et al., 2008; Camacho et al., 2009) with default parameters
38 to search the NCBI nucleotide collection (Johnson et al., 2008; Boratyn et al., 2013;
39 NCBI Resource Coordinators, 2015; Benson et al., 2015).
- 40 1.1.5 GNU Grep version 2.16 (Free Software Foundation, 2014) used to search the trimmed
41 and merged reads from lane 1 and 2 of the Nextera550bp library for reads that matched
42 the assembled sequence of *tRNA^{Phe}* or *tRNA^{Thr}*.
- 43 1.1.6 We found 3 reads that spanned *tRNA^{Phe}* (1 of which was a merged read pair (Hanna et al.,
44 2017)) and combined them using the Geneious version 9.1.4 (Kearse et al., 2012;

- 45 Biomatters, 2016) *de novo* assembler. The resulting contig contained spanned from the
46 control region through *tRNA^{Phe}* and into *12S*.
- 47 1.1.7 PRICE version 1.2 (Ruby, Bellare & DeRisi, 2013; Ruby, 2014) used with the assembled
48 contig spanning *tRNA^{Phe}* as the initial contig and the parameters “-spf<merged_reads>
49 300 600 -icf <initial_contig.fasta> 1 1 5 -mol 25 -mpi 85 -MPI 80 -nc 60 -lenf 40 5 -lenf
50 90 10 -a 10 -target 85 1 1 1 -maxHp 25 -o <output.fa> -o <output.priceq>”. The
51 “merged_reads” parameter referred to the merged overlapping sequences from lane 1 of
52 the Nextera550bp library.
- 53 1.1.8 BLATq version 1.0.2 (Henderson & Hanna, 2016a) to align our 150 bp read 1 sequences
54 from the Nextera700bp library to the assembly output by PRICE using default settings
55 other than “-stepSize=5 -repMatch=100000 -out=blast8”.
- 56 1.1.9 excerptByIds version 1.0.2 (Henderson & Hanna, 2016b) to extract the read 2 of the
57 paired-end sequences corresponding to the aligned read 1 sequences.
- 58 1.1.10 PRICE assembly with the same initial contig as before, but with more sequence data,
59 including the merged overlapping sequences from both lane 1 and 2 of the Nextera550bp
60 library and the matching paired-end sequences from the Nextera700bp library. We used
61 the default PRICE assembly settings with the following exceptions: “-fp <Nextera700bp
62 read 1 matches> <Nextera700bp read 2 matches> 809 -spf <Nextera550bp lane 1 merged
63 reads> 300 600 -spf <Nextera550bp lane 2 merged reads> 300 600 -icf
64 <initial_contig.fasta> 1 1 5 -mol 25 -mpi 85 -MPI 80 -nc 60 -lenf 40 5 -lenf 90 10 -a 10 -
65 target 85 1 1 1 -maxHp 25 -o <output.fa> -o <output.priceq>”.
- 66 1.1.11 MITOS WebServer version 605 (Bernt et al., 2013) specifying “genetic code = 02 -
67 Vertebrate”.

68 1.1.12 Tandem Repeats Finder version 4.07b (Benson, 1999, 2012) used with default options.

69 1.2 *S. o. caurina Sanger sequencing assembly confirmation*

70 1.2.1 We isolated genomic DNA using a DNeasy Blood & Tissue Kit (Qiagen, U.S.A.).

71 Polymerase chain reaction (PCR) conditions for primers 17589F and 41R included an
72 initial denaturation at 94°C for 1 min; then 30 cycles at 94°C for 30 s, 60°C for 30 s, and
73 72°C for 2 min; and a final extension at 72°C for 7 min. We then sequenced both ends of
74 the PCR-amplified fragment using BigDye terminator chemistry (Applied Biosystems,
75 U.S.A.) on an ABI 3130xl automated sequencer (Applied Biosystems, U.S.A.).

76 1.2.2 Polymerase chain reaction (PCR) conditions for primers 17572F and 41R were the same
77 as for primers 17589F and 41R (initial denaturation at 94°C for 1 min; then 30 cycles at
78 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
79 min). We sequenced the PCR products using BigDye terminator chemistry (Applied
80 Biosystems, U.S.A.) on an ABI 3130xl automated sequencer (Applied Biosystems,
81 U.S.A.).

82 1.2.3 We used the Geneious mapper through the “map to reference” function with default
83 options other than sensitivity set to “highest sensitivity / slow” to align the edited Sanger-
84 derived sequences to the 19,946 nt preliminary mitochondrial genome assembly.

85 1.2.4 Polymerase chain reaction (PCR) conditions for primers cytb-F1 and 17122R included an
86 initial denaturation at 94°C for 3 min; then 35 cycles at 94°C for 30 s, 50°C for 30 s, and
87 72°C for 2 min; and a final extension at 72°C for 10 min.

88 1.3 *S. o. caurina final assembly*

89 1.3.1 We removed scaffold-3674 from the draft whole nuclear genome assembly (Hanna et al.,
90 2017) using the “filterbyname.sh” tool in the BBMap tool suite version 36.02 (Bushnell,

91 2016) and replaced it with the 19,948 nt mitochondrial genome assembly from our
92 targeted assembly methodology. When referring to specific scaffolds here and in the
93 manuscript, we have inserted a dash (“-”) between the word “scaffold” and the scaffold
94 number for legibility. These dashes are not present in any of the assembly data files.
95 Thus, “scaffold-3674” referenced in the manuscript will appear as “scaffold3674” in the
96 assembly and other associated files.

97 1.3.2 We aligned all filtered Illumina sequences to this new draft reference genome using bwa
98 version 0.7.13-r1126 (Li, 2013a) using default options except parameters "bwa mem -M".
99 We separately aligned paired-end and unpaired reads. We then merged the paired-end
100 and unpaired read alignments using the Picard version 2.2.4
101 (<http://broadinstitute.github.io/picard>) function MergeSamFiles and sorted them using the
102 Picard function SortSam, employing default settings for both tools. We next marked
103 duplicate reads (both PCR and optical) using the Picard function MarkDuplicates,
104 employing default settings.

105 1.3.3 We then filtered out duplicate reads, low quality alignments, secondary alignments, and
106 alignments where both reads of a pair did not align to the mitochondrial assembly. We
107 used Samtools to filter out duplicate reads marked by Picard using the Samtools
108 parameters “-F 0x400”. We next used Samtools to filter out alignments of quality less
109 than 10 with the parameters “-q 10”. We then filtered out secondary alignments with the
110 Samtools parameters “-F 0x100”. We then used Samtools with GNU Awk (GAWK)
111 version 4.0.1 (Free Software Foundation, 2012) to filter out paired reads where one of the
112 reads mapped to a different contig/scaffold than the mitochondrial genome using

113 parameters “samtools view -h <input.bam> | awk ‘\$7 == “=” || \$7 == “*” || \$1 ~ “^@”’ |
114 samtools view -Sb - > <output.bam>”.

115 1.3.4 We used Samtools version 1.3 with HTSlib 1.3.1 (Li et al., 2009, 2015) with GAWK
116 version 4.0.1 (Free Software Foundation, 2012) to filter out all aligned sequences less
117 than 300 nt using parameters “samtools view -h <input.bam> | awk ‘\$1 ~ “^@”’ ||
118 length(\$10) >= 300’ | samtools view -Sb - > <output.bam>”.

119 1.3.5 We visually inspected all sites where there was lower coverage and any hint of
120 disagreement between reads and, except in the case of CR1 and CR2 where we relied on
121 the Sanger-derived sequence data, decided in favor of majority evidence, which matched
122 our preliminary assembly at all sites, providing confirmation of our assembly
123 methodology.

124 1.4 *S. o. caurina final annotation*

125 1.4.1 MITOS WebServer version 806 (Bernt et al., 2013) specifying “genetic code = 02 -
126 Vertebrate”.

127 1.4.2 Web version of Tandem Repeats Finder version 4.09 (Benson, 1999, 2016) employing
128 default options.

129 1.4.3 We used bioawk version 1.0 (Li, 2013b) and GAWK version 4.0.1 (Free Software
130 Foundation, 2012) to find goose hairpin sequences by searching for 7 C’s followed by 1
131 to 3 D nucleotides (A, G, or T) followed by 7 C’s in a FASTA format file (Pearson &
132 Lipman, 1988) with each control region input as a separate entry with the command
133 “bioawk -c fastx ‘{print \$seq}’ \$1 | awk ‘{pos=match(\$0,
134 /CCCCCCC[AGT]{1,3}CCCCCCC/);if(pos){print pos}}’”.

135 1.4.4 We used the NCBI BLAST+ version 2.4.0 (Altschul et al., 1990; Zhang et al., 2000;
136 Morgulis et al., 2008; Camacho et al., 2009) tool “makeblastdb” with options “-
137 parse_seqids -dbtype nucl” to create a database of the scaffold-3674 gene sequences and
138 then the tool “blastn” with default options except “-outfmt 6” to align the targeted
139 assembly gene sequences against this database.

140 1.4.5 We aligned the primers we developed to amplify control region 2 and the N1 primer used
141 by Barrowclough, Gutierrez & Groth (1999) to amplify a portion of control region 1 to
142 the final assembly using Geneious version 9.1.4 mapper through the “map to reference”
143 function (Kearse et al., 2012; Biomatters, 2016) with default options other than using the
144 sensitivity set at “highest sensitivity / slow”. We determined the position in the final
145 assembly of the D16 primer used by Barrowclough, Gutierrez & Groth (1999) to amplify
146 a portion of control region 1 by using the Geneious version 9.1.4 *de novo* assembler
147 (Kearse et al., 2012; Biomatters, 2016) with default parameters other than setting
148 sensitivity at “highest sensitivity / slow” to assemble control region 1 with the D16
149 primer.

150 1.4.6 We performed a multiple alignment of control regions 1 and 2 using the Geneious version
151 9.1.4 (Kearse et al., 2012; Biomatters, 2016) implementation of the MUSCLE version
152 3.8.425 (Edgar, 2004) aligner with default options.

153 1.4.7 We used the web version of NCBI’s BLAST+ Version 2.5.0 tool BLASTN (Altschul et
154 al., 1990; Zhang et al., 2000; Morgulis et al., 2008; Camacho et al., 2009) with default
155 parameters to search the NCBI nucleotide collection (Johnson et al., 2008; Boratyn et al.,
156 2013; NCBI Resource Coordinators, 2015; Benson et al., 2015) for sequences similar to

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

159 1.4.8 As a result of these searches, we aligned the primers used by Omote et al. (2013) to
160 amplify the control region in *Strix uralensis* to the final assembly using Geneious version
161 9.1.4 mapper with the “map to reference” function (Kearse et al., 2012; Biomatters, 2016)
162 with default options other than using the Geneious mapper with sensitivity set at “highest
163 sensitivity / slow”.

164 1.5 *Nuclear pseudogenation of S. o. caurina mitochondrial genes*

165 1.5.1 We first used the NCBI BLAST+ version 2.4.0 tool BLASTN (Altschul et al., 1990;
166 Zhang et al., 2000; Morgulis et al., 2008; Camacho et al., 2009) to align the final *S. o.*
167 *caurina* mitochondrial genome assembly to the draft nuclear genome assembly (Hanna et
168 al., 2017) using default parameters other than “-outfmt 6”.

169 1.5.2 We used GAWK version 4.0.1 (Free Software Foundation, 2012) to remove all
170 alignments to scaffold3674, which was the assembly of the mitochondrial genome in the
171 draft nuclear genome assembly.

172 1.5.3 We next used GAWK version 4.0.1 (Free Software Foundation, 2012) to reformat the
173 BLAST output into a BED (Browser Extensible Data) formatted file
174 (<http://genome.ucsc.edu/FAQ/FAQformat#format1>) with the parameters “cat
175 <filtered_BLAST_file> | awk 'BEGIN {OFS = "\t"} {print
176 "'Strix_Occidentalis'",\$7,\$8,\$2,\$9,\$10,\$12}' | awk 'BEGIN {OFS = \$"\\t"} {if (\$3<\$2)
177 print \$1,\$3,\$2,\$4,\$5,\$6,7; else print \$0}'”.

178 1.5.4 We used BEDTools version 2.26.0 tool “intersect” (Quinlan & Hall, 2010) to produce a
179 BED file of the intersection of the BED-formatted BLAST output with the BED file

180 output from the MITOS annotation of the final mitochondrial genome assembly using the
181 parameters “-a <BED-formatted BLAST output file> -b <MITOS annotation BED file> -
182 wo”. We then used the output of the intersection to determine the mitochondrial genes
183 spanned by each *Numt*.

184 1.6 *Strix varia sample CAS95964*

185 1.6.1 We extracted genomic DNA using a DNeasy Blood & Tissue Kit (Qiagen, U.S.A.). We
186 used 50 ng genomic DNA to prepare a whole-genome library using a Nextera DNA
187 Sample Preparation Kit (Illumina, U.S.A.). After fragmentation, we cleaned the reaction
188 with a DNA Clean & Concentrator -5 kit (Zymo Research, U.S.A.). We amplified the
189 reaction with 5 cycles of PCR using a KAPA Library Amplification kit (KAPA
190 Biosystems, U.S.A.) and then cleaned the reaction with a DNA Clean & Concentrator -5
191 kit (Zymo Research, U.S.A.). We used Dye-Free, 1.5% agarose, 250-1,500 base pair (bp)
192 cassette on a BluePippin (Sage Science, U.S.A.) to select library fragments in the size
193 range of 534-634 bp, which, after subtracting the 134 bp of adapters, corresponded to
194 selecting an average insert size of 450 bp. We next performed a real-time PCR (rtPCR)
195 using a KAPA Real-Time Library Amplification Kit (KAPA Biosystems, U.S.A.) on a
196 CFX96 Touch Real-Time PCR Detection System (Bio-Rad, U.S.A.) to further amplify
197 the library with 9 cycles PCR. We then cleaned the PCR products with a DNA Clean &
198 Concentrator -5 kit (Zymo Research, U.S.A.). We lastly assessed the library fragment
199 size distribution with a 2100 BioAnalyzer (Agilent Technologies, U.S.A.) and the
200 concentration of double-stranded DNA material with a Qubit 2.0 Flurometer (Invitrogen,
201 U.S.A.). Due to the presence of small peaks in the BioAnalyzer trace, we further cleaned
202 the library using 0.6X Agencourt AMPure XP (Beckman Coulter, U.S.A.) magnetic

203 beads. We obtained approximately one lane of 100 bp paired-end data sequenced in an
204 indexed pool using a 2-lane flow cell with a HiSeq PE Rapid Cluster Kit and a 200 cycle
205 HiSeq Rapid SBS Kit v1 on a HiSeq 2500 (Illumina, U.S.A.). The raw sequences are
206 available upon request.

207 1.6.2 We performed adapter and quality trimming of the sequence data using Trimmomatic
208 version 0.30 (Bolger, Lohse & Usadel, 2014) with the following options:
209 "ILLUMINACLIP:<FASTA format file of Illumina adapter sequences>:2:30:7
210 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36".

211 1.6.3 For use in only the SOAPdenovo2 assembly, we trimmed the sequences using a different
212 set of parameters and performed error-correction of the sequences. We performed adapter
213 and quality trimming using Trimmomatic version 0.30 (Bolger, Lohse & Usadel, 2014)
214 with the following options: "ILLUMINACLIP:<FASTA format file of Illumina adapter
215 sequences>:2:30:7 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36
216 HEADCROP:12". We merged into a single file all of the single reads that resulted after
217 their pair was dropped in the trimming process. Then we used the k-mer-based error
218 corrector in the SOAPdenovo2 toolkit, SOApEc version 2.01 (Luo et al., 2012), to correct
219 sequence errors. We first used the KmerFreq_HA tool to create a kmer frequency
220 spectrum with default options except "-k 27", which indicate that we used a kmer size of
221 27 for creating the spectrum. We then used the Corrector_HA tool along with the kmer
222 frequency spectrum that we created to correct all of our trimmed reads using default
223 options except "-k 27 -r 36", which indicate that we used a kmer size of 27 for the error
224 correction and kept trimmed reads as short as 36 bp.

225 1.7 *Strix varia sample CMCB41533*

226 1.7.1 We obtained tissue from a *S. varia* collected in Hamilton County, Ohio, United States of
227 America (CNHM<USA-OH>:ORNITH:B41533; Table 1), hereafter “CMCB41533”,
228 which is well outside of the zone of contact of *S. varia* and *S. occidentalis caurina* (Haig
229 et al., 2004). We obtained paired-end Illumina sequence data from a genomic library
230 constructed, sequenced, and the data processed as described in (Hanna et al., 2017). The
231 raw sequences are available from NCBI (SRA run accessions SRR5428115,
232 SRR5428116, and SRR5428117).

233 1.8 *Strix varia mitochondrial genome assembly*

234 1.8.1 We generated the mitochondrial genome assembly of the *S. varia* sample CMCB41533
235 by building a succession of assemblies that contributed information to the final assembly
236 from which we extracted the gene sequences. We used assemblies of sample CAS95964
237 to inform the process, but, as we had more sequence data for sample CMCB41533, we
238 chose to only produce a final assembly for this sample.

239 1.9 *Assembly of Strix varia ContigInput1*

240 1.9.1 We used bwa version 0.7.13-r1126 (Li, 2013a) with default options other than parameters
241 "bwa mem -M". We separately aligned paired-end and unpaired reads.

242 1.9.2 We merged the paired-end and unpaired read alignments using the Picard version 2.2.4
243 (<http://broadinstitute.github.io/picard>) function MergeSamFiles and sorted them using the
244 Picard function SortSam, employing default settings for both tools. We next marked
245 duplicate reads (both PCR and optical) using the Picard function MarkDuplicates,
246 employing default settings.

247 1.9.3 We filtered the alignment file for only alignments to the final mitochondrial genome
248 assembly using Samtools version 1.3 with HTSlip 1.3.1 (Li et al., 2009, 2015). We then

249 used Samtools to filter out duplicate reads marked by Picard using the Samtools
250 parameters “-F 0x400”. We next used Samtools to filter out alignments of quality less
251 than 10 with the parameters “-q 10”. We then filtered out secondary alignments with the
252 Samtools parameters “-F 0x100”. We then used Samtools with GNU Awk (GAWK)
253 version 4.0.1 (Free Software Foundation, 2012) to filter out paired reads where one of the
254 reads mapped to a different contig/scaffold than the mitochondrial genome using
255 parameters “samtools view -h <input.bam> | awk '\$7 == "=" || \$7 == "*" || \$1 ~ "^@"' |
256 samtools view -Sb - > <output.bam>”.

257 1.9.4 We visualized the alignment across the reference sequence in Geneious version 9.1.4
258 (Kearse et al., 2012; Biomatters, 2016). We used Geneious to generate a consensus
259 sequence with default parameters for the alignment to the mitochondrial genome.

260 1.9.5 We extracted 3 sequences from this consensus sequence based on the *S. o. caurina*
261 annotations to give a 142 nt fragment spanning from nucleotide 5 of *tRNA^{Phe}* part way
262 into *12S*; a longer, 844 nt fragment spanning from nucleotide 4 of *tRNA^{Phe}* part way into
263 *12S*; and a 1,042 nt fragment spanning from nucleotide 142 of *cyt b* part way into
264 *tRNA^{Thr}*. We then used these extracted sequences as three separate seed contigs in an
265 assembly using PRICE version 1.2 (Ruby, Bellare & DeRisi, 2013; Ruby, 2014). We
266 used the trimmed CMCB41533 paired read 1 and 2 sequences as the sequence data for
267 this run. Our PRICE assembly parameters were the defaults other than the following “-fp
268 <read 1 paired sequences> < read 2 paired sequences> 466 -icf <initial_contig.fasta> 1 1
269 5 -mol 30 -mpi 90 -MPI 85 -nc 60 -a 8 -target 85 1 1 1 -maxHp 25 -o <output.fa> -o
270 <output.priceq>”.

271 1.10 Assembly of ContigInput2 - CAS95964

272 1.10.1 We first used SOAPdenovo2 version 2.04 (Luo et al., 2012) to assemble all of the
273 trimmed, error-corrected CAS95964 sequences, employing default parameters other than
274 the options “SOAPdenovo all -m 63 -R”. In our configuration file for this assembly we
275 used the default minimum alignment length between a read and contig (32) and the
276 default minimum pair number cutoff (3) for both the paired-end and single-end reads. We
277 set the paired-end reads to be used for the assembly of both contigs and scaffolds and the
278 single-end reads for use only in contig assembly. We input both the paired-end and the
279 single-end reads with a rank of 1 and set the average insert size as 446 bp. This produced
280 a contig of length 15,019 nt.

281 1.10.2 We extended the contig using PRICE version 1.2 (Ruby, Bellare & DeRisi, 2013; Ruby,
282 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
283 the SOAPdenovo2 run above as the initial contig. Our PRICE assembly parameters were
284 the defaults other than the following “-fp <read 1 paired sequences> <read 2 paired
285 sequences> 446 -icf <initial_contig.fasta> 1 1 5 -mol 25 -mpi 85 -MPI 80 -nc 60 -lenf 40
286 5 -lenf 90 10 -a 10 -target 85 1 1 1 -maxHp 25 -o <output.fa> -o <output.priceq>”.
287

288 1.11 *Assembly of ContigInput2 - CMCB41533 PRICE*

289 1.11.1 We employed the trimmed CMCB41533 paired read 1 and 2 sequences as the sequence
290 data for this run. We used the 16,652 nt contig output from the CAS95964 run above as
291 the initial contig. Our PRICE version 1.2 (Ruby, Bellare & DeRisi, 2013; Ruby, 2014)
292 assembly parameters were the defaults other than the following “-fp <read 1 paired
293 sequences> < read 2 paired sequences> 350 -icf <initial_contig.fasta> 1 1 5 -mol 25 -mpi

294 85 -MPI 80 -nc 60 -lenf 40 5 -lenf 90 10 -a 10 -target 85 1 1 1 -maxHp 25 -o <output.fa>
295 -o <output.priceq>".

296 *1.12 Final Strix varia assembly*

297 1.12.1 We used PRICE version 1.2 (Ruby, Bellare & DeRisi, 2013; Ruby, 2014) the 9,690 nt
298 ContigInput1 contig output from cycle 16 of the initial CMCB41533 PRICE run and the
299 17,073 nt ContigInput2 as the initial contigs. We used the CMCB41533 paired read 1 and
300 2 sequences as well as the unpaired read 1 sequences that lost their mate as a result of
301 quality trimming for the sequence data input for this assembly. Our PRICE assembly
302 parameters were the defaults other than the following “-fp <read 1 paired sequences> <
303 read 2 paired sequences> 400 -spf <read 1 unpaired> 110 200 -icf <initial_contig.fasta> 1
304 1 5 -mol 25 -mpi 85 -MPI 80 -nc 60 -lenf 40 5 -lenf 90 10 -a 24 -target 85 1 1 1 -maxHp
305 25 -o <output.fa> -o <output.priceq>”.

306 1.12.2 PCR conditions for primers cytb-F1 and 17122R included an initial denaturation at 94°C
307 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
308 extension at 72°C for 10 min.

309 1.12.3 PCR conditions for primers ND6-ext1F and 12S-ext1R included an initial denaturation at
310 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
311 final extension at 72°C for 10 min.

312 1.12.4 We annotated the PRICE assembly using the MITOS WebServer version 605 (Bernt et
313 al., 2013) specifying “genetic code = 02 - Vertebrate”.

314 *1.13 Comparison of Strix occidentalis and Strix varia mitochondrial genes*

315 1.13.1 MAFFT version 7.305b (Katoh et al., 2002; Katoh & Standley, 2013; Katoh, 2016) used
316 with the default options other than parameters “--auto --clustalout”.

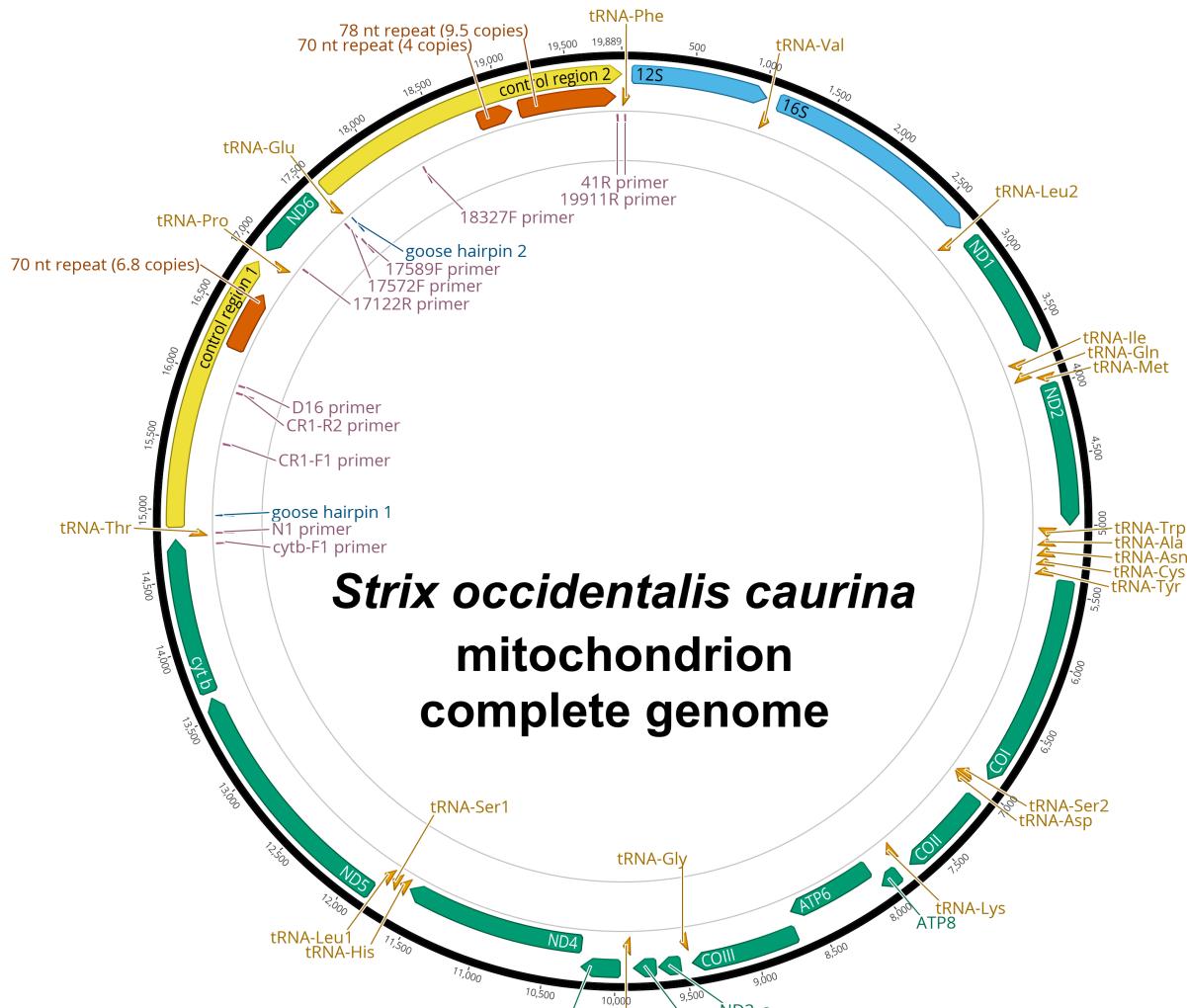
317 1.13.2 trimAl version 1.4.rev15 (Capella-Gutiérrez, Silla-Martínez & Gabaldón, 2009; Capella-
318 Gutiérrez & Gabaldón, 2013) with default options other than “-mega” used to convert the
319 alignments to MEGA format (Kumar, Tamura & Nei, 1994; Kumar, Stecher & Tamura,
320 2016).

321 1.13.3 MEGA version 7.0.18 (Kumar, Stecher & Tamura, 2016) used to calculate the p-distance
322 (an uncorrected pairwise distance that is the proportion of nucleotide sites at which two
323 sequences are different obtained by dividing the number of differences by the total
324 number of nucleotide sites) between *S. occidentalis caurina* and *S. varia* for each gene
325 with all alignment positions with gaps or missing data removed from the analysis.

326 1.13.4 MEGA version 7.0.18 (Kumar, Stecher & Tamura, 2016) used to calculate for each gene
327 the pairwise distance corrected by the Tamura-Nei model (TN93) of DNA sequence
328 evolution (Tamura & Nei, 1993) with rate variation among sites modeled using a gamma
329 distribution with shape parameter = 1, differences in the composition bias of sequences
330 considered in the comparisons (Tamura & Kumar, 2002), and with all alignment
331 positions with gaps or missing data removed from the analysis.

332 1.13.5 We weighted each distance by the length of the gene alignment from which it was
333 derived as a proportion of the total alignment length across all gene alignments and
334 calculated a weighted average pairwise distance across all of the genes.

335 2 Supplementary Figures



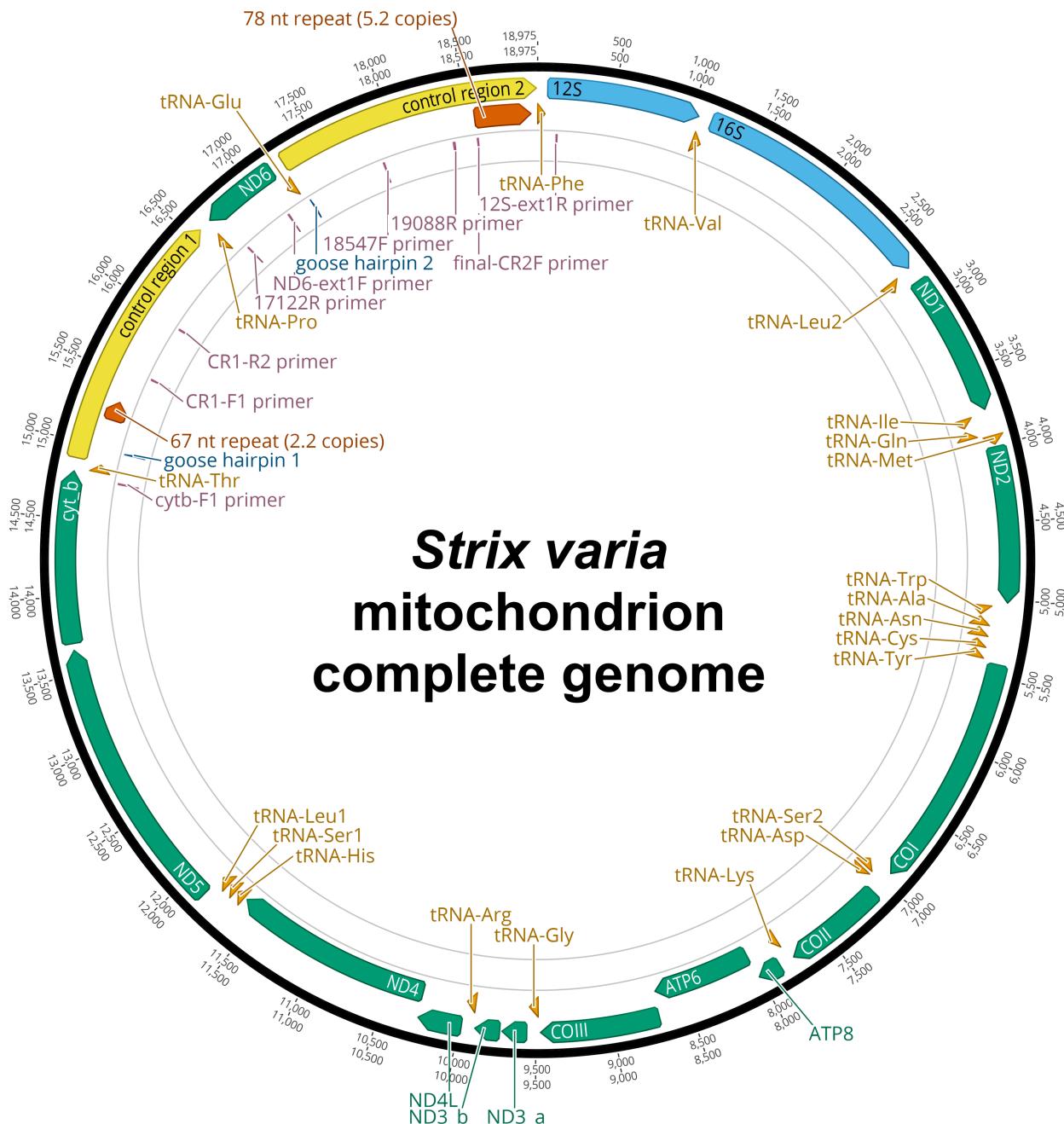
336

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

338 This is a graphical representation of the annotated complete genome of the northern
 339 spotted owl (*Strix occidentalis caurina*) mitochondrion. We have color-coded the various
 340 annotations, including genes for rRNA in sky blue, tRNA genes in orange, and all other genes in
 341 bluish green. The control regions are in yellow and the goose hairpin for each control region is
 342 depicted in blue. The locations of the primers we developed to amplify control regions 1 and 2 as
 343 well as the N1 and D16 primers used by Barrowclough et al. (1999) to amplify a portion of
 344 control region 1 are in reddish purple. The reverse complement versions of primers used (" -RC"

345 versions) are not shown. Regions with repetitive motifs are in vermillion. The base numbers
346 around the perimeter of the figure are in nucleotides. We used Geneious version 9.1.4 (Kearse et
347 al., 2012; Biomatters, 2016) to construct this figure.

348



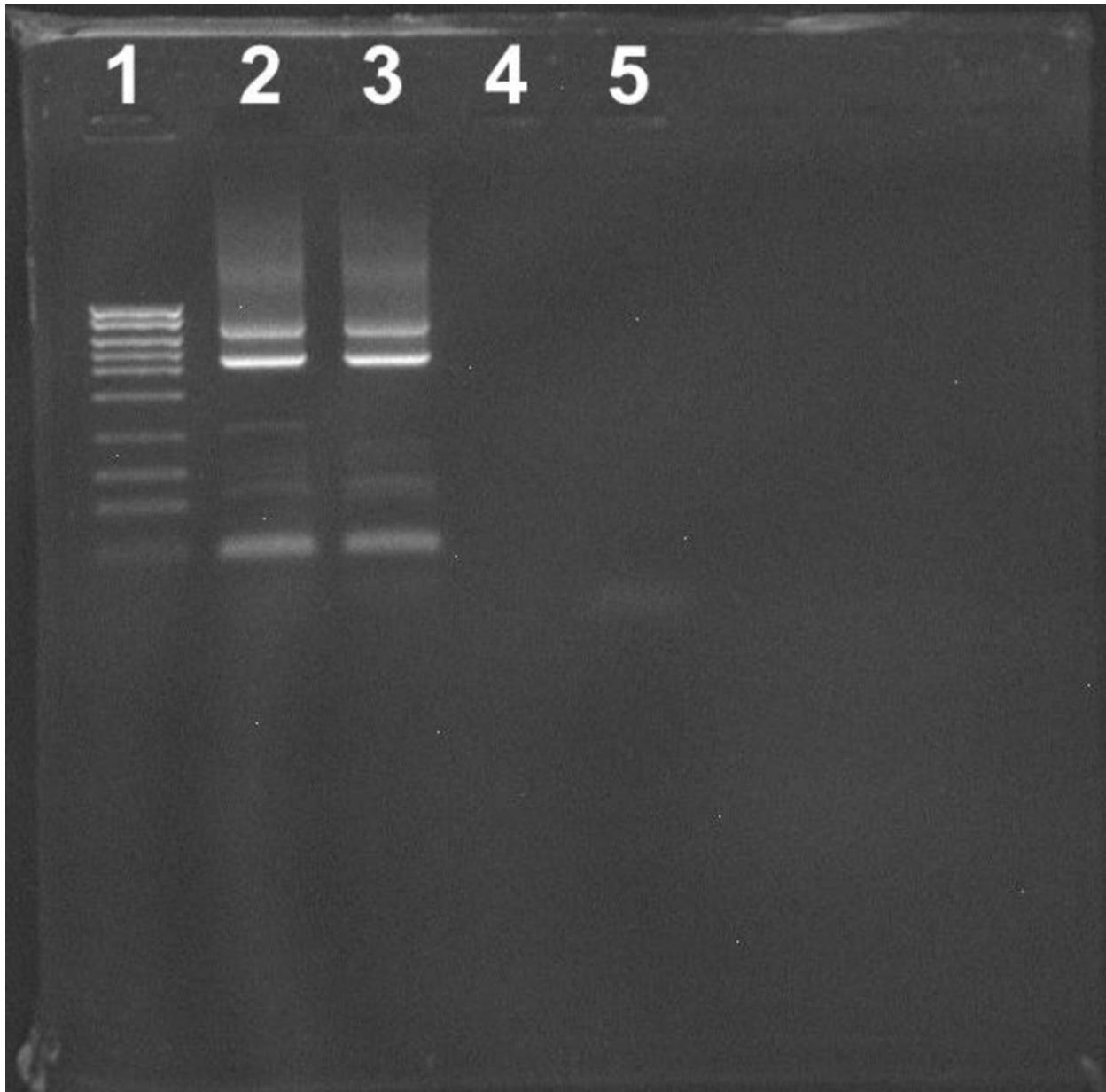
349

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

351 This is a graphical representation of the annotated complete genome of the barred owl
 352 (*Strix varia*) mitochondrion. We have color-coded the various annotations, including genes for
 353 rRNA in sky blue, tRNA genes in orange, and all other genes in bluish green. The control
 354 regions are in yellow and the goose hairpin for each control region is depicted in blue. The

355 locations of the primers we developed to amplify control regions 1 and 2 are in reddish purple.
356 The reverse complement versions of primers used (“-RC” versions) are not shown. Regions with
357 repetitive motifs are in vermillion. The base numbers around the perimeter of the figure are in
358 nucleotides. We used Geneious version 9.1.4 (Kearse et al., 2012; Biomatters, 2016) to construct
359 this figure.

360



363 This photograph of an agarose gel displays the lengths of the two products of PCR-

364 amplification of the *S. o. caurina* CR1 using primers cyb-F1 and 17122R. In lane 1 we loaded

365 Fisher BioReagents exACTGene DNA Ladder (Cat. No. BP2576100; Fisher Scientific) the ten

366 bands of which were of lengths 5,000; 4,000; 3,000; 2,500; 2,000; 1,500; 1,000; 700; 500; and

367 300 nt. Lanes 2 and 3 contained independent PCR replicates of CR1 amplification products.

368 Lane 4 was blank. In lane 5 we loaded the negative control for the PCR.

369

- 370 **3 References**
- 371 Altschul SF., Gish W., Miller W., Myers EW., Lipman DJ. 1990. Basic local alignment search
372 tool. *Journal of Molecular Biology* 215:403–410. DOI: 10.1016/S0022-2836(05)80360-2.
- 373 Barrowclough GF., Gutierrez RJ., Groth JG. 1999. Phylogeography of Spotted Owl (*Strix*
374 *occidentalis*) Populations Based on Mitochondrial DNA Sequences: Gene Flow, Genetic
375 Structure, and a Novel Biogeographic Pattern. *Evolution* 53:919–931. DOI:
376 10.2307/2640731.
- 377 Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids*
378 *Research* 27:573–580. DOI: 10.1093/nar/27.2.573.
- 379 Benson G. 2012. Tandem Repeats Finder. Version 4.07b. [Accessed 2016 Oct 1]. Available
380 from: <https://tandem.bu.edu/trf/trf.html>.
- 381 Benson G. 2016. Tandem Repeats Finder. Version 4.09. [Accessed 2016 Oct 1]. Available from:
382 <https://tandem.bu.edu/trf/trf.html>.
- 383 Benson DA., Clark K., Karsch-Mizrachi I., Lipman DJ., Ostell J., Sayers EW. 2015. GenBank.
384 *Nucleic Acids Research* 43:D30–D35. DOI: 10.1093/nar/gku1216.
- 385 Bernt M., Donath A., Jühling F., Externbrink F., Florentz C., Fritzsch G., Pütz J., Middendorf
386 M., Stadler PF. 2013. MITOS: Improved de novo metazoan mitochondrial genome
387 annotation. *Molecular Phylogenetics and Evolution* 69:313–319. DOI:
388 10.1016/j.ympev.2012.08.023.
- 389 Biomatters 2016. Geneious. Version 9.1.4. [Accessed 2016 Oct 1]. Available from:
390 <http://www.geneious.com>.
- 391 Bolger AM., Lohse M., Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence
392 data. *Bioinformatics* 30:2114–2120. DOI: 10.1093/bioinformatics/btu170.

393 Boratyn GM., Camacho C., Cooper PS., Coulouris G., Fong A., Ma N., Madden TL., Matten
394 WT., McGinnis SD., Merezhuk Y., Raytselis Y., Sayers EW., Tao T., Ye J., Zaretskaya I.
395 2013. BLAST: a more efficient report with usability improvements. *Nucleic Acids*
396 *Research* 41:W29–W33. DOI: 10.1093/nar/gkt282.

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

399 Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., Madden TL. 2009.
400 BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. DOI:
401 10.1186/1471-2105-10-421.

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

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

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

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

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

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

416 Hanna ZR., Henderson JB., Wall JD., Emerling CA., Fuchs J., Runckel C., Mindell DP., Bowie
417 RCK., DeRisi JL., Dumbacher JP. 2017. Northern spotted owl (*Strix occidentalis*
418 *caurina*) genome: divergence with the barred owl (*Strix varia*) and characterization of
419 light-associated genes. *Genome Biology and Evolution* evx158. DOI:
420 10.1093/gbe/evx158.

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

422 Henderson JB., Hanna ZR. 2016b. excerptByIDs. Version 1.0.2. Zenodo. DOI:
423 10.5281/zenodo.61134.

424 Johnson M., Zaretskaya I., Raytselis Y., Merezhuk Y., McGinnis S., Madden TL. 2008. NCBI
425 BLAST: a better web interface. *Nucleic Acids Research* 36:W5–W9. DOI:
426 10.1093/nar/gkn201.

427 Katoh K. 2016. MAFFT: a multiple sequence alignment program. Version 7.305b. [Accessed
428 2016 Oct 1]. Available from: <http://mafft.cbrc.jp/alignment/software>.

429 Katoh K., Misawa K., Kuma K., Miyata T. 2002. MAFFT: a novel method for rapid multiple
430 sequence alignment based on fast Fourier transform. *Nucleic Acids Research* 30:3059–
431 3066. DOI: 10.1093/nar/gkf436.

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

435 Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A.,
436 Markowitz S., Duran C., Thierer T., Ashton B., Meintjes P., Drummond A. 2012.
437 Geneious Basic: An integrated and extendable desktop software platform for the

- 438 organization and analysis of sequence data. *Bioinformatics* 28:1647–1649. DOI:
439 10.1093/bioinformatics/bts199.
- 440 Kumar S., Stecher G., Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis
441 version 7.0 for bigger datasets. *Molecular Biology and Evolution* 33:1870–1874. DOI:
442 10.1093/molbev/msw054.
- 443 Kumar S., Tamura K., Nei M. 1994. MEGA: Molecular Evolutionary Genetics Analysis software
444 for microcomputers. *Computer applications in the biosciences : CABIOS* 10:189–191.
445 DOI: 10.1093/bioinformatics/10.2.189.
- 446 Li H. 2013a. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
447 *arXiv:1303.3997 [q-bio]*.
- 448 Li H. 2013b. bioawk. Version 1.0. [Accessed 2016 Oct 1]. Available from:
449 <https://github.com/lh3/bioawk>.
- 450 Li H., Handsaker B., Marshall J., Danecek P. 2015. Samtools. Version 1.3 with HTSlip 1.3.1.
451 [Accessed 2016 Oct 1]. Available from: <http://www.htslip.org>.
- 452 Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin
453 R., Subgroup 1000 Genome Project Data Processing. 2009. The Sequence
454 Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. DOI:
455 10.1093/bioinformatics/btp352.
- 456 Luo R., Liu B., Xie Y., Li Z., Huang W., Yuan J., He G., Chen Y., Pan Q., Liu Y., Tang J., Wu
457 G., Zhang H., Shi Y., Liu Y., Yu C., Wang B., Lu Y., Han C., Cheung DW., Yiu S-M.,
458 Peng S., Xiaoqian Z., Liu G., Liao X., Li Y., Yang H., Wang J., Lam T-W., Wang J.
459 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo
460 assembler. *GigaScience* 1:18. DOI: 10.1186/2047-217X-1-18.

- 461 Morgulis A., Coulouris G., Raytselis Y., Madden TL., Agarwala R., Schäffer AA. 2008.
- 462 Database indexing for production MegaBLAST searches. *Bioinformatics* 24:1757–1764.
- 463 DOI: 10.1093/bioinformatics/btn322.
- 464 NCBI Resource Coordinators 2015. Database resources of the National Center for Biotechnology
- 465 Information. *Nucleic Acids Research* 43:D6–D17. DOI: 10.1093/nar/gku1130.
- 466 Omote K., Nishida C., Dick MH., Masuda R. 2013. Limited phylogenetic distribution of a long
- 467 tandem-repeat cluster in the mitochondrial control region in *Bubo* (Aves, Strigidae) and
- 468 cluster variation in Blakiston's fish owl (*Bubo blakistoni*). *Molecular Phylogenetics and*
- 469 *Evolution* 66:889–897. DOI: 10.1016/j.ympev.2012.11.015.
- 470 Pearson WR., Lipman DJ. 1988. Improved tools for biological sequence comparison.
- 471 *Proceedings of the National Academy of Sciences* 85:2444–2448.
- 472 Quinlan AR., Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic
- 473 features. *Bioinformatics* 26:841–842. DOI: 10.1093/bioinformatics/btq033.
- 474 Ruby JG. 2014. PRICE. Version 1.2. [Accessed 2016 Oct 1]. Available from:
- 475 <http://derisilab.ucsf.edu/software/price>.
- 476 Ruby JG., Bellare P., DeRisi JL. 2013. PRICE: Software for the Targeted Assembly of
- 477 Components of (Meta) Genomic Sequence Data. *G3: Genes|Genomes|Genetics* 3:865–
- 478 880. DOI: 10.1534/g3.113.005967.
- 479 Tamura K., Kumar S. 2002. Evolutionary Distance Estimation Under Heterogeneous
- 480 Substitution Pattern Among Lineages. *Molecular Biology and Evolution* 19:1727–1736.
- 481 Tamura K., Nei M. 1993. Estimation of the number of nucleotide substitutions in the control
- 482 region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and*
- 483 *Evolution* 10:512–526.

484 Zhang Z., Schwartz S., Wagner L., Miller W. 2000. A Greedy Algorithm for Aligning DNA
485 Sequences. *Journal of Computational Biology* 7:203–214. DOI:
486 10.1089/10665270050081478.
487