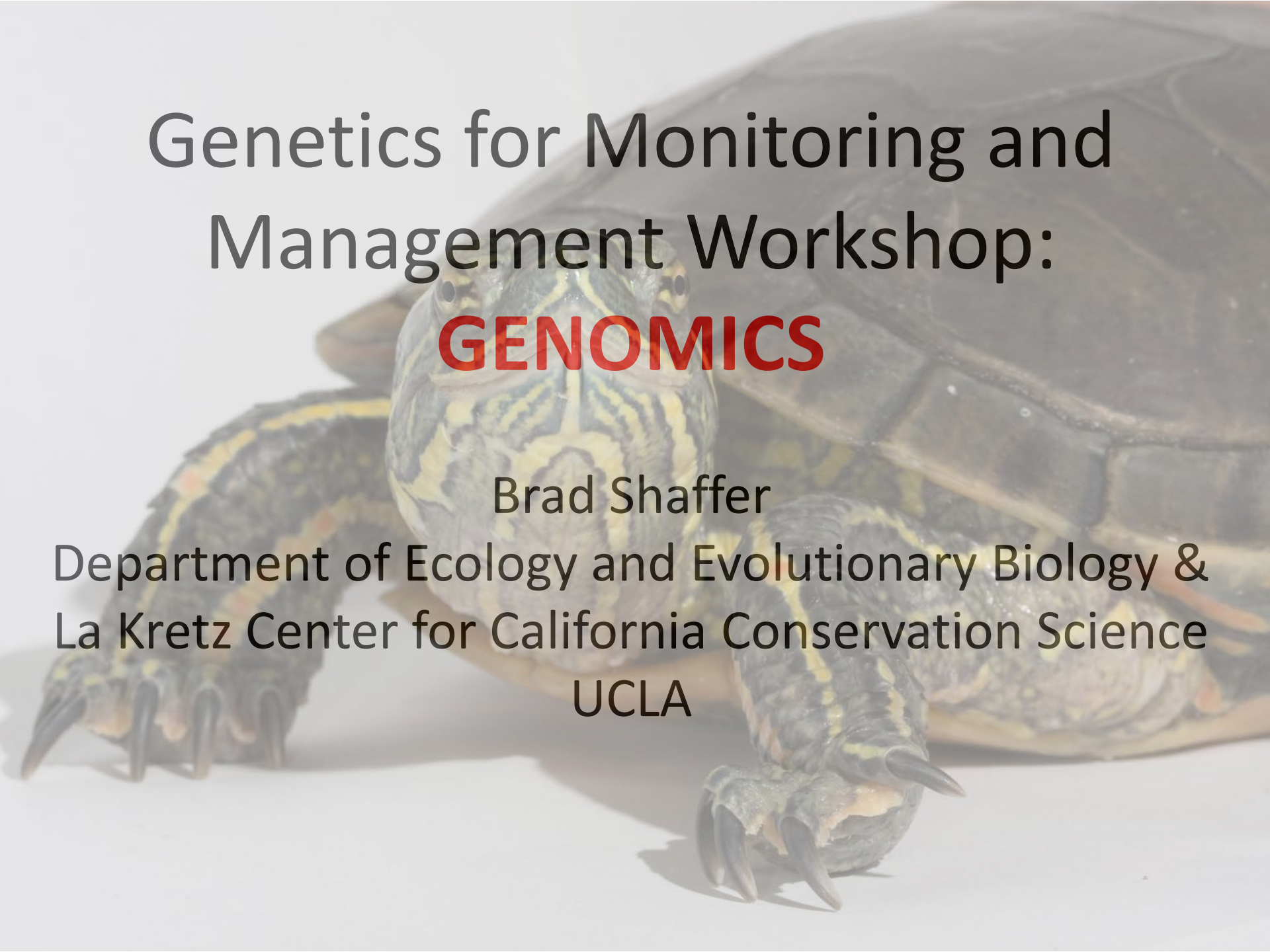


Genetics for Monitoring and Management Workshop: GENOMICS

Brad Shaffer
Department of Ecology and Evolutionary Biology &
La Kretz Center for California Conservation Science
UCLA

Genomics:

1) Genomics is a discipline in genetics that applies recombinant DNA, DNA sequencing methods, and bioinformatics to sequence, assemble, and analyze the function and structure of genomes (the complete set of DNA within a single cell of an organism). *Wikipedia*

2) Collecting a lot of sequence data for an organism or study (100's to 10,000,000's of markers).

Some genomic-y terms and concepts:

Next-generation sequencing (NGS): New sequencing machines and strategies that allow per-nucleotide sequencing costs to drop by ~10,000 fold. Illumina HiSeq is the most common.

Coverage: The average number of times that you sequence each nucleotide (1X, 10X, 100X).

Transcriptome: The RNA transcripts that float around in cells as they turn DNA into protein.

More genomic-y terms and concepts:

Gene capture arrays (AKA pull-down arrays, gene baits): A strategy used to isolate a specific set of sequences from each individual that will later be sequenced. Requires some knowledge of the genome already (to make the baits).

RADseq: An inexpensive approach that uses restriction enzymes to isolate a subset of the genome without baits or prior genomic knowledge. Relatively inefficient, but cheap and easy. Works within species pretty well.

Yet more genomic-y terms and concepts:

Whole genome resequencing: The emerging strategy where you just sequence the whole damn genome for each individual. Currently only feasible at *low coverage*.

Annotation: Putting a name, a boundary, and a function on each **gene** in the genome.

SNPs (Single Nucleotide Polymorphisms): A single site in the DNA that has two alternative states (nucleotides) present. The unit of choice for many analyses.

Cost:

(subject to change without notice)

Gene capture arrays: 1000-5,000 markers, ~$100 per individual

RADseq: 10,000+ markers, ~$20/individual

Whole genome resequencing: Depends on coverage. 0.5X coverage for a human-sized genome, potentially around $150/individual

Why genomics for management?

-More data allows for finer scale inferences, greater confidence

-Chasing important genes, understanding adaptation to climate change, disease, etc.

-Developing large marker panels for important taxa

At the end of the day, is it worth it?

Three examples from our lab:

1) Western pond turtle phylogeography and species delimitation

2) California tiger salamander hybridization

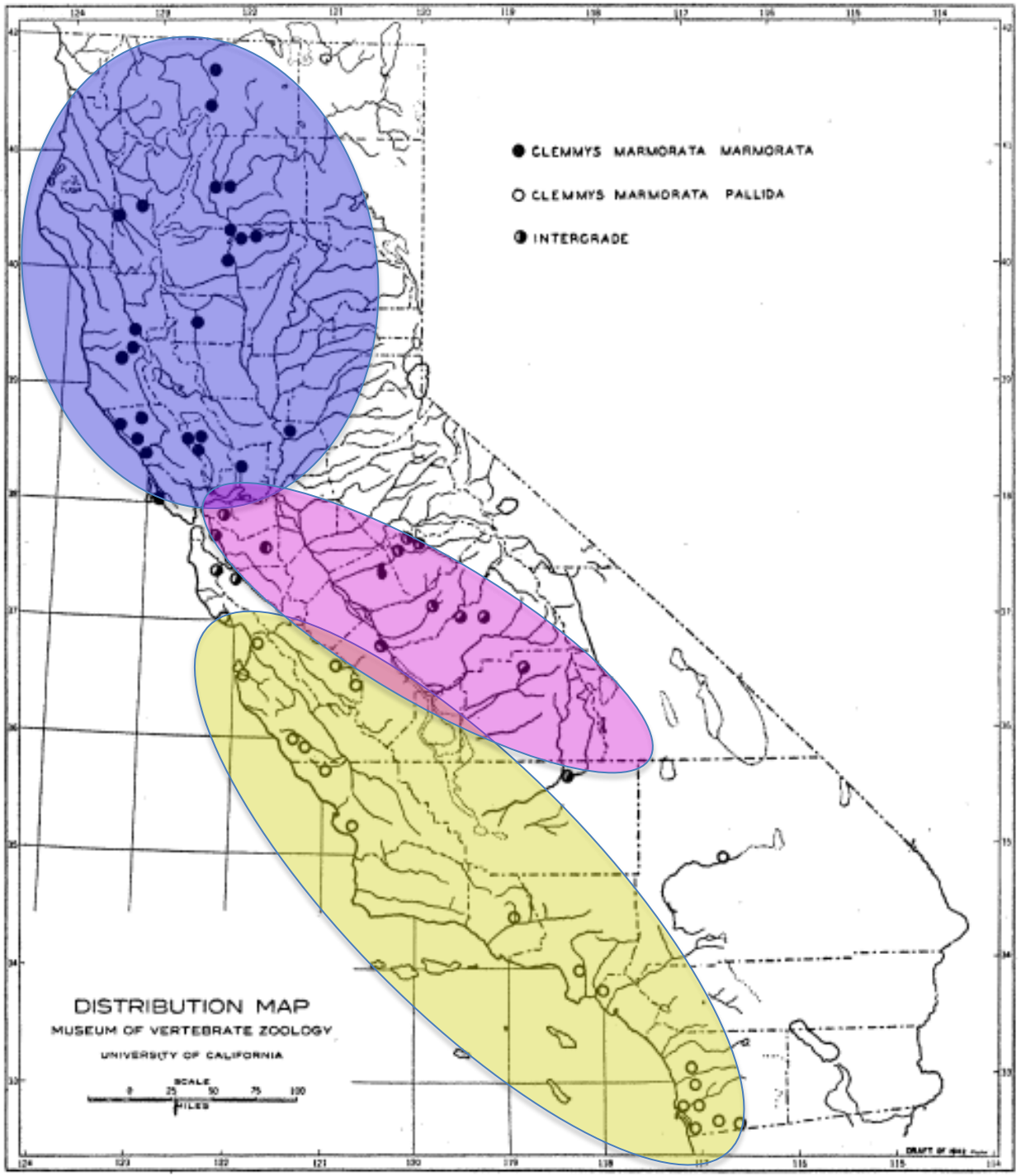
3) Desert tortoise population genomics

Consider *Emys marmorata*

(CA Species of Special Concern, candidate for ESA listing)

E. m. marmorata
Northern, healthy
populations in some areas

E. m. pallida
Southern, declining



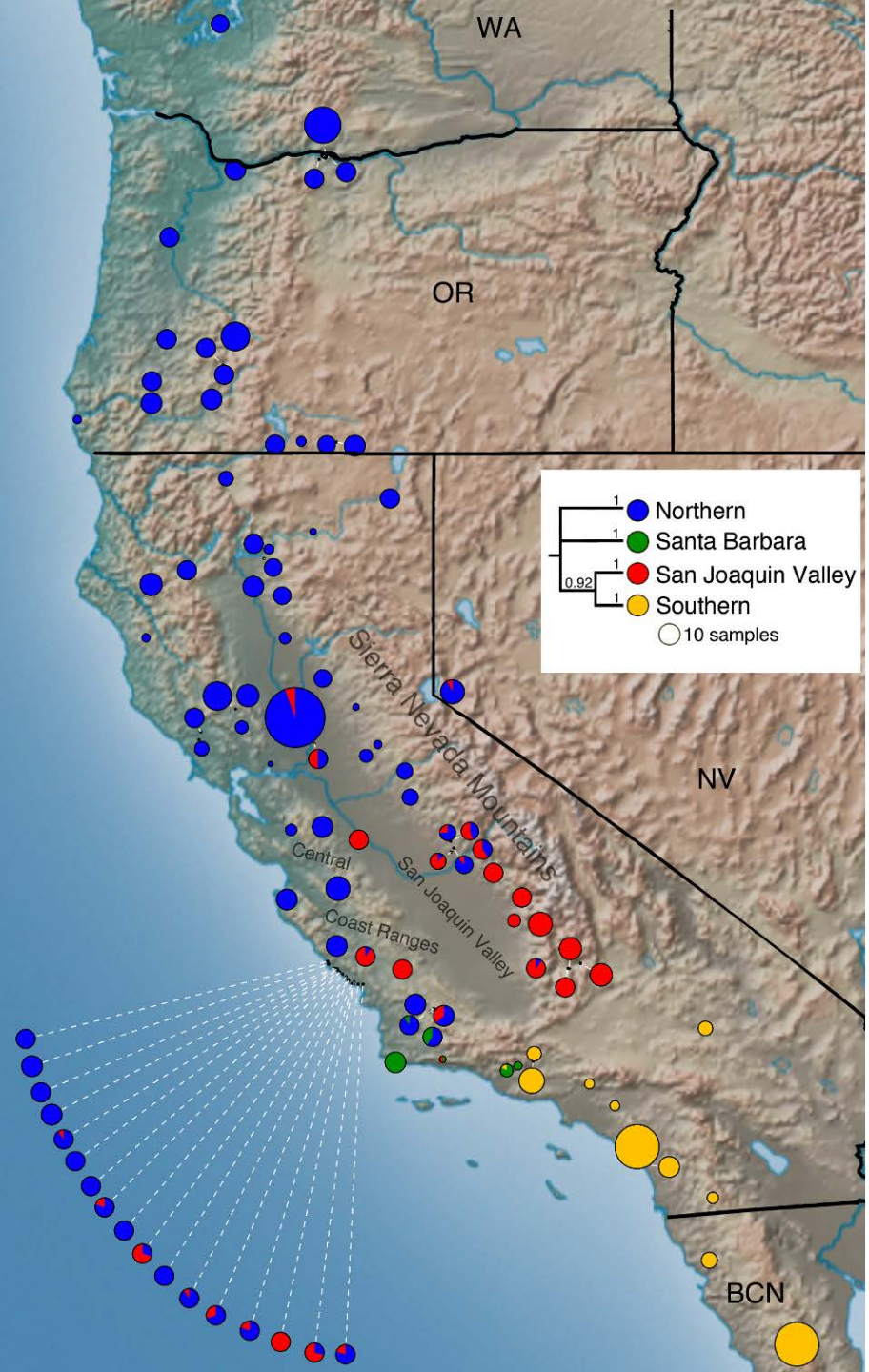
Seeliger 1945 (morphology)

North of SF bay = *marmorata*

South of Monterey – *pallida*

Bay Area, San Joaquin Valley = intergrades

Conclusion: single species, wide zone of intergrades



Spinks et al. 2005, 2010 & submitted

mtDNA only (nuclear data uninformative)

North of SF Bay *marmorata*

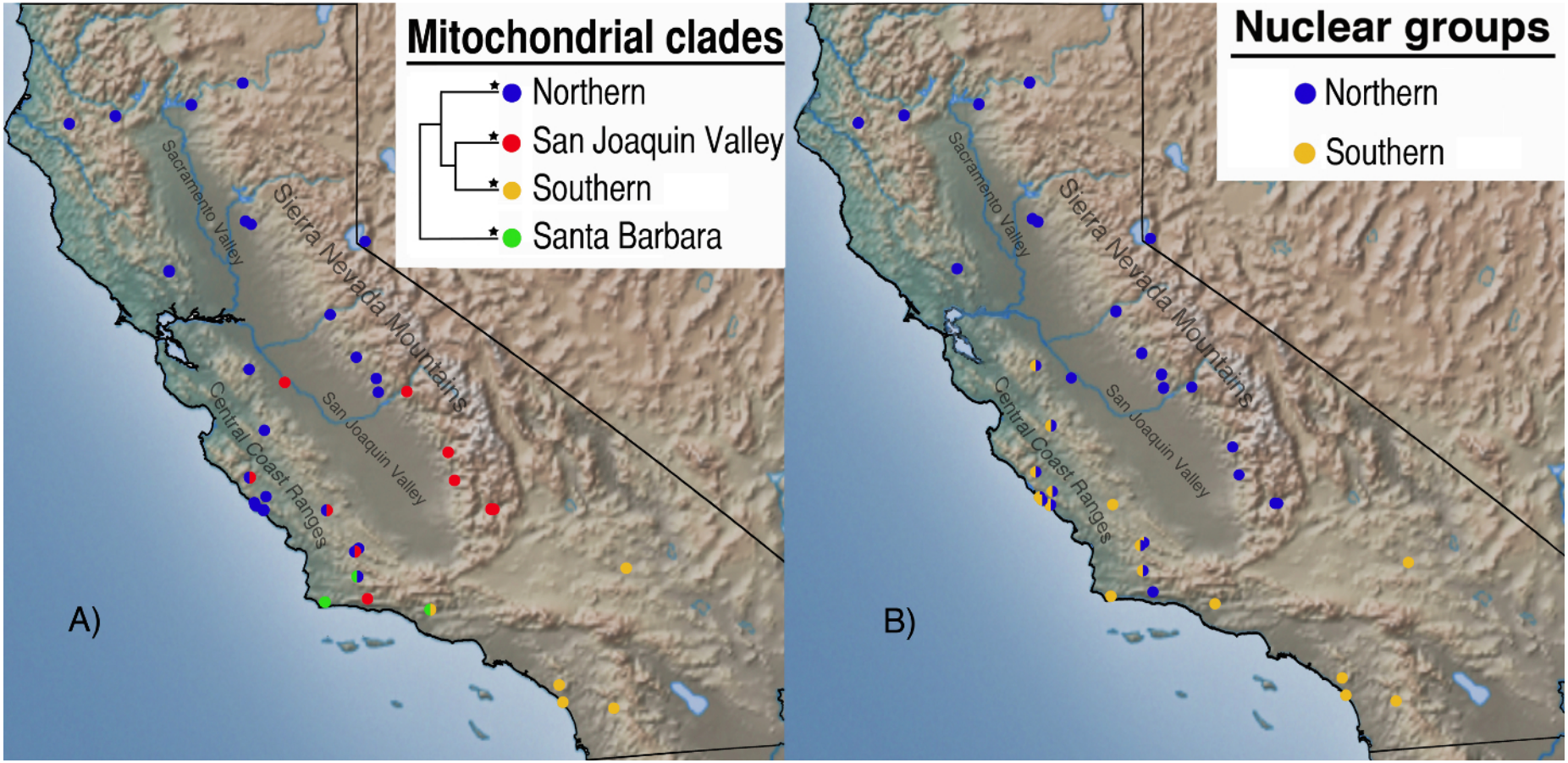
South of Santa Barbara *pallida*

San Joaquin Valley, Santa Barbara each distinct clades

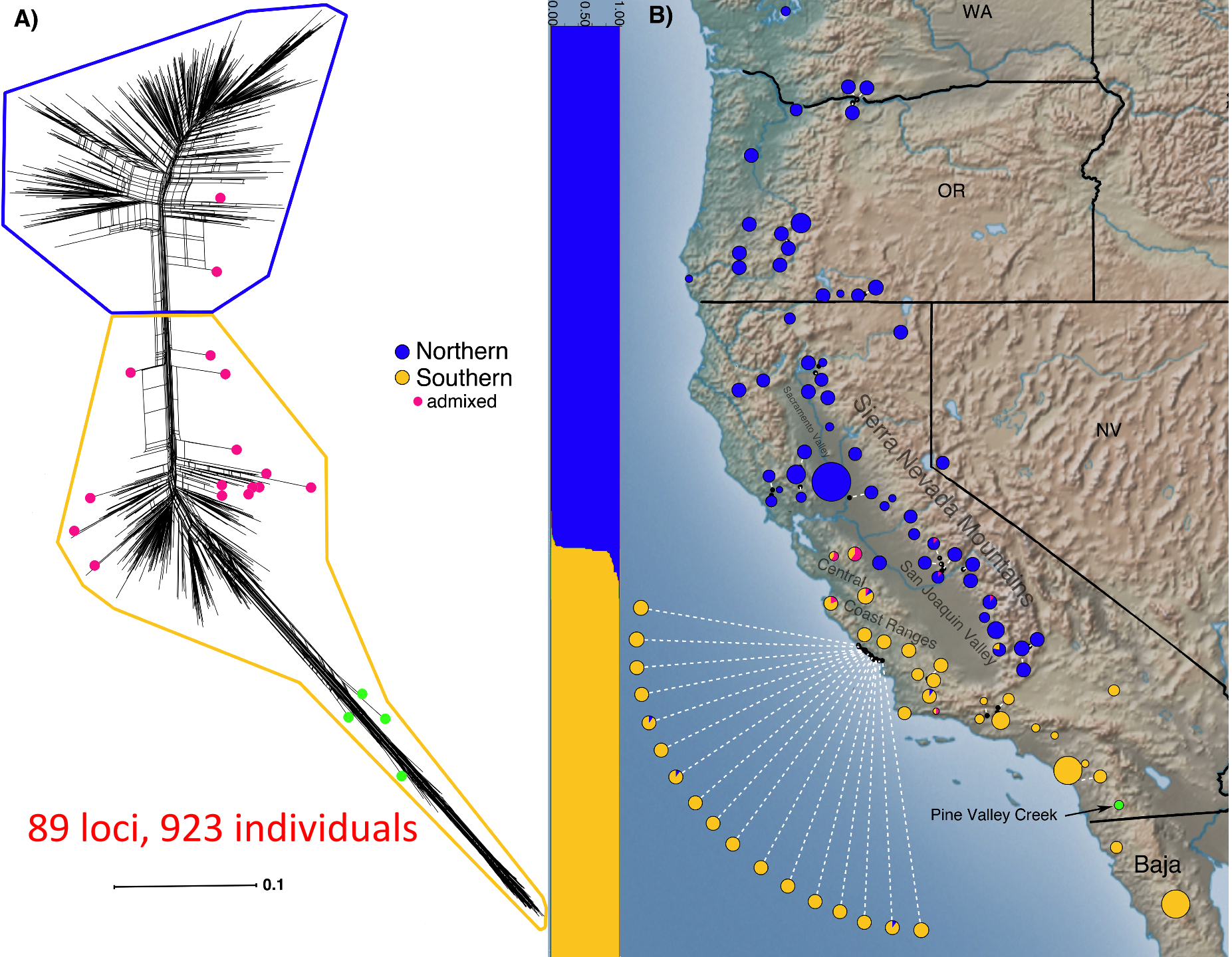
Central coast range admixed, but primarily *marmorata*

1000 bp mtDNA

5 loci nuDNA



Spinks et al.
2010


A)
Northern
Southern
admixed
89 loci, 923 individuals
0.1
B)
0.00
0.50
1.00
WA
OR
NV
Sacramento Valley
Sierra Nevada Mountains
Central
Coast Ranges
San Joaquin Valley
Pine Valley Creek
Baja

Will we learn more with 1000 markers?

For systematics, probably not.

For population differentiation, reintroductions, gene flow estimates in SoCal, probably.

The TIGER of vernal pools (*Ambystoma californiense*

Two species of tiger salamander in the Western US

Hybridization

Hybridization with an introduced species.

This is a huge conservation concern for California Tiger Salamander (CTS).

Its an equally huge opportunity for speciation studies.

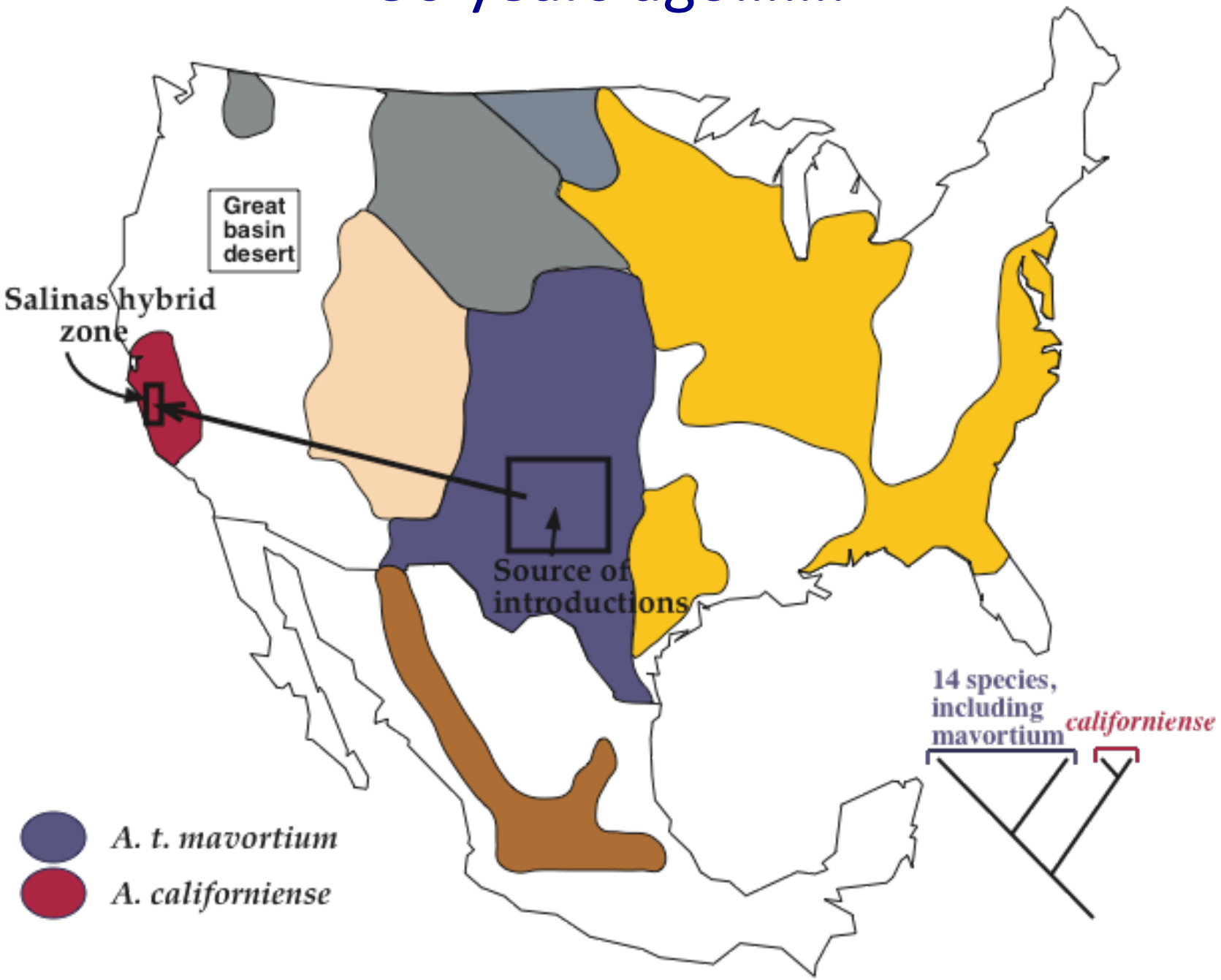
How we study hybrids with DNA

A A T T A G G T A C C G T

A A T T A G G C A C C G T

Species-specific marker (SNP)

50 years ago…….


Great basin desert
Salinas hybrid zone
Source of introductions
14 species, including mavortium
californiense
A. t. mavortium
A. californiense

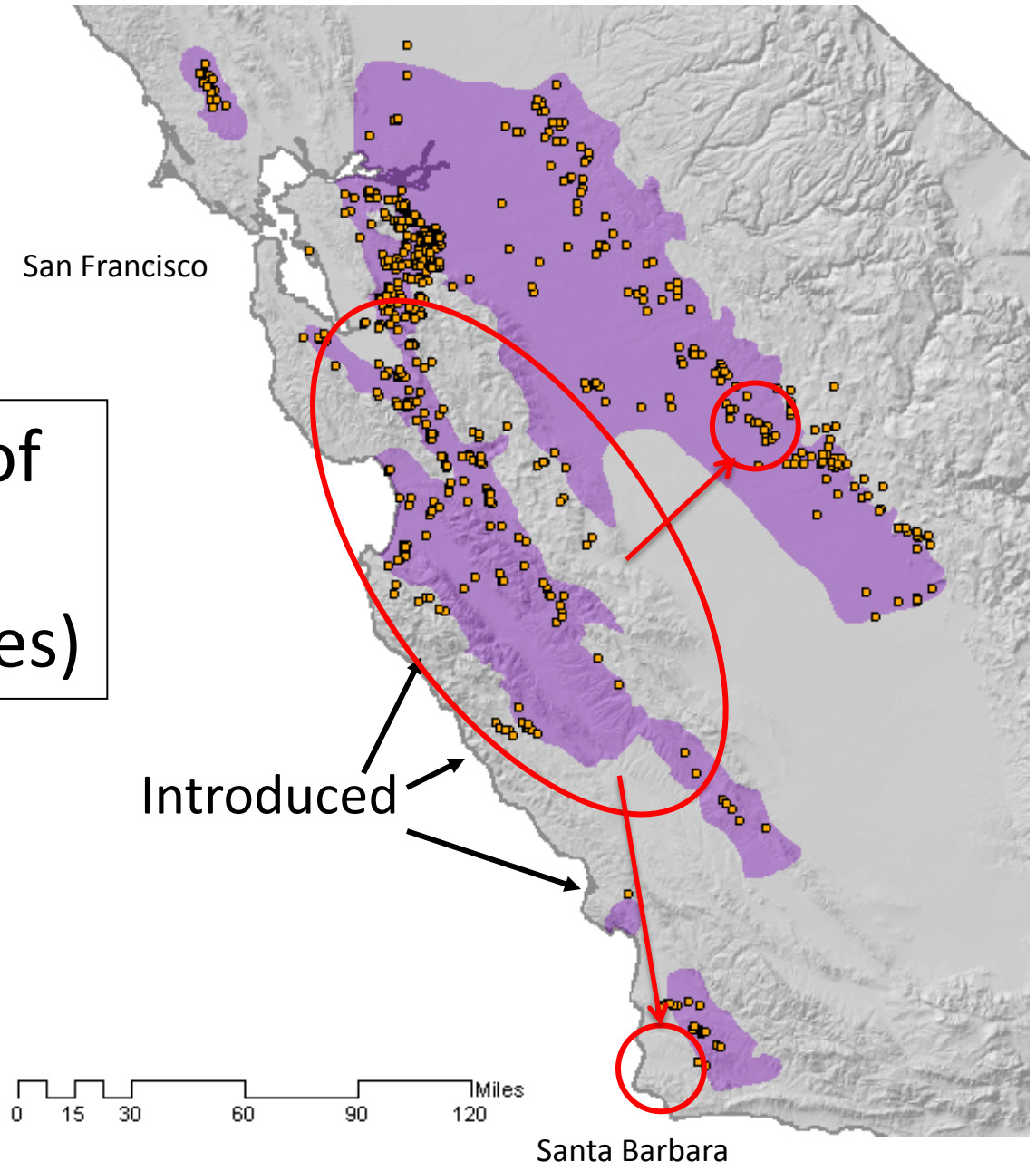
Waterdogs

"… 8- to 10-inch long salamanders are preferred by most trophy bass specialists."

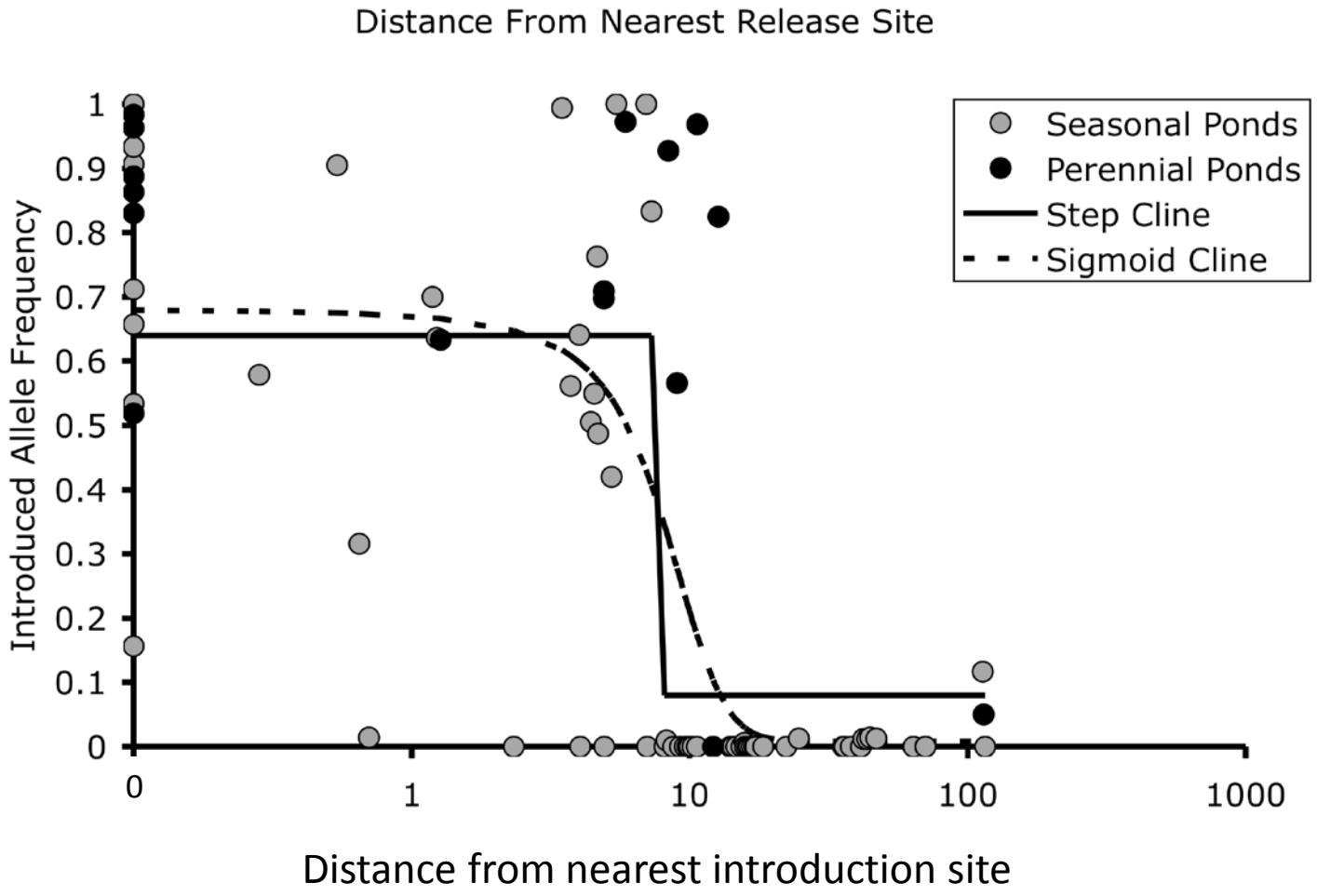
photo by Larry Larsen

Recent ~10-20 salamander generations old

Where are introduced genes?



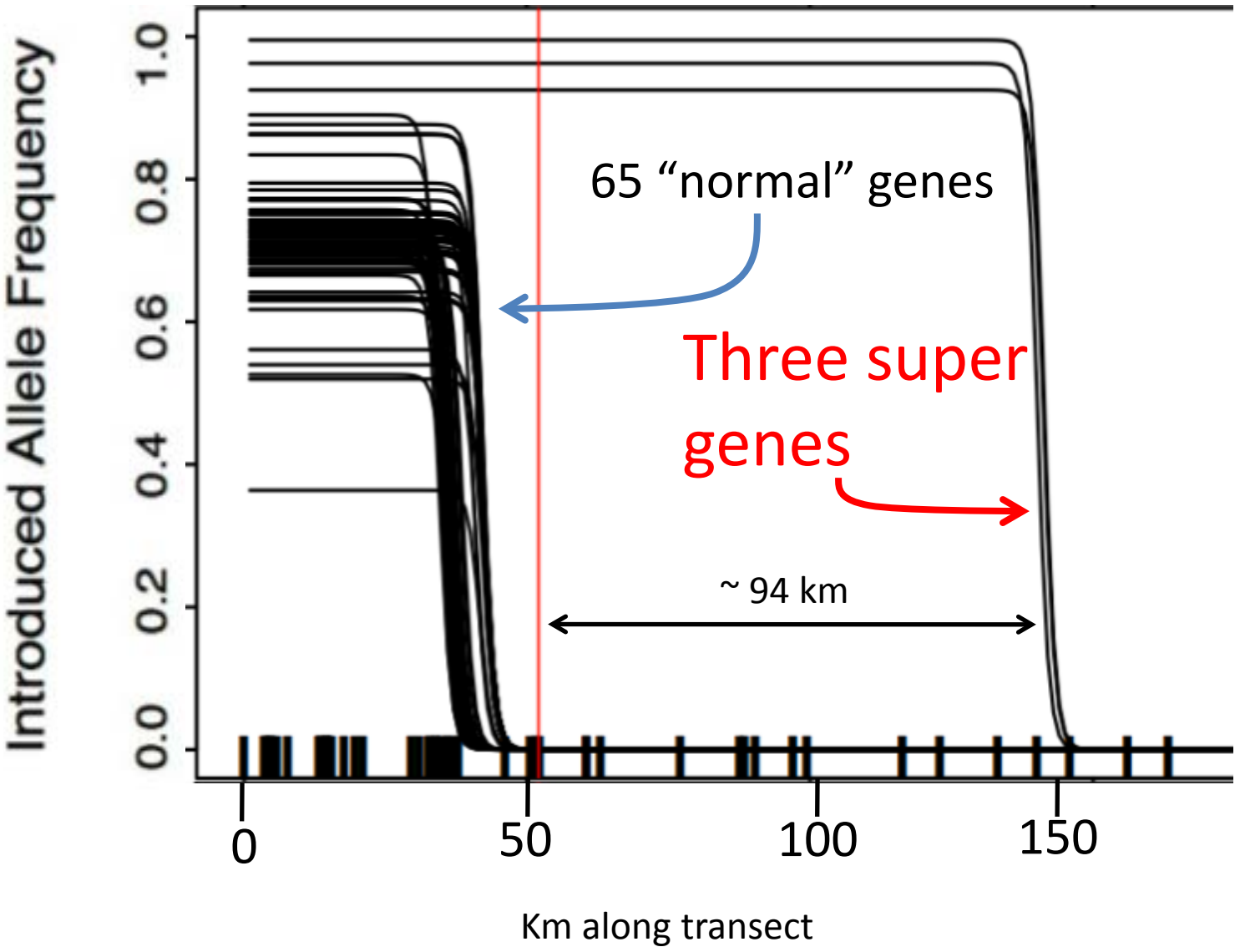
About 1/3 of the entire range (8 genes)



Introduction front moving ~ 1 km/generation (8-10 markers)

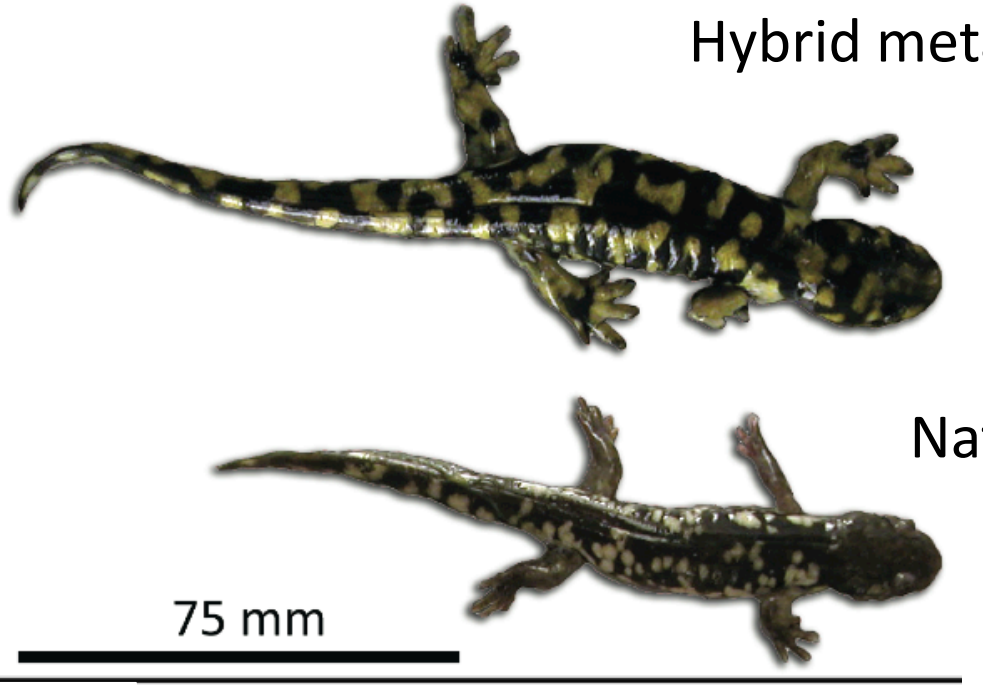
Fitzpatrick & Shaffer 2007 Ecological Applications

62 ponds, 68 genes

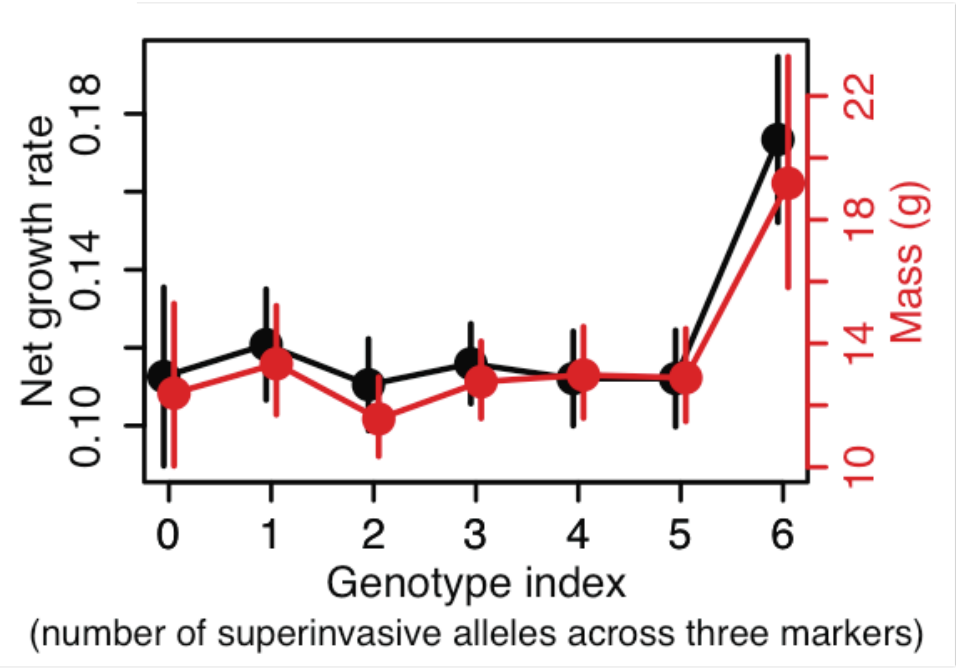


Fitzpatrick et al., 2010 PNAS

Hybrid metamorph

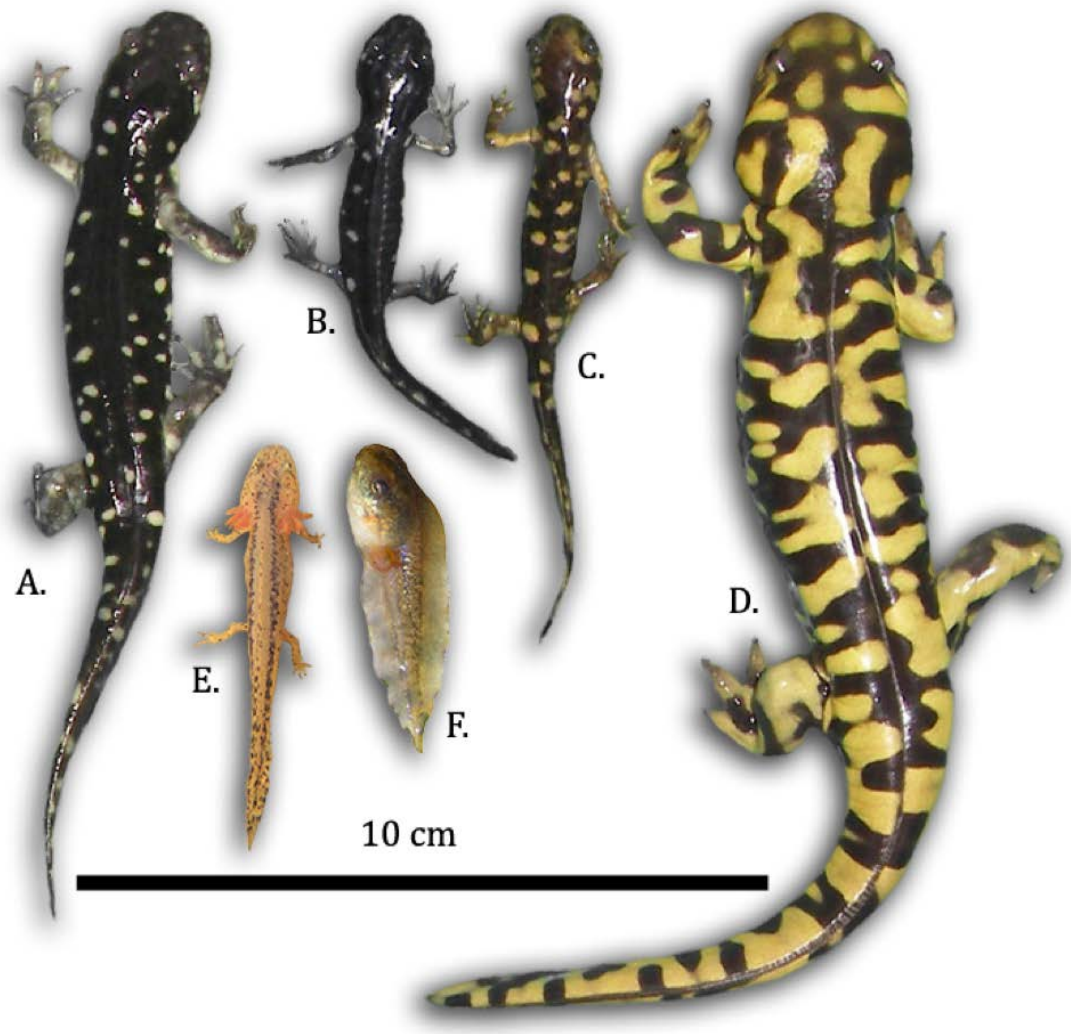


Native metamorph



Does it matter?

It TOTALLY matters



A: largest native B: smallest native C: smallest contemporary hybrid D: largest contemporary hybrid E: *Taricha torosa* larva E: *Pseudacris regilla* larva

Ryan et al. 2009 PNAS

Questions

1) Should hybrids be protected?

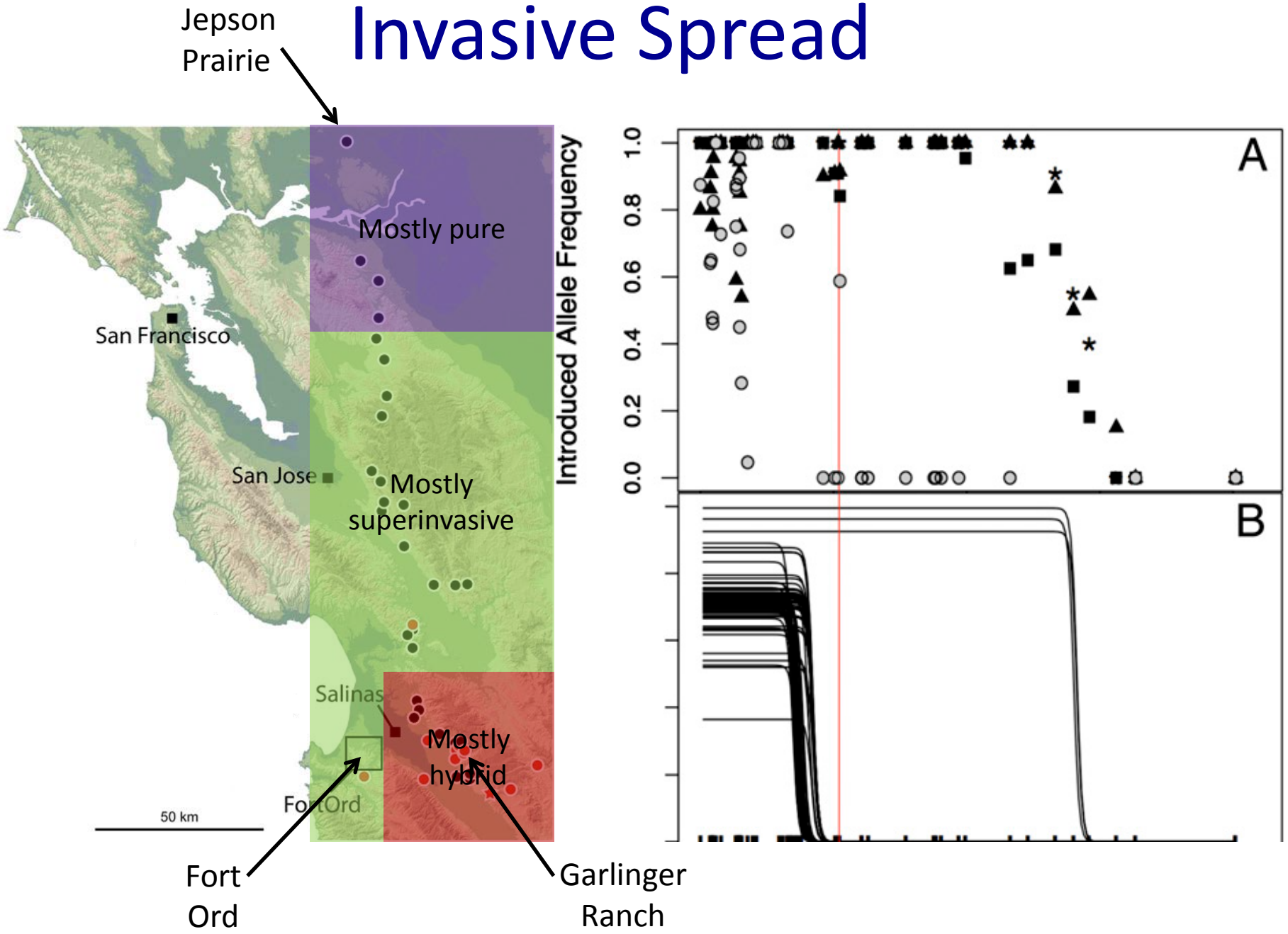
2) Does this answer change based on whether they are full hybrids or superinvasives?

Our Approach: Ecological Equivalency

- Mesocosm experiment (4 x 2 factorial)
- Treatments: larval genotype, larval density
- 4 levels of larval genotype:
  - pure CTS
  - superinvasive
  - full hybrid
  - no tiger salamander
- 2 levels of larval density:
  - 4 larvae per tank
  - 8 larvae per tank
- 5 replicates of each – 40 cattle tanks total

Searcy, C. A., H. B. Rollins, and H. B. Shaffer In prep. Ecological equivalency of endangered and invasive tiger salamanders.

Invasive Spread



Food Web

Tiger salamander larvae

Cyzicus

Gastropods

Pseudacris

Cladocera

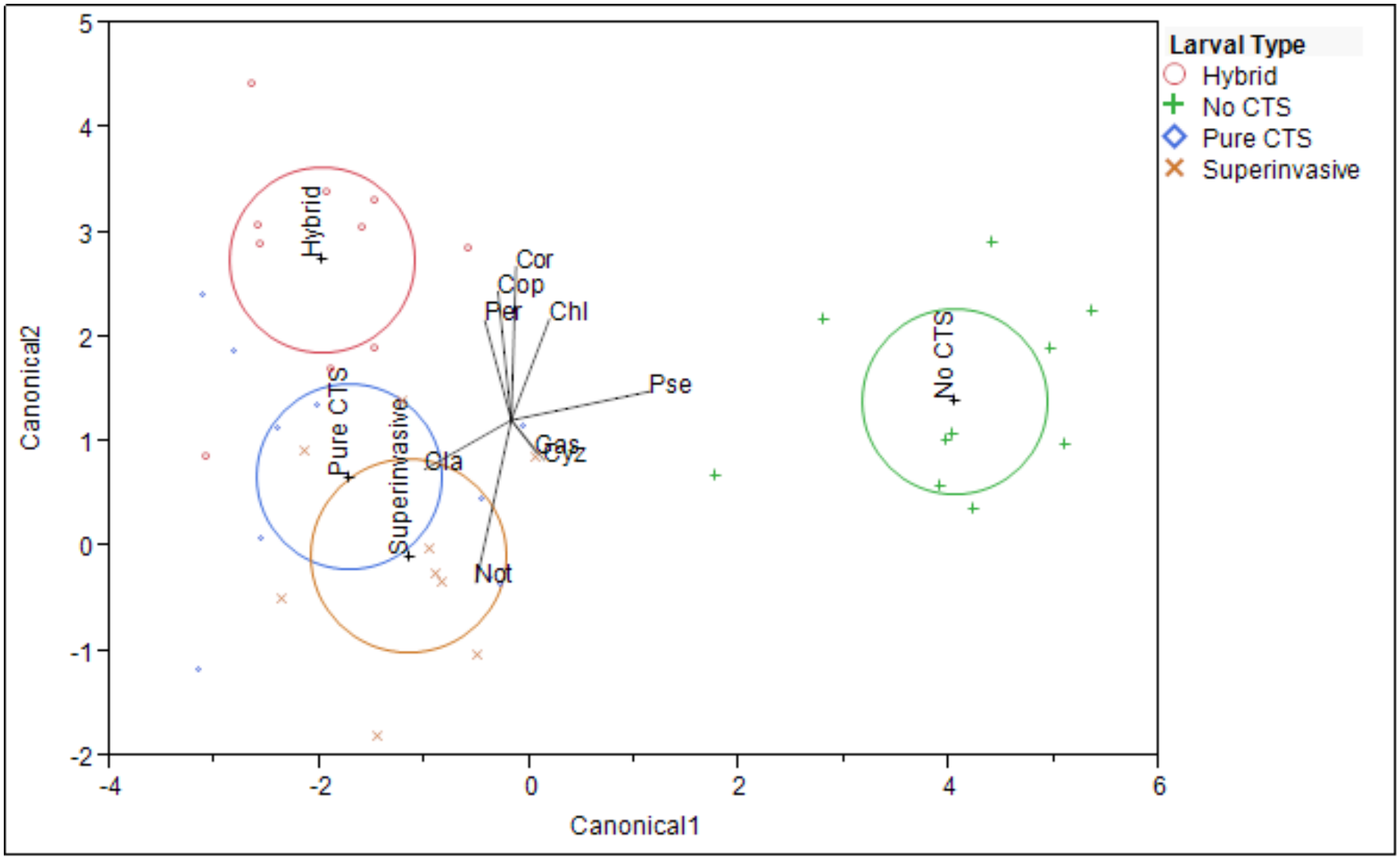
Periphyton

Community Metrics

Densities of:

1) Chlorophyll
2) Cladocera
3) Copepoda
4) Corixidae
5) *Cyzicus*
6) Gastropoda
7) Notonectidae
8) Periphyton
9) *Pseudacris*

Community Composition



Conclusions

1) We should conserve hybrids, because community composition is more similar to pure CTS tanks when hybrids are present than when there are no tiger salamanders at all.

2) We should manage habitat to decrease the percentage of invasive genes by decreasing hydroperiods.

3) We should not worry too much about superinvasives, because their community effects and life history strategy are not statistically distinguishable from pure CTS.

Desert tortoise landscape genomics

The question: How can we position solar panel farms and other developments to minimize impacts on tortoise movements?

The approach: Bradburd et al. *Evolution 2013 (*Bedassle)

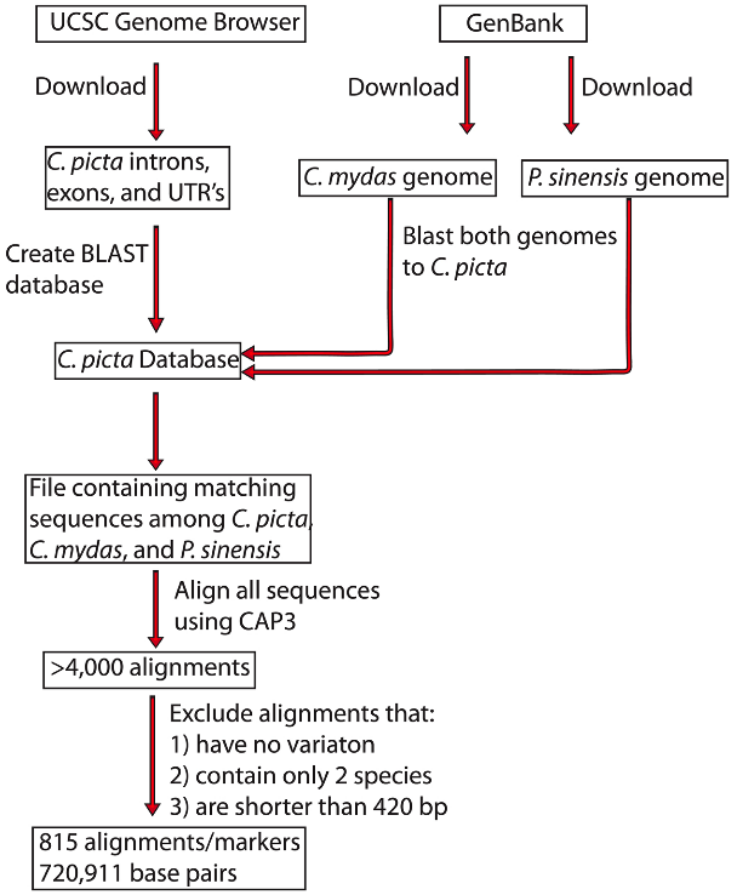
The data: Two strategies

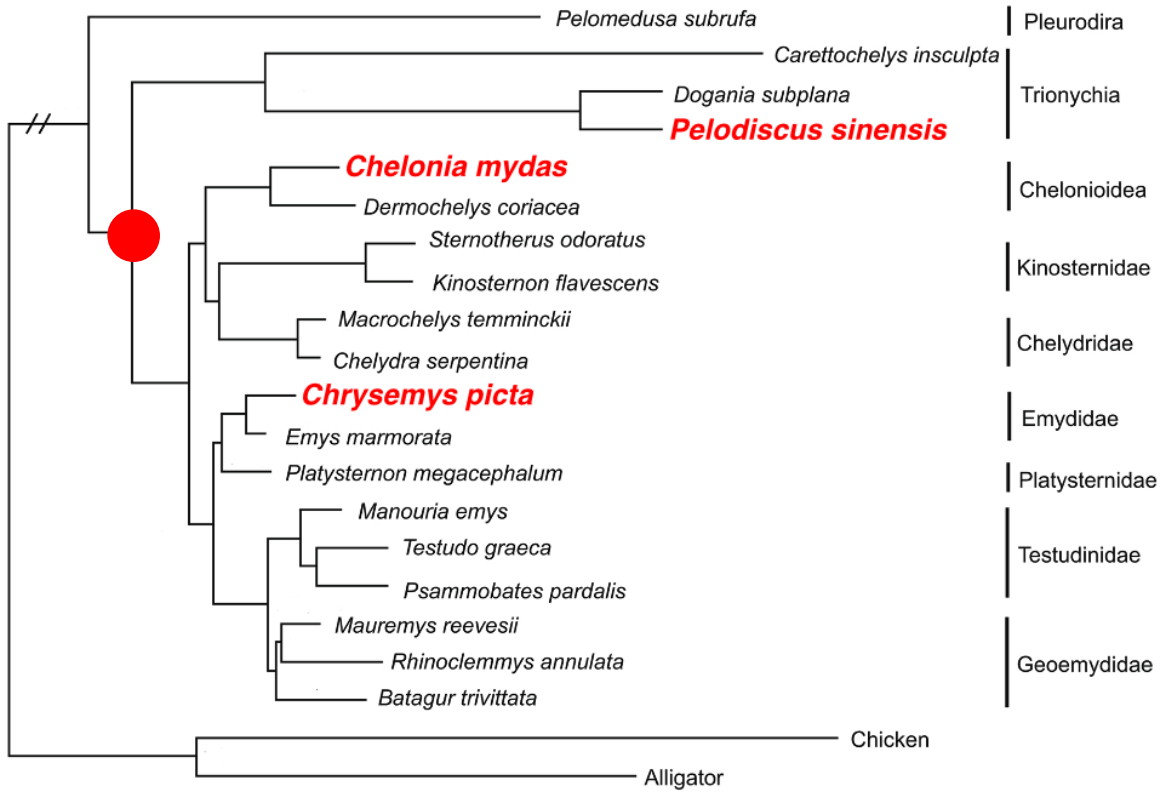
1) 800+ SNPs using baits

2) Whole genome, low coverage (0.5X) resequencing of 100-300 tortoises.

The result: Detailed habitat use/connectivity, with future projections for development scenarios (Peter Ralph, USC).

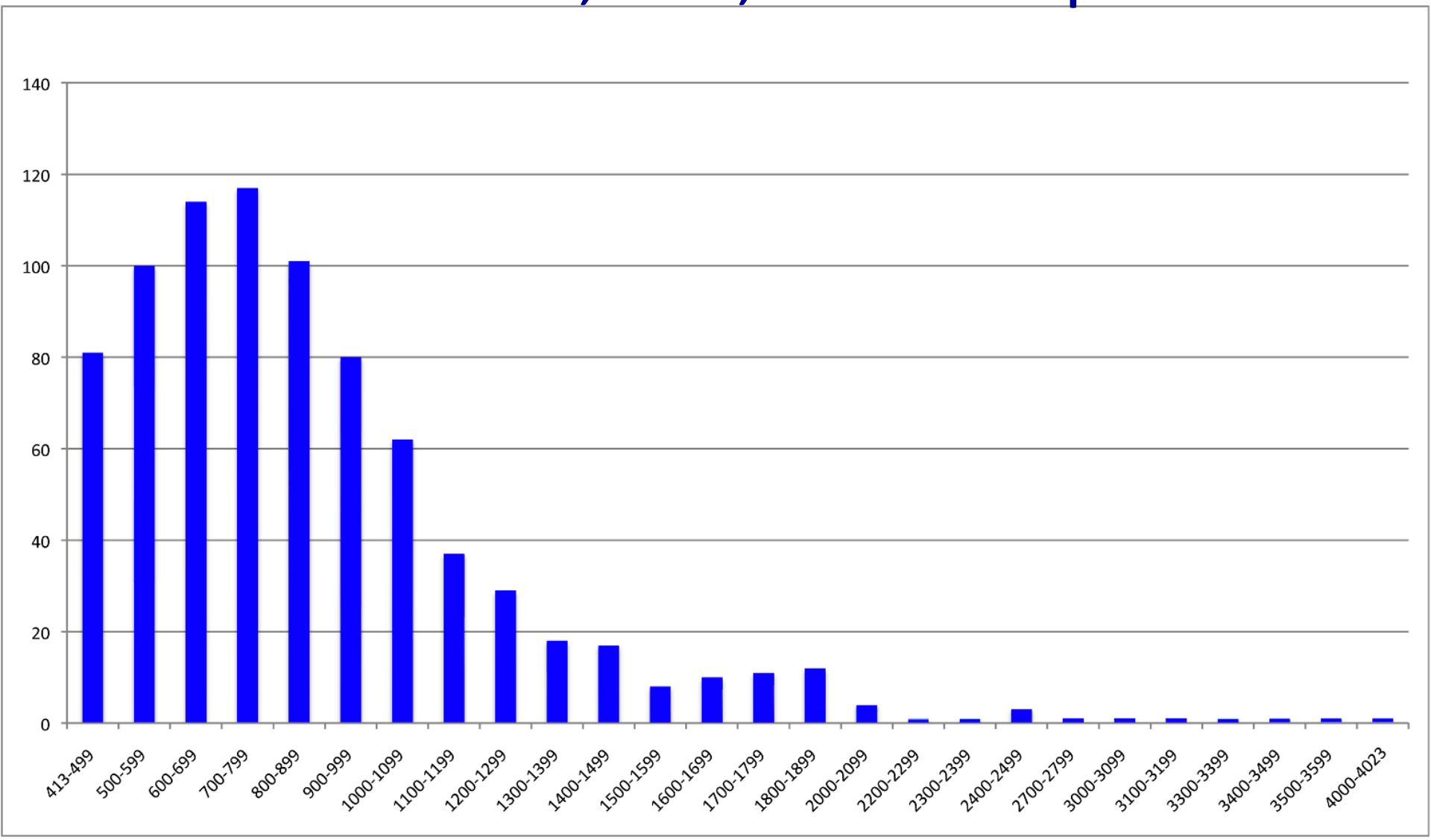
A Turtle-Specific Marker Set



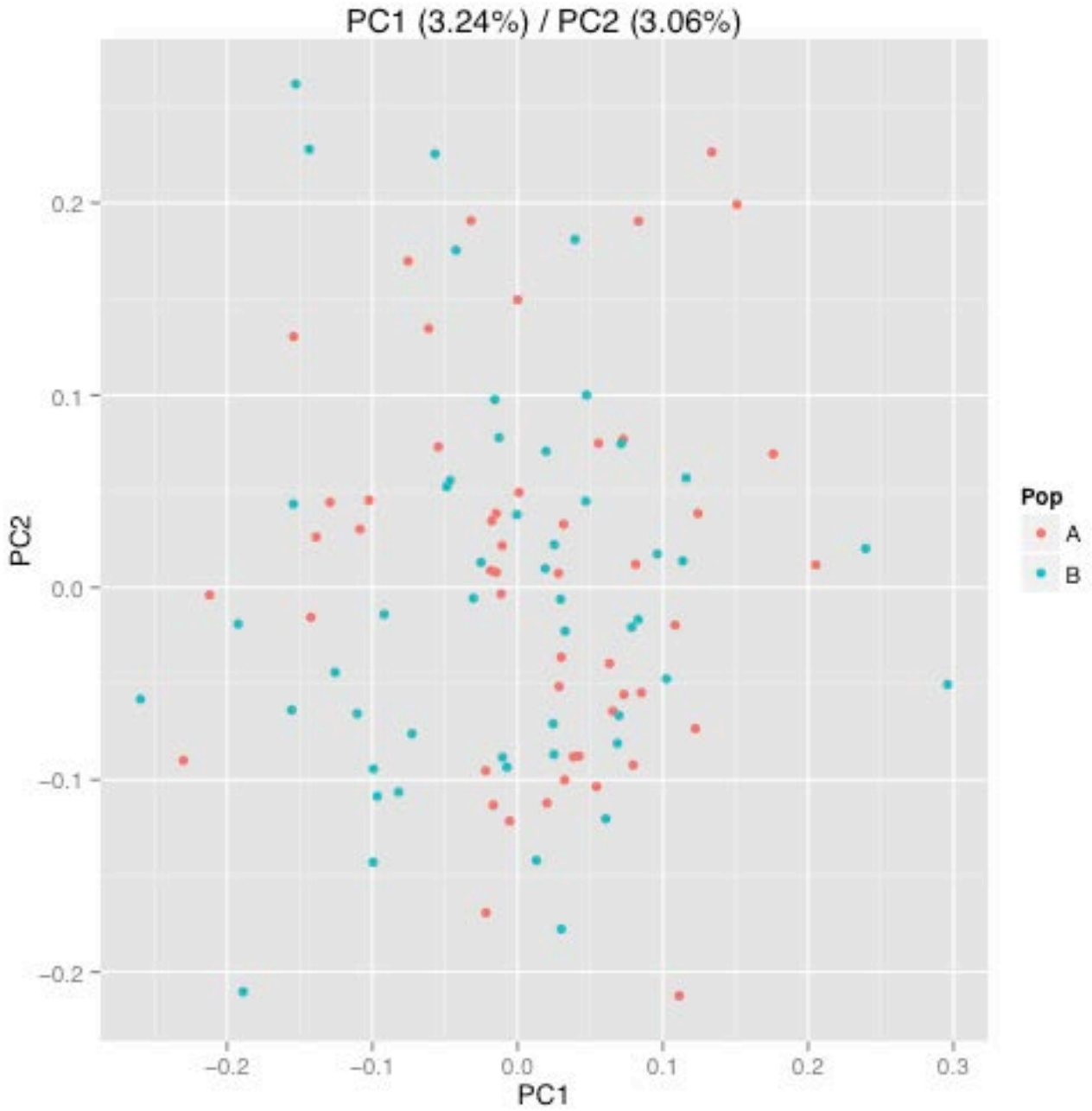


Pipeline developed by Phil Spinks @ UCLA

815 Genes, 750,000 base pairs



Roll out by 1 January 2014



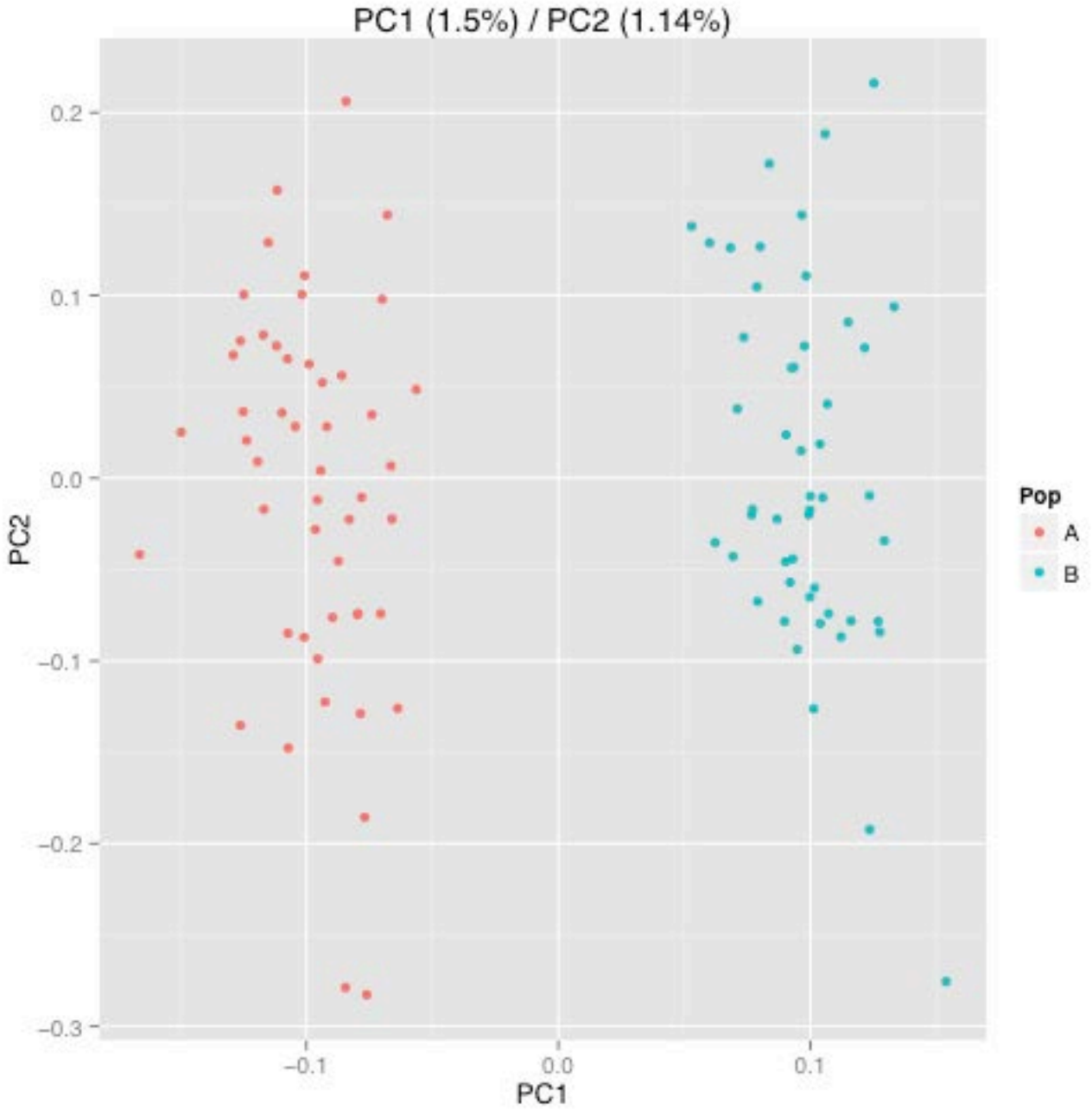
PCA plot

Simulated data

Fst = 0.01

100 SNPs

Complete overlap of populations



PCA plot

Simulated data

Fst = 0.01

10,000 SNPs

Complete separation of pops.

Re-sequencing should yield 100,000 – 1 M SNPs

Concluding thoughts

-We gain a lot in going even up to 100 markers, and lots more at 100,000

-For most vertebrates and plants, whole genome resequencing will be within reach in a year or two

-Very fine scale analyses of corridors, barriers, mating patterns within reach

-Ollie's point: tissues are limiting