

Measuring population genetic diversity: An example from declining populations of the California Red-legged Frog (*Rana draytonii*)

Jonathan Q. Richmond

US Geological Survey, 4165 Spruance Rd., Suite 200, San Diego, CA 92101

*email: jrichmond@usgs.gov



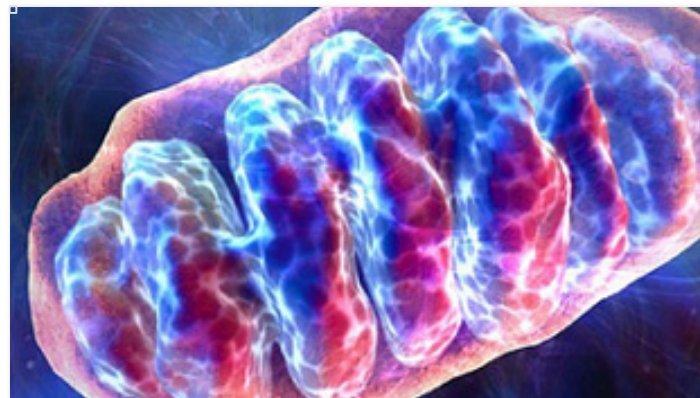
Outline:

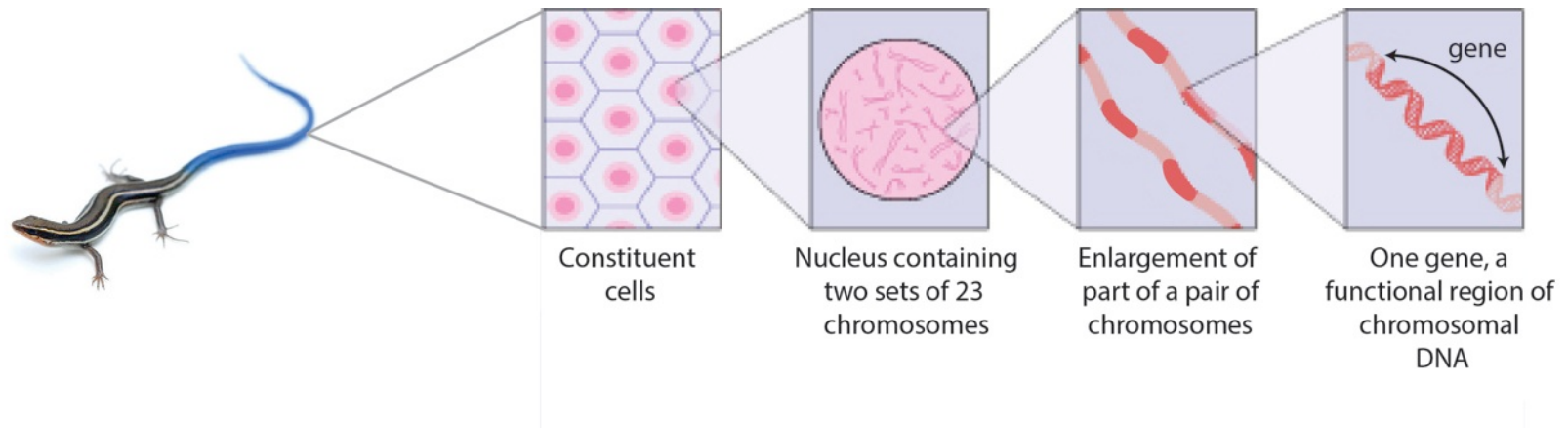
1. Sources of DNA in animals
2. The types of DNA data we collect:
 - Single nucleotide polymorphisms (SNPs)
 - Microsatellites
 - DNA sequence data
3. Visualization and interpretation of the data
 - Phylogenetic trees
 - Cluster assignment plots
4. Measuring genetic diversity:

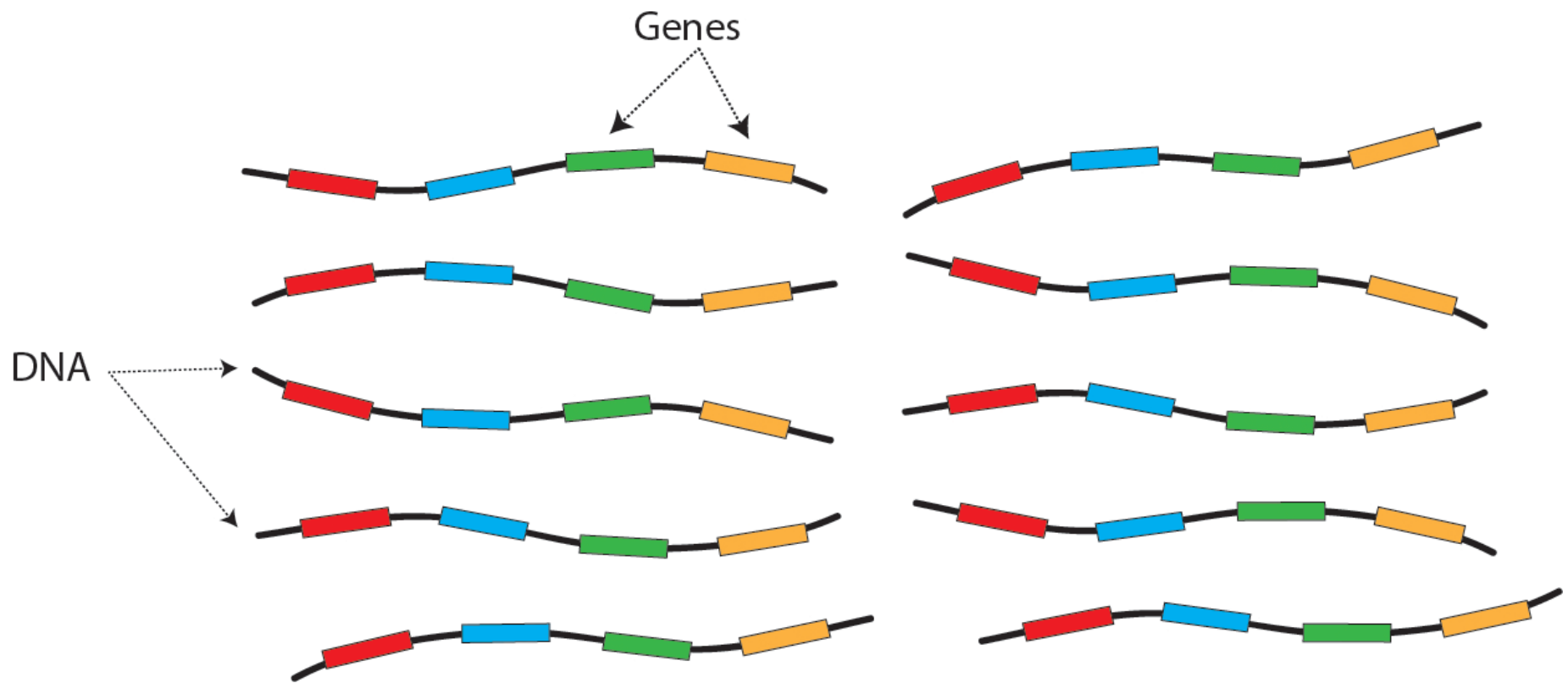
Case study: Conservation genetics of the California red-legged frog
Rana draytonii

Two main sources of DNA in animals:

- Nuclear DNA
 - Contained within the cell nucleus of eukaryotic organisms
 - Bi-parentally transmitted from parent to offspring
 - Undergoes recombination
- Mitochondrial DNA (mtDNA)
 - Located in organelles called mitochondria within cells
 - Derived from the circular genomes of the bacteria that were engulfed by the early ancestors of today's eukaryotic cells
 - Maternally inherited
 - No recombination

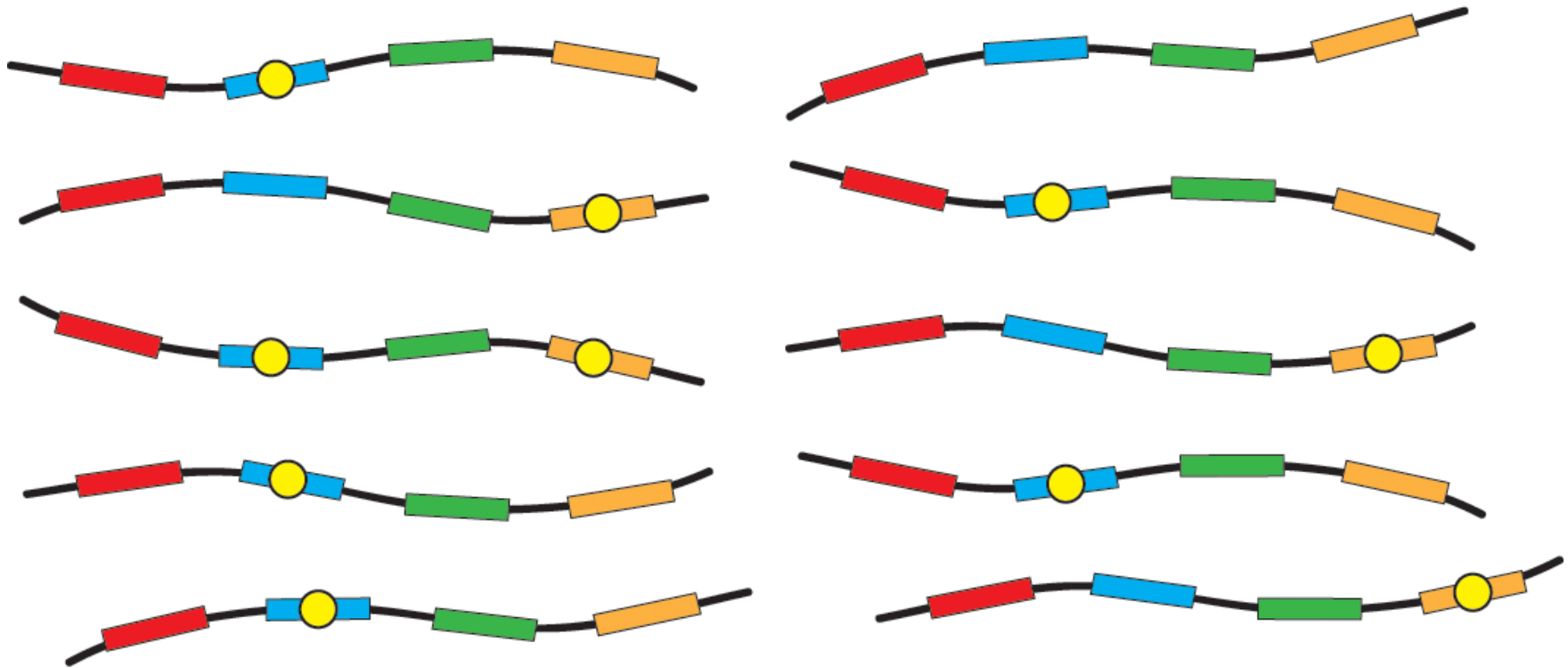




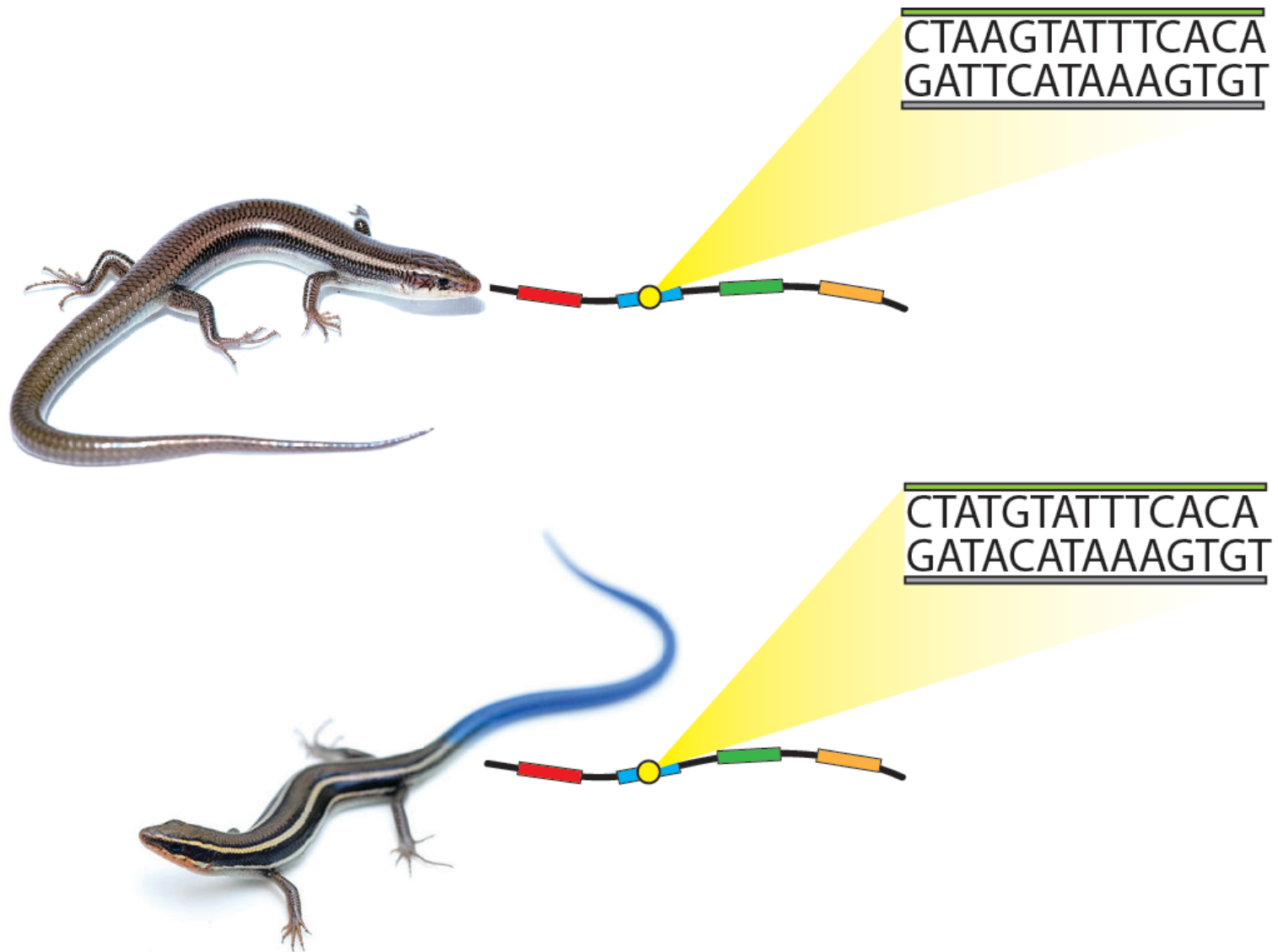


Genetic differences between organisms are evident in their DNA sequences.

Single Nucleotide Polymorphisms (SNPs)

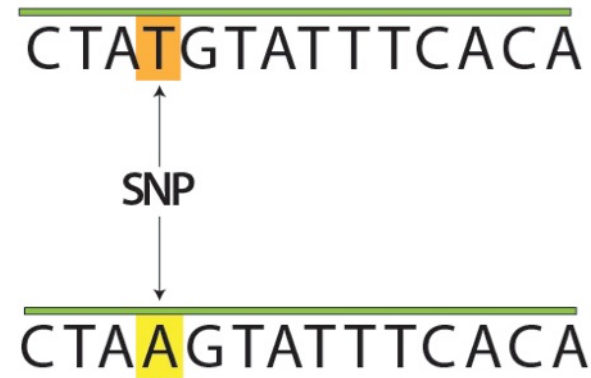


Many of the genetic differences are single nucleotide substitutions in the DNA sequence.



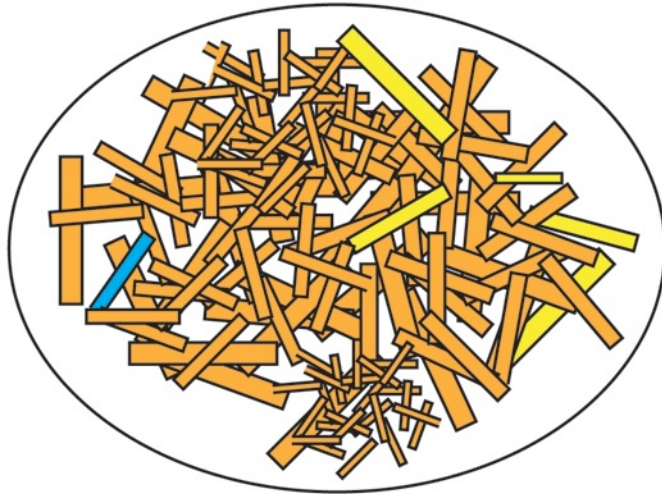
Here we have two short, double stranded DNA sequences taken from the same region of the genome in two different skinks.

Keeping it simple, we'll just show one strand....

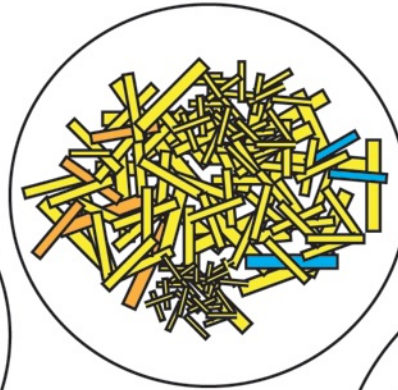


The sequences are almost identical except for one nucleotide position.

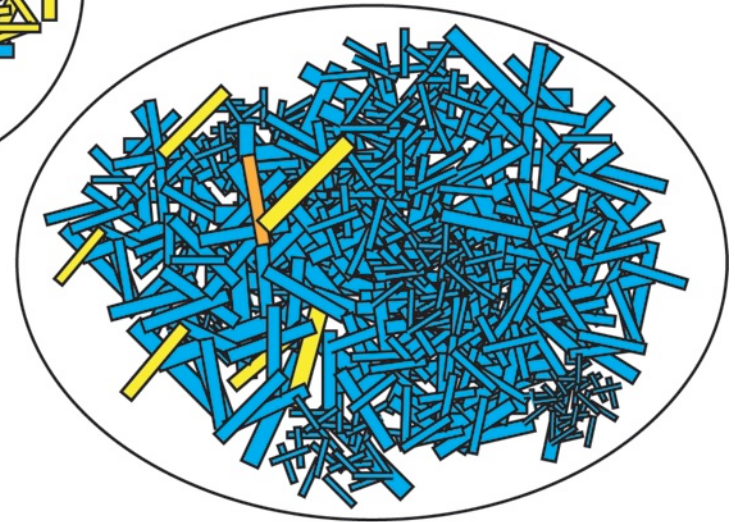
Pop A



Pop B



Pop C

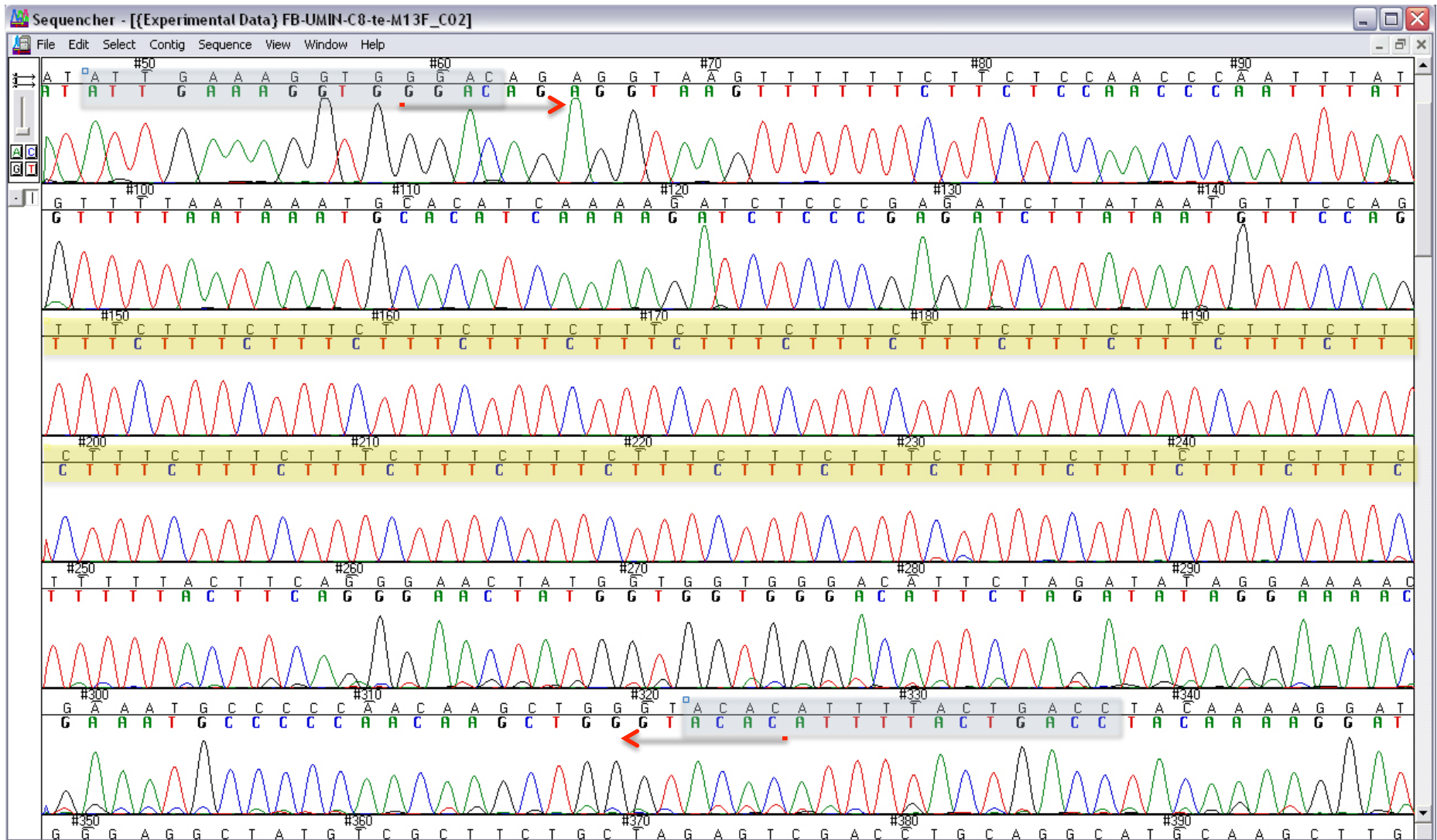


CTATGTATTTCACA = 

CTAAGTATTTCACA = 

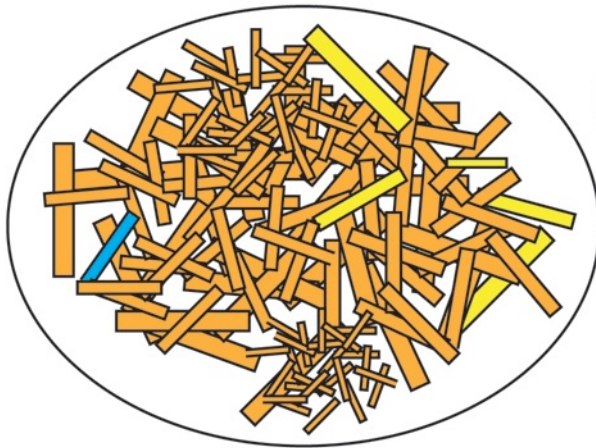
CTAGGTATTTCACA = 

What is a microsatellite?

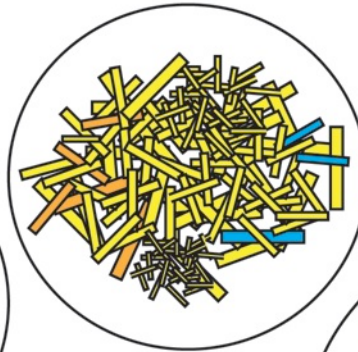


TTTC₂₅ tetramer motif

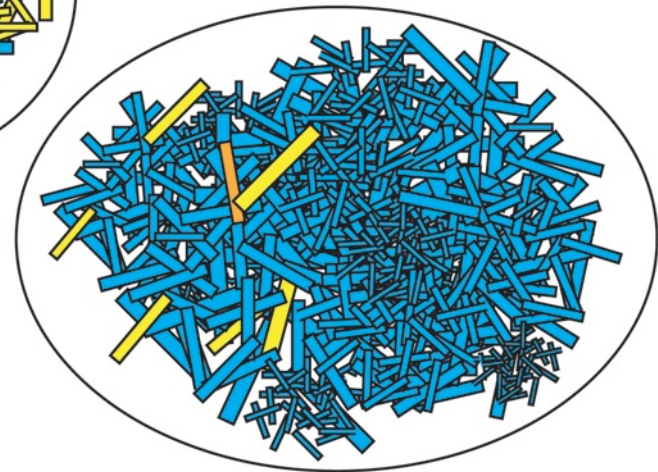
Pop A



Pop B



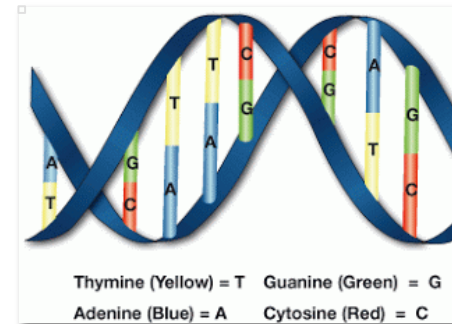
Pop C



 = CTACTACTACTACTA

 = CTACTACTACTA

 = CTACTACTA



Example of a DNA sequence alignment

```

Emoia_1 AACCTATTATCTCATTCCCACATCTGCCACACTAAGTATAATAACACCTCA
Emoia_2 AACCTATTATCTCATTCCCACATCTGCCACACTAAGTATAATAACACCTCA
Emoia_3 AACCTATTATCTCATTCCCTACAACCTGCCACACTAAGTATAATAACACCTCA
Emoia_4 AACCTATTATCTCATTCCCTACAACCTGCCACACTAAGTATAATAACACCTTA
Emoia_5 AACCTATTATTATATTCCCACAACCTGCCATACTGAGTATAATAACACTTCA
Emoia_6 AACCTATTATTATATTCCCACAACCTGCCATACTGAGTATAATAACACTTCA
Emoia_7 AACCTACTATTTTCATTCCCATAACCGCCACACTGAGTATAATAACACCTCA
Emoia_8 AACCTACTATTTTCATTCCCACAACCGCCACACTGAGTATAATAACACCTCA
Emoia_9 AACCGATTATTATATCCCCATAACTGCCACACTGAGTATAATAACACTTCA
Emoia_10 AACCGATTATTATATCCCCATAACTGCCACACTGAGTATAATAACACTTCA
  
```

The DNA sequences of closely related individuals are more similar than distantly related individuals.

Species C

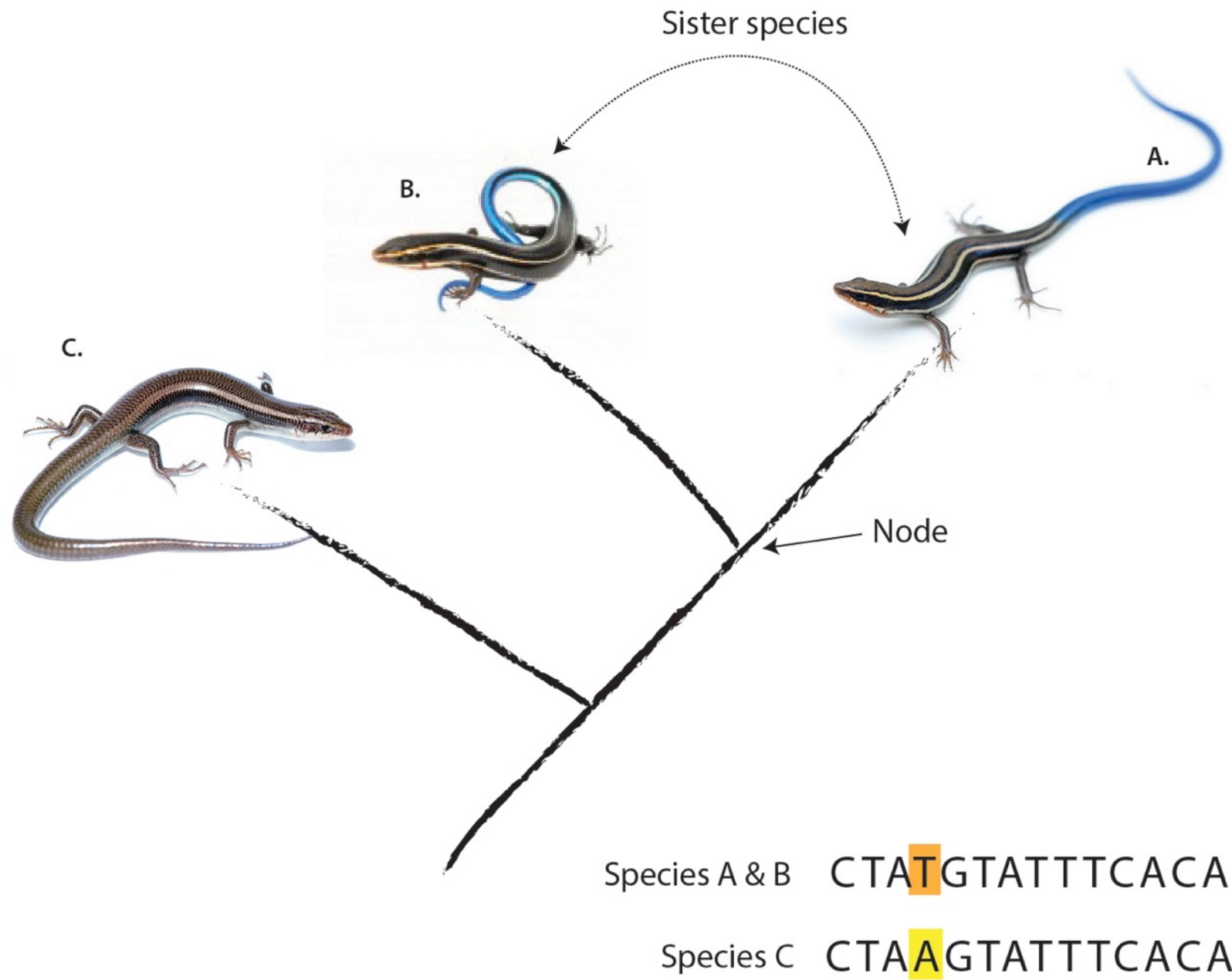


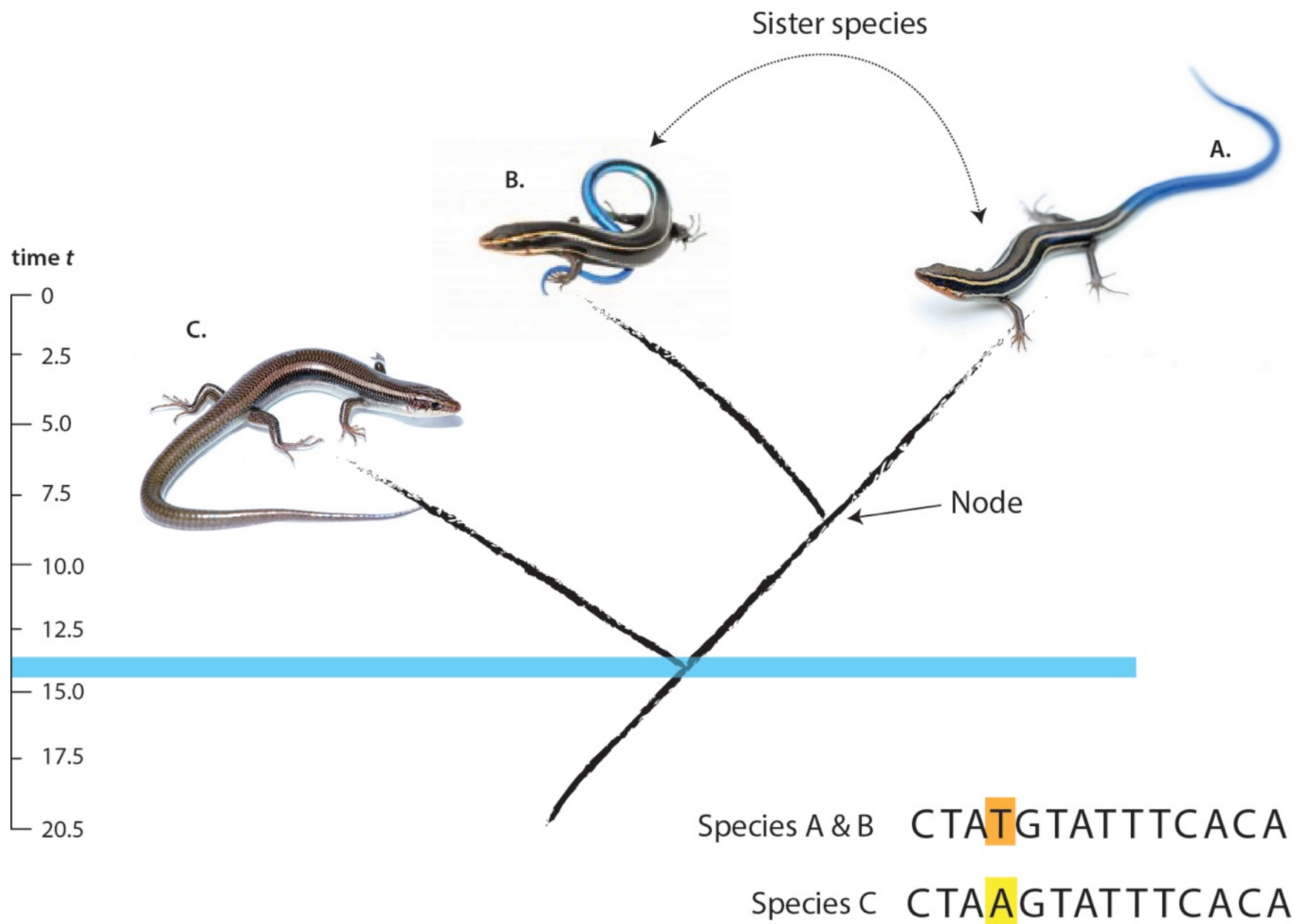
Species B



Species A





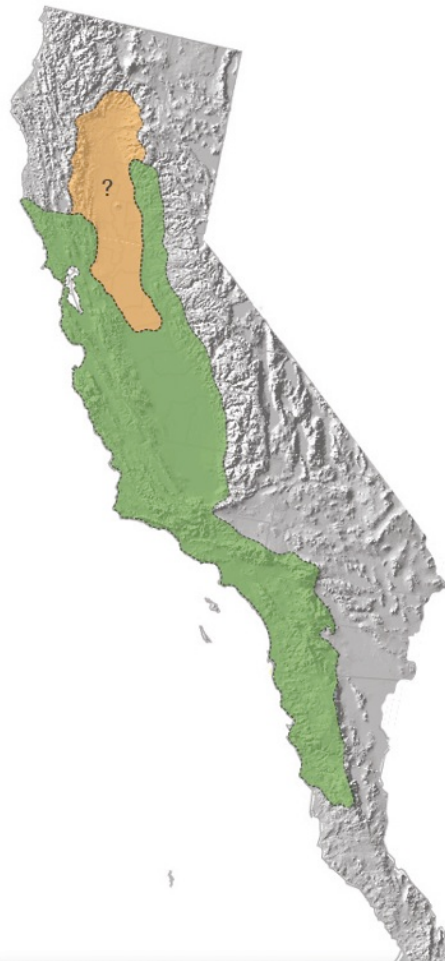




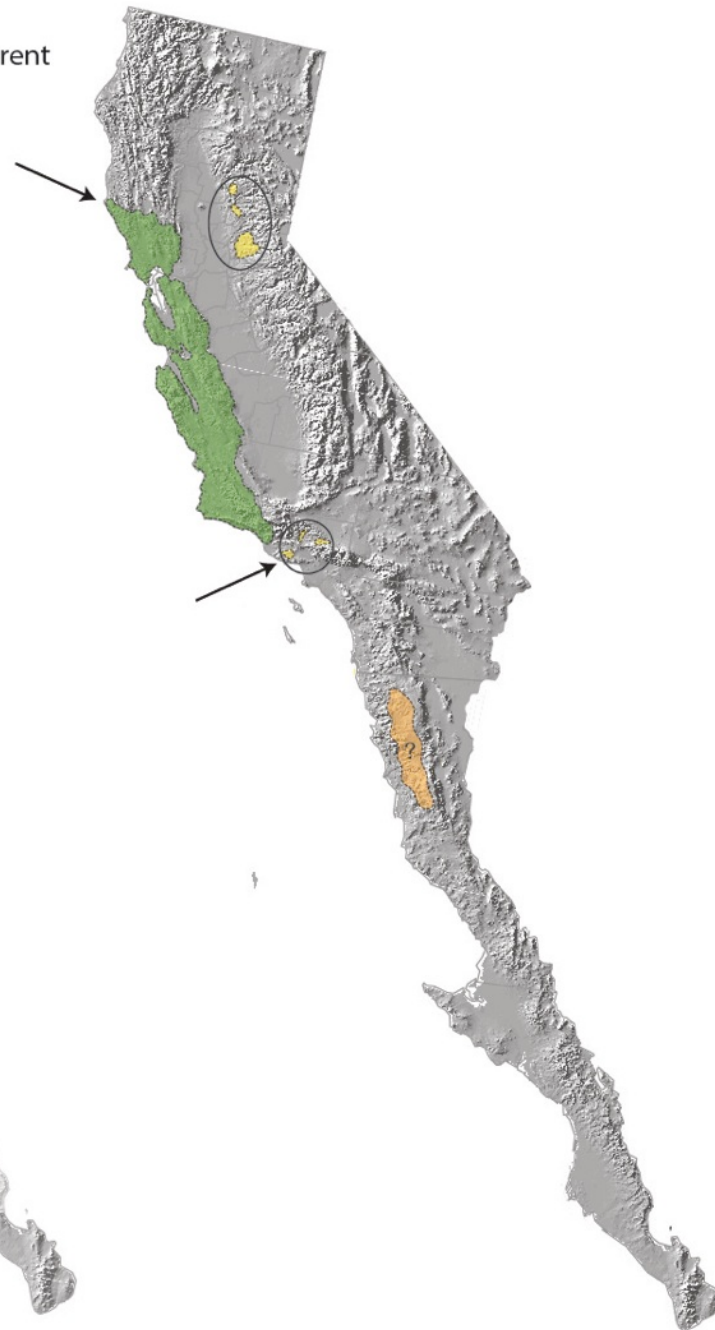




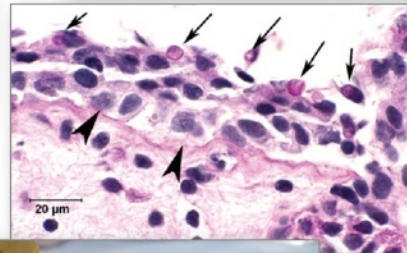
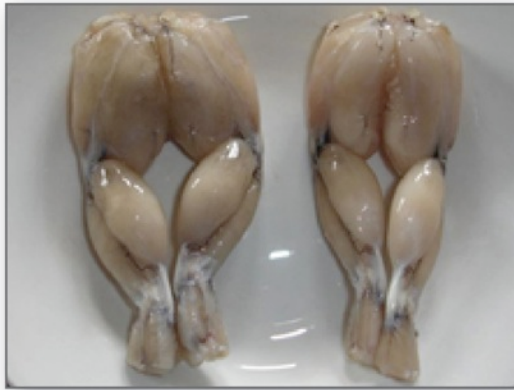
Historical



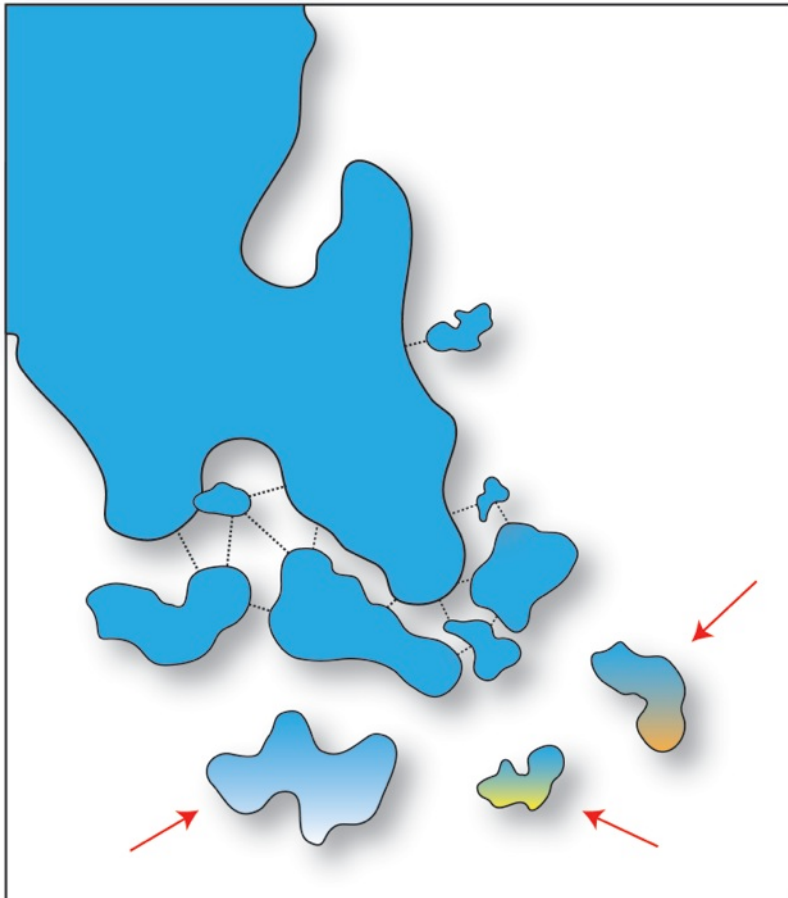
Current



Causes of decline...



Why are marginal populations often evolutionarily unstable?



margins of a hypothetical species distribution

1. Insufficient genetic variation to respond to changes in selective regimes at the species boundary...

or

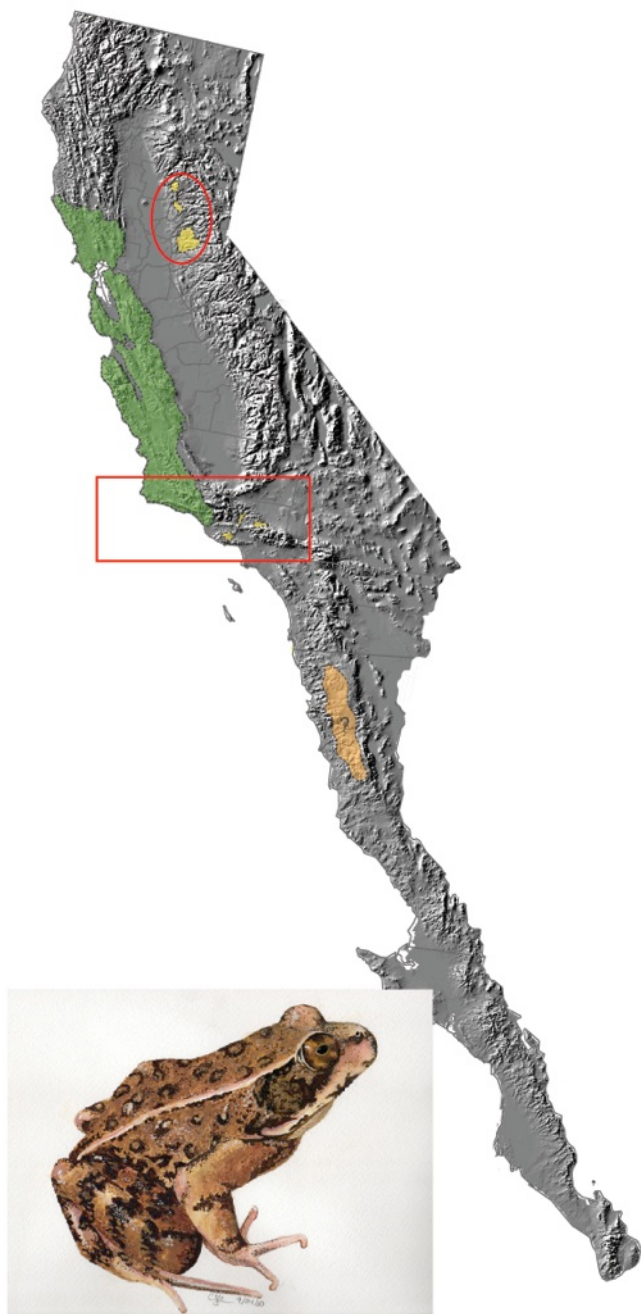
maladaptation to environments beyond the current species boundary.

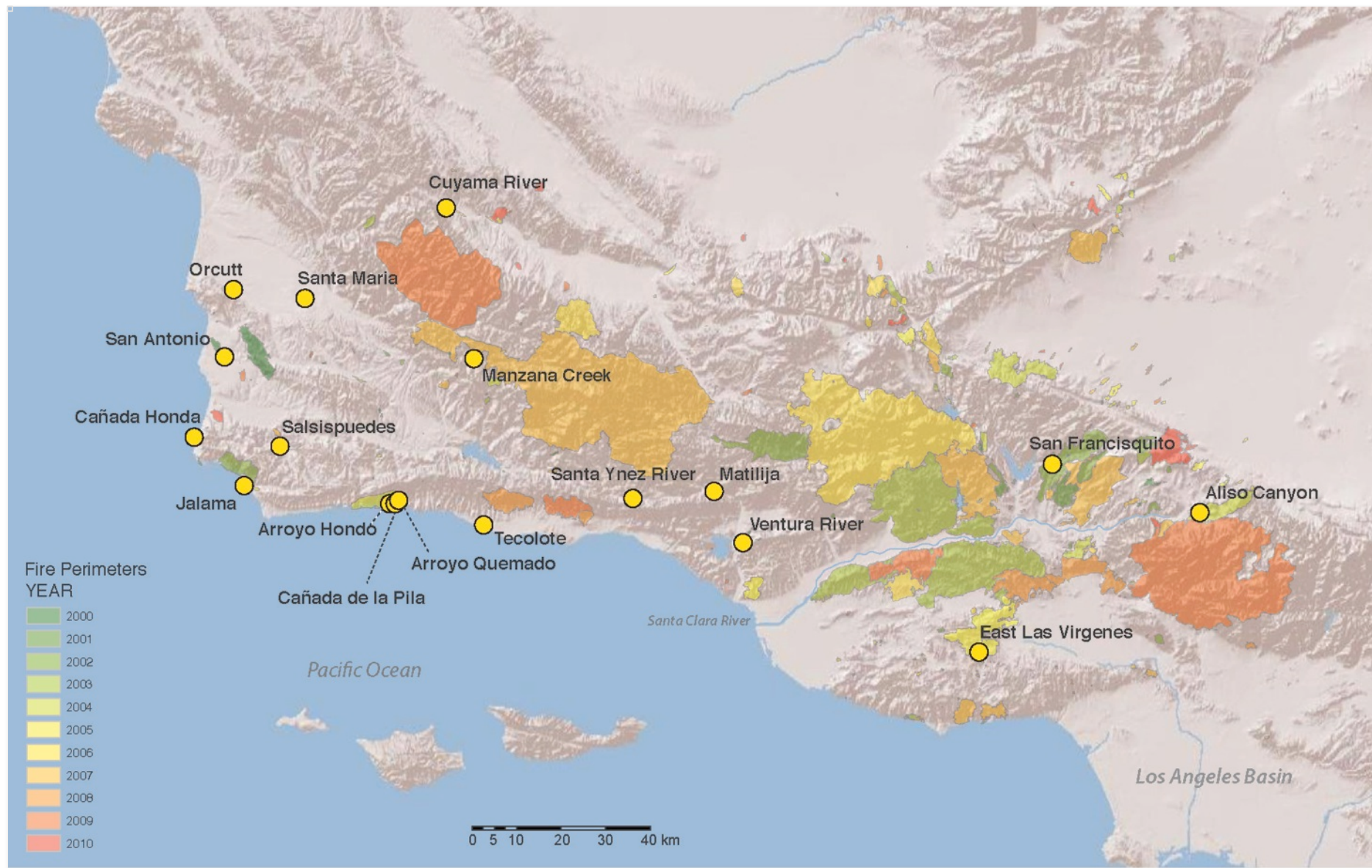
2. Increased competition with other species.

3. Reduced gene flow with 'internal' populations.

4. Fragmented populations of smaller size.

5. Interactions between these and other variables.



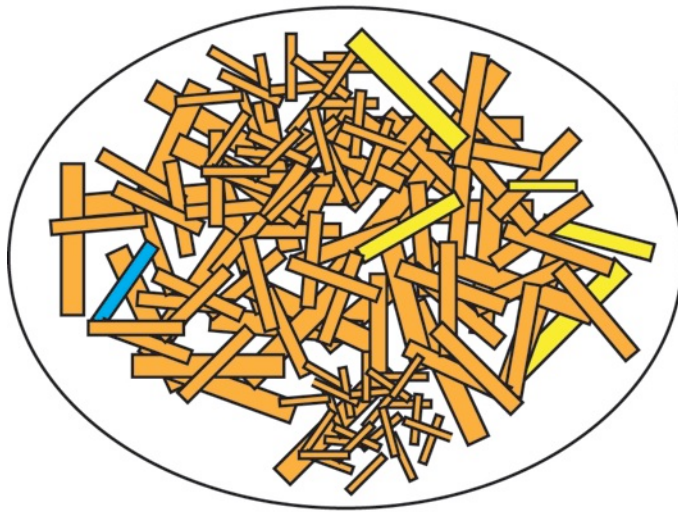


Some of our research questions:

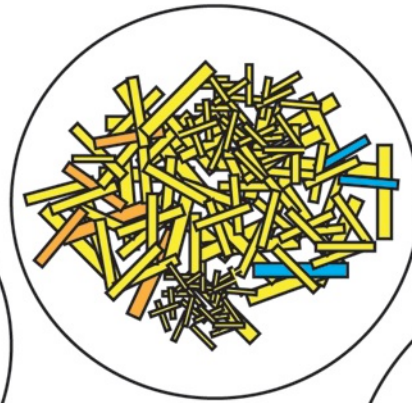
How are *R. draytonii* populations genetically structured across the landscape?



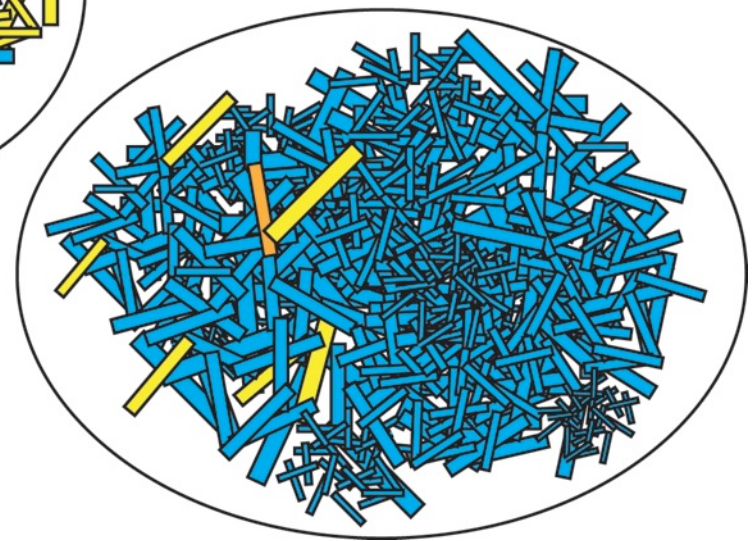
Pop A



Pop B



Pop C



Some of our research questions:

How are *R. draytonii* populations genetically structured across the landscape?

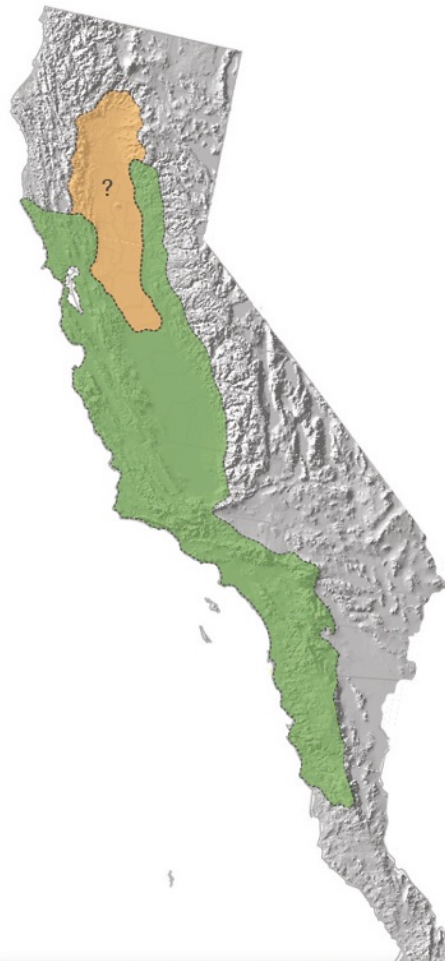
How does within-population genetic diversity change with respect to proximity to the range edge?

What factors determine the levels of within-population genetic diversity in California Red-legged frogs, particularly at the range edge?

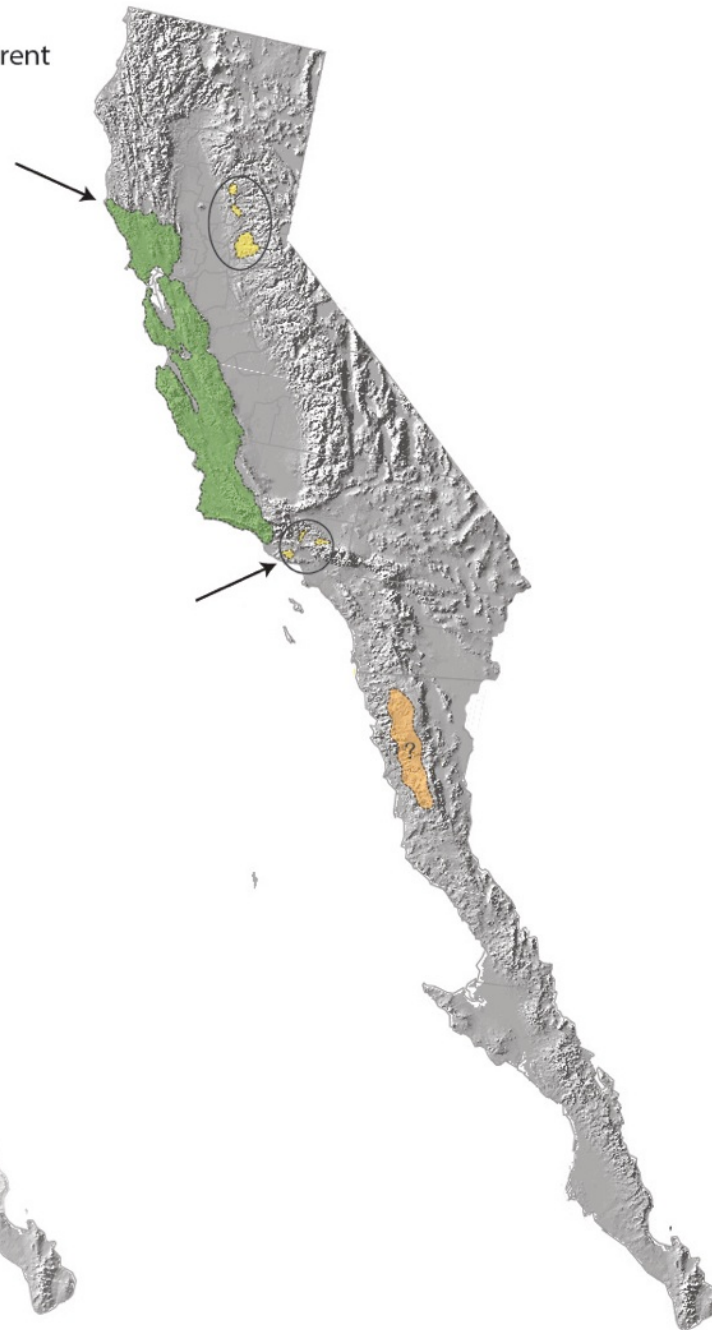
How can we identify suitable source populations for translocation to areas formerly occupied by *R. draytonii* in California?



Historical

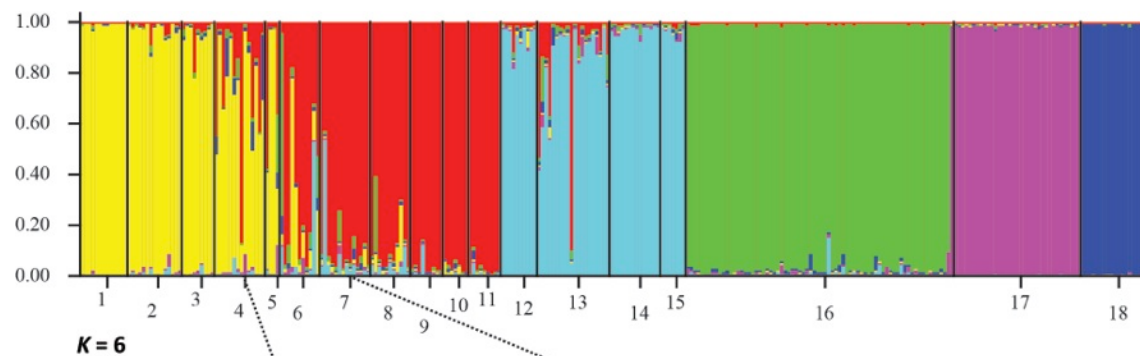


Current

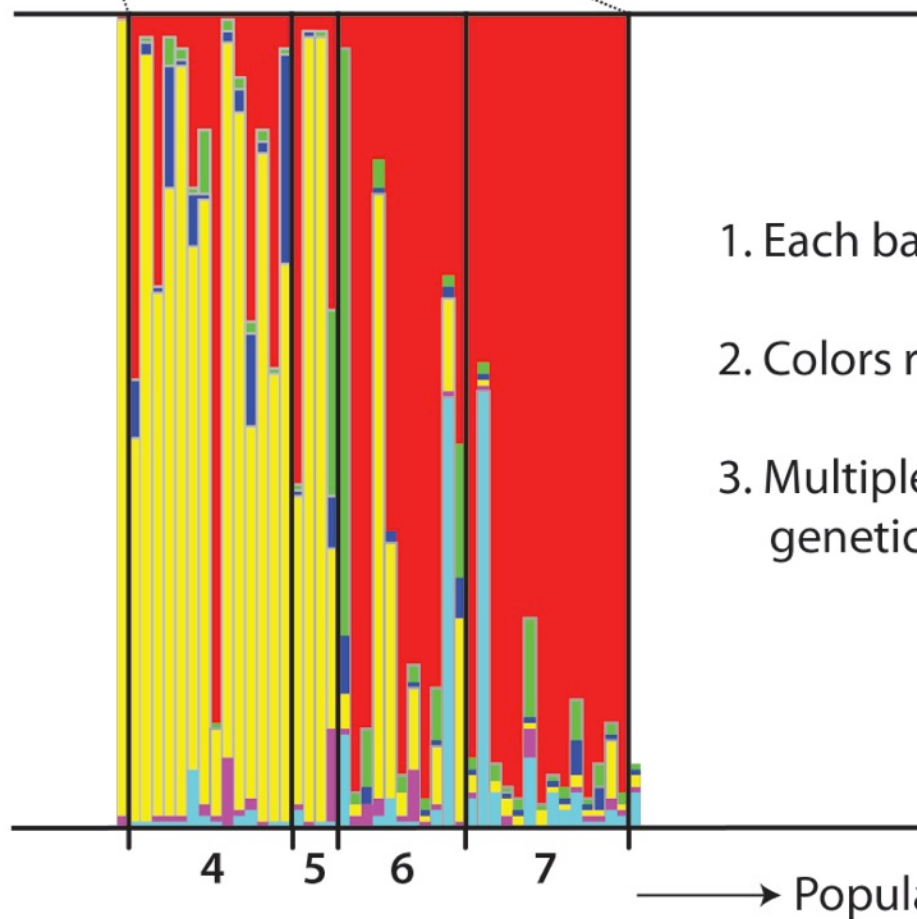


What types of data did we use?

1. Mitochondrial DNA sequence data (1013 nucleotides for the *cytb* gene)
 - Estimated a phylogenetic tree to assess the relationships of frogs from different sampling locations
 - Identify contact zones between evolutionarily divergent groups
2. Allele frequency data from 15 rapidly-evolving microsatellite loci
 - Identify natural genetic groups, or clusters based on the allele frequencies
 - Identify zones of admixture
 - Test for trends in genetic diversity with respect to proximity the range edge

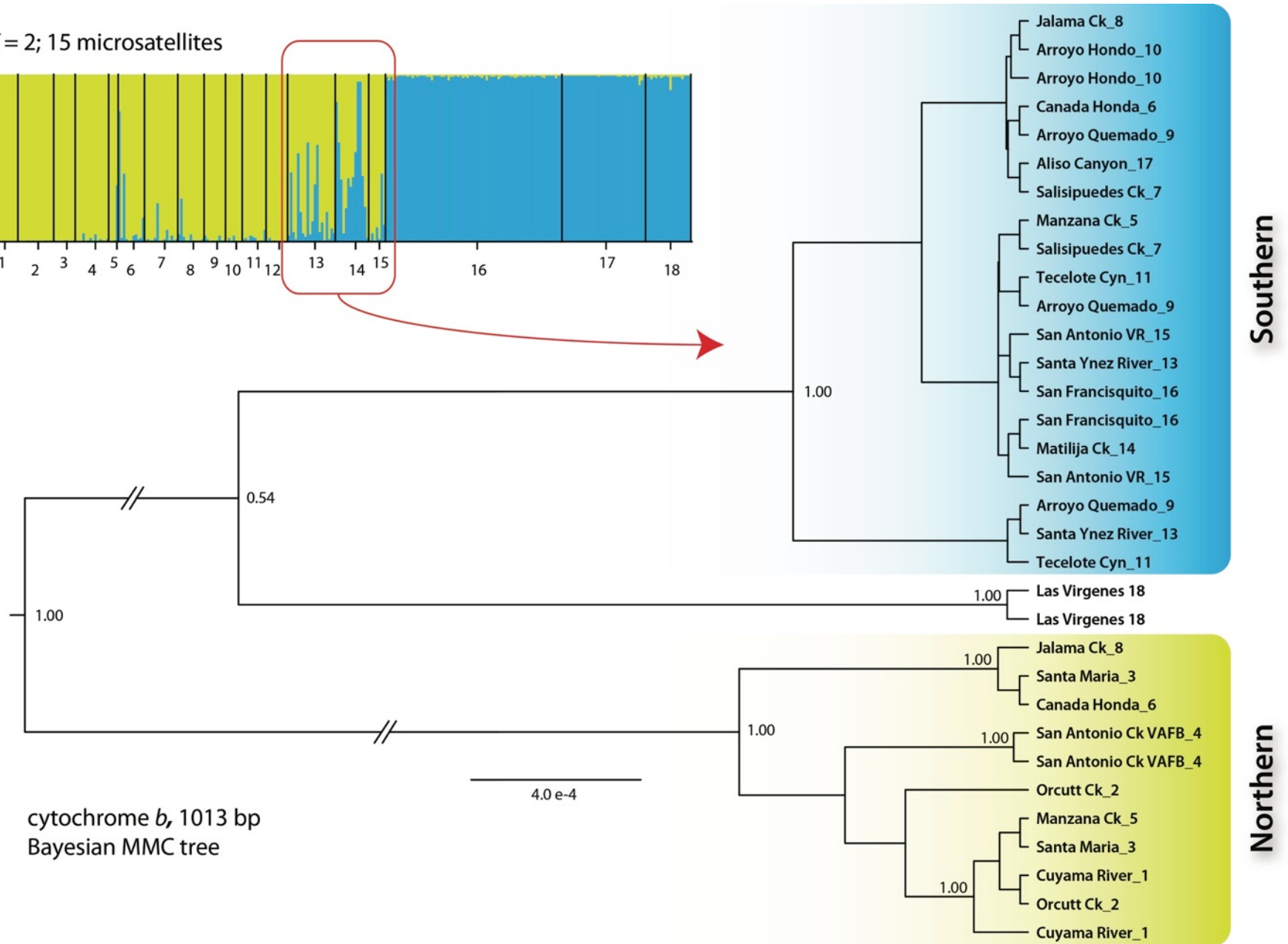
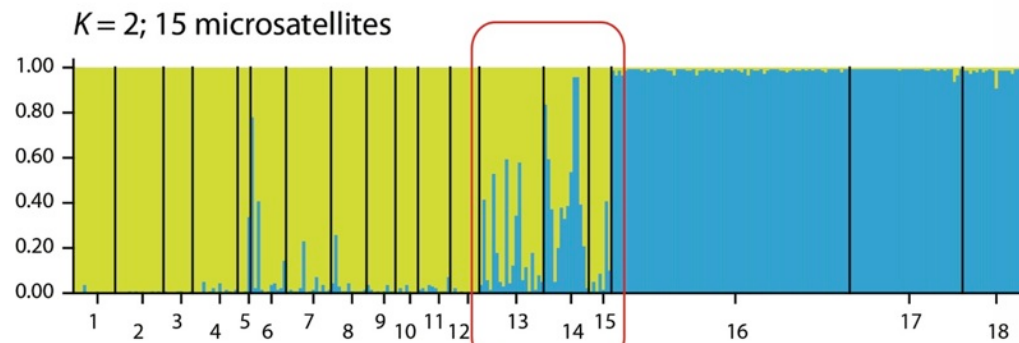


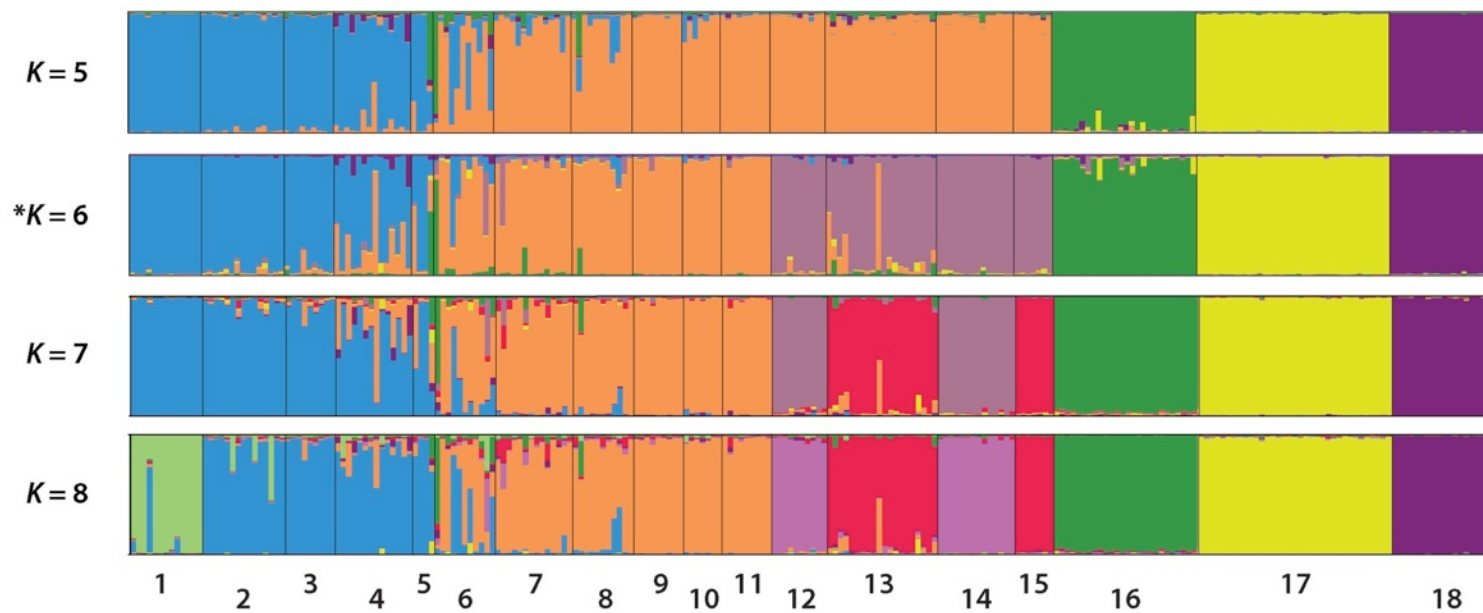
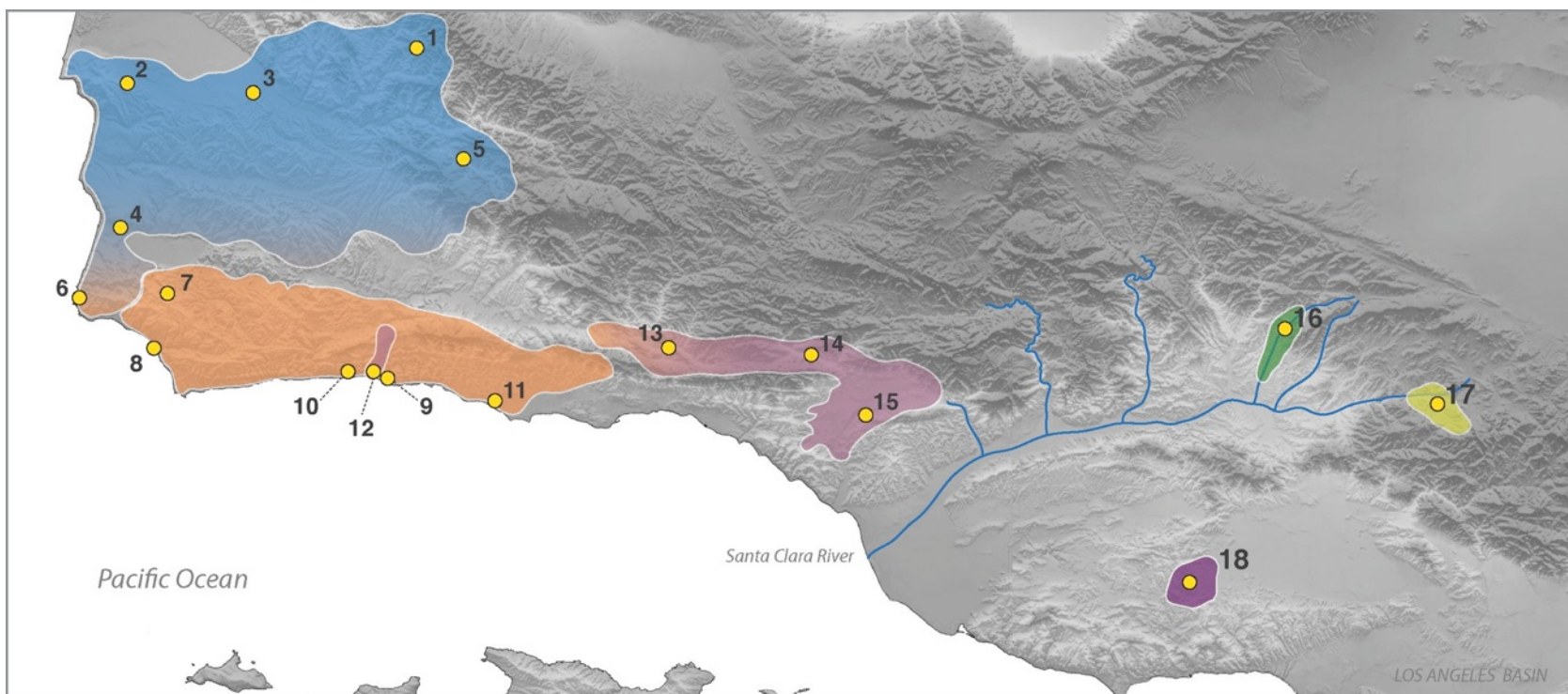
Assignment plot



1. Each bar in the histogram is an individual.
2. Colors represent genetic groups.
3. Multiple colors within a bar represent genetic admixture.

→ Population labels





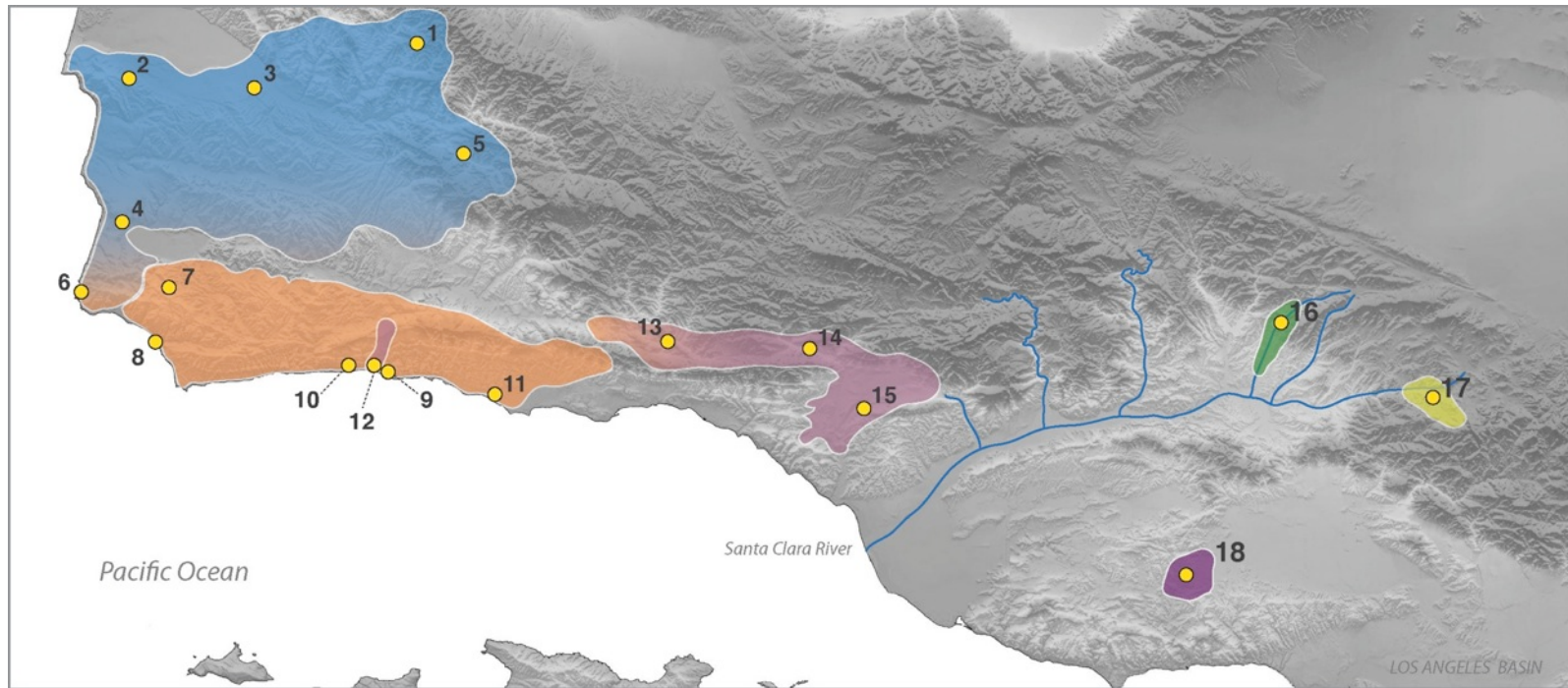
How can we measure genetic diversity?

Allelic richness A_r = number of alleles at locus for a given sample

Heterozygosity H = frequency of heterozygotes in a sample

- H ranges from zero (no heterozygosity) to nearly 1.0 (for a system with a large number of equally frequent alleles)
- H_O is the observed frequency of heterozygotes in a sample
- H_E is the expected heterozygosity under HWP

Relatedness = average pairwise measure of genetic similarity between individuals



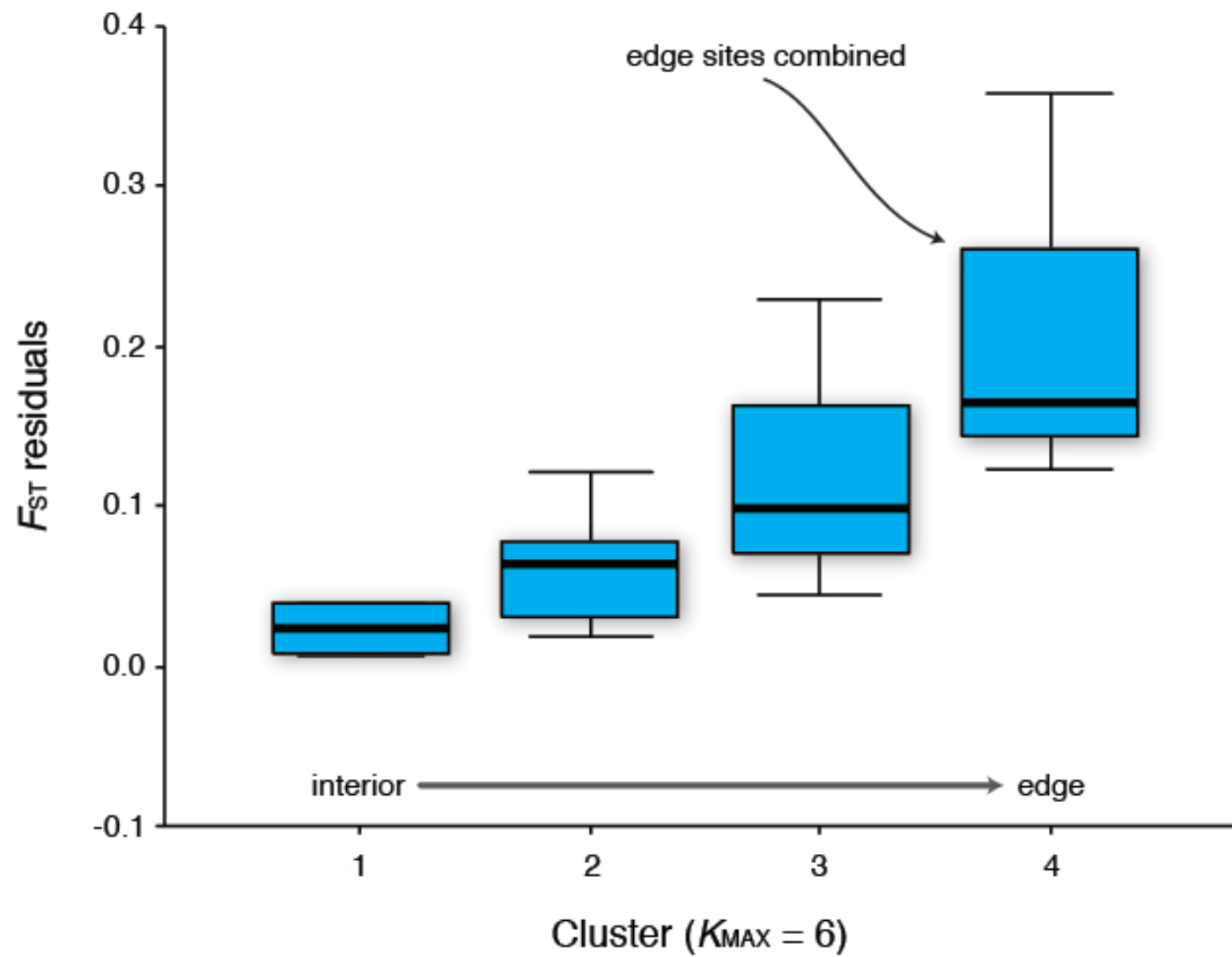
	Cluster	<i>N</i>	<i>A_R</i>	<i>H_O</i>	<i>H_E</i>	<i>r</i>	<i>M(M_c)</i>	<i>H_O excess</i>
W/NW ↓	Santa Maria 	55	5.00	0.54	0.63	0.11	0.88 (0.79)	*
	West Santa Ynez Mountains 	61	5.09	0.61	0.65	0.03	0.89 (0.79)	*
	East Santa Ynez Mountains 	49	4.79	0.51	0.59	0.05	0.85 (0.79)	-
	San Francisquito Canyon '09 	26	3.12	0.61	0.54	0.15	0.75 (0.77)*	**
	Aliso Canyon 	35	2.69	0.36	0.38	0.22	0.73 (0.77)	*
	East Las Virgenes Creek 	19	2.19	0.30	0.27	0.36	0.74 (0.75)*	-

longitude of sampling location vs.

A_R ($r^2 = 0.60$; $\beta = 0.77$; $P < 0.001$)

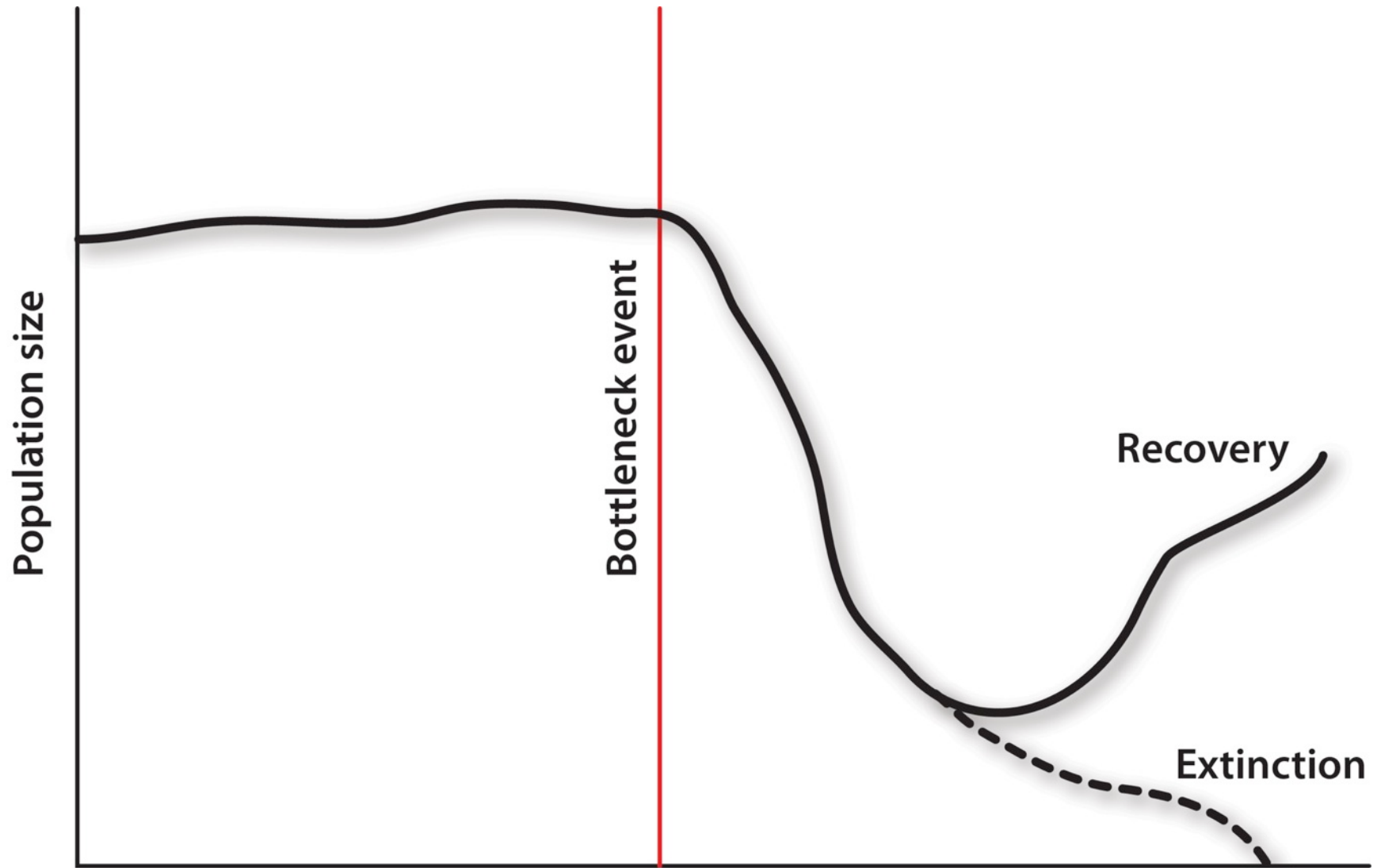
H_O ($r^2 = 0.37$; $\beta = 0.60$; $P < 0.008$)

R ($r^2 = 0.52$; $\beta = -0.72$; $P < 0.001$)





Station Fire north of Los Angeles, CA; Sept. 2009.

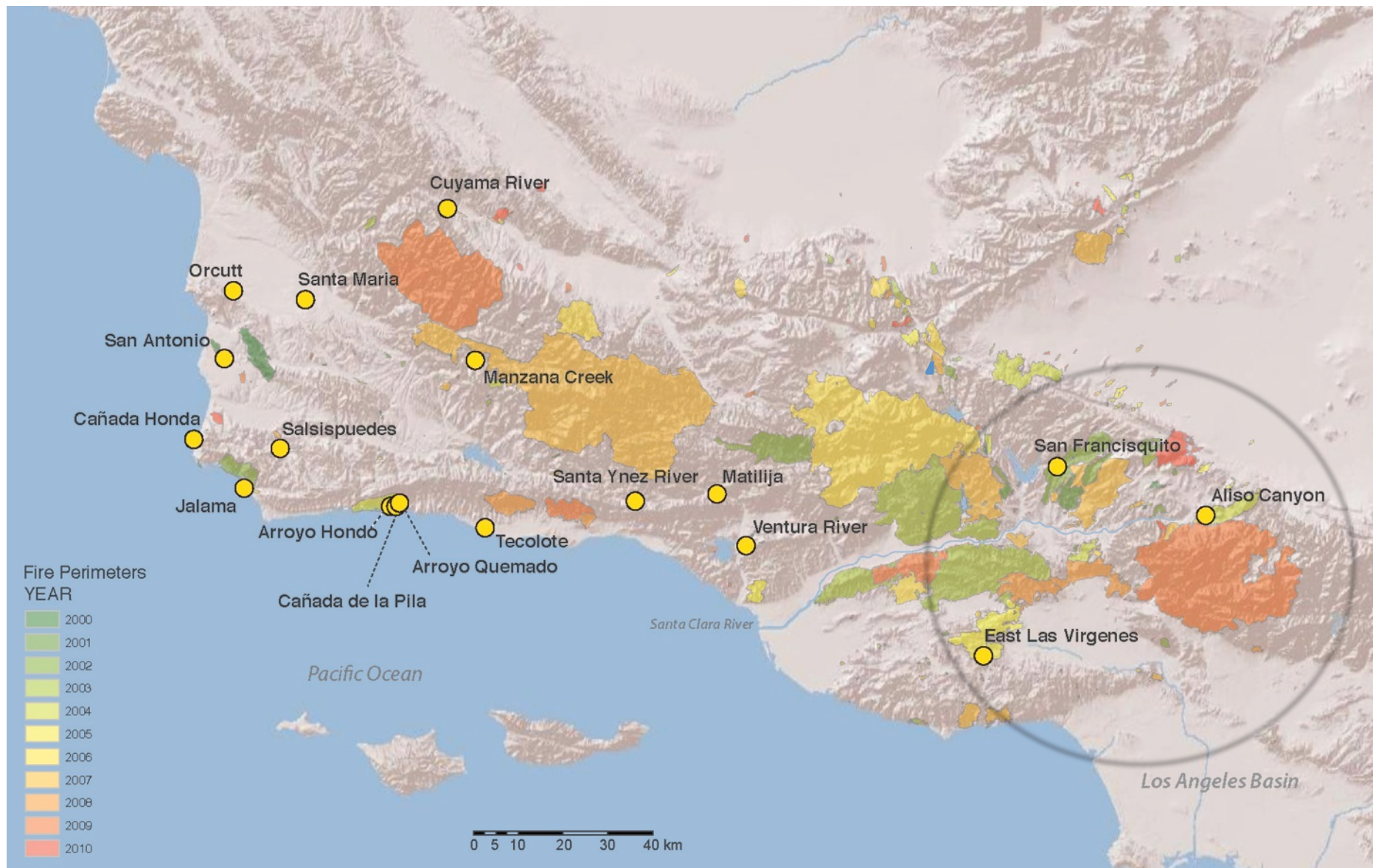


Sample Site	<i>N</i>	<i>Ar</i>	<i>H_o</i>	<i>H_e</i>	<i>F_{is}</i>	<i>R</i>	<i>H_o excess</i>
Cuyama River	13	2.23	0.40	0.41	0.04	0.52	
Orcutt Creek	15	3.23	0.62	0.61	-0.03	0.26	**
Santa Maria	9	2.80	0.43	0.49	0.12	0.42	
San Antonio Creek VAFB	14	3.16	0.63	0.60	-0.06	0.26	*
Manzana Creek	4	2.67	0.60	0.49	-0.19	0.44	*
Cañada Honda Creek	11	3.58	0.63	0.66	0.05	0.05	
Salsispuedes Creek	14	3.36	0.66	0.63	-0.06	0.15	
Jalama Creek	11	3.44	0.71	0.63	-0.12	0.15	
Arroyo Quemado	9	2.96	0.60	0.54	-0.13	0.32	*
Arroyo Hondo	7	3.03	0.66	0.58	-0.14	0.26	**
Cañada de la Pila	9	2.71	0.59	0.50	-0.18	0.44	
Tecolote Canyon	10	2.38	0.41	0.41	0.01	0.49	
Santa Ynez River	20	2.93	0.55	0.54	-0.02	0.30	*
Matilija Creek	14	2.43	0.49	0.46	-0.06	0.46	
Ventura River	7	2.15	0.33	0.36	0.10	0.53	
San Francisquito Canyon	26	2.53	0.61	0.54	-0.12	0.39	***
Aliso Canyon	35	2.17	0.36	0.38	0.05	0.61	*
East Los Virgenes Creek	19	1.79	0.30	0.27	-0.09	0.77	
Total	247						

**P* < 0.05

***P* < 0.01

****P* < 0.001





San Francisquito, Spring 2002



Copper Fire, June 5-11 2002



San Francisquito Canyon 2002, post-Copper Fire



Copper Fire June 5-11

3 Breeding Pools & Catchment Basin
Constructed

1999

USGS Monitoring Initiated

2002

CRLF Population Estimates
300-360 Adults

2003

Recruitment in Breeding Pools Only

2004

50 Egg Masses Observed
With No Recruitment in creek

2005

Low Rainfall Year

4 Additional Breeding Pools Constructed
Only 2 of 7 Breeding Pools Remain

Record High Rainfall & Flood Event

23 CRLF Adults Detected

2006

4 Additional
Breeding Pools
Constructed

Lowest Rainfall Year

2007

5 CRLF Adults Detected

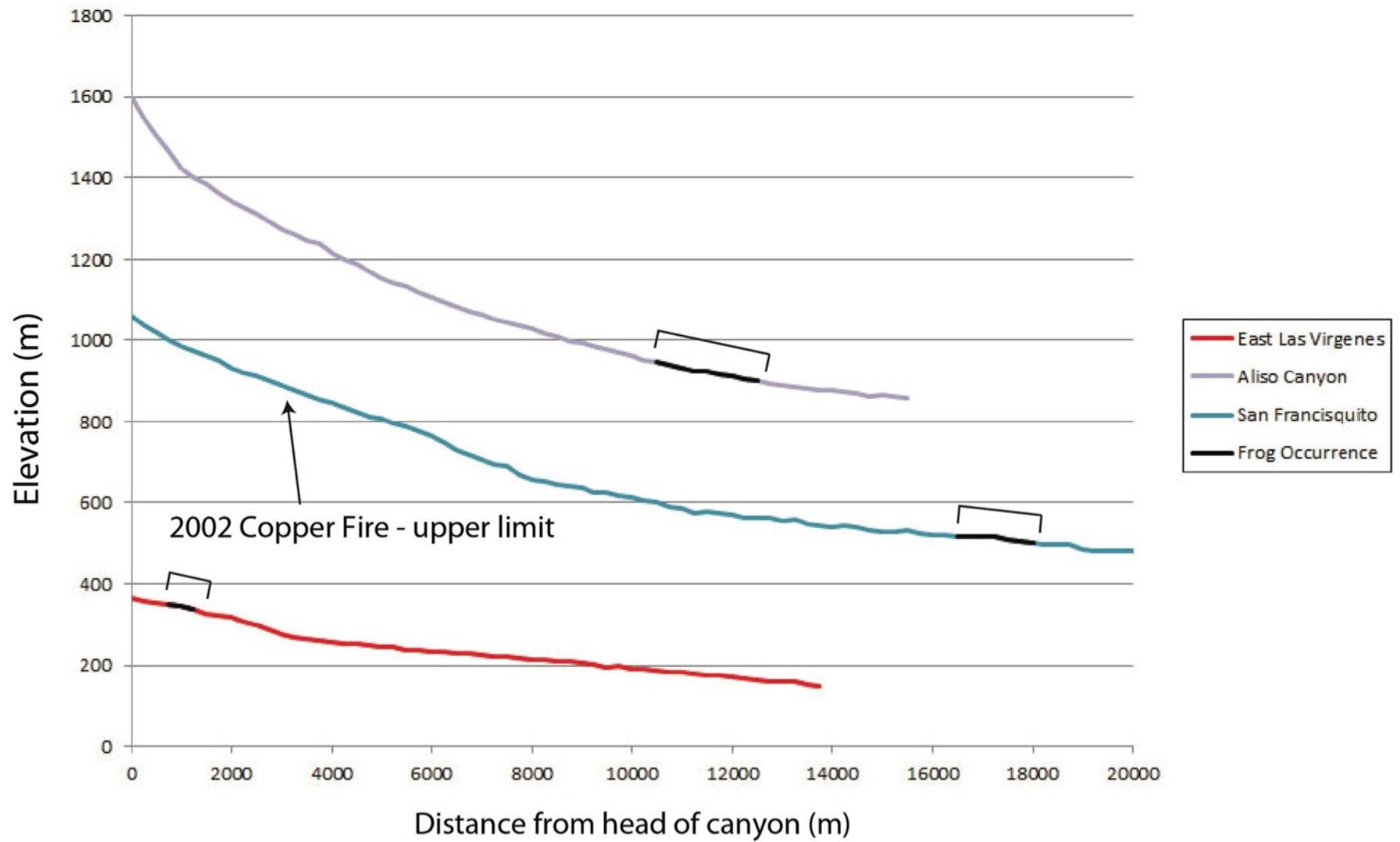
2008

33 Adults Detected

Only 4 CRLF Egg Masses
(No Recruitment)
& 1 Adult Detected

Tadpoles
Detected 1st time
since 2003

Profiles of creeks with *R. draytonii* occurrence



Comparison of genetic diversity indices by year (San Francisquito Canyon)

Year cohort	n	A_R	H_o	H_s	R
2002	28	3.04	0.61	0.59	0.09
2005-06	24	3.26	0.57	0.60	-0.13
2009	26	3.03	0.61	0.55	0.18
P -value (5000 perm.)		0.40	0.34	0.26	0.27

Comparison of pairwise F_{st} values across years

Year cohort	2002	2005-06	2009
2002	–	-0.0043	0.0473
2005-06		–	0.0229
2009			–

What about effective population size?

It's true that most populations are large, but they don't necessarily act large...

This is because individuals vary in their ability to pass their genes on to the next generation.

So it is important to distinguish between

- The census (actual) population size N
- and the effective population size N_e (used for genetic calculations)

Anything that increases the variance among individuals in reproductive success will reduce N_e

Estimates of effective population size (N_e)

Sample	N_{loci}	$N_{e(\text{ABC})}$ (95%CI)	$N_{e(\text{LD})}$ (95%CI)
Santa Maria	15	28.01 (23.24 - 51.98)	24.60 (20.8 – 29.95)
West Santa Ynez Mtn. 1	15	26.00 (23.74 - 29.95)	55.7 (42.9 – 75.6)
East Santa Ynez Mtn. 2	15	26.82 (21.63 - 36.56)	25.5 (19.9 – 33.4)
San Francisquito 2002	15	27.75 (24.82 - 34.65)	65.2 (25.9 - ∞)
San Francisquito 2005-06	15	25.66 (22.24 - 35.96)	46.1 (21.1 – 525.5)
San Francisquito 2009	15	20.27 (17.80 - 25.12)	11.1 (7.4 – 17.0)
Aliso Canyon	13	29.73 (25.14 - 44.79)	18.1 (8.4 – 47.1)
East Las Virgenes Canyon	12	15.93 (13.45 - 21.51)	12.3 (2.5 - ∞)

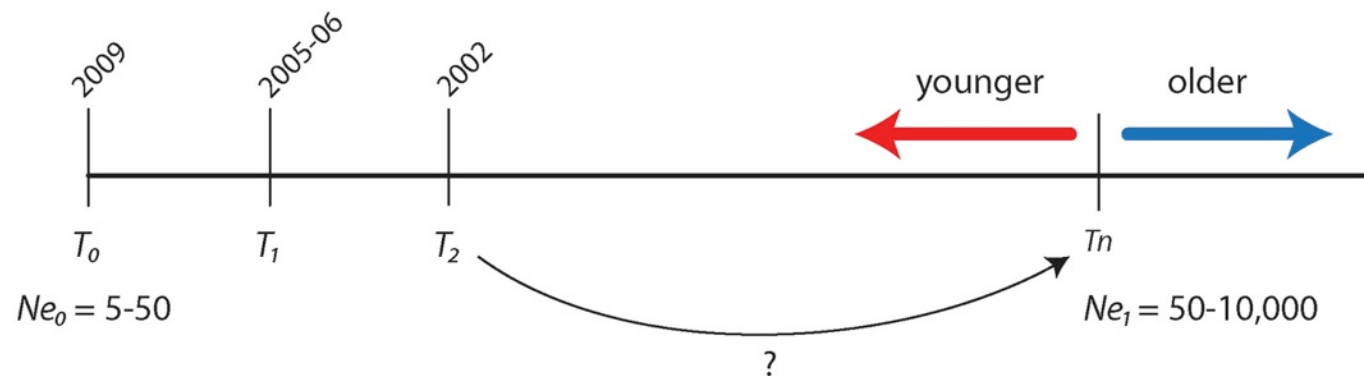
Why don't we see convincing evidence of a genetic decline following the fire-flood event, even though census surveys indicate that the population declined dramatically?

1. The population was bottlenecked to begin with, so only common alleles were present before the fire.
2. Frogs can and do survive the fires.
3. Long generation time causes genetic inertia, delaying the signature of a bottleneck in allele frequency data.

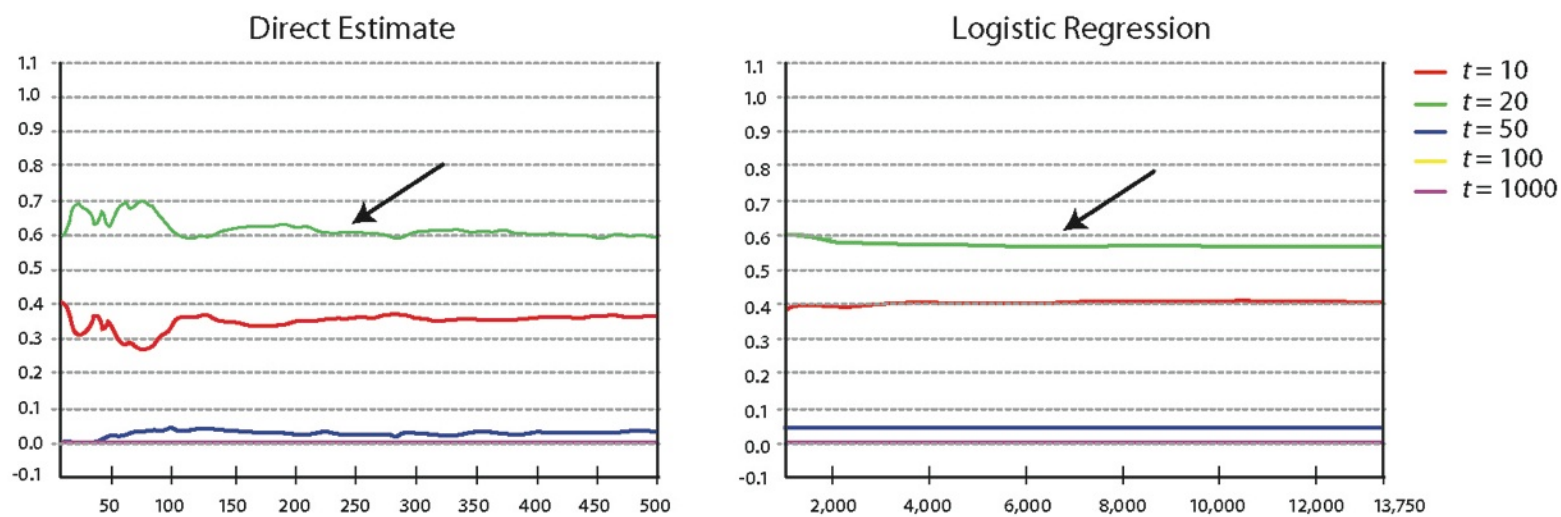
If bottleneck signatures cannot be traced to recent fires, then how long ago did the bottlenecks take place?

Approximate Bayesian Computation - DIYABC v1.0.4.46

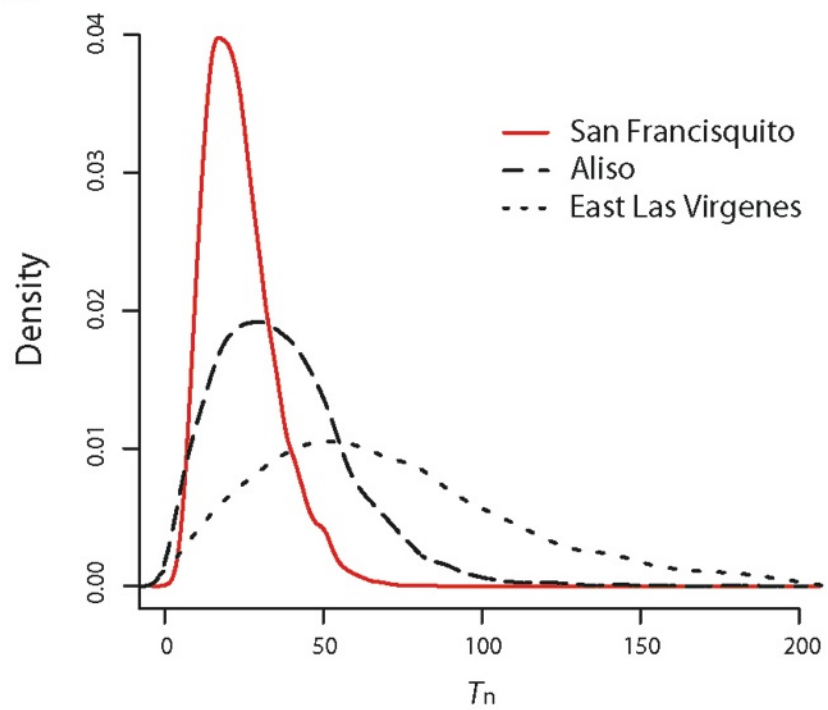
1. Simulate datasets assuming different demographic scenarios:
 T_n number of generations ago (1-200), the effective population size Ne_1 (50-10000) greater than the current effective size Ne_0 (5-50)
2. Select simulated data sets (i.e scenarios) closest to the observed data set and estimate the posterior distribution of parameters through a linear regression procedure
3. Posterior predictive checks using the best-fit model to compare simulated parameter values to the observed data.



A.



B.





Saint Francis Dam, 28 March 1928



St. Francis Dam, 29 March 1928

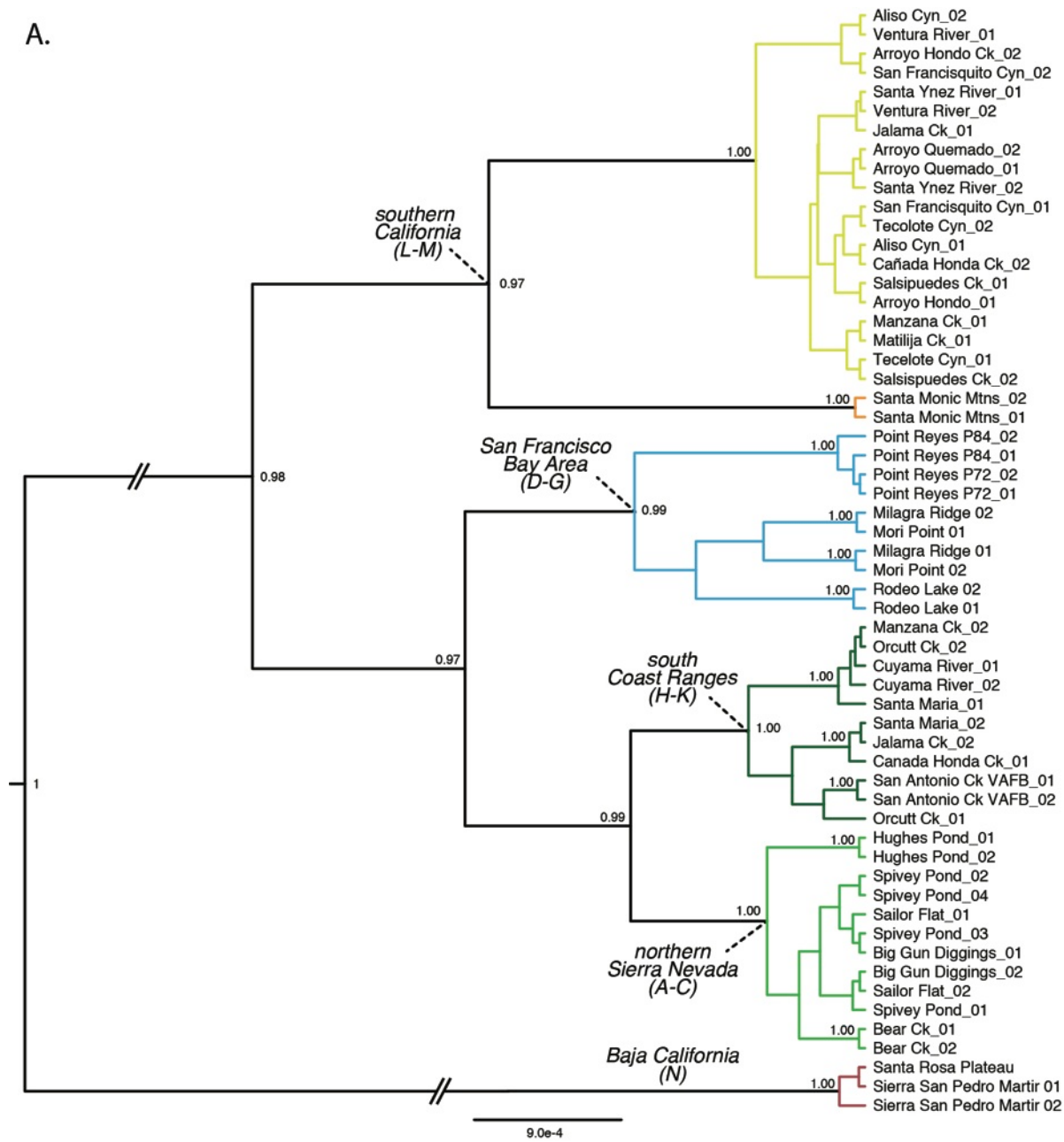




Current state of knowledge

1. Isolation, depauperate genetic diversity, lack of current gene flow, inability for new immigrants to enter range boundary populations, and susceptibility to fire/flood events suggests that peripheral populations in southern California are in an unstable condition and need to be managed.
2. Continued genetic monitoring for populations that have undergone known, severe declines would go a long way towards understanding the severity of bottlenecks and their implications for long term survival.
3. Spatial genetic data can and are being used to inform translocation efforts in manner consistent with the evolutionary dynamics of the species at the southern range boundary.

A.



B.

