FINAL PERFORMANCE REPORT

Federal Aid Grant No. F19AP00196 (E-91-R-1)

Genomic Assessment of Isolation and Connectivity Among Black-capped Vireo Populations in Oklahoma

Oklahoma Department of Wildlife Conservation

Grant Period: January 1, 2019 – December 31, 2019

Reporting Period: January 1, 2019 – December 31, 2019

FINAL PERFORMANCE REPORT

State: Oklahoma **Grant Number:** F19AP00196 (E-91-R-1)

Grant Program: Endangered Species Act Traditional Section 6

Grant Title: Genomic Assessment of Isolation and Connectivity Among Black-capped Vireo Populations in Oklahoma

Grant Period: January 1, 2019 – December 31, 2019

Reporting Period: January 1, 2019 – December 31, 2019

Principal Investigators: Matthew J. Miller, Sam Noble Oklahoma Museum of Natural History, University of Oklahoma

I. ABSTRACT:

The formerly-endangered Black-capped Vireo (*Vireo atricapilla*) was historically found across much of west-central and central Oklahoma. Due to widespread habitat fragmentation, Black-capped Vireos experienced a severe population bottleneck from isolation during the 1980s and 1990s. We sought to produce a high-quality, high-coverage reference genome for the Black-capped Vireo**.** Furthermore, we planned to calculate population genetic variation statistics and historical demography parameter estimates for Oklahoma Black-capped Vireos using whole genome sequences mapped to the reference genome. We were unable to complete the reference genome. Furthermore, we attempted to sequence the remaining 24 samples in late 2019. However, our libraries failed at the final quality control step. Due to various administrative complications, we were unsuccessful in accomplishing our goals.

II. BACKGROUND:

The Black-capped (*Vireo atricapilla*) was historically found across much of west-central and central Oklahoma (Graber 1957). However, since Graber's study in the 1950's, the vireo's population size and range in Oklahoma and elsewhere has declined. Surveys in the mid-1980s and 1990s noted them in only a few areas of Oklahoma and in low numbers (Blaine County, Caddo/Canadian counties, and the Wichita Mountains, Comanche County; Grzybowski et al. 1985). Fewer than 90 pairs occurred in Oklahoma, possibly fewer than 70, most of those in the Wichita Mountains.

The Wichita Mountains vireo population has been monitored and managed since the late 1980s (primarily through cowbird removal and managed or controlled burning), and has increased to an

estimate of over 5,000 pairs by 2010, although severe droughts in 2011 and 2012 cycled estimated population size down to about 3,000 pairs, recovering to almost 4,000 estimated pairs in 2015 (McDonald and Grzybowski, unpubl. data). The Blaine County group was monitored and managed through 2005 when 12-13 pairs were known present (Grzybowski 2005). Recently, a survey of Blaine County found them still present in the Salt Creek Canyon area (Grzybowski 2016, Ross, Bessozi Grzybowski, unpubl. data). A new group of vireos was recently discovered in isolated mountain outcrops in southwestern Kiowa County (Grzybowski 2016); these were likely present through this period. A small group located in central Oklahoma in the mid-1990s is likely extirpated (Stubbs, pers. comm.). Small pockets of vireos present in the Caddo Canyonlands of Caddo and Canadian counties were extirpated before the early 1990s (Grzybowski 1992).

Collectively, the Black-capped Vireos in Oklahoma have passed through a severe population bottleneck during the 1980s and 1990s. The Wichita Mountains group has increased substantially in this period. However, there is no evidence the two outlier populations in Blaine and Kiowa counties have ever been larger than 30 or fewer pairs. Only a few vireos have been encountered in other locations in Oklahoma since 2000.

Several questions arise from this observation. First, “did the bottleneck reduce genetic variation and cause inbreeding in Black-capped Vireo populations in Oklahoma?”; secondly, “is this effect more pronounced in the two peripheral populations (Kiowa and Blaine counties) relative to the larger population at Ft. Sill?”; and thirdly, “are the two peripheral populations genetically connected to the larger Ft. Sill population?”. Because Black-capped Vireos’ migration, it is possible that some of the birds arriving in the spring to the Blaine and Kiowa county populations originated from the Ft. Sill population. The answers to these questions have important implications for Black-capped Vireo management.

Prior studies have attempted to assess genetic diversity and genetic connectivity among Black-capped Vireo populations. The first study was completed by Fazio (2004) using allozymes and showed some differentiation between sampled areas, but only assessed two functional loci. Allozymes are low-power molecular markers and are no longer used in contemporary studies. Barr et al. (2008) found limited variation in microsatellite allele frequencies, interpreting local mixing, but with geographic structure at larger spatial scales. Zink et al. (2010) used short mtDNA sequences sampled across a broader range (including samples from Mexico) and showed results similar to Barr et al. (2008). His group interpreted their results as suggesting insignificant range-wide population structure. The consequence of that finding is that Oklahoma vireos should not be managed separately from other vireos across their range. However, the use of short mtDNA fragments for avian conservation genomics has been broadly criticized (e.g. McCormack & Maley 2015). Athrey et al. (2012a,b) refined their microsatellite analyses to distinguish differences in dispersal between males and females, and recovered substantial

population structure at a broader scale. Most recently, Vazquez-Miranda (2015) sequenced mitochondrial plus a few autosomal genetic loci, and continued the interpretation of low population structure. However, like Zink et al. (2010), these markers are typically used to determine systematic relationships among avian genera and families rather than within-species population structure, and are therefore inappropriate for questions of dispersal, isolation and potential management. While the Barr and Zink schools are still ambivalent or at odds on population structure in Black-capped Vireos, these studies generally used less than 20 markers. We feel as though this is wholly inappropriate for the conservation and management of endangered species, especially when whole genome sequencing is relatively affordable.

Here, we planned to leverage the incredible power of whole genome sequencing to answer questions of genomic variation, inbreeding, and connectivity in the three Black-capped Vireo populations included in this study.

III. OBJECTIVES:

1) Produce a high-quality, high-coverage genome for the Black-capped Vireo.

2) Perform ddRAD (Double-Digest Restriction-site Associated DNA sequencing (ddRAD-Seq) to reduce the genome to a common set of unique single-nucleotide markers (expected to be 1500-2500 markers) for up to 140 individuals from the three locations in Oklahoma.

3) Calculate population genetic variation statistics and historical demography parameter estimates for Oklahoma Black-capped Vireos.

IV. PROCEDURES:

We estimated genetic diversity in each of the three study populations by measuring observed heterozygosity (*Ho*) across the genome. We also estimated inbreeding by examining the distribution of short and long runs of homozygosity across the genome in individual Black-capped Vireos. Finally, we measured connectivity using single nucleotide polymorphisms from across the genome.

A. Identify best samples for each section of the study

Preceding the initiation of this grant, J. Gryzbowski and his colleagues collected over 200 blood samples from Black-capped Vireos across Oklahoma. We used the following criteria to identify which of these samples to include in the study. First, we wanted equal numbers from each of our study sites:

a) Salt Creek Canyon, Blaine Co., Oklahoma (36.08 -98.42)
b) Quartz Mountain, Kiowa Co., Oklahoma (34.92, -99.25)
c) Ft. Sill, Comanche Co., Oklahoma (34.75, -98.58)

Secondly, we decided to include only one female from each location in the study, with the remaining birds being males. This decision was made for two reasons. First, in birds, the females are the heterogametic sex, meaning that they have one Z and one W chromosome, while males have two Z chromosomes. The W chromosome is relatively small, has few functional genes, and is largely composed of repetitive regions; for these reasons it is often difficult to properly assemble and therefore is usually excluded from population-level studies (Smeds et al., 2015). Furthermore, because the W chromosome is only represented in any bird by at most one copy, it is impossible to estimate inbreeding within a single individual by comparing the degree of identity of the two chromosomal copies that are present in all other chromosomes, and in the two Z chromosomes of male. For this reason, males are often preferred in genomic studies because sequencing reads are a random sampling of the entire genome. By focusing on males, we also maximize our sequencing effort on chromosomes likely to give the strongest signal.

However, a female is preferable to a male bird for building a reference genome. This is because we do want to build into our reference genome both the Z and the W chromosome so that reads from all other birds in population studies are correctly mapped to the proper sex chromosome. We chose OK_Vireo01, a female from Ft. Sill, as the bird for our reference genome. This is because the raw material going into the genomic library preparation was best (highest DNA concentration: 30 ng/uL) and because the resulting sequence library had the largest number of recovered reads (>112,800,000 reads).

Blood samples were taken via a 26 g hypodermic needle inserted into the cutaneous ulnar vein where it passes over the humero-radioulnar joint. We swabbed vaseline onto the site using a Q-tip prior to venipuncture to facilitate blood droplet formation on the skin. Blood was collected via capillary action into 100 ml microhematocrit tubes, and then transferred to a 2 uL tube filled with 1.8 mL of Longmire's lysis. Because Black-capped Vireos weigh < 10 grams, and because best-practices recommend capping blood collection at 1% of body weight, we took no more than 100 ml (i.e. 1 tube).

We selected samples based on a visual inspection of the blood tubes. We looked for samples that were darkish pink, meaning that they have a reasonable amount of blood saved, but not so much blood that the recommended 1:9 ratio of blood:Longmire's lysis buffer was exceeded (Longmire et al., 1997). Too much blood for too little buffer has the perverse result that DNA degradation is more likely to occur, thereby resulting in poorer DNA yield during extraction and poorer genomic libraries. Table 1 lists the samples included in this study.

Table 1: Information about samples used to generate genomic data in this study
* indicates sample chosen for generating the reference whole genome

Sample ID	USFW Migratory Bird Band Number	Sampling Location	Sex	Number of base pairs sequenced
OK_Vireo01*	1570-67682	Blaine Co, OK	Female	16,930,898,400
OK_Vireo02	1570-67680	Blaine Co, OK	Male	16,290,059,400
OK_Vireo03	1570-67676	Blaine Co, OK	Male	
OK_Vireo04	2350-02416	Blaine Co, OK	Male	
OK_Vireo05	1570-67678	Blaine Co, OK	Male	
OK_Vireo06	2350-02421	Blaine Co, OK	Male	
OK_Vireo07	1570-67679	Blaine Co, OK	Male	
OK_Vireo08	1570-67677	Blaine Co, OK	Male	
OK_Vireo09	1570-67688	Blaine Co, OK	Male	
OK_Vireo10	2350-02423	Blaine Co, OK	Male	
OK_Vireo11	2350-02499	Quartz Mtn, OK	Female	18,089,005,200
OK_Vireo12	2350-02511	Quartz Mtn, OK	Male	
OK_Vireo13	2350-02505	Quartz Mtn, OK	Male	
OK_Vireo14	2350-02497	Quartz Mtn, OK	Male	
OK_Vireo15	2350-02508	Quartz Mtn, OK	Male	
OK_Vireo16	2350-02498	Quartz Mtn, OK	Male	
OK_Vireo17	2350-02503	Quartz Mtn, OK	Male	
OK_Vireo18	2350-02513	Quartz Mtn, OK	Male	
OK_Vireo19	2350-02509	Quartz Mtn, OK	Male	
OK_Vireo20	2350-02507	Quartz Mtn, OK	Male	15,718,332,900
OK_Vireo21	2350-02441	Ft. Sill, OK	Female	13,953,084,300
OK_Vireo22	2350-02368	Ft. Sill, OK	Male	18,089,005,200
OK_Vireo23	2350-02485	Ft. Sill, OK	Male	
OK_Vireo24	2350-02486	Ft. Sill, OK	Male	
OK_Vireo25	2350-02523	Ft. Sill, OK	Male	
OK_Vireo26	2350-02493	Ft. Sill, OK	Male	
OK_Vireo27	2350-02527	Ft. Sill, OK	Male	
OK_Vireo28	2350-02528	Ft. Sill, OK	Male	

OK_Vireo29	2350-02334	Ft. Sill, OK	Male	
OK_Vireo30	2350-02484	Ft. Sill, OK	Male	

B. DNA extraction, and genomic library preparation

We also generated one (1) long read sequencing library from sample OK_Vireo01 as part of our effort to generate a reference genome. We began by generating a separate high molecular weight (e.g. ultralong fragment) DNA extract from that sample. To do this we followed the manufacturer's recommended protocol for DNA extraction from blood, except that we added 150 uL of phosphate-buffered saline (PBS) to 5o uL of buffered blood. The resultant DNA was used to create an Oxford Nanopore long-read sequencing library. We used the Ligation Sequencing Kit SQK-LSK109 (Oxford Nanopore) and followed the manufacturer's protocol. We sequenced the library on one MinION 9.1 Flow Cell (Oxford Nanopore). That library generated ~ 3 gigabases of DNA sequence data with an N50 read length of 20 kilobases, meaning that half of the total DNA bases recovered are found in fragments larger than 20 kilobases. This is a typical yield for long-read bird DNA sequencing runs on the MinION. Based on our recent reference genome assembly of the Golden-fronted Woodpecker (Wiley and Miller 2020), we have learned that we can more efficiently generate the ~30 Gigabases of Oxford Nanopore data by sending high molecular weight DNA to the UC Davis Core Sequencing facility in California where they sequence high molecular weight DNA using Oxford Nanopore's larger Promethion sequencing device. The Promethion generates about 10× the data for about 3× the cost.

To generate the standard resequencing libraries for all other samples, we extracted total genomic DNA using the Monarch® Genomic DNA Purification Kit (New England Biolabs) following the manufacturer's protocol for blood samples. We evaluated whether DNA extraction was successful by measuring DNA samples on a Quantus Florometer (Promega Corporation). The Quantus device uses fluorescent detection of double stranded DNA. On average, our samples generated 21.5 ng/uL of double stranded DNA in 200 uL solution. We also visualized our DNA extracts on a TapeStation system (Agilent) which confirmed that most of our DNA was of high molecular weight.

To begin with we fragmented DNA on a Covaris S220 to a fragment size of approximately 350 base pairs. This is the preferred fragment size for genomic DNA libraries on the Illumina sequencing platform. We used the NEBNext® Ultra II DNA library prep kit to generate our sequencing libraries. This kit begins with DNA end repair and A-tailing followed by adapter ligation. We used NEBNext® Multiplex Oligos for Illumina® for our sample barcodes. These barcodes are unique dual index barcodes to single-index or combinatorial dual index barcodes because they use a non-redundant pair of barcodes on the left and right hand side of the DNA fragment. This greatly reduces the chance of index-hopping, which is when one DNA read is

improperly assigned to the wrong sample. Using unique dual indices is considered a best practice in genomics.

We quality checked our genomic libraries in the following ways. First, we measured library DNA concentration, expecting to see a concentration of at least 5 ng/uL. We also ran samples on the TapeStation to look for two things. First, the fragment size should increase because of the addition of barcodes and Illumina adapters. Second, we expect to see a more or less normal distribution of fragment sizes around 450-500 base pairs. It is important to have normal rather than skewed fragment sizes, because smaller fragments get preferentially sequenced on the Illumina sequencing machine. For this reason, it is important that all the samples have fragment peaks at around the same size, otherwise the samples with lower peaks will be over-represented in the data. After quality control, we pooled all six samples at equimolar concentrations and sent the libraries off to the sequencing facility.

Initially, we generated genomic libraries from the first six samples: OK_Vireo01 – OKVireo6 to ensure that our estimates of genome size, etc. were appropriate for the study. These six libraries were pooled and sequenced at the Oklahoma Medical Research Foundation Next Generation Sequencing Core Lab, which has an Illumina NovaSeq sequencing device. All six sequencing libraries came back with high quality data (see below) around our ideal coverage of ~12–15×, based on our estimate of a 1.25 gigabase genome size for the Black-capped Vireo. No sample had predicted coverage lower than 10×, which was our minimum cutoff for this study.

C. Bioinformatic data analysis

The bioinformatic pipeline to generate the high-quality reference genome is described in detail in Wiley and Miller (2020). Briefly, we used *flye* (Kolmogorov et al., 2019) to generate a first draft highly-contiguous but error-prone genome assembly. We used *racon (Vaser et al., 2017)* to improve the first draft genome, which results in a genome with fewer scaffolds. We used *pilon* (Walker et al., 2014) to correct base-calling errors in the *racon* assembly. We used *purge_haplotigs* to remove duplicated contigs inadvertently created in the previous steps of genome assembly. We used *BUSCO* (Simão et al., 2015) to evaluate the completeness of the genome by identifying conserved bird protein-coding genes in the assembly.

The pipeline to generate re-sequenced genomes mapped to the reference genome is as follows. We estimate the number of reads per sample using the following unix command:

```
echo $(gunzip -c Vireo10_LIB1_R2_trimmed.fastq.gz | wc -l)/4 | bc
```

Reads are trimmed with BBTools (Bushnell, 2017). We map the trimmed reads to the reference genome using BWA (Li and Durbin, 2009). We use GATK (Van der Auwera et al., 2013) to call variants in single samples and in the entire set of samples. We used vcftools (Danecek et al.,

2011) to calculate heterozygosity in population sets. We used PLINK (Purcell et al., 2007) to calculate the lengths and location of Runs of Homozygosity. We used STRUCTURE (Pritchard et al., 2000) to measure connectedness of populations. These pipelines can only be run when all samples have been sequenced.

V. RESULTS AND DISCUSSION:

We were not able to complete the reference genome. We also tried to sequence the remaining 24 samples in late 2019. In the first attempt, our libraries failed at the final quality control step. We redid the libraries in late December, and they passed quality control. Our intention was to send these repeated libraries to the NovaGene sequencing facility in California for sequencing, as part of a larger pool of sequenced samples to be run together. However, prior to the University of Oklahoma's COVID-19 shut down, we had not received the sole-source purchasing approval, and so we do not have the sequence data for the remaining samples.

The modification of our study design from ddRAD-seq to whole genome analysis required a larger budget than was contemplated in our 2018 proposal. As we discussed with ODWC, replacing ddRAD-Seq (which generates about 3,000–5,000 bases across the genome with whole genome data (which generates approximately 1 billion bases) would have given our study considerable power to best answer the original contemplated questions. However, we were unsuccessful in completing the project by the closing of the grant.

VI. SIGNIFICANT DEVIATIONS:

There were no significant deviations. However, we were unable to complete the project's original objectives by the closing of the grant.

VII. **Prepared by**: Dr. Matthew J. Miller
Sam Noble Oklahoma Museum of Natural History
University of Oklahoma

Date: 20 April 2020

Approved by: Matt Fullerton, Endangered Species Biologist
Russ Horton, Assistant Chief Wildlife Division
Oklahoma Department of Wildlife Conservation

Approved by: Andrea Crews, Federal Aid Coordinator
Oklahoma Department of Wildlife Conservation

REFERENCES

Athrey, G., R.F. Lance, and P.L. Leberg. 2012a. How far is too close? Restricted sex-biased dispersal in black-capped vireos. Molec. Ecol. 21:4359-4370.

Athrey, G., K.R. Barr, R.F. Lance, and P.L. Leberg. 2012b. Birds in space and time: genetic changes accompanying anthropogenic habitat fragmentation in the black-capped vireo. Evolutionary Applic. 5:540-552.

Barr, K.R., D.L. Lindsay, G. Athrey, R.L. Lance, T.J. Hayden, S.A. Tweddale, and P.L. Leberg. 2008. Population structure in an endangered songbird: maintenance of genetic differentiation despite high vagility and significant population recovery. Molec. Ecol. 17:3628-3639.

Bushnell, B., 2017. BBTools software package. URL http://sourceforge. net/projects/bbmap.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., 1000 Genomes Project Analysis Group, 2011. The variant call format and VCFtools. Bioinformatics 27, 2156–2158.

Fazio, V.F., III, D.B. Miles, and M.M. White. 2004. Genetic differentiation in the endangered Black-capped Vireo. Condor 106:377-385.

Graber, J.W. 1957. A bioecological study of the Black-capped Vireo (Vireo atricapillus). Ph.D. Diss., Univ. Oklahoma, Norman. 203p.

Grzybowski, J.A. 1992. Ecology and management of isolated populations of the Black-capped Vireo (Vireo atricapillus) in Oklahoma. Oklahoma Dept. Wildl. Conserv. Performance Rep., Project E-1-7. 16p.

Grzybowski, J.A. 2005. Reproductive enhancement and population monitoring of Black-capped Vireos in Blaine County, Oklahoma. Oklahoma Dept. Wildl. Conserv. Final Performance Rep., E-45-8.

Grzybowski, J.A. 2016. Searches for Black-capped Vireos on private lands in western Oklahoma, 2013-2016. Oklahoma Dept. Wildl. Conserv. Final Performance Rep., E-80-R-2, 33p.

Grzybowski, J.A., R.B. Clapp, and J.T. Marshall, Jr. 1986. History and population status of the Black-capped Vireo in Oklahoma. Am. Birds 40:1151-1161.

Kolmogorov, M., Yuan, J., Lin, Y., Pevzner, P.A., 2019. Assembly of long, error-prone reads using repeat graphs. Nat. Biotechnol. 37, 540–546.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754–1760.

Longmire, J.L., Maltbie, M., Baker, R.J., 1997. Use of "lysis buffer" in DNA isolation and its implication for museum collections. Museum of Texas Tech University.

McCormack, J. E., & Maley, J. M. 2015. Interpreting negative results with taxonomic and conservation implications: Another look at the distinctness of coastal California Gnatcatchers. Auk 132(2):380-388.

Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945–959.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212.

Smeds, L., Warmuth, V., Bolivar, P., Uebbing, S., Burri, R., Suh, A., Nater, A., Bureš, S., Garamszegi, L.Z., Hogner, S., Moreno, J., Qvarnström, A., Ružić, M., Sæther, S.-A., Sætre, G.-P., Török, J., Ellegren, H., 2015. Evolutionary analysis of the female-specific avian W chromosome. Nat. Commun. 6, 7330.

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K.V., Altshuler, D., Gabriel, S., DePristo, M.A., 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr. Protoc. Bioinformatics 43, 11.10.1–11.10.33.

Vaser, R., Sović, I., Nagarajan, N., Šikić, M., 2017. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 27, 737–746.

Vasquez-Miranda, H., K.R. Barr, C. C. Farquhar and R.M. Zink. 2015. Fluctuating fire regimes and their fluctuating effects on genetic variation in an endangered scrubland songbird. Ecol. and Evol. 5: 5487-5498.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., Earl, A.M., 2014. Pilon: an integrated tool for

comprehensive microbial variant detection and genome assembly improvement. PLoS One 9, e112963.

Zink R.M., A.W. Jones, C.C. Farquhar, M.C. Westberg, and J.I.G. Rojas. 2010. Comparison of molecular markers in the endangered Black-capped Vireo (Vireo atricapilla) and their interpretations in conservation. Auk 127: 797-806.