



Phylogenomics of the extinct Heath Hen provides support for sex-biased introgression among extant prairie grouse

Jeff A. Johnson^{a,b,*}, Ben Novak^c, Giridhar Athrey^d, Andrew G. Sharo^e, Tom Chase^f, John Toepfer^{g,1}

^a The Peregrine Fund, Boise, ID 83709, USA

^b Department of Biological Sciences, University of North Texas, Denton, TX 76203, USA

^c Revive & Restore, Sausalito, CA 94965, USA

^d Department of Poultry Science & Faculty of Ecology and Evolutionary Biology, Texas A&M University, College Station, TX 77843, USA

^e Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, Santa Cruz, CA 95064, USA

^f Village and Wilderness, Martha's Vineyard, MA 02557, USA

^g George Miksch Sutton Avian Research Center, Bartlesville, OK 74005, USA

ARTICLE INFO

Keywords:

Sex-biased introgression
Hybridization
Sex-chromosome
Historical DNA
Species tree
Extinct

ABSTRACT

Rapid divergence and subsequent reoccurring patterns of gene flow can complicate our ability to discern phylogenetic relationships among closely related species. To what degree such patterns may differ across the genome can provide an opportunity to extrapolate better how life history constraints may influence species boundaries. By exploring differences between autosomal and Z (or X) chromosomal-derived phylogenetic patterns, we can better identify factors that may limit introgression despite patterns of incomplete lineage sorting among closely related taxa. Here, using a whole-genome resequencing approach coupled with an exhaustive sampling of subspecies within the recently divergent prairie grouse complex (genus: *Tympanuchus*), including the extinct Heath Hen (*T. cupido cupido*), we show that their phylogenomic history differs depending on autosomal or Z-chromosome partitioned SNPs. Because the Heath Hen was allopatric relative to the other prairie grouse taxa, its phylogenetic signature should not be influenced by gene flow. In contrast, all the other extant prairie grouse taxa, except Attwater's Prairie-chicken (*T. c. attwateri*), possess overlapping contemporary geographic distributions and have been known to hybridize. After excluding samples that were likely translocated prairie grouse from the Midwest to the eastern coastal states or their resulting hybrids with mainland Heath Hens, species tree analyses based on autosomal SNPs consistently identified a paraphyletic relationship with regard to the Heath Hen with Lesser Prairie-chicken (*T. pallidicinctus*) sister to Greater Prairie-chicken (*T. c. pinnatus*) regardless of genic or intergenic partitions. In contrast, species trees based on the Z-chromosome were consistent with Heath Hen sister to a clade that included its conspecifics, Greater and Attwater's Prairie-chickens (*T. c. attwateri*). These results were further explained by historic gene flow, as shown with an excess of autosomal SNPs shared between Lesser and Greater Prairie-chickens but not with the Z-chromosome. Phylogenetic placement of Sharp-tailed Grouse (*T. phasianellus*), however, did not differ among analyses and was sister to a clade that included all other prairie grouse despite low levels of autosomal gene flow with Greater Prairie-chicken. These results, along with strong sexual selection (i.e., male hybrid behavioral isolation) and a lek breeding system (i.e., high variance in male mating success), are consistent with a pattern of female-biased introgression between prairie grouse taxa with overlapping geographic distributions. Additional study is warranted to explore how genomic components associated with the Z-chromosome influence the phenotype and thereby impact species limits among prairie grouse taxa despite ongoing contemporary gene flow.

* Corresponding author at: The Peregrine Fund, Boise, ID 83709, USA.

E-mail address: johnson.jeff@peregrinefund.org (J.A. Johnson).

¹ Deceased.

1. Introduction

Several phylogenetic studies with recently divergent avian taxa have identified elevated divergence among single nucleotide polymorphisms (SNPs) located on their sex chromosomes (ZW), specially the Z chromosome compared to their autosomal SNPs (Lavretsky et al., 2015; Bourgeois et al., 2020). This pattern, broadly described as the Faster-Z Effect, can be influenced by multiple factors such as selection on genes located on or linked to the Z chromosome that help maintain barriers to gene exchange (Dean et al., 2015; Fraïsse and Sachdeva, 2021) and stronger genetic drift because the effective population size of the Z chromosome (N_{eZ}) is $\frac{1}{4}$ that of autosomes (N_{eA}) (Charlesworth, 2009; Mank et al., 2010a,b; Ellegren, 2011). In species with high variance in male reproductive success and strong sexual selection, the effective size difference between Z and autosomal chromosomes (N_{eZ}/N_{eA}) is even further exaggerated (Laporte and Charlesworth, 2002;

Ellegren, 2009; Hayes et al., 2020), with ratios as low as 20% (e.g., Galla and Johnson, 2015) thereby further increasing divergence rate for Z-linked loci relative to autosomal.

The resulting differences in divergence rate and its effect on lineage sorting between Z and autosomal SNPs can be useful for investigating phylogenetic relationships among closely related taxa. However, not all studies have the resources available to partition their genomic datasets by chromosomes, let alone identify SNPs within exons, introns, or intergenic regions. Although it is becoming increasingly cost-effective to generate reduced representation (e.g., RADseq) or whole-genome resequencing data for large-scale phylogenetics, to what degree the strength of selection or drift may influence our ability to recover accurate species-level phylogenetic patterns deserves further study. Many phylogenetic studies focused on non-model species, for example, lack an annotated reference genome for partitioning SNPs (e.g., exons, introns, intergenic). An inability to partition SNPs within the genome may have important

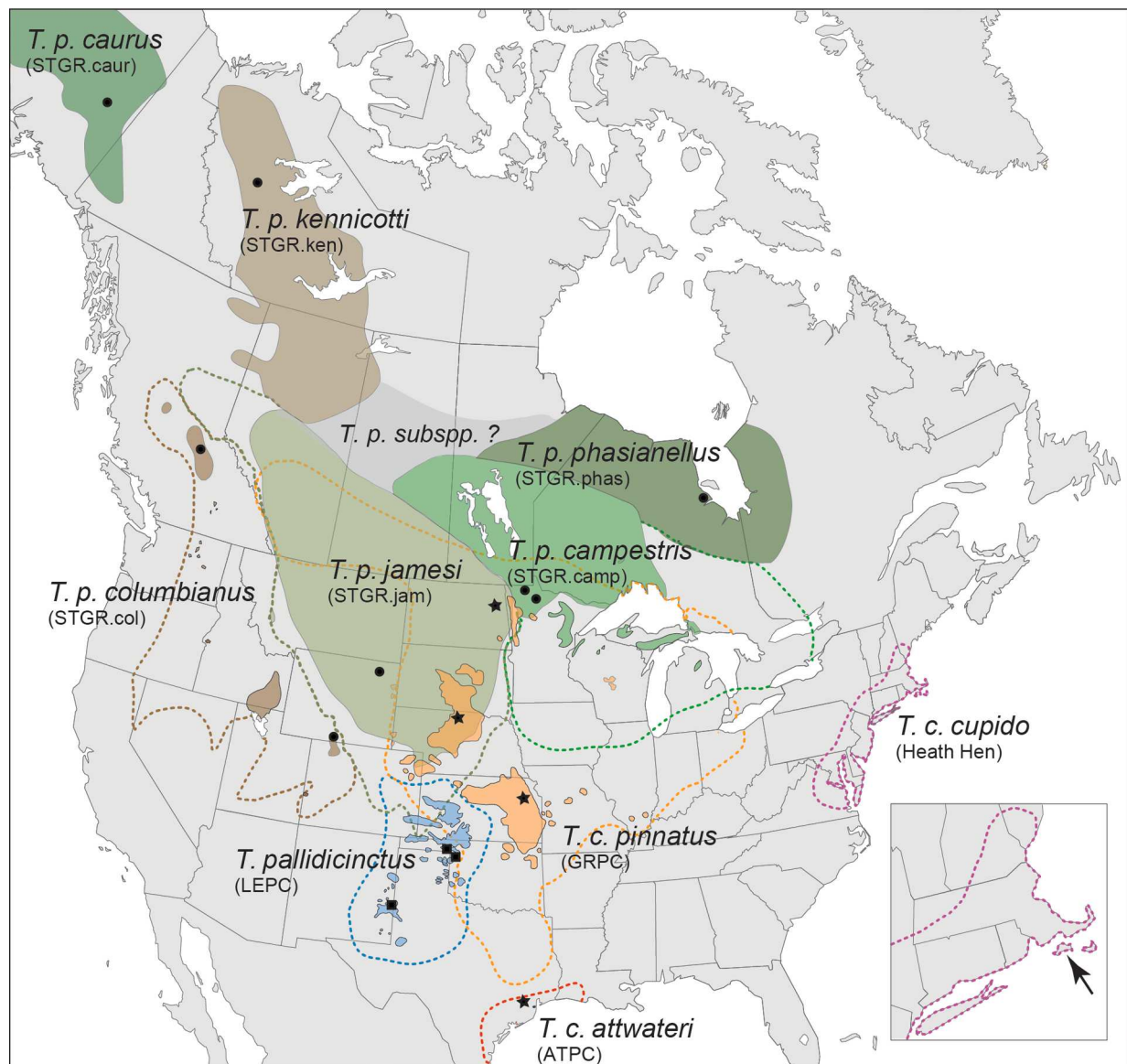


Fig. 1. North American prairie grouse species' and subspecies' historic (dotted) and contemporary (solid) distributions. Sharp-tailed Grouse (*T. phasianellus caurus*, STGR.caur; *T. p. kennicotti*, STGR.ken; *T. p. phasianellus*, STGR.phas; *T. p. campestris*, STGR.camp; *T. p. jamesi*, STGR.jam; *T. p. columbianus*, STGR.col-BC & col-WY), Heath Hen (*T. cupido cupido*), Greater Prairie-Chicken (*T. c. pinnatus*, GRPC), Attwater's Prairie-chicken (*T. c. attwateri*, ATPC), and Lesser Prairie-chicken (*T. pallidicinctus*, LEPC). Approximate sampling locations for each contemporary sample are shown a solid symbol for Greater and Attwater's Prairie-chicken (star), Lesser Prairie-chicken (square) and Sharp-tailed Grouse (circles). Map insert showing location of Martha's Vineyard (arrow) where the last Heath Hen was observed prior to its extinction in 1932. Map reproduced based information from: Johnsgard (2002); Connelly et al. (2020); Hagen and Giesen (2020); Johnson et al. (2020).

implications for our interpretation of phylogenetic patterns. We cannot assume, for example, that the recovered SNPs have experienced similar demographic histories across the entire genome because regions can differ with respect to how selection and other constraints may allow or prevent gene flow between taxa (Nosil et al., 2009; Ravinet et al., 2017). Therefore, the phylogenetic inferences made among select taxa may vary depending on the genomic dataset and the degree of reproductive isolation and its impact on the genome.

In this study, we further explored this effect within North American prairie grouse (genus *Tympanuchus*), a species complex that has experienced strong sexual selection following divergence from their common ancestor approximately 900,000 to 3.5 million years ago based on a fossil calibrated species tree using Z-linked sequence data (Galla and Johnson, 2015). North American prairie grouse include three species, two of which include multiple subspecies (Fig. 1). Sharp-tailed grouse (*T. phasianellus*) has the largest and most northern distribution that extends from the Canadian Arctic and Alaska south into Washington, Utah, Colorado, and Nebraska, and to the east through Minnesota, Wisconsin, and northern Michigan. The species includes at least six subspecies that largely correspond with geography (Connelly et al., 2020; Supplemental Table 1). The Lesser Prairie-chicken (*T. pallidicinctus*) is distributed from southwestern Kansas into southeastern Colorado through western Oklahoma and Texas into northeastern New Mexico with populations that are now highly fragmented (Hagen and Giesen, 2020).

The third and final species (*T. cupido*) includes three subspecies, each of which are commonly referenced by subspecies. The Greater Prairie-chicken (*T. cupido pinnatus*) was once found throughout the Midwest in short and long-grass prairie habitat from North Dakota south to Texas and east to Wisconsin, Michigan, and Ohio, but now with highly fragmented and locally extinct populations throughout much of its distribution (Johnson et al., 2020). The critically endangered Attwater's Prairie-chicken (*T. c. attwateri*) is currently restricted to two small management areas with numbers fewer than 100 in the wild (Silvy et al., 2004; M. Morrow, per. comm.) but previously extended along the gulf coastal plains of Texas and southwest Louisiana. The full extent of the historical distribution of Attwater's is not fully known and whether its range overlapped with other prairie grouse to the north including both Lesser and Greater Prairie-chickens (Silvy et al., 2004; Johnson et al., 2020). Lastly, the extinct Heath Hen (*T. c. cupido*) once occupied coastal scrub grassland habitat from Maine and south to Virginia but was extirpated from the mainland as early as the 1870s, and thereafter found only on the island of Martha's Vineyard, MA (Gross, 1928) until its extinction in 1932 (Johnsgard, 2002; Palkovacs et al., 2004; Johnson and Dunn, 2006).

In areas of sympatry, Sharp-tailed Grouse and Greater Prairie-chicken are known to hybridize (Bent, 1932; Johnsgard and Wood, 1968; Sparling, 1980; Augustine and Trauba, 2015; Huschle and Toepfer, 2020) as do Lesser and Greater Prairie-chickens (Bain and Farley, 2002; Oyler-McCance et al., 2016). Introgression among documented prairie-grouse appears largely sex-biased with no known hybrid males observed breeding in the wild, whereas hybrid females have produced fertile offspring that contribute to at least the F2 generation (J. Toepfer, unpubl. data; see also Huschle and Toepfer, 2020). This pattern was supported in a previous study showing strong differentiation of alleles from five introns on the Z chromosome among all three extant prairie grouse species, yet mitochondrial control region haplotypes and autosomal intron alleles were largely shared among species (Galla and Johnson, 2015). All three prairie grouse species possess a lek breeding behavior with males of each species having distinct sexual selected traits, yet females are more similar in appearance. Therefore, despite introgression as shown with mtDNA and autosomal loci, strong sexual selection and male-biased postzygotic behavioral isolation (i.e., "unsexy son") may help maintain barriers to gene exchange on the Z-chromosome among prairie grouse species with overlapping distributions (Galla and Johnson, 2015; see also Gould and Augustine, 2020).

It is not known if the same patterns exist with the extinct Heath Hen,

which was allopatric with the other prairie grouse taxa. Although taxonomically, the Heath Hen is conspecific with the Greater and Attwater's Prairie-chicken, mitochondrial control region sequence data suggested that it was more closely related to the Lesser Prairie-chicken (Johnson and Dunn, 2006; see also Palkovacs et al., 2004; Johnson, 2008). As the Heath Hen was geographically isolated (Johnsgard, 2002; Johnson et al., 2020), to what degree the pattern observed in Palkovacs et al. (2004) and Johnson and Dunn (2006) based on mtDNA reflects its phylogenetic history requires further study using nuclear DNA specifically.

Among prairie grouse, a high variance in male mating success and strong sexual selection has further reduced the effective population size of the Z chromosome compared to autosomes beyond what is expected based on the female heterogametic ZW inheritance pattern and female-biased dispersal alone (Ellegren, 2009). Consequently, this effect has thereby increased Z chromosome divergence rate compared to autosomes among prairie grouse due to stronger genetic drift (i.e., Faster-Z effect; Mank et al., 2010a,b; Hayes et al., 2020; see Galla and Johnson, 2015 for further discussion). In fact, the same could be said with mtDNA but in the opposite direction where the effective population size among mtDNA loci can exceed autosomal among species with high variance in male mating success and female biased dispersal (Chesser and Baker, 1996; Laporte and Charlesworth, 2002). As has been shown previously, phylogenetic analyses using Z-linked SNPs are more effective than autosomal and mtDNA for resolving phylogenetic relationships among extant prairie grouse taxa due to differences in their respective effective population sizes (Galla and Johnson, 2015). To our knowledge, no phylogenetic analyses using nuclear loci, let alone from the Z chromosome, have been published investigating the Heath Hen and its taxonomic relationship with other prairie grouse taxa.

Here we use a whole-genome resequencing approach to characterize the phylogenomic relationships among prairie grouse taxa, including the Heath Hen, using both mtDNA and nuclear datasets. We employed multiple methods to partition our nuclear whole-genome dataset into autosomal and Z chromosome SNPs, including the use of a transcriptome to further annotate SNPs according to their position within genes and intergenic regions. These approaches allowed us to provide a robust analysis exploring how different genomic partitioned SNP datasets, including demographic effects such as potential historic gene flow, influence our phylogenomic results of a recently divergent prairie grouse species complex. In so doing, we provide a more conclusive assessment of the taxonomic placement of the extinct Heath Hen relative to other prairie grouse taxa and generate additional evidence in support of historic gene flow in influencing our ability to resolve taxonomic relationships among prairie grouse taxa.

2. Material and methods

2.1. Sample collection and DNA extraction

A total of 53 prairie grouse samples from all eleven *Tympanuchus* subspecies, including Greater Sage-grouse (*Centrocercus urophasianus*), were collected for this study, which included 39 samples collected from contemporary blood or muscle tissues and 14 Heath Hen samples collected from museum toe-pad tissues (Table S1). All but three of the Heath Hen samples originated from the island of Martha's Vineyard, MA. The geographic origins of the remaining three Heath Hen samples are unknown as they were collected from markets in Philadelphia, Long Island, and Washington DC. Mainland (ML) and Martha's Vineyard (MV) Heath Hen samples were analyzed separately because initial analyses suggested differentiation between the two sets of samples.

Genomic DNA was extracted from all contemporary samples using the Qiagen DNeasy Blood & Tissue Kit using the manufacturer's protocol for nucleated blood. All DNA extractions and library preparation for museum samples were performed at the UCSC Paleogenomics Lab in a room designated for work only with historic specimens to prevent

potential cross-contamination with modern tissue samples. Clean suits, facial masks, and sterile gloves were worn at all times when handling historic specimens, per the standard protocols for working with ancient DNA (Cooper and Poinar, 2000). DNA extractions were performed according to the protocols proven for avian toe-pad specimens by Soares et al. (2016). DNA concentrations for all samples were quantified following extraction using a Qubit Fluorometer (ThermoFisher Scientific).

2.2. Reference genome

A high-quality reference genome using a female Greater Prairie-chicken collected in Nebraska was assembled by Dovetail Genomics (dovetailgenomics.com; Santa Cruz, CA, USA) using methods as described elsewhere (Putnam et al., 2016). The initial assembled genome was sequenced at 449 depth of coverage resulting in 12,186 scaffolds with a scaffold N50 of 12.1 Mb (Genbank assembly accession: GCA_001870855.1). A blood sample obtained from a male Attwater's Prairie-chicken at Houston Zoo was used to generate Hi-C data that was then aligned to the Dovetail Greater Prairie-chicken draft genome assembly to produce a chromosome-length Hi-C assembly resulting in 11,967 scaffolds with an N50 of 68.6 Mb (https://www.dnazoo.org/asssemblies/Tympanuchus_cupido; Dudchenko et al., 2017). The Hi-C assembly was further organized based on synteny alignment to the chicken genome assembly (GRCg6a; Genbank assembly accession: GCA_000002315.5) using the program Satsuma v. 3.1.0 (Grabherr et al., 2010). We performed this step to help facilitate genome partitioning of SNPs used in downstream analyses. The final reference genome used in this study was 989,026,393 bp in length consisting of 1,987 scaffolds with an N50 of 88.3 Mb. Ninety percent of the genome was contained within 15 scaffolds (minimum length of 13.6 Mb) with 99.7% of the genome within scaffolds > 100 kb.

Organ tissues (lung, kidney, heart, and brain) collected from two female Greater Prairie-chickens in Nebraska, one of which was the bird used for the high-coverage reference genome described above, were used to generate RNAseq reads for annotation of the reference genome. Tissues were stored in RNeasy Lysis Buffer (Qiagen) and kept frozen until total RNA isolation. RNA from all tissues from each bird were pooled prior to preparing Illumina libraries for sequencing. The libraries were enriched for mRNA prior to sequencing and the two libraries were sequenced on a HiSeq 2500 (PE 125 bp) at Texas A&M AgriLife Genomics and Bioinformatics Service. We generated a total of 150 M PE reads from each library. The resulting reads were quality filtered to retain high quality (Phred > Q30) reads using the tool Trim Galore (Babraham Software). High-quality reads were used to generate a *de novo* transcriptome assembly using the software Trinity using default settings. The resulting assembled transcripts were verified and annotated based on homology-based searches against the UniProt database using the tool TransDecoder, which uses HMM and BLAST-based searches on the translated sequences to annotate the most likely coding sequences. The resulting set of verified complete transcripts (longest ORF) was retained for further genome annotation.

Next we annotated the HiC-Satsuma assembly using the *ab initio* gene prediction and annotation program using Augustus with Maker v.3 pipeline (Cantarel et al., 2008). The longest ORFs output from TransDecoder were used as the transcript (expressed sequence tags, EST) evidence for genome annotation. We used three rounds of annotation, and at each successive round, we provided the improved predictions to develop the annotation set. The first training run with Augustus was performed without evidence, and the resulting GFF file was supplied along with EST evidence for subsequent refinement steps. We stopped after three rounds of annotation when over 80% of the generated annotation features had an annotation edit distance (AED) of <0.5. The resulting annotation set (GFF file) was used for downstream analyses.

2.3. Whole genome library preparation and resequencing

A total of 43 samples from all prairie grouse subspecies (34 contemporary and 9 museum) and two Greater Sage-grouse were selected for library preparation and whole genome resequencing based on samples of adequate concentration and quality following DNA extraction (Table S2). Each of the eleven prairie grouse subspecies were represented by between two to nine samples, with the Heath Hen having the largest number of samples.

Illumina sequencing libraries were prepared for a subset of samples (n = 18), including all 9 Heath Hen and 9 contemporary prairie grouse samples, following Meyer and Kircher (2010), and then pooled and sequenced on a HiSeq 2500 in 2x150bp paired-end (PE) mode at Texas A&M AgriLife Genomics Core Facility (Heath Hen samples, 5 lanes). After our initial analyses, a second set of libraries were prepared with additional contemporary samples (n = 25) using the Illumina Nextera DNA Flex library preparation kit and sequenced on one S4 150PE Flow Cell lane with samples from a separate study on an Illumina NovaSeq 6000 platform at North Texas Genome Center at the University of Texas in Arlington. Approximately 3.5 billion raw reads were generated from a total of 43 demultiplexed samples. All raw reads generated for this study are available in the NCBI Sequence Read Archive (BioProject accession: PRJNA927031). A third Greater Sage-grouse originally sampled in Colorado and available on Genbank (SRA accession SRS4596009) was included with all the raw sequenced reads for processing.

2.3.1. Nuclear genome data processing

All raw sequence reads were processed using standard bioinformatic tools and filtering criteria when working with contemporary and museum sample whole genome resequencing data. Specifically, adapters were removed using SeqPrep (<https://github.com/jstjohn/SeqPrep>) and CutAdapt (Martin, 2011) with a minimum Phred quality score of 20. SeqPrep was also used with default settings to merge overlapping paired-end reads for only the museum samples to increase fragment length and improve mapping efficiency.

Trimmed reads from the contemporary samples and merged reads from the museum samples were mapped to the HiC-Satsuma Greater Prairie-chicken assembly using BWA-MEM (Li, 2013) or BWA-ALN (Li and Durbin, 2009) v.0.7.17 with default parameters, respectively. Duplicates were removed from mapped reads, and then sorted and indexed using Samtools 1.6 (Li et al., 2009). Of the total set of prairie grouse prepared libraries, eight possessed a low average depth of coverage (<4x) following initial processing (GRPC, n = 3; LEPC, n = 2; HEHE, n = 1; STGR, n = 2; Table S2) and were excluded from further analysis to avoid any bias associated with high levels of missing sites.

The program mapDamage v2.2.0 (Jónsson et al., 2013) was used to confirm the presence of characteristic genomic DNA damage patterns in the Heath Hen samples that are often observed in museum or ancient specimens (Sawyer et al., 2012). Although a low level of DNA C to T and G to A misincorporations were observed at read termini of the Heath Hen mapped reads (Fig. S1), BAM base quality scores were rescaled using mapDamage and additional filtering steps were applied to help mitigate the effects of DNA damage in our Heath Hen consensus sequences and reduce potential bias in downstream analyses when including ancient DNA samples.

All mapped reads were realigned to minimize mismatched bases using RealignerTargetCreator and IndelRealigner with GATK v3.8 (McKenna et al., 2010). Variant calling was then performed using GATK HaplotypeCaller (in GVCF mode) and GenotypeGVCFs with default parameters (Poplin et al., 2018). Further quality filtering was performed using recommended GATK hard filters (QD < 2.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0 || SOR > 3 || DP < 3), with both indels and SNPs within 5 base pairs removed while retaining only biallelic SNPs using BCFtools v1.10.2 (Danecek et al., 2021) and custom scripts.

The resulting filtered SNP dataset was partitioned into genic and

intergenic VCF files using the generated transcriptome assembly and then separated into autosomal and Z chromosome datasets. We also excluded sites with heterozygosity >60% (to remove possible paralogs) and missingness above 10% among all samples and required heterozygous genotypes to have an allele balance of 20–80%. To further reduce the total number of SNPs for the autosomal datasets to adjust for a faster run time when conducting the phylogenetic analyses, SNPs were then pruned for linkage disequilibrium using BCFtools with the r^2 set to 0.6 and a window size of 1,000 sites. All singletons, or sites that are polymorphic at only one sample within the dataset, were also removed because such sites are likely to be more common in degraded samples (Axelsson et al., 2008).

2.3.2. Mitochondrial genome data processing

Raw fastq files were processed using fastp (Chen et al., 2018) using default parameters, and then mapped to the Greater Prairie-chicken mitochondrial (mtDNA) reference genome (GenBank accession: MW574394.1) using Mapping Iterative Assembler (MIA; Green et al., 2008). All samples were mapped to the reference using a substitution matrix file adjusted for working with damaged aDNA specifically (-s ancient.submat.txt) and a kmer filter of length 14 (-k 14) to help accommodate the mapping process. Following mapping, identical reads or duplicates were removed and a consensus mtDNA genome sequence was determined for each sample requiring a minimum of three reads to call a base at each site and 2/3 agreement between mapped reads. Sites not meeting those criteria were identified as a missing (N) within each samples' consensus sequence. The program MUSCLE v5.1 (Edgar, 2021) was then used to align all consensus mtDNA sequences using default parameters.

2.4. NeighborNet network and species tree inference

NeighborNet networks were constructed for the mtDNA, autosomal and Z-chromosome datasets using SplitsTree v.4.12.1 (Huson and Bryant, 2006) using uncorrected p-distances among all samples and 100 bootstrap replicates.

Nuclear genome species trees were reconstructed under the coalescent module using SVDquartets (SVDQ; Chifman and Kubatko, 2014) for each partitioned dataset (autosomal and Z chromosome) using all filtered, intergenic, or genic SNPs, and grouped according to species or subspecies taxonomy (see Tables S1-2) with Greater Sage-grouse identified as the outgroup. This species tree method has been shown to be robust to gene flow between closely related taxa (Long and Kubatko, 2018), which likely occurs among extant prairie grouse taxa. All contemporary samples were grouped according to their taxonomic species or subspecies groups, while the three mainland (ML) and the five Martha's Vineyard (MV) Heath Hen samples were analyzed as two separate groups.

The SVDQ analyses were run as implemented in PAUP* v4.0a (build 168) (Swofford, 2003) in "species tree" mode with exhaustive quartet sampling using the QFM algorithm. Each species tree analysis included a different number of variable sites due to differences in total SNPs available for each partitioned dataset and 100 bootstrap replicates to assess branch support among the sampled taxonomic groups.

2.5. Introgression analysis

To investigate for past introgression, or net effect of gene flow, between prairie grouse taxa and how its pattern may differ between autosomes and the Z chromosome based on an excess of shared variation on a genome-wide scale, we applied the ABBA/BABA test using Patterson's D and f_4 admixture ratio statistics (Patterson et al., 2012) among all possible combinations of trios among our ingroup taxa (P1, P2, and P3) using the Greater Sage-grouse as the outgroup (O) and a rooted tree ((P1,P2),P3),O). In this test the number of ancestral ("A") and derived ("B") SNPs were calculated for a four-taxon comparison that includes the

outgroup. In cases with incomplete lineage sorting or recurrent mutation with no gene flow, the number of SNPs showing the pattern "ABBA" (or where P2 and P3 share derived SNPs) and "BABA" (where P1 and P3 shared derived SNPs) should be equally frequent (or D statistic = 0), whereas an excess of either pattern would indicate possible gene flow or introgression (or a significant D statistic) with the proportion of their genomes shared as quantified using the f_4 admixture ratio (see Patterson et al., 2012; Malinsky et al., 2021 for further information concerning to two methods).

The analyses were performed for both the autosomal (31,820,592 SNPs) and Z chromosomal (2,158,190 SNPs) datasets that included genic and intergenic sites using the *Trios* program in *Dsuite* v. 0.5 r47 (Malinsky et al., 2021). *Trios* for calculating D and f_4 -ratio values were identified based on the Z-chromosome species tree topology to represent hypothesized relationships between prairie grouse taxa. This allowed for additional analyses using f-branch (f_b) tests to identify patterns of excessive allele sharing consistent with gene flow between non-sister taxa including internal branches of the species tree (Malinsky et al., 2018). Prior to calculating f_b , the f_4 -ratio value threshold for each trio was set to zero when their D statistic p -value was >0.001. Significance of the D statistic was determined by using a block jackknife procedure to obtain an approximate normally distributed standard error that was then used to calculate a Z -score and an associated p -value for each trio. A Z -score $|Z| > 3$ and a p -value < 0.001 was considered significant.

3. Results

3.1. Nuclear genomic processing

After excluding eight samples due to lower sequencing coverage, average depth of coverage ranged from 2.7x to 13.9x among the Heath Hen samples (mean = 6.4x, $n = 8$) and from 10.6x to 27.7x among the contemporary prairie grouse samples (mean = 15.0x, $n = 25$; Table S2). Because female grouse are the heterogametic sex (ZW), their average depth of coverage for Z-chromosome SNPs were reduced by half compared to their autosomal SNPs. The depth of coverage for Z-chromosome SNPs among Heath Hen samples ranged from 2.2x to 7.6x for females (mean = 4.3, $n = 3$) and 2.6x to 11.0x for males (mean = 5.2, $n = 5$), and for all other prairie grouse samples ranged from 6.5x to 10.9x for females (mean = 9.5, $n = 16$) and 12.4x to 15.6x for males (mean = 16.9, $n = 12$). We retained three Heath Hen samples with lower coverage (<4x) in our final datasets to increase sample size for that taxon. Additional analyses were conducted excluding the three Heath Hen samples with <4x coverage without any appreciable change in our results (data not shown).

After applying stringent filtering standards, a final set of 34,028,294 high-quality autosomal (31,820,592) and Z-chromosomal (2,158,190) SNPs were identified within and among all sampled taxa including the three Greater Sage-grouse samples used as our outgroup. Additional partitioning and filtering including LD-pruning were employed for downstream phylogenetic analyses resulting in six additional datasets that included autosomal and Z-chromosome datasets each with intergenic (1,628,622 and 620,068), genic (838,212 and 160,030, respectively), or all (2,439,756 and 780,098) SNPs, respectively.

3.2. Mitochondrial genomic processing

After excluding two Lesser Prairie-chicken samples due to low depth of coverage, average depth of coverage per site ranged from 123.8x to 1,775.3x among the Heath Hen samples (mean = 762.4x, $n = 9$) and from 5.4x to 779.5x among the contemporary prairie grouse samples (mean = 192.5x, $n = 30$; Table S3). A total of 16,697 bp of mtDNA sequence was produced following multi-sequence alignment using the generated consensus sequences for each sample.

3.3. Phylogenomic networks and species tree inference

Overall, limited concordance was observed between sequence similarity and prairie grouse taxonomy in the mitochondrial NeighborNet network with two groups consisting of multiple samples from different taxa connected by a relatively large reticulation with some taxa distributed in both groups (Fig. S3). Despite having small sample sizes for each prairie grouse taxa, only 4 of 39 samples from two Sharp-tailed Grouse subspecies clustered completely by taxonomy in the mitochondrial network. In contrast, most samples consistently clustered by taxonomy in both autosomal and Z-chromosome NeighborNet networks with the exception of Heath Hens, which formed two separate clusters based on geography (mainland and Martha's Vineyard samples; Fig. S2). This same pattern was also observed with the mitochondrial network, but each Heath Hen group had one sample that also clustered with other prairie grouse taxa (Fig. S3).

Differences in topology in both the NeighborNet networks and species tree analyses were also detected depending on autosomal or Z-

chromosome datasets. The primary differences were with the placement of Lesser Prairie-chicken and the two Heath Hen groups (mainland and Martha's Vineyard) (Figs. 2 and S2). Specifically, in the autosomal species tree analysis, regardless of partitioning strategy (i.e., all, genetic, or intergenic SNPs), the Heath Hen was paraphyletic with those from Martha's Vineyard being sister to a clade that included the Lesser Prairie-chicken, Attwater's and Greater Prairie-chickens, and the mainland sampled Heath Hens. In contrast, based on the Z-chromosome species trees, Lesser Prairie-chicken was sister to the clade that included Martha's Vineyard and mainland Heath Hens and both Attwater's and Greater Prairie-chicken. Sharp-tailed grouse subspecies also consistently formed a clade sister to all remaining prairie grouse taxa for both the autosomal and Z-chromosome partitioned dataset species tree analyses (Fig. 2). Variation did exist, however, among Sharp-tailed grouse subspecies branching pattern in both sets of species trees but was largely consistent with geographic distance. Nodal support values were low (<80) in many of those relationships particularly on the Z-chromosome species tree. These same patterns were largely observed with the

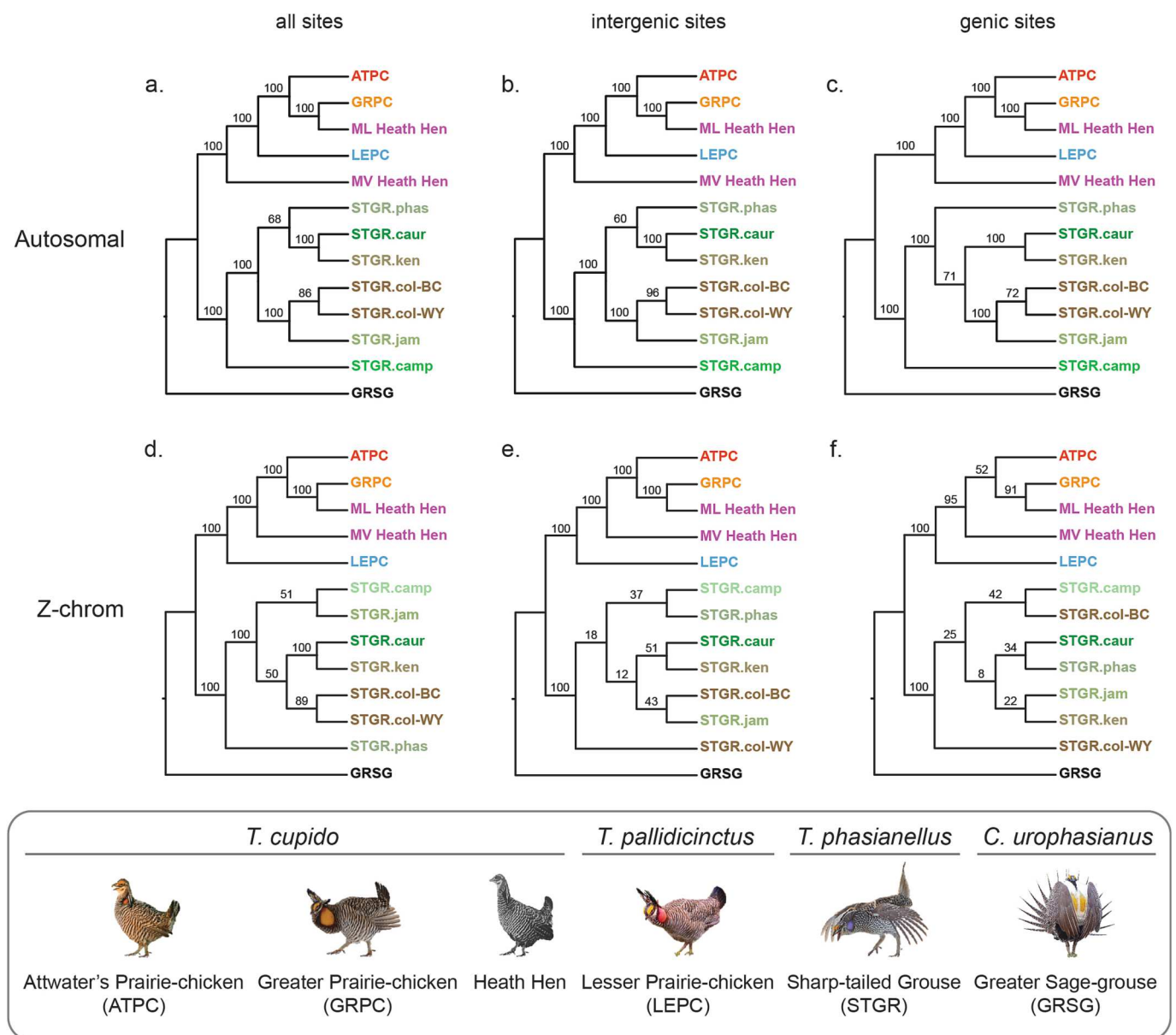


Fig. 2. SVDQ consensus species trees analyzed separately by (a-c) autosomal and (d-f) Z-chromosomal partitions (all, intergenic, and genic SNPs). Taxon abbreviations are similar to those provided in Table S1, and “ML” and “MV Heath Hen” labels correspond with mainland and Martha's Vineyard Heath Hen populations, respectively. Bootstrap support values are provided for each node.

autosomal and Z-chromosome NeighborNet networks, regardless of partition used in the analysis (Fig. S2).

3.4. Evidence for introgression

Results from the species tree and network analyses were corroborated by our D statistic and f_4 admixture ratio analyses with 51 of 220 trios with the autosomal dataset having a significant D statistic ($p < 0.001$) and excess of allele sharing between non-sister taxa (Fig. 3a; Table S4), whereas only four trios with the Z chromosome dataset were significant (Fig. 3c; Table S5). In fact, the only non-sister taxa to show significant admixture ratios (f_4) and an excess of shared alleles (f_b) with the Z chromosome dataset were among conspecific Greater Prairie-chicken and Martha's Vineyard and mainland Heath Hens and between two geographically proximate Sharp-tailed Grouse subspecies (*T. p. jamesi* and *T. p. campestris*; Fig. 3c, S4c). We also observed a consistent pattern indicating a higher admixture ratio and excess of shared alleles with the Z chromosome dataset between mainland Heath Hen and Greater Prairie-chicken ($f_4 = 0.159$) than between mainland and Martha's Vineyard Heath Hen groups ($f_4 = 0.1038$) and between Martha's Vineyard Heath Hen and Greater Prairie-chicken ($f_4 = 0.059$; Fig. 3c, S4c).

Not only did the autosomal dataset identify most of the same pattern as the Z chromosome dataset including significant admixture ratios between Greater Prairie-chicken and mainland Heath Hen ($f_4 = 0.041$), but also significant admixture ratios between Lesser Prairie-chicken and Greater ($f_4 = 0.074$) and Attwater's ($f_4 = 0.086$) Prairie-chickens and between multiple Sharp-tailed Grouse subspecies ($f_4 = 0.013$ – 0.035 ; Fig. 3a). Further analyses based on f_b statistics also indicated gene flow

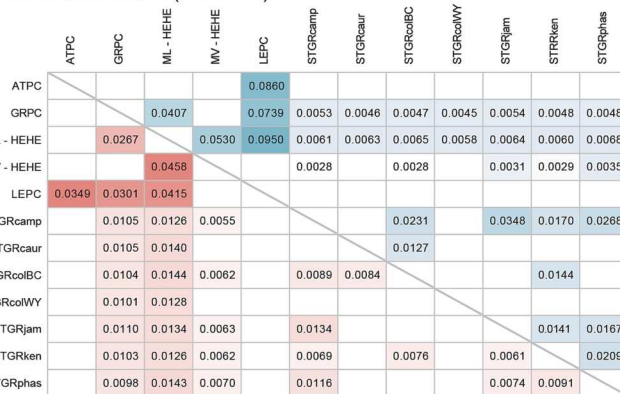
between Lesser Prairie-chicken and the ancestral lineage that included both Greater and Attwater's Prairie-chicken supporting long-term historic gene flow based on the autosomal but not the Z-chromosome datasets and between Sharp-tailed grouse subspecies and Greater Prairie-chicken, also based on the autosomal dataset, with no apparent pattern relative to geography (Fig. S4a).

Additional analyses were conducted using transversion-only variant datasets to address potential bias in our estimates of introgression influenced by postmortem cytosine deamination and an excess of transitions often observed among ancient DNA samples (Sawyer et al., 2012; Jónsson et al., 2013). Those results indicated similar patterns as described above for the Heath Hen samples but with the exception of those between Greater Prairie-chicken and the Martha's Vineyard Heath Hen. With the transversions-only dataset, no excessive allele sharing (f_b) was observed with the autosomal dataset (Fig. S4a,b) or a significant f_4 admixture ratio based on the Z-chromosome dataset between Greater Prairie-chicken and Martha's Vineyard Heath Hens (Fig. 3d).

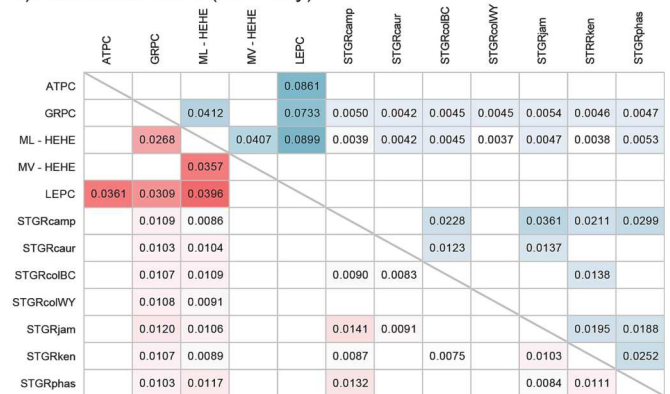
4. Discussion

Whole-genome resequencing data partitioned as either autosomal or Z-chromosome datasets provided the necessary resolution to investigate phylogenetic relationships among prairie grouse taxa using both species tree and NeighborNet network analyses, whereas previous studies based on sequence data from only a few nuclear loci or mitochondrial DNA (including this study) were largely inconclusive resulting in low nodal support values or paraphyletic relationships among taxa (Ellsworth et al., 1994, 1996; Drovetski, 2002; Oyler-McCance et al., 2010; Persons et al., 2016; but see Galla and Johnson, 2015). While the autosomal or Z-

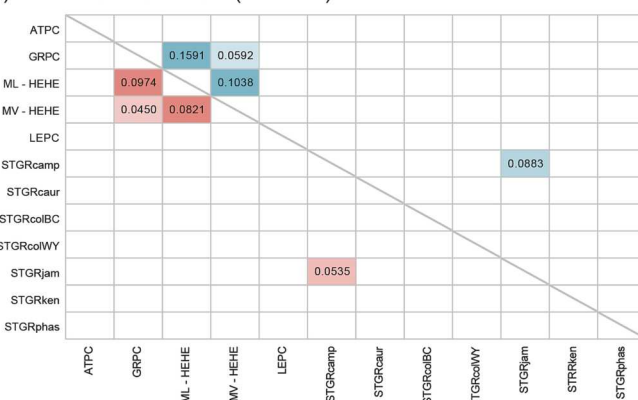
a) Autosomal SNPs (TS & TV)



b) Autosomal SNPs (TV - only)



c) Z-chromosome SNPs (TS & TV)



d) Z-chromosome SNPs (TV - only)

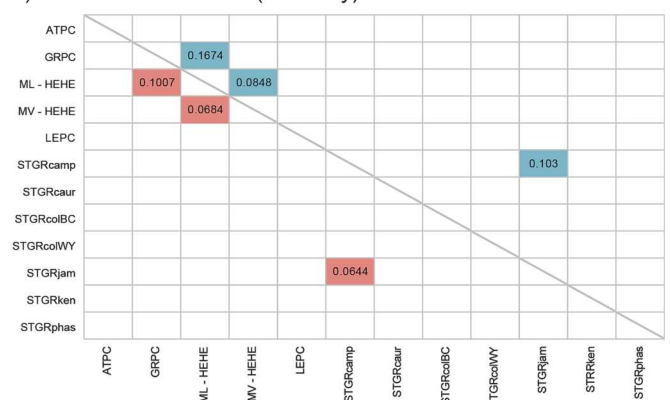


Fig. 3. Heatmaps of maximum pairwise D (below diagonal) and f_4 admixture ratio (above diagonal) statistics between taxon pairs for all a) all autosomal (transition (TS) and transversion (TV) and b) TV-only autosomal SNPs and c) all Z-chromosome and d) TV-only Z-chromosome SNPs with significant ($p < 0.001$) combinations of trios. Taxon abbreviations are similar to those provided in Fig. 2, and “ML-HEHE” and “MV-HEHE” labels correspond with mainland and Martha's Vineyard Heath Hen populations, respectively.

chromosome effective population sizes may have influenced the rate of divergence differently for the two dataset types and thereby potentially influence their ability to resolve species- or subspecies-level relationships (e.g., Corl and Ellegren, 2013; Galla and Johnson, 2015), both datasets regardless of further partitioning into intergenic or genic datasets, resulted in individual prairie grouse samples largely clustering by taxonomy as shown for the first time based on NeighborNet networks (Fig. S2). What did differ among datasets, however, was the topology or branching order among select prairie grouse taxa within the generated species trees and networks thereby providing further insight in how sex-biased gene flow and strong sexual selection has influenced patterns of divergence within this species complex.

4.1. Introgression among extant prairie grouse

Prairie grouse are a lek species, where strong sexual selection on male morphology and behavior influence reproductive success (Johnsgard, 2002). At the interspecific level, males of each species are easily distinguished both morphologically based on plumage characteristics and by their behavior and vocalizations (e.g., Bain and Farley, 2002), yet hybridization does occur among prairie grouse species in areas of sympatry (Bent, 1932; Johnsgard and Wood, 1968; Sparling, 1980; Bain and Farley, 2002; Augustine and Trauba, 2015; Oyler-McCance et al., 2016; Huschle and Toepfer, 2020). As previously described in Galla and Johnson (2015), we suspect that male F1 interspecific hybrids are less likely to reproduce due to their morphology and other behaviors related to reproduction being intermediate in form between the two parental species, thereby likely influencing female choice (i.e., unsexy-son hypothesis; see also Augustine and Trauba, 2015). In this way, for example, genes linked to traits associated with male reproductive success, possibly on the Z-chromosome (e.g., Sæther et al., 2007; Backström et al., 2010; see also Fraïsse and Sachdeva, 2021; Ottenburghs, 2022), may serve as reinforcement with selection to avoid maladaptive hybridization between prairie grouse species. Behavioral isolation may limit further introgression between prairie grouse species due to low male hybrid attractiveness and female preference for conspecifics. Female prairie grouse among all three species, however, are similar in morphology (Johnsgard, 2002), and therefore, reproductive success of hybrid females is less likely to be constrained than male hybrids, at least from a pre-zygotic perspective, and thereby allow for introgression as shown with autosomal and mitochondrial genomes (Galla and Johnson, 2015; this study). Previous research has suggested that initial hybridization events among prairie grouse are more likely to occur when populations are small or with naïve females (Augustine and Trauba, 2015), but more research is needed to further investigate what factors influence the frequency of hybridization among prairie grouse and determine if fitness is reduced among female hybrid individuals or their subsequent generations.

By including the Heath Hen in this study, we were better positioned to explore how patterns of historic gene flow have influenced results from previous phylogenetic analyses for extant prairie grouse taxa. The placement of Lesser Prairie-chicken relative to Heath Hen in our species tree and network analyses differed between autosomal and Z-chromosome datasets suggesting contrasting demographic histories. The Heath Hen formed a monophyletic clade with its conspecifics, the Greater and Attwater's Prairie-chickens based on the Z-chromosome species tree, yet formed a paraphyletic clade with autosomes with to the inclusion of Lesser Prairie-chicken regardless of partitioned dataset (genic, intergenic, or all SNPs; Fig. 2). Introgression analyses provided supporting evidence of historic gene flow between Lesser and Greater Prairie-chicken, but only based on those analyses using autosomal SNPs and not the Z-chromosome (Fig. 3). These results are consistent with those reported previously (Galla and Johnson, 2015; Oyler-McCance et al., 2016) that historic female-biased gene flow has occurred between Lesser and Greater Prairie-chickens in areas of sympatry, yet introgression on the Z-chromosome is lacking due to strong sexual selection and male-

biased behavioral isolation (Figs. 3, S4). The phylogenetic signal as shown with the Z-chromosome SNPs is therefore more likely to represent historic patterns of lineage divergence than autosomal SNPs due to historic female-biased gene flow influencing introgression patterns between prairie grouse in areas of sympatry.

Interestingly, despite hybridization being reported between Greater Prairie-chicken and Sharp-tailed Grouse in areas of sympatry, no such pattern was suggested based on the species tree or network analyses as was shown with Lesser and Greater Prairie-chicken. All of the Sharp-tailed Grouse subspecies formed a clade that was sister to a clade that included all of the remaining prairie grouse taxa including the Heath Hen for both the autosomal or Z-chromosome datasets (Fig. 2; see also Fig. S2). However, similar to comparisons with Lesser Prairie-chicken, the introgression analyses uncovered a pattern of excess allele sharing, or gene flow, between Sharp-tailed Grouse and Greater Prairie-chicken, but only with the autosomal dataset (Fig. 3). These results further corroborate those from Galla and Johnson (2015) that identified a higher proportion of population pairwise comparisons between Sharp-tailed Grouse and Greater Prairie-chicken that supported a strict isolation model of divergence based on introns on the Z-chromosome compared to those based on autosomal introns using an isolation-with-migration model analysis. Based on the corresponding f_b analyses using autosomal SNPs, the results suggest that the correlated excess allele sharing pattern observed between Sharp-tailed Grouse subspecies and Greater Prairie-chicken is likely due to shared ancestry between all Sharp-tailed Grouse subspecies lineages (Figs. 3, S4). Additional research is warranted with larger sample sizes investigating autosomal introgression between the two Sharp-tailed Grouse subspecies (*T. p. jamesi* and *T. p. campestris*) that overlap in geographic distribution with Greater Prairie-chicken.

A similar pattern was observed when comparing admixture ratios among non-sister Sharp-tailed Grouse subspecies despite low nodal support values in the species tree analyses. With the exception of one pairwise comparison between *T. p. jamesi* and *T. p. campestris*, Z-chromosome f_b analysis indicated no gene flow among the majority of non-sister Sharp-tailed Grouse subspecies, yet an excess of allele sharing was observed among at least nine subspecies comparisons based on the autosomal dataset (Fig. S4, see also Fig. 3). This was surprising since we assumed that more similar morphologies and behaviors shared among Sharp-tailed Grouse subspecies would more likely result in fewer apparent barriers to gene exchange with respect to the Z-chromosome than between two congeneric species such as Greater and Lesser Prairie-chicken. Of those subspecies that did possess significant admixture ratios based on the autosomal dataset, they were also largely geographically restricted to either side of the Continental Divide. Spaulding et al. (2006) reported a similar pattern based on AFLP and nuclear microsatellite loci while using a smaller subset of Sharp-tailed Grouse subspecies. Additional study is warranted using larger sample sizes in geographically proximate locations to further explore patterns of gene flow or lack thereof among Sharp-tailed grouse subspecies.

4.2. Heath hen phylogenomics

We also observed an incongruent pattern relative to taxonomy among the Heath Hens sampled either from the mainland or on Martha's Vineyard. In all species tree analyses and NeighborNet networks based on the nuclear dataset, regardless of partition, the mainland Heath Hens were sister to the Greater Prairie-chicken and separate from those collected on Martha's Vineyard. To our knowledge, most Heath Hen specimens within museum collections originated from the Martha's Vineyard population after the mainland population had become extirpated by the 1870 s. We know of only seven specimens presumed to be Heath Hens from the mainland population, all of which were labeled as obtained from public meat markets with the majority possessing no collection date. The few exceptions included two of the three specimens used in this study collected in 1872 and 1873 (see Table S1).

Compared to its conspecific, the Greater Prairie-chicken, [Brewster \(1885\)](#) described the Heath Hen as smaller in size, redder in color above, and with fewer and narrower and pointed neck pinnae feathers on males (4 to 5 vs 7 to 10). The three mainland specimens that were used in this study were all adult males and confirmed as more similar in morphology to Greater Prairie-chicken than Heath Hen when inspected more closely following our initial analyses. For that reason, we cannot discount that the mainland Heath Hen samples were translocated prairie grouse from the Midwest and erroneously identified as Heath Hens at time of capture or possibly post-translocation hybrids between the two taxa since their harvest dates were after reported translocation events. Starting as early as 1852, prairie grouse were commonly transported from the Midwest for release in areas previously occupied by Heath Hens in an attempt to establish prairie grouse in eastern states where previously harvested ([Gross, 1928; Phillips, 1928](#)).

In fact, our introgression analyses suggested that gene flow existed between Greater Prairie-chicken and both sampled Heath Hen geographic groups (Martha's Vineyard and mainland) based on the Z-chromosome analyses but not between Greater Prairie-chicken and Martha's Vineyard Heath Hens using the autosomal dataset ([Fig. 3](#)). When the same analyses were conducted using the transversion-only SNPs to address potential bias due to deamination with our aDNA samples, the same results were observed with the exception of no gene flow between Heath Hens on Martha's Vineyard and Greater Prairie-chicken using the Z-chromosome dataset similar to the autosomal datasets. Based on these results combined with relatively low admixture ratios (f_4 ; see [Fig. 3](#)), historic gene flow existed between mainland and Martha's Vineyard Heath Hens, yet with Greater Prairie-chicken, gene flow existed only with mainland Heath Hens and not with those on Martha's Vineyard. It cannot be discounted, however, based on these analyses that hybridization did occur on the island just prior to the subspecies' extinction in 1932 and its signal was not recovered based on the samples used in this study.

Therefore, we do not consider the mainland Heath Hen samples in our interpretations of Heath Hen phylogeny and argue that the samples originating from Martha's Vineyard are better representative of the subspecies from a historic genomic perspective. The separation of Martha's Vineyard from the mainland started approximately 5–6 k yr BP based on bathymetry and sea level rise ([Foster, 2017](#)), and even today the closest point to the mainland from Martha's Vineyard is only 5.5 km away. This is within distance for prairie grouse to disperse in either direction ([Earl et al., 2016; Johnson et al., 2020](#)) including across open water ([Bent, 1932](#)). We suspect, therefore, that the two areas were unlikely to be isolated from each other prior to Heath Hen extirpation on the mainland (see also [Gross, 1928; Bent, 1932](#)) and that connectivity with reoccurring dispersal events were likely important for maintaining a small viable population on Martha's Vineyard based on our understanding of prairie grouse demographics in the Midwest ([Westemeier et al., 1998; Johnson et al., 2004; Johnson and Dunn, 2006; Capel et al., 2022](#); see also [Lacy, 2000; Kardos et al., 2021](#)).

We did observe one Martha's Vineyard sample clustering with other prairie grouse haplotypes along with a mainland Heath sample in the mtDNA NeighborNet network ([Fig. S3](#); see also [Palkovacs et al., 2004; Johnson and Dunn, 2006](#)), yet no similar pattern was observed with the autosomal and Z-chromosome networks with all of the Martha's Vineyard Heath Hen samples forming a single cluster ([Fig S2](#)). These results suggest that the shared mtDNA haplotype may be due to ancestral polymorphism, but we cannot rule out the possibility of introgression as also suggested by our admixture analyses ([Fig. 3, S4a](#); but see discussion above). A few reports do exist stating that "western prairie chickens" obtained by local landowners at sportsman exhibitions or markets on the mainland were released on Martha's Vineyard that had also subsequently produced offspring on the island as early as 1898 ([Gross 1928](#)). It was not known, however, if any of the offspring that had been observed following those introductions were of hybrid origin, but [Gross \(1928\)](#) did state some concern that hybridization was a possibility. To

what extent those efforts to reestablish or supplement prairie grouse on the east coast may have been impacted by differences that may have existed with regard to local adaptation at the genomic level or possibly hybrid incompatibility between the Heath Hen and other prairie grouse deserves further study.

5. Conclusions

By using a whole-genomic resequencing approach and the inclusion of samples from the extinct Heath Hen, our results further support that sex-biased introgression between Lesser and Greater Prairie chickens in areas of sympatry have likely contributed to the observed phylogenetic variation and taxonomic discordance as shown with previous studies (e. g., [Ellsworth et al., 1994, 1996; Drovetski, 2002; Oyler-McCance et al., 2010; Persons et al., 2016](#)). Because the extinct Heath Hen was allopatric from other prairie grouse, gene flow should not contribute to such patterns and therefore help clarify our understanding of prairie grouse evolutionary history. Our study has suggested, however, that early efforts to establish prairie grouse in the eastern coastal states after Heath Hen decline and extirpation has likely complicated our interpretation, but additional analyses using partitioned genomic datasets have provided further insight concerning prairie grouse demographic history.

Here we have shown that the Heath Hen is more closely related to the Greater Prairie-chicken, its conspecific, then to other prairie grouse species despite previous studies based on mtDNA suggesting otherwise ([Palkovacs et al., 2004; Johnson and Dunn, 2006](#)). Our ability to confirm that relationship was made possible only by investigating their phylogenetic patterns using partitioned genomic sequence data on the Z-chromosome while recognizing or allowing for introgression to occur with the autosomal lineages. While prairie grouse are a unique group for studying contrasting patterns of genomic introgression relative to sex-biased life history constraints and strong sexual selection, other studies exploring phylogenetic patterns in recently divergent species should prioritize efforts in partitioning their nuclear genomic datasets accordingly to explore similar patterns with their focal taxa (see also [Bourgeois et al., 2020; Fraisse and Sachdeva, 2021; Ottenburghs, 2022](#)). As this study has shown, an arbitrary set of SNPs or sequence data without prior knowledge of their chromosomal origin can mislead or complicate our ability to uncover patterns of evolutionary history particularly among recently divergent taxa.

Author contributions

JAJ, BN, GA, TC and JT conceived the study; JAJ, BN and JT contributed samples; JAJ, GA, and AGS analyzed the data. JAJ wrote the article with input from all authors. All authors, with the exception of JT, read and approved the final manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

We thank Peter Dunn, Mike Morrow, Allen Spaulding, Sara Oyler-McCance, Lena Larsson, and Don Wolfe for their help in securing samples used in this study and André Soares for providing assistance for initial bioinformatic processing of museum sample raw sequence reads. We also thank Academy of Natural Sciences of Philadelphia (PA, USA), Royal Ontario Museum (Ontario, Canada), and Field Museum of Natural

History (Chicago, IL USA) for allowing us to use tissue samples from their Heath Hen study skins. This research was supported by funds generously provided by Revive & Restore and University of North Texas with early support provided by Warren Adams, G. Kenneth Baum and Ann Baum Philanthropic Fund, The Betsy & Jesse Fink Foundation, Peter & Gwen Norton, and Brad Palmer. AGS was supported by the NSF Postdoctoral Research Fellowships in Biology Program under Grant No. 2109912. The authors acknowledge the Talon3 system at The University of North Texas (www.research.unt.edu/research-services/research-computing) and the Texas Advanced Computing Center (TACC) at The University of Texas at Austin (www.tacc.utexas.edu) for providing High Performance Computing and storage resources that have contributed to the research results as reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ympev.2023.107927>.

References

- Augustine, J.K., Trauba, D.R., 2015. Potential for behavioral reproductive isolation between greater prairie-chickens and sharp-tailed grouse in west-central Minnesota. *J. Ethol.* 33, 15–24.
- Axelsson, E., Willerslev, E., Gilbert, M.T., Nielsen, R., 2008. The effect of ancient DNA damage on inferences of demographic history. *Mol. Biol. Evol.* 25, 2181–2187.
- Backström, N., Lindell, J., Zhang, Y., Palkopoulou, E., Qvarnström, A., Sætre, G.-P., Ellegren, H., 2010. A high-density scan of the Z chromosome in *Ficedula* flycatchers reveals candidate loci for diversifying selection. *Evolution* 64, 3461–3475.
- Bain, M.R., Farley, G.H., 2002. Display and apparent hybrid prairie-chickens in a zone of geographic overlap. *Condor* 104, 683–687.
- Bent, A.C., 1932. Life histories of North American gallinaceous birds (orders Galliformes and Columbiformes). *Bull. U.S. Natl. Mus.* 162, 1–490. <https://doi.org/10.5479/si.03629236.162.i>.
- Bourgeois, Y.X.C., Bertrand, J.A.M., Delahaie, B., Holota, H., Thébaud, C., Milá, B., 2020. Differential divergence in autosomes and sex chromosomes is associated with intra-island diversification at a very small spatial scale in a songbird lineage. *Mol. Ecol.* 29, 1137–1153.
- Brewster, W., 1885. The heath hen of Massachusetts. *Auk* 2, 80–84.
- Cantarel, B.L., Korf, I., Robb, S.M.C., Parra, G., Ross, E., Moore, B., Holt, C., Alvarado, A. S., Yandell, M., 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* 18, 188–196.
- Capel, S.L.R., Bouzat, J.L., Catchen, J.M., Johnson, J.A., Dunn, P.O., Paige, K.N., 2022. Evaluating the genome-wide impacts of species translocations: the greater prairie-chicken as a case study. *Conserv. Genet.* 23, 179–191.
- Charlesworth, B., 2009. Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* 10, 195–205.
- Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890.
- Chesser, R.K., Baker, R.J., 1996. Effective sizes and dynamics of uniparentally and diparentally inherited genes. *Genetics* 144, 1225–1235.
- Chifman, J., Kubatko, L., 2014. Quartet inference from SNP data under the coalescent model. *Bioinformatics* 30, 3317–3324.
- Connelly, J.W., Gratson, M.W., Reese, K.P., 2020. Sharp-tailed Grouse (*Tympanuchus phasianellus*), version 1.0. In: Poole, A.F., Gill, F.B. (Eds.), *Birds of the World*. Cornell Lab of Ornithology, Ithaca, NY, USA. 10.2173/bow.shtgro.01.
- Cooper, A., Poinar, H.N., 2000. Ancient DNA: Do it right or not at all. *Science* 289, 1139.
- Cori, A., Ellegren, H., 2013. Sampling strategies for species trees: the effects of phylogenetic inference of number of genes, number of individuals, and whether loci are mitochondrial, sex-linked, or autosomal. *Mol. Phylogenet. Evol.* 67, 358–366.
- Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitman, A., Keane, T., McCarthy, S.A., Davies, R.M., Li, H., 2021. Twelve years of SAMtools and BCFtools. *GigaScience* 10, giab008.
- Dean, R., Harrison, P.W., Wright, A.E., Zimmer, F., Mank, J.E., 2015. Positive selection underlies Faster-Z evolution of gene expression in birds. *Mol. Biol. Evol.* 32, 2646–2656.
- Drovetski, S.V., 2002. Molecular phylogeny of grouse: Individual and combined performance of W-linked, autosomal, and mitochondrial loci. *Syst. Biol.* 51, 930–945.
- Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P., Aiden, E.L., 2017. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* 356, 92–95.
- Earl, J.E., Fuhlendorf, S.D., Haukos, D., Tanner, A.M., Elmore, D., Carleton, S.A., 2016. Characteristics of lesser prairie-chicken (*Tympanuchus pallidicinctus*) long-distance movements across their distribution. *Ecosphere* 7, e01441.
- Edgar, R.C., 2021. MUSCLE v5 enables improved estimates of phylogenetic tree confidence by ensemble bootstrapping. *BioRxiv*. <https://doi.org/10.1101/2021.06.20.449169>.
- Ellegren, H., 2009. The different levels of genetic diversity in the sex chromosomes and autosomes. *Trends Genet.* 25, 278–284.
- Ellegren, H., 2011. Sex chromosome evolution: recent progress and the influence of male and female heterogamety. *Nat. Rev. Genet.* 12, 157–166.
- Ellsworth, D.L., Honeycutt, R.L., Silvy, N.J., Rittenhouse, K.D., Smith, M.H., 1994. Mitochondrial-DNA and nuclear-gene differentiation in North American prairie grouse (genus *Tympanuchus*). *Auk* 111, 661–671.
- Ellsworth, D.L., Honeycutt, R.L., Silvy, N.J., 1996. Systematics of grouse and ptarmigan determined by nucleotide sequences of the mitochondrial cytochrome-b gene. *Auk* 113, 811–822.
- Foster, D., 2017. *A Meeting of Land and Sea: Nature and the Future of Martha's Vineyard*. Yale University Press.
- Fraisse, C., Sachdeva, H., 2021. The rates of introgression and barriers to genetic exchange between hybridizing species: sex chromosomes vs autosomes. *Genetics* 217, iyaa025.
- Galla, S., Johnson, J.A., 2015. Differential introgression and effective size of marker type influence phylogenetic inference of a recently divergent avian group (Phasianidae: *Tympanuchus*). *Mol. Phylogenet. Evol.* 84, 1–13.
- Gould, G.M., Augustine, J.K., 2020. Multiple signals predict male mating success in the lek-mating lesser prairie-chicken (*Tympanuchus pallidicinctus*). *Behavioral Ecology and Sociobiology* 74, 137.
- Grabherr, M.G., Russell, P., Meyer, M., Mauceli, E., Alföldi, J., Di Palma, F., Lindblad-Toh, K., 2010. Genome-wide synteny through highly sensitive sequence alignment: Satsuma. *Bioinformatics* 26, 1145–1151.
- Green, R.E., Malaspina, A.-S., Krause, J., Briggs, A.W., Johnson, P.L.F., Uhler, C., Meyer, M., Good, J.M., Maricic, T., Stenzel, U., Prüfer, K., Siebauer, M., Burbano, H. A., Ronan, M., Rothberg, J.M., Egholm, M., Rudan, P., Brajković, D., Kučan, Z., Gusić, I., Wikström, M., Laakkonen, L., Kelso, J., Slatkin, M., Pääbo, S., 2008. A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* 134, 416–426.
- Gross, A.O., 1928. The heath hen. *Memoirs Boston Soc. Nat. Hist.* 6, 489–588.
- Hagen, C.A., Giesen, K.M., 2020. Lesser Prairie-Chicken (*Tympanuchus pallidicinctus*), version 1.0. In: Poole, A.F., Gill, F.B. (Eds.), *Birds of the World*. Cornell Lab of Ornithology, Ithaca, NY, USA. 10.2173/bow.lepchi.01.
- Hayes, K., Barton, H.J., Zeng, K., 2020. A study of faster-Z evolution in the Great Tit (*Parus major*). *Genome Biol. Evol.* 12, 210–222.
- Huschle, G., Toeffer, J.E., 2020. Trends in a Greater Prairie Chicken population established by translocation in North Dakota. *Prairie Nat.* 52, 76–79.
- Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23, 254–267.
- Johnsgard, P.A., 2002. *Grassland Grouse and Their Conservation*. Smithsonian Institution Press, Washington DC.
- Johnsgard, P.A., Wood, R.E., 1968. Distributional changes and interaction between prairie chickens and sharp-tailed grouse in the Midwest. *Wilson Bull.* 80, 173–188.
- Johnson, J.A., 2008. Recent range expansion and divergence among North American prairie grouse. *J. Hered.* 99, 165–173.
- Johnson, J.A., Bellinger, M.R., Toeffer, J.E., Dunn, P.O., 2004. Temporal changes in allele frequencies and low effective population size in greater prairie-chickens. *Mol. Ecol.* 13, 2617–2630.
- Johnson, J.A., Dunn, P.O., 2006. Low genetic variation in the Heath Hen prior to extinction and implications for the conservation of prairie-chicken populations. *Conserv. Genet.* 7, 37–48.
- Johnson, J.A., Schroeder, M.A., Robb, L.A., 2020. Greater Prairie-Chicken (*Tympanuchus cupido*), version 1.0. In: Poole, A.F., Gill, F.B. (Eds.), *Birds of the World*. Cornell Lab of Ornithology, Ithaca, NY, USA. <https://doi.org/10.2173/bow.grpchi.01>.
- Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L.F., Orlando, L., 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29, 1682–1684.
- Kardos, M., Armstrong, E.E., Fitzpatrick, S.W., Funk, W.C., 2021. The crucial role of genome-wide genetic variation in conservation. *Proc. Natl. Acad. Sci. U.S.A.* 118, e2104632118. 10.1073/pnas.2104642118.
- Lacy, R.C., 2000. Considering threats to the viability of small populations using individual-based models. *Ecol. Bull.* 48, 39–51.
- Laporte, V., Charlesworth, B., 2002. Effective population size and population subdivision in demographically structured populations. *Genetics* 162, 501–519.
- Lavretsky, P., Dacosta, J.M., Hernández-Baños, B.E., Engilis Jr, A., Sorenson, M.D., Peters, J.L., 2015. Speciation genomics and a role for the Z chromosome in the early stages of divergence between Mexican ducks and mallards. *Mol. Ecol.* 24, 5364–5378.
- Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760.
- Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bio.GN]* <https://arxiv.org/abs/1303.3997>.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* 25, 2078–2079.
- Long, C.L., Kubatko, L.S., 2018. The effect of gene flow on coalescent-based species-tree inference. *Syst. Biol.* 67, 770–785.
- Malinsky, M., Svardal, H., Tyers, A.M., Miska, E.A., Genner, M.J., Turner, G.F., Durbin, R., 2018. Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature Ecol. Evol.* 2, 1940–1955.
- Malinsky, M., Matschner, M., Svardal, H., 2021. Dsuite – fast D-statistics and related admixture evidence from VCF files. *Mol. Ecol. Resour.* 21, 584–595.
- Mank, J.E., Nam, K., Ellegren, H., 2010a. Faster-Z evolution is predominately due to genetic drift. *Mol. Biol. Evol.* 27, 661–670.

- Mank, J.E., Vicoso, B., Berlin, S., Charlesworth, B., 2010b. Effective population size and the Faster-X effect: empirical results and their interpretation. *Evol.* 64, 663–674.
- Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17, 10–12. <https://doi.org/10.14806/ej.17.1.200>.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303.
- Meyer, M., Kircher, M., 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* t5448. 10.1101/pdb.prot5448.
- Nosil, P., Funk, D.J., Ortiz-Barrientos, D., 2009. Divergent selection and heterogeneous genomic divergence. *Mol. Ecol.* 18, 375–402.
- Ottenburghs, J., 2022. Avian introgression patterns are consistent with Haldane's Rule. *J. Hered.* 113, 363–370. <https://doi.org/10.1093/jhered/esac005>.
- Oyler-McCance, S.J., St. John, J., Quinn, T.W., 2010. Rapid evolution in lekking grouse: implications for taxonomic definitions. *Ornithol. Monogr.* 67, 114–122.
- Oyler-McCance, S.J., DeYoung, R.W., Fike, J.A., Hagen, C.A., Johnson, J.A., Larsson, L. C., Patten, M.A., 2016. Rangewide genetic analysis of Lesser Prairie-chicken reveals population structure, range expansion, and possible introgression. *Conserv. Genet.* 17, 643–660.
- Palkovacs, E.P., Oppenheimer, A.J., Gladyshev, E., Toepfer, J.E., Amato, G., Chase, T., Caccone, A., 2004. Genetic evaluation of a proposed introduction: the case of the greater prairie chicken and the extinct heath hen. *Mol. Ecol.* 13, 1759–1769.
- Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., Reich, D., 2012. Ancient admixture in human history. *Genetics* 192, 1065–1093.
- Persons, N.W., Hosner, P.A., Meiklejohn, K.A., Braun, E.L., Kimball, R.T., 2016. Sorting out relationships among the grouse and ptarmigan using intron, mitochondrial, and ultra-conserved element sequences. *Mol. Phylogenet. Evol.* 98, 123–132.
- Phillips, J.C., 1928. Wild birds introduced or transplanted in North America. *USDA Tech. Bull.* No. p. 61.
- Poplin, R., Ruano-Rubio, V., DePristo, M.A., Fennell, T.J., Carneiro, M.O., Van der Auwera, G.A., Kling, D.E., Gauthier, L.D., Levy-Moonshine, A., Roazen, D., Shakir, K., Thibault, J., Chandran, S., Whelan, C., Lek, M., Gabriel, S., Daly, M.J., Neale, B., MacArthur, D.G., Banks, E., 2018. Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv* 10.1101/201178.
- Putnam, N.H., O'Connell, B.L., Stites, J.C., Rice, B.J., Blanchette, M., Calef, R., Troll, C.J., Fields, A., Hartley, P.D., Sugnet, C.W., Haussler, D., Rokhsar, D.S., Green, R.E., 2016. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Research* 26, 342–350.
- Ravinet, M., Faria, R., Butlin, R.K., Galindo, J., Bierne, N., Rafajlović, M., Noor, M.A.F., Mehlig, B., Westram, A.M., 2017. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *J. Evol. Biol.* 30, 1450–1477.
- Sæther, S.A., Sætre, G.-P., Borge, T., Wiley, C., Svedin, N., Andersson, G., Veen, T., Haavie, J., Servedio, M.R., Bureš, Kral, M., Hjernerquist, M.B., Gustafsson, L., Träff, J., Qvarnström, A., 2007. Sex chromosome-linked species recognition and evolution of reproductive isolation in flycatchers. *Science* 318, 95–97.
- Sawyer, S., Krause, J., Guschanski, K., Savolainen, V., Pääbo, S., 2012. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS ONE* 7, e34131.
- Silvy, N.J., Peterson, M.J., Lopez, R.R., 2004. The cause of the decline of pinnated grouse: the Texas example. *Wildl. Soc. Bull.* 32, 16–21.
- Soares, A.E.R., Novak, B.J., Haile, J., Heupink, T.H., Fjeldsø, J., Gilbert, M.T.P., Poinar, H., Church, G.M., Shapiro, B., 2016. Complete mitochondrial genomes of living and extinct pigeons revise the timing of the columbiform radiation. *BMC Ecol. Evol.* 16, 230.
- Sparling, D.W., 1980. Hybridization and taxonomic status of greater prairie chickens and sharp-tailed grouse (hybridization in grouse). *Prairie Nat.* 12, 92–102.
- Spaulding, A.W., Mock, K.E., Schroeder, M.A., Warheit, K.I., 2006. Recency, range expansion, and unsorted lineages: implications for interpreting neutral genetic variation in the sharp-tailed grouse (*Tympanuchus phasianellus*). *Mol. Ecol.* 15, 2317–2332.
- Swofford, D.L., 2003. PAUP*. Phylogenetic Analysis Using Parsimony (* and other methods). Version 4. <http://paup.phylosolutions.com>.
- Westemeier, R.L., Brawn, J.D., Simpson, S.A., Esker, T.L., Jansen, R.W., Walk, J.W., Kershner, E.L., Bouzat, J.L., Paige, K.N., 1998. Tracking the long-term decline and recovery of an isolated population. *Science* 282, 1695–1698.