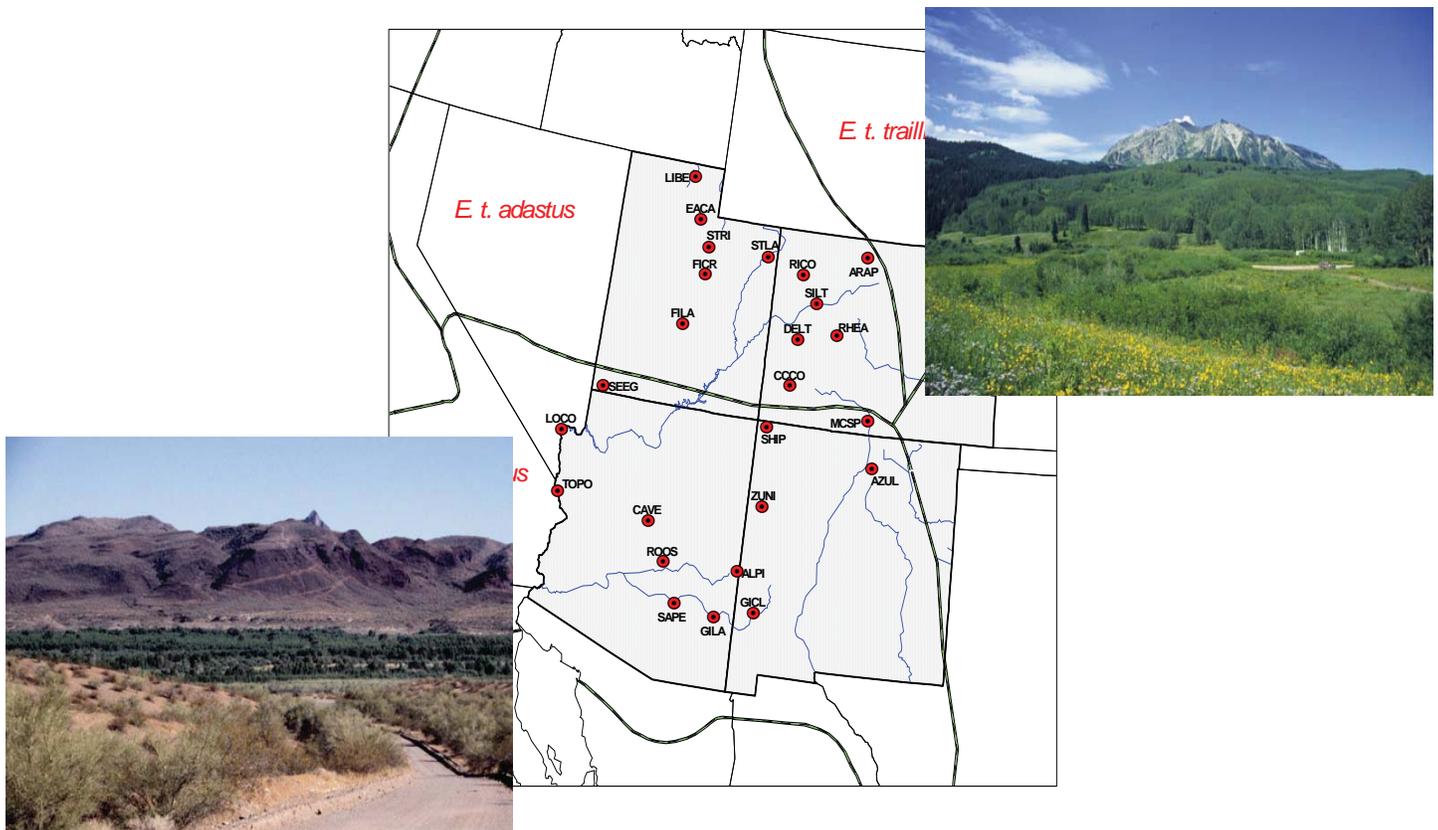


In cooperation with Northern Arizona University

Using Molecular Genetic Markers to Resolve a Subspecies Boundary: the Northern Boundary of the Southwestern Willow Flycatcher in the Four-corner States



Open-File Report 2008–1117

This page intentionally left blank



In cooperation with Northern Arizona University

Using Molecular Genetic Markers to Resolve a Subspecies Boundary: the Northern Boundary of the Southwestern Willow Flycatcher in the Four-corner States

By Eben H. Paxton^{1 2}, Mark K. Sogge², Tad C. Theimer¹, Jessica Girard¹, and Paul Keim¹

¹Northern Arizona University, Department of Biological Sciences, Flagstaff, Ariz.

²U.S. Geological Science Center, Southwest Biological Science Center, Colorado Plateau Research Station, Flagstaff, Ariz.

Open-File Report 2008–1117

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
DIRK KEMPTHORNE, Secretary

U.S. Geological Survey
Mark D. Myers, Director

U.S. Geological Survey, Reston, Virginia 2008

For product and ordering information:
World Wide Web: <http://www.usgs.gov/pubprod>
Telephone: 1-888-ASK-USGS

For more information on the USGS—the Federal source for science about the Earth,
its natural and living resources, natural hazards, and the environment:
World Wide Web: <http://www.usgs.gov>
Telephone: 1-888-ASK-USGS

Suggested citation:
Paxton, E.H., Sogge, M.K., Theimer, T.C., Girard, J., and Keim, P., 2008, Using molecular genetic markers to resolve a subspecies boundary: the northern boundary of the Southwestern Willow Flycatcher in the four-corner states: U.S. Geological Survey Open-File Report 2007–1117, 20 p.

Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this report is in the public domain, permission must be secured from the individual copyright owners to reproduce any copyrighted material contained within this report.

Contents

Executive Summary.....	1
Introduction	2
Methods	4
Results	7
Discussion.....	13
Acknowledgements	17
References.....	18

Figures

1. Location of Willow Flycatcher breeding sites sampled for this study.....	3
2. Willow Flycatcher breeding sites clustered out into two groups using a UPGMA analysis of genetic distances, based on mitochondrial DNA and nuclear DNA.....	10
3. Frequency of C-group haplotypes observed at breeding sites decreases with increasing latitude	12
4. Contours indicating the predicted probability of a breeding site having a C-group haplotype based on latitude and elevation.....	14
5. The amount of genetic information that each C-group probability line separates is highest at the 30% contour line.	15

Tables

1. Willow Flycatcher breeding sites sampled in Arizona, Colorado, New Mexico, and Utah, 1996–2004.....	5
2. Cytochrome-b sequence haplotype polymorphic positions and their frequency encountered at sampled breeding sites	8
3. Analysis of molecular variance results indicate that a significant amount of variation in mtDNA and nDNA frequencies is explained by grouping breeding sites into the two UPGMA clades	11

Using Molecular Genetic Markers to Resolve a Subspecies Boundary: the Northern Boundary of the Southwestern Willow Flycatcher in the Four-corner States

By Eben H. Paxton, Mark K. Sogge, Tad C. Theimer, Jessica Girard, and Paul Keim

Executive Summary

The northern boundary of the endangered Southwestern Willow Flycatcher (*Empidonax traillii extimus*) is currently approximated as running through southern Colorado and Utah, but the exact placement is uncertain because this subspecies shares a border with the more northern and non-endangered *E. t. adastus*. To help resolve this issue, we evaluated the geographic distribution of mitochondrial and nuclear DNA by sampling breeding sites across the four-corner states (Arizona, Colorado, New Mexico, and Utah). We found that breeding sites clustered into two major groups generally consistent with the currently designated boundary, with the exception of three sites situated along the current boundary. However, delineating a precise boundary that would separate the two subspecies is made difficult because (1) we found evidence for a region of intergradation along the boundary area, suggesting the boundary is not discrete, and (2) the boundary region is sparsely populated, with too few extant breeding populations to precisely locate a boundary. The boundary region encompasses an area where elevation changes markedly over relatively short distances, with low elevation deserts to the south and more mesic, higher elevation habitats to the north. We hypothesized that latitudinal and elevational differences and their concomitant ecological effects could form an ecological barrier that inhibited gene flow between the subspecies, forming the basis for the subspecies boundary. We modeled changes in geographic patterns of genetic markers as a function of latitude and elevation finding significant support for this relationship. The model was brought into a GIS environment to create multiple subspecies boundaries, with the strength of each predicted boundary evaluated on the basis of how much genetic variation it explained. The candidate boundary that accounted for the most genetic variation was situated generally near the currently recognized subspecies boundary, but should be more biologically meaningful because it incorporates the landscape features that may be driving separation of the subspecies. Even so, we caution that using any narrow boundary line as an indicator of subspecies identity could be misleading because biologically the boundary is a region of intergradation rather than a discrete line. Designating a boundary ultimately becomes a regulatory and management decision based on how much of the genetic variation unique to a subspecies should be protected. We discuss how the results of this study can help guide this decision process by wildlife policy makers.

Introduction

A wide array of national and international conservation laws and treaties, including the U.S. Endangered Species Act, provide for the recognition and protection of distinct portions of a species (e.g., subspecies, distinct populations); however, subspecific protection can raise challenging scientific, management, and conservation issues (Haig et al. 2006). One such issue is the difficulty of delineating the intraspecific range boundary between endangered and non-endangered subspecies or populations because of the management implications inherent to being on one side of the boundary line or the other. Delineating range boundaries is particularly difficult for intraspecific groups with contiguous, shared boundaries where a zone of introgression commonly occurs, increasing the difficulty in defining a distinct boundary (Avice 1994, Barrowclough et al. 2005).

The Willow Flycatcher (*Empidonax traillii*) is a neotropical migratory passerine that breeds across most of the United States and southern Canada, and winters from southern Mexico through Central America into northwestern South America (DeGraaf and Rappole 1995). The Willow Flycatcher is a polytypic species composed of four morphologically similar subspecies (Unitt 1987). In 1995, the USFWS declared the southwestern subspecies (*E. t. extimus*) endangered due to the extirpation from many historical breeding sites and a marked decrease in rangewide population size (USFWS 1995). Although there is general agreement on the approximate ranges of the subspecies among studies based on geographic patterns of morphological (Unitt 1987, Browning 1993), molecular genetic (Paxton 2000) and song (Sedgwick 2001) variation, range boundaries were for the most part approximated. Most of the boundary of *E. t. extimus* is discontinuous with other Willow Flycatcher subspecies: *E. t. extimus* occupies the southern most extent of the species range west to the Pacific Ocean, while the range of *E. t. extimus* to the east (Texas) and northwest (California and Nevada) is separated from other subspecies' populations by large geographic regions that are unoccupied due to either inhospitable terrain or historical extirpations (Unitt 1987). However, in Utah and Colorado the northern boundary of *E. t. extimus* is adjacent to the southern boundary of *E. t. adastus* (Fig. 1). Within this region, where the ranges of *E. t. extimus* and *E. t. adastus* meet and individuals of different subspecies may interbreed, the subspecies boundary is in need of resolution for legal and management purposes (USFWS 2002).

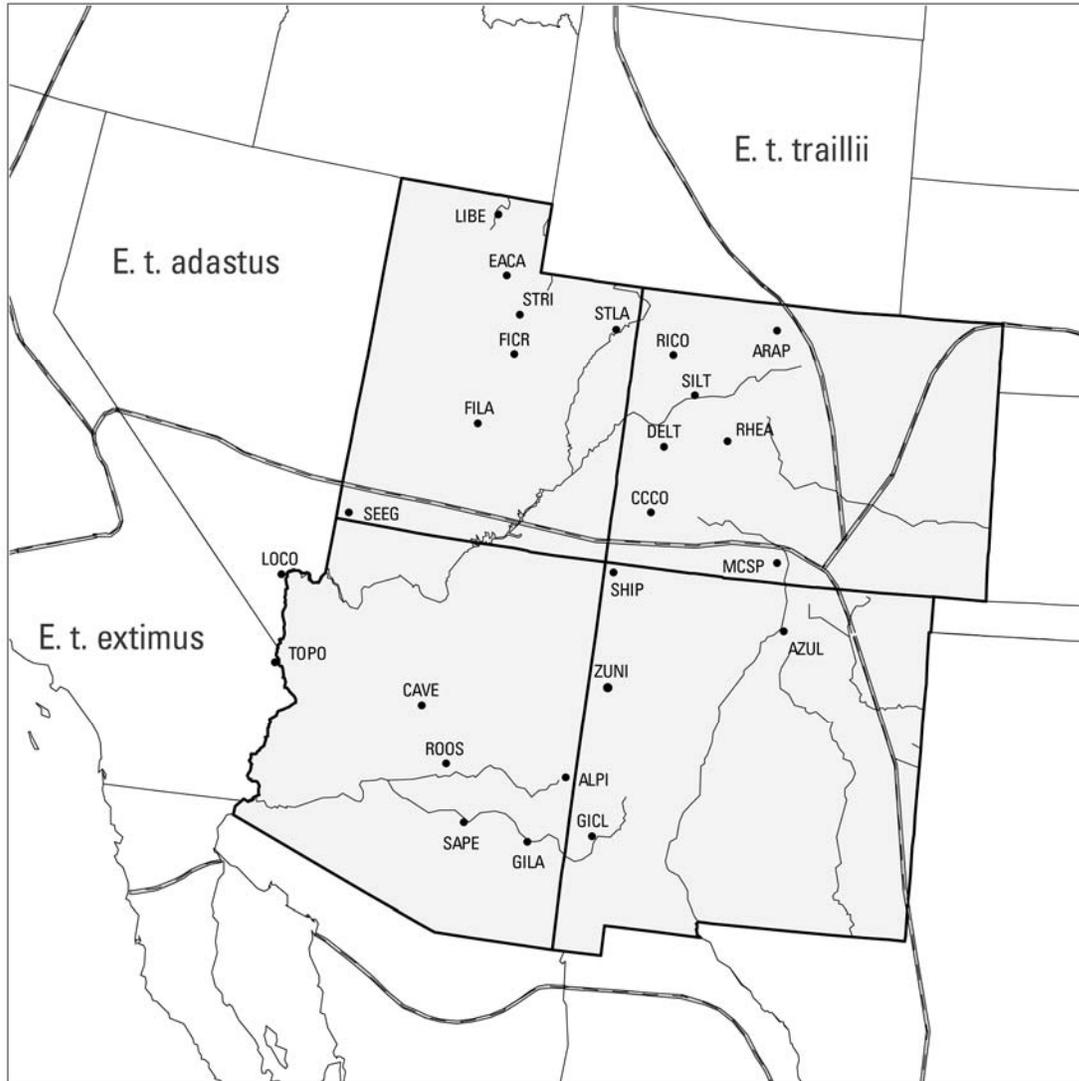


Figure 1. Location of Willow Flycatcher breeding sites sampled for this study. The four letter alpha code associated with each location is associated to the name of the site (see Table 1). The solid-dashed lines represent the approximate location of the currently managed subspecies boundaries (USFWS 2002).

Multiple studies have evaluated the geographic distribution of the two subspecies within this region, while only indirectly considering the issue of where the boundary between *E. t. extimus* and *E. t. adastus* lies. Virtually all studies found evidence of a region of intergradation separating the subspecies, though the proposed location of this region varied among studies. Based on morphological characteristics of museum specimens, Behle (1985) argued that flycatchers in northernmost Utah were *E. t. adastus*, those in the extreme southern part of the state were *E. t. extimus*, while most of Utah represented a gradual cline of morphological variation from one subspecies to the other. In contrast, Burleigh (1972) placed the location of *E. t. adastus* and *E. t. extimus* intergrades farther north into northern Utah, while Unitt (1987) and Browning (1993) believed the best boundary was through southern Utah, although Unitt (1987) noted

evidence for intergradation northward. In western Colorado, where very few museum specimens were collected, Unitt (1987) stated that there was not enough evidence to determine subspecies status, while Browning (1993) believed the region to be an intergrade zone. Based on geographic distribution of cytochrome-b sequences from flycatchers across their breeding range, Paxton (2000) found the strongest support for a southern Colorado/Utah boundary, but recognized that more sampling of the region was needed. Sedgwick (2001) analyzed geographic variation in Willow Flycatcher song patterns, which are innate (Kroodsma 1984), and found evidence for intergradation further south than other studies, into high elevation areas of Arizona and New Mexico. Based on the available evidence, the Southwestern Willow Flycatcher Recovery Team approximated the boundary as running along the southern portion of Utah and Colorado (Fig. 1), although it clearly acknowledged that more work is needed on exactly where to place the boundary (USFWS 2002).

The purpose of this study was to examine geographic patterns of Willow Flycatcher mitochondrial (mtDNA) and nuclear (nDNA) molecular genetic markers to determine the most appropriate location of the Southwestern Willow Flycatcher's northern range boundary in the four-corner states of Arizona, Colorado, New Mexico, and Utah. Mitochondrial DNA is ideal for evaluating genetic patterns across geographic space (Avice 1994), and the inclusion of nDNA provides additional independent information to help support or contrast patterns in the mitochondrial DNA (Moritz 1994). We assessed genetic relationships among flycatcher breeding sites distributed across the boundary area to understand where genetic structuring occurred. The boundary region is characterized by rapid changes in biotic communities, with desert lowlands to the south, and more mesic, high elevation habitats to the north. Because these environmental and biotic changes may inhibit gene flow between the two subspecies, and thus could help inform where a biologically meaningful boundary may lie, we built a statistical model associating the geographic distribution of the molecular markers with latitude and elevation as informative factors. Results from the model formed the basis of a GIS model that produced multiple candidate boundaries. Herein we evaluate the strength of each candidate boundary and discuss potential approaches to choosing one for management and conservation purposes.

Methods

Study Sites and Genetic Sampling

We sampled 145 individual flycatchers from 25 sites in Arizona, Colorado, New Mexico, and Utah (Fig. 1, Table 1). Sample sizes for sites varied from 2 to 16 individuals per site, with a mean of 6 per site (Table 1). Most known currently-occupied breeding sites in this region were sampled (Durst et al. 2007). Our objective was to sample at least five individuals from each site, but this was not always possible due to small breeding population sizes at many of the breeding locations, and nDNA markers were only generated for the larger breeding sites.

Table 1. Willow Flycatcher breeding sites sampled in Arizona, Colorado, New Mexico, and Utah, 1996–2004. For each breeding site there is a four letter site code (SITE), sample sizes for cytochrome-b sequences and AFLP analysis information, and information on the latitude, elevation. C-group refers to the predicted probability of finding a C-group haplotype at a particular breeding site, given latitude and elevation (see text).

Site	Name	State	Sample size mtDNA	Sample size nDNA	Latitude	Elevation	C-group
ALPI	Alpine Horse Pasture	AZ	9	0	33.830	2414	61%
CAVE	Camp Verde	AZ	3	0	34.558	942	76%
GILA	Solomon Bridge, Gila River	AZ	2	0	32.825	899	88%
LOCO	Lower Colorado River	AZ	7	9	36.050	200	73%
ROOS	Roosevelt Lake	AZ	12	8	33.767	640	85%
SAPE	San Pedro River	AZ	7	8	32.975	634	89%
TOPO	Topock Marsh	AZ	3	7	34.792	139	83%
ARAP	Arapahoe National Wildlife Refuge	CO	5	5	40.617	2470	7%
CCCO	Clear Creek	CO	9	6	37.792	2900	16%
DELT	Escalante State Wildlife Area	CO	6	0	38.756	1700	22%
MCSP	McIntire Springs	CO	16	6	37.283	2300	27%
RHEA	Gothic	CO	2	0	38.946	2830	11%
RICO	Rio Blanco Lake	CO	7	6	40.088	1830	12%
SILT	Colorado River at Silt	CO	4	0	39.541	1650	17%
AZUL	Tierra Azul	NM	5	0	36.304	2179	38%
GICL	West Fort Ditch	NM	5	9	33.039	1402	82%
SHIP	Shiprock	NM	2	0	36.867	1463	44%
ZUNI	Zuni/Nutria Diversion Reservoir	NM	4	0	35.235	2081	52%
EACA	East Canyon Reservoir	UT	4	11	40.874	1798	9%
FICR	Fish Creek	UT	5	6	39.775	2560	9%
FILA	Fish Lake	UT	4	5	38.693	3018	10%
LIBE	Little Bear River	UT	6	11	41.719	1372	9%
SEEG	St. George	UT	3	4	37.108	732	54%
STLA	Stewart Lake	UT	8	5	40.346	1158	17%
STRI	Strawberry River	UT	7	9	40.3468	2621	7%

We detected flycatchers via standard play-back surveys (Sogge et al. 1997), capturing them using mist nets (Sogge et al. 2001) for banding and genetic sampling. All captured flycatchers were banded with a Federal bird band imprinted with a unique number to assure we sampled each individual only once. To assure that no migrants were used in this study, we included only adults that were territorial at a site during the non-migrant period (15 June–20 July; Unitt 1987), or were known by field observations to be resident breeders (e.g., observed breeding activities). DNA samples were collected from 1996 to 2004, and included samples used in previous work (Busch et al. 2000; Paxton 2000).

Blood samples were taken by clipping a toenail to the vascularized tissue, and rinsing a drop of blood into a 1.5 mL tube with approximately 40 μ L of collection buffer (1xSSC, 50mM EDTA). Samples were stored on ice until frozen, with DNA subsequently isolated from the blood following the protocol of Mullenbach et al. (1989). Blood was digested overnight at 55o C in lysis buffer (10mM Tris, 1nM EDTA, 1 SDS,

100mM NaCl, pH 8) with 200 µg/mL proteinase K and 2 mM DTT (final concentration of both). This lysate was extracted with chloroform and followed by an isopropanol precipitation. An aliquot of each DNA extraction was then run on a 0.7 agarose gel to assess DNA quality and quantity.

DNA Sequencing

We sequenced 1,063 nucleotides of the cytochrome-b beginning 80 bases downstream from the start codon of the gene. All sequences for this region were confirmed with at least two aligned sequence reads. DNA was sequenced with primers obtained from Helm-Bychowski and Cracraft (1993; L14827: 5' CCACACTCCACACA GGCCTAATTAA 3', H16065: 5' GGAGTCTTCAGTCTCTGGTTTACAAGAC 3'). PCR reactions consisted of 50 ng of DNA, 1x PCR buffer, 3mM MgCl₂, 200µM dNTPs, 1µM of each primer, and 1 U of *Taq* DNA Polymerase. Cycling parameters were 35 cycles of 30 seconds at 94° C, 30 seconds at 55° C, and 2 minutes at 72° C. PCR products were concentrated using a QIAquick PCR purification kit (Qiagen), then sequenced on an ABI 377 DNA sequencer. We aligned the sequences manually and edited them using Sequence Navigator version 1.0.1 (Applied Biosystems). All unique sequences were deposited in Genbank (acquisition numbers: AF297237 to AF297276).

AFLP Analysis

AFLP markers were generated using the procedure of Vos et al. (1995) with modifications as described in Busch et al. (2000). Adenine was used as the selective nucleotide in the first amplification; while ACG/AGG, ACG/AGC, ACG/ACC, ACG/ACA, ACG/ACG, ACG/AAG (*EcoRI/MseI* primers, respectively) were used for the second selective amplification. Polymorphic AFLP markers were manually scored, with only distinct and unambiguous polymorphic markers that conformed to the 95% polymorphic rule being used for this study (Hartl and Clark 1997). A set of 20 individuals was blindly replicated and indicated 0% error in scoring and marker consistency of those loci used in this study.

Statistical Analysis

Individuals were grouped by their respective breeding site. Due to small number of individuals sampled at some sites, we grouped some breeding locations if they were within 30 km of one another and greater than 100 km from any other site. Determination of distances to group samples was based on banding studies in central Arizona which documented that flycatcher movements between breeding sites were frequent within 30–40 km but became increasingly rarer with increasing distance (Paxton et al. 2007).

To understand the genetic structure among breeding sites, we calculated pairwise F_{ST} genetic distance measures using Nei's (1972) minimum distance algorithm in TFPGA (Miller 1997), and graphically represented them using a UPGMA cluster analysis. To compare the degree of genetic separation between the two subspecies based on the currently recognized boundary and the one suggested from the genetic distances, we used an analysis of molecular variance (AMOVA) to obtain estimates of genetic variation explained and Θ_{ST} estimates (analogous to F_{ST} values, measuring genetic structure among groups; Excoffier et al. 1992), calculated using Arlequin v 2.0 (Schneider et al. 1997).

Mantel tests were run in TFPGA using genetic distances from TFPGA, geographic distances from the Distance Matrix Extension in ArcView (v. 3.2, ESRI, Inc.), and Euclidean distances for elevation and latitude.

We modeled the frequency of C-group haplotypes (a group of cytochrome-b haplotypes closely associated with *E. t. extimus*; Paxton 2000, this study) observed at each breeding site as a function of the breeding site's latitude (decimal degrees) and elevation (m) using a binary logistic regression in SAS (v 9.1, SAS, Inc.). The goodness-of-fit of the logistic regression was evaluated by a ratio of the deviation to the degrees of freedom, with a ratio of 1.5 or less taken as assurance of good fit (Agresti 2002). We tested latitude, elevation, and an interaction of the two factors. Those variables with a WALD test significance of $P \leq 0.05$ were included in the model. To evaluate how well the breeding site frequency of C-group haplotypes reflects overall genetic differences among sites, we created in Arlequin vs 2.0 a pairwise genetic distance matrix of all sites compared to one another which considered all the cytochrome-b haplotypes and AFLP information. Using the genetic distance of each breeding site from the Roosevelt Lake breeding site (chosen to represent a core *E. t. extimus* breeding population with a relatively large sample size), we evaluated the correlation between the genetic distance from Roosevelt Lake and the predicted occurrence of a C-group haplotype for each site using a Pearson's Correlation in JMP vs 6.0 (SAS, Inc.).

The logistic regression model was brought into a GIS environment to create a "landscape" of C-group haplotype probability estimates across the four-corner states. Using Spatial Analysis' Raster Calculator in ArcGIS (v 9.0, ESRI, Inc.), we modified USGS DEM grids and a simple grid with latitude values (created by authors) to create the probability grid using the following equation derived from the logistic regression:

$$\{e^{17.4399-0.0007(\text{elevation})-0.4519(\text{latitude})}/(1+e^{17.4399-0.0007(\text{elevation})-0.4519(\text{latitude})})\}.$$

Using ArcView Spatial Analysis Contour function, we created 10% increment contour lines (10–90%) to visually simplify the image, and to create lines that represent candidate subspecies boundaries. To help determine which line or lines were the most appropriate for designating a subspecies boundary, we quantify how much genetic variation was separated by each probability contour line. We ran multiple AMOVAs in Arlequin for each grouping of sites (e.g., all sites north of the 10% line versus everything south of that line) to obtain F_{ST} estimates, and graphed their values by the probability contour in SigmaPlot 8.0 (SPSS, Inc.). Statistical significance was assumed at the $P < 0.05$ level.

Results

Genetic Structuring Between Subspecies

We found no fixed differences between the two subspecies from either cytochrome-b haplotypes (Table 2) or AFLP polymorphic loci that would allow us to unambiguously distinguish individuals of one subspecies from the other; however, we did observe strong frequency differences between the subspecies. Based on genetic relationships among the breeding sites, a UPGMA dendrogram suggested two major groups of breeding sites for both mtDNA and nDNA markers (Fig. 2).

With mtDNA, which provided genetic information for all breeding sites sampled, the two dendrogram groups are generally consistent with the currently accepted boundary between the subspecies, with all the core *E. t. extimus* sites clustering together and all the core *E. t. adastus* sites clustering together. However, three sites that are generally considered to be within the range of *E. t. extimus*, but located along the currently accepted boundary, were clustered with the “*E. t. adastus*” group (Fig. 2). Two of these sites (SEEG and SHIP; Table 1), were breeding sites consisting of a small number of individuals relatively isolated from other known sites. Therefore, the clustering of these latter two sites with core *E. t. adastus* breeding locations may be a function of small samples sizes ($n=2$ and 3 individuals at each site), and their affiliation with one subspecies or the other should be viewed with some caution. However, the third site (MCSP) had a relatively large sample size analyzed ($n=16$; Table 1), and is rooted deep in the *E. t. adastus*-associated group.

The nDNA dendrogram provided generally similar results, with core *E. t. adastus* sites grouping together, and core *E. t. extimus* sites grouping together (Fig. 2). Similar to the mtDNA dendrogram, the MCSP site was deeply rooted in the “*E. t. adastus*” group, while SEEG and two other sites which were located in northern Colorado and Utah (ARAP and STLA) were unresolved (basal to the two groups; Fig. 2). No nDNA information was available for the SHIP breeding site. Because the two dendrograms generally agreed, but only the mtDNA dendrogram included all breeding sites sampled, we used the mtDNA UPGMA as the best genetic grouping of breeding sites.

We used an AMOVA to compare the amount of genetic variation accounted for if we grouped breeding sites into either the currently recognized boundary or the dendrogram grouping. The UPGMA clustering of breeding locations received more support, with 19% of the variation in the mtDNA markers explained by the dendrogram grouping versus 11% for the traditional boundary, and 7% of nDNA variation explained for the dendrogram grouping versus 4% for the traditional boundary (Table 3). All four values were significantly different from 0% ($p<0.001$).

Although both the UPGMA clustering and AMOVA analysis indicated strong genetic frequency differences between the two subspecies, the geographic distribution of haplotypes suggested a region of introgression in the boundary region. For example, the C-group haplotypes (C1, C2, C3, C6) are dominate within the range of *E. t. extimus* (Table 2) but the proportion of individuals possessing a C-group haplotype changed from high frequencies at low latitudes within the core breeding range of *E. t. extimus* to low frequencies at the northern latitudes of the study area (Fig. 3). No C-group haplotypes were detected in previously sampled breeding sites further north in Idaho, Wyoming, and Montana (Paxton 2000). A Mantel test indicated a significant relationship between genetic and geographic distance when all sites were considered simultaneously, (mtDNA, $r=0.23$, $p=0.002$; nDNA, $r=0.30$, $p=0.007$); however, Mantel tests of genetic distance versus geographic distance indicated no relationship among breeding sites within the range of *E. t. extimus* (mtDNA, $r=0.05$, $p=0.32$; nDNA, $r=0.24$, $p=0.19$) or *E. t. adastus* (mtDNA, $r=0.0$, $p=0.78$; nDNA, $r=0.03$, $p=0.39$). These results suggest that the relationship between genetic and geographic distance is not due to a gradual frequency cline but rather a rapid change in frequencies at the boundary.

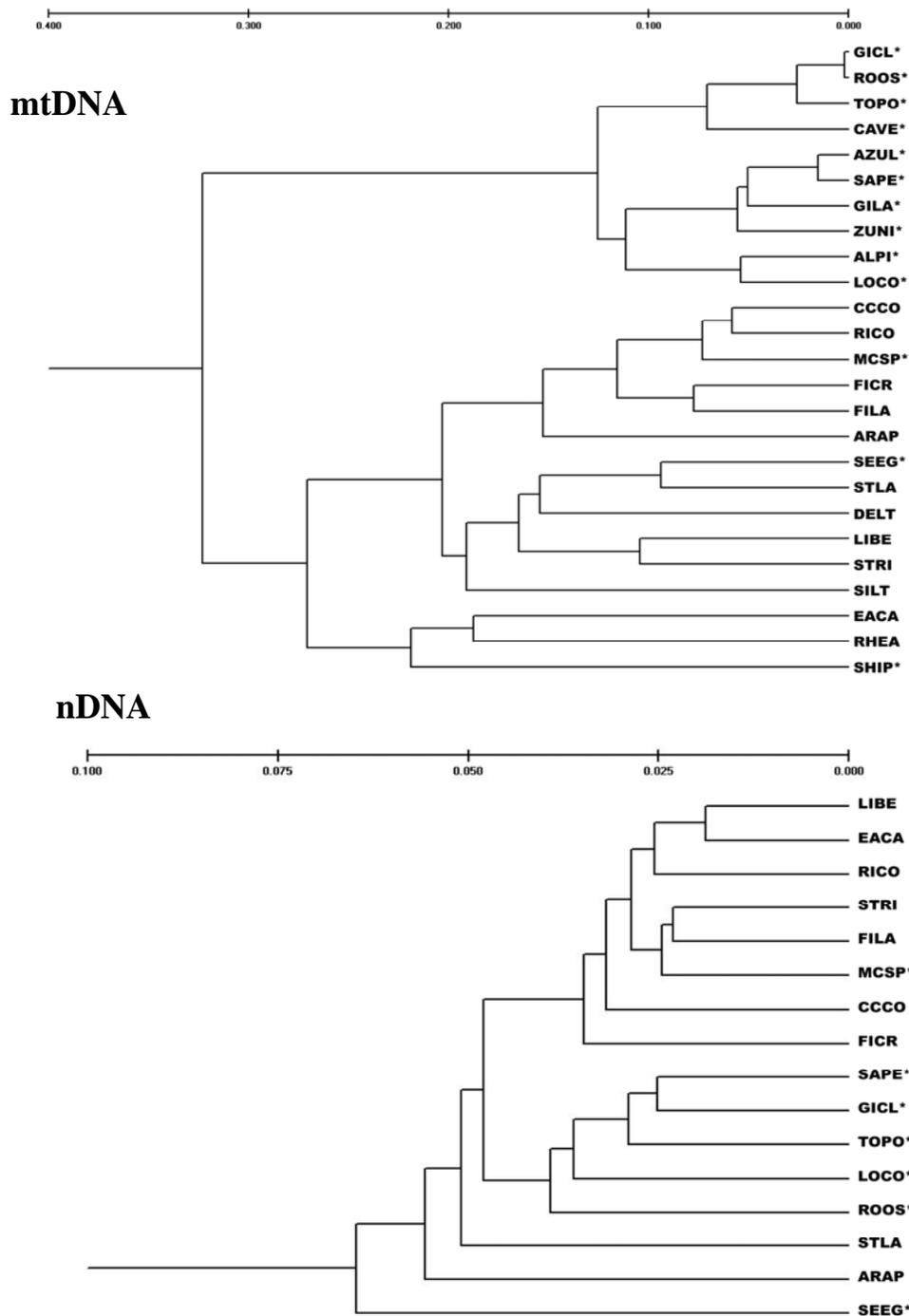


Figure 2. Willow Flycatcher breeding sites clustered out into two groups using a UPGMA analysis of genetic distances, based on mitochondrial DNA (cytochrome-b sequences; top) and nuclear DNA (AFLP; bottom). Four letter alpha codes designate breeding sites (see Table 1), with sites currently considered *E. t. extimus* followed by an asterisk.

Table 3. Analysis of molecular variance (AMOVA) results indicate that a significant amount of variation in mtDNA and nDNA frequencies is explained by grouping breeding sites into the two UPGMA clades (Fig. 2). An AMOVA partitions genetic variation among individuals, breeding sites, by grouping the individuals into the three categories and subspecies, with those percentages interpreted as the percent of genetic variation explained.

Source of variation	d.f.	Current boundary			UPGMA cluster		
		Sum of squares	Variance components	Explained variation (%)	Sum of squares	Variance components	Explained variation (%)
Mitochondrial DNA							
Subspecies	1	9.894	0.115	10.63	15.966	0.211	18.63
Breeding sites	23	31.909	0.089	8.24	25.837	0.043	3.79
Individuals	121	106.368	0.879	81.13	106.368	0.879	77.58
Total	145	148.171	1.084		148.171	1.133	
Nuclear DNA							
Subspecies	1	51.426	0.583	4.36	68.616	0.975	7.18
Breeding sites	14	248.994	0.821	6.14	231.805	0.649	4.78
Individuals	99	1183.736	11.957	89.49	1183.736	11.957	88.04
Total	114	1484.157	13.361		1484.157	13.581	

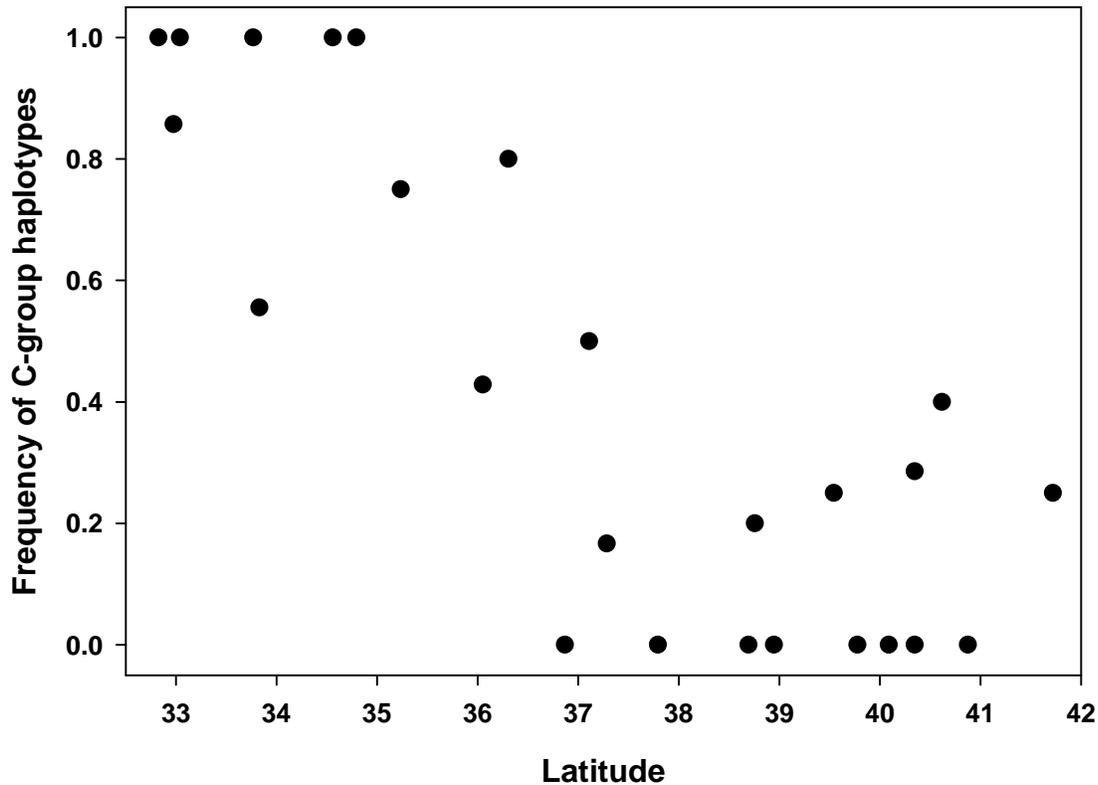


Figure 3. Frequency of C-group haplotypes (see Table 2) observed at breeding sites decreases with increasing latitude. Frequencies are high at low latitudes within the core range of *E. t. extimus*, and occur in low frequencies within the core range of *E. t. adastus* at high latitudes, with intermediate frequencies within the region between the subspecies.

Effects of Latitude and Elevation

The putative divide between the two subspecies lies within a geographic region of relative rapid transition from low elevation sites in the south, to higher elevation breeding sites in the north. To assess whether these landscape features (latitude and elevation) might influence genetic distances among sites, we ran Mantel tests of the pairwise genetic distances among sites and latitude and elevation. Latitude was significantly related to genetic distance for mtDNA ($r=0.43$, $p<0.001$), but only marginally with nDNA ($r=0.15$, $p=0.1$). Elevation was strongly associated with nDNA genetic distances ($r=0.35$, $p=0.002$), but weakly with mtDNA ($r=0.09$, $p=0.14$). Although the strength of the relationships varied between the variables and mtDNA and nDNA, we concluded that a relationship existed between latitude, elevation and geographic patterns of molecular markers, and that these landscape features could be informative for predicting a boundary.

Using the strong association of C-group haplotypes with the *E. t. extimus* subspecies, we built a logistic model evaluating the frequency of C-group haplotypes occurring at breeding sites with latitude and elevation as predictive factors. The model indicated that both latitude and

elevation were important explanatory factors, with no support for an interaction between the two ($p=0.78$). Latitude had the strongest effect ($\beta=-0.4519$, $p < 0.001$), with the odds of finding a C-group haplotype decreasing 36% with each increase in a degree of latitude. Elevation also contributed ($\beta=-0.0007$, $p=0.029$), with the odds of finding a C-group haplotype decreasing 50% with every thousand meter increase in elevation. The model fit the data well, with a goodness-of-fit ratio of 1.27. The logistic model predicted the probability of a C-group haplotype occurring at a particular site, given a specific latitude and elevation, and should reflect gene flow as influenced by latitude and elevation. To evaluate this basic assumption, we evaluated the correlation between the predicted probability of C-group occurrence for each breeding location (based on the logistic model; Table 1) with each location's genetic distance from Roosevelt Lake, a breeding site chosen for comparison because it is within the core-range of *E. t. extimus* and was well sampled. The resulting correlations (mtDNA: -0.80 , $P < 0.001$; nDNA: -0.71 , $P=0.003$) supported this general relationship, indicating that as the predicted probability of a C-group haplotype occurring at a site decreases, its observed genetic distance from a core *E. t. extimus* site (Roosevelt Lake) increases.

We incorporated the logistic model into a GIS environment, projecting a complex "landscape" of C-group probability values across this topographically rich region (graphically shown in 10% probability contours; Fig. 4). The currently recognized boundary lies primarily between the 20% and 40% probability contour lines, while the genetically-derived dendrogram groups were between the 30% and 40% contour lines. Each 10% probability contour divides different groups of breeding sites; to evaluate how much genetic variability is explained by different groupings of breeding sites, we grouped sites north and south of each probability line (10–80%) and obtained measures of genetic distance (F_{ST}) for both mtDNA and nDNA per grouping. Genetic distance, based on mitochondrial DNA, had a strong inflection point at the 30% contour line (Fig. 5), suggesting that the line of greatest geographic-genetic separation lies at or near this probability contour. Partitioning nuclear DNA information in a similar way was less informative because multiple breeding sites from the greater boundary region were not represented, but genetic structuring appears to increase moving toward the boundary region (Fig. 5).

Discussion

The geographic distribution of Willow Flycatcher mitochondrial and nuclear molecular genetic markers within the study region suggests two distinct groups (subspecies) based on strong frequency differences, with the two groups geographically separated by a region roughly overlapping the currently recognized boundary. However, the region that separates the two subspecies shows evidence of intergradation, such as the frequency change in cytochrome-b haplotypes moving south to north across the boundary region. Thus, the subspecies boundary does not appear to be distinct, but rather a region of intergradation approximately centered along the southern areas of Utah and Colorado, and the northern areas of Arizona and New Mexico. These findings, based on molecular genetic information, are consistent with morphological (e.g., Unitt 1987, Browning 1993) and song-based studies (Sedgwick 2001) which also found evidence for an intergradation zone. It is not uncommon or unexpected to find some degree of introgression along the shared boundary of two subspecies (Mayer 1982, O'Brien and Mayr 1991, Crandall et al. 2000), and different genes may show different patterns of geographic distribution in such an intergrade region (Latta 2004).

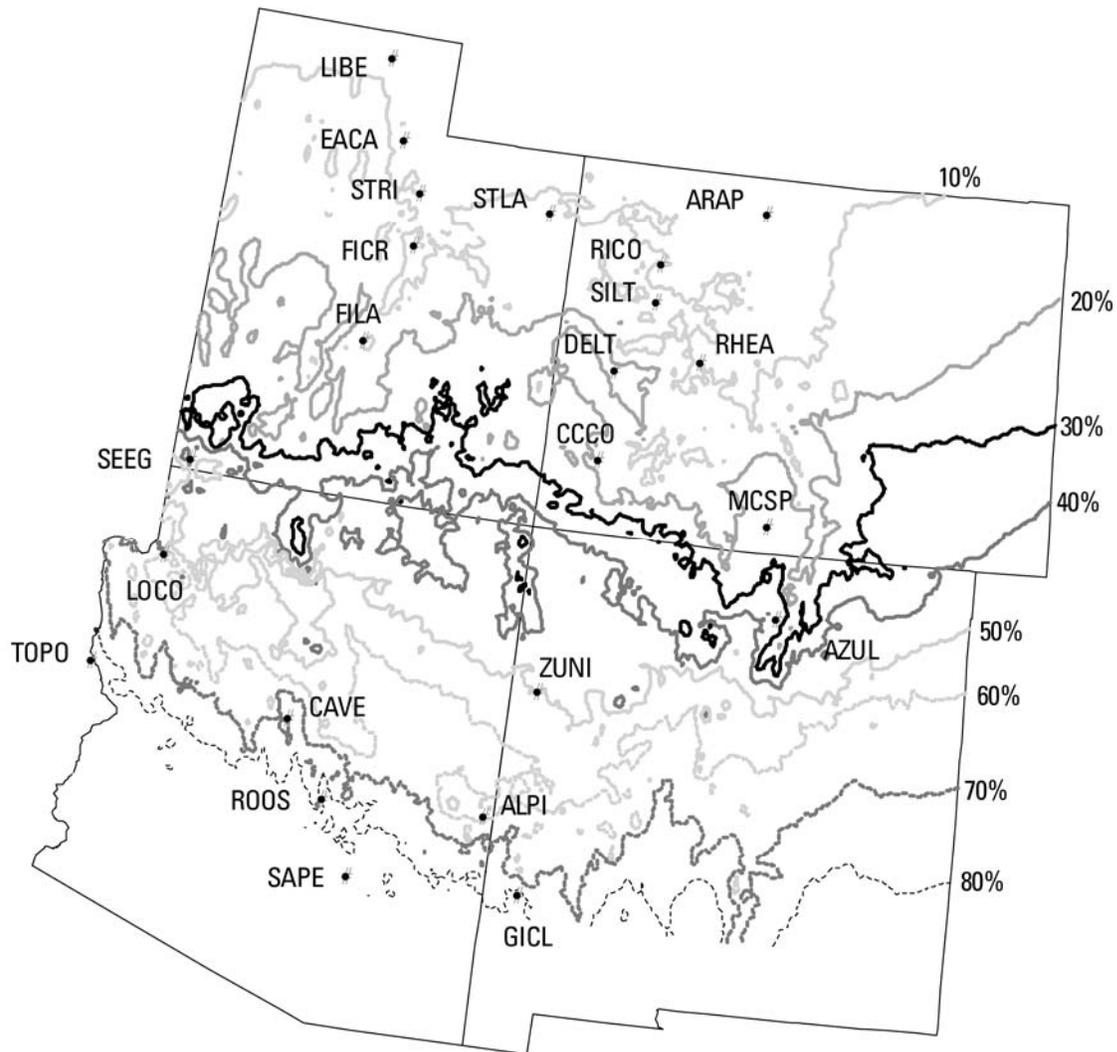


Figure 4. Contours indicating the predicted probability of a breeding site having a C-group haplotype based on latitude and elevation (shown in 10% increment contours). Each contour line represents 10% increment probabilities of encountering a C-group haplotype from 10–90%. The exact probabilities associated with each site are given in Table 1.

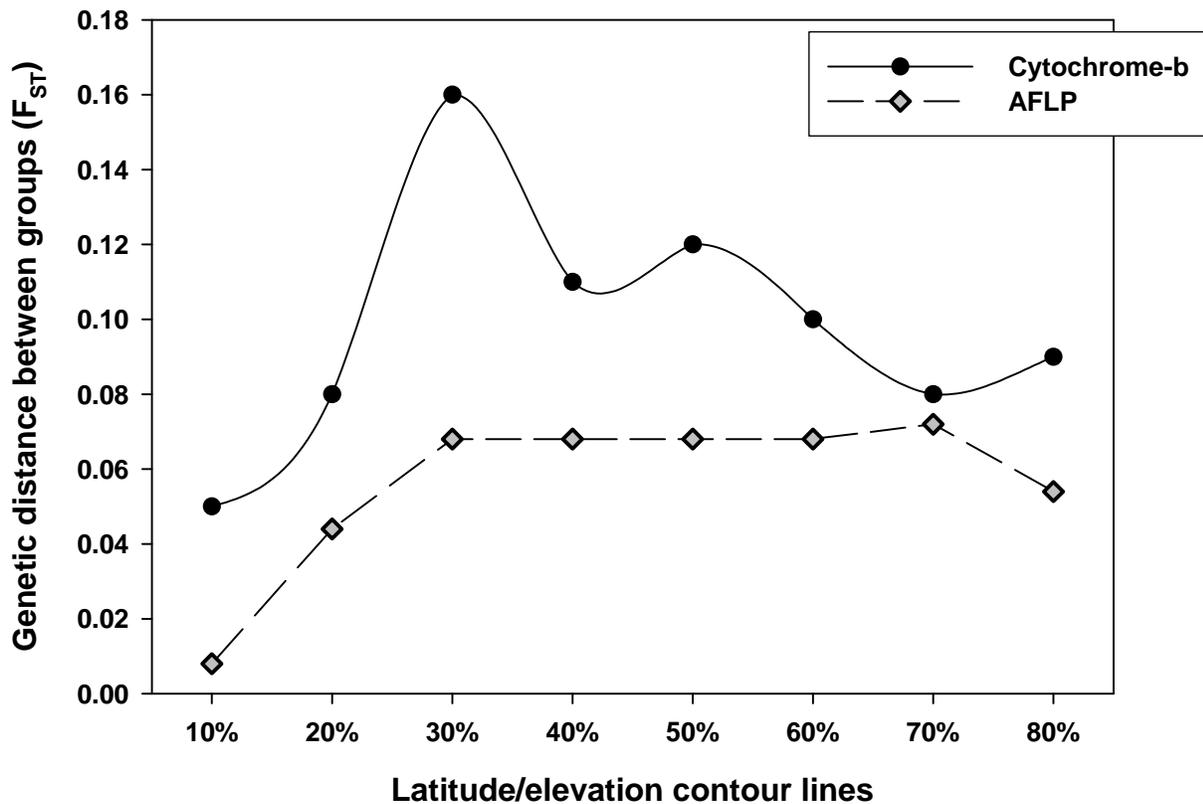


Figure 5. The amount of genetic information that each C-group probability line separates is highest at the 30% contour line. For each 10%-probability contour line (see Fig. 4), sites were grouped by their geographic placement on either side of the dividing line, and a measure of genetic distance (F_{ST}) was calculated for each grouping and plotted, for both cytochrome-b sequences (mtDNA) and AFLP data (nDNA). AFLP contour values from 30–60% are the same due to no change in the division of sites.

Effects of Latitude and Elevation on Distribution of Genetic Markers

Landscape features such as elevation and latitude have long been recognized as shaping the distribution of bird communities (Morton 2002, Hawkins and Diniz-Filho 2006). In particular, landscape features can influence gene flow between populations, thus influencing the location of a boundary between groups such as subspecies (Manel et al. 2003, Storfer et al. 2007). Landscape-scale habitat differences are believed to be important drivers of speciation within the genus *Empidonax*, with subspecies of multiple *Empidonax* species showing strong habitat differences (Johnson and Cicero 2002). For the Willow Flycatcher specifically, mountain ranges (Cascade, Sierra Nevada, and Rocky Mountains) form the boundaries between the other Willow Flycatcher subspecies, and the biotic changes due to latitude and elevation at the northern range of *E. t. extimus* potentially form an ecological barrier that inhibits geneflow between the two subspecies. Much of the breeding range of *E. t. extimus* is characterized by riparian habitat within low elevation desert habitats, whereas the habitat of *E. t. adastus* is characterized primarily by riparian habitat in more mesic, higher altitude locations. The Willow Flycatcher breeding sites formed two genetic

groups separated geographically within the region of change between these two types of habitat. Additionally, the relationship between genetic distance among sites and differences in latitude and elevation also suggest that latitude and elevation co-vary with genetic patterns across this area, suggesting these geographic features either directly influence, or are associated with other factors that influence, gene flow in Willow Flycatchers. This pattern is consistent with the idea that animal populations spanning altitudinal extremes may show genetically-based physiological adaptations to elevation that could lead to selection for locally adapted subpopulations (Storz and Dubach 2004), as well as influence of habitat selection on dispersal patterns (Kisdi 2002).

Inclusion of landscape features such as latitude and elevation in determining the location of the northern boundary of *E. t. extimus* helps to overcome two difficulties of this study: (1) the boundary is not discrete, but rather a region of intergradation, and (2) the boundary region is sparsely populated providing minimal information that would help narrow down the location of the boundary. If the landscape features form an ecological barrier across which gene flow is inhibited, then identifying those features that are driving the ecological barrier can help inform where a divide between the subspecies occurs. Additionally, landscape-informed boundaries can be used to predict genetic patterns that might be expected in areas where there are no longer extant populations. For example, the Colorado River running through southeastern Utah represents a low elevation corridor extending northward that once supported breeding populations of flycatchers, from which at least one study skin collected in Glen Canyon (now under Lake Powell) was classified as *E. t. extimus* based on morphological characters (Unitt 1987). However, despite extensive surveys, no breeding populations have been identified from this area in recent times (USFWS 2002), making decisions about where a boundary line may lie in this area exceedingly difficult. Our model suggests that low elevation sites in this area would have been more likely to have been occupied by *E. t. extimus*.

Where Should the Northern Boundary of *E. t. extimus* be Placed?

While both mtDNA and nDNA distinguished between two genetic lineages that roughly reflected current designations of *E. t. extimus* versus *E. t. adastus*, with the greatest amount of genetic differentiation near the 30% contour line, the evidence for a region of intergradation suggests there is little support biologically for the legal necessity of designating a boundary line. As a result, designating the location of a discrete line separating the two subspecies cannot be made unambiguously based on genetic data and thus becomes a value judgment dependent on the philosophy of policy makers. Yet resolution of this issue is important for protection and conservation of *E. t. extimus*, because its listing was based on subspecies status. We describe three alternative approaches that could be taken in making such a decision based on the findings of this study:

1. Place the boundary in the zone of intergradation at a point that maximizes the degree of genetic differentiation based on our analysis of genetic markers. This line would best correspond with the 30% probability contour (Fig. 4) in the case of the GIS model used in this study.
2. If management priorities instead favor protecting as much of the *E. t. extimus* genetic stock as possible, then the boundary line should be drawn farther north, and at higher elevations (i.e., contour lines <30%). How much further north would be determined based on the degree of priority for protecting *E. t. extimus* genes in areas that become increasingly dominated by *E. t. adastus* genes.
3. Alternatively, management priorities may favor protecting only those breeding sites that are dominated by *E. t. extimus* genes, and therefore would place the range boundary further south

and at lower elevations than the 30% probability line (i.e., contour lines >30%). Again, the distance beyond this line would be related to the level of stringency one placed on designating a population as *E. t. extimus*.

Because there is no biological basis for a distinct boundary line, the location of a boundary line ultimately becomes a policy and regulatory decision, ideally one that integrates the best available science, taxonomic principles, policy objectives of the responsible agencies, and an examination of the intent of the Endangered Species Act (Haig et al. 2006).

An alternative to designating a specific geographic line would be to manage the boundary region as a special status area. Biologically, the boundary region is a dynamic area that is unlikely to be static over time with respect to the distribution of genetic lineages (Rohwer et al. 2000) and in face of distributional changes of many bird species brought on by climatic changes (Johnson 1994). Through the process of geneflow, especially driven by the extirpation and colonization of new breeding sites, it is likely that the affinity of breeding sites to one or the other of the two subspecies may change over time, with the region likely widening and contracting over time. The boundary region is an area where there is the exchange of genes between the two subspecies, and may be an important area of evolution (Barrowclough 1980, Crandall et al. 2000, Haig et al. 2006).

Conclusions

This study examined the possible location of the northern boundary of the Southwestern Willow Flycatcher (*E. t. extimus*) in the four-corner states. The genetic markers indicated strong mitochondrial and nuclear DNA frequency differences among subspecies, with breeding sites clustering into two groups separated approximately along the currently recognized boundary; however, the geographic pattern of the molecular markers indicated that a distinct genetic boundary line between the subspecies does not exist. Thus, the boundary between the two subspecies should be thought of as a region of genetic overlap as previous work based on museum skins and song variation also suggested. Our GIS model, which used latitude and elevation to produce a number of potential boundary lines based on the geographic distribution of mtDNA haplotypes, indicated that the line that separates the greatest amount of genetic information among our sample sites is the 30% probability contour. Although this line approximates the currently recognized management boundary, policy and management goals may favor designation of a boundary line farther north or south of this approximate mid-point. Alternatively, management of the region of subspecies intergradation as a special management unit in itself may more closely reflect the biological reality of the subspecies distributions. Given that the molecular genetic data suggests that there is no biological basis for a distinct boundary, the final decision of where to place the boundary, for the purpose of the Endangered Species Act management, will ultimately be a policy-based choice.

Acknowledgements

We thank the many persons, agencies, and private organizations and companies that helped in coordination, sharing of information, and land access permission. Funding was provided by the U.S. Fish and Wildlife Service, U.S. Geological Survey, and Northern Arizona University. We greatly appreciate the hard work of the USGS field biologists who collected samples over the years. This manuscript has been greatly improved through reviews of earlier drafts by Karen Mock, Kristina Paxton, and Dustin Wood.

References

- Agresti, A. 2002. Categorical data analysis, second edition. Wiley-Interscience, John Wiley and Sons, Hoboken, New Jersey.
- Avise, J. C. 1994. Molecular Markers, Natural History and Evolution. Chapman and Hall, New York, NY.
- Barrowclough, G. F. 1980. Genetic and phenotypic differentiation in wood warbler (genus *Dendroica*) hybrid zone. *Auk* 97:655–668.
- Barrowclough, G. F., J. G. Groth, L. A. Mertz, and J. Gutierrez. 2005. Genetic structure, introgression, and a narrow hybrid zone between northern and California spotted owls (*Strix occidentalis*). *Molecular Ecology* 14:1109–1120.
- Behle, W. H. 1985. Utah birds: Geographic distributions and systematics. Utah Museum of Natural History Occasional Publications 5. 147 pp.
- Browning, M. R. 1993. Comments on the taxonomy of *Empidonax traillii* (Willow Flycatcher). *Western Birds* 24:241–257.
- Burleigh, T. D. 1972. The birds of Idaho. Caxton, Caldwell, Idaho.
- Busch, J. D., M. P. Miller, E. H. Paxton, M. K. Sogge, and P. Keim. 2000. Genetic variation in the endangered southwestern Willow Flycatcher. *Auk* 117:586–595.
- Crandall, K. A., O. R. P. Bininda-Emonds, G. M. Mace, and R. K. Wayne. 2000. Considering evolutionary processes in conservation biology. *Trends in Evolution and Ecology* 15:290–295.
- DeGraaf, R. M. and J. H. Rappole. 1995. Neotropical migratory birds: Natural history, distribution, and population change. Comstock Publishing Associates, Ithaca, New York.
- Durst, Scott L., Sogge, Mark K., Stump, Shay D., Williams, Sartor O., Kus, Barbara E., and Sferra, Susan J., 2007, Southwestern Willow Flycatcher breeding site and territory summary; 2006: U.S. Geological Survey Open-File Report 2007–1391, 28 p. [<http://pubs.usgs.gov/of/2007/1391/>].
- Excoffier, L., P. E. Smouse, and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* 131:479–491.
- Haig, S. M., E. A. Beever, S. M. Chambers, H. M. Draheim, B. D. Dugger, S. Dunham, E. Elliot-Smith, J. B. Fontaine, D. C. Kesler, B. J. Knaus, I. F. Lopes, P. Locshl, T. D. Mullins, and L. M. Sheffield. 2006. Taxonomic considerations in listing subspecies under the U.S. Endangered Species Act. *Conservation Biology* 20:1584–1594.
- Hartl, D. L., A. G. Clark. 1997. Principles of population genetics. 3rd ed. Sinauer Associates, Massachusetts.
- Hawkins, B. A., and J. A. F. Diniz-Filho. 2006. Beyond Rapoport's rule: evaluating range size patterns of New World birds in a two-dimensional framework. *Global Ecology and Biogeography* 15:461–469.
- Helm-Bychowski, K. and J. Cracraft. 1993. Recovering phylogenetic signal from DNA sequences: Relationships within the Corvine assemblages (Class Aves) as inferred from complete sequences of the mitochondrial DNA *cytochrome-b* gene. *Molecular Biology and Evolution* 10:1196–1214.

- Johnson, N. K. 1994. Pioneering and natural expansion of breeding distributions in western North American birds. *Studies in Avian Biology* 15:27–44.
- Johnson, N. K. and C. Cicero. 2002. The role of ecological diversification in sibling speciation of *Empidonax* flycatchers (Tyrannidae): multigene evidence from mtDNA. *Molecular Ecology* 11:2065–2081.
- Latta, R. G. 2004. Relating process to patterns of genetic variation across landscapes. *Forest Ecology and Management* 197:91–102.
- Kisdi, E. 2002. Dispersal: Risk spreading versus local adaptation. *American Naturalist* 159:579–596.
- Kroodsma, D. E. 1984. Songs of the Alder Flycatcher (*Empidonax alnorum*) and Willow Flycatcher (*Empidonax traillii*) are innate. *Auk* 101:13–24.
- Manel, S., M. K. Schwartz, G. Luikart, and P. Taberlet. 2003. Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology and Evolution* 18:189–197.
- Mayr, E. 1982. Of what use are subspecies? *Auk* 99:593–595.
- Miller, M. P. 1997. Tools for Population Genetic Analyses (TFPGA) version 1.3: A windows program for the analysis of allozyme and molecular population genetic data. Computer software distributed by author.
- Moritz, C. 1994. Applications of mitochondrial DNA analysis in conservation: A critical Review. *Molecular Ecology* 3:401–411.
- Morton, M. 2002. The Mountain White-crowned Sparrow: migration and reproduction at high altitude. *Studies in Avian Biology* 24:1–236.
- Mullenbach, R., J. P. L. Lagoda, and C. Welter. 1989. An efficient salt-chloroform extraction of DNA from blood and tissue. *Trends in Genetics* 5:391.
- Nei, M. 1972. Genetic distance between populations. *American Naturalist* 106:283–292.
- O'Brien, S. J. and E. Mayr. 1991. Bureaucratic mischief: recognizing endangered species and subspecies. *Science* 251:1187–1188.
- Owen, J. C. and M. K. Sogge. 1997. Banding and genetic sampling of Willow Flycatchers in Colorado – 1996 and 1997 summary report. USGS Colorado Plateau Field Station Report.
- Paxton, E. H. 2000. Molecular genetic structuring and demographic history of the Willow Flycatcher (*Empidonax traillii*). MS thesis, Northern Arizona University, Flagstaff, Arizona.
- Paxton, E.H., M.K. Sogge, S.L. Durst, T.C. Theimer, and J.R. Hatten. 2007. The Ecology of the Southwestern Willow Flycatcher in central Arizona: A 10-year Synthesis Report. USGS Open File Report 2007–1381.
- Rohwer, S., E. Bermingham, and C. Wood. 2001. Plumage and mitochondrial DNA haplotype variation across a moving hybrid zone. *Evolution* 55:405–422.
- Schneider, S., J. M. Kueffer, D. Rosessli, and L. Excoffier. 1997. Arlequin. [Http://anthropologie.unige.ch/arlequin](http://anthropologie.unige.ch/arlequin), University of Geneva, Geneva, Switzerland.
- Sedgwick, J. A. 2001. Geographic variation in the song of Willow Flycatchers: Differentiation between *Empidonax traillii adastus* and *E. t. extimus*. *Auk* 118:366–379.

- Sogge, M. K., R. M. Marshall, T. J. Tibbitts and S. J. Sferra. 1997. A Southwestern Willow Flycatcher Natural History Summary and Survey Protocol. National Park Service Technical Report NPS/NAUCPRS/NRTR-97/12.
- Sogge, M. K., J. C. Owen, E. H. Paxton, S. M. Langridge and T. J. Koronkiewicz. 2001. A Targeted Mist Net Capture Technique for the Willow Flycatcher. *Western Birds* 32:167–172.
- Storfer, A., M. A. Murphy, J. S. Evans, C. S. Goldberg, S. Robinson, S. F. Spear, R. Dezzani, E. Delmelle, L. Vierling, and L. P. Waits. 2007. Putting the ‘landscape’ in landscape genetics. *Heredity* 98:128–142.
- Storz, J. F., and J. M. Dubach. 2004. Natural selection drives altitudinal divergence at the albumin locus in deer mice, *Peromyscus maniculatus*. *Evolution* 58:1342–1352.
- United States Fish and Wildlife Service. 1995. Final rule determining endangered status for the southwestern Willow Flycatcher. *Federal Register* 60:10694–10715.
- United States Fish and Wildlife Service. 2002. Southwestern Willow Flycatcher Recovery Plan. U.S. Fish and Wildlife Service Region 2, Albuquerque, NM.
- Unitt, P. 1987. *Empidonax traillii extimus*: an endangered subspecies. *Western Birds* 18:137–162.
- Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* 23:4407–4414.

This page intentionally left blank

This page intentionally left blank

Paxton and others—Using Molecular Genetic Markers to Resolve a Subspecies Boundary: the Northern Boundary of the Southwestern Willow Flycatcher in the Four-corner States—Open-File Report 2008-1117