USGS
science for a changing world

Genetic Connectivity in the Coastal Cactus Wren

By Kelly R. Barr, Amy G. Vandergast, and Barbara E. Kus

Data Summary Report prepared for SDMMP.

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

U.S. Department of the Interior
KEN SALAZAR, Secretary

U.S. Geological Survey
Marcia K. McNutt, Director

U.S. Geological Survey, Reston, Virginia 2012

.

Genetic Connectivity in the Coastal Cactus Wren

By Kelly R. Barr, Amy G. Vandergast, and Barbara E. Kus

Introduction

The coastal cactus wren (*Camphylorynchus brunneicapillus sandiegensis*) is one of numerous species in decline in San Diego County. Limited to prickly pear (*Opuntia* sp.) and cholla (*Cylindropuntia* sp.) cacti for nesting, the resident songbird's persistence in the county relies upon the existence of such habitat. Urbanization, agriculture, and fire have reduced cactus in San Diego County, leaving only a remnant of the once abundant habitat for the coastal cactus wren (Shuford & Gardali 2008). Large aggregations of cactus wrens exist in areas where urbanization and agriculture have been excluded, such as on the Fallbrook Naval Weapons Station (NWS), on several sites in San Pasqual Valley, and around both Lake Jennings and the Sweetwater Reservoir. Smaller groups dwell in urban canyons, nature reserves, and otherwise undeveloped areas around the county as well. On the order of 200 known coastal cactus wren territories currently exist on public and otherwise preserved properties in San Diego County, likely representing a major reduction from historical population sizes (Shufard & Gardali 2008).

The coastal cactus wren has long been recognized as a species in decline, and is designated both as a California State Species of Special Concern (Shuford and Gardali 2008) and is a targeted species in California's Natural Communities Conservation Planning Program (Pollak 2001). It is also one of four birds species prioritized for connectivity monitoring by the San Diego Management and Monitoring Program. Connectivity describes movement between habitat patches, whether during migration, dispersal, or as part of regular behavioral activity. With high levels of connectivity between populations, genetic diversity, the maintenance of which is vital for species persistence over evolutionary time (Reed & Frankham 2003), is less likely to be lost. If connectivity with nearby populations is too low, small isolated areas affected by stochastic processes (e.g., fire) will not be repopulated with new individuals in the absence of direct human intervention (e.g., translocation). Small and isolated populations naturally lose genetic diversity through genetic drift, leaving less adaptive potential to environmental change and novel disease (Quattro & Vrijenhoek 1989, Leberg & Vrijenhoek 1994). As populations become exceptionally small, a lack of connectivity with other groups may also lead to inbreeding depression, reducing the genetic health of individuals (Charlesworth & Charlesworth 1987, Hemmings et al 2012). For all of these reasons, an understanding of connectivity is essential in species conservation (Lowe & Allendorf 2010).

There are many methods for examining how individuals in a species move between habitat patches, such as colorband resighting in birds (Villard & Hache 2012) and radio telemetry (Doerr et al 2011). Gene flow, the movement of individuals between groups

followed by successful breeding, is a vital component of connectivity (Lowe & Allendorf 2010), and cannot be measured with these techniques. Genetic methods, however, can infer gene flow. Using many polymorphic loci, current population structure in a species can be described, patterns of dispersal can be inferred, and genetic diversity can be estimated. These inferences can be vital for understanding connectivity in a species.

When employing genetic methods for exploring genetic connectivity, the selection of an appropriate genetic marker is of great importance. Different markers require varying effort and expense in discovery and genotyping; furthermore, their validity varies greatly by study question (Sunnucks 2000). To understand very recent changes in gene flow and connectivity, high variation, such as is available with microsatellites, is necessary (Zhang & Hewitt 2003). With high variation, the resolution needed to detect subtle differences between groups can be obtained. Variation is a product of mutation rate and the number of markers, and microsatellites have the highest rates of any genetic marker and many can be easily genotyped. Microsatellites are short repeating regions of DNA, the structure of which confers upwards of 1000X the mutation rates of other regions in the genome. Since they are largely located in non-coding areas, microsatellites are assumed to be selectively neutral, an important quality for population genetic analyses that aim to quantify movement and demographic changes. While development of microsatellite libraries for individual species was historically expensive and time-consuming, the recent advent of next generation sequencing technology has greatly reduced the time and expense of their discovery (Guichoux et al 2011).

With a large suite of microsatellites developed for the species and an extensive sampling regime, we describe genetic connectivity in the coastal cactus wren in San Diego County. To gain a broader perspective, we also obtained samples from the Nature Reserve of Orange County, where many of the remaining coastal cactus wrens are found in that county. These data provide excellent resolution for describing current population structure in the species, reveal the gene flow regime, and provide insight on current levels of genetic diversity within populations. Understanding these patterns will aid in management of current coastal cactus wren populations and future efforts in habitat restoration.

Methods

Samples

All significant and accessible (those on public lands or private lands that provided permission) cactus patches were visited to find cactus wrens and nests through western San Diego County. Potential sites were identified using data from recent surveys by cooperators and known mapped cactus (data not shown; pers. comm. C. Winchell). Nestlings were monitored and sampled at 6 to 12 days in age. Adults and hatch-year birds were captured using standard mist-netting techniques with song playback. All handled individuals were given a numbered metal federal band and 1-3 plastic bands to yield unique color combinations (Federal Bird Banding Permit 22372 to B. Kus). Birds were sampled for growing feathers, in the case of nestlings, or blood via a toenail clip, in the case of adults and hatch-years. Samples

were stored in Queen's Lysis Buffer at -20^{o}C until DNA extraction. A few deceased individuals detected during nest monitoring were collected as well, and either muscle tissue, if available, or toepads were used for DNA extractions from these. All sample collection in San Diego County was authorized by a Memorandum of Understanding between the California Department of Fish and Game and B. Kus, and permit SC-001504 held by B. Kus. Samples from Orange County sites were provided to us by Kristine Preston (Nature Reserve of Orange County; NROC). A slight modification of the standard protocol with the DNA Tissue Extraction Kit (Qiagen) was used for all DNA extractions. Namely, 20 uL of dithiothreitol was added to the extraction buffer, and the tissue digestion was extended to 48 hours for most extractions. For blood, all extraction volumes were doubled. All DNA extractions were quantified on a Nanodrop spectrophotometer and diluted to 50 ng/µL to ensure consistent PCR amplifications across samples.

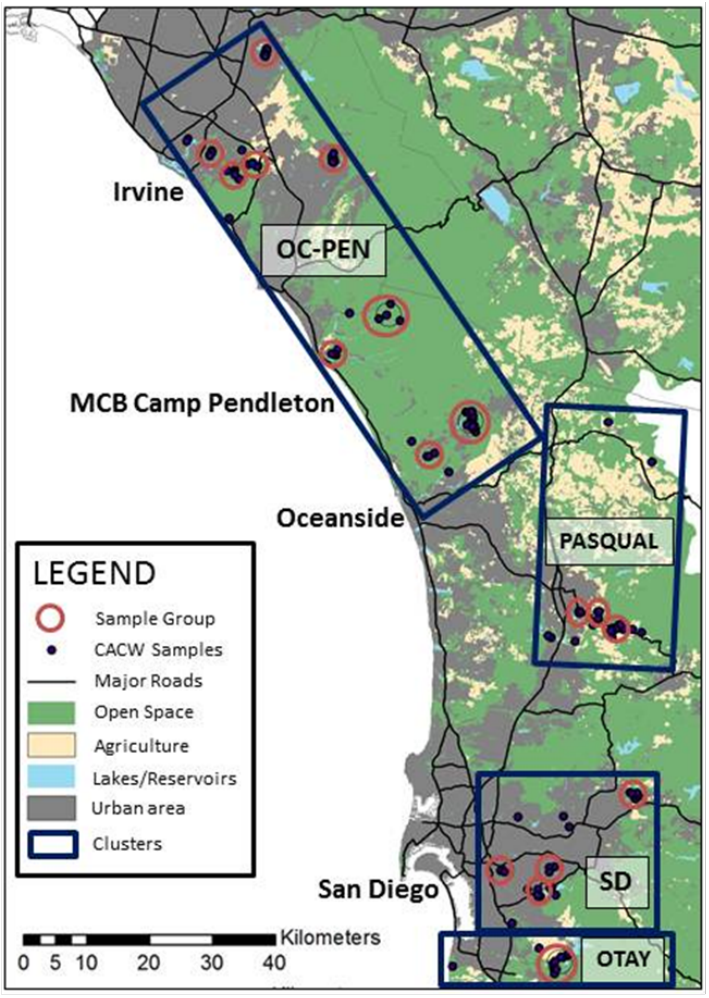


Figure 1. Map of study area. Samples used for isolation by distance analysis are circled in red, and the four major clusters discussed throughout this report in dark blue.

Library Development and Genotyping

We discovered microsatellite loci in the coastal cactus wren genome using a slight modification of the standard development techniques of Hamilton et al. (1999). Libraries were

constructed by excising genomic DNA using the restriction enzyme *Hinc*II, and these fragments were subsequently ligated to an SNX linker. Biotinylated oligonucleotide probes that included both trinucleotide and tetranucleotide repeats were then used to isolate and separate microsatellite repeat regions. These fragments were PCR amplified and sequenced on a next-generation DNA sequencer, a Roche 454 GL FLX, in the Evolutionary Genetics Core Facility (EGCF) at Cornell University. In 3,350 resulting sequences, 414 had microsatellite regions. These sequences were mapped to the Zebra Finch (*Taeniopygia guttata)* genome, thereby providing information on their physical locations and generally facilitating library development. Mapping allowed conserved loci (those that did not differ from the Zebra Finch) and those on the Z chromosome to be dropped from consideration. Since birds are heterogametic, that is one sex has two Z chromosomes (males) and the other a Z and a W (females), loci located on the Z chromosome would confound statistical analyses. W-linked loci were effectively excluded as well, as that chromosome has yet to be sequenced in the Zebra Finch genome. Further, loci without appropriate flanking region for primer design, some that were found to be redundant, and those with complex repeats were eliminated, leaving 52 potential loci. Tests for variation were conducted across each of these loci using a three-primer technique (Schuelke 2000) on seven individuals spread evenly across the sampling area. All genotyping runs occurred on an ABI 3730 DNA Analyzer in the CSUPERB Microchemical Core Facility at San Diego State University.

Twenty-eight loci were divided into three sets, and these were co-amplified across the entire sample set using standard conditions with the Multiplex Kit (Qiagen). These conditions include a 15 minute initial denaturation step at 95 ^{o}C followed by 30 PCR cycles with a 30s denaturation step at 94 ^{o}C, a 90s annealing step at 60 ^{o}C, and a 90s extension step at 72 ^{o}C, and a final extension of 10min at 72^{o}C. Loci were combined as indicated in Table 1. Approximately 10% of samples were amplified and genotyped twice to obtain an error rate. Loci were checked for stepwise mutation consistency using MICRO-CHECKER (van Oosterhout et al 2004), Hardy-Weinberg Equilibrium and linkage disequilibrium by sampling group in GENEPOP ON THE WEB (Raymond & Rousset 1995, Rousset 2008), and for selection using the heterozygosity versus F_{ST} technique of Beaumont (2000) as implemented in LOSITAN (Antao et al 2008). Many of the analyses used here assume loci do not deviate significantly from Hardy-Weinberg Equilibrium, have no significant linkage disequilibrium, and are not under selection.

Genetic Analyses

Several analyses were conducted to explore patterns of population structure and genetic diversity across the study area. First, several Bayesian clustering analyses were used to determine if individuals were arranged in distinct gene pools or populations. For consistency, we refer to these as clusters. These results were verified using several methods, including using multiple clustering algorithms and by attempting to detect very recent movement between identified clusters. Next, a series of techniques for exploring the effects of the landscape on genetic structure patterns were conducted. These analyses help to determine which factors are most important for explaining observed levels of connectivity in the species. Finally, both current and recent past patterns of genetic diversity are described. These measures may provide some inference on the effects of recent events on genetic diversity in the species.

Cluster Inference

Bayesian clustering analyses are individual-based, which means all samples in the dataset are included. These algorithms search for combinations of individuals that can best be grouped together while conforming to expectations of Hardy-Weinberg Equilibrium and linkage disequilibrium. These qualities are expected when a group of individuals is essentially a common gene pool in population genetics terms. STRUCTURE (Pritchard et al 1999) was implemented to explore the structuring of individuals into hierarchical clusters using the correlated alleles model with admixture and without a location prior. For clusters (K) one to 10, a burn-in of 10,000 Markov chain Monte Carlo steps were followed by 100,000 additional steps, and 20 repetitions of the analysis were conducted at each K. The top 10 highest likelihood runs were then analyzed using STRUCTURE HARVESTOR (Earl & van Holdt 2012), which averages and visualizes likelihoods across runs. CLUMPP (Jakobsson & Rosenberg 2007) was then used to combine results of the top 10 most likely runs, and these results were visualized in DISTRUCT (Rosenberg 2004).

Many clustering algorithms exist, each with varying assumptions and strengths. Typically, datasets are analyzed with multiple algorithms to boost the confidence in individual assignments to clusters (Pearse & Crandall 2004). To complement the results from STRUCTURE, another Bayesian clustering analysis was performed in GENELAND (Guillot et al 2005). This analysis takes geographic information into consideration along with genetic data, and has been shown to be accurate for determining recently developed clusters (Guillot 2008). Since we found strong isolation-by-distance (IBD) in our data, the uncorrelated alleles model with admixture was used, testing for Ks between 1 to 10 with 1M Markov chain Monte Carlo repetitions and a 20% burn-in.

Subsets of the data were also analyzed in STRUCTURE guided by the results of the overall clustering analyses. These analyses were conducted with the same parameters as with the total dataset, but only with samples from the highest likelihood clusters identified in the original analysis. Information about genetic substructure and gene flow regimes within these clusters can be gleaned from such hierarchical analyses.

COLONY (Wang 2009) was used to identify closely related individuals in the dataset using a Markov chain Monte Carlo algorithm. Since the pattern of strong genetic isolation by geographic distance would lead to errant conclusions in this analysis using all of the samples, the dataset was analyzed by cluster and in a pairwise manner over sets of clusters. The former strategy allows for family groups to be identified and excluded in the verification of the clustering analyses described above, as family structure can lead to false identification of clusters. The latter strategy was used as a proxy for determining if individuals had recently dispersed across the boundaries between clusters. For this latter strategy, all 420 samples in the dataset were analyzed to increase the power of the COLONY analysis. Sampled adults that are identified as full siblings can be assumed to share the same two parents. COLONY also identifies half sibs (i.e., individuals that share a single parent); however, the power of the algorithm for identifying such relationships decreases with family size in a dataset. A randomly

permuted pseudo-dataset was used to determine the strength of COLONY for identifying full and half siblings among the cactus wrens, with the assumption that none would be identified in this dataset if the original provides enough information for reliable inference.

Landscape Correlations

Both group and individual based metrics of genetic differentiation were calculated, and observed patterns were compared to various landscape factors to determine which are most influential on connectivity. F_{ST}, a measure of genetic differentiation, was measured between the groups designated in Fig 1 using GENEPOP ON THE WEB, and isolation-by-distance determined using a Mantel test in IBDWS (Jensen et al 2005). Groups were determined by combining individuals that were relatively close to one another and without any obvious potential barrier to movement. Isolation by distance is a significant relationship between genetic and geographic distances, whereby the level of genetic differentiation increases with spatial distance. Such a pattern is expected when gene flow is localized with individuals only dispersing between proximate sites, as opposed to panmixia where all aggregations throughout a species' distribution are well-connected.

To evaluate the effects of landscape characteristics on dispersal and gene flow, we approximated suitable cactus wren habitat with a cactus distribution model (Appendix I) created by Kristine Preston (NROC; USGS). Cactus distribution was modeled using a partitioned Mahalanobis D^2 model with one kilometer grid cells from 227 random cactus survey points in both San Diego and Orange Counties, and validated with 97 additional points (Rotenberry et al. 2002, 2006). Variables included in the final model were minimum temperature in January, maximum temperature in July, annual precipitation, aspect, slope, topographical heterogeneity (Sappington's index; Sappington et al. 2007), and percent coastal sage scrub and chaparral within $1km^2$. Developed and agricultural areas were not excluded. A cost surface was created from the model as the inverse of habitat suitability scores and with both urban (derived from USGS impervious surfaces layer; USGS National Map) and agricultural (FRAP vegetation data) areas penalized 10X, effectively masking out human development (Appendix II). Mantel tests were conducted on matrices of genetic differentiation as F_{ST} and least cost path distance through the cost surface. The potential effects of highways and urban areas as barriers to dispersal were assessed using partial Mantel tests controlling for geographic distance. Presence of intervening highways and extensive urban areas were coded binomially, with a 1 if sites were fragmented and 0 if sites were connected through open space. These analyses were conducted within identified genetic clusters as well (Appendix III) using genetic differentiation between individuals, a_r (Rousset 2000). To increase the number of samples for this analysis, two of the clusters were combined (SD and OTAY; see results).

Genetic spatial autocorrelation, which describes fine-scale genetic structure, was calculated in GENALEX (Peakall and Smouse 2012), with 999 permutations to establish significance and 999 bootstraps to obtain a confidence interval. Spatial autocorrelation quantifies the average genetic similarity between each individual and all of those within geographic distance bins from that individual. These patterns can provide inferences on the dispersal regime. Since large-scale genetic structure can confound this analysis (Banks and

Peakall 2012), two analyses were conducted: one with samples from Orange County and Marine Corps Base (MCB) Camp Pendleton/Fallbrook NWS (OC-PEN; see results), and another with those from across the southern portion of the study area (SD and OTAY; see results). Samples around the San Pasqual Valley area (PASQUAL; see results) were too closely aggregated for this analysis, which is most powerful when samples are taken across an even matrix of distances. An initial analysis was performed with size bins increasing by 1000m up to the greatest distance between samples; however, to better display the results, a smaller subset of bins are presented here.

Genetic Diversity

Diversity within identified clusters was measured as allelic richness in FSTAT v. 2.9.3.2 (Goudet 1995) and heterozyosity, both observed and expected, in GENALEX. These statistics were also calculated for major aggregations as well. Tests for heterozygote excesses were conducted in BOTTLENECK (Piry et al 1999). This test is based upon the expectation that allelic diversity is lost more rapidly during a bottleneck than heterozygosity, and hence determines if a significant population decline has recently occurred. Finally, LDNe (Waples & Do 2008) and COLONY were implemented to calculate current effective population sizes, N_e. The former calculates effective population size based upon linkage disequilibrium, and the latter uses a sibship approach. Effective population size is an important parameter in population genetics, as it determines inbreeding rates, the strength of genetic drift, the potential for selection, and the effect of migration. It is closely associated with the number of successful breeding individuals in a population.

Results

Data Quality

Though 420 coastal cactus wrens were sampled, multiple nestlings from the same nest do not represent independent genetic samples; consequently, the dataset was reduced to 168 individuals for analyses here except where noted. Samples provide thorough coverage of the full range of the coastal cactus wren in Orange and San Diego County (Fig 1).

All individuals were genotyped at 28 loci. After testing for departures from Hardy-Weinberg Equilibrium and linkage disequilibrium within all of the major sampling groups and eliminating those with troublesome amplification, 20 loci were used for all subsequent analyses (Table 1). These loci are located across the genome, falling on nine different chromosomes (Table 1). By designing primers to amplify the loci across a wide range of lengths, large numbers could be co-amplified and genotyped in multiplex (MP; Table 1). Total numbers of alleles ranged from three to 18, and overall heterozygosities were generally large, as would be expected with highly polymorphic microsatellites (Table 1). After re-runs, the error rate was found to be negligible (<0.1%), and there are very few missing data from failed amplifications (<1%). Finally, there was no evidence for selection in the 20 loci analyzed here.

Table 1. Information on the 20 loci used for all analyses presented here. Chr = chromosome; MP = multiplex membership; A = total alleles; H_O = overall heterozygosity

Locus	Chr	Forward Primer	Reverse Primer	Repeat Type	Length	MP	A	H_O
CACW3-01	1	ACTGTTCACCCTTGGACCTG	TGTCTGGAAACCACTGAAGAAC	Trinucleotide	250	1	7	0.85
CACW3-02	1	AATGGAAAGGAGCATCAACTG	TTCATGGTGCATACAAGATAGC	Trinucleotide	117	1	5	0.59
CACW3-03	1A	TCCTGAAATGTAATTCAGACACC	CAGAGTGCTACTTAAATTGATTCTTTC	Trinucleotide	262	1	9	0.73
CACW3-04	2	CATGGATAGAGTGAGAACAATATGC	CATGAGATGGACATTATGAGCTG	Trinucleotide	125	1	4	0.31
CACW3-05	2	GATGCATATTGTCAGAGTTCCAC	CTGGACTGAGCTAACAAATGATG	Trinucleotide	141	2	7	0.63
CACW3-06	3	CTCTTTGTTTGACTTAGGAGAACC	AAACCCACCAACCTCTTCC	Trinucleotide	190	2	3	0.52
CACW3-07	4	GCTCAAACTCCTGACCAAGG	TTTTGTACTTTGCTGAAGTCAATTT	Trinucleotide	199	2	5	0.51
CACW3-08	5	GCCCAGGCTCCATCACAG	ATGTCTGCTGCTCCCTCAG	Trinucleotide	98	1	4	0.36
CACW3-09	5	AGGAAGAAATAGAGGTGAGGGAAC	TGACGACTGAACAAAAGTACGAG	Trinucleotide	126	2	5	0.3
CACW3-11	22	TTCTCCTCCCTCTACCTCCTTT	GTGACAACAGAAAATTCCCTTTA	Trinucleotide	183	1	9	0.6
CACW3-12	24	CAGCAGGAGTCTGGAACAGG	TTGGCTGGCAGTGAGGATG	Trinucleotide	290	2	10	0.8
CACW4-01	1	TTTTGCCTAATAAACTGGCTGAC	CACAGAACCACAACCTACATGG	Tetranucleotide	162	3	9	0.74
CACW4-03	1	CCTTACCGAAGTATGCAACAAG	TTGAGATAGAGTGTAGCCATGTG	Tetranucleotide	284	2	10	0.83
CACW4-04	1	TCTCACGTCTTACCATCCTGTG	TTGATACTTGAAACTCTCCTTCTGTC	Tetranucleotide	284	2	8	0.59
CACW4-05	2	GCTCTAAAACTCTGTGGGCAAC	CGAGAACAAGATCATTAACAGCAG	Tetranucleotide	135	2	6	0.69
CACW4-06	2	TTCCTAAGCTCTCTCAATTTCTTACTG	GACTGAATCAAATATGTTATGGCAAC	Tetranucleotide	223	1	16	0.85
CACW4-09	3	GCTAACTGAAAGGGATTGTTGG	TTTCTGGCATGTTTCCTGTC	Tetranucleotide	180	3	18	0.81
CACW4-10	5	GGGTTGGACAAGGTGACATC	TCAATGTGCTTTGCAGGAAG	Tetranucleotide	221	3	16	0.85
CACW4-12	5	CCTGCCACCACTGTATTTCTG	AGAGGCCAAAGACTGAATGG	Tetranucleotide	300	1	4	0.55
CACW4-13	28	GCAGAACTTGGGACTTCGAC	ACTGGGCTTGTTATGGATGG	Tetranucleotide	108	1	6	0.62

Inference of Clusters

Clustering analyses results from STRUCTURE are displayed in Figure 2. The histograms display the assignment probabilities, Q, in columns for each individual arranged from the northernmost on the left to the southernmost on the right. Divisions between individuals are generally imperceptible; rather, general patterns are considered. Common colors represent shared assignment to particular clusters. Partial assignment of an individual is referred to as admixture, which can be from recent gene flow, an indication that clusters have yet to differentiate completely, or simply occur for the lack of enough information to exclusively assign an individual with high probability to a single cluster. An individual that exhibits 50% admixture with one cluster and the same to another could be the offspring of parents from each of the clusters. Unfortunately these analyses cannot definitively discern this from a lack of complete genetic assortment between the clusters.

In coastal cactus wrens, STRUCTURE provided evidence for four current clusters with the highest likelihood (Fig 2): 1, NROC and MCB Camp Pendleton/Fallbrook NWS (OC-PEN), 2, San Pasqual Valley and Lake Hodges (PASQUAL), 3, San Diego and El Cajon (SD), and 4, Otay River (OTAY). Notably, individuals sampled near Bonsall and Pauma Valley (Appendix VI) both grouped into the PASQUAL. While this configuration had the highest likelihood, more information about the relationships of the clusters can also be obtained by examining other STRUCTURE results hierarchically. For instance, at K = 2, a deep division between the general northern and southern areas is identified (Fig 2). At K = 3, a division between San Pasqual Valley and the rest of the southern groups becomes evident, and at K =4, the most southern samples, those near Otay Lakes, are separated from those around San Diego (Fig 2). Throughout each level of K, the northern group, composed of samples from throughout MCB

Camp Pendleton and Fallbrook NWS northward into Orange County, remained static (Fig 2). At K = 5, the 3 southern clusters are evident, but the northern group exhibits partial admixture at the sites intermediate to the most distant ones, a signal of isolation-by-distance (Fig 2).

The results of additional analysis in GENELAND are perfectly congruent with the evidence from STRUCTURE for 4 current clusters and with the same individuals assigned to each of those clusters (Appendix IV). These results from do not provide the same opportunity for

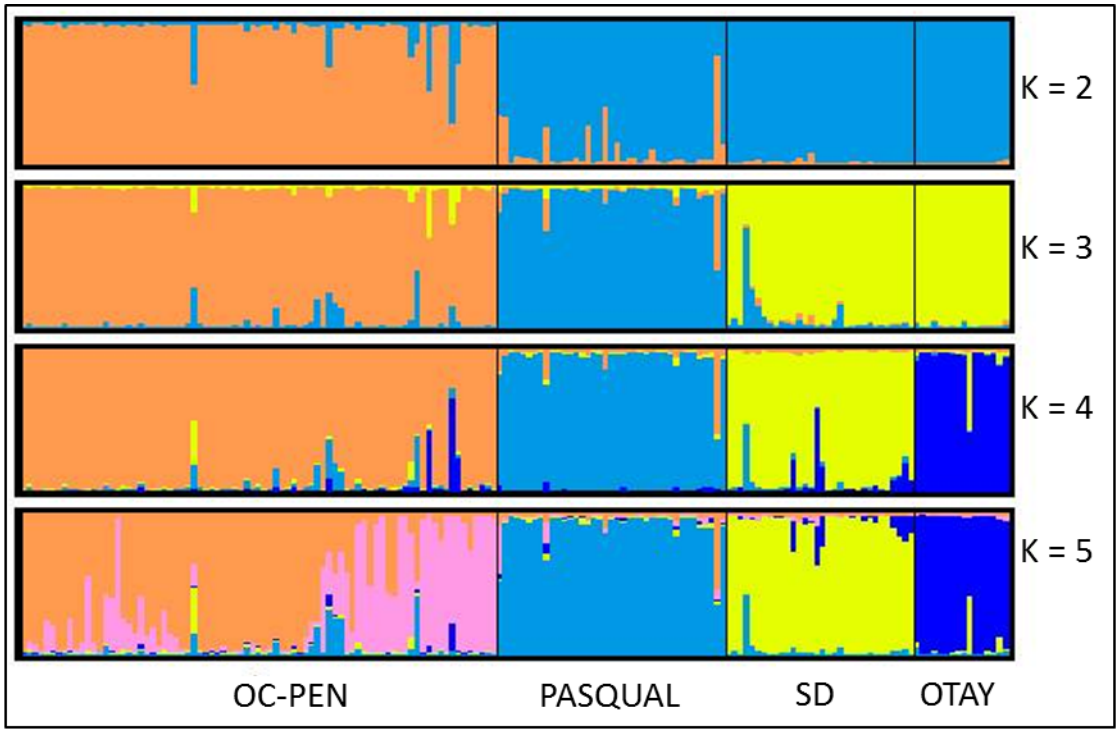


Figure 2. STRUCTURE results. OC-PEN = all samples from Orange County and Marine Corps Base Camp Pendleton, PASQUAL = samples in and near San Pasqual Valley, SD = samples generally around the city of San Diego, Otay = samples around the Otay River. K = number of clusters.

exploratory analyses as described in the STRUCTURE analyses above, and hence alternative levels of K are not described here. The consistency between the two different methods provides further confidence in the veracity of the results from STRUCTURE.

All full sibships detected in COLONY were composed of individuals sampled within each of the 4 clusters. None were detected among clusters, indicating that no sampled individuals had recently dispersed across the boundaries indicated between clusters at K = 4. Testing the validity of COLONY sibship assignments with the permuted dataset revealed that while full sibships could be reliably identified, our data do not provide the power to detect half sibships without spurious results. Hence, we did not consider the half sibship results.

STRUCTURE results from within cluster analyses exhibit contrasting results (Fig 3). In OC-PEN, for instance, the signal of isolation by distance is evident as an admixture cline at K=2

(Fig 3), particularly between samples east of I-5 and throughout MCB Camp Pendleton/Fallbrook NWS. At K=2 in PASQUAL, all samples are evenly admixed, providing evidence for genetic panmixia between the sample sites (Fig 3). Conversely, in an analysis that combines SD and OTAY to boost the number of samples (only 16 are available in the latter cluster) and assuming K =3, a handful of individuals are admixed between the three. Most individuals, on the other hand, cluster strongly into geographic groups (Fig 3). The number of clusters presented here are from the analyses that exhibited high likelihoods and low variance between runs in the cases of OC-PEN and SD & OTAY, and to display the signal of panmixia in PASQUAL where the highest likelihood is actually at K = 1.

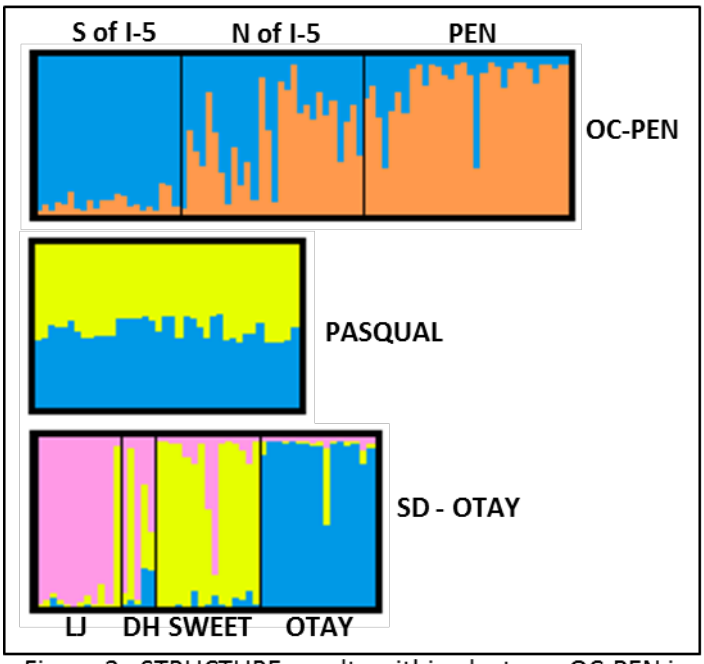


Figure 3. STRUCTURE results within clusters. OC-PEN is divided into sites south of I-5, north of I-5, and Camp Pendleton/Fallbrook Naval Weapons Station (PEN). SD and OTAY were combined for analysis, and SD is divided into Lake Jennings and nearby sites (LJ), Dictionary Hill (DH) and Sweetwater Reservoir, Encanto Canyon, and nearby sites (SWEET).

Landscape Analyses

Pairwise F_{ST} among the 17 groups ranged from 0.019 to 0.2, and overall IBD was significant ($r = 0.5545$, $p < 0.001$; Fig 4), indicating that genetic differentiation generally increases with geographic distance. Least cost path distance (Appendix II) was also positively correlated with genetic differentiation, and with a similar correlation coefficient (Table 2). While controlling for geographic distance, partial Mantels revealed significant effects of fragmentation both by highways and urbanization (Table 2).

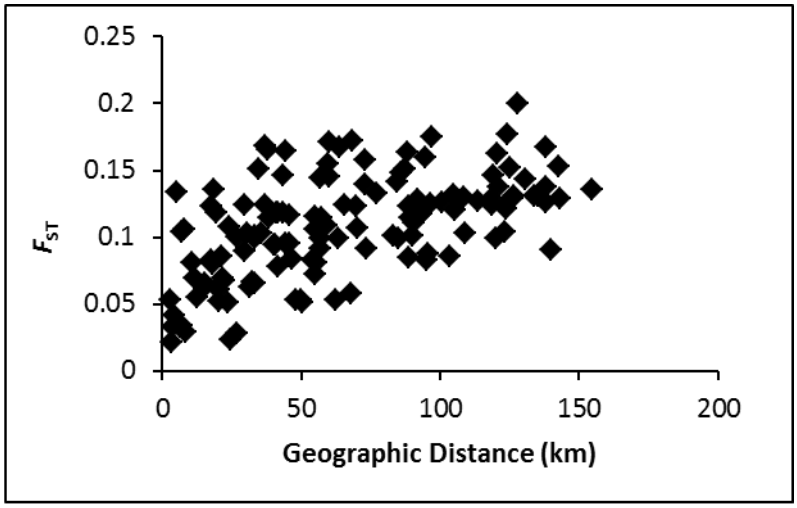


Figure 4. Pairwise genetic distance versus geographic distance between the 17 groups encircled in Figure 1.

Results within clusters varied. Significant relationships between geographic distance and genetic differentiation were detected in each (Table 2). Within OC-PEN, least cost path distance showed a slightly better fit than simple Euclidean geographic distance; however, in the SD-OTAY combined analysis, Euclidean distance was more strongly correlated with genetic differentiation than least cost path distance (Table 2). Furthermore, a significant effect of highways and urbanization was detected in OC-PEN using the fragmentation matrices, but not in either PASQUAL or the combined SD-OTAY analysis (Table 2).

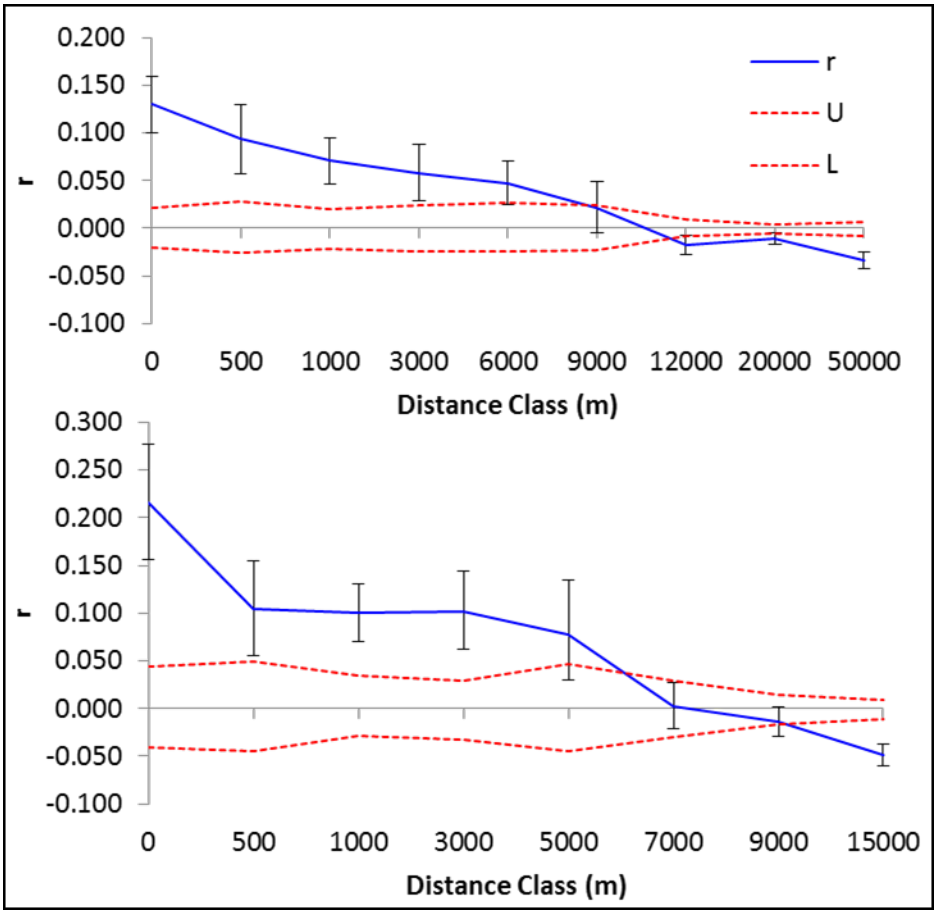
Table 2. Mantel test results on population or individual genetic differentiation and either geographic or least cost path distance, and partial Mantel tests results on fragmentation while controlling for geographic distance.

Site	Euclidean Distance		Least Cost Path		Highways		Urban*	
	r	p-value	r	p-value	r	p-value	r	p-value
Across Sampling Areas	0.5545	<0.0001	0.5533	<0.0001	0.1814	<0.0001	0.2914	<0.0001
Within Clusters								
OC-PEN	0.1496	0.0002	0.1762	< 0.0001	0.1685	< 0.0001	0.2409	<0.0001
PASQUAL	0.2271	0.006	0.2435	0.0015	0.0241	NS	0.0597	NS
SD - OTAY	0.3907	< 0.0001	0.3418	< 0.0001	-0.0326	NS	-0.0237	NS

*Pasqual includes agricultural development as well.

Spatial autocorrelation analyses indicate very strong positive relationships up to 9km in OC-PEN and 5km in SD-OTAY (Fig 5), each of these being the farthest extents at which *r* is significant (OC-PEN: r = 0.022, p = 0.037; SD-OTAY: r = 0.078, p = 0.001). These distances can be assumed to be related to effective dispersal distances. While some dispersers may go farther than 9km in OC-PEN or 5km in SD-OTAY, many more individuals move those distances or

less. There are too few individuals dispersing farther than these distances to exceed the signal created by unrelated cactus wrens beyond those distances.



Genetic Diversity

Basic indices of genetic diversity, heterozygosity, both observed and expected, and allelic richness, were generally similar across the sampling range (Table 3). Effective population sizes as estimated via the linkage disequilibrium method and the sibship methods were larger in OC-PEN and PASQUAL than in either SD or OTAY (Table 3). The methods were relatively congruent in their estimations, with the only notable discrepancy at PASQUAL (Table 3). General patterns follow expectations given the number of territories known to exist within clusters. Heterozygote excesses at three of the clusters indicate recent severe population declines in these areas (Table 3). Results by major aggregation are presented in Appendix V.

Table 3. Indices of genetic diversity. H = average heterozygosity, both observed (o) and expected (e), A = average alleles, *A* = allelic richness, LDNE = Ne estimated from LDNE, SibNe = Ne from the sibship method, Bottleneck = * if there is a significant test for a recent decline in Ne.

Cluster	Samples	H_O	H_E	A	*A*	LDNE	SibNe	Bottleneck
OC-PEN	81	0.60	0.64	6.7	5.5	83.1 (63.3 - 115.3)	110 (80 - 151)	*
PASQUAL	39	0.66	0.66	5.8	5.1	159.5 (83.2 - 895.9)	58 (38 - 93)	*
SD	32	0.60	0.63	6.1	5.4	48.2 (36.4 - 68.3)	44 (27 - 75)	
OTAY	16	0.69	0.65	5.0	5.0	28.6 (20.1 - 46)	40 (19 - 120)	*

Discussion

Coastal cactus wrens have previously been suggested to have a relatively restricted dispersal regime, based on direct observations that documented only a handful of movements over 5km (Unitt 2004). The genetic data support this implication, as a strong signal of isolation by distance illustrates that dispersing movements are not spreading genes across the full study area in Orange and San Diego Counties. Rather, coastal cactus wren movements are spatially constrained. Since this indicates gene flow would like only occur between relatively close sites, connectivity between disparate sites would be considerably low even in a natural landscape. This conclusion is further strengthened by significant signals of spatial autocorrelation up to 9km in OCPEN and 5km in SD and OTAY. Both the IBD and the spatial autocorrelation support a stepping stone gene flow model, wherein distant sites may be connected with intervening habitat.

Despite the signal of isolation by distance, gene flow does appear to be disrupted between major sites. For instance, while genetic connectivity is high enough for sites in Orange County and both MCB Camp Pendleton and Fallbrook NWS to form a single cluster, three other clusters were identified across the rest of San Diego County (Fig 2). Both isolation by distance (Meirmans 2012) and the presence of family structure (Anderson & Dunham 2008) can confound these clustering analyses by artificially raising the detected number of groups. Isolation by distance, for instance, naturally creates clines in allele frequencies. Samples at disparate points along such a cline would have very different allele frequencies than one another, and thus may be incorrectly designed as independent clusters by the algorithms in STRUCTURE and GENELAND. Understanding that IBD is significant, as it is here, helps to choose the correct allele frequency model for each of these analyses (see methods); additionally, understanding the system being studied helps to determine if the identified clusters are accurate. For instance, PASQUAL is geographically isolated from the other clusters; however, it is important to note that little to no occupied habitat exists in the unsampled area between this cluster and the others. Furthermore, OC-PEN covers a spatial range of 80km while PASQUAL, SD, and OTAY cover 60 km combined. If the IBD signal was causing incorrect clustering of samples, it would certainly be expected to appear within OC-PEN as well. In the K = 5 analysis of the overall dataset (Fig 2) and in the individual cluster analyses (Fig 3), OC-PEN exhibits the pattern of admixture that would be expected if localized, stepping-stone gene flow was

presently on-going across PASQUAL, SD, and OTAY. The issue of family groups is much simpler to resolve. Several recent simulation studies have illustrated that removing one member of any identified full sibship relationships from a dataset and conducting clustering analyses is the best solution (Anderson & Dunham 2008, Rodriguez-Ramilo & Wang 2012). Several such relationships were identified in COLONY in each of the clusters. To be robust, a member of each full sibship was removed regardless of the probability of that relationship, and this reduced dataset was analyzed in both STRUCTURE and GENEPOP with the same models and parameters described in the methods section. Four clusters were still identified, although OTAY is less differentiated from SD in STRUCTURE than with the full dataset; however, the group is distinct in GENELAND (data not shown). It is unlikely that IBD combined with a poor sampling scheme or the presence of family groups would be better explanations for the identified clusters than a lack of gene flow among them.

With the current number of clusters established, further exploration of various levels of K can help to make some inferences about substructure among them. Such inferences require disentangling the effect of geographic distance, which is established here a major factor in population differentiation in the cactus wren, from landscape alteration. For instance, a seemingly deeper relationship between PASQUAL, SD, and OTAY versus OC-PEN evident at K = 2 (Fig 2) may be caused by a rift in suitable habitat along southern MCB Camp Pendleton, as predicted in the cactus habitat model (Appendix I). The model also predicts an extensive band of poor habitat between SD and PASQUAL in the vicinity of Scripps Ranch and Marine Corps Air Station Miramar (Appendix I), which might explain the closer relationship between the former cluster and OTAY at K = 3 (Fig 2). The model predicts that most of the high suitability cactus habitat occurs along the coast and overlaps extensively with urban development (Appendix II); hence any connectivity between SD and PASQUAL that may have historically circumvented this predicted potential barrier to the west has likely been significantly reduced or eliminated by habitat loss.

Considering the STRUCTURE and landscape correlation results within the clusters reveals highly variable gene flow regimes across the sample range. For instance, the intermediate admixture signal evident in OC-PEN (Fig 3) would be expected under IBD, with distant sites connected by aggregations between them. While disparate sites are well-differentiated genetically, gene flow would still be expected between them over several generations of dispersal. It should be noted, however, that habitat fragmentation is having a discernible genetic effect in this cluster. Sites west of I-5 (Appendix VI), for instance, are more genetically differentiated from one another and from other OC-PEN sites than would be predicted by geographic distance alone (Table 2). The extensive admixture exhibited in PASQUAL (Fig 3) implies that gene flow is evenly distributed throughout the cluster, and connectivity is presumably high between each one of the sample sites. Hence, Lake Hodges and the rest of San Pasqual Valley, including Bear Valley, the San Diego Zoo Safari Park, and the San Pasqual Battlefield (Appendix VI) appear to be genetically panmictic. These analyses reflect relatively recent changes in population structure and gene flow, but any decline in connectivity between Lake Hodges and the San Diego Zoo Safari Park that may have resulted from the 2007 Witch Creek Fire may not be evident yet. An entirely different gene flow regime is apparent in the

analysis of samples from the southern part of San Diego County. Rather than a cline in admixture implying IBD, three groups are evident: one around Lake Jennings, another near Sweetwater Reservoir and Encanto Canyon, and a third near Otay Lakes (Appendix VI) that is also identified in the overall clustering analyses (Fig 3). Interestingly, while some individuals throughout the sample set appear admixed between the clusters, implying gene flow in some recent generations, other samples are identified that may have been dispersers themselves or the offspring of very recent dispersers. For instance, one adult sampled around the Sweetwater Reservoir and another in the Dictionary Hill area clustered with Lake Jennings individuals (Fig 3). Rather than regular connectivity through multiple-generation stepping stone movements, such as is evident across OC-PEN, or free-flowing gene flow, such as in PASQUAL, dispersing movements around these southern clusters are likely more intermittent, with individuals only making successful long distance movements infrequently. The lack of such longer distance movements may also explain the much shorter significant autocorrelation distance detected in SD-OTAY (5km) versus that in OC-PEN (9km). These patterns and levels of differentiation observed between groups in SD-OTAY, among the highest across the full study extent (data not shown), indicate that gene flow between the aggregations around Lake Jennings and Sweetwater Reservoir-Encanto Canyon is relatively low.

While geographic distance is a strong factor in explaining overall levels of differentiation, landscape features are implicated as well. For instance, two major aggregations of cactus wrens in southern San Diego County, one around Sweetwater Reservoir and another near Otay Lakes (Appendix VI), would be predicted to be connected since they are only 10km apart and with nearby open space around Otay Lakes and the San Diego National Wildlife Refuge (Appendix III); however, these sites are identified as being part of different clusters in both STRUCTURE (Fig 2, 3) and GENELAND (Appendix IV). No full sibships were identified between the clusters, and clustering analysis did not identify any recently dispersed individuals between the sites (Fig 2, 3). One individual nestling sampled in OTAY did exhibit some admixture with the Sweetwater Reservoir-Encanto Canyon group (Fig 3), possibly the product of some gene flow in prior generations. Another consideration is that if gene flow were recently disrupted, not enough time may have passed for the clusters to completely differentiate. Considering the strong results in all clustering analyses, it would seem that gene flow between these sites has been significantly disrupted.

Determining the causes of the disruption in connectivity between SD and OTAY using statistical evaluations of landscape features is inconclusive here, as genetic differentiation is significantly correlated with geographic distance but not with fragmentation by urbanization (Table 2). It should be noted, however, that the power of these analyses is very low without a relatively large number of samples. With only 16 samples from OTAY and very few additional territories known to exist in the area, there may simply not be enough cactus wrens for this analysis. The power of partial Mantel tests can be significantly reduced when there are too few pairwise comparisons. Furthermore, the open space just to the east of OTAY and the Sweetwater Reservoir was predicted to have poor habitat suitability (Appendix I) and has burned in recent wildfires (Appendix III; the 2003 Mine/Otay and 2007 Harris Fires). Lack of suitable habitat would likely limit the use of the area by cactus wrens. Therefore, all of the

sampled aggregations in SD and OTAY are probably subject to some amount of habitat fragmentation, either by urbanization or unsuitable habitat (Appendix II), which also reduces the power of the partial Mantel test. Fragmentation effects may be more detectable in the partial Mantel test in OC-PEN simply because there are a large number of samples that are well-spaced throughout suitable contiguous habitat as well as some within fragmented sites (Appendix II). Hence we are reticent to conclude that fragmentation effects are not present in SD and OTAY.

It is possible that the levels of differentiation observed among highly fragmented sites may result from a lack of successful breeding by dispersing individuals, rather than only a lack of movement. Some of these areas have very limited available habitat, and therefore all potential territories may be occupied. Analyses indicate that cactus wrens do exhibit extra-pair paternity (data not shown; K. Preston, pers. comm.), as do many songbirds, and may practice egg-dumping; hence, individuals without territories, or floaters, may still be able to contribute to the next generation. On the other hand, if individuals can disperse between sites but are not breeding, those individuals do not confer gene flow between those sites. These are questions that warrant more study.

Genetic diversity is evenly distributed across the four current clusters (Table 3); however, disruptions in gene flow are evident in population structure long before genetic diversity is affected (Leberg et al 2010). This is because genetic drift, the random survival of alleles from one generation to the next (i.e., some individuals pass on their genes successfully, others do not), causes populations to differentiate from one another much more rapidly than it confers loss of alleles. Strong indications of bottlenecks in three of four clusters may provide some insight on future patterns, as declines in genetic diversity would be expected in these clusters (Table 3). These recent drops in population sizes could be the result of major wildfires within two of the clusters, OC-PEN and PASQUAL. Though no bottleneck was detected in SD, it is likely that populations here are also decreased from historical sizes (Shuford & Gardali 2008). The method used here to detect declines in population size is only sensitive to very recent events, and hence would not detect historical drops in N_e. Finally, because N_e can be reduced by emerging population structure (England et al 2010), the bottleneck detected in the OTAY cluster may be a consequence of a recent reduction in gene flow with surrounding sites.

Effective population sizes mirror expectations given the number of known birds in the sample areas and the expectation that N_e should be less than census sizes (Frankham 1995). The discrepancy between the LD method and the sibship methods in PASQUAL should not be discouraging in terms of their accuracy. Typically, contrasting results such as these are averaged using their harmonic mean (Waples & Do 2010), which would be 85 in PASQUAL. Estimations of N_e are interpreted in a comparative manner, and to determine the extent to which populations have lost adaptive potential (Leberg 2005). Theory predicts minimum N_e thresholds of 50 to avoid the negative effects of inbreeding, 500 to prevent the loss of diversity through genetic drift, and 5000 to persist in evolutionary time (Traill et al 2010); however, it should be noted that gene flow has been shown to counter the loss of genetic diversity even when weak (Palstra & Ruzzante 2008). Estimates of N_e can also infer gene flow from

unsampled, large populations (Waples & Do 2010), which would be indicated with larger than expected estimations given known census sizes. Considering that the N_e estimates in OTAY (28.6 and 40) are similar to the known number of adults there (42 - 44), for instance, it may be unlikely that the group is well connected to significant unsampled aggregations across the border in Baja California, Mexico. The confidence intervals for OTAY from both LDNe (20.1 – 46) and the sibship method (19 – 120), however, do afford the possibility that some other groups are contributing gene flow into the cluster. This is not a question that can be definitively answered without samples from cactus wrens in northern Baja, MX.

Importantly, populations with lower effective sizes experience more rapid change in population structure and genetic diversity, meaning that populations would be more susceptible to recent processes than would larger ones. This may explain the surprising sorting of relatively proximate aggregations into different clusters at the most southern portion of the sample range. With such strong IBD and low effective population sizes, the removal of stepping stones between groups may have led to rapid differentiation. In theory, if the gene flow regime observed in the OC-PEN cluster were in place across southern San Diego County, the southern sites would cluster together and exhibit a similar clinal pattern (Fig 3). Conversely, if a barrier to gene flow formed between Orange County and MCB Camp Pendleton/Fallbrook NWS, those sites would be expected to rapidly differentiate from one another, with the alleles that were once shared between those areas drifting out of the populations with time. Such a barrier is evident in the area separating SD and OTAY (Fig 3).

Management Implications

Perhaps the most important implication of these genetic analyses is localized gene flow. Distant cactus wrens are only genetically connected through intermediate sites. In the absence of such sites, it would be expected that distant aggregations will rapidly differentiate because of relatively small effective population sizes across the sample range. Consequently, it appears that while southern Orange County and MCB Camp Pendleton/Fallbrook NWS retain genetic connectivity, this group, San Pasqual Valley, and southern San Diego County are well differentiated from one another likely for the lack of intervening habitat and cactus wrens.

Within southern San Diego County, Dictionary Hill and Encanto Canyon are enveloped by urban development, and these sites seem to retain some genetic connectivity with Sweetwater Reservoir (Appendix VI). Given the low level of admixture in the largest aggregations in SD, namely Lake Jennings and Encanto Canyon-Sweetwater Reservoir (Fig 3), it seems that genetic connectivity between the sites is relatively low. Management and restoration of cactus habitat between these sites might help to reduce the potential for local extirpation from stochastic processes. As it is, gene flow might restore either population, but it may be a very slow process given the level of movement predicted from analyses here.

A stronger barrier seems to be in place between Sweetwater Reservoir and Otay (Fig 2, 3). Given the observed connectivity elsewhere over a relatively limited range but across some urban areas, restoration of cactus habitat between these aggregations may restore connectivity. While protected land exists to the east of these sites, much of it recently burned;

furthermore, the habitat model predicts the area to be poorly suited for cactus (Appendix II). Finer spatial scale modeling would help to identify potential movement corridors that could warrant conservation attention in the future to re-establish connectivity between OTAY and SD.

Future Study

During the 2012 breeding season, additional cactus wrens have been sampled from aggregations throughout Orange, Riverside, Los Angeles, San Bernardino and Ventura Counties. These samples will provide a broader perspective on general population structure and genetic diversity in the cactus wren throughout its entire southern California range, with completion of analyses anticipated in 2013. These patterns will also help to make further inferences on the effects of habitat fragmentation and genetic isolation in the declining songbird.

Using a large number of coastal cactus wrens sampled in the early 1900s and now in museum collections, the patterns of structure reported here will also be compared to those of historical populations. Analyzing these historical samples will provide information on population structure in the coastal cactus wren prior to the extensive urbanization of San Diego County in the latter half of the 20th century. Comparing these patterns with the population structure reported here will further elucidate the causes of reductions in genetic connectivity, whether by geographic distance, fragmentation, or natural, historical barriers.

Finally, an on-going resighting study will help to provide further insight on dispersal in the coastal cactus wren. Sample sites and all other known, accessible cactus habitat in San Diego County are being visited during the 2012 breeding season. These results will provide direct information on recent movement between aggregations. Combining resighting, genetic data, and finer resolution habitat suitability models will improve our ability to identify optimal linkages among current aggregations of cactus wrens for protection and restoration.

Acknowledgements

Funding for this project was provided by SANDAG, the Nature Reserve of Orange County, California Department of Fish and Game and the U.S. Geological Survey Western Ecological Research Center. We thank Kristine Preston for providing samples, expertise on cactus wren behavior and movement, and creating the cactus habitat model. K. Moore and D. Kamada of NROC, and K. Allen, L. Allen, P.J. Falatek, A. Gallagher, A. Houston, S. Howell, and M. Madden-Smith of USGS, as well as the authors, shed ample sweat, and a little blood, to acquire samples. Many organizations and private landowners helped by providing site access and information on cactus habitat and cactus wren presence including: City of San Diego, County of San Diego, City of Chula Vista, San Diego Gas and Electric, San Dieguito River Park, CalTrans, Pala Band of Mission Indians, Helix Water District, California State Parks and Recreation, U.S. Fish and Wildlife Service, CACW Working Group, San Diego National Wildlife Refuge, California Department of Fish and Game, San Diego Zoo Institute for Conservation Research, Conservation Biology Institute, The Nature Conservancy, AECOM, Fallbrook NWS, MCB Camp Pendleton, San Diego Audubon Society, Sweetwater Authority, San Diego Monitoring and Management Program, San Dieguito River Valley Conservancy, City of Irvine, Crystal Cove State Park, Orange County Parks, Southern California Edison Viejo Conservation Easement, and UC-Irvine Ecological Preserve. The use of trade names does not imply endorsement by the U.S. Geological Survey.

References

Anderson, EC and KK Dunham (2008) The influence of family groups on inferences made with the program STRUCTURE. Molecular Ecology Resources 8: 1219-1229.

Antao, T et al (2008) LOSITAN: A workbench to detect molecular adaptation based on a FST-outlier method. BMC Bioinformatics 9: 323.

Banks, SC and R Peakall (2012) Genetic spatial autocorrelation can readily detect sex-biased dispersal. Molecular Ecology 21: 2092-2105.

Charlesworth, D and B Charlesworth (1987) Inbreeding depression and its evolutionary consequences. Annual Review of Ecology and Systematics 19: 237-268.

Doerr, VAJ, ED Doerr, and MJ Davies (2011) Dispersal behavior of brown treecreepers predicts functional connectivity for several other woodland birds. Emu 111: 71-83.

Earl, DA and BM von Holdt (2012) STRUCTURE HARVESTER: a website program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4: 359-361.

England, PR, G Luikart, and RS Waples (2010) Early detection of population fragmentation using linkage disequilibrium estimation of effective population size. Conservation Genetics 11: 2425-2430.

Frankham, R (1995) Effective population size/adult population size ratios in wildlife: a review. Genetical Research 66: 95-107.

Goudet, J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Online.

Guichoux, E et al (2011) Current trends in microsatellite genotyping. Molecular Ecology Resources 11: 591-611.

Guillot, G (2008) Inference of structure in subdivided populations at low levels of genetic differentiation—the correlated allele frequencies model revisited. Bioinformatics 24: 2222-2228.

Guillot, G, F Santos, and A Estoup (2008) GENELAND: a new algorithm to deal with null alleles and a friendly graphical user interface. Bioinformatics 24: 1406-1407.

Hamilton, MB, EL Pincus, A DiFiore, and RC Fleischer (1999) Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. *BioTechniques* 27: 500-507.

Hemmings, NL, J Slate, and TR Birkhead (2012) Inbreeding causes early death in a passerine bird. Nature Communications 3.

Jakobsson, M and NA Rosenberg (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23: 1801-1806.

Jensen, JL, AJ Bohonak, and ST Kelley (2005) ISOLATION BY DISTANCE, web service. BMC Genetics 6: 13.

Leberg, PL and RC Vrijenhoek (1994) Variation among desert topminnows in their susceptibility to attack by exotic parasites. Conservation Biology 8: 419-424.

Leberg, PL (2005) Genetic approaches for estimating the effective size of populations. Journal of Wildlife Management 69: 1385-1399.

Leberg, PL et al (2010) Implications of landscape alteration for the conservation of genetic diversity of endangered species. In Molecular Approaches in Natural Resource Conservation and Management (eds JA DeWoody et al), p. 212-238. Cambridge University Press, New York, NY.

Lowe, WH and FW Allendorf (2010) What can genetics tell us about population connectivity? Molecular Ecology 19: 3038-3051.

Meirmans, PG (2012) The trouble with isolation by distance. Molecular Ecology 21: 2839-2846.

Palstra, FP and DE Ruzzante (2008) Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for wild population persistence? Molecular Ecology 17: 3428-3447.

Peakall, R and P Smouse (2012) GENALEX 6.5: Genetic analysis in EXCEL. Population genetic software for teaching and research—an update. BIOINFORMATICS: in press.

Pearse, DE and KA Crandall (2004) Beyond FST: analysis of population genetic data for conservation. Conservation Genetics 5: 585-602.

Pollak, D (2001) Natural Community Conservation Planning: the origins of an ambitious experiment to protect ecosystems. California Research Bureau, Sacramento, CA.

Piry, S, G Luikart, and JM Cornuet (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, JK, M Stephens, and P Donnelly (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

Quattro, JM and RC Vrijenhoek (1989) Fitness differences among remnant populations of the endangered Sonoran topminnow. Science 245: 967-978.

Raymond, M and F Rousset (1995) GENEPOP v1.2: population genetics software for exact tests and ecumenism. Journal of Heredity 86: 248-249.

Reed, DH and R Frankham (2003) Correlation between fitness and genetic diversity. Conservation Biology 17: 230-237.

Rosenberg, NA (2004) DISTRUCT: a program for the graphic display of population structure. Molecular Ecology Notes 4: 137-138.

Rotenberry, J.T., S.T. Knick and J.E. Dunn. 2002. A minimalist approach to mapping species' habitat: Pearson's planes of closest fit. Pages 281-289 in J.M. Scott, P.J. Heglund, M.L. Morrison, J.B. Haufler, M.G. Raphael, W.A. Wall and F.B. Samson, eidtors. Predicting species occurrences: issues of accuracy and scale. Island Press, Washington, D.C., USA.

Rotenberry, J.T., K.L. Preston and S.T. Knick. 2006. GIS-based modeling for mapping species' habitat. Ecology 87:1458-1464.

Rousset, F (2000) Genetic differentiation between individuals. Journal of Evolutionary Biology 13: 58-62.

Rousset, F (2008) GENEPOP '007: a complete reimplementation of the GENEPOP software for Windows and Linux. Molecular Ecology Resources 8: 103-106.

Sappington, J.M., K.M. Longshore and D.B. Thompson. 2007. Quantifying landscape ruggedness for animal habitat analysis: A case study using bighorn sheep in the Mojave Desert. Journal of Wildlife Management 71:1419-1426.

Schuelke, M (2000) An economic method for the fluorescent labeling of PCR fragments. Nature Biotechnology 18: 233-234.

Shuford, WD and T Gardali (2008) 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 No. 1, Western Field Ornithologists, Camarillo, CA, and the California Department of Fish and Game, Sacramento, CA.

Sunnucks (2000) Efficient genetic markers for population biology. TREE 15: 199-203.

Traill, LW et al (2010) Pragmatic population viability targets in a rapidly changing world. Biological Conservation 143: 28-34.

Unitt, P (2004) San Diego county bird atlas. Proceedings of the San Diego Society of Natural History 39.

van Oosterhout, C et al (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes 4: 535-538.

Villard, MA and S Hache (2012) Conifer plantations consistently act as barriers to movement in deciduous forest songbird: a translocation experiment. Biological Conservation 155: 33-37.

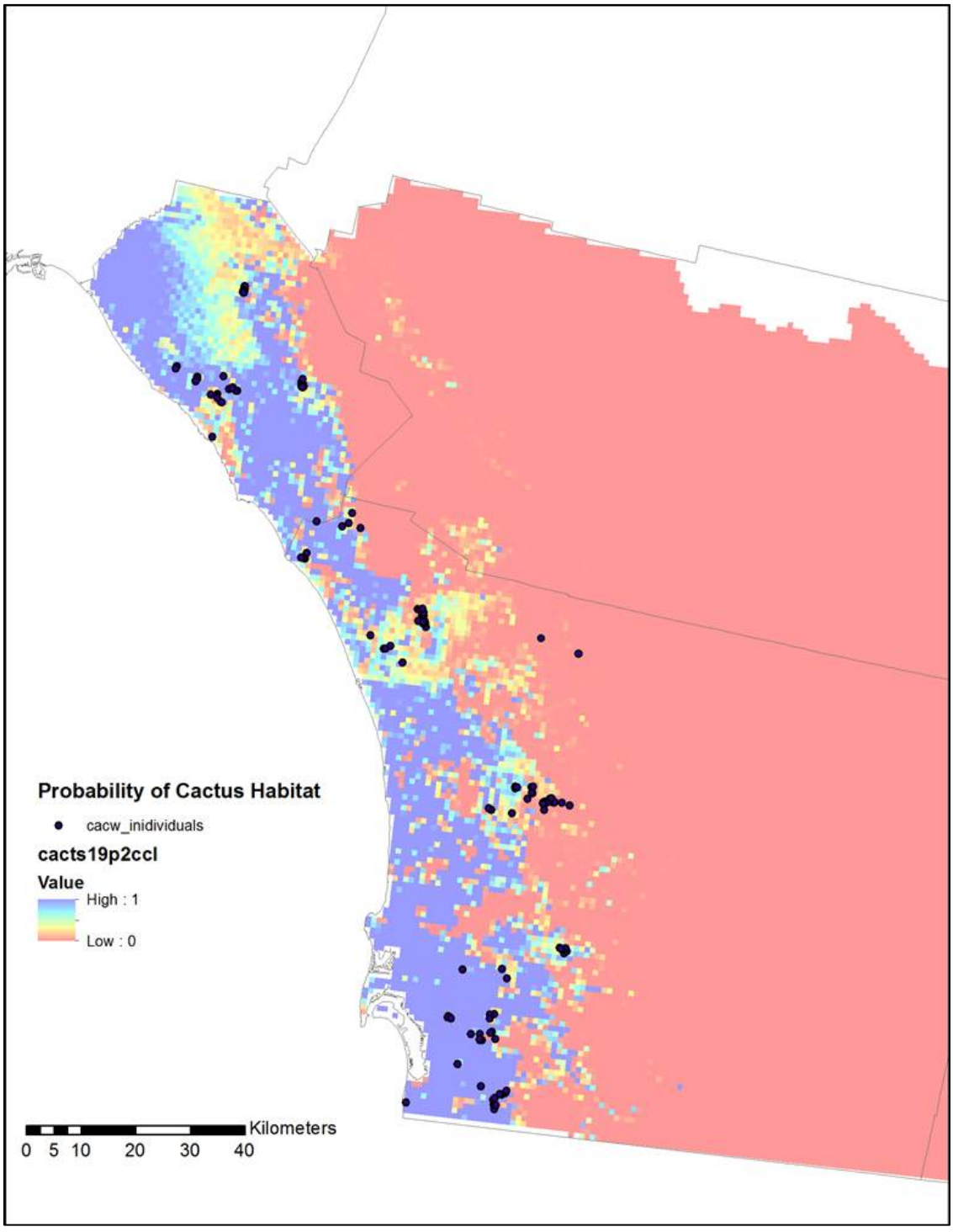
Wang, J (2009) A new method for estimating effective population sizes from a single sample of multilocus genotypes. Molecular Ecology 18: 2148-2164.

Waples, RS and C Do (2008) LDNE: a program for estimating effective population size from data on linkage disequilibrium. Molecular Ecology Resources 8: 753-756.

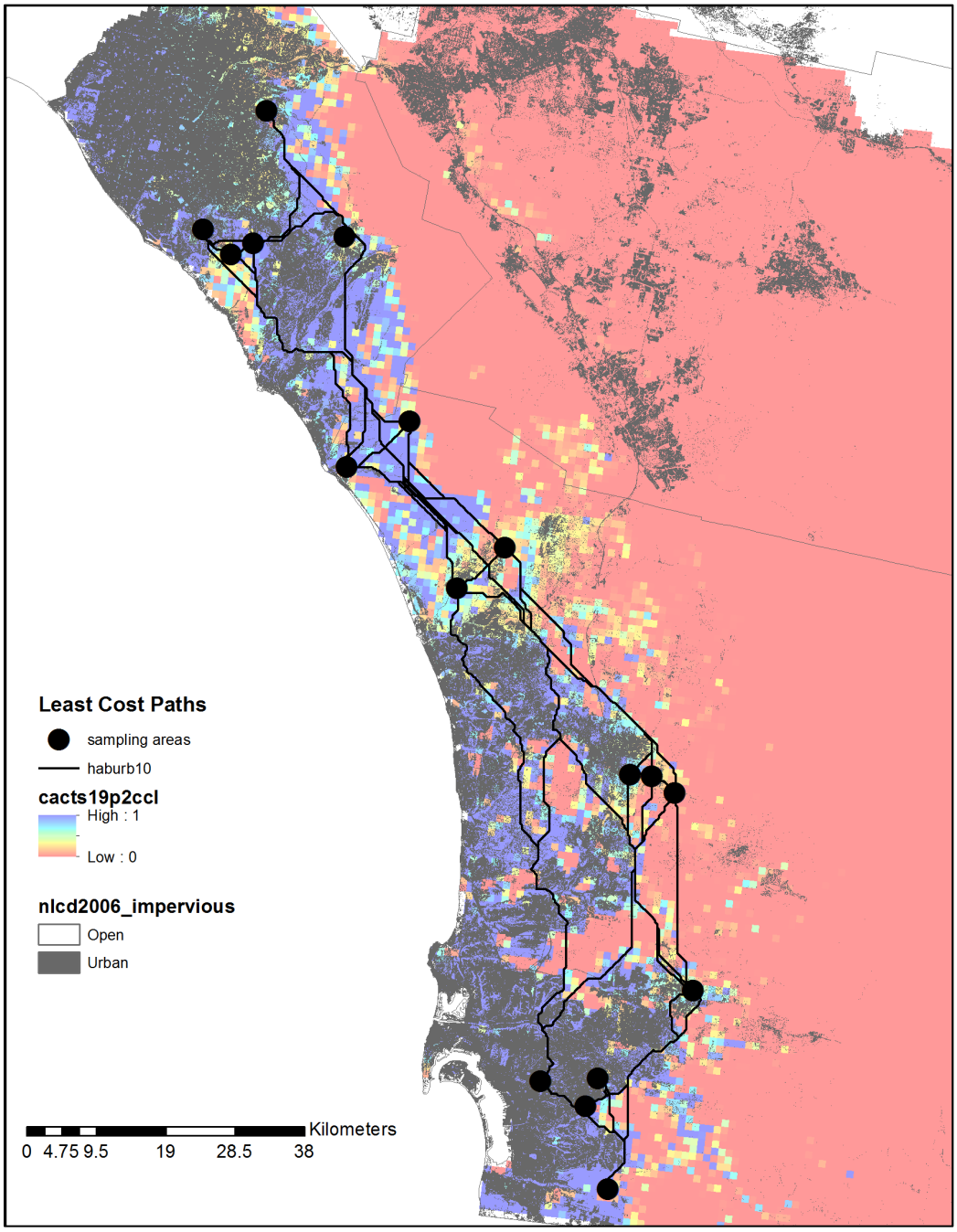
Waples, RS and C Do (2010) Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: a largely untapped resource for applied conservation evolution. Evolutionary Applications 3: 244-262.

Zhang, DX and GM Hewitt (2003) Nuclear DNA analyses in genetics studies of populations: practice, problems, and prospects. Molecular Ecology 12: 563-584.

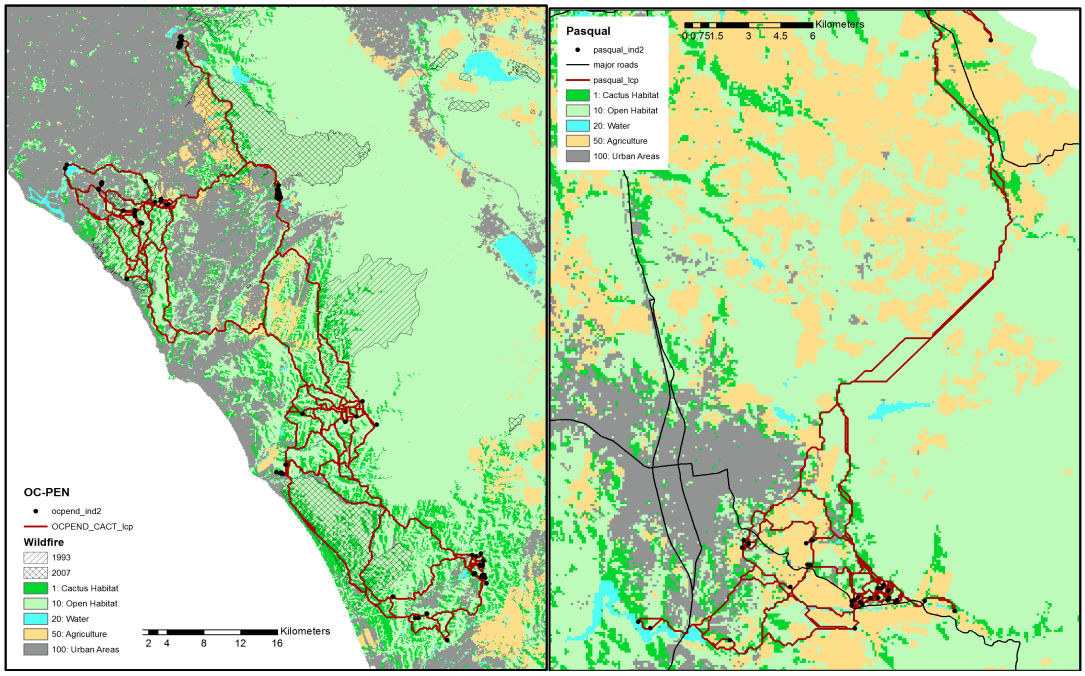
Appendix I. Preliminary cactus habitat model developed by Kristine Preston (NROC; USGS). Cactus distribution was modeled using a partitioned Mahalanobis D2 model with one kilometer grid cells from 227 random cactus survey points in both San Diego and Orange Counties, and validated with 97 additional points. Variables included in the final model were minimum temperature in January, maximum temperature in July, annual precipitation, aspect, slope, togographical heterogeneity (Sappington's index), and percent coastal sage scrub and chaparral. Developed and agricultural areas were not excluded.

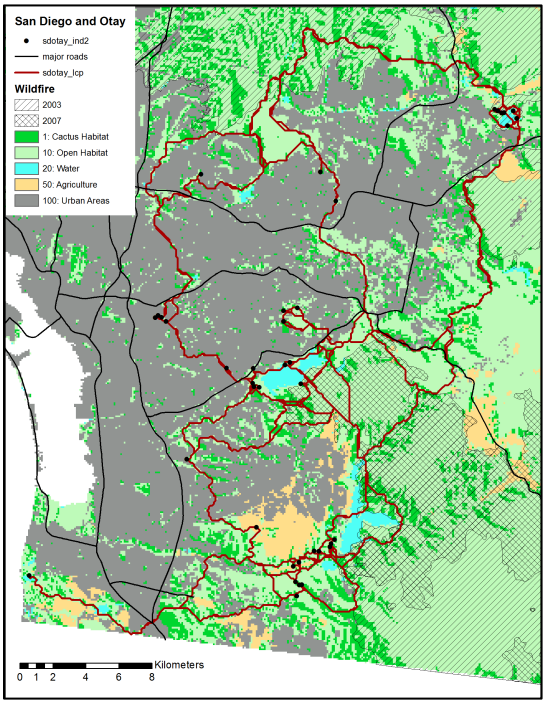


Appendix II. Least cost paths among sampled areas. The cost surface was created from the cactus habitat model as the inverse of habitat suitability scores and with both urban (derived from USGS impervious surfaces layer; USGS National Map) and agricultural (FRAP veg data) areas penalized 10X, effectively masking out human development.

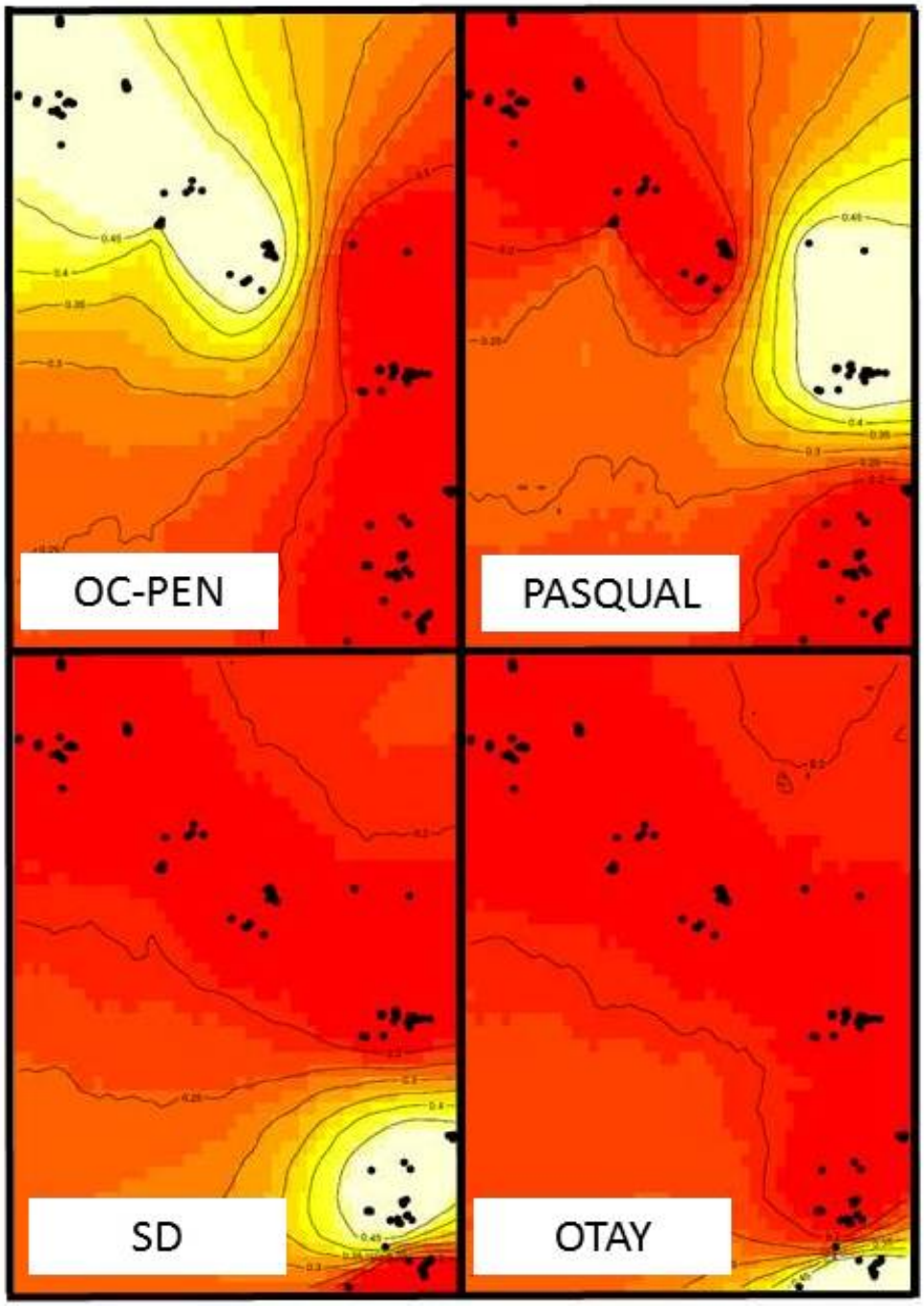


Appendix III. Least cost paths among individuals within OC-PEN, PASQUAL and the combined SD-OTAY clusters. Cost surfaces were created from the model as the inverse of habitat suitability scores and with both urban (derived from USGS impervious surfaces layer; USGS National Map) and agricultural (FRAP veg data) areas penalized, effectively masking out human development.





Appendix IV. GENELAND results. The clustering algorithm in GENELAND incorporates spatial information and genetic data to estimate the number of clusters and assign individuals to them. Results are presented as heat maps, seen here, with black dots representing samples arranged by spatial position. Probability of cluster membership ranges from high in white and yellow to low in red. These results conform to those from STRUCTURE analyses, with four clusters identified: 1, OC-PEN, composed of individuals from Orange County and Marine Corps Base Camp Pendleton/Fallbrook Naval Weapons Station; 2, PASQUAL, composed of individuals from San Pasqual Valley, Lake Hodges, and some near Bonsall and Pauma Valley; 3, SD, composed of individuals from Lake Jennings, Encanto Canyon, Sweetwater Reservoir, and nearby; 4, OTAY, composed of individuals around Otay Ranch and the Tijuana River Estuary.



Appendix V. Genetic diversity indices by major aggregation. These are sites with >5individuals sampled. Sites with fewer than 5 individuals sampled are excluded, as diversity estimates are less accurate and informative when based on few samples.

Area	Cluster Membership	Samples	H_O	H_E	A
El Modena	OC-PEN	12	0.592	0.567	2.88
Whiting Ranch	OC-PEN	16	0.653	0.618	3.01
UC-Irvine	OC-PEN	5	0.530	0.520	2.69
San Joaquin Foothills	OC-PEN	14	0.623	0.595	2.95
Northern Camp Pendleton	OC-PEN	11	0.575	0.557	2.99
Fallbrook Naval Weapons Stations	OC-PEN	15	0.553	0.567	2.94
Southern Camp Pendleton	OC-PEN	5	0.640	0.519	2.77
Bear Valley	PASQUAL	6	0.725	0.585	2.96
Rockwood/Via Rancho	PASQUAL	6	0.642	0.571	2.9
SD Zoo Safari Park	PASQUAL	18	0.648	0.643	3.08
Lake Jennings	SD	9	0.591	0.574	2.89
Sweetwater Reservoir	SD	9	0.661	0.599	2.98
Dictionary Hill	SD	5	0.548	0.509	2.79
Encanto Canyon	SD	5	0.550	0.491	2.55
Otay Ranch	OTAY	13	0.669	0.639	3.05

Appendix VI. Map of general geographic references and sites mentioned individually in this report. MCB is Marine Corps Base.