Subspecies differentiation and range-wide genetic structure are driven by climate in the California gnatcatcher, a flagship species for coastal sage scrub conservation

Amy Vandergast, Barbara Kus, Dustin Wood,
Liz Milano, Kristine Preston

USGS Western Ecological Research Center
avandergast@usgs.gov

Evolutionary Applications https://doi.org/10.1111/eva.13429
USGS Data Release https://doi.org/10.5066/P9MB2YE2


Alex Houston

Acknowledgements

J. Smith

A. Mitelberg

S. Lynn

S. Howell

L. Allen

A. Houston

M. Treadwell

M. Madden

R. Pottinger

R. Rodriguez Estrella (CIBNOR)

A. Vegas (CIBNOR)


Alex Houston

USGS

SANDAG

NATURE RESERVE of ORANGE COUNTY

Protecting our natural resources for future generations

U.S. FISH & WILDLIFE SERVICE

CALIFORNIA STATE PARKS SINCE 1864

CALIFORNIA DEPARTMENT OF FISH & WILDLIFE

Subspecies in Avian Conservation

- Generally defined as a breeding population or collection of populations that occupies a distinct segment of the geographic range of the species and that is measurably distinct in phenotype, genotype or a combination of these traits (Mayr 1969, Avise 2004, Haig and D'Elia 2010, Patten 2010, Remsen 2010).
- 44% of federally listed avian taxa are listed at the subspecies level.
- Considerable ongoing debate on criteria for defining subspecies – how to define “distinctness” or “diagnosability”.

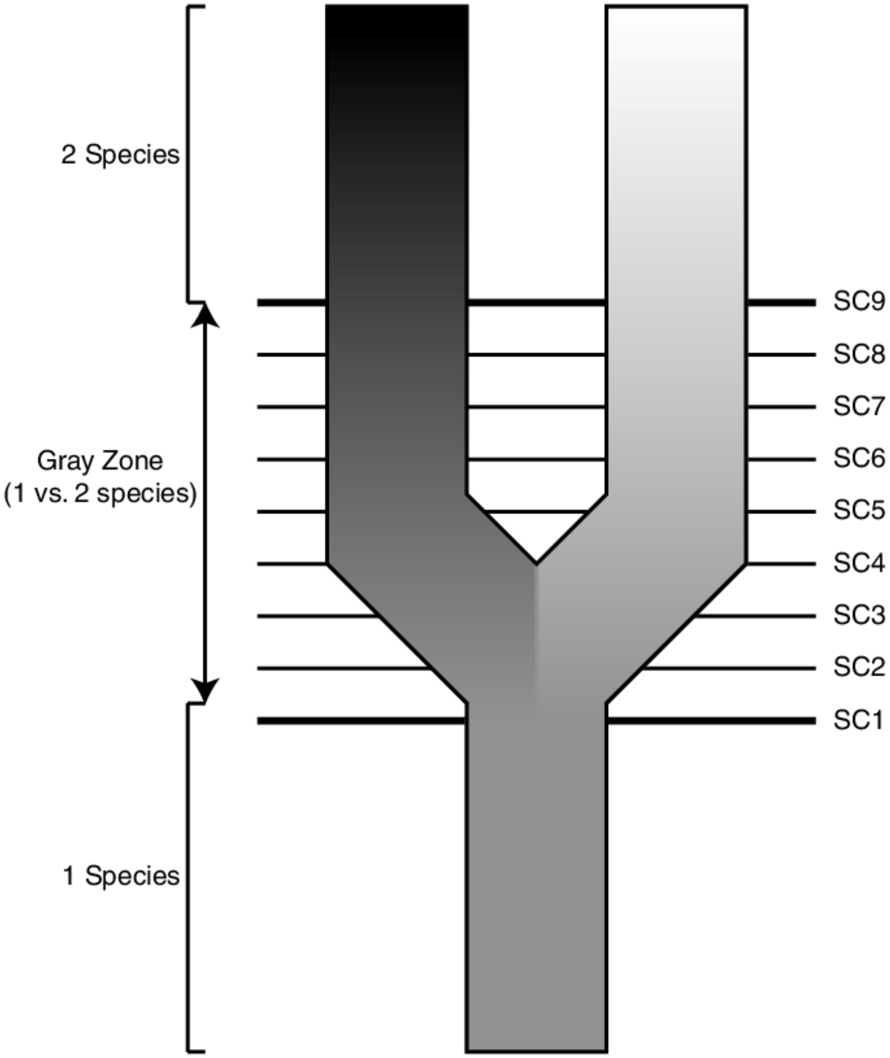
Avian Subspecies. Ornithological Monographs Volume (2010), No. 67

USGS

Genetic & Genomic Data & Subspecies

- Genetic data have distinguished among morphologically similar taxa and important in identifying previously cryptic groups (subspecies and species) (Fleischer et al. 2006, Bickford et al. 2007, Funk et al. 2007a, Klicka et al. 2016).
- In many cases genetic data are not congruent with morphologically defined subspecies.
- In cases where few genes code for phenotypic traits under selection, and/or the time scale of differentiation is recent, discordance between morphological traits and neutral genetic markers is more likely (Winker 2010, McCormack and Maley 2015).

Genetic & Genomic Data & Subspecies



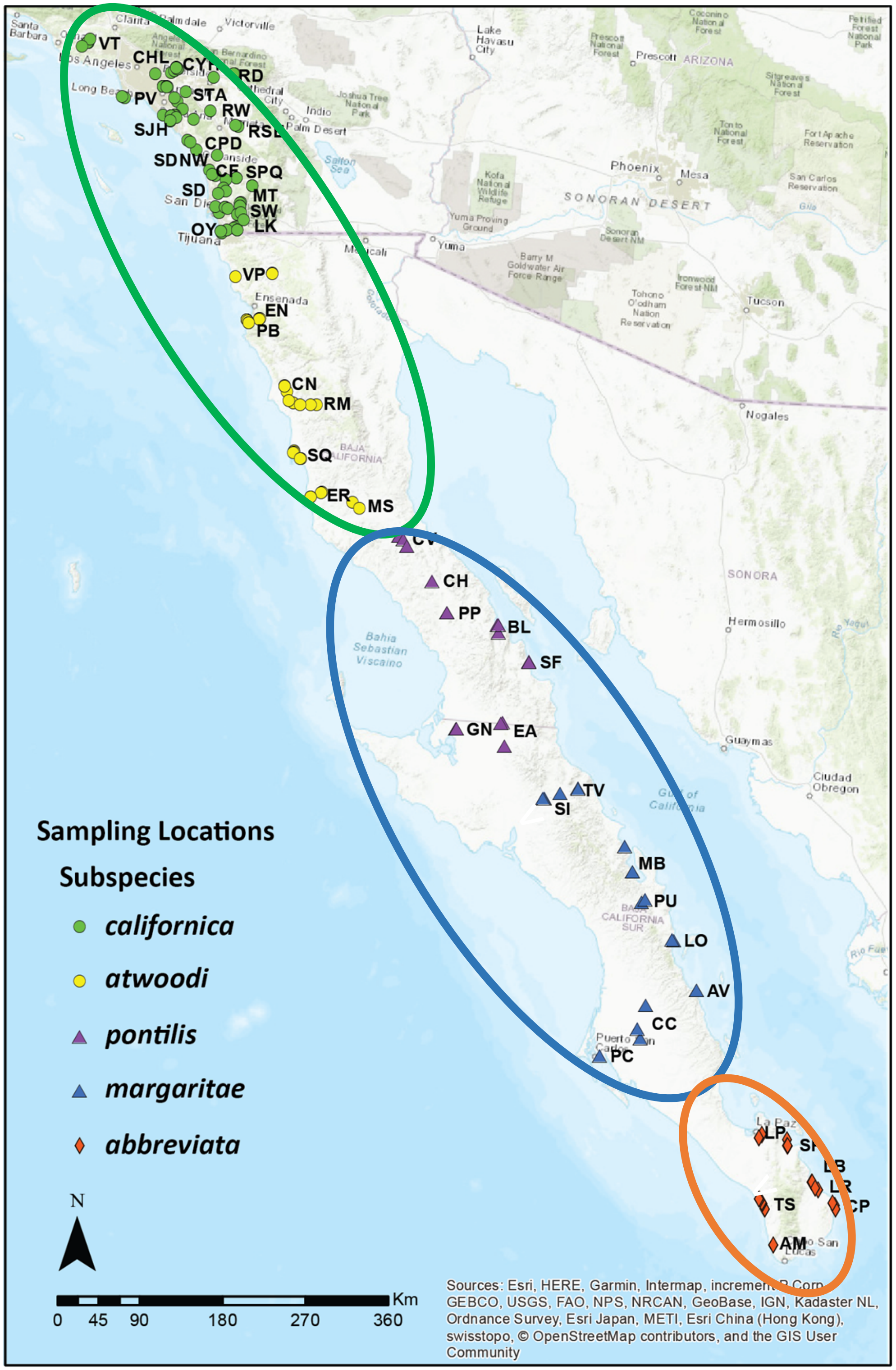
K. de Queiroz Systematic Biology 2007

CA Gnatcatcher

- Distributed from the southern tip of Baja CA Peninsula north to Ventura County CA.
- 1993 *P. c. californica* (Atwood 1991) listed as threatened

3-5 morphological subspecies:

- Atwood (1991)
  - ***californica*** (to 30° N)
  - ***margaritae*** (to 24° N)
  - ***abbreviata***
- Mellink & Rea (1994)
  - *californica* (to 32° N)
  - *atwoodi* (to 30° N)
  - *pontilis* (to 28° N)
  - *margaritae* (Atwood)
  - *abbreviata* (Atwood)



Sources: Esri, HERE, Garmin, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, © OpenStreetMap contributors, and the GIS User Community

Previous Genetic Data

- Two range-wide genetic studies using mtDNA and a small set of nuDNA did not find monophyletic clades or structure and concluded that there was no evidence of evolutionary significant divisions, and therefore no valid subspecies (Zink et al. 2000, Zink et al. 2013).
- These conclusions were criticized in the literature (McCormack and Maley 2015; Patten 2015) and by a panel of scientists convened to review all existing data for USFWS (AMEC 2015).
  - Existing genetic data not the appropriate type or quality to overturn long-standing conclusions based on phenotypic data.

Scientific Panel Recommendations

1. Develop Genome-wide data (SNPs)

2. Test for adaptive loci

3. Use a hypothesis testing framework for distinctness

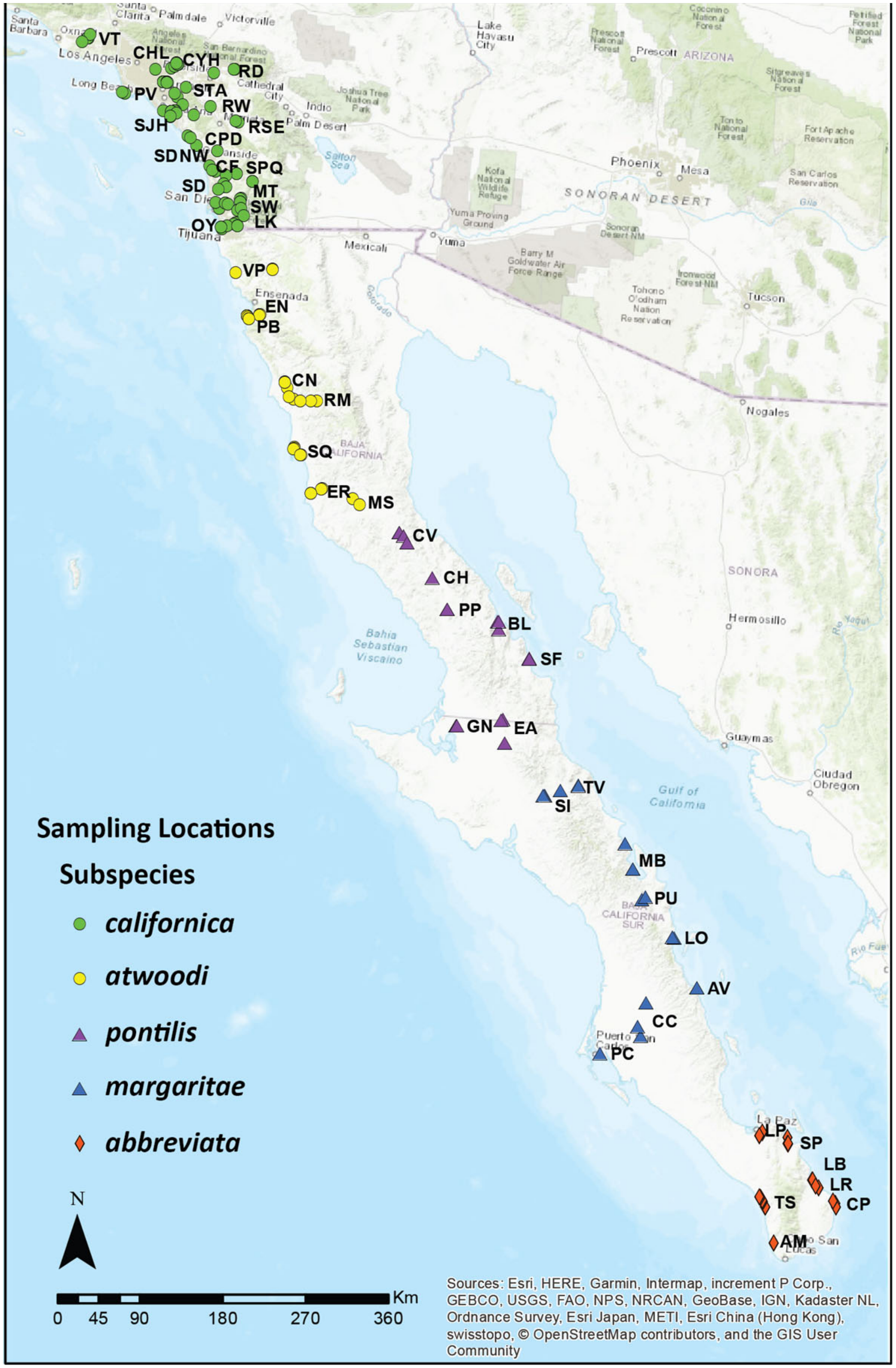
Study Goals

1. Develop Genome-wide data (SNPs)
    - Double digest restriction associated high throughput sequencing
    - Clustering Analyses
2. Test for adaptive loci
    - Test for gene-environment associations (using climate variables hypothesized to be important in range limitations).
3. Use a hypothesis testing framework for distinctness

<u>Other Goals:</u>

4. Determine landscape factors associated with genetic differentiation
5. Explore demographic history
6. Assess contemporary patterns of genetic diversity

USGS



Sources: Esri, HERE, Garmin, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, © OpenStreetMap contributors, and the GIS User Community

Field Sampling

- CA sampling 2012-2013
  - Msat study
- Baja sampling 2018-2019
  - Sites selected overlapped with Zink's
  - additional sites selected to make sure contact zones were well covered.

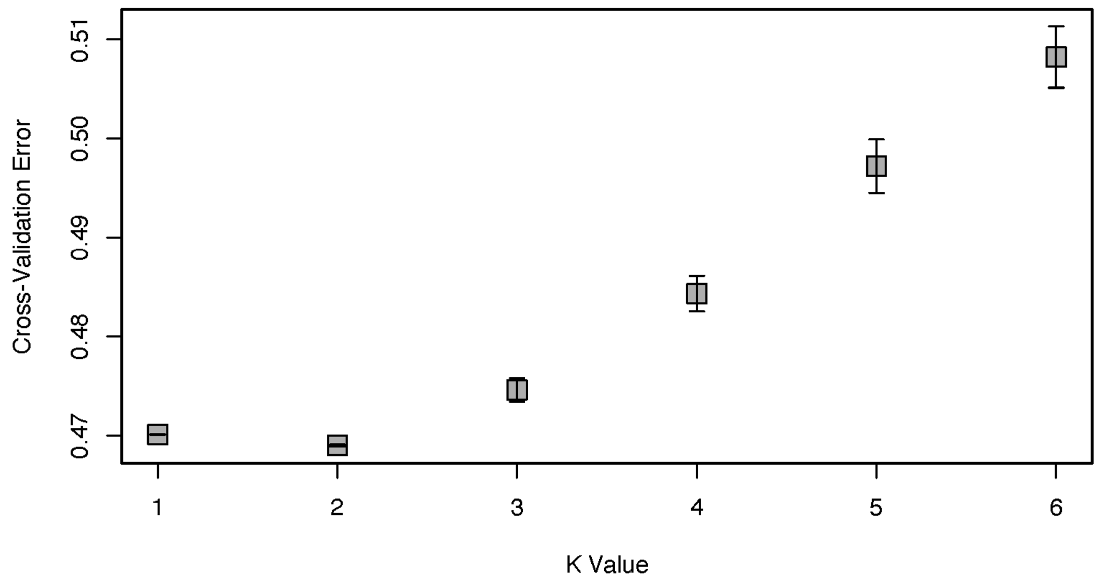
Final Dataset

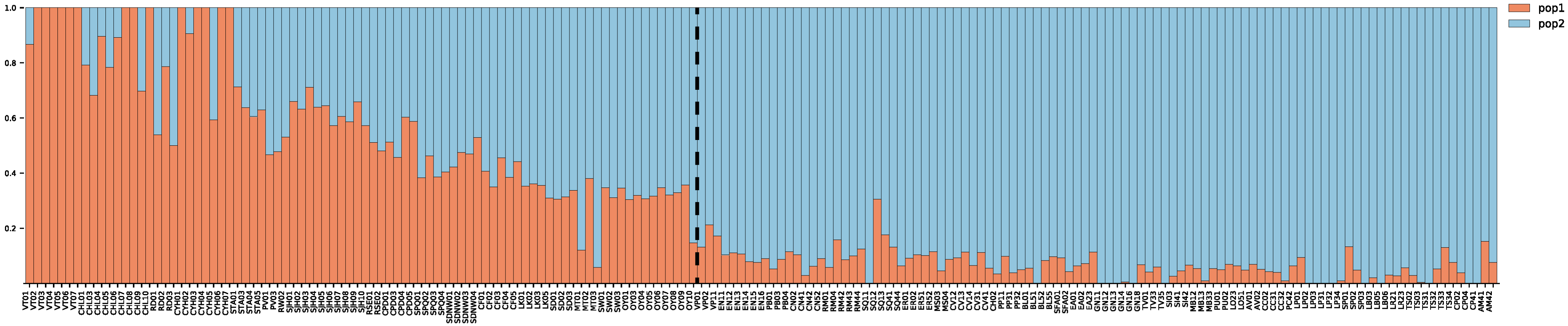
- 184 birds from 48 locations
- ddRAD
  - 84,125 loci (snps)
  - 7% missing data

Previous Zink et al. datasets

- 47-82 birds, 13 locations
- 8 loci

Clustering Results: Admixture

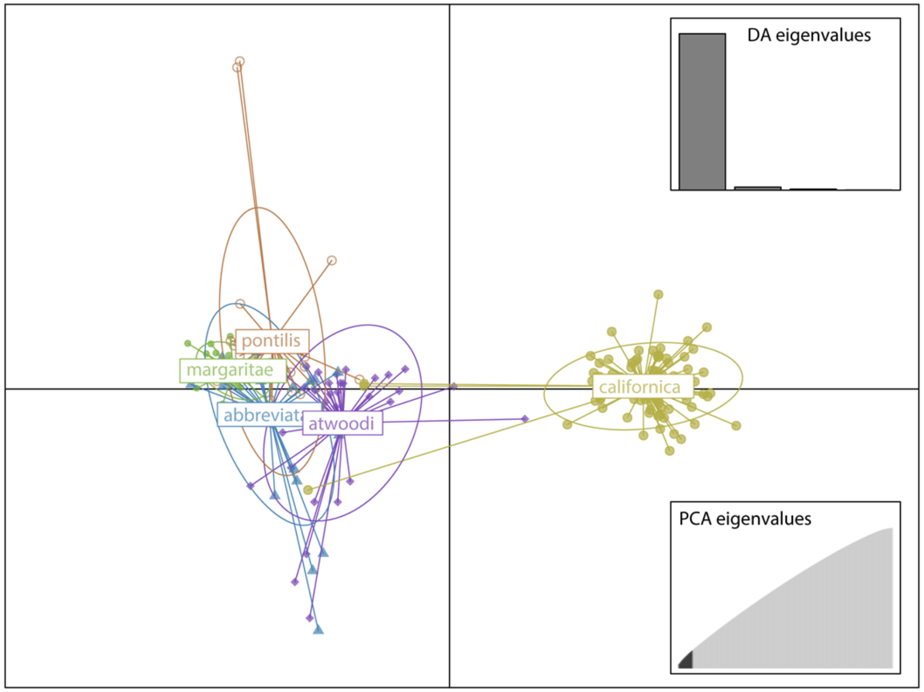




North 32 N South

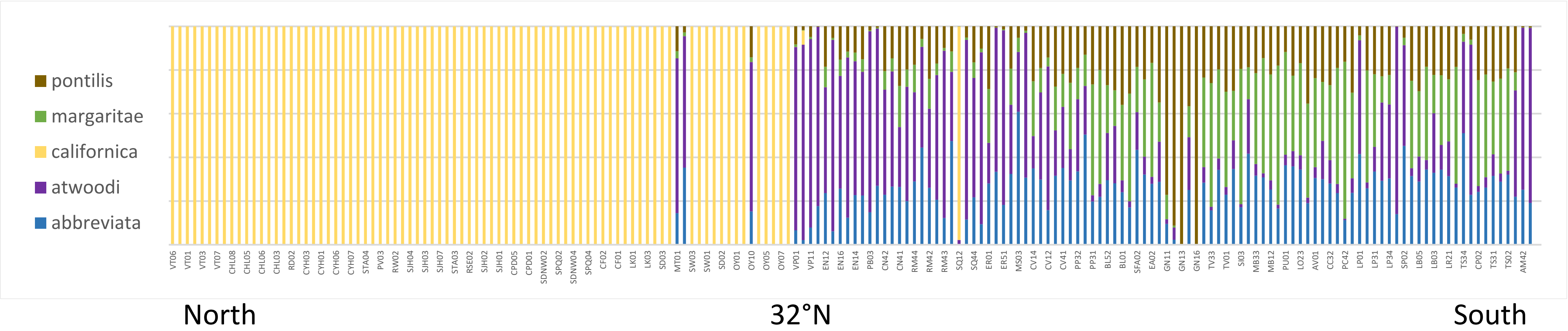
*35,440 snps using a MAF of 0.05: Clustering, DAPC & IBD analyses.

DAPC: Mellink + Atwood Subspecies



Subspecies	prop. of ind. with >90% assignment	avg. assignment prob
californica	0.96	0.96
atwoodi	0.03	0.58
pontilis	0.13	0.36
margaritae	0.00	0.44
abbreviata	0.00	0.3

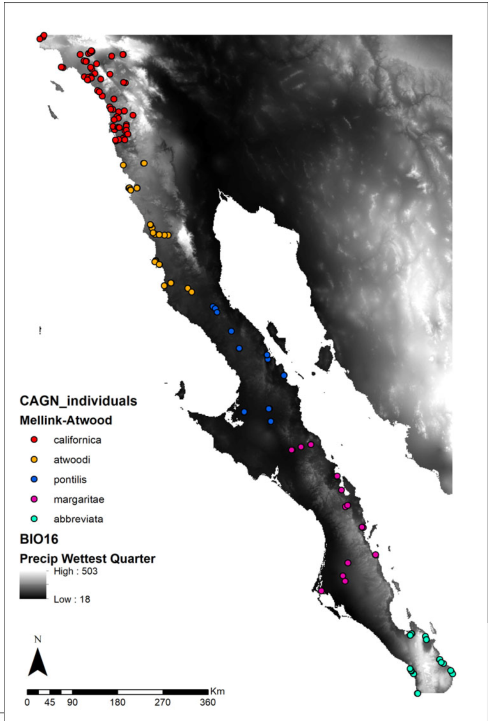
Posterior probabilities of assignment

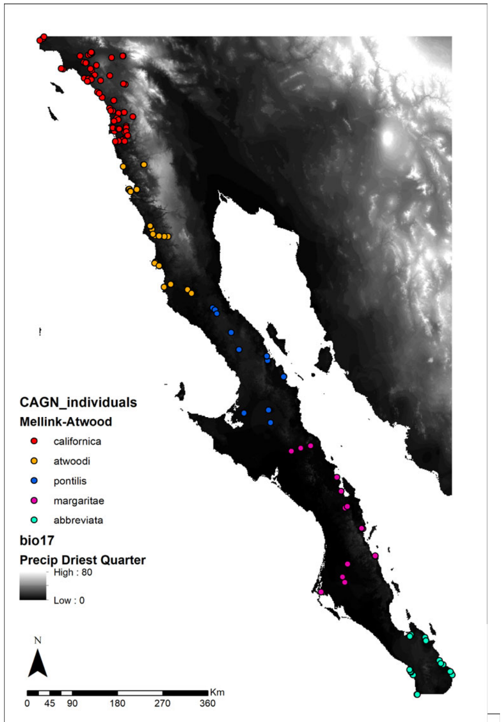


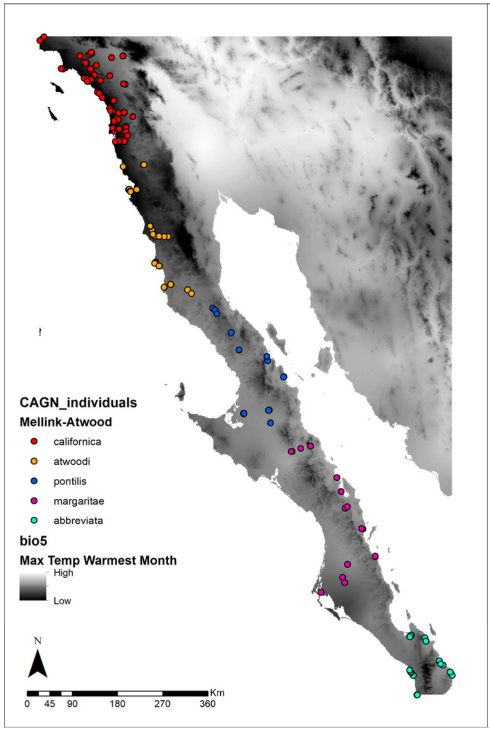
*3 individuals in CA that assign to Baja group, 1 in San Quintin assigns to CA group

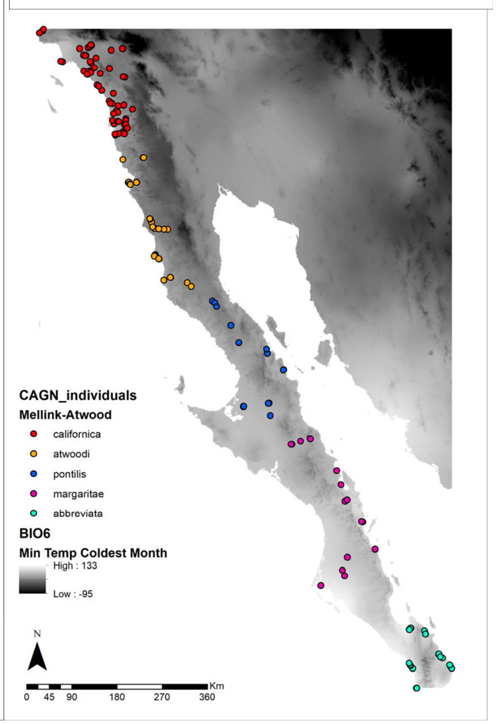
Adaptation to Climate?

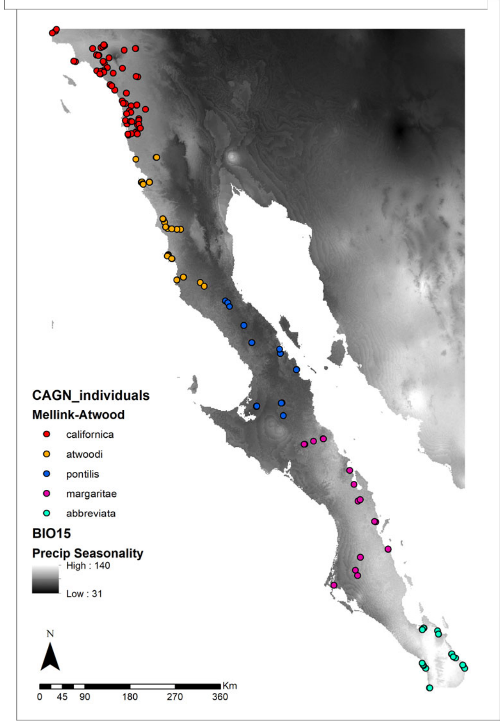
- High climate variability over range.
- Winter precip and low temperatures limit range (Mock 1998).
- High mortality documented over cold wet winters.
- Climate could be a strong driver of selection and evolution.
- Climate Variables (Precip WQ, DQ, seasonality. Tmax, Tmin).





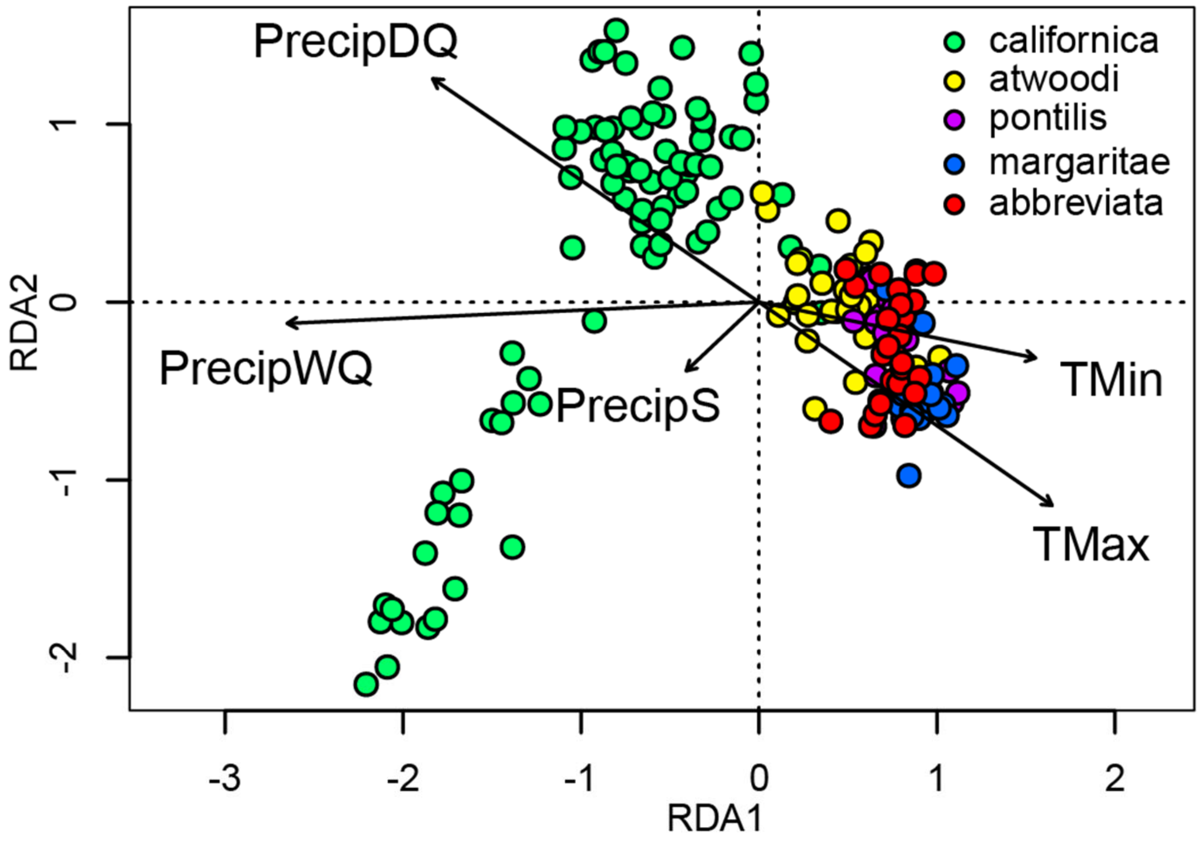


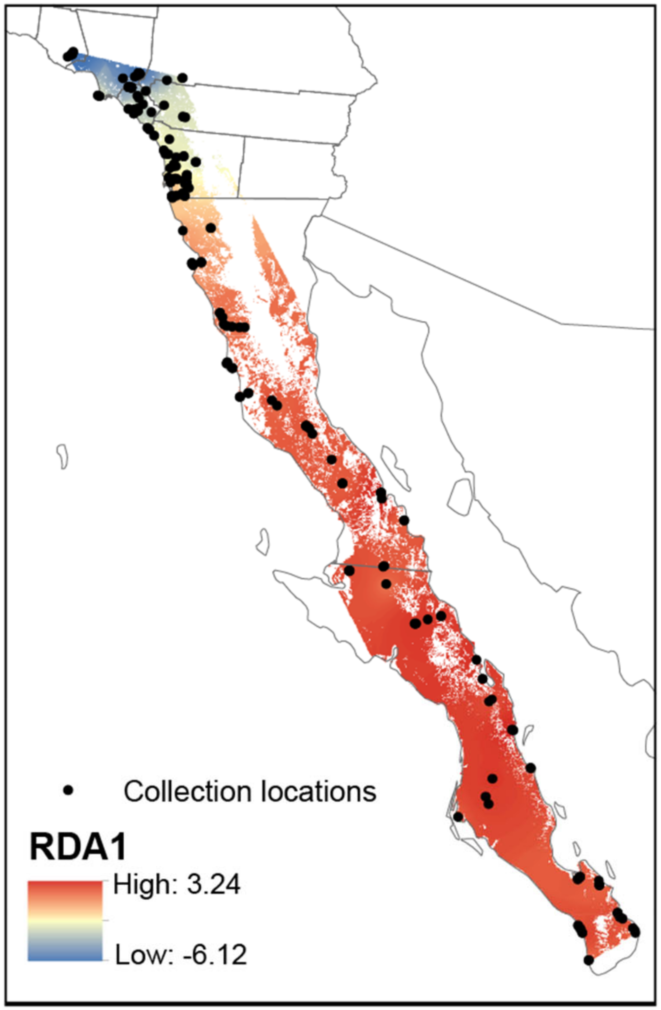




Gene-Environment: RDA

Redundancy analysis (RDA) to determine how groups of loci covary in response to climate



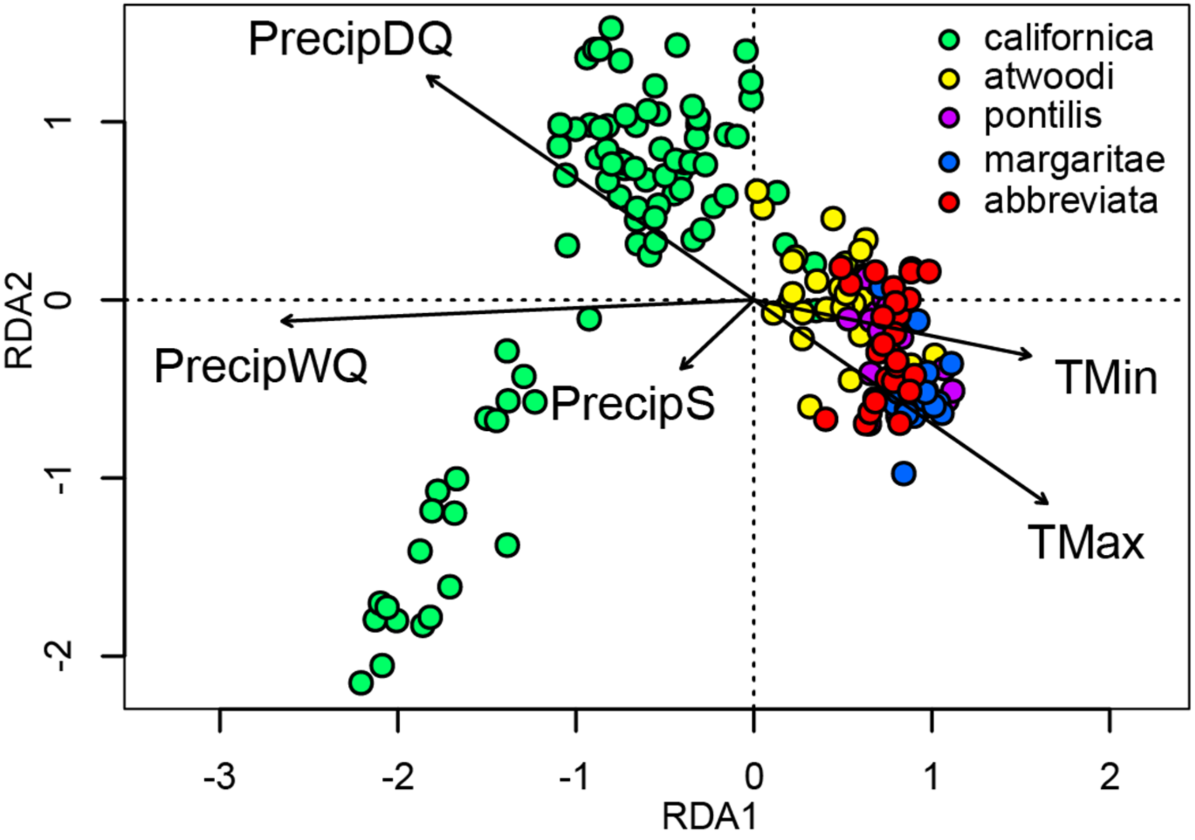


RDA 1 35% of the total variation

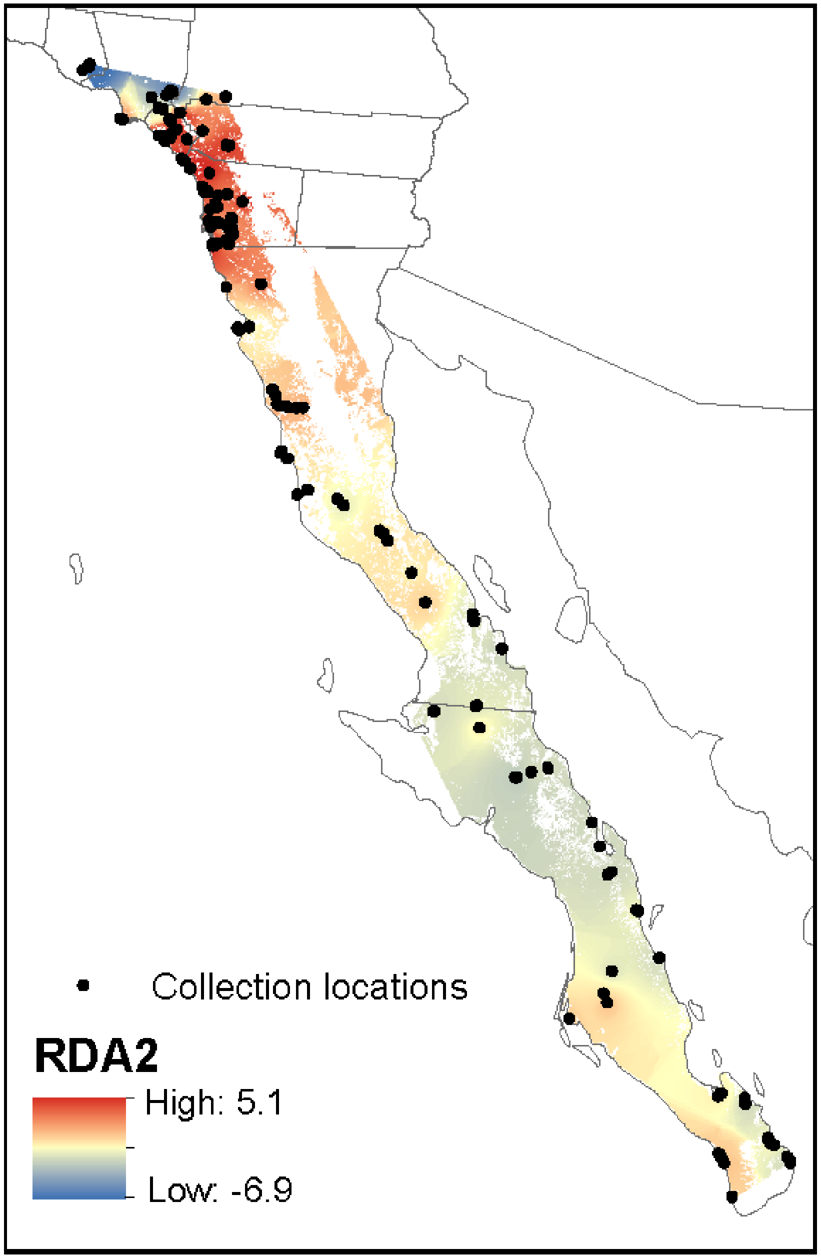
TMAX	0.58
TMIN	0.55
PRECIPS	-0.14
PRECIPWQ	-0.94
PRECIPDQ	-0.65

Gene-Environment: RDA

Redundancy analysis (RDA) to determine how groups of loci covary in response to climate

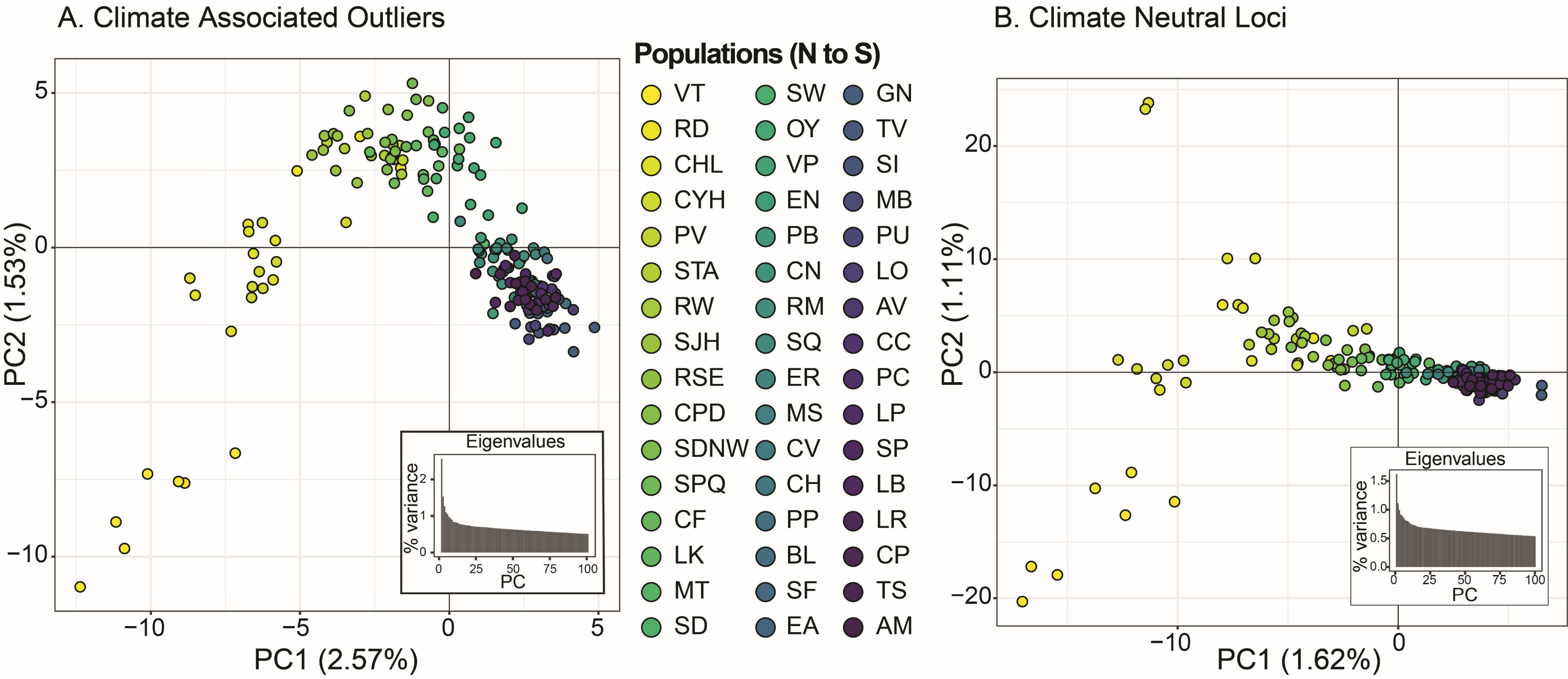


RDA 2 40% of the total variation	
TMAX	-0.46
TMIN	-0.13
PRECIPS	-0.16
PRECIPWQ	-0.05
PRECIPDQ	0.51



Clustering: Climate Loci

- 6,339 putatively adaptive SNPs



2 clusters supported with climate outliers; 1 cluster supported with climate neutral snps

Hypothesis tests for subspecies distinctiveness

4 approaches

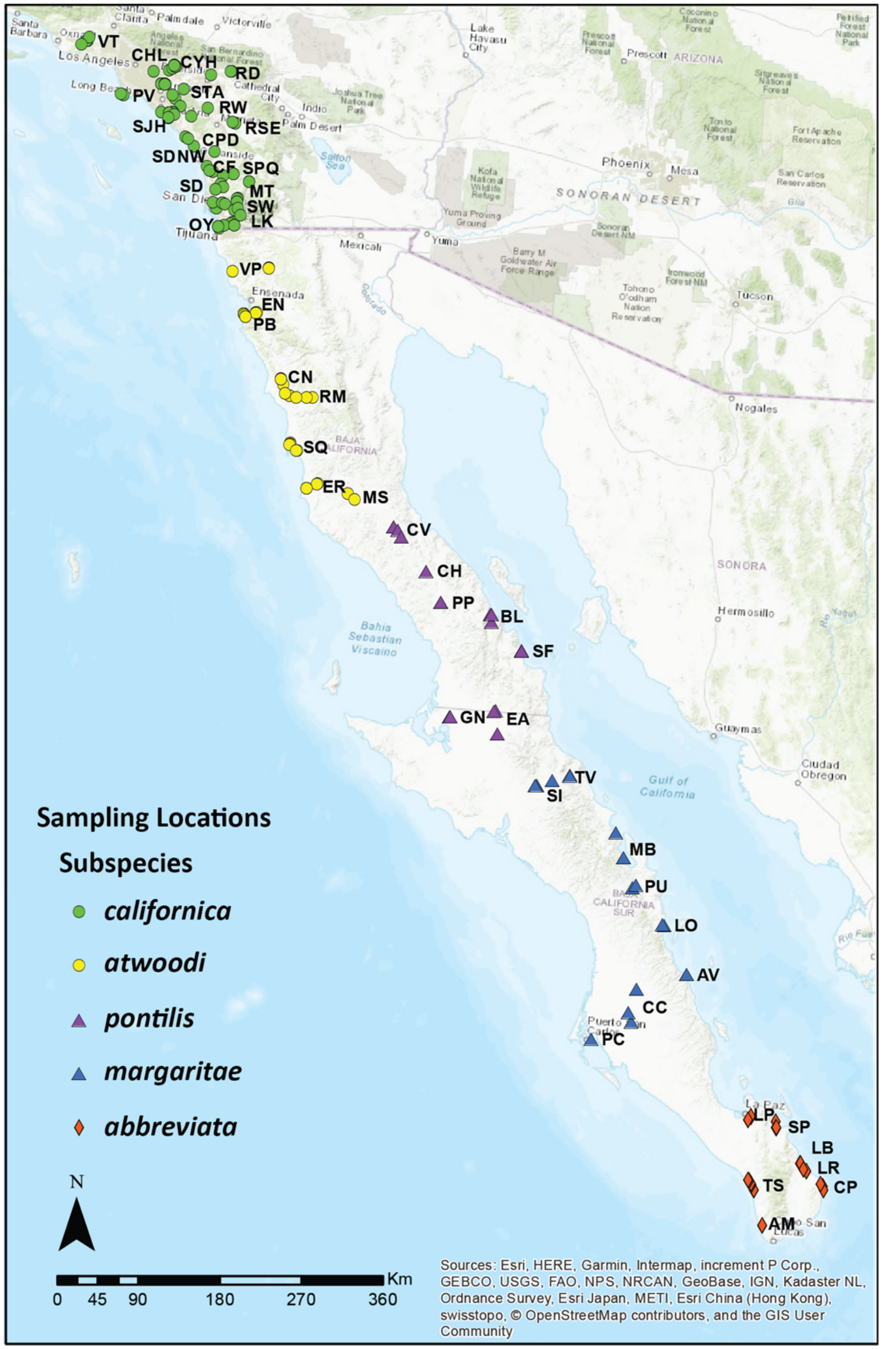
1) DAPC posterior assignments– what proportion of the range is assigned to the correct subspecies group with high confidence (75% rule)?
2) Individual genetic distances – are there significant group differences across subspecies boundaries (positive Mantel R)?
3) Analysis of Molecular Variance (AMOVA) – is a significant amount of genetic variation partitioned among subspecies groups, and if so which ones? (post-hoc pairwise tests).
4) Analysis of Molecular Variance (AMOVA) using climate associated outlier loci

Tested using the latitudinal breaks for the 2 most recent and data-robust subspecies hypotheses: 5 subspecies (Mellink and Rea 1993 plus Atwood 1991), 3 subspecies (Atwood 1991).

DAPC posterior assignments

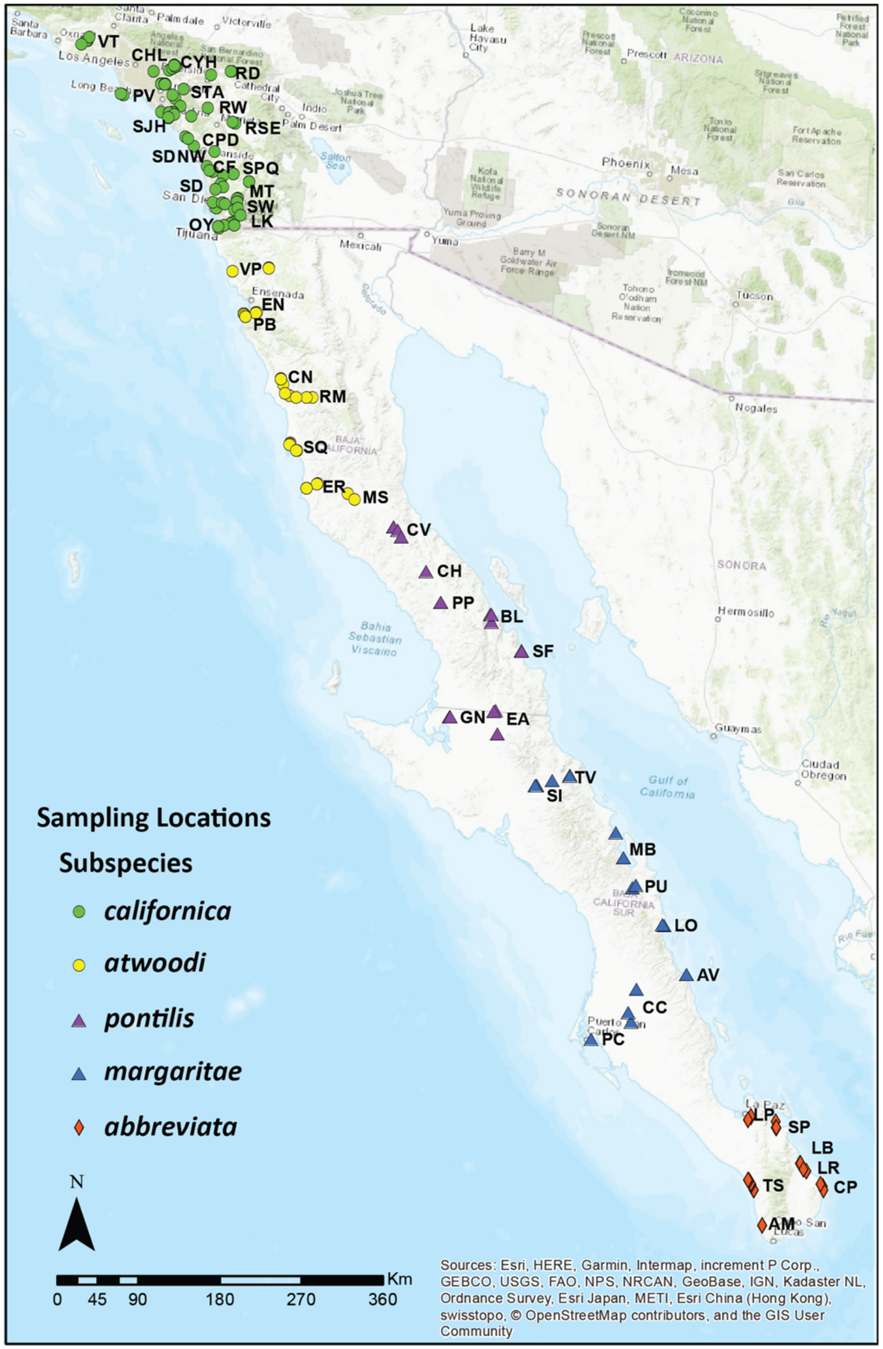
75% Rule (Patten and Unitt 2002)

Subspecies	prop. of ind. with >90% assignment	avg. assignment prob
Hypothesis 1		
californica (32 N)	0.96	0.96
atwoodi	0.03	0.58
pontilis	0.13	0.36
margaritae	0.00	0.44
abbreviata	0.00	0.3
Hypothesis 2		
californica (30 N)	0.73	0.70
margaritae	0.00	0.32
abbreviata	0.00	0.30



USGS

IBD: Mantel Tests for Group Differences



Sources: Esri, HERE, Garmin, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, © OpenStreetMap contributors, and the GIS User Community

Hypothesis 1: 5 subspecies

Mantel R = -0.25, NS

Hypothesis 2: 3 subspecies

Mantel R = -0.18, NS

CA vs. Baja

Mantel R = 0.19

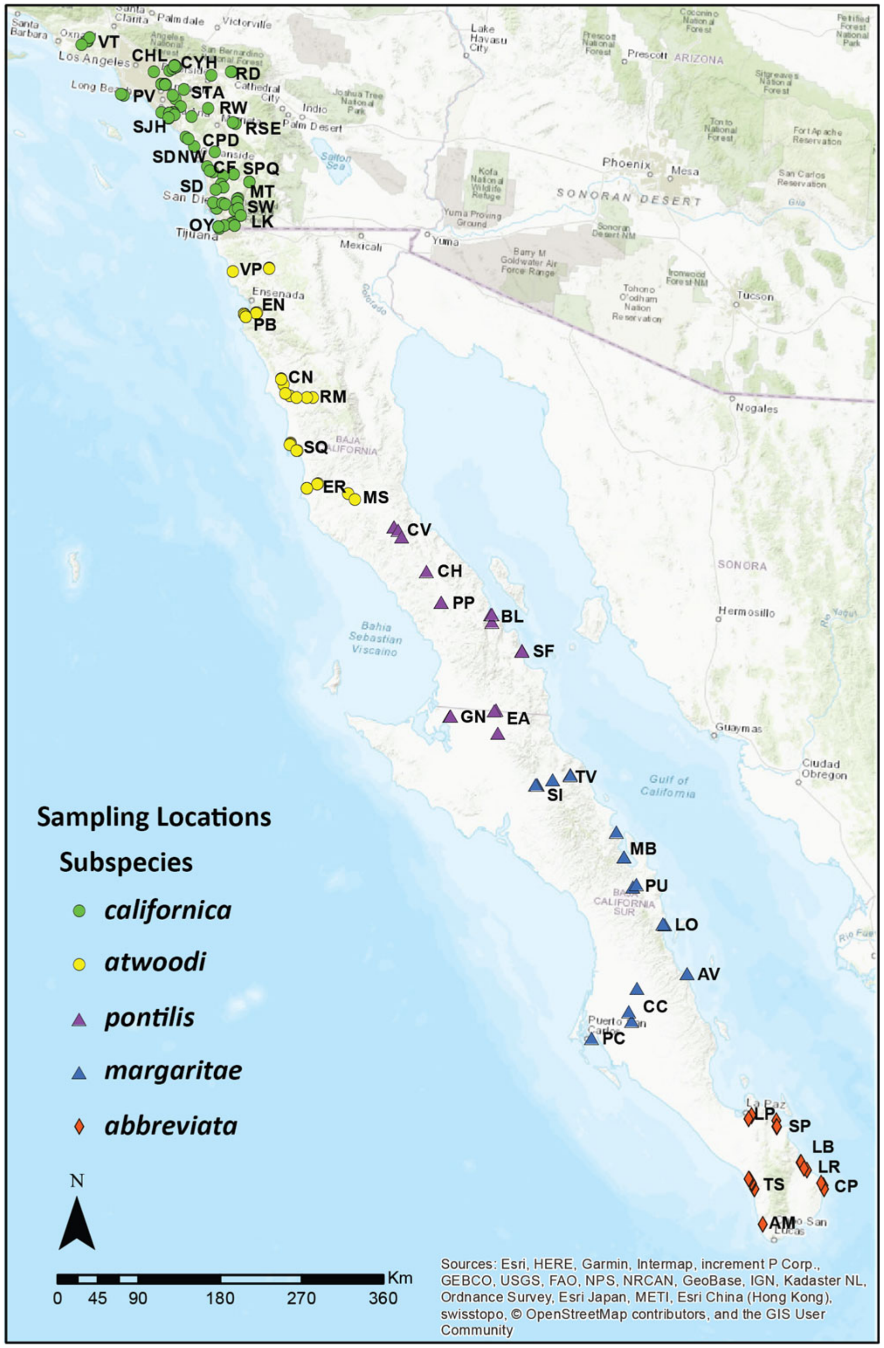
$P \leq 0.001$

AMOVA: all loci

- Small but significant proportion of variation partitioned among subspecies for both hypotheses.
- *Groups should be significantly different from their neighboring groups in pairwise tests* (McCormack and Maley 2015).

Pairwise ΦCT	*californica*	*atwoodi*	*pontilis*	*margaritae*	*abbreviata*
californica	-				
atwoodi	**0.0033**	-			
pontilis	**0.0076**	0.0005	-		
margaritae	**0.0091**	**0.0022**	0.0005	-	
abbreviata	**0.0053**	0.0001	-0.0003	-0.0005	-

Pairwise ΦCT	*californica*	*margaritae*	*abbreviata*
californica	-		
margaritae	**0.006**	-	
abbreviata	0.0025	-0.0014	-



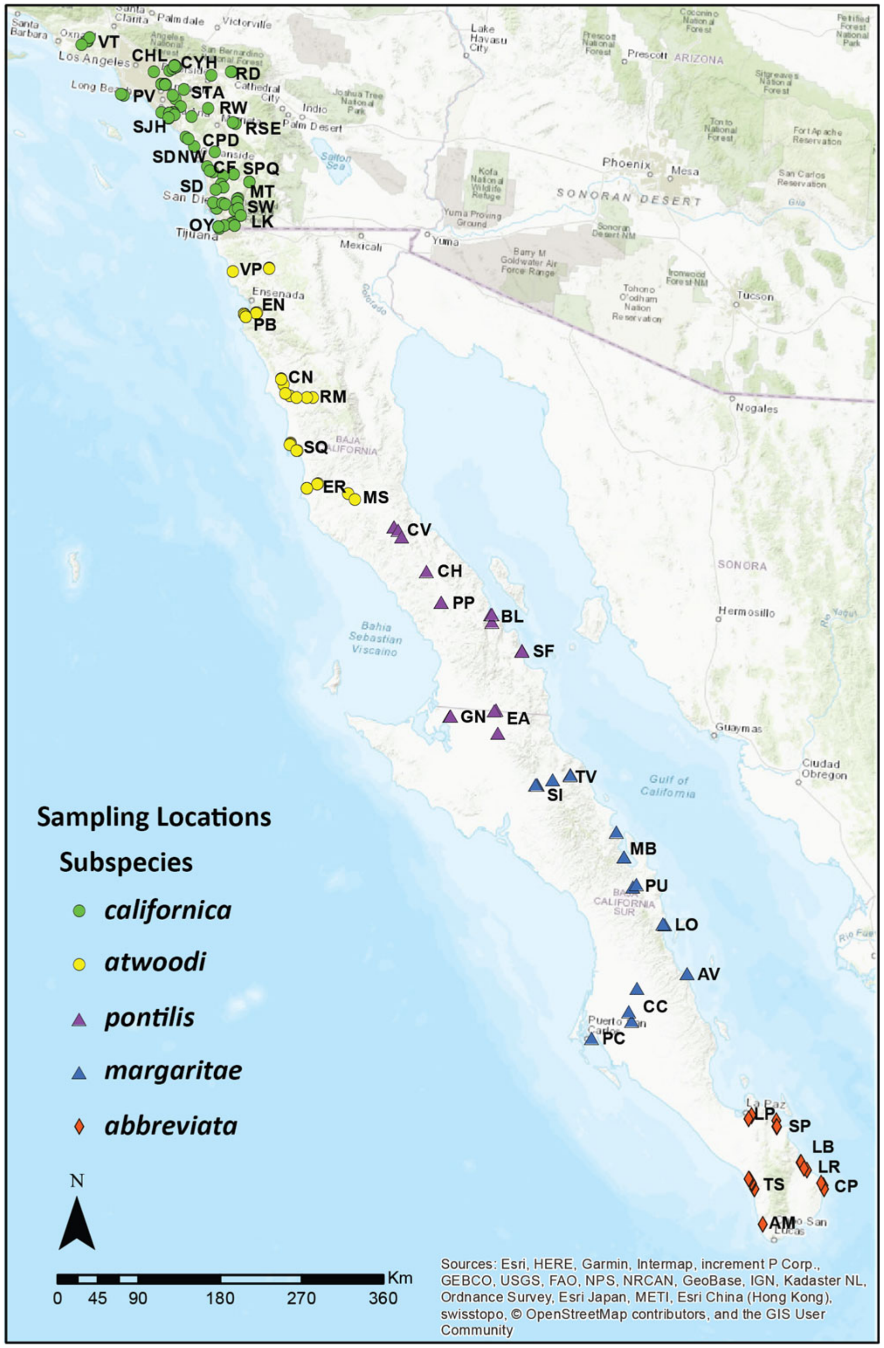
Sources: Esri, HERE, Garmin, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, © OpenStreetMap contributors, and the GIS User Community

AMOVA: climate outliers

- Small but significant proportion of variation partitioned among subspecies for both hypotheses.
- *Climate loci above diagonal in blue*
- Bolded values are significant correction
- Underlined indicate adjacent pairs

Pairwise ΦCT	*californica*	*atwoodi*	*pontilis*	*margaritae*	*abbreviata*
californica	-	**0.0078**	**0.016**	**0.0164**	**0.0117**
atwoodi	**0.0033**	-	**0.0051**	**0.0041**	0.0013
pontilis	**0.0076**	0.0005	-	0.0026	**0.0044**
margaritae	**0.0091**	**0.0022**	0.0005	-	0.0011
abbreviata	**0.0053**	0.0001	-0.0003	-0.0005	-

Pairwise ΦCT	*californica*	*margaritae*	*abbreviata*
californica	-	**0.011**	**0.0064**
margaritae	**0.006**	-	0.0011
abbreviata	0.0025	-0.0014	-



Sources: Esri, HERE, Garmin, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, © OpenStreetMap contributors, and the GIS User Community

Evidence for significant genetic groups

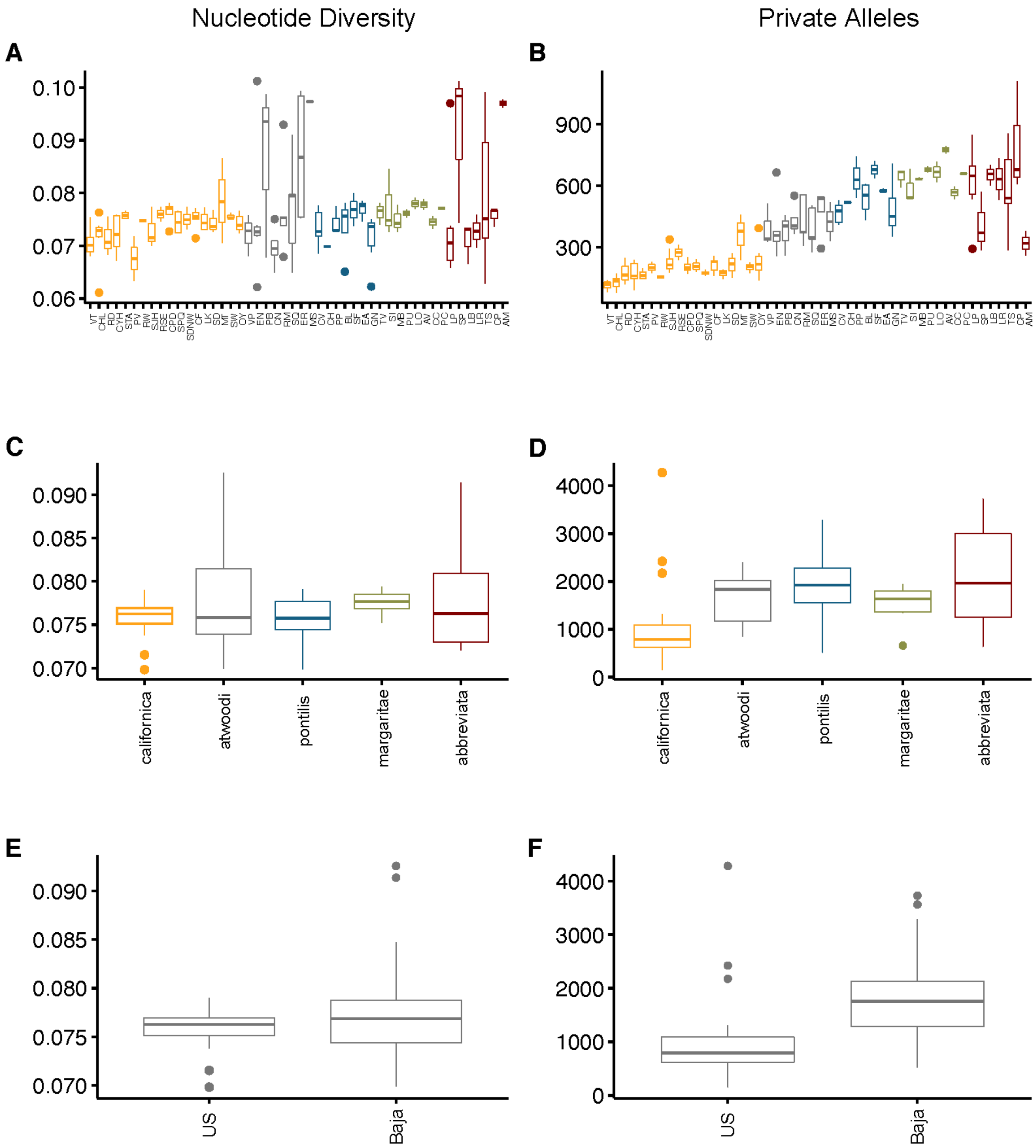
Hypothesis	Subspecies	**DAPC (All Loci)**	**Mantel Tests (All Loci)**	**AMOVA (All Loci)**	**AMOVA (Climate Loci)**
1 (Mellink and Rea plus Atwood)	*californica 32°N*	✓	✓	✓	✓
	atwoodi	✘	✘	✘	✓
	pontilis	✘	✘	✘	✘
	margaritae	✘	✘	✘	✘
	abbreviata	✘	✘	✘	✘
2 (Atwood)	*californica 30°N*	✘	✘	✓	✓
	margaritae	✘	✘	✘	✘
	abbreviata	✘	✘	✘	✘

Demographic History

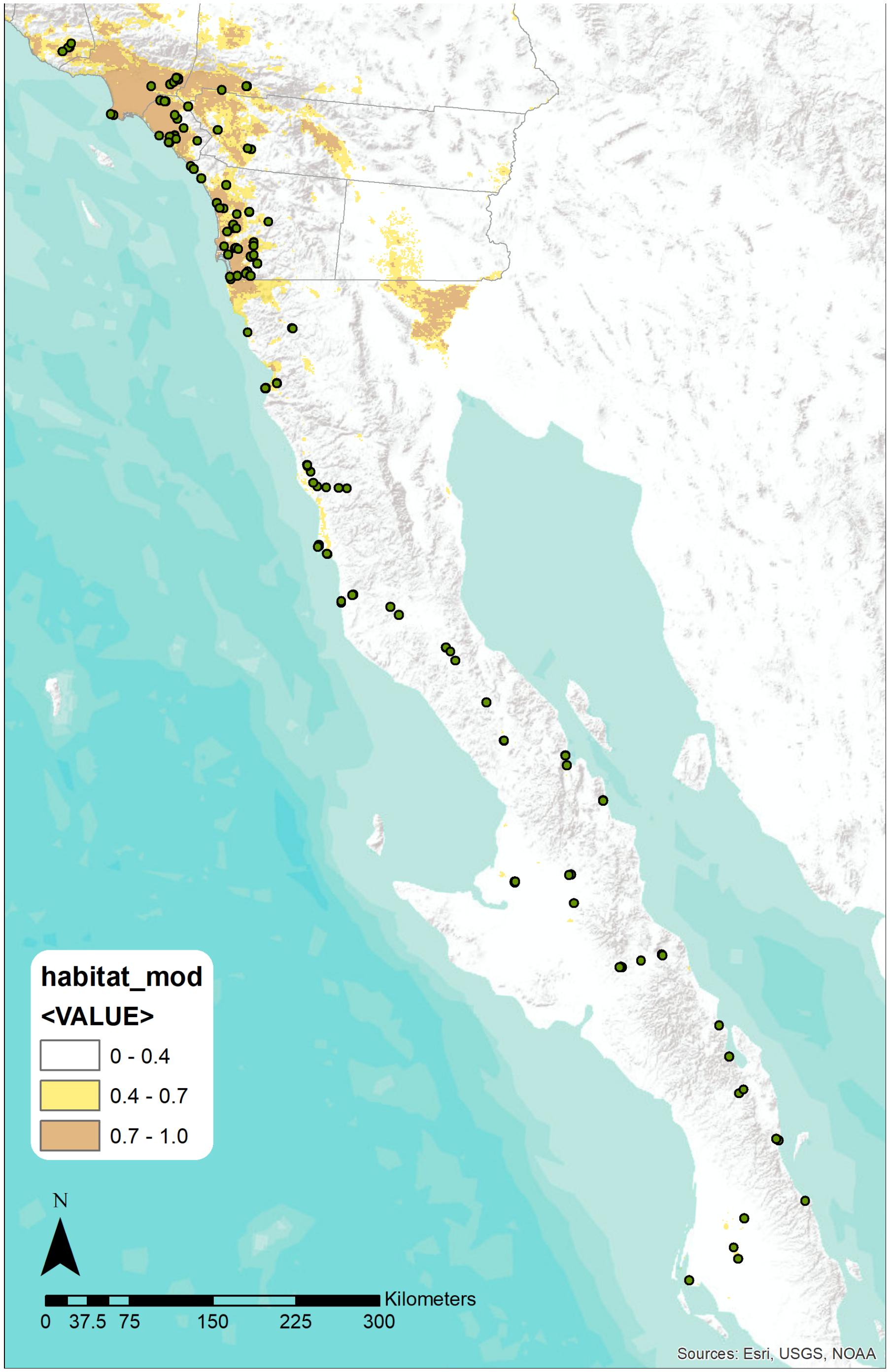
- Data best fit a multiple expansion model
  - Late Pleistocene-Early Holocene
  - Late Holocene

- Contemporary Effective Population Size (Ne)
  - California: 451 - 2063
  - Baja: 281 – 5181

Genetic Diversity



What factors influence genetic distance?



- Individual genetic distances (Nei's)
- Euclidean Distance
- Human Modified Habitat Cost Distance
  - 1 km resolution, NASA SEDAC
- 5 climate variables

- Multiple Regression of Matrices analysis framework

What factors influence genetic distance?

- Different factors are associated with genetic distance in different portions of the range.

Full Range: Genetic Dist~ CA/Baja Groups + PrecipWQ

$R^2 = 0.058$, $P \leq 0.001$

CA Only: Genetic Dist ~ Mod. Habitat Cost + Tmin + PrecipWQ

$R^2 = 0.23$, $P \leq 0.001$

Baja Only: Genetic Dist ~PrecipS

$R^2 = 0.012$, $P \leq 0.042$

Summary of Results

1. Genetic data most strongly support a difference between *P. c. californica* and birds in Baja California, with split near the US/Mexico Border.
    - Indication of recent movement and gene flow between San Diego and Northern Baja.
    - Previously, we found evidence of long-distance dispersal in CA (Vandergast et al 2019).

2. Differences are stronger in climate associated loci than in neutral loci and may be of adaptive significance.
    - Genotypes of birds in Ventura and other northern sites associated with a unique climate space

3. No detectable genetic differentiation among the other previously described subspecies in Baja California

4. Young lineage: more than one range expansion in the mid-Pleistocene and Early Holocene

Alex Houston, USGS