With a large number of available loci, we were able to rigorously limit analyses to only those with reliable and consistent amplification. We also utilized analyses that explicitly account for the presence of allelic dropout. With these data, we assessed historical population structure and genetic diversity in cactus wrens in coastal southern California and compared these to contemporary patterns.

METHODS

Samples

We searched museum collections for cactus wren skins via ORNIS (ornis2.ornisnet.org), and selected museums with extensive material from southern California to contact and visit. Skins were sampled following Mundy *et al*. (1997), slicing off small toepads (~1-2 mm^2) using new scalpel blades for each bird. We stored dry toepads at cool temperatures prior to DNA extraction. All DNA extractions were conducted in a fume hood in a separate laboratory from other avian DNA-based projects. We incubated chopped toepads in 500 uL of extraction solution (10 mg/mL DTT, 1% SDS, 0.1 mg/mL proteinase K, 0.02M EDTA, 0.01M Tris-HCl, and 0.01M NaCL) at 55^{o} C for 5 days and regularly agitated these by hand. After incubation, DNA was purified using phenol-chloroform extractions and Millipore Amicon Ultra 30K centrifugal filters. We further purified these products using an ethanol-sodium acetate precipitation. All disposables and pipettes used in DNA extractions were decontaminated with a CL-1000 UV Crosslinker (UVP) prior to all extractions. Blanks (negative controls) were included during each extraction session to allow testing for cross contamination downstream.

Data quality

Anticipating degraded genomic DNA, we focused on loci that were <200 basepairs and hence more likely to amplify. In our original cactus wren microsatellite library, six loci were already of an appropriate size. We also redesigned primers to amplify smaller products for an additional 11 loci (Table S1). These 17 loci were amplified using the Multiplex PCR Kit (Qiagen) with 10 µL reactions with bovine serum albumin (BSA) as an additive. We assessed data quality at several steps to focus genotyping efforts on higher quality extractions and more reliable loci. In the first step, all samples were amplified across all loci. After eliminating loci with spurious amplification patterns (e.g., slippage) and samples that did not provide any data across loci, we genotyped the remaining dataset twice more with 20 µL reactions. Individuals that genotyped as homozygous in the first run but heterozygous in the later, larger reactions were assumed to be heterozygous for analytical purposes. The larger PCRs resulted in larger amounts of product, hence we assumed these were amplifying alleles that were not detected using smaller reactions. In the few instances of discrepancies in alleles between repeated genotypes (e.g., alleles of differing sizes were amplified in different runs) genotypes were discarded and coded as missing. Negative controls were checked on electrophoresis gels.

Prior to analyses, we converted loci that were reduced in size to the number of repeats to make alleles match in the historical and contemporary datasets. We checked loci for allelic dropout in MICRO-CHECKER (van Oosterhout *et al*. 2004) and Hardy-Weinberg Equilibrium (HWE) in GENALEX (Peakall & Smouse 2006, 2012); but we did not test for linkage disequilibrium since these loci have already been shown to be unlinked (Barr *et al*. 2012, 2013).

Our contemporary dataset was reduced to locations overlapping those where historical samples were captured for comparative analyses.

To assess data quality and the potential impacts of missing data on analysis results, we plotted individual proportions of missing data and individual heterozygosity (proportion of heterozygous loci found within an individual) by collection year and tested for significant correlations by calculating Pearson's *r*. We also tested whether individual heterozygosity levels were significantly different in the total historical sample compared to the contemporary sample using a two sample t-test. We would expect to see significantly fewer heterozygous individuals in older samples if allelic dropout rates were high due to poor template DNA quality.

Identifying population structure

Geographic localities for cactus wrens in museum collections were usually too general or antiquated to assign accurate global position system (GPS) points; hence, very broad groupings were used for analyses (Figure 1). Furthermore, captures in the study area largely occurred over a 50 year period, 1885-1923. To help account for the sampling variance and genetic drift that might confound polymorphic marker datasets over numerous generations, we further divided the dataset into smaller time frames in a sliding window-type format for genetic structure analyses. For each genetic structure analysis, we analyzed samples captured 1885-1905, 1899-1910, and 1906-1923. These intervals were chosen to maximize the number of available samples (1899-1910) and to include all available data without overlapping years (1885-1905 and 1906-1923).

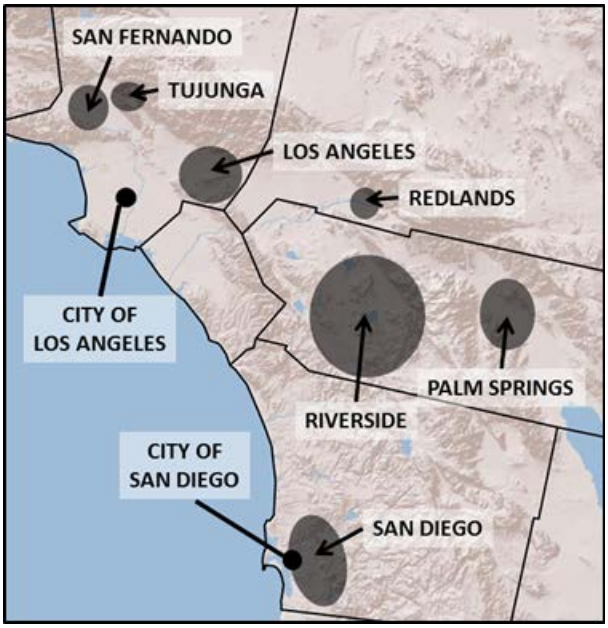

Figure 1. General locations of cactus wrens sampled in museum collections.

Pairwise Weir and Cockerham's (1984) F_{ST} and Cavalli-Sforza and Edwards (1967) distance (D_c), two measures of genetic differentiation, were calculated between groupings in FreeNA (Chapuis & Estoup 2007), both with and without a correction for null alleles. F_{ST} and D_c both scale 0 to 1, with higher numbers representing great levels of genetic differentiation. While F_{ST} accounts for allele frequency differences, D_c also takes into account the number of mutational steps between alleles. The measures are differentially affected by missing data and null alleles. Calculating both can help to interpret results. We tested for associations between geographic distance and genetic distance, or isolation by distance (IBD), using a Mantel test as implemented in IBDWS (Jensen *et al*. 2005) with 1,000 randomizations to test for significance. For these tests for IBD, we used F_{ST}s corrected for null alleles for the historical samples, but used the uncorrected calculation for the

contemporary samples. Since there is no evidence for allelic dropout in the contemporary dataset, there is no reason to use the corrected calculation. IBD can be related to dispersal patterns and levels of genetic isolation. In instances of significant IBD, it might be assumed that stepping-stone gene flow is occurring between populations, for instance. A lack of an association between genetic and geographic distances, on the other hand, might be the product of many generations of isolation, with little or no gene flow occurring. We also implemented the Bayesian clustering analyses STRUCTURE (Pritchard *et al*. 2000, Falush *et al*. 2003, Hubisz *et al*. 2009) and GENELAND (Guillot *et al*. 2005, Guillot *et al*. 2008). Analyses in STRUCTURE were conducted assuming a correlated alleles model with admixture and a location prior based upon the general capture area (Figure 1). In each analysis, we ran 1M Markov chain Monte Carlo (MCMC) steps after a 500,000 step burn-in for potential numbers of clusters (*K*) 1 – 10 with five repetitions at each *K*. These results were summarized using STRUCTURE HARVESTOR (Earl & van Holdt 2012). In GENELAND, analyses were conducted using the uncorrelated, null alleles model (a feature unique to this program) and testing for *K*s 1 - 10 with 2M MCMC steps and a 20% burn-in. Since individual GPS points were not available for historical samples, we used a very broad spatial uncertainly of 1×10^6 meters. Both of these programs use differing algorithms to detect "clusters" of individuals based upon Hardy-Weinberg Equilibrium (HWE), the product of sharing gene flow over time, while maximizing linkage equilibrium, the product of recent gene flow.

Genetic diversity

We calculated expected heterozygosity (H_E) in GENALEX, allelic richness (*A*) in HP-RARE (Kalinowski 2005), and effective population size (N_e) using the linkage disequilibrium method (Waples & Do 2010) in NeEstimator (Do *et al*. 2013). Heterozygosity is the percentage of loci that have two different alleles, and H_E is calculated to account for sampling variance. *A* is the total number of alleles scaled to account for sample size. Populations with higher H_E and *A* have higher levels of genetic diversity. N_e reflects the rate of genetic drift and inbreeding (Caballero 1994) and approximates the number of individuals that contribute equally to offspring in an idealized population (Wright 1938). Based upon genetic structure results, we pooled sites to boost sample size for an overall historical versus contemporary comparison in genetic diversities in the forms of both H_E and *A,* and tested for significance using a paired t-test. We did not make the comparisons in N_e since pooling samples over numerous generations would severely violate a basic assumption of the linkage disequilibrium method (Waples & England 2011).

RESULTS

Data quality

We genotyped a total of 225 cactus wrens over 17 loci (Table S1); however, analyses here are limited to 111 total individuals genotyped at 11 loci. There was no evidence for cross contamination as we observed neither bands from PCRs on negative controls on electrophoresis gels nor extra alleles in genotypes. Many individuals were dropped due to missing data at four or more loci after three amplification attempts. Locations with less than four individuals were also dropped since these do not provide enough information to warrant inclusion in analyses. Six loci could not be amplified consistently in the historical samples. The historical datasets exhibited much higher levels of missing data and allelic dropout than the