

State of California
The Resources Agency
Department of Fish and Game
Wildlife Branch

and

Audubon California
4225 Hollis Street
Emeryville, CA 94608

**Population structure of the Tricolored Blackbird (*Agelaius tricolor*)
in California: are northern and southern populations genetically
distinct?**

by

Dr. Elena C. Berg
Dr. John P. Pollinger
Dr. Thomas B. Smith
Center for Tropical Research
University of California, Los Angeles
619 Charles Young Drive South
Los Angeles, CA 90095

**Population structure of the Tricolored Blackbird (*Agelaius tricolor*) in California:
are northern and southern populations genetically distinct?**

Final Report to:

California Department of Fish and Game
Wildlife Branch, Nongame Wildlife Program
1812 Ninth Street
Sacramento, California 95811

Audubon California
4225 Hollis Street
Emeryville, CA 94608

Los Angeles Audubon Society
PO Box 931057
Los Angeles CA 90093-1057

Pomona Valley Audubon Society
W.M. Keck Science Center
925 North Mills Avenue
Claremont, CA 91711

Sea and Sage Audubon Society
P.O. Box 5447
Irvine, CA 92616-5447

CDFG Agreement # P0780028
Supported by State Wildlife Grant T-13-1; R0685021

Contractor:

Center for Tropical Research
University of California, Los Angeles

Principal Investigator: Dr. Thomas B. Smith
Project Scientists: Dr. Elena C. Berg, Dr. John P. Pollinger

Suggested Citation: Berg, E.C., Pollinger, J.P. and Smith, T.B. 2010. Population structure of the Tricolored Blackbird (*Agelaius tricolor*) in California: are northern and southern populations genetically distinct? Calif. Dept. Fish and Game, Nongame Wildlife Program Rpt. 2010-05 and Audubon California, Sacramento, CA. 25 pp.

INTRODUCTION

The Tricolored Blackbird (*Agelaius tricolor*; TRBL) is a colonial breeder that forms the largest colonies of any North American passerine bird. It is one of only two bird species endemic to the Central Valley and adjacent coastal California. More than 99% of TRBLs live in California, with just a few scattered populations in Oregon, Washington, coastal Baja California, Mexico and a single breeding colony in western Nevada (Beedy and Hamilton 1999). Surveys have estimated that the abundance of this species has declined by at least 70% since the 1930's, mainly as a result of the destruction and fragmentation of wetland habitat and subsequent conversion of these lands to human settlements and agriculture (Beedy and Hamilton 1999). Owing to its restricted distribution as well as declining population trend it is a species of great conservation importance and is listed as a California Bird Species of Special Concern (Shuford and Gardali 2008).

Although most of the largest TRBL colonies are found in the San Joaquin and Sacramento valleys, the species also breeds in several southern California counties, where a century ago, it was considered to be the most abundant bird species (Baird in Cooper 1870). The status of the southern California TRBL population is of great concern as both the number of breeding colonies and the number of breeding birds have shown steep declines. The late W. J. Hamilton (pers. comm.) documented behavioral differences between Central Valley and southern populations, and he suggested that the Tehachapi Mountains in southern California might serve as a dispersal barrier for the species. The banding evidence provided by DeHaven and Neff (1973), who with their collaborators banded many thousands of birds on both sides of the Tehachapis, also suggests an absence of interchange of individuals, with the Tehachapis serving as a dispersal barrier. Hamilton wondered whether the northern and southern TRBLs might even represent two different subspecies. We know very little about the ecology of southern California TRBLs, and nothing at all about their population genetics. DNA analyses may be useful to determine whether there are genetic differences not only between southern California and the Central Valley but also among sites within southern California.

Effective management of the TRBL would benefit from a better knowledge of both the distribution and connectedness of populations across the state. Without knowing how populations are connected (both demographically and genetically), it is impossible to identify the spatial scale at which they should be managed. Some remnant populations may represent unique genetic units and be high conservation priorities; alternatively, small populations at risk of extirpation may be genetically similar to larger, stable populations elsewhere and therefore represent lower conservation priority. Because demographically independent populations should be managed as separate units, accurate identification of these units is essential.

Molecular genetic techniques have been extremely useful in identifying population units for management and have provided essential data for making informed conservation decisions. Recently, our lab used two complementary molecular markers, nuclear DNA microsatellites and mitochondrial DNA sequences, to examine the genetic structure of

seven colonies of TRBLs in the Central Valley. Our microsatellite analysis showed no evidence for any population structuring within the seven sample areas. There was also no correlation between genetic distance and geographic distance. Similarly, our sequencing analyses revealed low overall genetic diversity and no evidence for population structuring. These results suggest that at least on a genetic level, the Central Valley colonies represent a single population. However, we still know nothing about the genetic structure of southern California populations or their relationship to the colonies north of the Tehachapi Pass.

Information on the genetic structure of southern populations is critical for several reasons. First, it will tell us whether there is genetic support for Hamilton's suggestion that there is little or no gene flow between northern and southern populations. If northern and southern populations are indeed genetically distinct, then even more urgent efforts will be necessary to protect the small and ever-declining colonies of southern California. Second, evidence for further population structuring within southern populations (as well as between north and south) would indicate that southern California populations are behaving quite differently from the genetically homogenous populations of the Central Valley. We studied the genetic structure within this species to help evaluate whether there is a single genetically homogenous population to conserve, or instead whether there are indeed two (or more) distinct populations that should be managed separately.

METHODS

In April, 2008, we obtained tissue samples of southern California TRBLs from abandoned dead nestlings collected in spring 2006 from western Riverside County, currently housed in the collection of the San Bernardino County Museum, and abandoned dead nestling samples collected by the California Department of Fish and Game (CDFG) from San Diego County in spring 2007.

Between April and June, 2008, we obtained tissue samples of southern California TRBLs by collection of dead nestlings, dead adults, eggs and feathers from recently-abandoned colonies.

Between June and December, 2008, labwork focused on DNA microsatellite and mitochondrial sequence analysis of approximately 169 TRBL samples. Specifically, we genotyped 10 microsatellites and sequenced the mitochondrial ATPase 6 and 8 genes.

Analysis of southern California TRBL population structure was carried out using various genetic analysis software methods including Bayesian cluster analysis, F statistics, and allele-sharing distance trees for microsatellite data and minimum spanning networks and phylogenetic trees for mitochondrial sequence data. We quantified and contrasted population structure of the southern California population(s) with the previously analyzed Central Valley population and estimated migration rates between the two.

SAMPLE COLLECTION

We first identified recently collected samples of southern California populations of TRBL. Tom Paulek (CDFG, retired) collected 90 carcasses of brood-reduced nestlings from the perimeter of a large TRBL colony at Ramona Farms in May 2006. Ramona Farms is located in the San Jacinto Valley, Riverside County. These samples were archived at the San Bernardino County Museum and we received permission and assistance from biological collections manager Gerald Braden to sample frozen tissue from the dead nestlings in their collection. In addition, Paulek and Susan Nash had performed a post-abandonment survey of the 2007 TRBL colony nesting site at the Rancho Jamul Ecological Reserve in San Diego County in July 2007 and collected 15 dead nestlings, which were provided to UCLA for archiving/testing.

During the spring and early summer of 2008, we coordinated with the 2008 statewide TRBL statewide survey (statewide coordinator Rodd Kelsey southern California coordinator Jon Feenstra, (both of Audubon California) to identify and monitor nesting colonies in southern California suitable for dead nestling/adult and feather salvage after nesting site abandonment for use in this study. There were very few nesting sites of any size identified in southern California during 2008 suitable for post-abandonment sampling (Kelsey 2008). As we already had samples from San Diego County (2007) and Riverside County (2006) representing the southernmost and central/eastern distributions of the known extant TRBL distribution in southern California, we targeted northern Los Angeles County and western San Bernardino County in 2008 for possible suitable sized nesting colonies to survey the rest of their known southern California distribution. Two suitably large nesting colonies were located at Holiday Lake Reservoir, West Valley County Water District, in northern Los Angeles County (estimated population size of 550 individuals) and Newberry Springs near the Hanson Gravel Mine facility in San Bernardino County (estimated population size of 500 individuals) (Figure 1).

With the help of Paulek and Nash, we monitored the sites. Allison Alvarado of our lab worked with Paulek and Nash to collect material from the colony at Holiday Lake after the birds departed post-breeding in late June. Paulek and Nash collected material at the Newberry Springs site after it was abandoned in early July. At each site we collected feathers, unhatched eggs, dead nestlings, and dead adults from nest sites. We were ultimately able to collect 22 specimens at Holiday Lake and 42 at Newberry Springs suitable for genetic testing (Table 1).

LABORATORY METHODS AND ANALYSES

After extracting DNA from a subset of samples using Qiagen QIAmp DNA Mini Kits (Qiagen, Inc., Valencia, California, U.S.A.), we analyzed both nuclear microsatellites and mitochondrial DNA sequences. We first evaluated DNA amplification from each extracted sample using both microsatellite markers and the mtDNA ATPase segment. Since many of the samples were dead nestlings, abandoned eggs or feathers with probable DNA degradation, we were able to robustly amplify microsatellites and mtDNA sequence from only 13 of the 22 Holiday Lake samples, and 12 of the 15 Rancho Jamul

samples due to their degree of degradation. We were able to successfully amplify DNA from all samples from the other two sites.

Microsatellites

We used microsatellites to detect differences among populations. Microsatellites evolve rapidly and are highly variable; therefore, they are effective at determining the amount of gene flow among populations. All samples were screened at 10 loci (Table 2).

All PCR reactions were carried out on an MJ Research PC-200 Thermocycler. PCR was performed in 10 μ L reactions with 1.5 μ L template DNA, 5.0 μ L Qiagen Multiplex Mastermix (contains a mixture of *Taq*, dNTPs, $MgCl_2$, and reaction buffer), 0.4 μ L BSA (at 10mg/ μ L concentration), and 0.02 μ L of each primer (10 μ M). The PCR profile was as follows: incubation at 95°C for 15 min; 12 cycles of amplification at 94°C for 30 s, 60°C-0.5°C per cycle for 90 s, and 72°C for 60 s; 33 cycles of 89°C for 30 s, 55°C for 90 s, 72° for 60 s; and a final extension step at 60°C for 30 min. The 30 minute final hold was performed to reduce reverse split peaks. One primer within each pair was labeled at the 5' end with one of two fluorescent dyes, 6-FAM or NED, each emitting a different wavelength. PCR products were diluted to optimize product signal and visualized on an Applied Biosystems (ABI, Foster City, California, U.S.A.) 3730XL automated genetic analyzer, using the ABI GS500 LIZ internal size standard. Electropherograms of each DNA fragment were analyzed using the ABI GeneMapper (version 3.7) fragment analysis software. In each run, at least five of the 96 samples were positive controls (i.e. samples for which allele sizes had been calculated on one or more previous runs). In each run, at least one of the samples was a negative control (distilled water) to test for contamination of the PCR product.

For all analyses, we included only those samples for which at least 75% of the loci that were run for that species were successfully scored. We used GENEPOP version 3.4 (Raymond and Rousset 1995) to test for significant departure from Hardy-Weinberg equilibrium and linkage disequilibrium. All loci that showed an excess of homozygotes were excluded from subsequent analyses. Before running analyses of population structure, we quantified genetic relatedness between all individuals using the program GENALEX (Version 6, Peakall and Smouse 2006), calculating the pairwise relatedness estimator R (Queller and Goodnight 1989). To ensure that we had sampled evenly across the sample populations and were not skewing our analyses towards highly related individuals, we plotted pairwise relatedness of all individuals and confirmed that the distribution was normal and centered around zero.

We employed two kinds of population assignments, one at a regional scale, and one at a finer scale. To examine relative genetic distance and structure between individuals, we used the program POPULATIONS (Version 1.2.28, Langella 1999) to construct neighbor-joining trees based on shared allele distance (DAS; Jin and Chakraborty 1993). Trees were visualized using the program TREEVIEW (Version 1.6.6, Page 1996). Next, using GENEPOP, we measured overall genetic distance and genetic distance between sample groups by calculating overall F_{ST} and pairwise F_{ST} values, respectively (Weir and Cockerham 1984). We corrected for multiple comparisons in all statistical tests using a

sequential Bonferroni adjustment (Rice 1989). Statistical significance ($p < 0.05$) of F_{ST} values was determined using the G-test as implemented in the program FSTAT (Version 2.9.3, Goudet 2001). In order to test for the null expectation of isolation by distance in the absence of fragmentation effects, we performed Mantel tests on log-transformed geographical and $F_{ST}/(1 - F_{ST})$ distance matrices (Mantel 1967; Rousset 1997).

To complement analyses based upon F_{ST} , we also implemented a Bayesian clustering method using STRUCTURE 2.3 (Pritchard et al. 2000) to infer the number of genetic clusters (K) without *a priori* assumptions about sample location. The stability of the inferred clusters was evaluated using three independent runs at $K = 1-11$ (the no. of sampled southern California TRBL populations (4) and previously characterized Central Valley TRBL populations (7)) with a burn-in period of 50,000 steps followed by 500,000 Markov chain Monte Carlo (MCMC) cycles. We ran STRUCTURE assuming correlated allele frequencies and admixture to account for possible gene flow and mixed ancestry within the population. We also used the locality option which uses the designation of independent populations (11 in this case), but not population location, as a prior for the Bayesian analyses to maximize detection sensitivity to small levels of population structure. We used the estimated log likelihood of the data $\ln P(D)$ and the Evanno ΔK (2005) statistic to determine the most best fit true number of clusters to use for subsequent analyses (e.g. population differentiation and migrant identification).

Using the program GENALEX (Version 6, Peakall and Smouse 2006), we also conducted spatial autocorrelation analyses (Smouse and Peakall 1999) to investigate whether individuals from populations that were located closer together in physical space were also more genetically similar. This genetic-distance-based, multivariate approach compares pairwise genetic distance derived from microsatellite data with geographical distance between individuals or populations. The program estimates an autocorrelation coefficient (r) from two pairwise matrices, one containing geographic distances and the other containing squared genetic distances. Geographical distance between sample sites was derived from global positioning system coordinates taken at each sampling location (one waypoint for each of the 11 populations). Correlograms were produced by plotting r as a function of distance. Two methods were used to assign statistical significance (Peakall et al. 2003): random permutation and bootstrap estimates, with numbers of permutations and bootstraps set to 1000. In this study, we conducted two separate analyses, one including individuals sampled from the four southern Californian populations, and a second one including individuals from the seven Central Valley populations studied by our research group under a previous study funded by CalFed.

Populations undergoing declines in effective population size exhibit reductions in allele numbers (K) and gene diversity (H_e), with the alleles reducing in number faster than the gene diversity. The analysis program BOTTLENECK (Piry et al. 1999) evaluates evidence for bottlenecks by comparing observed (H_o) and expected (H_e) gene diversity. We tested for bottlenecks using both the IAM and SMM models of microsatellite mutation using a Wilcoxon test for significance.

We also examined evidence for inbreeding within and among southern California and Central Valley populations. A general rule is that F_{IS} values of 0.100 or above represent significant levels of inbreeding and are thus a cause for concern.

Historic (long-term) migration rates between the southern California and Central Valley populations were estimated in MIGRATE 2.4 (Beerli and Felsenstein 1999). The program estimates θ ($4N_e\mu$, where N_e is the effective population size and μ is the mutation rate) and M (m/μ), where m is the unscaled migration rate independent of mutation). We used 10 Markov chains of 10,000 steps and three chains of 100,000 steps and an adaptive heating scheme (temperatures 1.0, 1.2, 1.5, 3.0). Runs were repeated until the confidence intervals for the posterior probabilities of θ and M overlapped and using prior estimates of θ and M as starting parameters for subsequent runs (three runs were sufficient).

Recent migration (over the last several generations) between the southern California and Central Valley populations was estimated using a Bayesian MCMC analysis in BAYESASS 1.3 (Wilson and Rannala 2003). Individuals were preassigned to the two populations based on sampling location and the program returned means of the posterior probabilities for migration rates into each population. The confidence intervals derived by BAYESASS were compared to those that occur when there is no information in the data to determine if the BayesAss results are reliable.

Finally, we used the program MSVAR (Beaumont 1999) to examine the most probable demographic and genealogical histories of southern California and Central Valley populations. A sample of chromosomes typed at a microsatellite locus can be assumed to have a genealogical history consisting of coalescence events and mutation events going back in time until the most recent common ancestor of the sample. The assumed demographic history is of a single stable population that was of size N_1 chromosomes at some time t_a ago and subsequently changed gradually in size to N_0 chromosomes over the period from t_a to the current time. The program estimates the posterior distribution of parameters describing the genealogical and demographic history of the sample using Markov Chain Monte Carlo simulation.

Mitochondrial DNA sequencing

We analyzed the combined mitochondrial DNA sequence data from ATPase 6 and 8 genes. All PCR reactions were carried out on an MJ Research PC-200 Thermocycler. PCR was performed in 25 μ L reactions with 1 μ L template DNA, 15.9 μ L H_2O , 1.5 μ L $MgCl_2$, 2.5 μ L 10x PCR buffer, 2.0 μ L dNTPs (2.5 μ M), 0.1 μ L Taq Gold, and 1 μ L each of primers L8929 (10 μ M) and C03HMH (10 μ M). The PCR profile was as follows: incubation at 94°C for 15 min; 35 cycles of amplification at 94°C for 30 s, 59°C for 30 s, and 72°C for 60 s; and a final extension step at 72°C for 10 min. PCR products were electrophoresed on 1-2% agarose gels. Bands were manually cut from gels and purified using a Zymoclean Gel DNA Recovery Kit, and the product was sequenced on a Applied Biosystems (ABI, Foster City, California, U.S.A.) 3730XL automated genetic analyzer following the complementary sequencing protocol. Sequences were manually aligned and edited in the program Sequencher 2.0 (Gene Codes Corporation, Ann Arbor,

MI USA). Contigs were then exported to the program MacClade (Version 4.0.8, Maddison & Maddison 2003), where subsequent editing was performed.

RESULTS

After extracting DNA, we used two complementary molecular markers, nuclear DNA microsatellites and mitochondrial DNA partial gene sequences, to examine population genetic structure across the four southern California sites. We then compared our results to samples we collected and analyzed between 2001 and 2005 from seven Central Valley colonies from Colusa, Merced, Tulare, and Kern counties under a CalFed funded project (Figure 1).

We were able to robustly amplify microsatellites and mtDNA sequence from 13 of the 22 Holiday Lake samples due their degree of degradation. We were able to successfully amplify DNA from all samples from the other three sites. Our final sample sizes for microsatellite genotyping and mtDNA sequencing (in parentheses) analyses were:

Holiday Lake: 13 (5)

Newberry Springs: 40 (9)

Ramona Farms: 30 (10)

Rancho Jamul: 12 (10)

Microsatellites

Of the ten microsatellite loci that were screened (Table 2), eight conformed to Hardy-Weinberg expectations, and only these eight were included in subsequent analyses. A total of 95 individuals from four populations in southern California and 122 individuals from 7 populations in the Central Valley were analyzed at >75% of microsatellite loci (Tables 1 and 3, Figure 1). Pairwise relatedness (R) values were low (mean = -0.015) and conformed to normal distributions. Interestingly, observed heterozygosity values were higher within southern California than within the Central Valley ($H_o = 0.682$ and 0.583 , respectively; Table 3). Also, the total allele count was higher for southern California than for the Central Valley (10.38 and 8.5 alleles, respectively).

Allele-sharing distance neighbor-joining trees for both individuals and populations indicate there was no genetic structuring within or between the Central Valley and southern California populations (Figure 2). There was no relationship between genetic distance and sampling location. Overall and pairwise F_{ST} values were low: overall F_{ST} was 0.0074, and pairwise values ranged from -0.024 to 0.037, (Table 4). None of the 55 pairwise F_{ST} values was significant (Table 4). There was no evidence whatsoever for isolation-by-distance ($R^2 = 0.0005$, $P = 0.466$; Figure 3).

STRUCTURE revealed no genetic structure among populations (Figure 4). Log likelihood and ΔK statistics indicated $K=4$ as the best fit of clusters for the dataset. Manual review of cluster assignments also indicated that no additional resolution of individual and population clustering was observed at higher K values. The large majority of assignment is to the blue cluster which is distributed broadly across both regions and throughout each

population sampled. We also see the yellow and red clusters assigning to individuals in both regions, indicating possible migrants and connectivity between the regions.

Spatial autocorrelation analyses show little evidence that genetically similar individuals are grouping together in space (Figure 5). In southern California, at close distance intervals (below 75 km), individuals are significantly more similar than expected by chance ($P = 0.001$) which may indicate some small degree of isolation from each other. However, since only four populations are included in the Southern California analysis, the power to detect such variation is limited.

Overall heterozygosity levels for the microsatellites were high (average = 0.59; Table 3), and there was no evidence for recent bottlenecks ($P > 0.05$).

The inbreeding coefficient (F_{IS}) values were significantly positive ($P < 0.05$) for both the southern California ($F_{IS} = 0.121$) and Central Valley ($F_{IS} = 0.090$) populations, with the value for the southern California population ~33% larger. The positive values of F_{IS} are indirect genetic signatures of possible significant inbreeding in each population.

Estimates of recent migration rates between the southern California and Central Valley regions are compromised by the fact that BAYESASS utilizes population assignment for the migration rate estimates and we do not detect significant population structure between the two regions. Nevertheless, the results provide insight into directionality of migration as estimated from genetic analyses. BAYESASS estimates the migration per generation from southern California to the Central valley as 32.3% (30.4-33.2 CI) and the migration rate from the Central Valley to southern California as 9.5% (5.3-14.9 CI). Both estimates are significant when compared to simulated estimates of values where there is no information in the data. We again caution that the lack of population structure compromises this analysis as to order of magnitude of migration rates, but the directionality of estimated migration (~3X greater from southern California to the Central Valley than from the Central Valley to southern California) provides a qualitative indication of asymmetrical geneflow and migration.

The long-term/historical estimates of migration rate m generated by MIGRATE are 1.2×10^{-3} for migration from southern California to the Central Valley and 1.7×10^{-4} for migration from the Central Valley to southern California. Note that these results are migration rate based on effective population size N_e , not census size (that is why the numbers are significantly smaller for MIGRATE results compared to BAYESASS results). As in the BAYESASS results, we see asymmetrical migration rate estimates with the directionality of higher estimated rates from southern California to the Central Valley. MIGRATE also estimates historical effective population size (N_e), with estimates of $N_e = 2,365$ for the Central Valley population and 19,620 for the southern California population.

Finally, the MSVAR analyses suggest that there have been significant declines in effective population size (N_e) over recent generations in southern California, but not in the Central Valley. In southern California, simulations suggest that effective population size has declined from an historical N_e of 36,308 (± 953) to a current N_e of 813 (± 98), over 63 (± 8.5)

generations. In contrast, in the Central Valley, simulations show that N_e has remained more or less constant, with an historical N_e of 9120(\pm 761) and a current N_e of 12,303(\pm 884), over 117(\pm 9.1) generations (95% confidence intervals).

Mitochondrial DNA sequencing

We obtained sequence from ATPase 6 and 8 genes from 34 southern Californian individuals representing all four populations. We then compared these data to sequences previously obtained for 52 Central Valley specimens (see Table 1 for sample sizes). The TRBL samples contained low numbers of mtDNA haplotype sequences relative to overall sample sizes (Figure 6). The minimum spanning tree was star-like, with a few uncommon haplotypes branching off from one very common haplotype; this pattern is consistent with post-Pleistocene population expansion with no indications of longterm isolation of the southern California and Central Valley populations. The vast majority of haplotypes from both regions was haplotype A, with southern California having two unique haplotypes and the Central Valley having 5 unique haplotypes, all at very low frequency.

DISCUSSION

Microsatellite and sequencing results reveal no evidence of significant population structuring between the southern California and Central Valley populations of the TRBL, suggesting either that there is considerable movement between sites and few if any isolated populations, or isolation is very recent and is not yet reflected in the population genetic signatures. Microsatellite analyses strongly suggest that there has been significant gene flow between Central Valley and southern California populations, maintaining a single interbreeding population. We also do not find any significant genetic differentiation between populations within either region, other than some specific cluster assignments distributed amongst populations within each region detected with the most sensitive analysis settings of the STRUCTURE program. These are most likely individuals that are relatively closely related lineages and may be a result of our sampling methodology (limited sampling opportunities of dead nestling and abandoned egg samples) but do not constitute unique populations.

We do not find any relationship of physical distance between populations and calculated genetic distances between those populations, indicative of a lack of restriction of geneflow within and between the two regions.

Genetic diversity (as measured by microsatellite heterozygosity) for the southern California ($H_o=0.60$) and Central Valley ($H_o=0.58$) populations are at moderately high levels and genetic tests for signatures of recent population bottleneck events is negative. So there are not any indications of significant reductions in genetic diversity from microsatellite analysis (which can be correlated to reduction in fitness in a number of vertebrate studies when H_o drops below ~ 0.5). The microsatellite derived inbreeding coefficient values (F_{IS}) of 0.090 for the Central Valley and 0.121 for the southern California population are significant. Values ~ 0.1 are considered an indicator of a

moderate level of inbreeding, but future monitoring is warranted to see if the values increase over time considering TRBL recent demographics.

The southern California population has a higher average number of microsatellite alleles per marker than the Central Valley population (10.38 vs. 8.5 alleles on average per marker). This higher allelic diversity of the southern California population, despite its much smaller census population size compared to the Central Valley population, suggests that the southern California population is an important reservoir of genetic variation for the species overall.

Genetic demographic trend analyses indicate that the southern California population has gone through a significant population size decline over the last 50 to 100 generations, from an effective population size (N_e) estimate of ~36,000 to a current estimated N_e of ~60, almost 3 orders of magnitude (MSVAR results). N_e is a genetic estimate of the number of breeding individuals. This trajectory of a rapidly decreasing population size in southern California is consistent with the severe reduction in observed bird count population size estimates over the last few decades (Tricolored Blackbird Working Group 2009). In contrast, for the Central Valley population the genetic demographic trend analyses do not detect a decline in N_e , with an estimate of N_e at ~9,100 approximately 100 to 140 generations ago, and estimated at a current N_e of ~12,300. Observed bird counts of trends of significant reductions in population size are not detected in the genetic analyses yet, possibly due to the still relatively large populations in the Central Valley, compared to southern California's much smaller and fragmented populations. MIGRATE estimates of historical N_e ($N_e=2,365$ for Central Valley and 19,260 for southern California) are consistent with MSVAR results in both order of magnitude and the fact that the historical N_e estimate for southern California is significantly larger than the Central valley population. This is consistent with the observations of early workers in southern California, who considered the tricolor to be the most abundant bird species in southern California (Baird in Cooper 1870).

We can also estimate the ratio of effective population size (N_e) to estimated actual population sizes (N_{census} or N_c) as a possible insight into the unique colonial behavior and breeding of this species. For reference, most species with sophisticated breeding behavior (non-random) have N_e/N_c ratios ranging from ~0.2 to 0.4). The N_e/N_c ratio for massively colonial birds such as the Tricolored blackbird has not been quantified before, because the breeding behavior (number of males involved, number of females involved, multiple breeding attempts per season, extra pair copulations, etc.) has not been effectively determined. We utilized MSVAR estimates of contemporary N_e for the southern California and Central Valley populations and estimated population sizes based on the recent TRBL conservation plan (Tricolored Blackbird Working Group 2009) of 10,000 for the southern California population and 250,000 for the Central Valley populations. The resulting N_e/N_c ratio for the southern California population is $N_e/N_c=0.081$ and for the Central Valley it is $N_e/N_c=0.049$. Other than being significantly lower than generally observed for the general observed range of N_e/N_c mentioned above, we can not quantitatively put this into context of the tricolored blackbird breeding behavior and system at this time. Combined detailed observational and genetic studies are

required to tease apart the origin of these low ratios estimated from genetic data, but are extremely difficult due to the colonial nature of this species.

Mitochondrial DNA sequence variation was extremely low in this species. There were very few distinct haplotypes, with almost all individuals sharing the same type. This lack of variation is of some concern, since the tricolored blackbirds have an extremely restricted range and have been declining in numbers in recent years (Cook and Toft 2005). As a colonial species, they are more vulnerable to destruction of key nesting habitats as well as the spread of disease. Understanding the relationship between intraspecific variation in genetic diversity and reproductive fitness would be useful for determining whether low genetic variation in this species is merely a signature of anthropogenic habitat alteration, or is itself a cause of population decline.

Finally, the genetic signature of a recent and dramatic decrease in effective population size in southern California is of high concern, since it suggests that despite the lack of evidence for recent bottlenecks in this species, there are many fewer birds breeding in southern California than in the recent past. However, these numbers are difficult to place in context, since we have almost no empirical data on individual lifetime reproductive success or variation in reproductive success across different age and sex classes. Information on individually marked birds will help us understand the relationship between census size – i.e. the number of birds counted at different colonies from year to year – and the numbers of breeding birds needed to maintain a healthy viable population.

RECOMMENDATIONS

Marking/banding studies such as the one currently being carried out by Robert Meese at UCD in the Central Valley would also be extremely beneficial in identifying possible migrants between the southern California and Central Valley populations. We recommend a banding program of birds in the southern California population.

We recommend continued genetic monitoring of the Central Valley and southern California populations to provide demographic, connectivity and genetic fitness information to the Tricolored Blackbird Working Group and other conservation stakeholders. Genetic testing of ~100 individuals from each of two regions (southern California and the Central valley), distributed across the specific population sites, every three years should be a reasonable time period to capture trends measurable by genetic assays, and sync with the TRBL statewide survey conducted approximately triennially.

Finally, we recommend a genetic analysis of historical TRBL samples from the early 1900's to attempt to quantify the historical demography of this species as a baseline to compare and contrast to the current populations' status. As an example, the Museum of Vertebrate Zoology at the University of California, Berkeley (UCB), has 290 TRBL specimens dating back to 1861. DNA can be extracted from toepads, minimizing impact on the museum voucher specimens. The DNA will be significantly degraded, preventing microsatellite assays or cost-effective sequencing of mtDNA gene segments, but new

genetic characterization techniques assaying single nucleotide changes (polymorphisms or SNPs) and sequencing only 50-60bp of DNA can be applied to these historical samples. A number of research groups now have this historical sample DNA extraction and SNP assaying capability including UCLA, UCD, and UCB.

ACKNOWLEDGEMENTS

The UCLA Center for Tropical Research gratefully acknowledges the funding that enabled this study from the California Department of Fish and Game, Audubon California, and three of its chapters – the Los Angeles Audubon Society, the Pomona Valley Audubon Society, and the Sea and Sage Audubon Society. We also would like to acknowledge and thank Tom Paulek and Susan Nash for their southern California TRBL nesting site monitoring and sample collection efforts, which were crucial to this study.

REFERENCES

Baird, S. F. 1870. Ornithology of California. Land birds, Vol. 1. From the manuscripts and notes of J. G. Cooper. U. S. Geological Survey of California. Welch Bigelow and Co. Cambridge, MA.

Beaumont, M. A. 1999. Detecting population expansion and decline using microsatellites. *Genetics*. 153: 2013-2029.

Beedy, E. C. and Hamilton, W. J. I. 1999. Tricolored Blackbird (*Agelaius tricolor*). In: *The Birds of North America, No. 423* (Ed. by Poole, A. & Gill, F.). Philadelphia, PA: The Birds of North America, Inc.

Beerli P., and Felsenstein J. 1999. Maximum likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* 152:763–773

Cook, L. F. and Toft, C. A. 2005. Dynamics of extinction: population decline in the colonially nesting Tricolored Blackbird *Agelaius tricolor*. *Bird Conservation International*. 15: 73-88.

Dawson, R. J. G., Gibbs, H. L., Hobson, K. A. and Yezerinac, S. M. 1997. Isolation of microsatellite DNA markers from a passerine bird, *Dendroica petechia* (the yellow warbler), and their use in population studies. *Heredity*. 79: 506-514.

DeHaven, R.W. and Neff, J. A. 1973. Recoveries and returns of tricolored blackbirds, 1941-1964. *Western Bird Bander* 48: 10 – 11.

Goudet, J. 2001. FSTAT, a program to estimate and test gene diversities and fixation indices, Version 2.9.3. Available at <http://www2.unil.ch/izea/software/fstat.html>.

- Hughes, C. R., Kavlie, R. and Johnson, K. 1998. Characterization of polymorphic trinucleotide microsatellite loci in the great-tailed grackle. *Molecular Ecology*. 7: 783-784.
- Jin, L. and Chakraborty, R. 1993. Estimation of genetic distance and coefficient of gene diversity from single-probe multilocus DNA fingerprinting data. *Molecular Biology and Evolution*. 11: 120-127.
- Kelsey, R. 2008. Results of the Tricolored Blackbird 2008 census. Audubon California. <http://ca.audubon.org/pdf/Tricolored-Blackbird-2008-Status-Report.pdf>.
- Langella, O. 1999. Populations, Version 1.2.28. Available at <http://www.cnrs-gif.fr/pge/bioinfo/populations/index.php?lang=en>.
- Maddison, D. R. and Maddison, W. R. 2003. MacClade 4: analysis of phylogeny and character evolution. Version 4.06. Sunderland, MA: Sinauer.
- Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Research*. 27: 209-220.
- Page, R. D. M. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences*. 12: 357-358.
- Peakall, R., Ruibal, M. and Lindenmayer, D. B. 2003. Spatial autocorrelation analysis offers new insights into gene flow in the Australian bush rat *Rattus fuscipes*. *Evolution*. 57.
- Peakall, R. and Smouse, P. E. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*. 6: 288-295.
- Petren, K. 1998. Microsatellite primers from *Geospiza fortis* and cross-species amplification in Darwin's finches. *Molecular Ecology*. 7: 1782-1784.
- Piertney, S. B., Marquiss, M. and Summers, R., M. 1998. Characterization of tetranucleotide microsatellite markers in the Scottish crossbill (*Loxia scotica*). *Molecular Ecology*. 7: 1261-1263.
- Piry, S., Luikart, G. and Cornuet, J. M. 1999. BOTTLENECK, a computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity*. 90: 502-503.
- Pritchard, J. K., Stephens, M. and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics*. 155: 945-959.
- Queller, D. C. and Goodnight, K. F. 1989. Estimating relatedness using genetic markers. *Evolution*. 43: 258-275.

- Raymond, M. and Rousset, F. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*. 86: 248-249.
- Rice, W. R. 1989. Analyzing tables of statistical tests. *Evolution*. 43: 223-225.
- Richardson, D. S., Jury, F. L., Dawson, D. A., Salgueiro, P., Komdeur, J. and Burke, T. 2000. Fifty Seychelles warbler (*Acrocephalus sechellensis*) microsatellite loci polymorphic in Sylviidae species and their cross-species amplification in other passerine birds. *Molecular Ecology*. 9: 2225-2230.
- Rousset, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*. 145: 1219-1228.
- Shuford, W. D. and Gardali, T., editors. California Bird Species of Special Concern: A ranked assessment of species, subspecies, and distinct populations of birds of immediate conservation concern in California. Studies of Western Birds 1. Western Field Ornithologists, Camarillo, CA and California Department of Fish and Game, Sacramento, CA.
- Smouse, P. E. and Peakall, R. 1999. Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity*. 82: 561-573.
- Tricolored Blackbird Working Group. 2009. Conservation Plan for the Tricolored Blackbird (*Agelaius tricolor*). 2.0 update, Jan. 2009. Susan Kester (ed.). Sustainable Conservation, San Francisco CA.
- Unitt, P. 2004. *San Diego County Bird Atlas*. San Diego Natural History Museum. San Diego, CA
- Webster, M. S., Chuang-Dobbs, H. C. and Holmes, R. T. 2001. Microsatellite identification of extrapair sires in a socially monogamous warbler. *Behavioral Ecology*. 12: 439-446.
- Weir, B. and Cockerham, C. C. 1984. Estimating F-statistics for the analysis of population structure. *Evolution*. 38: 1358-1370.
- Wilson GA., and Rannala B. 2003. Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* 163:1177-1191
- Winker, K., Glenn, T. C. and Graves, G. R. 1999. Dinucleotide microsatellite loci in a migratory wood warbler (Parulidae: *Limnothlypis sivainsonii*) and amplification among other songbirds. *Molecular Ecology*. 8: 1553-1556.

Table 1. Sampling sites within southern California and the Central Valley for analysis of Tricolored Blackbird population genetic structure.

Site	County	Latitude	Longitude	Year sampled	Samples collected	Microsat analysis	Sequence analysis
<i>Southern California</i>							
Ramona Farms	Riverside	33.8291	-117.0181	2006	90	30	10
Rancho Jamul	San Diego	32.6852	-116.8560	2007	15	12	10
Newberry Springs	S. Bernardino	34.8334	-116.6955	2008	42	40	9
Holiday Lake	Los Angeles	34.7997	-118.5767	2008	22	13	5
<i>Central Valley</i>							
Capitol Outing Club	Colusa	39.2311	-122.0992	2002	38	30	11
O'Neill Forebay	Merced	37.0760	-121.0210	2005	8	5	2
Merced NWR	Merced	37.1710	-120.6060	2005	17	13	13
Te Velde Ranch	Tulare	36.3000	-119.3000	2002	26	20	8
Costa Dairy	Kern	35.6000	-119.3000	2002	13	13	7
Kern River Preserve	Kern	35.6694	-118.3153	2001-02	23	23	2
Wind Wolves	Kern	35.0500	-119.1700	2001	50	18	9

Table 2. Microsatellite loci used in this study and species for which these loci were initially developed.

Locus	Primer Sequences (5' → 3')	Repeat	Source species	Reference
Dpμ1	TGGATTCACACCCCAAAATT AGAAGTATATAGTGCCGCTTGC	(CA) ₂₂	Yellow warbler (<i>Dendroica petechia</i>)	Dawson <i>et al.</i> (1997)
Dpμ16	ACAGCAAGGTCAGAATTTAAA AACTGTTGTGTCTGAGCCT	(AC) ₁₂ (GC) ₄ ACGCAC(GC) ₂	Yellow warbler	Dawson <i>et al.</i> (1997)
Gf5	AAACTCTGGGAGTGAAGTCT AACTATTCTGTGATCCTGTTACAC	(AC) ₁₄	Medium ground finch (<i>Geospiza fortis</i>)	Petren (1998)
Gf12	AATCCTTCTCGTCCCTCTTGG TTTGAGTGTGCAGCAGTTGG	(AC) ₁₇	Medium ground finch	Petren (1998)
Ase18	ATCCAGTCTTCGCAAAAAGCC TGCCCCAGAGGGAAGAAG	(GT) ₁₂	Seychelles warbler (<i>Acrocephalus sechellensis</i>)	Richardson <i>et al.</i> (2000)
Dca32	GGACACAAGCACATCACAATC CCCATGCNTTCCACANACTCT	CA	Black-throated blue warbler	Webster <i>et al.</i> (2001)
Lsw μ7	GATGTGACAAGTGTGCTCTCC TTTATATCTAGTGACGCTCTA	(GT) ₁₄	Swainson's warbler	Winker <i>et al.</i> (1999)
QmAAT10	GGAATTCAGTATGTGAATGAGTC ATTGCAAAAAACAGAAGCATTTTAAC	AAT	Great-tailed grackle	Hughes <i>et al.</i> (1998)
Lox1*	ATGATGGTAAGTCTAATGAAAGC CCACACACATTCACTCTATTG	(CTTT) ₃₀	Scottish Crossbill (<i>Loxia scotica</i>)	Piertney <i>et al.</i> (1998)
Dpμ3*	GAATTACCCATTATTGGATCC AGCAGCAAAAACAAACCAG	(GAGg/cA) ₅ (GAGAGGa/g) ₆	Yellow warbler	Dawson <i>et al.</i> (1997)

* = these two loci did not conform to Hardy-Weinberg equilibrium and were excluded from final analyses.

Table 3. Eight polymorphic microsatellite loci used in final analyses of population genetic structure of the Tricolored Blackbird. Listed for each locus is the number of alleles found, expected heterozygosity H_E , and observed heterozygosity H_O for both Southern California and Central Valley populations. Data refer to the results from the 95 Southern Californian and 122 Central Valley individuals for which at least 75% of alleles were scored. All of the loci listed here conformed to Hardy-Weinberg equilibrium.

Locus	<i>Southern California</i>			<i>Central Valley</i>		
	Alleles	H_E	H_O	Alleles	H_E	H_O
Ase18	15	0.748	0.681	12	0.676	0.694
Dp μ 16	13	0.799	0.630	11	0.721	0.610
Dp μ 1	3	0.223	0.189	2	0.178	0.180
QmAAT10	8	0.794	0.784	7	0.738	0.694
Gf12	18	0.908	0.843	17	0.911	0.873
Lsw μ 7	5	0.514	0.374	3	0.485	0.344
Gf5	10	0.630	0.533	5	0.573	0.504
Dca32	11	0.855	0.778	11	0.846	0.762
Average	10.38	0.684	0.602	8.5	0.641	0.583

Table 4. Pairwise estimates of F_{ST} values of the Tricolored Blackbird based on microsatellite DNA genotypes (below diagonal). NS above the diagonal denotes that there were no statistically significant differences at the $P < 0.05$ level.

	Kern River	Wind Wolves	Te Velde Ranch	Costa Dairy	Capitol Outing Club	O'Neill Forebay	Merced NWR	Ramona Farms	Holiday Lake	Rancho Jamul	Newberry Springs
Kern River	-	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
Wind Wolves	0.0069	-	NS	NS	NS	NS	NS	NS	NS	NS	NS
Te Velde Ranch	0.0181	-0.0054	-	NS	NS	NS	NS	NS	NS	NS	NS
Costa Dairy	0.0067	-0.0137	-0.0125	-	NS	NS	NS	NS	NS	NS	NS
Capitol Outing Club	0.0228	0.0003	-0.0016	-0.0143	-	NS	NS	NS	NS	NS	NS
O'Neill Forebay	0.0182	0.0216	-0.0236	-0.0239	-0.0149	-	NS	NS	NS	NS	NS
Merced NWR	0.0212	0.0110	0.0117	0.0199	0.0374	0.0298	-	NS	NS	NS	NS
Ramona Farms	0.0248	0.0048	0.0026	0.0005	0.0108	0.0289	0.0204	-	NS	NS	NS
Holiday Lake	0.0060	-0.0176	-0.0075	0.0057	0.0144	-0.0181	0.0072	0.0073	-	NS	NS
Rancho Jamul	0.0070	-0.0213	-0.0103	-0.0210	-0.0092	-0.0212	0.0144	-0.0046	-0.0074	-	NS
Newberry Springs	0.0204	-0.0017	0.0029	-0.0001	0.0149	-0.0190	0.0208	0.0253	0.0052	-0.0008	-



Figure 1. Tricolored Blackbird (*Agelaius tricolor*) colonies of the Central Valley (from previous CalFed funded study) and southern California (this study) sampled for population genetic analysis. Numbers in parentheses represent sample sizes for individuals genotyped. Red and blue circles/writing represent Central Valley and southern California sampling locations, respectively.

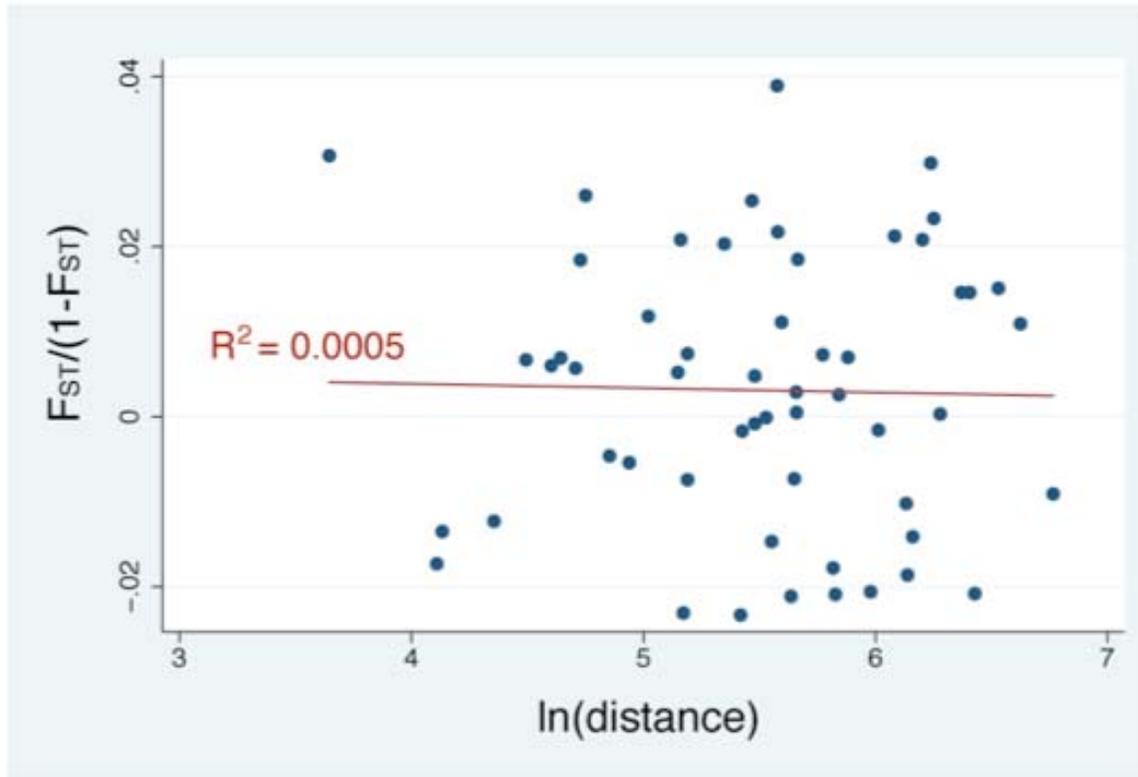


Figure 3. Pairwise estimates of genetic distance plotted against log-transformed geographical distance for 11 sample groups of Tricolored Blackbirds. Regression line and R^2 value are shown.

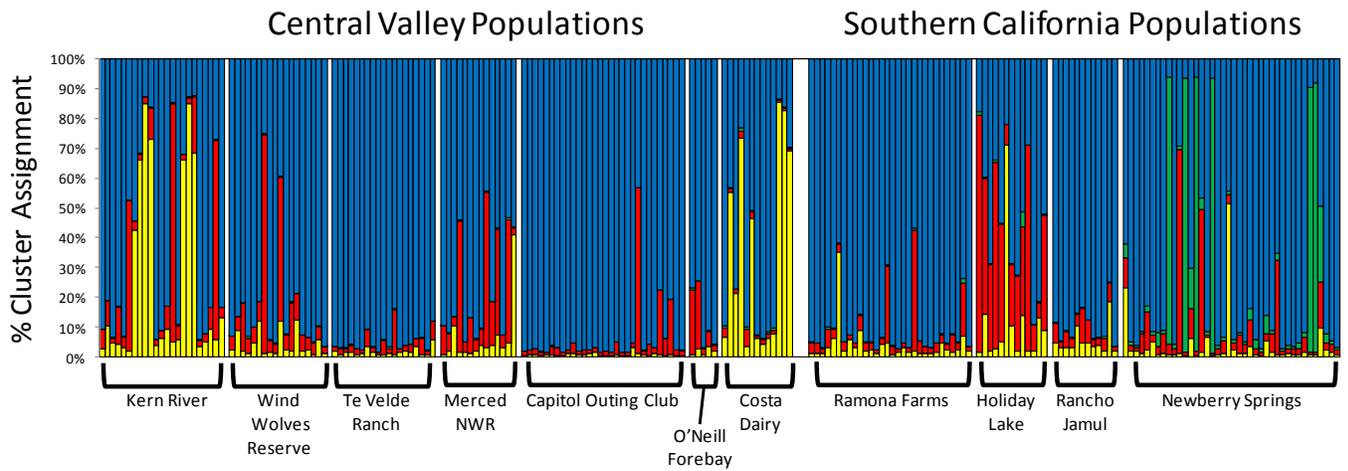


Figure 4. Results of STRUCTURE clustering analysis for the best fit K value of 4 clusters for the combined dataset of 95 Southern California individuals (4 populations) and 122 Central Valley individuals. Percent cluster assignment is represented for each of four clusters (represented by red, blue, yellow and green colors).

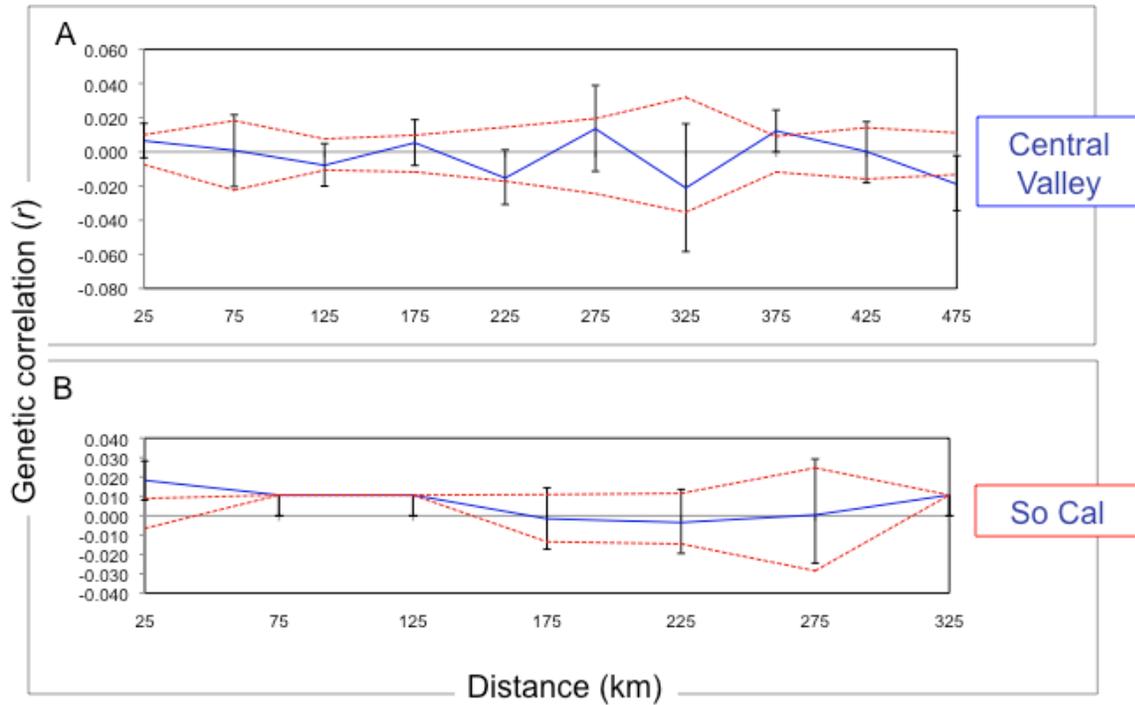


Figure 5. Correlogram plots of the spatial genetic autocorrelation coefficient (r) as a function of geographical distance for (A) Central Valley and (B) southern California Tricolored Blackbird populations. Red dotted lines represent the permuted 95% confidence intervals. Bootstrapped 95% confidence error bars are also shown.

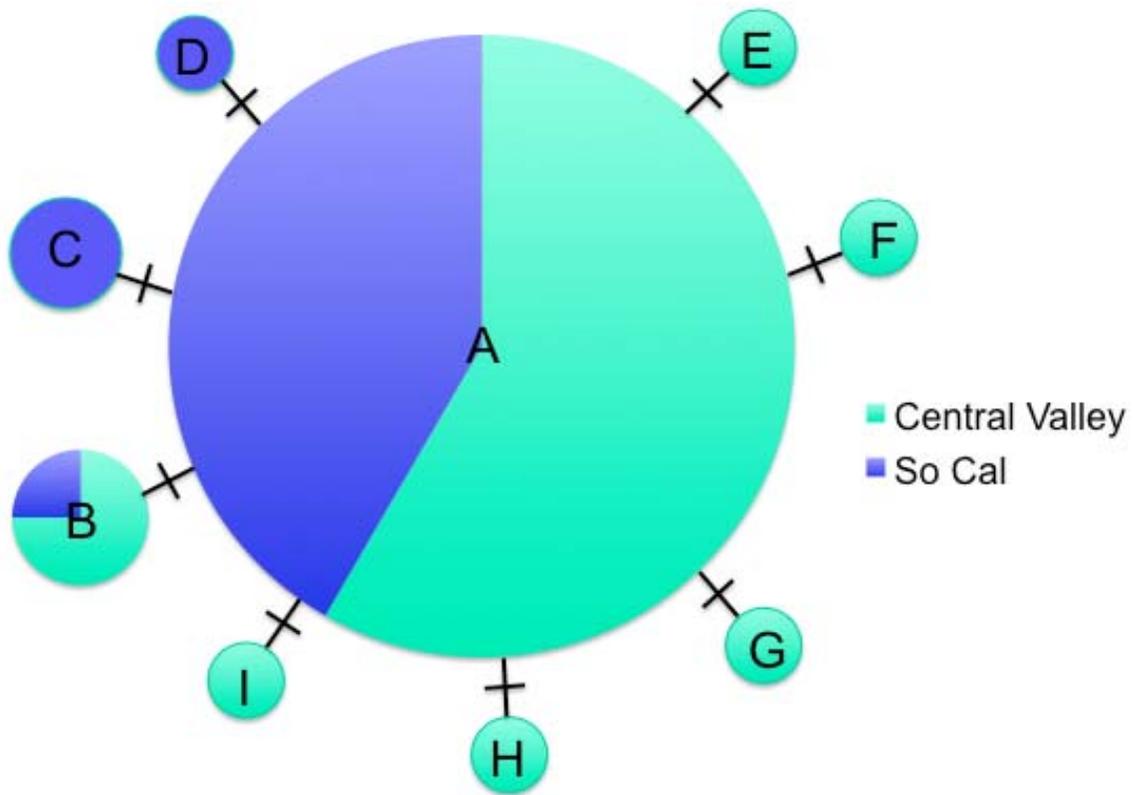


Figure 6. Minimum spanning tree for the Tricolored Blackbird. Letters indicate haplotypes, sizes of circles are proportional to haplotype frequency, and individual bars represent base pair substitutions.