**A peer-reviewed version of this preprint was published in PeerJ on 10 October 2017.**



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

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

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

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

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

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

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

[6] UMR 7205 Institut de Systématique, Evolution, Biodiversité, CNRS, MNHN, UPMC, EPHE, Sorbonne Universités, Département Systématique et Evolution, Muséum National d'Histoire Naturelle, Paris, France

Corresponding Author: Zachary R Hanna
Email address: zachanna@berkeley.edu

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 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. 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*).

# 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,[1,2,3,4,7] James B. Henderson,[3,4] Anna B. Sellas,[4,5] Jérôme Fuchs,[3,6] Rauri C. K. Bowie,[1,2] John P. Dumbacher[3,4]

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

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

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

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

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

[6] UMR 7205 Institut de Systématique, Evolution, Biodiversité, CNRS, MNHN, UPMC, EPHE, Sorbonne Universités, Département Systématique et Evolution, Muséum National d'Histoire Naturelle, Paris, France

[7] Corresponding Author:

Zachary Hanna

Email address: zachanna@berkeley.edu

## 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 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. 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*).


## 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 lead to the conclusion that the mitochondrial genome of the chicken to be representative of the ancestral avian gene order (Desjardins & Morais, 1990; Mindell, Sorenson & Dimcheff, 1998a; Haddrath & Baker, 2001; Gibb et al., 2007). Almost a decade after publication of the chicken mitochondrial genome, Mindell, Sorenson & Dimcheff (1998a) described an alternative or, to use their terminology,

47 "novel" avian gene order from that of the chicken, which included a different positioning of
48 *tRNA*$^{Pro}$, *ND6*, and *tRNA*$^{Glu}$ relative to the control region sequence as well as an additional
49 noncoding segment that they hypothesized was a degraded copy of the control region. A few
50 years later, researchers first described the presence of an intact, duplicated control region in the
51 mitochondrial genomes of *Amazona* parrots (Eberhard, Wright & Bermingham, 2001) and the
52 common buzzard *Buteo buteo* (Haring et al., 2001).
53 Mindell, Sorenson & Dimcheff (1998a) detected their novel avian gene order in the
54 mitochondrial genomes of taxa in multiple avian orders that spanned a significant portion of
55 Neoaves, but did not detect it in the single owl species that they studied, *Otus asio* (Mindell,
56 Sorenson & Dimcheff, 1998a). However, further investigation of owl (Strigiformes)
57 mitochondrial genomes has revealed several surprises.
58 First, there is evidence that the duplicated control region structure and Mindell, Sorenson
59 & Dimcheff (1998a) novel gene order is present in the mitochondrial genomes of at least three
60 wood owl species, *Strix aluco*, *S. uralensis* (Brito, 2005), and *S. varia* (Barrowclough et al.,
61 2011). The use of a primer in *tRNA*$^{Thr}$ to amplify a fragment of the control region suggests that
62 the novel gene order is present in two additional wood owl species, *S. occidentalis*
63 (Barrowclough, Gutierrez & Groth, 1999) and *S. nebulosa* (Hull et al., 2010). However, the
64 novel gene order was not reported as present in the mitochondrial genome of *S. leptogrammica*
65 (Liu, Zhou & Gu, 2014).
66 Second, some species of eagle-owls (genus *Bubo*) have a large control region (up to ~
67 3,800 nucleotides) relative to *Strix*, their putative sister genus (Fuchs et al., 2008; Wink et al.,
68 2009), largely due to a tandem repeat structure in the 3' end of the control region (Omote et al.,
69 2013). Such control region tandem repeat blocks appear to be widespread in Strigidae (Xiao et

al., 2006; Omote et al., 2013). These results suggest that the structures of owl mitochondrial genomes are surprisingly dynamic and in need of further investigation, particularly for species of conservation concern for which portions of the control region are used in population genetic studies (Barrowclough, Gutierrez & Groth, 1999; Haig et al., 2004; Hull et al., 2010, 2014).

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

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

Population-level studies of the genetics of *S. occidentalis* and *S. varia* have mainly used two mitochondrial markers, a partial control region sequence (Barrowclough, Gutierrez & Groth,

1999; Haig et al., 2004; Barrowclough et al., 2005) and *cytochrome b* (*cyt b*) (Haig et al., 2004), although a phylogeographic study of *S. varia* also utilized portions of *ND6* and *COIII* (Barrowclough et al., 2011). The sequences of the complete genomes of the mitochondria of these two species will aid researchers in utilizing additional mitochondrial markers in population genetic studies of these owls.

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

## Methods

### *Strix occidentalis* sample

We utilized blood 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 (CAS:ORN:98821).

### *Illumina data*

We utilized paired-end Illumina sequence data from nine different genomic libraries constructed, sequenced, and processed as described in (Hanna et al., 2017). The raw sequences are available from the NCBI Sequence Read Archive (SRA) (SRA run accessions SRR4011595,

SRR4011596, SRR4011597, SRR4011614, SRR4011615, SRR4011616, SRR4011617, SRR4011618, SRR4011619, and SRR4011620).

*Initial assembly*

We modified BLAT version 35 (Kent, 2002, 2012) to take fastq files as input for alignments. We used our modified BLAT, which we called BLATq version 1.0.2 (Henderson & Hanna, 2016a), to find Illumina reads that aligned matches to the *Ninox novaeseelandiae* mitochondrial genome (GenBank Accession AY309457.1) (Harrison et al., 2004) (Supplementary Materials (SM) 1.1.1) and extracted those matching reads using excerptByIds version 1.0.2 (Henderson & Hanna, 2016b) (SM 1.1.2). We then used SOAPdenovo2 version 2.04 (Luo et al., 2012) to assemble those sequences (SM 1.1.3).

We used the web version of the NCBI BLAST+ version 2.2.29 tool BLASTN (Altschul et al., 1990; Zhang et al., 2000; Morgulis et al., 2008; Camacho et al., 2009) to search the NCBI nucleotide collection (Johnson et al., 2008; Boratyn et al., 2013; Benson et al., 2015; NCBI Resource Coordinators, 2015) (NCBI-nt) to assess the completeness of the resulting assembled continuous sequences (contigs) by aligning them to available mitochondrial genome sequences (SM 1.1.4). We confirmed that we had assembled a contig with the genes for $tRNA^{Phe}$ through *cyt b* to $tRNA^{Thr}$ that was approximately 18,000 nucleotides (nt) in length, but lacked the complete control region sequence. We used GNU Grep version 2.16 (Free Software Foundation, 2014) to search the Illumina reads for matches to the assembled sequence of $tRNA^{Phe}$ or $tRNA^{Thr}$ (SM 1.1.5). We found three reads that spanned $tRNA^{Phe}$ and combined them using the Geneious version 9.1.4 *de novo* assembler (Kearse et al., 2012; Biomatters, 2016) (SM 1.1.6). We then extended this assembled contig using a targeted assembly approach with the software PRICE version 1.2 (Ruby, Bellare & DeRisi, 2013; Ruby, 2014) (SM 1.1.7). This PRICE run produced

an improved and lengthened assembly after 31 cycles, but the assembly still lacked the complete control region sequence.

We used BLATq version 1.0.2 (Henderson & Hanna, 2016a) to align Illumina sequences to the assembly output by PRICE (SM 1.1.8) and extracted aligned reads using excerptByIds version 1.0.2 (Henderson & Hanna, 2016b) (SM 1.1.9). We then performed another PRICE assembly with the same initial contig as before, but with the extracted additional Illumina sequence data (SM 1.1.10). This run produced an assembly of one contig of length 18,489 nt after 26 cycles.

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

*Sanger sequencing assembly confirmation*

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 polymerase chain reaction (PCR) with primers 17589F and 41R (Table 1), which primed in $tRNA^{Glu}$ and $tRNA^{Phe}$, respectively. We then sequenced both ends of the amplified fragment using BigDye terminator chemistry (Applied Biosystems, Foster City, Calif., U.S.) on an ABI 3130xl automated sequencer (Applied Biosystems, Foster City, Calif., U.S.; SM 1.2.1). We also used primer 17572F, which primed in $tRNA^{Glu}$, and primer 41R (Table 1) to successfully amplify a 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 1), which primed outside of the repetitive region (SM 1.2.2).

We edited the sequences using Geneious version 9.1.4 (Kearse et al., 2012; Biomatters, 2016) and then used the Geneious mapper to align the sequences to the 19,946 nt preliminary mitochondrial genome assembly (SM 1.2.3). These Sanger 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 sequencing data by first amplifying CR1 with primers cytb-F1 and 17122R (Table 1), which primed in *cyt b* and *ND6*, respectively (SM 1.2.4). We visualized the PCR products on a 1% agarose gel, which revealed two PCR products approximately 2,250 and 3,500 nt in length. We re-ran the PCR and gel visualization to confirm this result, which was consistent. We then excised each band from a 1% low melting point agarose gel, performed gel purification using a Zymoclean Gel DNA Recovery Kit (Zymo Research, Irvine, Calif., U.S.), 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 (Table 1) with BigDye terminator chemistry (Applied Biosystems, Foster City, Calif., U.S.) on an ABI 3130xl automated sequencer (Applied Biosystems, Foster City, Calif., U.S.). We edited the sequences using Geneious version 9.1.4 (Kearse et al., 2012; Biomatters, 2016) 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.

*Final assembly*

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 in the draft whole genome assembly, with the 19,889 nt mitochondrial genome assembly from our targeted assembly methodology (SM 1.3.1).

We aligned all filtered Illumina sequences to this new draft reference genome using bwa version 0.7.13-r1126 (Li, 2013a) and then merged, sorted, and marked duplicate reads using Picard version 2.2.4 (http://broadinstitute.github.io/picard) (SM 1.3.2). We filtered the alignment file to only retain alignments to the preliminary targeted mitochondrial genome assembly using Samtools version 1.3 with HTSlib 1.3.1 (Li et al., 2009, 2015). We then used Samtools and GNU Awk (GAWK) version 4.0.1 (Free Software Foundation, 2012) to filter out duplicate reads, low quality alignments, secondary alignments, and alignments where both reads of a pair did not align to the mitochondrial assembly (SM 1.3.3). We next visualized the alignment across the reference sequence in Geneious version 9.1.4 (Kearse et al., 2012; Biomatters, 2016). We visually inspected all sites where there was lower coverage and any hint of disagreement between reads and, except in the case of CR1 and CR2 where we relied on the Sanger sequence data, decided in favor of majority evidence, which matched our preliminary assembly at all sites, providing confirmation of our assembly methodology.

*Final Annotation*

We annotated the final assembly using the MITOS WebServer version 806 (Bernt et al., 2013) (SM 1.4.1) and annotated the repetitive regions using the web version of Tandem Repeats Finder version 4.09 (Benson, 1999, 2016) (SM 1.4.2). We used bioawk version 1.0 (Li, 2013b) and GAWK version 4.0.1 (Free Software Foundation, 2012) to find goose hairpin sequences in CR1 and CR2 (SM 1.4.3). We compared the sequences of the annotated genes in this final assembly with those of the incomplete mitochondrial genome assembly included in the *S. occidentalis caurina* draft whole genome assembly (Hanna et al., 2017) by aligning all of the gene nucleotide sequences against a database of the scaffold-3674 gene nucleotide sequences using NCBI BLAST+ version 2.4.0 tool BLASTN (Altschul et al., 1990; Zhang et al., 2000; Morgulis et al., 2008; Camacho et al., 2009) (SM 1.4.4).

In order to visualize the binding sites of the primers that we developed to amplify CR1 and CR2 as well as the primers used by Barrowclough, Gutierrez & Groth (1999) to amplify a portion of CR1 we used Geneious version 9.1.4 (Kearse et al., 2012; Biomatters, 2016) (SM 1.4.5). We assessed the similarity of CR1 and CR2 by performing a multiple alignment using the Geneious version 9.1.4 (Kearse et al., 2012; Biomatters, 2016) implementation of MUSCLE version 3.8.425 (Edgar, 2004) (SM 1.4.6). In order to assess whether published control region sequences of related species are more similar to CR1 or CR2, we used the web version of NCBI's BLAST+ version 2.5.0 tool BLASTN (Altschul et al., 1990; Zhang et al., 2000; Morgulis et al., 2008; Camacho et al., 2009) to search NCBI-nt for sequences similar to CR1 and CR2 (SM 1.4.7). As a result of these searches, we aligned the primers used by Omote et al. (Omote et al., 2013) to amplify the control region in *Strix uralensis* to the final assembly using Geneious version 9.1.4 (Kearse et al., 2012; Biomatters, 2016) (SM 1.4.8).

We downloaded the mitochondrial genome sequences of *Alligator mississippiensis* (GenBank Accession NC_001922.1) (Janke & Arnason, 1997), *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 mitochondrial genome of *Alligator mississippiensis* served to illustrate the gene order of non-avian vertebrates as it shares the same protein-coding gene order (but not tRNA order) with mammals, amphibians, and fishes (Janke & Arnason, 1997). The *Gallus gallus* mitochondrion represents 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* is illustrative of the novel gene order first described by Mindell, Sorenson & Dimcheff (1998a) with a remnant CR2 (Gibb et al., 2007). We visualized the mitochondrial genome sequences and the accompanying annotations using Geneious version 9.1.4 (Kearse et al., 2012; Biomatters, 2016) and removed regions of the mitochondrial genomes that shared the same gene order (although the *A. mississippiensis* genome differed in the positions of tRNAs in this removed region) amongst all of the mitochondrial genomes and only retained those regions relevant to illustrating the gene rearrangements. We then used Geneious to

visualize the relevant annotations, classify the sequences into gene order categories, and create a graphic with a representative sequence from each gene order category. For a coarse assessment of gene similarity, we next used the Geneious version 9.1.4 (Kearse et al., 2012; Biomatters, 2016) implementation of MUSCLE version 3.8.425 (Edgar, 2004) 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.

*Pseudogenation of 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 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 (Altschul et al., 1990; Zhang et al., 2000; Morgulis et al., 2008; Camacho et al., 2009) (SM 1.5.1) using the default threshold Expect value (E-value) of 10. We then used GAWK version 4.0.1 (Free Software Foundation, 2012) 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 (Free Software Foundation, 2012) to reformat the BLAST output into a Browser Extensible Data (BED) formatted file (SM 1.5.3). In order to determine the mitochondrial genes spanned by each *Numt*, we used BEDTools version 2.26.0 (Quinlan & Hall, 2010) to produce a BED file of the intersection of the BED-formatted

BLAST output with the BED file output from the MITOS annotation of the final mitochondrial genome assembly (SM 1.5.4).

*Strix varia mitochondrial genome assembly*

In order to assess divergence across all genes of the mitochondrial genome between *S. occidentalis* and *S. varia*, 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. occidentalis caurina* (Haig et al., 2004). The first individual was from Marion County, Indiana, United States of America (CAS:ORN:95964), 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 (raw sequences available upon request; SM 1.6.1). We performed adapter and quality trimming of these sequence data using Trimmomatic version 0.30 (Bolger, Lohse & Usadel, 2014) (SM 1.6.2). For use in only the SOAPdenovo2 assembly, we trimmed the sequences using a different set of parameters and performed error-correction of the sequences using SOAPec version 2.01 (Luo et al., 2012) (SM 1.6.3).

The second set of *S. varia* whole-genome Illumina data originated from sequencing of a tissue sample collected in Hamilton County, Ohio, United States of America (CMC:ORNI-T:B41533), 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).

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*.

*Assembly of Strix varia ContigInput1*

In order to generate ContigInput1, which was an input to our final *S. varia* assembly, we used bwa version 0.7.13-r1126 (Li, 2013a) to align all of the trimmed CMCB41533 paired read 1 and 2 sequences to a reference sequence that included the draft *S. o. caurina* whole nuclear genome along with our final mitochondrial genome assembly (SM 1.9.1). We then merged the paired-end and unpaired read alignments, sorted the reads, and marked duplicate reads using Picard version 2.2.4 (http://broadinstitute.github.io/picard; SM 1.11.2).

We filtered the alignment file to only retain alignments to the final mitochondrial genome assembly using Samtools version 1.3 with HTSlib 1.3.1 (Li et al., 2009, 2015). We then used Samtools and GNU Awk (GAWK) version 4.0.1 (Free Software Foundation, 2012) to filter out duplicate reads, low quality alignments, secondary alignments, and alignments where both reads of a pair did not align to the mitochondrial assembly (SM 1.9.2-1.9.3). We next visualized the alignment across the reference sequence in Geneious version 9.1.4 (Kearse et al., 2012; Biomatters, 2016) and generated a consensus sequence from the alignment (SM 1.9.4). We extracted three sequences from this consensus sequence based on the *S. o. caurina* mitochondrial genome annotations and then used these extracted sequences as three separate seed contigs in an assembly using PRICE version 1.2 (Ruby, Bellare & DeRisi, 2013; Ruby, 2014) (SM 1.9.5). This run produced one contig (ContigInput1) of length 9,690 nt after 16 cycles.

*Assembly of Strix varia ContigInput2*

The series of assemblies that resulted in ContigInput2, an input to our final *S. varia* assembly, involved first using SOAPdenovo2 version 2.04 (Luo et al., 2012) to assemble all of the trimmed, error-corrected CAS95964 sequences (SM 1.10.1). This produced a contig of length 15,019 nt. We extended the contig using PRICE version 1.2 (Ruby, Bellare & DeRisi, 2013; Ruby, 2014) (SM 1.10.2). After seven cycles, this run produced an assembly of one contig of length 16,652 nt, which included the sequence for $tRNA^{Phe}$ through $tRNA^{Thr}$ and part of CR1. We used this CAS95964 contig to seed a more complete assembly using PRICE version 1.2 (Ruby, Bellare & DeRisi, 2013; Ruby, 2014) with the larger CMCB41533 Illumina sequence dataset (SM 1.11.1). After four cycles, this assembly produced one contig of length 17,073 nt, which we will refer to as "ContigInput2" below.

*Final Strix varia assembly*

We performed a final assembly using PRICE version 1.2 (Ruby, Bellare & DeRisi, 2013; Ruby, 2014) and the 9,690 nt ContigInput1 and the 17,073 nt ContigInput2 as the initial contigs (SM 1.12.1). After two cycles, this assembly produced one contig of length 19,589 nt. We then used Sanger sequencing to confirm the sequences of CR1 and CR2.

We amplified CR1 with primers cytb-F1 and 17122R (Table 1), which primed in *cyt b* and *ND6*, respectively (SM 1.12.2). We then sequenced the fragment using the original external primers as well as the internal primers CR1-F1, CR1-F1-RC, CR1-R2, CR1-R2-RC, and N1 (Table 1). We amplified CR2 with primers ND6-ext1F and 12S-ext1R (Table 1), which primed in *ND6* and *12S*, respectively (SM 1.12.3). We then sequenced the amplified fragment using the original external primers as well as the internal primers final-CR2F, 18547F, 19088R, and 19088R-RC. We performed all sequencing using BigDye terminator chemistry (Applied

Biosystems, Foster City, Calif., U.S.) on an ABI 3130xl automated sequencer (Applied Biosystems, Foster City, Calif., U.S.).

We edited the sequences using Geneious version 9.1.4 (Kearse et al., 2012; Biomatters, 2016) 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 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 (Bernt et al., 2013) (SM 1.12.4).

*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 from our final *S. o. caurina* and *S. varia* assemblies. We aligned them using MAFFT version 7.221 (Katoh & Standley, 2013; Katoh, 2014) (SM 1.13.1). We verified the alignments by eye and then used trimAl version 1.4.1 (Capella-Gutiérrez, Silla-Martínez & Gabaldón, 2009; Capella-Gutiérrez & Gabaldón, 2013) to convert the alignments to MEGA format (Kumar, Tamura & Nei, 1994; Kumar, Stecher & Tamura, 2016) (SM 1.13.2). We then used MEGA version 7.0.18 (Kumar, Stecher & Tamura, 2016) to calculate the p-distance (SM 1.13.3) and the corrected pairwise distance (Tamura & Nei, 1993) (SM 1.13.4) between *S. occidentalis caurina* and *S. varia* for each gene. We calculated a weighted average pairwise distance across all of the genes (SM 1.13.5).

## Results

The lengths of the final *S. o. caurina* and *S. varia* mitochondrial genome assemblies were 19,889 nt and 18,975 nt, respectively. As for all typical avian mitochondrial genomes, they are circular and code for 2 rRNAs, 22 tRNAs, and 13 polypeptides (Figure 1 and Figure 2). 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 Strigiformes (Fuchs et al., 2008).

Both the *S. o. caurina* and *S. varia* mitochondrial genomes contain a duplicated control region. In both genomes, CR1 and CR2 each include a C-rich sequence near the 5' end, the goose hairpin (Quinn & Wilson, 1993), which is identical across the two species and across CR1 and CR2. The *S. o. caurina* CR1 contains a 70 nt motif repeated 6.8 times near the 3' end while CR2 includes two sets of tandem repeats near the 3' end of the region, a 70 nt motif repeated 4 times followed by 9.5 repetitions of a 78 nt motif (Table 2).

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

We obtained an alignment (88.37% identity) of 1,429 nt from the 5' ends of the *S. o. caurina* and *S. varia* CR1 sequences, but it included fifteen gaps (Figure 4). 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.

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

The *S. o. caurina* CR1 is 2,021 nt in length and the *S. varia* CR1 is 1,686 nt long. In both species, the 5' end of CR1 borders $tRNA^{Thr}$ and the 3' end is adjacent to $tRNA^{Pro}$, then *ND6*, and then $tRNA^{Glu}$ (Figure 1 and Figure 2). 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 (Altschul et al., 1990; Zhang et al., 2000; Morgulis et al., 2008; Camacho et al., 2009) 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 (Altschul et al., 1990; Zhang et al., 2000; Morgulis et al., 2008; Camacho et al., 2009) searches of the *S. o. caurina* CR1 to NCBI-nt, which had the highest total score (2,177) and query coverage (96% versus 36-41% for the other matches) of the top 100 matches was to a *S. uralensis* control region sequence (GenBank Accession AB743794.1) (Omote et al., 2013). The majority of the primers

used by Omote et al. (2013) to amplify the control region in *S. uralensis* align within and around the *S. o. caurina* CR2. Four of the control-region-specific primers align to the middle of CR2 in our *S. o. caurina* sequence, which is identical to the middle of the *S. o. caurina* CR1 sequence. Perhaps most crucially, the primer L16728 aligns in the forward direction in $tRNA^{Glu}$ such that it would amplify CR2, if present in the species.

As we mentioned in the methodology, our PCR amplification of the *S. o. caurina* CR1 using primers that spanned from *cyt b* to *ND6* yielded two products approximately 2,250 and 3,500 nt in length (Figure 5). 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. This 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* (GenBank Accession EU410491.1) (Pratt et al., 2009), *Ninox novaeseelandiae* (GenBank Accession AY309457.1) (Harrison et al., 2004), *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) 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) (Panel B in Figure 6). *Melopsittacus*

*undulatus* (GenBank Accession NC_009134.1) (Guan, Xu & Smith, 2016) shares the duplicate control region mitochondrial gene order of *S. o. caurina* (Panel C in Figure 6). 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 3). The *Numts* ranged in length from 226-19,522 nt and had an average length of 3,280 nt. The *Numts* provided evidence of nuclear copies of all mitochondrial genes, except $tRNA^{Pro}$, *ND6*, and $tRNA^{Glu}$, the three genes between CR1 and CR2. *Numt* #9 (Table 3) 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, while 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 3).

*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 4). The lowest number of base substitutions per site occurs within *16S* and the highest within *ATP8* (Table 4).

## 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). In such situations, additional types of sequencing data that complement short-read data may be necessary in order to obtain a complete assembly of the mitochondrial genome. This proved to be the case in our study where the longer Sanger 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) elucidated 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, but rather made inferences 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.

Regarding the positions of *tRNA$^{Pro}$*, *ND6*, and *tRNA$^{Glu}$*, 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 amplify the control region (CR1) fragment used in their study. Hull et al. (2010) also used the Barrowclough, Gutierrez & Groth (1999) N1 primer to 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 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

498 distance from the binding site of primer N1 in *tRNA*$^{Thr}$ to the CR2 sites of primers D16 and D20
499 (3,742 nt and 3,392 nt, respectively, in our *S. o. caurina* assembly) likely reduces the probability
500 of this second primer binding site causing problems in the PCR-amplification of the CR1
501 fragment.

502 The second control region appears intact, not degraded as found in other avian taxa
503 (Mindell, Sorenson & Dimcheff, 1998a). This gene order corresponds to the "Type D Duplicate
504 CR gene order" of Gibb et al. (2007) and the "Duplicate CR gene order I" of Eberhard & Wright
505 (2016). The goose-hairpin structure is typically found near the beginning of the control region in
506 avian mitochondria (Marshall & Baker, 1997; Randi & Lucchini, 1998; Bensch & H, 2000) and,
507 in agreement with what we found, it appeared in the beginning of the intact, duplicated control
508 region sequences in the genomes of *Amazona* parrot mitochondria (Eberhard, Wright &
509 Bermingham, 2001).

510 The lengths of the *S. o. caurina* CR1 and CR2 (2,021 nt and 2,319 nt, respectively) and of
511 the *S. varia* CR1 and CR2 (1,686 nt and 1,719 nt , respectively) are all shorter than the length
512 reported for the control regions of some species in the sister genus of owls, *Bubo* (Wink et al.,
513 2009), which have lengths up to approximately 3,800 nt due to tandem repeats in the 3' end of
514 the control region (Omote et al., 2013). Similar tandem repeat blocks occur in the control regions
515 of several other owl species in the family Strigidae (Xiao et al., 2006; Omote et al., 2013). The
516 duplicated control region structure, unreported in Strigiformes outside of *Strix*, appears to have
517 arisen in the common ancestor of *Strix*, but proper phylogenetic testing of this hypothesis is
518 warranted. The length of the tandem repeat motif unit is 78 nt in the 3' end of the control region
519 sequences of *Bubo blakistoni*, *Bubo virginianus*, *Strix uralensis* (Omote et al., 2013), and *Strix*

520 *aluco* (Xiao et al., 2006); 78 nt is also the length of the motif in the longest tandem repeat block
521 in both the *S. o. caurina* and *S. varia* CR2 (Table 2).

522 As we previously mentioned, both *S. uralensis* and *S. aluco* exhibit a duplicated control
523 region structure in their mitochondrial genomes (Brito, 2005). Neither Omote et al. (2013) nor
524 Xiao et al. (2006) report the presence of a duplicated control region structure in either *S.*
525 *uralensis* or *S. aluco*, respectively, in their discussions of the repetitive content of the control
526 region sequences of these two species. It is not overtly clear from their methodologies which
527 control region they sequenced.

528 The precise primer combinations used for the amplification and sequencing of the control
529 region of *S. uralensis* are not provided by Omote et al. (2013), but mapping the primer sequences
530 used by the researchers to our *S. o. caurina* genome suggests that, if the structure of the *S.*
531 *uralensis* mitochondrial genome shares that of *S. o. caurina*, they likely sequenced CR2 in at
532 least *S. uralensis* and in the other owl species if a CR2 was present. We are unsure how
533 placement of the primers in *cyt b* and *12S*, as reported in the methodology of Xiao et al. (2006)
534 could amplify a single control region sequence for *S. aluco*, given the duplicated control region
535 structure (Brito, 2005). Further work on the structure of control region sequences in *Strix* and
536 related taxa is warranted to elucidate the pattern of evolution of this region across the Strigidae
537 phylogeny.

538 Although inconclusive and warranting further investigation, our evidence for two
539 versions of the 3' repetitive region of CR1 suggests that mitochondrial heteroplasmy is present in
540 this *S. o. caurina* individual. Mitochondrial heteroplasmy due to tandem repeat variability in the
541 control region has been shown to occur in other bird species (Berg, Moum & Johansen, 1995;
542 Mundy, Winchell & Woodruff, 1996). Previous work has suggested that the most likely

543 mechanism by which the gain and loss of such tandem repeat elements occurs in the
544 mitochondrial control region is that the repetitive region forms a stable, single-stranded
545 secondary structure and there is slippage during replication (Levinson & Gutman, 1987;
546 Wilkinson & Chapman, 1991; Fumagalli et al., 1996; Faber & Stepien, 1998). Greater numbers
547 of repeats may improve the stability of the secondary structure (Faber & Stepien, 1998).
548 Utilizing sequence from the 3' region of CR1 for population genetic study of *S. o. caurina* is not
549 likely to be useful due to the variability (in terms of the number of copies of the tandem repeat
550 motifs in this region) that is potentially present within a single individual.

551 The 29,520 nt of *Numt* sequence in the draft *S. o. caurina* nuclear genome assembly is
552 more than triple the 8,869 nt of *Numt* sequence found in a *Gallus gallus* draft nuclear genome
553 assembly (Pereira & Baker, 2004). The 3,280 nt average *Numt* size exceeds the average size in
554 all of the eukaryotic genomes examined by Richly & Leister (2004). There are markedly fewer
555 control region *Numts* in the *S. o. caurina* draft genome assembly than found in a *Gallus gallus*
556 draft genome assembly (Pereira & Baker, 2004). We only found one control region *Numt* (Table
557 3). Indeed the longest *Numt*, *Numt* #1, extends through almost the entire mitochondrial genome
558 sequence including from $tRNA^{Phe}$ through $tRNA^{Thr}$, immediately adjacent to, but ending at CR1.
559 The percentage identity of the mitochondrial pseudogenes in the nuclear genome range from
560 77.5-87.81%, so care must be taken to insure that *Numts* are not PCR-amplified in place of
561 mitochondrial gene sequences. The paucity of *Numts* including CR1 or CR2 and the shortness of
562 the one *Numt* which does (Table 3) is comforting as those are the mitochondrial regions that have
563 been used most often in studies of the population genetics of *Strix* species (Barrowclough,
564 Gutierrez & Groth, 1999; Haig et al., 2004; Barrowclough et al., 2005, 2011; Brito, 2005). As

long as researchers amplify sequences that span beyond the 592 nt *Numt* #9, they should have confidence in PCR-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% that we calculated across the non-tRNA mitochondrial genes (Table 4), 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 than the CR1 portion in *ND3*, *ND4L*, *ND6*, and *ATP8* (Table 4). We anticipate that these whole mitochondrial genome resources will be useful to those with interest in developing new mitochondrial markers to study the genetics of *S. occidentalis caurina*, *S. varia*, and related taxa.

**Acknowledgements**

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

## References


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

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

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

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

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

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

Benson G. 2012. *Tandem Repeats Finder. Version 4.07b. Available at <https://tandem.bu.edu/trf/trf.html>.*

Benson G. 2016. *Tandem Repeats Finder. Version 4.09. Available at <https://tandem.bu.edu/trf/trf.html>.*

609 Benson DA., Clark K., Karsch-Mizrachi I., Lipman DJ., Ostell J., Sayers EW. 2015. GenBank.
610 *Nucleic Acids Research* 43:D30–D35. DOI: 10.1093/nar/gku1216.
611 Berg T., Moum T., Johansen S. 1995. Variable numbers of simple tandem repeats make birds of
612 the order Ciconiiformes heteroplasmic in their mitochondrial genomes. *Current Genetics*
613 27:257–262. DOI: 10.1007/BF00326158.
614 Bernt M., Donath A., Jühling F., Externbrink F., Florentz C., Fritzsch G., Pütz J., Middendorf
615 M., Stadler PF. 2013. MITOS: Improved de novo metazoan mitochondrial genome
616 annotation. *Molecular Phylogenetics and Evolution* 69:313–319. DOI:
617 10.1016/j.ympev.2012.08.023.
618 Biomatters 2016. Geneious. Version 9.1.4. Available at <http://www.geneious.com>.
619 Bolger AM., Lohse M., Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence
620 data. *Bioinformatics* 30:2114–2120. DOI: 10.1093/bioinformatics/btu170.
621 Boratyn GM., Camacho C., Cooper PS., Coulouris G., Fong A., Ma N., Madden TL., Matten
622 WT., McGinnis SD., Merezhuk Y., Raytselis Y., Sayers EW., Tao T., Ye J., Zaretskaya I.
623 2013. BLAST: a more efficient report with usability improvements. *Nucleic Acids*
624 *Research* 41:W29–W33. DOI: 10.1093/nar/gkt282.
625 Brito PH. 2005. The influence of Pleistocene glacial refugia on tawny owl genetic diversity and
626 phylogeography in western Europe. *Molecular Ecology* 14:3077–3094. DOI:
627 10.1111/j.1365-294X.2005.02663.x.
628 Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., Madden TL. 2009.
629 BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. DOI:
630 10.1186/1471-2105-10-421.

631 Capella-Gutiérrez S., Gabaldón T. 2013. *trimAl. Version 1.4.1. Available at*
632 *<https://github.com/scapella/trimal>*.
633 Capella-Gutiérrez S., Silla-Martínez JM., Gabaldón T. 2009. trimAl: a tool for automated
634 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973.
635 DOI: 10.1093/bioinformatics/btp348.
636 CAS:ORN:95964 *CAS Ornithology (ORN). California Academy of Sciences, San Francisco,*
637 *California, United States of America.*
638 *<http://ipt.calacademy.org:8080/ipt/resource.do?r=orn> (accessed on 2016-08-15).*
639 CAS:ORN:98821 *CAS Ornithology (ORN). California Academy of Sciences, San Francisco,*
640 *California, United States of America.*
641 *<http://ipt.calacademy.org:8080/ipt/resource.do?r=orn> (accessed on 2016-08-15).*
642 CMC:ORNI-T:B41533 *CMC ORNI-T. Museum of Natural History & Science, Cincinnati*
643 *Museum Center, Cincinnati, Ohio, United States of America.*
644 De Wit P., Pespeni MH., Ladner JT., Barshis DJ., Seneca F., Jaris H., Therkildsen NO.,
645 Morikawa M., Palumbi SR. 2012. The simple fool's guide to population genomics via
646 RNA-Seq: an introduction to high-throughput sequencing data analysis. *Molecular*
647 *Ecology Resources* 12:1058–1067. DOI: 10.1111/1755-0998.12003.
648 Desjardins P., Morais R. 1990. Sequence and gene organization of the chicken mitochondrial
649 genome. *Journal of Molecular Biology* 212:599–634. DOI: 10.1016/0022-
650 2836(90)90225-B.
651 Eberhard JR., Wright TF. 2016. Rearrangement and evolution of mitochondrial genomes in
652 parrots. *Molecular Phylogenetics and Evolution* 94, Part A:34–46. DOI:
653 10.1016/j.ympev.2015.08.011.

654 Eberhard JR., Wright TF., Bermingham E. 2001. Duplication and Concerted Evolution of the
655 Mitochondrial Control Region in the Parrot Genus *Amazona*. *Molecular Biology and*
656 *Evolution* 18:1330–1342.
657 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high
658 throughput. *Nucleic Acids Research* 32:1792–1797. DOI: 10.1093/nar/gkh340.
659 Faber JE., Stepien CA. 1998. Tandemly Repeated Sequences in the Mitochondrial DNA Control
660 Region and Phylogeography of the Pike-Perches *Stizostedion*. *Molecular Phylogenetics*
661 *and Evolution* 10:310–322. DOI: 10.1006/mpev.1998.0530.
662 Free Software Foundation 2012. GNU Awk . Version 4.0.1. Available at
663 <https://www.gnu.org/software/gawk/>.
664 Free Software Foundation 2014. *GNU Grep. Version 2.16. Available at*
665 *<https://www.gnu.org/software/grep/>*.
666 Fuchs J., Pons J-M., Goodman SM., Bretagnolle V., Melo M., Bowie RC., Currie D., Safford R.,
667 Virani MZ., Thomsett S., Hija A., Cruaud C., Pasquet E. 2008. Tracing the colonization
668 history of the Indian Ocean scops-owls (Strigiformes: *Otus*) with further insight into the
669 spatio-temporal origin of the Malagasy avifauna. *BMC Evolutionary Biology* 8:197. DOI:
670 10.1186/1471-2148-8-197.
671 Fumagalli L., Taberlet P., Favre L., Hausser J. 1996. Origin and evolution of homologous
672 repeated sequences in the mitochondrial DNA control region of shrews. *Molecular*
673 *Biology and Evolution* 13:31–46.
674 Funk WC., Mullins TD., Forsman ED., Haig SM. 2007. Microsatellite loci for distinguishing
675 spotted owls (*Strix occidentalis*), barred owls (*Strix varia*), and their hybrids. *Molecular*
676 *Ecology Notes* 7:284–286. DOI: 10.1111/j.1471-8286.2006.01581.x.

677 Gibb GC., Kardailsky O., Kimball RT., Braun EL., Penny D. 2007. Mitochondrial Genomes and
678 Avian Phylogeny: Complex Characters and Resolvability without Explosive Radiations.
679 *Molecular Biology and Evolution* 24:269–280. DOI: 10.1093/molbev/msl158.
680 Guan X., Xu J., Smith EJ. 2016. The complete mitochondrial genome sequence of the
681 budgerigar, *Melopsittacus undulatus*. *Mitochondrial DNA Part A* 27:401–402. DOI:
682 10.3109/19401736.2014.898277.
683 Haddrath O., Baker AJ. 2001. Complete mitochondrial DNA geonome sequences of extinct
684 birds: ratite phylogenetics and the vicariance biogeography hypothesis. *Proceedings of*
685 *the Royal Society of London B: Biological Sciences* 268:939–945. DOI:
686 10.1098/rspb.2001.1587.
687 Haig SM., Mullins TD., Forsman ED., Trail PW., Wennerberg L. 2004. Genetic identification of
688 spotted owls, barred owls, and their hybrids: legal implications of hybrid identity.
689 *Conservation Biology* 18:1347–1357.
690 Hanna ZR., Henderson JB., Wall JD., Emerling CA., Fuchs J., Runckel C., Mindell DP., Bowie
691 RCK., DeRisi JL., Dumbacher JP. 2017. Northern spotted owl (*Strix occidentalis*
692 *caurina*) genome: divergence with the barred owl (*Strix varia*) and characterization of
693 light-associated genes. *Manuscript submitted for publication*.
694 Haring E., Kruckenhauser L., Gamauf A., Riesing MJ., Pinsker W. 2001. The Complete
695 Sequence of the Mitochondrial Genome of *Buteo buteo* (Aves, Accipitridae) Indicates an
696 Early Split in the Phylogeny of Raptors. *Molecular Biology and Evolution* 18:1892–1904.
697 Harrison GL (Abby)., McLenachan PA., Phillips MJ., Slack KE., Cooper A., Penny D. 2004.
698 Four New Avian Mitochondrial Genomes Help Get to Basic Evolutionary Questions in

699 the Late Cretaceous. *Molecular Biology and Evolution* 21:974–983. DOI:
700 10.1093/molbev/msh065.
701 Hausknecht R., Jacobs S., Müller J., Zink R., Frey H., Solheim R., Vrezec A., Kristin A., Mihok
702 J., Kergalve I., Saurola P., Kuehn R. 2014. Phylogeographic analysis and genetic cluster
703 recognition for the conservation of Ural Owls (*Strix uralensis*) in Europe. *Journal of*
704 *Ornithology* 155:121–134. DOI: 10.1007/s10336-013-0994-8.
705 Henderson JB., Hanna ZR. 2016a. *BLATq. Version 1.0.2. Zenodo. DOI: 10.5281/zenodo.61136*.
706 Henderson JB., Hanna ZR. 2016b. *excerptByIDs. Version 1.0.2. Zenodo. DOI:*
707 *10.5281/zenodo.61134*.
708 Hengjiu T., Jianwei J., Shi Y., Zhiming Z., Laghari MY., Narejo NT., Lashari P. 2016. Complete
709 mitochondrial genome of Eagle Owl (*Bubo bubo*, Strigiformes; Strigidae) from China.
710 *Mitochondrial DNA Part A* 27:1455–1456. DOI: 10.3109/19401736.2014.953090.
711 Hull JM., Englis A., Medley JR., Jepsen EP., Duncan JR., Ernest HB., Keane JJ. 2014. A New
712 Subspecies of Great Gray Owl (*Strix nebulosa*) in the Sierra Nevada of California, U.S.A.
713 *Journal of Raptor Research* 48:68–77. DOI: 10.3356/JRR-13-35.1.
714 Hull JM., Keane JJ., Savage WK., Godwin SA., Shafer JA., Jepsen EP., Gerhardt R., Stermer C.,
715 Ernest HB. 2010. Range-wide genetic differentiation among North American great gray
716 owls (*Strix nebulosa*) reveals a distinct lineage restricted to the Sierra Nevada, California.
717 *Molecular Phylogenetics and Evolution* 56:212–221. DOI:
718 10.1016/j.ympev.2010.02.027.
719 Janke A., Arnason U. 1997. The complete mitochondrial genome of *Alligator mississippiensis*
720 and the separation between recent archosauria (birds and crocodiles). *Molecular Biology*
721 *and Evolution* 14:1266–1272.

722 Johnson M., Zaretskaya I., Raytselis Y., Merezhuk Y., McGinnis S., Madden TL. 2008. NCBI
723 BLAST: a better web interface. *Nucleic Acids Research* 36:W5–W9. DOI:
724 10.1093/nar/gkn201.
725 Katoh K. 2014. *MAFFT: a multiple sequence alignment program. Version 7.221. Available at*
726 *<http://mafft.cbrc.jp/alignment/software/>*.
727 Katoh K., Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7:
728 Improvements in Performance and Usability. *Molecular Biology and Evolution* 30:772–
729 780. DOI: 10.1093/molbev/mst010.
730 Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A.,
731 Markowitz S., Duran C., Thierer T., Ashton B., Meintjes P., Drummond A. 2012.
732 Geneious Basic: An integrated and extendable desktop software platform for the
733 organization and analysis of sequence data. *Bioinformatics* 28:1647–1649. DOI:
734 10.1093/bioinformatics/bts199.
735 Kelly EG., Forsman ED. 2004. Recent Records of Hybridization Between Barred Owls (*Strix*
736 *varia*) and Northern Spotted Owls (*S. occidentalis caurina*). *The Auk* 121:806–810. DOI:
737 10.1642/0004-8038(2004)121[0806:RROHBB]2.0.CO;2.
738 Kent WJ. 2002. BLAT—The BLAST-Like Alignment Tool. *Genome Research* 12:656–664.
739 DOI: 10.1101/gr.229202.
740 Kent WJ. 2012. *BLAT—The BLAST-Like Alignment Tool. Version 35. Available at*
741 *<https://users.soe.ucsc.edu/~kent>*.
742 Krosby M., Rohwer S. 2009. A 2000 Km Genetic Wake Yields Evidence for Northern Glacial
743 Refugia and Hybrid Zone Movement in a Pair of Songbirds. *Proceedings of the Royal*
744 *Society B: Biological Sciences* 276:615–621. DOI: 10.1098/rspb.2008.1310.

745 Kumar S., Stecher G., Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis
746 version 7.0 for bigger datasets. *Molecular Biology and Evolution*:msw054. DOI:
747 10.1093/molbev/msw054.
748 Kumar S., Tamura K., Nei M. 1994. MEGA: Molecular Evolutionary Genetics Analysis software
749 for microcomputers. *Computer applications in the biosciences : CABIOS* 10:189–191.
750 DOI: 10.1093/bioinformatics/10.2.189.
751 Levinson G., Gutman GA. 1987. Slipped-strand mispairing: a major mechanism for DNA
752 sequence evolution. *Molecular Biology and Evolution* 4:203–221.
753 Li H. 2013a. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
754 *arXiv:1303.3997 [q-bio]*.
755 Li H. 2013b. *bioawk. Version 1.0. Available at <https://github.com/lh3/bioawk>*.
756 Li H., Handsaker B., Marshall J., Danecek P. 2015. *Samtools. Version 1.3 with HTSlib 1.3.1.*
757 *Available at <http://www.htslib.org>*.
758 Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin
759 R., Subgroup 1000 Genome Project Data Processing. 2009. The Sequence
760 Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. DOI:
761 10.1093/bioinformatics/btp352.
762 Liu G., Zhou L., Gu C. 2014. The complete mitochondrial genome of Brown wood owl *Strix*
763 *leptogrammica* (Strigiformes: Strigidae). *Mitochondrial DNA* 25:370–371. DOI:
764 10.3109/19401736.2013.803540.
765 Lopez JV., Yuhki N., Masuda R., Modi W., O'Brien SJ. 1994. Numt, a recent transfer and
766 tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat.
767 *Journal of Molecular Evolution* 39:174–190.

768 Luo R., Liu B., Xie Y., Li Z., Huang W., Yuan J., He G., Chen Y., Pan Q., Liu Y., Tang J., Wu
769 G., Zhang H., Shi Y., Liu Y., Yu C., Wang B., Lu Y., Han C., Cheung DW., Yiu S-M.,
770 Peng S., Xiaoqian Z., Liu G., Liao X., Li Y., Yang H., Wang J., Lam T-W., Wang J.
771 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo
772 assembler. *GigaScience* 1:18. DOI: 10.1186/2047-217X-1-18.
773 Marshall HD., Baker AJ. 1997. Structural conservation and variation in the mitochondrial control
774 region of fringilline finches (*Fringilla* spp.) and the greenfinch (*Carduelis chloris*).
775 *Molecular Biology and Evolution* 14:173–184.
776 Marthinsen G., Wennerberg L., Solheim R., Lifjeld JT. 2009. No phylogeographic structure in
777 the circumpolar snowy owl (*Bubo scandiacus*). *Conservation Genetics* 10:923–933. DOI:
778 10.1007/s10592-008-9581-6.
779 Mazur KM., James PC. 2000. Barred Owl (*Strix varia*). *The Birds of North America Online* (A.
780 Poole, Ed.) Ithaca: Cornell Lab of Ornithology; Retrieved from the Birds of North
781 America Online: https://birdsna.org/Species-Account/bna/species/brdowl. DOI:
782 10.2173/bna.508.
783 Mindell DP., Sorenson MD., Dimcheff DE. 1998a. Multiple independent origins of
784 mitochondrial gene order in birds. *Proceedings of the National Academy of Sciences*
785 95:10693–10697.
786 Mindell DP., Sorenson MD., Dimcheff DE. 1998b. An extra nucleotide is not translated in
787 mitochondrial ND3 of some birds and turtles. *Molecular Biology and Evolution* 15:1568–
788 1571.

789 Mindell DP., Sorenson MD., Dimcheff DE., Hasegawa M., Ast JC., Yuri T. 1999. Interordinal
790 Relationships of Birds and Other Reptiles Based on Whole Mitochondrial Genomes.
791 *Systematic Biology* 48:138–152. DOI: 10.1080/106351599260490.
792 Mindell DP., Sorenson MD., Huddleston CJ., Miranda HC., Knight A., Sawchuk SJ., Yuri T.
793 1997. Phylogenetic Relationships among and within Select Avian Orders Based on
794 Mitochondrial DNA. In: *Avian Molecular Evolution and Systematics*. Elsevier, 211–245.
795 Morgulis A., Coulouris G., Raytselis Y., Madden TL., Agarwala R., Schäffer AA. 2008.
796 Database indexing for production MegaBLAST searches. *Bioinformatics* 24:1757–1764.
797 DOI: 10.1093/bioinformatics/btn322.
798 Mundy NI., Winchell CS., Woodruff DS. 1996. Tandem Repeats and Heteroplasmy in the
799 Mitochondrial DNA Control Region of the Loggerhead Shrike (*Lanius ludovicianus*).
800 *Journal of Heredity* 87:21–26.
801 NCBI Resource Coordinators 2015. Database resources of the National Center for Biotechnology
802 Information. *Nucleic Acids Research* 43:D6-17. DOI: 10.1093/nar/gku1130.
803 Omote K., Nishida C., Dick MH., Masuda R. 2013. Limited phylogenetic distribution of a long
804 tandem-repeat cluster in the mitochondrial control region in *Bubo* (Aves, Strigidae) and
805 cluster variation in Blakiston's fish owl (*Bubo blakistoni*). *Molecular Phylogenetics and*
806 *Evolution* 66:889–897. DOI: 10.1016/j.ympev.2012.11.015.
807 Pereira SL., Baker AJ. 2004. Low number of mitochondrial pseudogenes in the chicken (*Gallus*
808 *gallus*) nuclear genome: implications for molecular inference of population history and
809 phylogenetics. *BMC Evolutionary Biology* 4:17. DOI: 10.1186/1471-2148-4-17.

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

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

Quinn TW., Wilson AC. 1993. Sequence evolution in and around the mitochondrial control region in birds. *Journal of Molecular Evolution* 37:417–425. DOI: 10.1007/BF00178871.

Randi E., Lucchini V. 1998. Organization and evolution of the mitochondrial DNA control region in the avian genus *Alectoris*. *Journal of Molecular Evolution* 47:449–462.

Richly E., Leister D. 2004. NUMTs in Sequenced Eukaryotic Genomes. *Molecular Biology and Evolution* 21:1081–1084. DOI: 10.1093/molbev/msh110.

Ruby JG. 2014. *PRICE. Version 1.2. Available at <http://derisilab.ucsf.edu/software/price/>*.

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

Ruegg K. 2008. Genetic, Morphological, and Ecological Characterization of a Hybrid Zone That Spans a Migratory Divide. *Evolution* 62:452–466. DOI: 10.1111/j.1558-5646.2007.00263.x.

Sorenson MD., Quinn TW. 1998. Numts: A Challenge for Avian Systematics and Population Biology. *The Auk* 115:214–221. DOI: 10.2307/4089130.

Sun X., Zhou W., Sun Z., Qian L., Zhang Y., Pan T., Zhang B. 2016. The complete mitochondrial genome of *Glaucidium brodiei* (Strigiformes: Strigidae). *Mitochondrial DNA Part A* 27:2508–2509. DOI: 10.3109/19401736.2015.1036252.

833 Tamura K., Nei M. 1993. Estimation of the number of nucleotide substitutions in the control
834 region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and*
835 *Evolution* 10:512–526.
836 Wilkinson GS., Chapman AM. 1991. Length and sequence variation in evening bat D-loop
837 mtDNA. *Genetics* 128:607–617.
838 Williford D., DeYoung RW., Honeycutt RL., Brennan LA., Hernández F., Heffelfinger JR.,
839 Harveson LA. 2014. Phylogeography of the Gambel's Quail (*Callipepla gambelii*) of
840 western North America. *The Wilson Journal of Ornithology* 126:218–235. DOI:
841 10.1676/13-111.1.
842 Wink M., El-Sayed A-A., Sauer-Gürth H., Gonzalez J. 2009. Molecular Phylogeny of Owls
843 (Strigiformes) Inferred from DNA Sequences of the Mitochondrial Cytochrome *b* and the
844 Nuclear *RAG-1* gene. *Ardea* 97:581–591. DOI: 10.5253/078.097.0425.
845 Xiao B., Ma F., Sun Y., Li Q-W. 2006. Comparative Analysis of Complete Mitochondrial DNA
846 Control Region of Four Species of Strigiformes. *Acta Genetica Sinica* 33:965–974. DOI:
847 10.1016/S0379-4172(06)60131-5.
848 Zhang Z., Schwartz S., Wagner L., Miller W. 2000. A Greedy Algorithm for Aligning DNA
849 Sequences. *Journal of Computational Biology* 7:203–214. DOI:
850 10.1089/10665270050081478.
851 Zhang Y., Song T., Pan T., Sun X., Sun Z., Qian L., Zhang B. 2016. Complete sequence and
852 gene organization of the mitochondrial genome of *Asio flammeus* (Strigiformes,
853 strigidae). *Mitochondrial DNA Part A* 27:2665–2667. DOI:
854 10.3109/19401736.2015.1043538.

855 Zink RM. 1994. The Geography of Mitochondrial DNA Variation, Population Structure,
856 Hybridization, and Species Limits in the Fox Sparrow (*Passerella iliaca*). *Evolution*
857 48:96–111. DOI: 10.2307/2410006.

858

859

## Table 1(on next page)

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 / Internal | Primer sequence (5' → 3') |
|---|---|---|---|---|
| cytb-F1 | CR1 | *S. o. caurina* and *S. varia* | External | ATCCTCATTCTCTTCCCCGT |
| 17122R | CR1 | *S. o. caurina* and *S. varia* | External | GGTGGGGGTTATTATTAACTTT |
| CR1-F1 | CR1 | *S. o. caurina* and *S. varia* | Internal | CTCSASCAAATCCCAAGTTT |
| CR1-F1-RC | CR1 | *S. o. caurina* and *S. varia* | Internal | AAACTTGGGATTTGSTSGAG |
| CR1-R2 | CR1 | *S. o. caurina* and *S. varia* | Internal | GGAGGGCGAGAATAGTTGRT |
| CR1-R2-RC | CR1 | *S. o. caurina* and *S. varia* | Internal | AYCAACTATTCTCGCCCTCC |
| N1 | CR1 | *S. o. caurina* | Internal | AACATTGGTCTTGTAAACCAA |
| 41R | CR2 | *S. o. caurina* | External | GCATCTTCAGTGCCATGCTT |
| 17572F | CR2 | *S. o. caurina* | External | ATTATCCAAGGTCTGCGGCC |
| 17589F | CR2 | *S. o. caurina* | Internal | GCCTGAAAAACCGCCGTTAA |
| 18327F | CR2 | *S. o. caurina* | Internal | CACTTTTGCGCCTCTGGTTC |
| 19911R | CR2 | *S. o. caurina* | Internal | AGAGAGGCTCTGATTGCTTG |
| ND6-ext1F | CR2 | *S. varia* | External | ACAACCCCATAATAYGGCGA |
| 12S-ext1R | CR2 | *S. varia* | External | GGTAGATGGGCATTTACACT |
| final-CR2F | CR2 | *S. varia* | Internal | TCAAACCAAACGATCGAGAA |
| 18547F | CR2 | *S. varia* | Internal | CTCACGTGAAATCAGCAACC |
| 19088R | CR2 | *S. varia* | Internal | ATTCAACTAAAATTCGTTACAAATCTT |
| 19088R-RC | CR2 | *S. varia* | Internal | AAGATTTGTAACGAATTTTAGTTGAAT |

1

## Table 2 (on next page)

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. “Entropy” is based on percent composition of bases with larger values indicating more even distribution of base composition. 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 (%) | Alignment Score | A (%) | C (%) | G (%) | T (%) | Entropy (0-2) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *S. o. caurina* | 10,267-10,309 | *ND4* | 18 | 2.3 | 19 | 84 | 4 | 52 | 25 | 46 | 0 | 27 | 1.53 |
| *S. o. caurina* | 15,066-15,162 | CR1 | 22 | 4.3 | 22 | 70 | 7 | 61 | 37 | 27 | 6 | 28 | 1.81 |
| *S. o. caurina* | 15,169-15,311 | CR1 | 67 | 2.1 | 67 | 83 | 8 | 175 | 40 | 31 | 6 | 21 | 1.78 |
| *S. o. caurina* | 16,243-16,715 | CR1 | 70 | 6.8 | 70 | 98 | 1 | 907 | 39 | 21 | 4 | 33 | 1.74 |
| *S. o. caurina* | 16,245-16,715 | CR1 | 139 | 3.4 | 139 | 99 | 0 | 924 | 39 | 22 | 4 | 33 | 1.75 |
| *S. o. caurina* | 16,403-16,515 | CR1 | 37 | 3.2 | 37 | 61 | 27 | 89 | 40 | 23 | 3 | 32 | 1.71 |
| *S. o. caurina* | 17,679-17,795 | CR2 | 44 | 2.6 | 45 | 87 | 4 | 157 | 43 | 30 | 4 | 21 | 1.72 |
| *S. o. caurina* | 17,719-17,795 | CR2 | 22 | 3.5 | 22 | 89 | 0 | 118 | 45 | 31 | 3 | 19 | 1.68 |
| *S. o. caurina* | 18,798-19,076 | CR2 | 70 | 4.0 | 70 | 99 | 0 | 551 | 39 | 21 | 4 | 34 | 1.73 |
| *S. o. caurina* | 18,800-19,076 | CR2 | 139 | 2.0 | 139 | 100 | 0 | 554 | 39 | 21 | 4 | 34 | 1.73 |
| *S. o. caurina* | 18,958-19,070 | CR2 | 37 | 3.2 | 37 | 61 | 27 | 89 | 40 | 23 | 3 | 32 | 1.71 |
| *S. o. caurina* | 19,110-19,853 | CR2 | 78 | 9.5 | 78 | 99 | 0 | 1470 | 41 | 15 | 15 | 27 | 1.88 |
| *S. varia* | 15,126-15,209 | CR1 | 22 | 3.8 | 22 | 82 | 4 | 107 | 36 | 27 | 4 | 30 | 1.78 |
| *S. varia* | 15,193-15,340 | CR1 | 67 | 2.2 | 68 | 83 | 1 | 190 | 37 | 32 | 8 | 22 | 1.83 |
| *S. varia* | 17,384-17,482 | CR2 | 22 | 4.4 | 23 | 87 | 5 | 123 | 41 | 34 | 5 | 19 | 1.73 |
| *S. varia* | 18,548-18,951 | CR2 | 78 | 5.2 | 77 | 93 | 2 | 598 | 40 | 17 | 15 | 26 | 1.89 |

1

## Table 3 (on next page)

Mitochondrial pseudogenes (*Numts*) identified in *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. We characterized some of the *Numts* by examining more than one alignment and concluding that a *Numt* spanned across those individual alignments.

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

1

## Table 4(on next page)

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

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

| Gene | Number of sites in alignment (nt) | p-distance | Distance with TN93 model |
|---|---|---|---|
| *12S* | 984 | 5.79% | 6.61% |
| *16S* | 1,589 | 5.48% | 6.14% |
| *ATP6* | 678 | 9.00% | 10.91% |
| *ATP8* | 162 | 14.81% | 21.29% |
| *COI* | 1,533 | 7.96% | 9.42% |
| *COII* | 675 | 9.04% | 11.12% |
| *COIII* | 783 | 7.54% | 8.89% |
| *cyt_b* | 1,125 | 9.33% | 11.54% |
| *ND1* | 948 | 10.65% | 13.47% |
| *ND2* | 1,029 | 9.43% | 11.73% |
| *ND3*_a | 174 | 10.92% | 13.96% |
| *ND3*_b | 174 | 11.49% | 14.86% |
| *ND4* | 1,368 | 10.23% | 12.99% |
| *ND4L* | 294 | 11.22% | 14.29% |
| *ND5* | 1,806 | 9.14% | 11.21% |
| *ND6* | 516 | 9.69% | 14.71% |

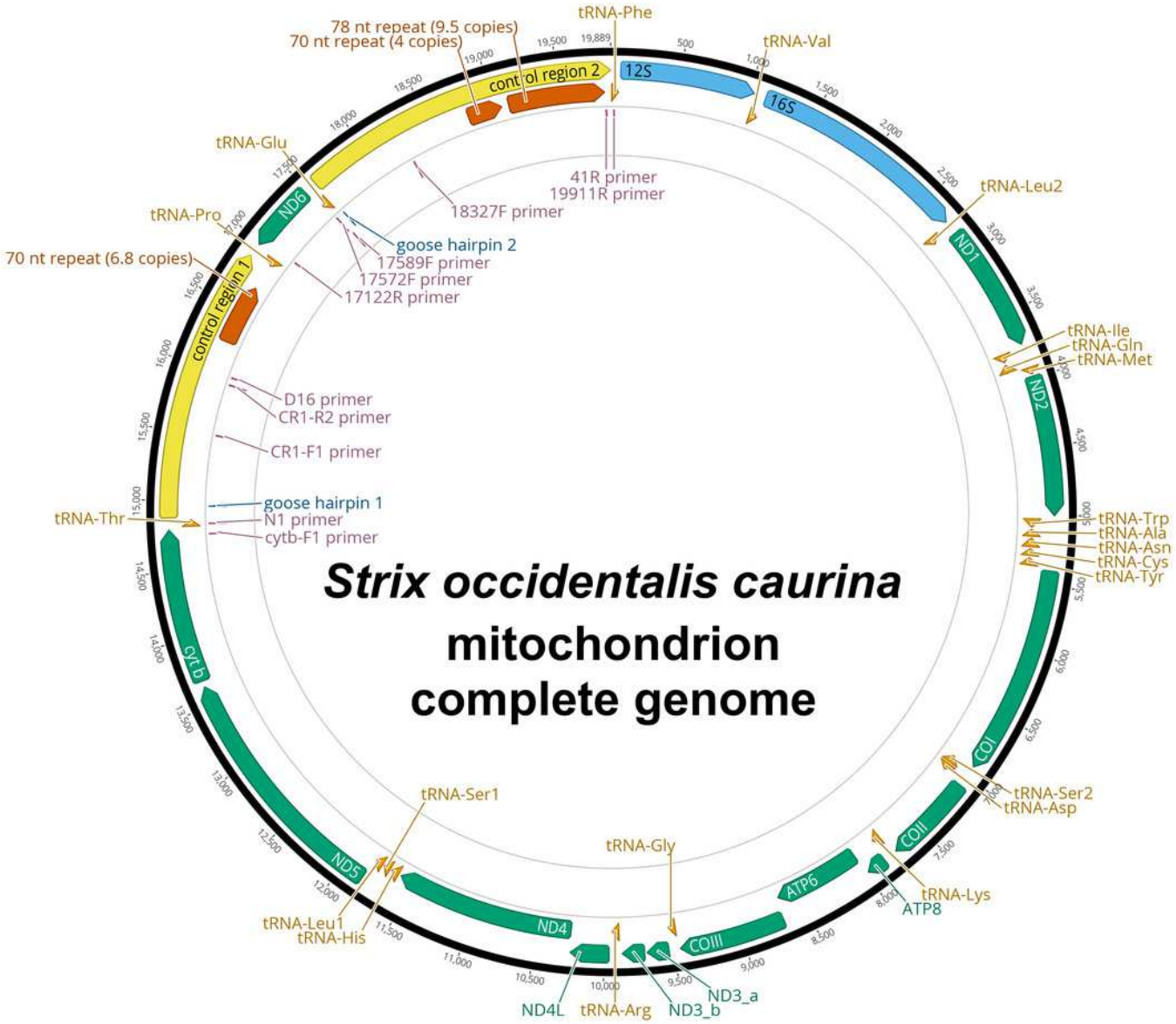
1

# Figure 1

Complete genome of *Strix occidentalis caurina* mitochondrion

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

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3120v1 | CC BY 4.0 Open Access | rec: 2 Aug 2017, publ: 2 Aug 2017

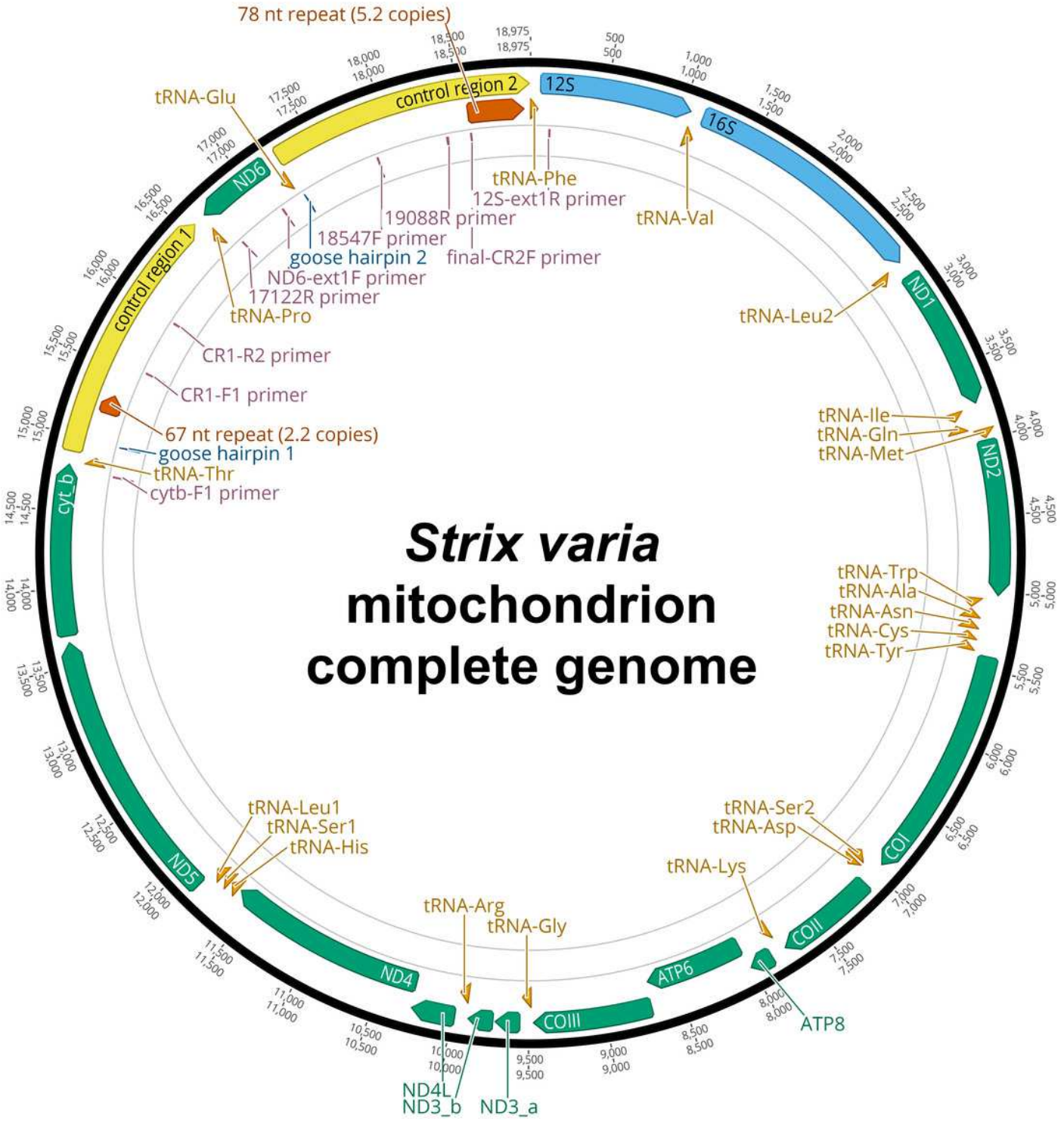

Strix occidentalis caurina
mitochondrion
complete genome
tRNA-Phe
78 nt repeat (9.5 copies)
70 nt repeat (4 copies)
control region 2
12S
tRNA-Val
16S
tRNA-Glu
ND6
41R primer
19911R primer
18327F primer
tRNA-Leu2
ND1
tRNA-Pro
goose hairpin 2
17589F primer
17572F primer
17122R primer
70 nt repeat (6.8 copies)
control region 1
tRNA-Ile
tRNA-Gln
tRNA-Met
ND2
D16 primer
CR1-R2 primer
CR1-F1 primer
tRNA-Trp
tRNA-Ala
tRNA-Asn
tRNA-Cys
tRNA-Tyr
goose hairpin 1
N1 primer
cytb-F1 primer
tRNA-Thr
COI
cyt b
tRNA-Ser2
tRNA-Asp
tRNA-Ser1
tRNA-Gly
COII
ATP6
tRNA-Lys
ATP8
ND5
ND4
COIII
tRNA-Leu1
tRNA-His
ND4L
tRNA-Arg
ND3_b
ND3_a

# Figure 2

Complete genome of *Strix varia* mitochondrion

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

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3120v1 | CC BY 4.0 Open Access | rec: 2 Aug 2017, publ: 2 Aug 2017



# Figure 3

Alignment of control regions 1 and 2 within *Strix occidentalis caurina* and *Strix varia*

Panel A depicts an alignment of the *Strix occidentalis caurina* control regions 1 and 2. Panel B displays an alignment of the *Strix varia* control regions 1 and 2. Grey 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.

# Figure 4

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

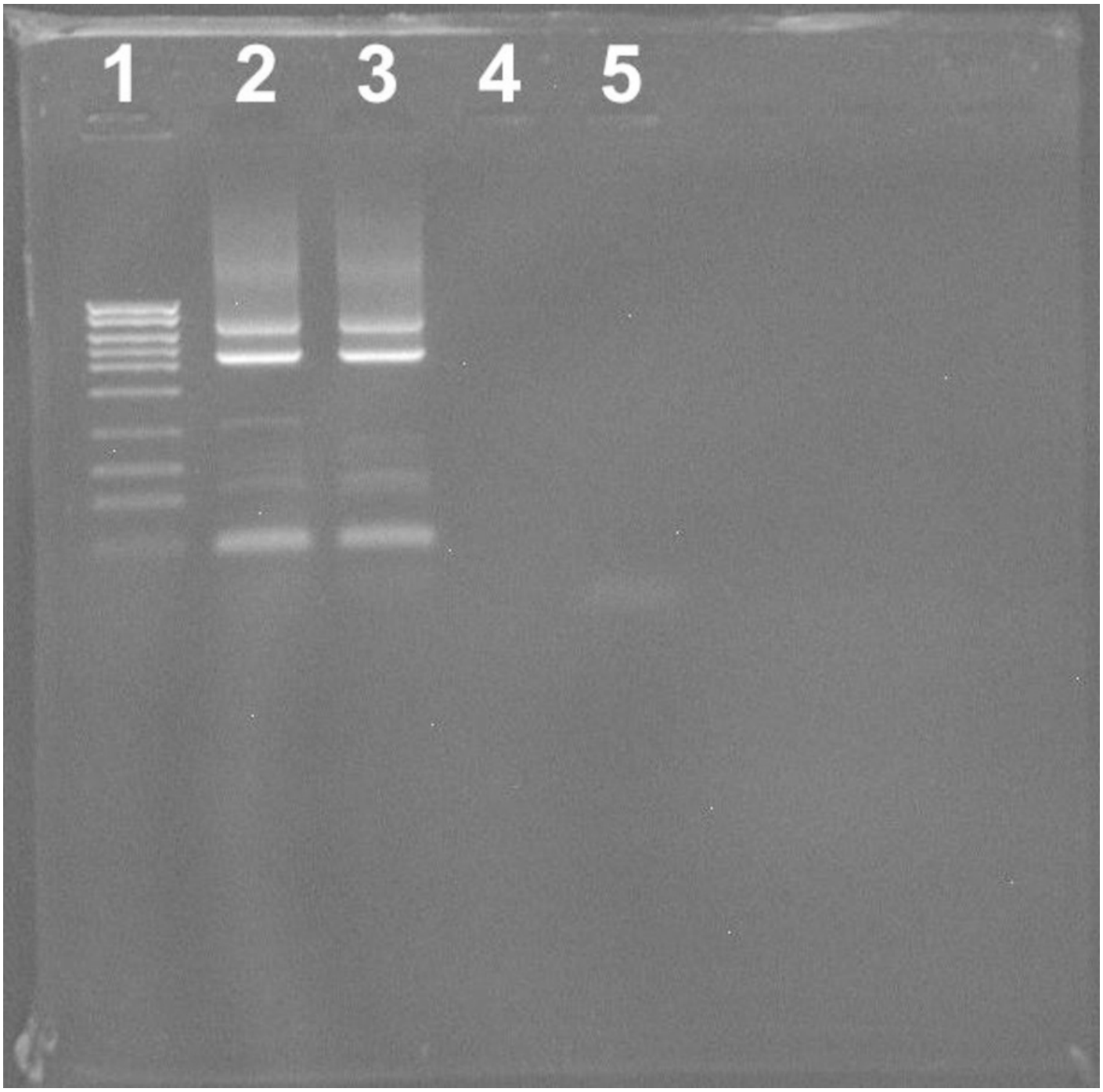
Panel A depicts an alignment of the *Strix occidentalis caurina* control region 1 with that of *Strix varia*. Panel B displays an alignment of the *Strix occidentalis caurina* control region 2 with that of *Strix varia*. Grey rectangles for each control region denote continuous sequence, whereas intervening 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 annotation of primer final-CR2F is elongated as it is situated across a gap region in the alignment.

# Figure 5

*Strix occidentalis caurina* CR1 amplification products

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

**Note: Auto Gamma Correction was used for the image. This only affects the reviewing manuscript. See original source image if needed for review.*




PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3120v1 | CC BY 4.0 Open Access | rec: 2 Aug 2017, publ: 2 Aug 2017

# Figure 6

Comparison of mitochondrial gene orders

These are four vertebrate mitochondrial gene orders. 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. “rCR2” refers to a remnant control region 2. The Alligator panel depicts the gene order of *Alligator mississippiensis*, which is representative of the non-avian gene order. The Chicken panel B 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 *Strix varia* with their duplicate control region structure. The Peregrine Falcon panel depicts the order of *Falco peregrinus*, which is representative of the “novel” gene order first described by Mindell et al. (1998a) with a control region 2 remnant following $tRNA^{Glu}$ instead of the intact control region 2 seen in the Spotted Owl panel (Gibb et al., 2007). 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 base numbers on the top of the figure in nucleotides.