Golden-cheeked Warbler Population Genomics

A report on the current population genetic status of golden-cheeked warblers

Prepared By:
Giridhar Athrey, Ph.D., Texas A&M University
Laurel Moulton, Ph.D., The City of Austin
Lisa O'Donnell, The City of Austin

March 2023

This project was supported by the Department of Defense Legacy Resource Management Program

Contents

11 Appendix I I

1 Objectives of the Study

As of 2023, the endangered Golden-cheeked Warbler, hereafter GCWA) remains a priority species for conservation under the Endangered Species Act and the IUCN Red List. The species numbers across the range and on military installations have expanded since their historic lows, but the overall genetic diversity and connectivity with other populations on public and private lands are unclear. The species action plan (SAP) lists the identification of population structure and genetics as a critical assessment component. The last thorough genetic assessment was carried out over a decade ago. Therefore, a comprehensive population genetic assessment is critical for informing future species management. To address these needs, we performed a study with two primary objectives for this study:

Objective 1: To generate a de novo assembly of the GCWA genome using a combination of short-read and long-read sequencing technologies.
The absence of a reference genome for species such as the GCWA remains one of the barriers to assessing how genetic variation at few marker loci is representative of whole genome patterns and whether such patterns are functionally important for the survival and fitness of individuals. In a nutshell, we often need to associate genetic diversity data with fitness consequences for the bird. Secondly, the lack of standardized genomic resources makes it difficult to perform studies that can be compared over time due to changes in molecular technologies. The availability of a genome reference greatly improves our options for assessing variation at functionally important regions (such as those involved in immunity or fertility). Therefore, our first objective in this study was to generate a new reference genome assembly for the GCWA to enable whole-genome studies of GCWA for the present and future.

Objective 2: Assess the current population genetic diversity, structure, and demographic history across the range of the GCWA. As the previous study that characterized demographic history and population structure was published in 2011 [5], with the samples collected between 2006-2008, that dataset is nearly two decades old. Therefore, our study aimed to assess the extent of genetic diversity within and between populations and the degree of genetic differentiation among them, using the genome reference (Objective 1) as the foundational standard for such an analysis.

2 Executive Summary of Findings

1. Genome Assembly
    - We assembled the GCWA genome from a single individual captured at Balcones Canyonlands National Wildlife Refuge.
    - Used a Hybrid Assembly approach to build a new draft genome assembly for GCWA
    - The draft assembly has a total size 817Mbp with high levels of contiguity and completeness.
2. We performed Whole-Genome Sequencing (WGS) on 239 field-sampled GCWA from across 14 Texas counties, representing 11 sites.
    - We performed various population genetic analyses using open-source bioinformatics tools
3. We estimated genetic diversity using multiple approaches
    - GCWA has a nucleotide diversity π of 0.0014
    - GCWA populations show low levels of heterozygosity, averaging 0.03 across the species.
    - Populations show a high level of inbreeding, with a species average of 0.09 for inbreeding coefficients.
4. We assessed population genetic structure using multiple approaches
    - Pairwise F_{ST} values show high differentiation levels (0.008 to 0.02) among populations indicating reduced gene flow.
    - Hierarchical clustering and ordination analyses indicate a high degree of population structuring.
    - There was no significant isolation by distance.
5. We reconstructed ancient and recent demographic histories
    - Ancient (>10Kya) populations of GCWA were much larger than recent demographic or genetic effective sizes.
    - Recent demographic estimates of$N_e$show small genetic effective sizes across the range, suggesting recent severe bottlenecks.
6. We assessed genome-wide neutrality statistics
    - We found an excess of positive Tajima's D and Fu & Li's D values, suggesting recent bottlenecks with ongoing effects on genetic variation.
7. Overall, we found low genetic diversity and high population structuring, with ongoing effects of bottlenecks. The totality of genetic evidence does not indicate a recovery from population bottlenecks in the 20th century.

3 Study Design

The GCWA is an endangered songbird species that has long been the face of conservation in Texas. Over two decades of intensive monitoring and ecological studies have yielded actionable data for informing the demographic recovery of this species from its historic lows in the 1990s. Even as continued ecological and modeling approaches may indicate demographic recovery for some sites, data from the ecological studies do not always present a clear picture of the recovery and prescriptions for success [40, 3, 46, 29], or an accurate representation of range-wide trends. For example, populations may decline at sites with less contiguous habitat patches. A genetic survey will objectively measure local and range-wide population trends, but details about the genetic status or recovery were unclear before our project.

The most spatially comprehensive genetic assessment was performed over a decade ago [28]. The most temporally comprehensive genetic assessment was also performed a decade ago and indicated that contemporary populations had significantly reduced genetic diversity and increased genetic fragmentation compared to a century prior [5, 25]. Athrey et al. [5] compared genetic data from 100-year-old museum specimens against modern samples to generate this information. One of the more troubling results from this study was the low effective population size estimates - a central population parameter of much significance in conservation biology. The effective population size (Ne) describes the number of individuals contributing genetically to the following generations. These declines were so great that population genetic theory predicts several generations before the lost genetic diversity can be recovered to historic levels. These initial genetic studies provided much-needed genetic context to the discussion of GCWA conservation and successfully demonstrated the importance of such data in endangered species management plans.

There is a critical need to revisit the status of genetic diversity and population genetic structure and determine how the genetic attributes relate to demographic trends. In this project, we built on the template of past studies and utilized new genetic tools and technologies to help make such knowledge a standalone resource, as well as a benchmark for future monitoring efforts of the species.

Genetic monitoring at regular intervals has been demonstrably valuable in conserving and managing various species [43, 1]. While assessing genetic diversity at neutral molecular markers and estimating population size and structure remain priorities, the evolution of affordable next-generation sequencing technology has opened avenues for evaluating adaptive genetic variation in a conservation context [51, 4]. Furthermore, a consensus is developing in the scientific community supporting the integration of adaptive potential into U.S. Endangered Species Act decisions [14]. Adaptive potential and adaptive genetic variation (such as loci important in survival and fitness traits) are important considerations for species that have experienced past or ongoing population declines. Fortunately, assessing these emergent necessities is increasingly feasible due to the low-cost sequencing approaches available [14, 2]. In this study, we made these types of assessments available for GCWA.

3.1 Population Sampling

We sample GCWAs from across the current breeding distribution of the species. We commenced sampling in 2019. While our original plan was to complete sampling in 2020, we were delayed by a year due to COVID. We completed our sampling in spring 2021. In addition to the samples collected for this project, we also included samples previously collected for other projects. Therefore, we collected 282 samples (20 females, 262 males) across the GCWA breeding range from 2018-2021. The samples were collected across 14 counties at Palo Pinto State Park, Fossil Rim Wildlife Center, Meridian State Park, Fort Hood, Colorado Bend State Park, Balcones Canyonlands National Wildlife Refuge, Balcones Canyonlands Preserve, Camp Bullis, Government Canyon State Natural Area, Guadalupe River State Park, Garner State Park, Love Creek Preserve, Kickapoo Caverns State Park, and from private ranches in Somervell, Uvalde, Bandera, Kinney, Real, and Edwards counties (Figure 1). Samples from these 14 counties represented 11 sites based on their proximity (eg. Bell/Coryell).

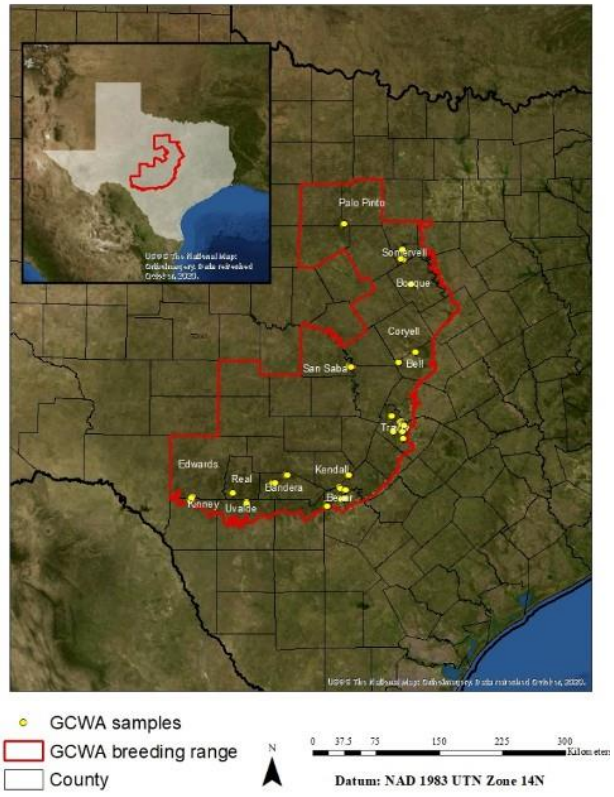

Figure 1: A map of the breeding range of GCWA, with points denoting the sampling locations for this study.

We used conspecific audio playback to target specific individuals for capture and attraction to the mist net during the species breeding season (March 5 to May 30). We banded each bird with a USGS silver band to avoid re-sampling the same individual. We collected blood samples using a PrecisionGlide™ 30-gauge beveled needle to prick the brachial vein and a capillary tube to collect 15-25μL of blood (2-3 drops). We stored the blood samples in RNAlater™ in a refrigerator (~40^{0}F) until DNA extraction.

Of the 282 total samples collected, 238 were included in the genomic analysis, as the remainder did not yield sufficient quantity/quality DNA for inclusion in genome sequencing. More details about these samples are provided below, and a complete listing of samples can be found in Appendix I. This research was conducted under federal bird banding permit 23615, federal scientific permit TE59231C-2, state scientific research permit SPR-0219-028, and state park scientific study permit 2019-R5-01.

4 Genome Assembly and Sequencing of Population Samples

4.1 Background

Table 1: A summary of the sampled counties and the number of GCWA samples collected at each site.

Sites/Counties	Sampled
Bandera	27
Bexar	45
Bosque	15
Coryell/Bell	23
Kendall	27
Kinney/Edwards	15
Palo Pinto	16
San Saba	20
Somervell	19
Travis	50
Uvalde/Real	25
Total	282

Genome assembly is the process of reconstructing an organism's complete genomic sequence from fragmented DNA sequences [34]. Using advanced computational algorithms, these smaller sequences, called reads, are aligned and merged into longer contiguous sequences known as contigs [20]. The quality of an assembly is evaluated based on parameters such as contiguity, completeness, and accuracy, which are assessed through metrics like N50, the proportion of conserved genes, and the rate of misassemblies. Genome assemblies play a crucial role in population genetic studies by providing a reference for variant identification, discovering population-specific adaptations, and understanding the genetic basis of complex traits. By comparing genomes across diverse populations, scientists can unveil the history of species, track migration patterns, and uncover the drivers of genetic variation [54].

Genome assemblies are valuable tools for assessing genetic variation, estimating population size, and measuring gene flow, which are crucial aspects of conservation genetics. By comparing assembled genomes from different individuals within a population, scientists can identify single nucleotide polymorphisms (SNPs), insertions/deletions (INDELS), and structural variations. These variations can provide insight into the genetic diversity and differentiation of populations.

In applying genetics to inform the management of species, especially endangered species, understanding genetic variation, population size, and gene flow is vital for implementing effective strategies to conserve endangered species and maintain biodiversity. Genome assemblies can help identify unique populations, prioritize conservation efforts, and guide future management and species action plans [42, 14].

4.2 Description of Methods

4.2.1 Genome assembly

A hybrid assembly is an approach in genome assembly that utilizes multiple types of sequencing technologies to generate a high-quality assembly [23]. One popular hybrid assembly approach combines long-read sequencing data from platforms like Oxford Nanopore Technologies (ONT) with short-read sequencing data from platforms like Illumina. This approach leverages each technology's strengths to overcome weaknesses and generate an accurate and contiguous assembly [30, 49].

The hybrid assembly process typically involves three main steps: error correction of the long reads using the short reads, assembly of the corrected long reads, and polishing the assembly with the short reads. The corrected long reads serve as a backbone for the assembly and help bridge gaps in the short-read assembly, while the short reads provide high accuracy and help correct errors in the long-read assembly.

Haslr is a software tool that implements this hybrid assembly approach by first generating an assembly with the long reads and then using the short reads to correct errors and polish the assembly [16]. This tool has been shown to produce highly contiguous and accurate assemblies, with contig N50 values up to 5.5 Mb and error rates as low as 0.01% [23].

The benefits of hybrid assembly with haslr include:

- Improved assembly continuity and accuracy: Hybrid assembly using long and short reads can help overcome the challenges of each technology and generate a highly contiguous and accurate assembly [15].
- Resolving complex regions: Hybrid assembly can help resolve complex regions of the genome that are difficult to assemble with short reads alone, such as repetitive regions and structural variations [32].
- Cost-effective: Hybrid assembly is more cost-effective than long-read sequencing alone and can generate high-quality assemblies with relatively small short-read sequencing data [23].

Overall, hybrid assembly is a robust approach for generating high-quality genome assemblies that are more complete and accurate than assemblies generated with a single sequencing technology.

We sequenced a single individual male (ASY) captured at the Balcones Canyonlands National Wildlife Refuge for the purpose of genome assembly in May 2019. The Texas Institute for Genome Sciences and Society (TIGSS) on the Texas A&M University campus performed the library preparation and sequencing for the genome assembly. Following sequencing, the raw data (.fastq) was checked for adapter content and quality filtered for a minimum base quality score of Q28. This quality is high enough for assembly purposes, as most assembly algorithms need the ability to separate base pair variants from sequencing errors during assembly. The quality-filtered short-read data was used for the hybrid assembly generation step with Haslr.

4.2.2 Whole genome resequencing of population samples

We submitted the whole blood samples collected from the field for full-service genome sequencing at the TIGSS lab. This included DNA isolation, library preparation for whole genome sequencing, and sequencing of the samples. Forty-three samples did not meet the quality and quantity criteria to generate libraries for sequencing and did not make it past the quality control step for sequencing. The remainder of the samples were sequenced and delivered to the Athrey lab for further analysis.

While the sequencing data was of very high quality, with most libraries showing average quality scores above Q35 (<1 base error per 1000 observations for the base position), we

followed the best practices for sequence analyses. The raw sequence data were first quality filtered to trim bases with quality scores <Q30 using the tool trim galore [19]. Following this, the data were aligned to the draft GCWA assembly using the 'bwa mem' algorithm for paired-end data [26]. The aligned reads were then used for subsequent analyses workflows for population genetic analyses (details below).

4.3 Results

4.3.1 Assembly Description and Reference Free Metrics

We generated Oxford Nanopore long reads and Illumina short reads (151bp paired-end sequencing) on the HiSeq platform. These runs yielded a total of 9.8M Nanopore reads, with a total throughput of approximately 12x genome coverage. We generated 714M paired-end reads with Illumina (approximately 70x depth of sequencing). The depth of sequencing estimates was based on the assumption of 1Gbp genome size, which is conserved in birds. Our assembly strategy required us to generate a long-read assembly first, followed by gap-filling, base corrections, polishing, and scaffolding using the short-read data. The hybrid Haslr assembly generated an 817.25Mbp long assembly with 2450 contigs greater than 500bp in length. The longest contig in the assembly is 12.6Mbp, with an N50 value of 1.518Mbp. The N50 value is the contig size that describes 50% of the assembled contigs. Therefore, 50% of the assembled contigs in this assembly are 1.51Mbp in length or greater. The genome has a GC (guanine-cytosine) content of 41.52%, and 1195 of the contigs were over 50,000bp in length. These numbers compare favorably to other draft genome assemblies, including those for the songbirds sequenced recently.

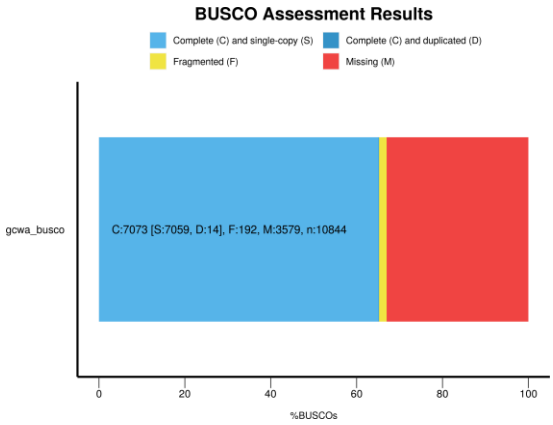

Figure 2: A graphic showing the BUSCO analysis result for the GCWA draft assembly. The results showed that we successfully reconstructed

Another approach to understanding assembly completeness is characterizing the recovery of known orthologs from the assembly. We used the Benchmarking Universal Single-Copy Orthologs (BUSCO) approach [45]to assess the assembly contiguity and completeness using the AvesDB ortholog database. This analysis showed that we recovered about 70% of known avian orthologs that are complete, with 30% of orthologs that are missing from the assembly. This is typical for a draft genome assembly and shows that the genome assembly can be further improved. For comparison, the chicken genome assembly has reached a high level of completeness (>95%) twenty years after it was first sequenced [18].

4.3.2 Whole Genome Sequencing (WGS)

Of the 282 GCWA blood samples collected, 239 yielded sufficient DNA quantity to proceed with next-generation sequencing on the Illumina NovaSeq platform. For each sample, we

generated an average of 15x sequencing coverage (i.e., every base pair was observed on average 15 times). The alignment step produced genome alignments, with an average (across 239 samples) alignment rate of 84%. This number shows a high degree of alignment to the draft assembly, which is sufficient for further analyses, requiring variant discovery, genotyping, and determination of haplotypes. We performed these analyses with well-established software tools.

5 Measures of Genetic Diversity

We assessed genome-wide measures of heterozygosity and the inbreeding coefficient (F_{IS}) to determine contemporary measures of genetic variation in these populations. Heterozygosity or gene diversity values are informative about the frequency of heterozygotes, which directly indicate how diverse a population is. In this case, we used whole-genome estimates of heterozygosity, which estimate the occurrence of heterozygous genotypes at every variable locus across the genome. Genome-wide heterozygosity values (<0.1) suggest lower levels of genetic variation. Another critical measure of a population's diversity is the inbreeding coefficient, which estimates non-random mating among relatives within populations. Positive F_{ST}values indicate an excess of homozygosity (reduced genetic diversity) due to inbreeding, genetic drift, or founder effects, which can result in negative fitness consequences such as decreased disease resistance and reproductive success [17]. Negative F_{IS} values indicate an excess of heterozygotes. Inbreeding can also indicate low gene flow between populations, leading to genetic differentiation and reduced potential for local adaptation [55].

5.1 Bioinformatics Methods

We used the open-source software tool ANGSD (Analysis Next Generation Sequencing Data) [24] for most of our genome data analyses. The ANGSD tool (citation) is a widely used software package for analyzing next-generation sequencing data in population genetics studies. ANGSD can generate genotype likelihoods from sequencing data, which are used to estimate site frequency spectra (SFS), genetic diversity estimates, and pairwise F_{ST} values. To generate SFS, ANGSD considers sequencing errors, sample contamination, and other noise sources in the data. The estimated SFS can then be used to calculate pairwise F_{ST} values between populations, which measures the degree of genetic differentiation between populations. We used only sequencing reads with high base quality (Phred score $\geq$ 30) and Mapping Quality (minMap $\geq$ 40) to ensure that only accurate alignments were used for generating genotype likelihoods. This is especially important, considering we used a draft genome assembly here.

An example code snippet used for generating genotype likelihoods is reproduced below.

```
$angsd -bam ''$homedir/$population.list'' \
-GL 1 -doGlf 2 -doMajorMinor 1 -doMaf 1 -doSaf 1\
-out ''$SAMPLE_OUT_DIR/$population'' \
-P 5 -SNP_pval 1e-6 -C 50 -minMapQ 40 -minQ 30 \
-ref ''$gcref''
```

Table 2: Summary of the sample sizes per population, the population-wise inbreeding coefficients (Fi), and the observed heterozygosity (Ho). Genome-wide averages are presented by population, and a final species average is presented.

Population	Sample Size (N)	Avg. F_{IS}	Avg. Ho
Bandera	25	0.097	0.025
Bell/Coryell	23	0.095	0.033
Bexar	42	0.075	0.022
Bosque	15	0.114	0.040
Kendall	25	0.070	0.024
Kinney Edwards	15	0.112	0.041
Palo Pinto	16	0.119	0.034
San Saba	19	0.101	0.035
Somervell	18	0.084	0.030
Travis	15	0.124	0.035
Uvalde/Real	25	0.096	0.025
Total	239	0.099 (Avg.)	0.031 (Avg.)

To estimate the inbreeding coefficient (F_{IS}), we ran ANGSD separately to assess deviations from Hardy-Weinberg Equilibrium and report values for F_{IS}.

```
$angsd -bam test.list -HWE_pval_F 1  -GL 1 -doMaf 1\
 -SNP_pval 1e-6 -minQ 30 -minMapQ 40 -doMajorMinor 1 -P 30 -out
```

We estimated nucleotide diversity based on the alignment of an individual high-coverage sample to the reference genome and variant calling with bcftools mpileup, followed by statistics estimation using bcftools [9].

5.2 Results

5.2.1 Genetic Diversity

We found that the genome-wide nucleotide diversity, π, was estimated as 0.0014, within the range of values seen in other avian species, such as Ficedula flycatchers (0.0039) [33, 10] and hooded crows (0.0011) [56]. There are no universally accepted thresholds for π, as it can vary across the genome due to various evolutionary factors. However, a genome-wide average value of 0.0014 is on the lower end of the spectrum. For example, recent work on Tasmanian Silvereye shows an average $\pi > 0.11$ [44].

Next, we estimated Observed Heterozygosity (Ho) using the SFS data generated from ANGSD, using the method described in the ANGSD documentation. The western populations (Uvalde, Bandera, and Kendall) were unique in showing the lowest Ho values among the 11 populations. These low values of Ho suggest that the populations are likely still recovering from past population declines. These values showed that the average heterozygosity across all populations was 0.031, ranging from 0.02 to 0.04. These values are also much lower

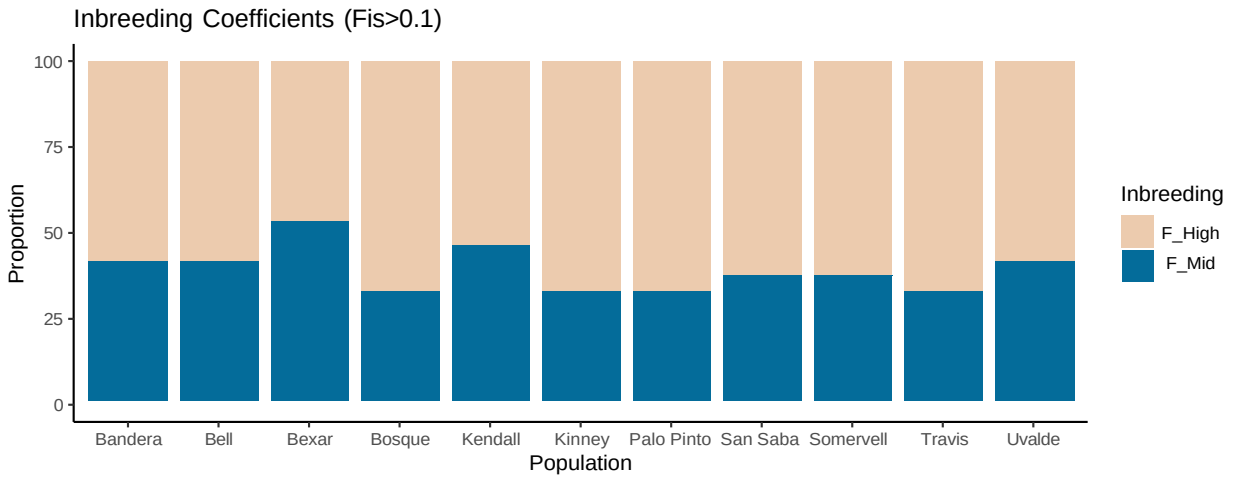

Figure 3: Relative proportions of inbreeding coefficients across whole genome regions in GCWA populations. NB: Only proportions of F_{IS} values above 0.1 are plotted.

than usually observed in widespread species, such as the seaside sparrows [11], with an average Ho of 0.05 or higher. In the Tasmanian Silvereye study [44], species-wide heterozygosity values were >0.10. As heterozygosity increases gradually in a population after bottlenecks, it may take tens of generations before new alleles and random mating restore heterozygosity to pre-decline levels.

5.2.2 Inbreeding

Our WGS analyses revealed that the eleven populations showed consistently positive F_{IS} values, with a significant proportion of values above 0.25 and a species average of 0.09. These values are much higher than those reported in other birds. For example, F_{IS} values of 0.02 in wild Superb Parrots [47] and 0.066 in Island Scrub Jays [7]. This suggests that inbreeding, due to a prolonged history of low population sizes, has resulted in losses of genetic diversity in GCWA populations and has not recovered from their bottleneck events. Positive values of F_{IS}, especially thousands of loci with values >0.25, indicate high homozygosity levels, which can be traced back to common ancestry due to a historical bottleneck event. Furthermore, the populations have distinct distributions of F_{IS} values (Figure 3). These differences were statistically significant (based on contingency analyses, P<1e-04). This result suggests two things: 1) the populations have distinct recent demographic histories with different levels of inbreeding, and b) the inbreeding and low heterozygosity values across populations will limit the signature of gene flow or differentiation due to the limited genetic variation within populations.

Possible causes for high inbreeding coefficients include habitat fragmentation, and geographic isolation, which can disrupt gene flow [1].

6 Population Genetic Structure

6.1 Estimates of population genetic structure

The fixation index, F_{ST}, is a widely used statistic in population genetics to quantify genetic differentiation among populations. Classical studies have used pairwise F_{ST} values to infer population structure and evolutionary history. For example, F_{ST} estimates have been used to investigate the genetic diversity and population structure of various species [35, 6, 8, 21]. Recently, more advanced methods have been developed to estimate F_{ST} values from whole-genome data, such as the Weir and Cockerham estimator and the Hudson estimator.

The range of F_{ST} values can vary widely depending on the degree of population differentiation. F_{ST} values can range from 0, indicating no differentiation between populations, to 1, indicating complete differentiation. Significant F_{ST} values suggest genetic differentiation between populations may be due to geographic isolation or genetic drift.

Population structure can be inferred from genetic data using various analytical methods, including hierarchical clustering and principal component analysis (PCA) based on genetic distances. Hierarchical clustering can group individuals into clusters based on their genetic similarity, each representing a distinct population. PCA can also be used to visualize the genetic variation among individuals, with each principal component representing a different source of variation.

Besides human population studies [50], hierarchical clustering and PCA have also been used to study the genetic structure of various other species, including crops [38], livestock [31], and wildlife [14]. Hierarchical clustering and PCA are powerful tools for understanding population structure and genetic variation based on genetic distances. These methods have been widely used in population genetics studies and can provide insights into populations' evolutionary history and genetic diversity across various species.

6.2 Description of Methods

We first used the population-wise genotype likelihood to calculate the folded SFS (-fold 1 option) and then calculated the pairwise F_{ST} values using the 'realSFS fst' option in ANGSD.

6.3 Results

6.3.1 Pairwise measures of differentiation

Pairwise measures of F_{ST} showed relatively high levels of F_{ST} across all pairs of the sampled population, ranging from 0.8% to just over 2% (Figure 4).

These values are quite high, considering the relatively short distances spanning the entire range of the GCWA (<300Km between any two points). Among the sampled locations, samples from Palo Pinto were notable for showing the greatest levels of differentiation, even between other sites short distances away (such as Somervell, San Saba, and Bell/Coryell). Furthermore, these values are higher than those in other highly mobile species. Geraldes et al.

We also used pairwise genetic distances to characterize the clustering of populations using both hierarchical clustering and ordination analyses, and figures 5 and 6 show these results.

	bellcoryell	bexar	bosque	kendall	kinneyedwa	palopinto	sansaba	somervell	travis	uvalde
bandera	0.011533	0.008467	0.01347	0.011498	0.012311	0.018657	0.013071	0.014068	0.013174	0.009711
bellcoryell		0.009931	0.013541	0.012485	0.014223	0.019261	0.013647	0.014064	0.014116	0.011708
bexar			0.012681	0.009546	0.011795	0.017957	0.011622	0.013032	0.012009	0.008513
bosque				0.014256	0.015239	0.019498	0.01474	0.014396	0.015427	0.01398
kendall					0.013639	0.019651	0.013336	0.014099	0.014147	0.011492
kinneyedwards						0.020535	0.015435	0.015709	0.015215	0.012226
palopinto							0.019875	0.020342	0.020711	0.018936
sansaba								0.01443	0.015119	0.013225
somervell									0.014837	0.014478
travis										0.013552
uvalde										

Figure 4: Table of Pairwise F_{ST} values among sampled populations.

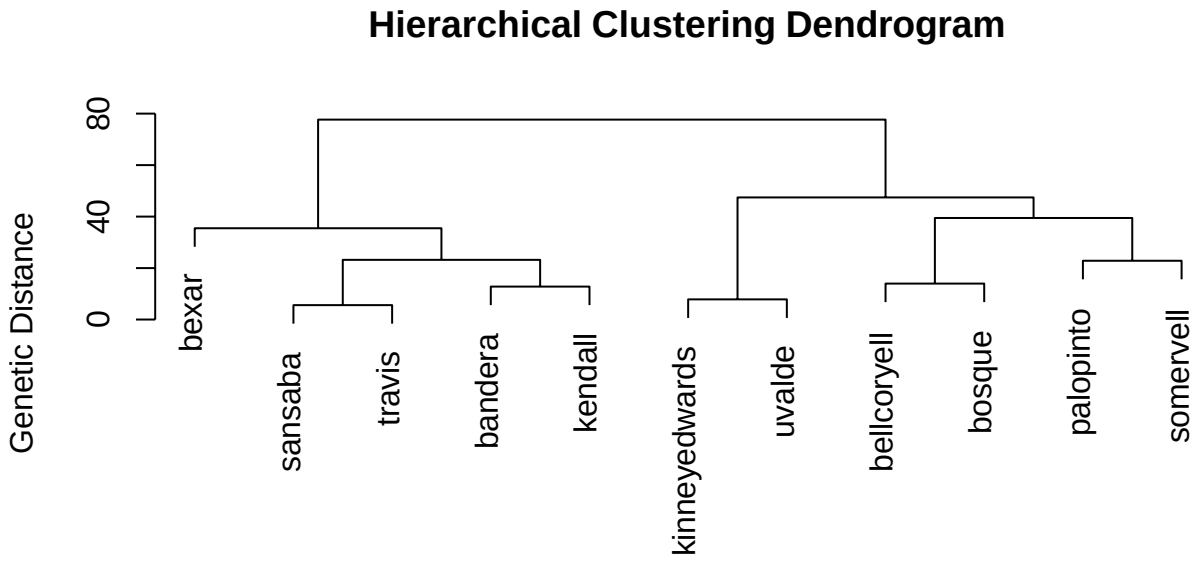

Figure 5: Hierarchical clustering dendrogram showing the grouping of populations based on genetic similarity.

Both of these show some expected and unexpected patterns. For example, based on hierarchical clustering (Figure 5), the Bexar population is the least like other populations (based on genetic distance and branching pattern). Among the remaining population, the order of branching shows that Travis clustered with San Saba. In contrast, Bandera clustered with Kendall, with these dyads showing more similarity to each other than other populations. Uvalde and Kinney/Edwards (the Western populations) clustered together, whereas the northern populations (Bell-Coryell, Palo Pinto, Bosque, and Somervell) formed a subcluster. In contrast, the ordination analysis (Figure 6) shows little similarity between most populations, with notable overlapping populations being 1) Uvalde/Real and Kinney/Edwards, 2) Travis and San Saba, and 3) Bell/Coryell and Bosque, and 4) to a limited extent Bandera and Kendall. Notable is that Bexar, Palo Pinto, and Somervell are most distinct from the other populations.

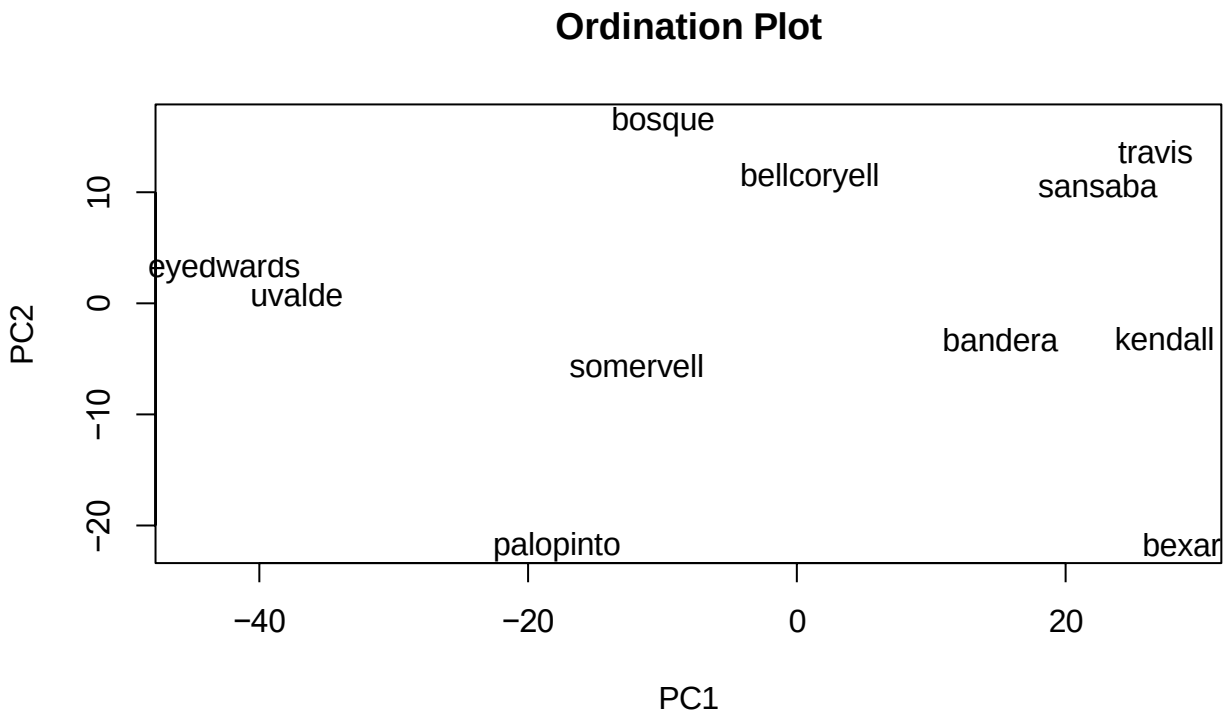

Figure 6: PCA plot showing the relative overall patterns and similarity of GCWA populations. NB, the name 'KinneyEdwards' is partially cut off in the graphic.

6.3.2 Isolation by Distance

Isolation by distance (IBD) is a pattern commonly observed in population genetics, where the genetic similarity between individuals or populations decreases as geographic distance increases. This pattern can be attributed to limited gene flow, genetic drift, and local adaptation. Isolation by distance has been observed in various species [37, 53].

We tested for IBD using Mantel's correlation test between the pairwise genetic distance (F_{ST}/(1-F_{ST}) matrix against the pairwise geographic distance matrix. This dataset had no significant IBD (Figure 7), and the regression had a positive (non-significant) slope (0.28).

Reduced genetic variation and increased inbreeding can significantly impact the pattern of isolation by distance across populations. Inbreeding reduces genetic diversity within populations and can lead to increased genetic differentiation between populations, contributing to stronger isolation by distance [41]. On the other hand, reduced genetic variation within and between populations can also weaken the pattern of isolation by distance, as there may be fewer genetic differences to be affected by geographic distance [58, ?, 41].

7 Demographic History

7.1 Background

Genetic data can reveal a population's demographic history by analyzing patterns of genetic variation and using computational methods to reconstruct historical scenarios. We can esti-

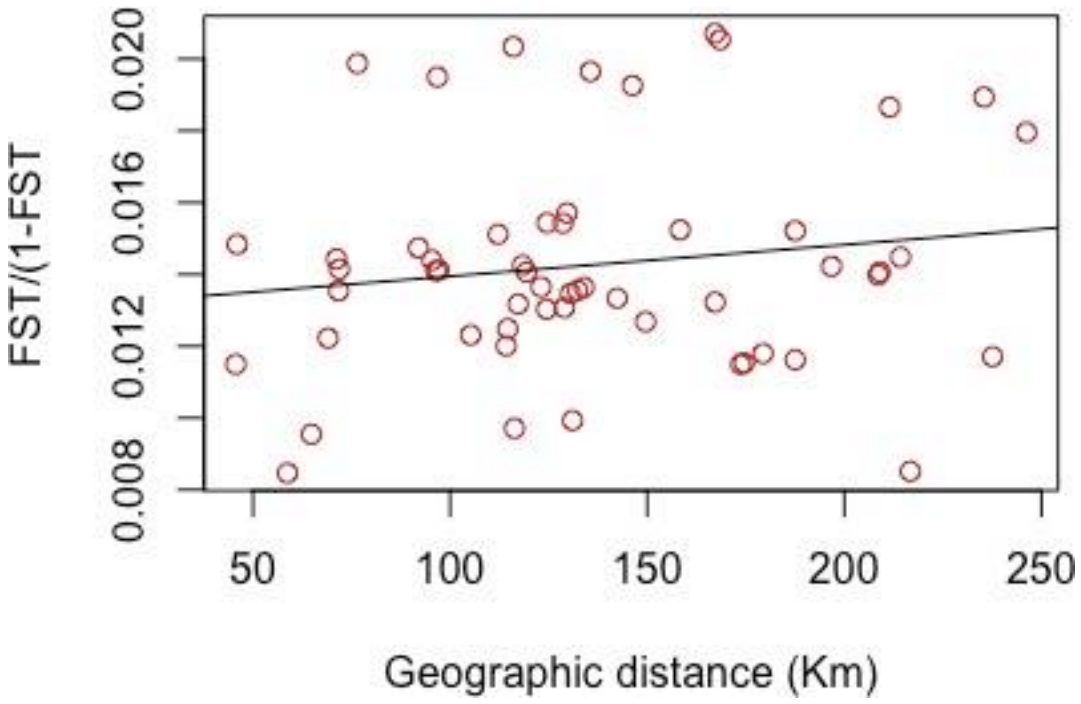

Figure 7: Isolation By Distance plot. We did not find significant isolation by distance in this study.

mate critical parameters such as effective population size and migration rates by comparing the observed genetic data to the expected patterns under different demographic models. The effective population size (Ne) is the most critical parameter for population demographic history. Effective population size (Ne) is defined as the size of the ideal population that experiences genetic drift at the same rate as the census population size [57]. These analyses can provide insights into the historical events that shaped the genetic diversity of a population, such as population bottlenecks, range expansions, or changes in migration patterns. Accurate inference of demographic history requires careful consideration of factors such as sample size, selection, and the complex interactions between demographic and evolutionary processes [12].

Furthermore, from the perspective of management decisions, it would be beneficial to understand both recent and ancient trends of population size histories. Genetic datasets allow us to query both these timeframes to illuminate population size histories of ancient or contemporary populations.

7.2 Description of Methods

Estimating the$N_e$using genetic data is a well-used and documented process, and several approaches are available to estimate Ne. While many approaches rely on understanding variance in allele frequencies to reveal the recent population size history, other methods, particularly coalescent-based methods, explore ancient demographic trends. Methods based on allele frequencies apply to the former, whereas the Pairwise Sequential Markovian Co-

alescent (PSMC) [27] is a method that fits the latter category.$N_e$estimators calculate the long-term average size of the breeding population over the last few generations. PSMC infers the effective population size changes over time based on the distribution of coalescent times between pairs of sequences, typically thousands of generations ago.

Ne estimates may be influenced by population structure and non-random mating [27, 59]. Various methods exist to estimate$N_e$from single samples, including the Heterozygote Excess method [39], Nomura's Coancestry method [36], and Waples and Do's Linkage disequilibrium-based method [52] from multilocus genotype data. We used these three approaches to generate $N_e$estimates per population. The$N_e$estimations used ten independently sampled datasets of 5000 randomly selected loci without missing data from each dataset. In contrast, the PSMC method used a whole genome alignment to the reference assembly. It is important to note that the results of the PSMC analyses, due it the deep time it queries, apply to the species as a whole. In contrast, contemporary$N_e$estimators query the recent demographic histories of individual populations.

7.3 Results

7.3.1 Ancient Demographic history

Analysis of a single whole genome (sampled at Balcones Canyonlands Preserve) was used for the PSMC analyses. The data were aligned to the draft Haslr assembly using the 'bwa mem' algorithm, and then the resulting alignment (bam file) was used as input for the PSMC analysis steps. We assumed a generation time of 1 year for the species. The results show (Figure 8) that between 1Mya to 100kya, the population was stable at relatively large population sizes of $2x10^4$. Then, between 100kya to 10kya, the species experienced a steady expansion in population size, reaching between 8-$9x10^4$. The 10000 years before the present is the typical boundary for PSMC analyses, as the analyses are based on the coalescence of individual alleles in the genome to a single ancestor. Therefore, this approach does not illuminate more recent demographic trends, and we need to rely on other approaches to understand more recent timescales, such as those presented below.

7.3.2 Effective Population Sizes

Given the recorded history of the GCWA leading to its listing as an endangered species, we know that the number of breeding pairs had reached a low point in the late 20th century. While intensive management and habitat protection has helped expand GCWA populations, continuing efforts must consider the genetic impacts and consequences of the severe bottleneck experienced in the 20th century. In their 2011 paper, Athrey et al., [5] used temporal samples spanning a century to show that the species had significantly reduced genetic diversity and low $N_e$estimates. Ten years hence, the present study evaluated these questions using whole genome data.

In this study, we found that based on three different estimators, each population shows relatively low $N_e$estimates (Table 3), on par with the results from the 2011 study and magnitudes smaller than the ancient population sizes for this species. Excepting samples from

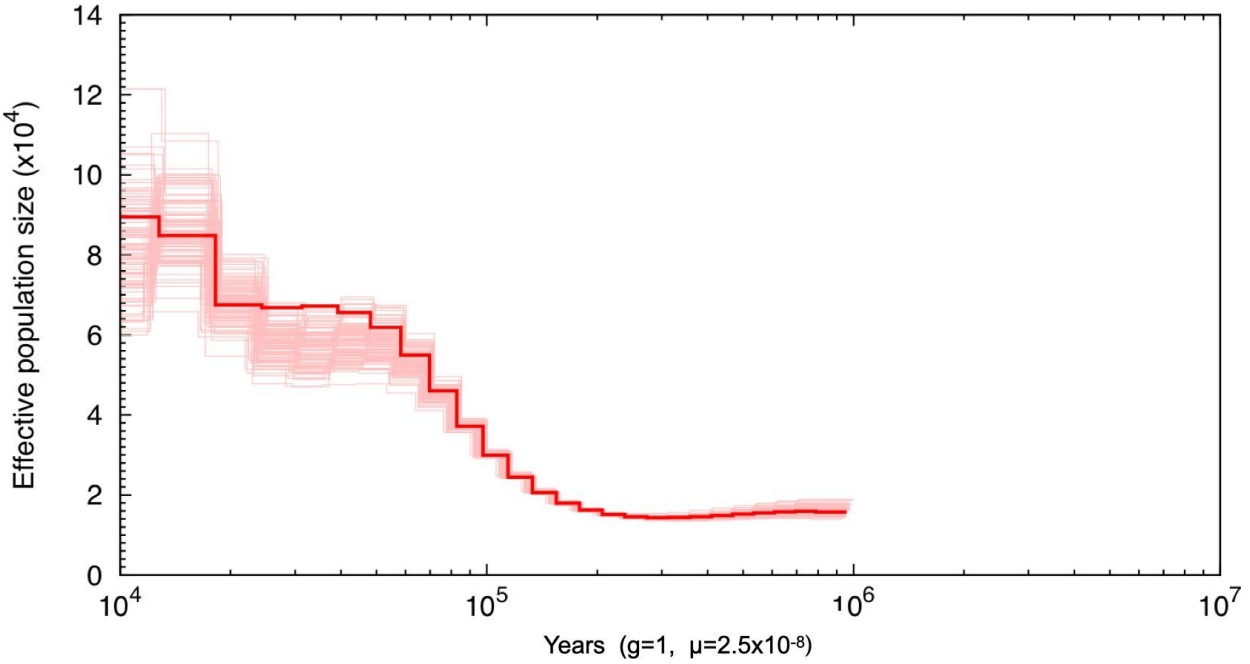

Figure 8: The demographic and effective population size history of GCWAs revealed by PSMC analyses.

Bexar county, all the other locations show N_e estimates of 150 or less (based on the coancestry method). The LD and Heterozygosity based methods were all much lower. Furthermore, the heterozygosity-based estimates were the lowest of the three estimates, suggesting that the population has experienced/continues to experience the effects of a prolonged reduction in population size. As inbreeding accumulates relatively slowly, compared to the rate at which alleles are lost due to bottlenecks and genetic drift, it is reasonable to interpret the low heterozygosity, high inbreeding, and low N_e values as, together, suggesting prolonged periods of small effective population size. Furthermore, comparing these estimates to the temporal N_e estimates (based on multiple time points) shows that these genetic population sizes do not show any signs of expansion; this pattern is not surprising given the inability of populations to recover quickly from severe bottlenecks. It may take tens or hundreds of generations for new variation to emerge within the species that can counteract the effects of inbreeding, *if* no new or additional factors act to reduce population size or gene flow.

8 Signals of selection

8.1 Neutrality Tests

Tajima's D [48] and Fu and Li's D [13] are neutrality statistics that are used as a way of testing the neutral theory of molecular evolution [22]. Tajima's D compares observed versus expected nucleotide diversity, assuming a constant population size and that polymorphisms are selectively neutral. The statistic ranges between negative and positive, with positive values indicating higher levels of common variation and low values indicating higher levels

Table 3: Summary of the estimates of effective population sizes using three different methods in this study. When available, the numbers from the temporal estimates (MLNe) generated in the 2011 study are reported alongside. 95% confidence intervals are shown in parentheses when available.

Population	2011 study	N_e (Coancestry)	N_e (LD)	N_e (Het)
Bandera		152 (108-191)	78 (29-103)	5.9
Bell Coryell		140 (73-178)	72 (34-97)	4.8
Bexar	49 (24-212)	254 (159-361)	130 (73-161)	6.1
Bosque		82 (41-104)	46 (31-53)	4.3
Kendall		156 (110-181)	80 (54-201)	3.9
Kinney Edwards		92 (76-123)	47 (24-77)	2.17
Palo Pinto		98 (66-134)	51 (19-81)	7.75
San Saba		116 (91-133)	59 (41-73)	5.17
Somervell		110 (101-145)	52 (38-89)	4.7
Travis	273 (46-621)	89 (53-99)	49 (23-108)	3.47
Uvalde/Real		110 (93-128)	70 (49-103)	4.8

of rare variation in a given region.

An excess of common variation can indicate balancing selection or recent population contraction, while an excess of rare variation can indicate directional selection or recent population expansion. To distinguish between natural selection and demographic change, many unrelated loci must be sampled. Fu and Li's D measures the number of singleton mutations (the number of individuals within a population with a novel and unique mutation) and compares the difference between these singleton mutations and the total number of mutations in the population. When the statistic is strongly negative, this indicates an excess of singleton mutations in the population. This can signal either a selective sweep where strong directional selection resulted in an overrepresentation of a specific mutation or a rapid population growth event where individuals are closely related, and mutation rates have not yet caught up. When the statistic is strongly positive, this indicates an excess of ancestral variants selected in the past and few unique variants. This can occur either due to balancing selection or a demographic bottleneck event.

8.2 Description of Methods

As with other analyses, we used the ANGSD package to estimate population-scaled mutation rates θ and neutrality test statistics. We started with the population-wise genotype likelihoods, from which we calculated the site frequency spectra (SFS). The folded SFS was then used to estimate the neutrality statistics using the 'thetaStat' function in ANGSD. An example of the code used for this is reproduced below.

```
$angsdir/misc/realSFS $population".saf.idx" -P 10 -fold 1 > $population".sfs"
$angsdir/misc/thetaStat  do_stat  $population".thetas.idx"
```

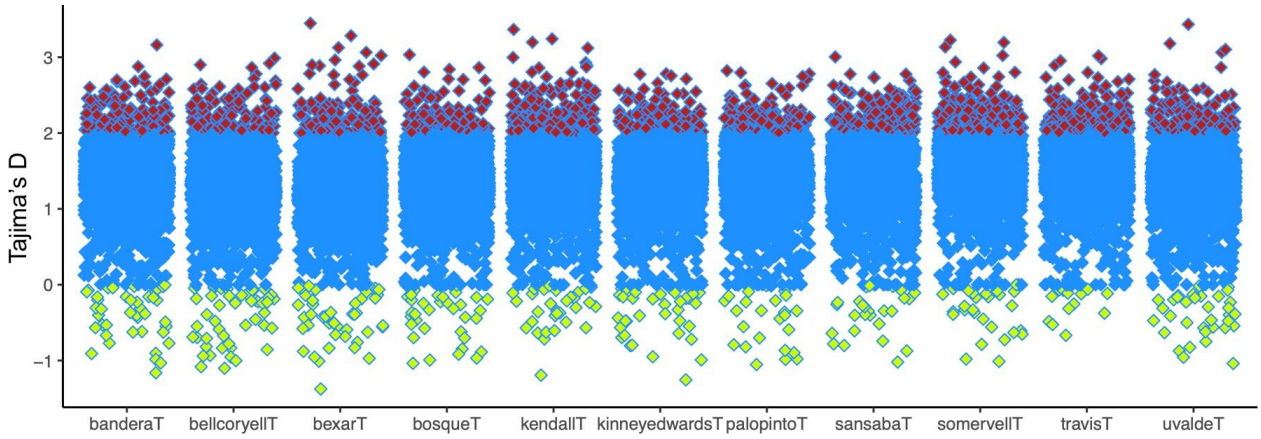

Figure 9: Side-by-side comparison of Tajima's D estimates among GCWA populations. Y-axis shows the Tajima's D values. The positive values of Tajima's D suggest balancing selection or a demographic bottleneck. In contrast, negative values suggest directional selection or population expansion Sites with estimates ≥2 are highlighted in red to show extreme values, and negative values are shown in green.

8.3 Results

All eleven GCWA populations show excess positive values for Tajima's D and Fu and Li's D statistics (Figures 9 and 10), with Bexar, Kendall, Somervell, and Uvalde/Real counties showing the highest values and suggesting acute levels of recent population contraction. Both these estimators analyze different but dependent components of genetic variation in the same population. Therefore, the concordance between these two differing metrics for mutational processes in populations suggests common factors influencing these genetic parameters. Viewing these measures with other information, such as heterozygosity, F_{IS} values, and N_e estimates, is essential to determine whether balancing selection or population decline is explanatory.

9 Inferences

The totality of evidence generated and considered here leaves little doubt that genetic population size and genetic diversity estimates remain small for the GCWA. The population differentiation data show that the differentiation among GCWA populations is high for a species with such a limited regional distribution and without major geophysical barriers to movement. Furthermore, the high inbreeding coefficients provide insights into the extent and duration of these population declines. Inbreeding accumulates rather slowly in populations, as it takes several generations before non-random mating between relatives increases; the estimates observed here point to small effective sizes over several generations.

Reversing these trends may take much longer than we might expect. We can use the Wright-Fisher model to estimate the time (in generations) needed for a population to recover from such levels of inbreeding (from 9% to 2%), accounting for the current average

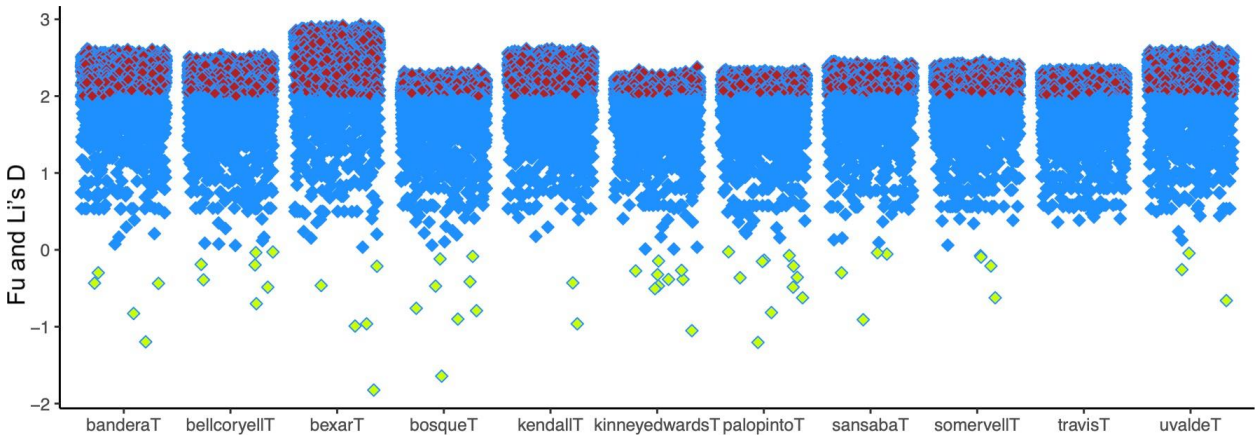

Figure 10: Comparison of Fu and Li's D statistic values among GCWA populations. Fu and Li's D values are on the Y-axis. The positive values of Fu & Li's D suggest balancing selection or a demographic bottleneck. In contrast, negative values suggest directional selection or population expansion Sites with estimates ≥2 are highlighted in red to show extreme values, and negative values are shown in green.

heterozygosity of 0.03 and an average effective size of 150. This can be calculated using the formula $T \approx (\ln(0.5) / \ln(1 - F)) / (2N_e \cdot H)$, Where T is the time in generations, N_e is the effective size, H is the observed heterozygosity, and F is the inbreeding coefficient. Using the values generated in this study, we can estimate the number of generations required, T, to be about 104 generations. Based on the timing of the bottleneck events in the 20th century, we are likely in the first half, or at best, in the middle of the recovery. It is important to remember that continued recovery assumes the maintenance and expansion of current effective sizes. Therefore, approaches to expand the total demographic population size and, by extension, the effective population sizes and connectivity among these populations must be central in future management plans for this species.

10 Acknowledgements

First and foremost, we thank the Department of Defense and the U.S. Army Corps of Engineers for funding this work and the City of Austin for funding the data collection and providing logistical support. The work presented here was supported by Grant #W9126G-19-2-0014. We especially thank Chris Harper (USFWS), Ryan Orndorff (Department of Interior), Elizabeth Galli-Noble (DoD Legacy Program), and Mel Coleman (USACE) for their support of the project. Thank you to L. Porter, G. Terry, S. Carrasco, J. Levac, C. Decker, J. Chenowith, J. Scalise, C. Sperry, C. Campbell, S. Summers, J. Alfieri, and C. Strickland for their assistance with fieldwork. We also thank the private and public landowners that provided access so that we could collect samples: T. Hatfield, G. Marsh, B. Marsh, J. Porcher-Streitwieser, B. Murden, R. Phillips, C. Smith, B. Fisher, Texas Parks and Wildlife, The Nature Conservancy, Fossil Rim Wildlife Center, Camp Bullis, and Fort Hood, and U.S. Fish and Wildlife. Thank you to the numerous TPWD and USFWS staff, D. Gimnich, S. Rowin, J. Mueller, A. Eyres, H. Haefele, T. Edwards, C. Reemts, R. Neill, and J. Florence for assisting with access, logistics, permits, and support. We thank the Texas Institute for Genome Sciences & Society (TIGSS) for the sequencing work. Last but not least, we thank Dr. Byron Stone of Austin, TX, and Patsy and Tom Inglet of San Antonio, TX, for their help in intangible ways.

References

[1] Fred W Allendorf, Paul A Hohenlohe, and Gordon Luikart. Genomics and the future of conservation genetics. *Nature Reviews. Genetics*, 11(10):697–709, oct 2010.

[2] Kimberly R Andrews, Jeffrey M Good, Michael R Miller, Gordon Luikart, and Paul A Hohenlohe. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews. Genetics*, 17(2):81–92, feb 2016.

[3] Christina M. Andruk and Norma L. Fowler. Conflicting short and long-term management goals: Fire effects in endangered golden-cheeked warbler (setophaga chrysoparia) habitat. *In: Keane, Robert E.; Jolly, Matt; Parsons, Russell; Riley, Karin. Proceedings of the large wildland fires conference; May 19-23, 2014; Missoula, MT. Proc. RMRS-P-73. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. p. 22-29.*, 2015.

[4] Giridhar Athrey, Nikolas Faust, Anne-Sophie Charlotte Hieke, and I Lehr Brisbin. Effective population sizes and adaptive genetic variation in a captive bird population. *PeerJ*, 6:e5803, oct 2018.

[5] Giridhar Athrey, Denise L. Lindsay, Richard F. Lance, and Paul L. Leberg. Crumbling diversity: comparison of historical archived and contemporary natural populations indicate reduced genetic diversity and increasing genetic differentiation in the golden-cheeked warbler. *Conservation genetics (Print)*, 12(5):1345–1355, oct 2011.

[6] J A M Bertrand, B Delahaie, Y X C Bourgeois, T Duval, R Garca-Jimnez, J Cornuault, B Pujol, C Thbaud, and B Mil. The role of selection and historical factors in driving population differentiation along an elevational gradient in an island bird. *Journal of Evolutionary Biology*, 29(4):824–836, apr 2016.

[7] Rebecca G Cheek, Brenna R Forester, Patricia E Salerno, Daryl R Trumbo, Kathryn M Langin, Nancy Chen, T Scott Sillett, Scott A Morrison, Cameron K Ghalambor, and W Chris Funk. Habitat-linked genetic variation supports microgeographic adaptive divergence in an island-endemic bird species. *Molecular Ecology*, 31(10):2830–2846, may 2022.

[8] Sonya M Clegg, Sandie M Degnan, Jiro Kikkawa, Craig Moritz, Arnaud Estoup, and Ian P F Owens. Genetic consequences of sequential founder events by an island-colonizing bird. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12):8127–8132, jun 2002.

[9] Petr Danecek, James K Bonfield, Jennifer Liddle, John Marshall, Valeriu Ohan, Martin O Pollard, Andrew Whitwham, Thomas Keane, Shane A McCarthy, Robert M Davies, and Heng Li. Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2), feb 2021.

[10] Ludovic Dutoit, Reto Burri, Alexander Nater, Carina F Mugal, and Hans Ellegren. Genomic distribution and estimation of nucleotide diversity in natural populations:

perspectives from the collared flycatcher (ficedula albicollis) genome. *Molecular ecology resources*, 17(4):586–597, jul 2017.

[11] Carolyn Enloe, W. Andrew Cox, Akanksha Pandey, Sabrina S. Taylor, Stefan Woltmann, and Rebecca T. Kimball. Genome-wide assessment of population structure in floridacoastal seaside sparrows. *Conservation genetics (Print)*, nov 2021.

[12] Laurent Excoffier, Matthieu Foll, and Rmy J. Petit. Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics*, 40(1):481–501, dec 2009.

[13] Y X Fu and W H Li. Statistical tests of neutrality of mutations. *Genetics*, 133(3):693–709, mar 1993.

[14] W. C. Funk, Brenna R. Forester, Sarah J. Converse, Catherine Darst, and Steve Morey. Improving conservation policy with genomics: a guide to integrating adaptive potential into u.s. endangered species act decisions for conservation practitioners and geneticists. *Conservation genetics (Print)*, 20(1):1–20, aug 2018.

[15] Sara Goodwin, John D McPherson, and W Richard McCombie. Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews. Genetics*, 17(6):333–351, may 2016.

[16] Ehsan Haghshenas, Hossein Asghari, Jens Stoye, Cedric Chauve, and Faraz Hach. HASLR: fast hybrid assembly of long reads. *iScience*, 23(8):101389, aug 2020.

[17] Philip W Hedrick. Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology*, 22(18):4606–4618, sep 2013.

[18] Zhen Huang, Zaoxu Xu, Hao Bai, Yongji Huang, Na Kang, Xiaoting Ding, Jing Liu, Haoran Luo, Chentao Yang, Wanjun Chen, Qixin Guo, Lingzhan Xue, Xueping Zhang, Li Xu, Meiling Chen, Honggao Fu, Youling Chen, Zhicao Yue, Tatsuo Fukagawa, Shanlin Liu, Guobin Chang, and Luohao Xu. Evolutionary analysis of a complete chicken genome. *Proceedings of the National Academy of Sciences of the United States of America*, 120(8):e2216641120, feb 2023.

[19] Babraham Institute. Babraham bioinformatics - FastQC a quality control tool for high throughput sequence data, oct 2018.

[20] Rei Kajitani, Kouta Toshimoto, Hideki Noguchi, Atsushi Toyoda, Yoshitoshi Ogura, Miki Okuno, Mitsuru Yabana, Masayuki Harada, Eiji Nagayasu, Haruhiko Maruyama, Yuji Kohara, Asao Fujiyama, Tetsuya Hayashi, and Takehiko Itoh. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Research*, 24(8):1384–1395, aug 2014.

[21] Jaana Kekkonen, Ilpo K. Hanski, Henrik Jensen, Risto A. Visnen, and Jon E. Brommer. Increased genetic differentiation in house sparrows after a strong population decline: From panmixia towards structure in a common bird. *Biological conservation*, 144(12):2931–2940, dec 2011.

[22] M Kimura. The neutral theory of molecular evolution. *Scientific American*, 241(5):98–100, 102, 108 passim, nov 1979.

[23] Sergey Koren, Brian P Walenz, Konstantin Berlin, Jason R Miller, Nicholas H Bergman, and Adam M Phillippy. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research*, 27(5):722–736, may 2017.

[24] Thorfinn Sand Korneliussen, Anders Albrechtsen, and Rasmus Nielsen. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics*, 15:356, nov 2014.

[25] Paul L. Leberg, Giridhar N. R. Athrey, Kelly R. Barr, Denise L. Lindsay, and Richard F. Lance. Implications of landscape alteration for the conservation of genetic diversity of endangered species. In J. Andrew DeWoody, John W. Bickham, Charles H. Michler, Krista M. Nichols, Gene E. Rhodes, and Keith E. Woeste, editors, *Molecular approaches in natural resource conservation and management*, pages 212–238. Cambridge University Press, Cambridge, 2010.

[26] Heng Li and Richard Durbin. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, 25(14):1754–1760, jul 2009.

[27] Heng Li and Richard Durbin. Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357):493–496, jul 2011.

[28] Denise L Lindsay, Kelly R Barr, Richard F Lance, Scott A Tweddale, Timothy J Hayden, and Paul L Leberg. Habitat fragmentation and genetic diversity of an endangered, migratory songbird, the golden-cheeked warbler (dendroica chrysoparia). *Molecular Ecology*, 17(9):2122–2133, may 2008.

[29] Ashley M. Long, Melanie R. Coln, Jessica L. Bosman, Tiffany M. Mcfarland, Anthony J. Locatelli, Laura R. Stewart, Heather A. Mathewson, John C. Newnam, and Michael L. Morrison. Effects of road construction noise on golden-cheeked warblers: An update. *Wildlife Society bulletin*, 41(2):240–248, jun 2017.

[30] Zhanshan Sam Ma, Lianwei Li, Chengxi Ye, Minsheng Peng, and Ya-Ping Zhang. Hybrid assembly of ultra-long nanopore reads augmented with 10x-genomics contigs: Demonstrated with a human genome. *Genomics*, 111(6):1896–1901, dec 2019.

[31] Sithembile O. Makina, Farai C. Muchadeyi, Este van Marle-Kster, Jerry F. Taylor, Mahlako L. Makgahlela, and Azwihangwisi Maiwashe. Genome-wide scan for selection signatures in six cattle breeds in south africa. *Genetics, selection, evolution : GSE*, 47(1):92, dec 2015.

[32] Ann M Mc Cartney, Kishwar Shafin, Michael Alonge, Andrey V Bzikadze, Giulio Formenti, Arkarachai Fungtammasan, Kerstin Howe, Chirag Jain, Sergey Koren, Glennis A Logsdon, Karen H Miga, Alla Mikheenko, Benedict Paten, Alaina Shumate, Daniela C Soto, Ivan Sovi, Jonathan M D Wood, Justin M Zook, Adam M Phillippy, and Arang Rhie. Chasing perfection: validation and polishing strategies for telomere-to-telomere genome assemblies. *Nature Methods*, 19(6):687–695, jun 2022.

[33] Krystyna Nadachowska-Brzyska, Cai Li, Linnea Smeds, Guojie Zhang, and Hans Ellegren. Temporal dynamics of avian populations during pleistocene revealed by whole-genome sequences. *Current Biology*, 25(10):1375–1380, may 2015.

[34] Niranjan Nagarajan and Mihai Pop. Sequence assembly demystified. *Nature Reviews. Genetics*, 14(3):157–167, mar 2013.

[35] S. R. Narum, D. Hatch, A. J. Talbot, P. Moran, and M. S. Powell. Iteroparity in complex mating systems of steelhead oncorhynchus mykiss (walbaum). *Journal of Fish Biology*, 72(1):45–60, jan 2008.

[36] Tetsuro Nomura. Estimation of effective number of breeders from molecular coancestry of single cohort sample. *Evolutionary applications*, 1(3):462–474, aug 2008.

[37] John Novembre and Montgomery Slatkin. Likelihood-based inference in isolation-by-distance models using the spatial distribution of low-frequency alleles. *Evolution*, 63(11):2914–2925, nov 2009.

[38] J K Pritchard, M Stephens, and P Donnelly. Inference of population structure using multilocus genotype data. *Genetics*, 155(2):945–959, jun 2000.

[39] AI Pudovkin, OL Zhdanova, and D Hedgecock. Sampling properties of the heterozygote-excess estimator of the effective number of breeders. *Conservation Genetics*, 2010.

[40] Dianne H. Robinson, Heather A. Mathewson, Michael L. Morrison, and R. Neal Wilkins. Habitat effects on golden-cheeked warbler productivity in an urban landscape. *Wildlife Society bulletin*, 42(1):48–56, mar 2018.

[41] F Rousset. Genetic differentiation and estimation of gene flow from f-statistics under isolation by distance. *Genetics*, 145(4):1219–1228, apr 1997.

[42] Michael A Russello, Matthew D Waterhouse, Paul D Etter, and Eric A Johnson. From promise to practice: pairing non-invasive sampling with genomics in conservation. *PeerJ*, 3:e1106, jul 2015.

[43] Michael K Schwartz, Gordon Luikart, and Robin S Waples. Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology & Evolution*, 22(1):25–33, jan 2007.

[44] Ashley T Sendell-Price, Kristen C Ruegg, Bruce C Robertson, and Sonya M Clegg. An island-hopping bird reveals how founder events shape genome-wide divergence. *Molecular Ecology*, 30(11):2495–2510, jun 2021.

[45] Felipe A Simo, Robert M Waterhouse, Panagiotis Ioannidis, Evgenia V Kriventseva, and Evgeny M Zdobnov. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19):3210–3212, oct 2015.

[46] Laura R. Stewart, Michael L. Morrison, Mark R. Hutchinson, David N. Appel, and R. Neal Wilkins. Effects of a forest pathogen on habitat selection and quality for the endangered golden-cheeked warbler. *Wildlife Society bulletin*, 38(2):279–287, jun 2014.

[47] D. Stojanovic, E. McLennan, G. Olah, M. Cobden, R. Heinsohn, A. D. Manning, F. Alves, C. Hogg, and L. Rayner. Reproductive skew in a vulnerable bird favors breeders that monopolize nest cavities. *Animal Conservation*, jan 2023.

[48] F Tajima. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3):585–595, nov 1989.

[49] Mun Hua Tan, Christopher M Austin, Michael P Hammer, Yin Peng Lee, Laurence J Croft, and Han Ming Gan. Finding nemo: hybrid assembly with oxford nanopore and illumina reads greatly improves the clownfish (amphiprion ocellaris) genome assembly. *GigaScience*, 7(3):1–6, mar 2018.

[50] Sarah A Tishkoff and Brian C Verrelli. Patterns of human genetic diversity: implications for human evolutionary history and disease. *Annual Review of Genomics and Human Genetics*, 4:293–340, 2003.

[51] Wendy Vandersteen Tymchuk, Patrick O'Reilly, Jesse Bittman, Danielle Macdonald, and Patricia Schulte. Conservation genomics of atlantic salmon: variation in gene expression between and within regions of the bay of fundy. *Molecular Ecology*, 19(9):1842–1859, may 2010.

[52] Robin S Waples, Tiago Antao, and Gordon Luikart. Effects of overlapping generations on linkage disequilibrium estimates of effective population size. *Genetics*, 197(2):769–780, jun 2014.

[53] Robin S Waples and Oscar Gaggiotti. What is a population? an empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, 15(6):1419–1439, may 2006.

[54] Wesley C Warren, LaDeana W Hillier, Chad Tomlinson, Patrick Minx, Milinn Kremitzki, Tina Graves, Chris Markovic, Nathan Bouk, Kim D Pruitt, Francoise Thibaud-Nissen, Valerie Schneider, Tamer A Mansour, C Titus Brown, Aleksey Zimin, Rachel Hawken, Mitch Abrahamsen, Alexis B Pyrkosz, Mireille Morisson, Valerie Fillon, Alain Vignal, William Chow, Kerstin Howe, Janet E Fulton, Marcia M Miller, Peter Lovell, Claudio V Mello, Morgan Wirthlin, Andrew S Mason, Richard Kuo, David W Burt, Jerry B Dodgson, and Hans H Cheng. A new chicken genome assembly provides insight into avian genome structure. *G3 (Bethesda, Md.)*, 7(1):109–117, jan 2017.

[55] M C Whitlock and D E McCauley. Indirect measures of gene flow and migration: FST not equal to 1/(4Nm + 1). *Heredity*, 82 (Pt 2):117–125, feb 1999.

[56] Jochen B W Wolf, Till Bayer, Bernhard Haubold, Markus Schilhabel, Philip Rosenstiel, and Diethard Tautz. Nucleotide divergence vs. gene expression differentiation: comparative transcriptome sequencing in natural isolates from the carrion crow and its hybrid zone with the hooded crow. *Molecular Ecology*, 19 Suppl 1:162–175, mar 2010.

[57] S Wright. *The roles of mutation, inbreeding, crossbreeding, and selection in evolution*. 1932.

[58] S Wright. Isolation by distance. *Genetics*, 28(2):114–138, mar 1943.

[59] Sewall Wright. On the roles of directed and random changes in gene frequency in the genetics of populations. *Evolution*, 2(4):279, dec 1948.

11 Appendix I

Table 4: Sample Table

Sample ID	Location	County
2830_17850_S169	Love Creek Preserve	Bandera
2830_17851_S170	Love Creek Preserve	Bandera
2830_17857_S171	Love Creek Preserve	Bandera
2830_17859_M_61_S175	Love Creek Preserve	Bandera
2830_17862_S176	Love Creek Preserve	Bandera
2830_17863_S177	Love Creek Preserve	Bandera
2830_17864_S178	Love Creek Preserve	Bandera
2830_17865_S179	Love Creek Preserve	Bandera
2830_17866_S180	Love Creek Preserve	Bandera
2830_17867_S181	Hatfield Ranch	Bandera
2830_17868_S182	Hatfield Ranch	Bandera
2830_17869_S183	Hatfield Ranch	Bandera
2830_17871_S185	Hatfield Ranch	Bandera
2830_17872_S186	Hatfield Ranch	Bandera
2830_17873_S187	Hatfield Ranch	Bandera
2830_17874_S188	Hatfield Ranch	Bandera
2830_19028_S217	Hatfield Ranch	Bandera
2830_19029_S1	Hatfield Ranch	Bandera
2830_19030_S2	Hatfield Ranch	Bandera
2830_19031_S3	Hatfield Ranch	Bandera
2830_19196_S44	Love Creek Preserve	Bandera
2830_19197_S45	Love Creek Preserve	Bandera
2830_19198_S46	Love Creek Preserve	Bandera
2830_19199_S47	Love Creek Preserve	Bandera
2830_19200_S48	Love Creek Preserve	Bandera
2760_11891_S50	Camp Bullis	Bexar
2760_11892_S51	Camp Bullis	Bexar
2760_11893_S52	Camp Bullis	Bexar
2760_11897_S132	Government Canyon SNA	Bexar
2760_11898_S133	Government Canyon SNA	Bexar
2760_11899_S245	Government Canyon SNA	Bexar
2790_35292_S218	Government Canyon SNA	Bexar
2790_35293_S219	Government Canyon SNA	Bexar
2790_35294_S237	Government Canyon SNA	Bexar
2790_35304_S227	Camp Bullis	Bexar
2790_35305_S228	Camp Bullis	Bexar
2790_35306_S229	Camp Bullis	Bexar
2790_35307_S230	Camp Bullis	Bexar

Continued on next page

Table 4 – continued from previous page

Sample ID	Location	County
2790_55303_S226	Camp Bullis	Bexar
2830_17815_S149	Government Canyon SNA	Bexar
2830_17816_S150	Government Canyon SNA	Bexar
2830_17817_S151	Government Canyon SNA	Bexar
2830_17881_S189	Camp Bullis	Bexar
2830_17882_S190	Camp Bullis	Bexar
2830_17883_S191	Camp Bullis	Bexar
2830_17884_S192	Camp Bullis	Bexar
2830_17885_S193	Camp Bullis	Bexar
2830_17886_S194	Camp Bullis	Bexar
2830_17887_S195	Camp Bullis	Bexar
2830_17888_S196	Camp Bullis	Bexar
2830_17889_S197	Camp Bullis	Bexar
2830_17890_S198	Camp Bullis	Bexar
2830_17891_S199	Camp Bullis	Bexar
2830_17892_S200	Camp Bullis	Bexar
2830_17893_S201	Camp Bullis	Bexar
2830_17894_S202	Camp Bullis	Bexar
2830_19110_S60	Government Canyon SNA	Bexar
2830_19111_S61	Government Canyon SNA	Bexar
2830_19113_S62	Government Canyon SNA	Bexar
2830_19114_S63	Government Canyon SNA	Bexar
2830_19121_S70	Government Canyon SNA	Bexar
2830_19122_S71	Government Canyon SNA	Bexar
2830_19130_S79	Government Canyon SNA	Bexar
2830_19131_S80	Government Canyon SNA	Bexar
2830_19132_S81	Government Canyon SNA	Bexar
2830_19134_S82	Government Canyon SNA	Bexar
2830_19135_S83	Government Canyon SNA	Bexar
2830_17818_S152	Meridian State Park	Bosque
2830_17819_S153	Meridian State Park	Bosque
2830_17820_S154	Meridian State Park	Bosque
2830_17900_S207	Meridian State Park	Bosque
2830_19019_S208	Meridian State Park	Bosque
2830_19144_S91	Meridian State Park	Bosque
2830_19145_S92	Meridian State Park	Bosque
2830_19146_S93	Meridian State Park	Bosque
2830_19147_S94	Meridian State Park	Bosque
2830_19148_S95	Meridian State Park	Bosque
2830_19149_S96	Meridian State Park	Bosque
2830_19150_S114	Meridian State Park	Bosque
2830_19151_S115	Meridian State Park	Bosque

Continued on next page

Table 4 – continued from previous page

Sample ID	Location	County
2830_19152_S116	Meridian State Park	Bosque
2830_19153_S117	Meridian State Park	Bosque
2830_19173_S22	Ft Hood	Coryell/Bell
2830_19174_S23	Ft Hood	Coryell/Bell
2830_19175_S24	Ft Hood	Coryell/Bell
2830_19176_S25	Ft Hood	Coryell/Bell
2830_19177_S26	Ft Hood	Coryell/Bell
2830_19178_S27	Ft Hood	Coryell/Bell
2830_19179_S28	Ft Hood	Coryell/Bell
2830_19180_S29	Ft Hood	Coryell/Bell
2830_19181_S30	Ft Hood	Coryell/Bell
2830_19182_S31	Ft Hood	Coryell/Bell
2830_19183_S32	Ft Hood	Coryell/Bell
2830_19184_S33	Ft Hood	Coryell/Bell
2830_19185_S34	Ft Hood	Coryell/Bell
2830_19186_S247	Ft Hood	Coryell/Bell
2830_19187_S35	Ft Hood	Coryell/Bell
2830_19188_S36	Ft Hood	Coryell/Bell
2830_19189_S37	Ft Hood	Coryell/Bell
2830_19190_S38	Ft Hood	Coryell/Bell
2830_19191_S39	Ft Hood	Coryell/Bell
2830_19192_S40	Ft Hood	Coryell/Bell
2830_19193_S41	Ft Hood	Coryell/Bell
2830_19194_S42	Ft Hood	Coryell/Bell
2830_19195_S43	Ft Hood	Coryell/Bell
2760_11884_S110	Guadalupe River State Park	Kendall
2760_11885_S111	Guadalupe River State Park	Kendall
2760_11886_S112	Guadalupe River State Park	Kendall
2760_11887_S113	Guadalupe River State Park	Kendall
2760_11888_S223	Guadalupe River State Park	Kendall
2760_11890_S49	Guadalupe River State Park	Kendall
2760_11900_S134	Guadalupe River State Park	Kendall
2830_17801_S135	Guadalupe River State Park	Kendall
2830_17827_S161	Guadalupe River State Park	Kendall
2830_17828_S162	Guadalupe River State Park	Kendall
2830_17829_S163	Guadalupe River State Park	Kendall
2830_17830_S164	Guadalupe River State Park	Kendall
2830_19102_S53	Guadalupe River State Park	Kendall
2830_19103_S54	Guadalupe River State Park	Kendall
2830_19104_S55	Guadalupe River State Park	Kendall
2830_19105_S56	Guadalupe River State Park	Kendall
2830_19106_S57	Guadalupe River State Park	Kendall

Continued on next page

Table 4 – continued from previous page

Sample ID	Location	County
2830_19107_S58	Guadalupe River State Park	Kendall
2830_19108_S59	Guadalupe River State Park	Kendall
2830_19115_S64	Guadalupe River State Park	Kendall
2830_19116_S65	Guadalupe River State Park	Kendall
2830_19117_S66	Guadalupe River State Park	Kendall
2830_19118_S67	Guadalupe River State Park	Kendall
2830_19119_S68	Guadalupe River State Park	Kendall
2830_19120_S69	Guadalupe River State Park	Kendall
2830_19020_S209	Kickapoo Caverns State Park	Kinney/Edwards
2830_19021_S210	Kickapoo Caverns State Park	Kinney/Edwards
2830_19022_S211	Kickapoo Caverns State Park	Kinney/Edwards
2830_19023_S212	Kickapoo Caverns State Park	Kinney/Edwards
2830_19024_S213	Kickapoo Caverns State Park	Kinney/Edwards
2830_19025_S214	Kickapoo Caverns State Park	Kinney/Edwards
2830_19026_S215	Kickapoo Caverns State Park	Kinney/Edwards
2830_19027_S216	Kickapoo Caverns State Park	Kinney/Edwards
2830_19159_S123	Kickapoo Cavern State Park	Kinney/Edwards
2830_19160_S124	Kickapoo Cavern State Park	Kinney/Edwards
2830_19161_S125	Kickapoo Cavern State Park	Kinney/Edwards
2830_19162_S126	Kickapoo Cavern State Park	Kinney/Edwards
2830_19163_S127	Dobbs Run Ranch	Kinney/Edwards
2830_19164_S128	Dobbs Run Ranch	Kinney/Edwards
2830_19165_S129	Dobbs Run Ranch	Kinney/Edwards
2830_17821_S155	Palo Pinto State Park	Palo Pinto
2830_17822_S156	Palo Pinto State Park	Palo Pinto
2830_17823_S157	Palo Pinto State Park	Palo Pinto
2830_17824_S158	Palo Pinto State Park	Palo Pinto
2830_19037_S9	Palo Pinto State Park	Palo Pinto
2830_19038_S10	Palo Pinto State Park	Palo Pinto
2830_19039_S11	Palo Pinto State Park	Palo Pinto
2830_19040_S12	Palo Pinto State Park	Palo Pinto
2830_19041_S13	Palo Pinto State Park	Palo Pinto
2830_19042_S14	Palo Pinto State Park	Palo Pinto
2830_19043_S15	Palo Pinto State Park	Palo Pinto
2830_19154_S118	Palo Pinto State Park	Palo Pinto
2830_19155_S119	Palo Pinto State Park	Palo Pinto
2830_19156_S120	Palo Pinto State Park	Palo Pinto
2830_19157_S121	Palo Pinto State Park	Palo Pinto
2830_19158_S122	Palo Pinto State Park	Palo Pinto
255_99667_S107	Colorado Bend State Park	San Saba
255_99668_S108	Colorado Bend State Park	San Saba
255_99669_S244	Colorado Bend State Park	San Saba

Continued on next page

Table 4 – continued from previous page

Sample ID	Location	County
255_99670_S109	Colorado Bend State Park	San Saba
2830_17844_S167	Colorado Bend State Park	San Saba
2830_17845_S168	Colorado Bend State Park	San Saba
2830_19136_S84	Colorado Bend State Park	San Saba
2830_19137_S85	Colorado Bend State Park	San Saba
2830_19138_S86	Colorado Bend State Park	San Saba
2830_19140_S87	Colorado Bend State Park	San Saba
2830_19141_S88	Colorado Bend State Park	San Saba
2830_19142_S89	Colorado Bend State Park	San Saba
2830_19143_S90	Colorado Bend State Park	San Saba
2830_19166_S16	Colorado Bend State Park	San Saba
2830_19167_S17	Colorado Bend State Park	San Saba
2830_19168_S18	Colorado Bend State Park	San Saba
2830_19169_S19	Colorado Bend State Park	San Saba
2830_19170_S20	Colorado Bend State Park	San Saba
2830_19172_S21	Colorado Bend State Park	San Saba
255_99657_S100	Fossil Rim Wildlife Center	Somervell
255_99658_S242	Fossil Rim Wildlife Center	Somervell
255_99659_S101	Cahopa Ranch	Somervell
255_99661_S102	Cahopa Ranch	Somervell
2790_35295_S238	Cahopa Ranch	Somervell
2790_35297_S220	Cahopa Ranch	Somervell
2790_35298_S221	Fossil Rim Wildlife Center	Somervell
2790_35299_S240	Fossil Rim Wildlife Center	Somervell
2790_35300_S241	Fossil Rim Wildlife Center	Somervell
2830_17825_S159	Cahopo Ranch	Somervell
2830_17826_S160	Cahopo Ranch	Somervell
2830_17842_S165	Marsh Ranch	Somervell
2830_17843_S166	Marsh Ranch	Somervell
2830_17895_S203	Marsh Ranch	Somervell
2830_17896_S204	Marsh Ranch	Somervell
2830_17897_S205	Marsh Ranch	Somervell
2830_17898_S206	Marsh Ranch	Somervell
2830_17899_S246	Marsh Ranch	Somervell
255_99662_S222	BCNWR (Victoria tract)	Travis
255_99663_S103	BCNWR (Victoria tract)	Travis
255_99664_S104	BCNWR (Victoria tract)	Travis
255_99665_S105	BCNWR (Victoria tract)	Travis
255_99666_S106	BCNWR (Victoria tract)	Travis
2790_35242_032319_S239	BCP	Travis
2790_35288_S234	BCNWR (Victoria)	Travis
2790_35289_S235	BCNWR (Victoria)	Travis

Continued on next page

Table 4 – continued from previous page

Sample ID	Location	County
2790_35290_S99	BCNWR (Victoria)	Travis
2790_35291_S236	BCNWR (Victoria)	Travis
2790_35308_S231	BCNWR (Victoria)	Travis
2790_35309_S232	BCNWR (Victoria)	Travis
2790_35310_S233	BCNWR (Victoria)	Travis
2790_35322_S97	BCP	Travis
2830_17701_S98	BCNWR (Victoria)	Travis
2830_17802_S136	Shield Ranch	Real
2830_17803_S137	Shield Ranch	Real
2830_17804_S138	Shield Ranch	Real
2830_17805_S139	Shield Ranch	Real
2830_17806_S140	Shield Ranch	Real
2830_17807_S141	Shield Ranch	Real
2830_17808_S142	Garner State Park	Uvalde
2830_17809_S143	Garner State Park	Uvalde
2830_17810_S144	Garner State Park	Uvalde
2830_17811_S145	Garner State Park	Uvalde
2830_17812_S146	Garner State Park	Uvalde
2830_17813_S147	Garner State Park	Uvalde
2830_17814_S148	Garner State Park	Uvalde
2830_19032_S4	Garner State Park	Uvalde
2830_19033_S5	Garner State Park	Uvalde
2830_19034_S6	Garner State Park	Uvalde
2830_19035_S7	Garner State Park	Uvalde
2830_19036_S8	Garner State Park	Uvalde
2830_19123_S72	Shield Ranch	Real
2830_19124_S73	Shield Ranch	Real
2830_19125_S74	Shield Ranch	Real
2830_19126_S75	Garner State Park	Uvalde
2830_19127_S76	Garner State Park	Uvalde
2830_19128_S77	Garner State Park	Uvalde
2830_19129_S78	Shield Ranch	Real