The American Kestrel (*Falco sparverius*) Genoscape: Implications for Monitoring, Management, and Subspecies Boundaries

Michaela Brinkmeyer  
*Boise State University*

Julie A. Heath  
*Boise State University*

Publication Information


This is a pre-copyedited, author-produced PDF of an article accepted for publication in *Ornithology* following peer review. The version of record  
is available online at https://doi.org/10.1093/auk/ukaa051
Title: The American Kestrel (Falco sparverius) Genoscape: Implications for Monitoring, Management, and Subspecies Boundaries

Kristen C. Ruegg1,2*, Michaela Brinkmeyer3*, Christen M. Bossu1,2, Rachael A. Bay2,4, Eric C. Anderson5, Clint W. Boal6, Russell D. Dawso7, Amber Eschenbauch8, Christopher JW McClure9, Karl E. Miller10, Lance Morrow11, Jill Morrow11, M. David Oleyar12, Bill Ralph13, Sarah Schulwitz9, Ted Swem14, Jean-Francois Therrien15, Thomas B. Smith2,15, Julie A. Heath3

*These authors contributed equally.

Author Mailing Addresses:
1 Biology Department, Colorado State University, 251 W. Pitkin St, Fort Collins, CO 80521
2 Center for Tropical Research, Institute of the Environment and Sustainability, University of California, 619 Charles E Young Drive East, Los Angeles, CA 90095, USA
3 Department of Biological Sciences and Raptor Research Center, Boise State University, 1910 University Dr., Boise, ID 83725, USA.
4 Department of Evolution and Ecology, University of California Davis, Davis, CA 95616
5 Fisheries Ecology Division, Southwest Fisheries Science Center, National Marine Fisheries Service, 110 McAllister Way, Santa Cruz, CA 95060
6 US Geological Survey, Texas Cooperative Fish and Wildlife Research Unit, Texas Tech University, Lubbock, TX 79409
7 Ecosystem Science and Management Program, University of Northern British Columbia, 3333 University Way, Prince George, BC V2N 4Z9, Canada
8 Central Wisconsin Kestrel Research, 10115 Yellow Brick Road, Amherst WI 5440, USA.
9 The Peregrine Fund, 5668 West Flying Hawk Lane, Boise, ID 83709 USA.
10 Fish and Wildlife Research Institute, Florida Fish and Wildlife Conservation Commission, 1105 SW Williston Rd., Gainesville, FL 32601, USA.
11 Shenandoah Valley Raptor Study Area, Timberville, VA 22853, USA.
12 HawkWatch International, 2240 South 900 East, Salt Lake City, UT, 84106, USA.
13 Yosemite Area Audubon Society, 41932 Bessie Jacobs Rd., Raymond, CA 93653, USA.
14 U.S. Fish and Wildlife Service, 101 12th Avenue, Fairbanks, AK 99701, USA.
15 Hawk Mountain Sanctuary, 410 Summer Valley Road, Orwigsburg, PA, USA.
16 Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095, USA.

Corresponding author: Kristen Ruegg
Department of Biology
Colorado State University
Fort Collins, CO 80521
Kristen.Ruegg@colostate.edu

Running Head: Migrant Conservation Prioritization

Key Words: Genoscape, Conservation Unit, Migratory Bird
ACKNOWLEDGMENTS

KCR, JAH, MDO, SS, CJWM, TBS and MB conceived of the idea design and experiment.

KCR, MB, CMB, RB, and ECA collected the genomic data and conducted the bioinformatic and population genetic analyses. KCR, MB, JAH, and CMB wrote the manuscript with contributions from all co-authors. MB, CWB, RDD, AE, CJWM, KEM, LM, JM, MDO, BR, SS, TS, JFT, and JAH contributed blood samples for the genome-wide genetic analysis. We thank Joel Pagel and Little Gap Raptor Research for contributing samples to the genomic analysis, Anjolene Hunt for coordinating sample collection for the Kestrel Genoscape Project, the many generous supporters who contributed to the American Kestrel genoscape via a crowd-sourcing campaign, and the many citizen scientists who contributed feather samples for genetic analysis through the Peregrine Fund’s American Kestrel Partnership as well as other channels. This work was made possible by a generous gift from J. Ellis, a California Energy Commission grant to K. Ruegg and T. Smith (EPC-15-043), a National Geographic grant to K. Ruegg (WW-202R-17), a donation from First Solar Incorporated, the Boise State Raptor Research Center, the Raptor Research Foundation’s Dean Amadon Grant, and a grant to JAH from the Strategic Environmental Research and Development Program (RC-2702). Funding for the sample collection was provided by Natural Sciences and Engineering Research Council of Canada, Canada Foundation for Innovation, British Columbia Knowledge Development Fund, and University of Northern British Columbia. We thank the DNA Technologies and Expression Analysis Cores at the UC Davis Genome Center (supported by NIH Shared Instrumentation Grant 1S10OD010786-01) for their assistance with the Next-Generation Sequencing.
ABSTRACT: Identifying population genetic structure is useful for inferring evolutionary process and comparing the resulting structure with subspecies boundaries can aid in species management. The American Kestrel (*Falco sparverius*) is a widespread and highly diverse species with 17 total subspecies, only two of which are found north of United States/ Mexico border (*F. s. paulus* is restricted to southeastern United States, while *F. s. sparverius* breeds across the remainder of the United States and Canadian distribution). In many parts of their US and Canadian range, American Kestrels have been declining, but it has been difficult to interpret demographic trends without a clearer understanding of gene flow among populations. Here we sequence the first American Kestrel genome and scan the genome of 197 individuals from 12 sampling locations across the US and Canada in order to identify population structure. To validate signatures of population structure and fill in sampling gaps across the US and Canada range, we screened 192 outlier loci in an additional 376 samples from 34 sampling locations. Overall, our analyses support the existence of 5 genetically distinct populations of American Kestrels—Eastern, Western, Texas, Florida, and Alaska. Interestingly, we found that while our genome-wide genetic data support the existence of previously described subspecies boundaries in the US and Canada, genetic differences across the sampled range correlate more with putative migratory phenotypes (resident, long-distance, and short-distance migrants) rather than a priori described subspecies boundaries per se. Based on our results, we suggest the resulting five genetically distinct populations serve as the foundation for American Kestrel conservation and management in the face of future threats.
An important application of population genetics is the identification of genetically distinct populations within species that can be used to guide conservation and management efforts. Depending on the context, such groups are often referred to as subspecies, management units (MUs), evolutionary significant units (ESUs), conservation units, or genetically distinct populations (Moritz 1994, Allendorf and Luikart 2007, Funk et al. 2007). Population genetic structure below the species level has frequently been used to delineate units for conservation and management (Moritz 1994; Allendorf & Luikart 2007; Funk et al. 2007), but other factors including behavior and morphological variations are also important, particularly in species for which genetic data are absent or lacking in resolution (Mayr 1982, Waples et al. 2007). In highly mobile species it has historically been difficult to identify subspecies that correlate with genetically distinct populations because gene flow often homogenizes the diversifying effects of local adaptation and drift (Waples 1998, Willoughby et al. 2017, Medina et al. 2018, Doyle et al. 2018). As a result, identifying genetically distinct populations in migratory animals, such as migratory birds, remains a challenge (Larson et al. 2014, Zink 2014, Freer et al. 2015, Mura-Jornet et al. 2018).

Traditionally, genetic studies focused on identifying genetically distinct populations relied on a limited number of molecular markers (e.g., microsatellites, mitochondrial (mt) DNA sequences, and allozymes) to make inferences about population genetic structure (Ryman et al. 2006, Morin et al. 2009, Rowe et al. 2011, Mura-Jornet et al. 2018). However, recent advances in sequencing technology have made it possible to screen tens
of thousands to millions of genetic markers and reveal patterns of population structure that may have previously gone undetected (Rowe et al. 2011). For many species, greatly increasing the number of loci included in population genetic analyses has improved the precision of population genetic parameters (Egger et al. 2017), increased the resolution of detectable population genetic structure (Ruegg et al. 2014, Benestan et al. 2015, Jahner et al. 2016), and provided opportunities for fine scale investigations of genetically distinct populations and their relationship to subspecies boundaries (Larson et al. 2014, Fredrickson et al. 2015, Bussche et al. 2017, Mura-Jornet et al. 2018). In migratory birds, Ruegg et al. (2014) coined the term genoscapes to refer to maps of genetic variation across geographic space, but the relationship among genetically distinct populations within a genoscape and previously defined subspecies boundaries has yet to be explored.

Here we use Next Generation Sequencing (NGS) technology to create a genoscape for the American Kestrel, assess its relationship to current subspecies boundaries, and provide a framework for conservation and management of this and other highly mobile species with a high capacity for dispersal.

The American Kestrel is a widely distributed species that breeds throughout North and South America (Smallwood and Bird 2002) and has upwards of 17 recognized subspecies (Ferguson-Lees and Christie 2006). American Kestrels show highly variable migration strategies across their range, including individuals that migrate long distances, short distances, or do not migrate, and populations that are completely migratory, partially migratory, or non-migratory (Layne 1982, Bird and Palmer 1988, Henny and Brady 1994, Smallwood and Bird 2002). Here we focus on two American Kestrel subspecies found
north of the Mexico / United States border, the non-migratory subspecies (F. s. paulus) found breeding in the southeastern United States and the widespread subspecies (F. s. sparverius) found throughout the remainder the United States and Canada (Hoffman and Collopy 1988, Smallwood 1990). In general, it is thought that populations of F. s. sparverius follow a pattern of leap-frog migration, where migratory distance decreases on a latitudinal gradient, with birds in the northernmost part of the breeding range migrating the farthest and birds in the southernmost part of the range remaining year-round residents (Heath et al. 2012). In addition, there is growing evidence that American Kestrel populations are declining (Bird 2009, Farmer and Smith 2009, Smallwood et al. 2009, Hinnebusch et al. 2010), but estimates of demographic trends differ regionally (McClure et al. 2017). One hypothesis to explain regional variation in demographic trends is that genetically distinct populations with different migratory strategies are exposed to different stressors across the annual cycle. As a result, identifying genetically distinct populations in American Kestrels and how they correspond with previously defined subspecies will improve our ability to interpret recent demographic trends and appropriately focus conservation actions.

Previous American Kestrel genetic research using five microsatellite loci and one mitochondrial DNA marker identified no strong signal of population genetic structure across the U.S. range, with only subtle differences in allele frequencies between the two recognized subspecies (Miller et al. 2012). Here we employ NGS-sequencing technology to screen three times the number of samples and 10,000 times the number of genetic loci relative to previous work, and use the resulting data to re-evaluate patterns of population
structure in American Kestrels across their U.S. and Canadian breeding distribution. Specifically, we ask the following research questions: (1) Does genome-wide SNP data provide higher resolution of population structure than previous work based on fewer markers?, (2) How does the resulting population genetic structure relate to previously defined subspecies boundaries and variation in migratory behavior across the range?, and (3) How can the resulting genoscape be used help identify genetically distinct populations and develop hypotheses to explain regional variation in demographic trends?

METHODS

Sample collection and DNA extraction

Genetic samples were collected from 683 breeding adult or nestling American Kestrels from across the breeding range in North America in collaboration with several non-profit organizations, state agencies, university researchers, and citizen scientists (Table 1, Appendix 1). Blood (~ 30 µl) or pin feather samples were collected from 287 individuals for the construction of Restriction Site Associated DNA sequencing (RADseq) libraries. Blood was collected via brachial or jugular venipuncture, preserved in Queen’s lysis buffer (Seutin et al. 1991) and stored at -80°C or 3-4 pin feathers containing a small amount of blood in the base of the feather were collected from the breast of one nestling per brood and stored in envelopes at room temperature. The remainder of the feather samples used for high throughput SNP analyses (n=396) were collected from the breast of adult birds and stored in envelopes at room temperature. All samples were extracted using a QIAGEN DNeasy® Blood and Tissue Kits (San Francisco, CA) and blood and pin feather extractions were further quantified using Qubit dsDNA HS157 Assay kits.
(ThermoFisher Scientific) and visually inspected via gel electrophoresis to ensure selection of high quality, intact DNA for construction of RADseq libraries. Remaining tissue and blood samples, as well as remaining extractions, were curated and made available for future use in -20 and -80 freezers respectively in the Conservation Genomics Laboratory at Colorado State University.

**Genome sequencing, assembly and annotation**

To create a reference genome, the Illumina TruSeq DNA PCR-Free LT kit (Illumina) was used to prepare a genomic DNA library from a single individual from Boise, Idaho, following the adjustments made by Ruegg et al. (2018). The resulting library was sequenced on two lanes of an Illumina HiSeq2500 using 250 base-pair (bp) paired-end sequencing at the DNA Technologies and Expression Analysis Cores at the UC Davis Genome Center (Davis, CA). Initial contigs were assembled with the Discovar DeNovo assembler from the Broad Institute (http://www.broadinstitute.org), discarding contigs less than 1000bp in length. We also sequenced mate pair libraries with two insert sizes (4kb and 8kb) on one third of an Illumina HiSeq2500 2x100bp lane at the University of Utah Huntsman Cancer Center. Mate pair reads were trimmed with NxTrim (O’Connell et al. 2015) and scaffolds were generated with both paired end and mate pair libraries with SSPACE (overlap requirement k=3; Boetzer et al. 2011). The assembly was then broken at likely error regions using REAPR (Hunt et al. 2013) and SSPACE scaffolding was repeated with k=5 and scaffolds less than 1000bp were discarded for the final assembly.
For annotation purposes, REPEATMASKER (-species birds) (Tarailo-Graovac and Chen 2009) was used to replace repetitive regions of the final genome assembly with N’s. Two different ab initio gene predictions were used within the MAKER pipeline (Cantarel et al. 2008): SNAP and AUGUSTUS. The Zebra Finch (Taeniopygia guttata) cDNA and protein sequences were downloaded from Ensembl and used to train SNAP, and the available chicken (Gallus gallus domesticus) training dataset was used to train AUGUSTUS. INTERPROSCAN (Zdobnov and Apweiler 2001) was used to add Pfam protein annotation and gene ontology (GO) terms and identified 13,342 genes.

**SNP discovery and SNP filtering**

High-density RADseq was carried out on 287 individuals from 12 sampling locations following a modified version of the bestRAD library preparation protocol (SI Table 1; Ali et al. 2016). In short, DNA was normalized to a final concentration of 100ng in a 10ul volume, digested with restriction enzyme SBfl (New England Biolabs, NEB). The fragmented DNA was then ligated with SBfl specific adapters prepared with biotinylated ends and samples were pooled and cleaned using 1X Agencourt® AMPure XP beads (Beckman Coulter). Pooled and clean libraries were sheared to an average length of 400bp with 10 cycles on the Bioruptor NGS sonicator (Diagenode) to ensure appropriate length for sequencing and an Illumina NEBNext Ultra DNA Library Prep Kit (NEB) was used to repair blunt ends and ligate on NEBNext Adaptors to the resulting DNA fragments. Agencourt® AMPure XP beads (Beckman Coulter) were then used to select DNA fragments with an average length of 500bp, libraries were enriched with PCR, and cleaned again with Agencourt® AMPure XP beads. The resulting libraries were
sequenced on three lanes of an Illumina HiSeq 2500 at the UC Davis Genome Center using 250 base pair, paired-end sequencing, and 66 individuals with low coverage were re-sequenced on a fourth lane.

The program STACKS (Catchen et al. 2013) was used to demultiplex, filter and trim adapters from the data with the process_radtags function and remove duplicate read pairs using the clone_filter function. BOWTIE2 was used to map reads to the genome (Langmead and Salzberg 2012), and the HaplotypeCaller in the Genome Analysis Toolkit was used to identify SNPs (McKenna et al. 2010, Auwera et al. 2013). VCFTOOLS (Danecek et al. 2011) was used to remove indels, non-biallelic SNPs, and low quality and rare variants (genotype quality 20; coverage depth 10; minor allele frequency 0.05). The final number of SNPs and individuals to be retained for further analyses was assessed by visualizing the tradeoff between discarding low-coverage SNPs and discarding individuals with missing genotypes using custom scripts within the R-package GENOSCAPETOOLS (Anderson 2019). Because preliminary analyses revealed outliers in the principal components analysis and we were concerned about sample contamination among individuals during the library preparation stage, we filtered out individuals with >40% heterozygosity as heterozygosity is expected to be higher than expected in cases where multiple individuals are combined into a single well (SI Figure 2).

Identification of outlier SNPs for Population Assignment

Population genomic analyses were conducted on all SNPs that passed our filters to assess genome-wide patterns of genetic divergence and identify SNPs for population assignment.
and assay design. Population genetic structure was assessed by calculating pairwise population level $F_{ST}$ (with different sampling sites representing different populations) with bootstrapped confidence intervals using the R package ASSIGNER (Gosselin et al. 2019), and principal components analysis (PCA) using SNPRELATE (Zheng et al. 2012).

To test for isolation by distance we compared linearized $F_{ST}$ with pairwise geographic distance calculated from the central longitudinal and latitudinal coordinates of each location using the Vincenty ellipsoid method in the R package GEOSPERE (Hijmans 2019). Because the PCA of all SNPs from the genome-wide analysis revealed five major groups, including Alaska, Texas, the west, the east and Florida (see results, Figure 1), subsequent analyses focused on developing SNPs for population assignment within and among these groups.

To identify SNPs useful for population assignment between the 5 genetically distinct populations, we used VCFTOOLS (Danecek et al. 2011) to calculate site-wise $F_{ST}$ between populations and identify individual SNPs with the most power for discriminating between populations (SNPs with the biggest allele frequency differences). It is important to note that population genetic summary statistics were based on the full RADseq data set (see above) rather than downstream SNP dataset in order to avoid potential biases associated with selecting SNPs with the highest discriminatory power for population assignment.

Custom R-scripts were used to evaluate which of our top-ranking SNPs would generate designable assays based on the following parameters: 1) Guanine – Cytocene content was less than 0.65, 2) there were no insertions or deletions (indels) within 30bp of the variable site, or 3) there were no ambiguous codes within 20bp of the variable site.
Additionally, we used BWA-MEM (Burrows-Wheeler Aligner; Li and Durbin 2009) to determine which of our designable SNPs mapped uniquely to the reference genome. Fluidigm SNPtype assays (Fluidigm Inc.) were then developed in the 216 top ranking SNPs that passed our filters.

**Genetic screening and building the genoscape**

Ninety-three samples and three non-template controls were screened on the Fluidigm Corporation EP1 Genotyping System (Fluidigm Inc.) and assays were ranked by variability and call rate to identify the most reliable 192 SNP assays of 216 that were designed. The 192 variable SNP assays with the highest call rate were used to screen an additional 396 American Kestrel feather samples from 34 breeding locations in the United States and Canada in order to fill in sampling gaps and refine the resulting map of population genetic structure (Table 1). Following the methods described in Ruegg et al. (2014), we amplified PCR products using fluorescently labelled allele-specific primers and then used the EP1 Array Reader and Fluidigm’s automated Genotyping Analysis Software (Fluidigm Inc.) to call alleles with a confidence threshold of 90%. Each genotype was also visually inspected for potential irregularities and uncertain genotype calls were removed from the analysis. Samples with missing genotypes at >25% of SNP assays were removed from the analyses and SNP loci with >25% missing genotypes were removed, resulting in a total of 376 additional individuals at 186 SNP loci that could be used to identify genetic structure across the range (Table 1).
The final analysis of population genetic structure at 186 loci was conducted on a subset of the loci from the RADseq dataset combined with the SNP genotype-only dataset for a total of 683 individuals. The program *structure* (version 2.3.4; Pritchard et al. 2000) was used to assess how genetic variation is distributed across geographic space. The admixture model with the locprior option was run with uncorrelated allele frequencies, a burn-in period of 50,000, a total run length of 150,000 and assuming the number of genetic clusters (K) ranged from 1 to 10 (with 5 iterations run at each assumed value of K). We used the Evanno method to determine the number of genetic clusters. The Evanno method (Evanno et al. 2005), implemented in POPHELPER in R (Francis 2017), is an *ad hoc* method to determine the most probable number of population genetic clusters based on the rate of change in the log probability of data between successive K values. We used this algorithm to detect the uppermost hierarchical level of structure across the Kestrel breeding range and visually inspected subsequent structure plots to identify regions where geographic barriers to gene flow exist and/or where admixture homogenizes population structure. The resulting posterior probabilities of genetic group membership estimated from structure were visualized as transparency levels of different colors overlaid upon a base map from Natural Earth (naturalearthdata.com) and clipped to a map of the American Kestrel breeding range (NatureServe 2012), making use of the R packages SP, RGDAL, and RASTER (Bivand et al. 2013, 2017; Hijmans 2017). We scaled the transparency of colors within each distinguishable group, so that the highest posterior probability of membership in the group according to structure is opaque and the smallest is transparent. This creates a spatially-explicit map of the population structure analysis that we call the genoscape of the American Kestrel (Figure 2).
RESULTS

Genome sequencing, assembly and annotation

The final American Kestrel genome assembly is 1.23 Gb in length and consists of 5,096 scaffolds with an N50 of 941kb.

SNP discovery and SNP filtering

RAD-sequencing data from 287 individuals resulted in the identification of 199,705 bi-allelic loci with a minor allele-frequency greater than 5%, minimum quality score greater than 20, and minimum per individual sequencing depth greater than 10. After assessing the tradeoff between low coverage SNPs and missing genotypes (SI Figure 1), the data were further filtered to include 197 individuals, 12 populations, and 75,000 loci; 2 populations, one in CA and one in Idaho (Table 1), were subsequently dropped for the purposes of population genetic analyses as a result of low sample size (n<3). Seven outlier individuals with greater than >40% heterozygosity were also subsequently removed to avoid inclusion of samples potentially subject to contamination (as indicated by a histogram of the distribution of heterozygosity across all individuals which showed individuals above this threshold to be clear outliers, SI Figure 2). The final RADseq data set consisted of 197 individuals and 72,263 SNPs.

Population genetic structure

Significant pairwise $F_{ST}$ between the 10 sampling locations (2 were filtered out because they had fewer than 4 individuals, see above) ranged from 0.0010-0.0162 (Table 2). The positive correlation between $F_{ST}$ and geographic distance ($r^2 = 0.123; p$-value = 0.01)
suggests that isolation by distance contributes to genetic differentiation across the range. Overall, $F_{ST}$ was highest between non-migratory Florida and Texas sampling locations, while the genetic differentiation was the lowest among sampling locations within Eastern and Western breeding areas (Table 2). Principal components analysis based on 72,263 RAD-sequence loci revealed 4 main clusters with East, West, and Florida falling out separately, while Alaska and Texas overlapped (Figure 1). Principal component 1 was strongly influenced by data missingness, while PC2 and PC3 reflected differences in geography. Although there was overlap in principal component space between Texas and Alaska along PC2 and PC3, these groups were separated in the subsequent PC axes and in the search for loci representative of the observed population structure based on significant pairwise $F_{STs}$ and geographic distance. Overall, the first three principal components explained less than 3% of the total variation in allele frequencies.

Genoscape Construction / Structure Results

We successfully genotyped 376 samples collected from 34 breeding locations in the United States and Canada using the final panel of 192 SNP-type assays for population assignment (Table 1). Running structure with $K$ ranging from 1 to 10 revealed the strongest support for $K = 5$, where the plateau of delta $K$ (i.e. the greatest change in $K$) supports the uppermost hierarchical level of structure being $K = 3$ (SI Figure 3A; Evanno et al. 2005, Pritchard et al. 2000); however, subsequent increases of $K$, until $K = 5$, reduced the log likelihood of the model and the addition of the Florida ($K=4$) and Alaskan populations ($K=5$) were biologically feasible (SI Figure 4), suggesting that American Kestrel’s can be separated into 5 genetically distinct populations. Ultimately,
we find genetic distinctness of the residential Texas population, the residential Florida *F. s. paulus* subspecies, as well as distinct clustering of the Eastern and Western migratory populations, and an Alaskan migratory population (Figure 2).

**DISCUSSION**

Identifying distinct units for conservation is an important first step in the management of declining populations (Allendorf and Luikart 2007, Funk et al. 2012). Historically, conservation biologists have used a combination of morphological, behavioral, and genetic variation to define management or conservation units within species and these have sometimes, but not always, corresponded to subspecies boundaries. Here we generate a genoscape for the American Kestrel by sequencing the first American Kestrel genome, assessing population structure at 72,263 SNP markers screened in 12 populations from across the US and Canada migratory and non-migratory range, and validating patterns of population structure at 192 SNP markers screened in 34 populations. In contrast to previous work based on a more limited number of samples and markers that detected no major signals of population structure across the breeding range (Miller et al. 2012), our genoscape supports the existence of five genetically-distinct populations within American Kestrels found breeding across Canada and the United States (East, West, Alaska, Texas and Florida), one of which correlates with the previously identified southeastern subspecies (*F. s. paulus*). Overall, the most significant genetic differences occurred between the two resident populations (Texas and Florida), followed by differences between resident and migratory populations, and regional separation of Eastern and Western breeding populations. Here we discuss the utility of
the resulting genoscape for clarifying the relationship between previously defined
subspecies boundaries and genetically distinct populations identified using genome-wide
genetic data, as well as for providing a framework for developing hypotheses regarding
drivers of regional variation in demographic trends.

The question of whether subspecies represent defensible taxonomic units has been
controversial in the past because some molecular studies have failed to identify
subspecies as phylogenetically distinct (Barrowclough 1980, Mayr and Ashlock 1991,
O’brien and Mayr 1991, Ball and Avise 1992, Burbrink et al. 2000). Further, discord in
location of subspecies boundaries often arises when there are mismatches in the
timescales over which divergence occurs in various datasets; for example, subspecies
boundaries based on neutral genetic markers often diverge from subspecies boundaries
identified based on genetic or morphological markers that may be under selection (Haig
and Winker 2010). Using genome-wide sequencing, we found support for genetic
differentiation between the two US and Canadian subspecies of American Kestrels
breeding North of Mexico, *F. s. paulus* (southeastern) and *F. s. sparverius* (remainder of
the US and Canadian breeding range), but also found that divergence between the
subspecies is similar in magnitude to the degree of divergence detected between resident
*F. s. sparverius* in Texas and their migratory counterparts to the North. In general, levels
of genetic differentiation across the range were low and pairwise genetic distance versus
geographic distance suggest that patterns of divergence are in large part explained by
isolation by distance. Such low levels of differentiation are consistent with past studies of
American Kestrels based on fewer loci (Miller et al. 2012) and suggest that gene flow
may homogenize the diversifying effects of local adaptation and drift in high dispersal species like Kestrels (Willoughby et al. 2017; Doyle et al. 2018; Medina et al. 2018). While our genome-wide genetic analysis supports the existence of *F. s. paulus* as a genetically distinct subspecies, it also suggests that weak population structure within American Kestrels relates as much to migratory phenotype as it does to subspecies boundaries per se.

One possible explanation for the putative relationship between genetic differentiation and migratory phenotype in American Kestrels is that dispersal is limited between distinct migratory phenotypes, as has been found in other migratory systems (reviewed within Turbek et al. 2018). Like most raptors, Kestrels in Eastern and Western North America follow a strong north-to-south pattern of migration, with little longitudinal drift (Mueller and Berger 1967, Evans and Rosenfield 1985, Goodrich and Smith 2008). Additionally, the frequency of long-distance migration into Mexico is thought to increase from east to the west (Mueller & Berger 1967; Evans & Rosenfield 1985; Goodrich & Smith 2008), supporting the idea that Eastern and Western populations have different overwintering locations. One explanation for the observed genetic break in central North America is that separate Eastern and Western migratory routes and overwintering locations have resulted in the evolution of a weak migratory divide where gene flow is limited as a result of reproductive isolation between distinct migratory phenotypes, as has been documented across migratory divides in other avian taxa (reviewed within Turbek et al. 2018). Alternatively, the observed genetic break between Eastern and Western populations may have nothing to do with migratory phenotype per se, but instead may result from low
366 population density in Central North America (an additive effect of Isolation-by-Distance)
367 limiting gene flow between eastern and western groups (Winker et al 2010). Future work
368 will focus on quantifying migratory phenotypes in American Kestrels across their North
369 American range and assessing the relative contribution of migration and isolation by
370 distance to patterns of genetic divergence in this and other species with similar variation
371 in migratory phenotypes.
372
373 An alternative explanation for higher levels of divergence between resident and
374 migratory *F. sparverius* populations in our study is that divergence is not caused by
375 differences in migratory phenotype, but instead results from gene flow between resident
376 forms in Texas and Florida (*F. s. Paulus*) and resident subspecies further to the south (*F.
377 s. peninsularis* from Baja and W. Mexico, and *F. s. sparverioides* from Cuba). Previous
378 work suggests *F. s. sparverius* may hybridize with *F. s. peninsularis* in northern and
379 eastern Arizona (Bond 1943), but here we successfully genotyped 15 samples from
380 breeding migratory *F. s. sparverius* from Arizona and found that they all assigned clearly
381 to the western migratory group of *F. s. sparverius* rather than to the resident Texas
382 group. This result is opposite of what we would expect if high levels of divergence
383 between Texas residents and their migratory counterparts to the North was due to
384 hybridization between Texas birds and the *F. s. peninsularis* subspecies to the
385 South. While to the best of our knowledge, there are no known records of hybridization
386 between *F. s. sparverioides* from Cuba and *F. s. paulus* from Florida, eBird records of
387 kestrels from extreme southeastern Florida during the summer support the possibility that
388 *F. s. sparverioides* vagrants may occur within the same region. However, gene flow is
unlikely, given the infrequent (<1 per year) sightings of kestrels in southeastern Florida
and the fact that the nearest breeding population of *F. s. paulus* is >150 km to the north (FWC 2003). Thus, while on-going gene flow with resident subspecies to the south seems
like an unlikely explanation for the observed patterns of divergence between resident and
migratory populations north of Mexico, more extensive sampling south of the United
States border is needed to fully test all alternative hypotheses.

Heterogeneity in patterns of American Kestrel population decline across North America
suggest that regional populations are experiencing different threats and / or are
responding to the same threats differently (Butcher 1990, Smallwood et al. 2009, Sauer et
al. 2014, McClure et al. 2017), but past analyses have been limited by the lack of
genetically distinct populations. The results presented herein demonstrate the utility of
the genoscape approach for identifying five genetically distinct populations of American
Kestrels – East, West, Alaska, Texas and Florida, which can serve as the foundation for
the development of hypotheses to explain regional variation in demographic trends. For
example, while interpreting patterns of population decline from existing datasets is
complicated by known northward shifts in distribution (Paprocki et al. 2014), migration
count data from the Raptor Population Index project between 2006-2016 support the idea
that western populations have largely remained stable or are increasing, while eastern
populations are largely declining (SI Figure 3; Brandes et al. 2016, Crewe et al. 2016). In
addition, work focused specifically on understanding responses of American Kestrels to
climate change in the last decade supports the hypothesis that western populations are
migrating shorter distances and breeding earlier (Heath et al. 2012), while corresponding
changes in the east have not been documented. In light of the genoscape results presented herein, one hypothesis that warrants further exploration is that genetically-based differences in phenology between Eastern and Western groups affect population-specific responses to changing climate conditions, resulting in population decreases in the east, but not the west.

In conclusion, the American Kestrel genoscape reveals previously undetected levels of population structure among Eastern, Western, Alaska, Texas and Florida populations. While our data support the existence $F. s. sparverius$ and $F. s. paulus$ subspecies as genetically distinct groups, it also suggests that genetic differentiation is more closely tied to migratory phenotype (resident, long-distance, and short-distance migrants) than to previously defined subspecies boundaries. Based on our results, we suggest it would be ecologically appropriate to establish five management areas corresponding to the five genetically unique populations identified by our genoscape. More importantly, when the resulting genetically distinct populations are paired with data from existing long-term monitoring efforts, such as the Raptor Population Index, the results can be used to test hypotheses regarding drivers of observed population-specific responses to climate change and other stressors.

**Literature Cited**


Bivand, R. S., E. Pebesma, and V. Gómez-Rubio (2013). Applied spatial data analysis with R. Springer.


Table 1. Sampling locations and sample type used in the construction of the genoscape map for American Kestrels. The samples are broken down into those used for the RADsequencing analysis (pre and post filters) and the SNP genotyping (pre and post filtering). State and Province location codes correspond to Figure 1, while letter location codes correspond to the locations on Figure 2.

<table>
<thead>
<tr>
<th>STUDY SITE</th>
<th>LATITUDE</th>
<th>LONGITUDE</th>
<th>SAMPLE TYPE</th>
<th>N_{RAD_NoFilter}</th>
<th>N_{RAD_Filter}</th>
<th>N_{SNP_NoFilter}</th>
<th>N_{SNP_Filter}</th>
<th>LOCATION CODE</th>
<th>Fig 1 / Fig 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALBERTA, CANADA</td>
<td>56.16262</td>
<td>-117.34</td>
<td>Feather</td>
<td>12</td>
<td>12</td>
<td></td>
<td></td>
<td>E</td>
<td></td>
</tr>
<tr>
<td>ALASKA, USA</td>
<td>63.96041</td>
<td>-145.136</td>
<td>Blood, feather</td>
<td>25</td>
<td>18</td>
<td>7</td>
<td>7</td>
<td>AK / A</td>
<td></td>
</tr>
<tr>
<td>ARIZONA, USA</td>
<td>33.45007</td>
<td>-111.943</td>
<td>Feather</td>
<td>20</td>
<td>15</td>
<td></td>
<td></td>
<td>G</td>
<td></td>
</tr>
<tr>
<td>CALIFORNIA (1), USA</td>
<td>37.23551</td>
<td>-120.244</td>
<td>Blood, feather</td>
<td>26</td>
<td>10</td>
<td>22</td>
<td>22</td>
<td>CA / D1</td>
<td></td>
</tr>
<tr>
<td>CALIFORNIA (2), USA</td>
<td>33.7315</td>
<td>-117.878</td>
<td>Blood, feather</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>CA / D2</td>
<td></td>
</tr>
<tr>
<td>CONNECTICUT (1), USA</td>
<td>41.95001</td>
<td>-73.364</td>
<td>Feather</td>
<td>10</td>
<td>10</td>
<td></td>
<td></td>
<td>Y1</td>
<td></td>
</tr>
<tr>
<td>CONNECTICUT (2), USA</td>
<td>41.94374</td>
<td>-72.5072</td>
<td>Feather</td>
<td>10</td>
<td>7</td>
<td></td>
<td></td>
<td>Y2</td>
<td></td>
</tr>
<tr>
<td>DELAWARE, USA</td>
<td>39.75047</td>
<td>-75.6969</td>
<td>Feather</td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
<td>W</td>
<td></td>
</tr>
<tr>
<td>FLORIDA, USA</td>
<td>29.54151</td>
<td>-82.2334</td>
<td>Blood, feather</td>
<td>18</td>
<td>16</td>
<td>20</td>
<td>20</td>
<td>FL / AA</td>
<td></td>
</tr>
<tr>
<td>IDAHO (1), USA</td>
<td>43.51209</td>
<td>-116.377</td>
<td>Blood, feather</td>
<td>33</td>
<td>29</td>
<td>18</td>
<td>11</td>
<td>ID / F1</td>
<td></td>
</tr>
<tr>
<td>IDAHO (2), USA</td>
<td>43.27597</td>
<td>-115.003</td>
<td>Blood, feather</td>
<td>9</td>
<td>2</td>
<td>6</td>
<td>5</td>
<td>ID / F2</td>
<td></td>
</tr>
<tr>
<td>ILLINOIS, USA</td>
<td>38.30563</td>
<td>-89.7969</td>
<td>Feather</td>
<td>5</td>
<td>3</td>
<td></td>
<td></td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>KANSAS, USA</td>
<td>38.77533</td>
<td>-95.2486</td>
<td>Feather</td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
<td>M</td>
<td></td>
</tr>
<tr>
<td>MASSACHUSETTS, USA</td>
<td>42.15064</td>
<td>-71.8469</td>
<td>Feather</td>
<td>20</td>
<td>18</td>
<td></td>
<td></td>
<td>Z</td>
<td></td>
</tr>
<tr>
<td>MICHIGAN (1), USA</td>
<td>42.3216062</td>
<td>-86.182447</td>
<td>Feather</td>
<td>6</td>
<td>6</td>
<td></td>
<td></td>
<td>R1</td>
<td></td>
</tr>
<tr>
<td>MICHIGAN (2), USA</td>
<td>44.9774725</td>
<td>-85.712026</td>
<td>Feather</td>
<td>6</td>
<td>6</td>
<td></td>
<td></td>
<td>R2</td>
<td></td>
</tr>
<tr>
<td>MICHIGAN (3), USA</td>
<td>44.9034875</td>
<td>-83.4873</td>
<td>Feather</td>
<td>8</td>
<td>8</td>
<td></td>
<td></td>
<td>R3</td>
<td></td>
</tr>
<tr>
<td>MISSOURI (1), USA</td>
<td>39.04824</td>
<td>-94.3663</td>
<td>Feather</td>
<td>8</td>
<td>5</td>
<td></td>
<td></td>
<td>N1</td>
<td></td>
</tr>
<tr>
<td>MISSOURI (2), USA</td>
<td>38.3804</td>
<td>-90.925</td>
<td>Feather</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td>N2</td>
<td></td>
</tr>
<tr>
<td>MONTANA, USA</td>
<td>45.76764</td>
<td>-111.154</td>
<td>Feather</td>
<td>19</td>
<td>15</td>
<td></td>
<td></td>
<td>I</td>
<td></td>
</tr>
<tr>
<td>NORTH CAROLINA, USA</td>
<td>35.73285</td>
<td>-82.6776</td>
<td>Feather</td>
<td>17</td>
<td>16</td>
<td></td>
<td></td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>NEBRASKA, USA</td>
<td>40.99115</td>
<td>-96.8957</td>
<td>Feather</td>
<td>6</td>
<td>6</td>
<td></td>
<td></td>
<td>L</td>
<td></td>
</tr>
<tr>
<td>NEW JERSEY, USA</td>
<td>40.4297</td>
<td>-74.9271</td>
<td>Feather</td>
<td>30</td>
<td>28</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>NEW YORK, USA</td>
<td>41.8773</td>
<td>-74.7309</td>
<td>Feather</td>
<td>4</td>
<td>3</td>
<td></td>
<td></td>
<td>U</td>
<td></td>
</tr>
<tr>
<td>OREGON (1), USA</td>
<td>45.57362</td>
<td>-123.048</td>
<td>Blood, feather</td>
<td>14</td>
<td>11</td>
<td>3</td>
<td>3</td>
<td>OR / C1</td>
<td></td>
</tr>
<tr>
<td>OREGON (2), USA</td>
<td>44.36466</td>
<td>-121.412</td>
<td>Feather</td>
<td>25</td>
<td>24</td>
<td></td>
<td></td>
<td>C2</td>
<td></td>
</tr>
<tr>
<td>PENNSYLVANIA, USA</td>
<td>40.60183</td>
<td>-75.5961</td>
<td>Blood, feather</td>
<td>37</td>
<td>23</td>
<td>13</td>
<td>13</td>
<td>PA / V</td>
<td></td>
</tr>
<tr>
<td>SASKATCHEWAN, CANADA</td>
<td>55.01968</td>
<td>-105.893</td>
<td>Blood, feather</td>
<td>40</td>
<td>28</td>
<td>12</td>
<td>11</td>
<td>SK / K</td>
<td></td>
</tr>
<tr>
<td>TENNESSEE, USA</td>
<td>35.16944</td>
<td>-90.0489</td>
<td>Feather</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
<td>O</td>
<td></td>
</tr>
<tr>
<td>TEXAS, USA</td>
<td>33.63184</td>
<td>-101.889</td>
<td>Blood, feather</td>
<td>13</td>
<td>12</td>
<td>1</td>
<td>1</td>
<td>TX / B</td>
<td></td>
</tr>
<tr>
<td>UTAH, USA</td>
<td>40.62255</td>
<td>-111.987</td>
<td>Feather</td>
<td>30</td>
<td>28</td>
<td></td>
<td></td>
<td>H</td>
<td></td>
</tr>
<tr>
<td>VIRGINIA, USA</td>
<td>38.6457</td>
<td>-78.7522</td>
<td>Blood, feather</td>
<td>38</td>
<td>29</td>
<td>17</td>
<td>17</td>
<td>VA / T</td>
<td></td>
</tr>
<tr>
<td>WISCONSIN, USA</td>
<td>43.77114</td>
<td>-89.4059</td>
<td>Blood, feather</td>
<td>30</td>
<td>18</td>
<td>24</td>
<td>24</td>
<td>WI / Q</td>
<td></td>
</tr>
<tr>
<td>WYOMING, USA</td>
<td>43.60488</td>
<td>-110.658</td>
<td>Feather</td>
<td>17</td>
<td>15</td>
<td></td>
<td></td>
<td>j</td>
<td></td>
</tr>
</tbody>
</table>

TOTAL               | 287      | 197       | 396          | 376              |
Table 2. Weighted mean pairwise Weir and Cockerham (1984) FST values (upper triangle) and confidence intervals (lower triangle) between all American kestrel sampling locations with greater than five individuals. Sampling locations correspond to those listed in Figure 1 and Table 1 and are organized by approximate migration strategy. Color coding indicates degree of differentiation between locations with the greatest values of genetic divergence between nonmigratory populations in Florida and Texas.

<table>
<thead>
<tr>
<th>Complete Migrant</th>
<th>SK1</th>
<th>OR1</th>
<th>ID1</th>
<th>CA1</th>
<th>WI1</th>
<th>PA1</th>
<th>VA1</th>
<th>Resident</th>
</tr>
</thead>
<tbody>
<tr>
<td>AK1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0096</td>
</tr>
<tr>
<td>SK1</td>
<td>0.0028 - 0.0034</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0078</td>
</tr>
<tr>
<td>OR1</td>
<td>0.0036 - 0.0046</td>
<td>0.0015 - 0.0023</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0077</td>
</tr>
<tr>
<td>ID1</td>
<td>0.0044 - 0.0050</td>
<td>0.0014 - 0.0019</td>
<td>0.0021 - 0.0030</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0088</td>
</tr>
<tr>
<td>CA1</td>
<td>0.0048 - 0.0058</td>
<td>0.0010 - 0.0019</td>
<td>0.0025 - 0.0037</td>
<td>0.0015 - 0.0023</td>
<td></td>
<td></td>
<td></td>
<td>0.0099</td>
</tr>
<tr>
<td>WI1</td>
<td>0.0056 - 0.0065</td>
<td>0.0012 - 0.0018</td>
<td>0.0040 - 0.0052</td>
<td>0.0037 - 0.0043</td>
<td>0.0035 - 0.0044</td>
<td></td>
<td></td>
<td>0.0100</td>
</tr>
<tr>
<td>PA1</td>
<td>0.0058 - 0.0065</td>
<td>0.0021 - 0.0027</td>
<td>0.0046 - 0.0056</td>
<td>0.0048 - 0.0054</td>
<td>0.0047 - 0.0057</td>
<td>0.0006 - 0.0012</td>
<td></td>
<td>0.0103</td>
</tr>
<tr>
<td>VA1</td>
<td>0.0082 - 0.0090</td>
<td>0.0043 - 0.0048</td>
<td>0.0068 - 0.0079</td>
<td>0.0069 - 0.0075</td>
<td>0.0065 - 0.0074</td>
<td>0.0025 - 0.0032</td>
<td>0.0025 - 0.0030</td>
<td></td>
</tr>
<tr>
<td>Resident</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0156 - 0.0168</td>
</tr>
</tbody>
</table>

| TX1              | 0.0091 - 0.0102 | 0.0074 - 0.0083 | 0.0070 - 0.0082 | 0.0083 - 0.0092 | 0.0092 - 0.0106 | 0.0094 - 0.0105 | 0.0099 - 0.0108 | 0.0121 - 0.0132 |     | 0.0162  |
| FL1              | 0.0120 - 0.0129 | 0.0099 - 0.0106 | 0.0105 - 0.0116 | 0.0121 - 0.0129 | 0.0122 - 0.0134 | 0.0076 - 0.0088 | 0.0072 - 0.0081 | 0.0089 - 0.0098 |     |         |

**Key**
- FST < 0.005
- 0.005 ≤ FST < 0.01
- 0.01 ≤ FST < 0.015
- FST ≥ 0.015
Figure 1. Principal components analysis of 72,263 Single Nucleotide Polymorphism markers from across the US and Canadian breeding range of American Kestrels showing separation among Eastern, Western, Alaska, and Texas genetically distinct populations, and the Florida subspecies. Each genetically distinct population is encircled by an ellipse. Separation between Alaska and Texas conservation units occurred along Principle Components axes not shown and was also evident in pairwise $F_{ST}$ calculations.
Figure 2. The American Kestrel genoscape. 
A) Structure plot showing support for K= 5 
genetic groups. Letters correspond to 
population locations on the map as well as 
sample numbers listed in Table 1. B) A 
spatially explicit representation of the 
population structure results showing the 
biggest genetic differences among eastern, 
western, Texas, Florida and Alaska genetic 
groups. Dots with circles around them 
indicate sampling locations where both 
Restriction-site associated-DNA sequencing 
and Single Nucleotide Polymorphism 
genotyping was conducted. The dashed line 
indicates the hypothesized southern 
boundary of F. s. sparvarius (Lane & Fischer 
1997).
Figure 3. The relationship between pairwise linearized $F_{ST}$ and geographic distance for American Kestrel based on analysis of Restriction-Site-Associated DNA sequencing data. The significant $r^2$ supports the idea that genetic divergence across the range is largely a result of Isolation-by-Distance.
We determined the final number of single nucleotide polymorphisms and individuals to be retained in analyses of population genetic structure of American Kestrels by visualizing the tradeoff between discarding low coverage single nucleotide polymorphisms (y-axis) and individuals with missing genotypes (x-axis). Based on this analysis, we retained 75,000 SNPs and 204 individuals for further filtering using percent heterozygosity per individual and per locus.
Figure SI 2. Histogram of heterozygosity per individual. The above plot was used to identify high heterozygosity outliers above 0.40 which we subsequently removed from the analysis due to suspected contamination of samples.
SI Figure 3. Population Structure in American Kestrels assuming different numbers of populations. Ancestry plots generated in *structure* corresponding to K values of 2 – 6, illustrating support for 5 genetically distinct lineages within American Kestrels. The sampling locations along the top bar correspond to the locations illustrated in Figure 2 and Table 1.
**SI Figure 4.** The Evanno method for determining the appropriate number of genetic clusters within the data (K). A) The change in K versus the number of clusters demonstrates K= 3 is the uppermost structure in the data. B) Visual inspection of the log likelihood values illustrate that K=5 represents the best likelihood that is also a biologically reasonable value of K for the number of populations within the American Kestrel data.
Figure SI 5. Population trends for American Kestrels from Raptor Population Index. A summary of population trend estimates based on migration count data from the Raptor Population Index (RPI) project (Crewe et al. 2016; Brandes et al. 2016). Green arrows represent areas of significant increase, blue arrows represent areas of slight increase (pointing up) or decrease (pointing down), red arrows represent areas of significant decrease, and blue circles signify areas with stable population trends. These data demonstrate the variability of population trend estimates of American Kestrels in North America, with the greatest declines occurring in the east. No trends are available for Alaska or Canada.