

POPULATION GENOMICS OF THE MOJAVE DESERT TORTOISE

Peter Ralph

University of Oregon
Biology & Mathematics

Evolution!
Austin, June 18, 2016

Collaborators:

Evan McCartney-Melstad



Gideon Bradburd



Brad Shaffer



Roy Averill-Murray (US F&WS)

Erik Lundgren (USC)

Jannet Vu (UCLA)

Data: Fran Sandmeier, Chava Weitzman, Dick Tracy

Funding: US F&WS

NSF: ABI // Sloan Foundation

THE MOJAVE DESERT TORTOISE: *Gopherus agassizii*



THE MOJAVE DESERT TORTOISE:



THE MOJAVE DESERT TORTOISE:

World's Largest Solar Plant Opens

Gabriela Quirós, KQED Science | February 13, 2014 | 0 Comments

Share:



Print



The Ivanpah solar project in the Mojave Desert, the largest solar farm in the world. (Lauren Sommer/KQED)

The largest solar plant in the world officially starts generating electricity on Thursday. The Ivanpah solar farm, in California's Mojave Desert about 40 miles south of Las Vegas, will produce enough electricity to power 140,000 homes per year.

THE MOJAVE DESERT TORTOISE:

Endangered tortoises delay Mojave Desert solar plant

APRIL 28, 2011 | 12:18 PM



The Obama administration has halted the building of two-thirds of a massive solar project in San Bernardino's Mojave Desert as a new federal assessment found that more than 600 endangered desert tortoises would die as a result of construction.

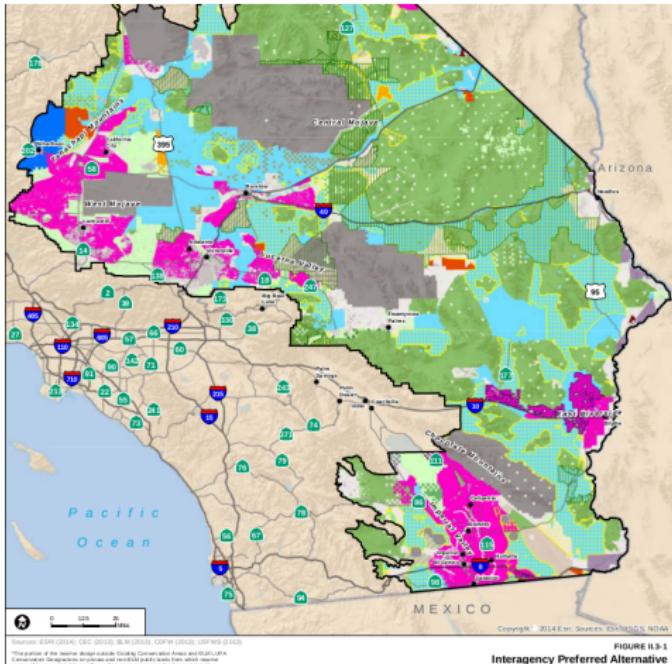
The U.S. Bureau of Land Management assessment this week disputed the estimate by BrightSource Energy, developer of the 392 MW solar thermal plant, that only 38 of the reptiles would be disturbed by construction at the 5.6-square mile Ivanpah Valley site near Primm, Nev. [corrected: an earlier version of this post said 5.6 acres]

Questions concerning the California tortoises highlight the friction between wilderness conservation and the quest for cleaner power. Many environmentalists contend it would be preferable to subsidize smaller solar arrays on commercial and residential rooftops, or on industrial acreage, than offer government loan guarantees to large complexes on wildlands that require transmission lines to transport the electricity to urban areas.

THE QUESTION(S)

How will changes
to the landscape
affect population
viability and gene
flow?

How do tortoises move
around on the
landscape?



(Hagerty, Nussear, Esque, & Tracy 2011)

DATA, PART I

Thanks to **lots of hard work**,
we have

- ▶ 83 GIS layers at 30m –
Jannet Vu (UCLA)
- ▶ tissue samples – *Dick Tracy, Chava Weitzman (UNR), Fran Sandmeier (U. Lindenwood)*
- ▶ 270 tortoises chosen for sequencing – *Evan McCartney-Melstad (UCLA)*

| layer name | Layer Category | Layer Description |
|---------------|----------------|-------------------------------------|
| imperv_30 | anthropogenic | percent impervious surfaces |
| road_30 | anthropogenic | euclidean distance to nearest road |
| agp_250 | biotic | annual growth potential |
| grass_herb_30 | biotic | grassland/herbaceous cover |
| shrub_30 | biotic | shrub cover |
| m2_08_precip | climate | avg. precip (Aug) |
| m2_ann_precip | climate | avg. annual precip |
| m2_anntmax | climate | avg. annual max temp |
| m2_anntmean | climate | avg. annual mean temp |
| m2_anntmin | climate | avg. annual min temp |
| win_precip | climate | avg. winter precip |
| avg_rough_30 | landscape | average surface roughness |
| aspect_30 | landscape | direction of slope face |
| barren_30 | landscape | percent barren land |
| dem_30 | landscape | elevation |
| eastness_30 | landscape | degree to which slope faces east |
| lat_gcs_30 | landscape | latitude |
| lon_gcs_30 | landscape | longitude |
| northness_30 | landscape | degree to which slope faces north |
| slope_30 | landscape | inclination of landscape in degrees |
| surfarea_30 | landscape | surface area of a grid cell |
| bd_ss2_st_30 | soils | bulk soil density |
| bdrck_ss2_st | soils | depth to bedrock |
| pr_ss2_st | soils | percent rocks |
| | | TOTAL |

DATA, PART I

Thanks to **lots of hard work**,
we have

- ▶ 83 GIS layers at 30m –
Jannet Vu (UCLA)
- ▶ tissue samples – *Dick Tracy, Chava Weitzman (UNR), Fran Sandmeier (U. Lindenwood)*
- ▶ 270 tortoises chosen for sequencing – *Evan McCartney-Melstad (UCLA)*

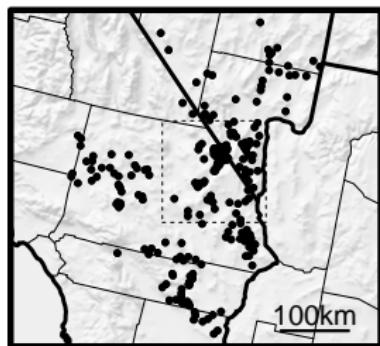


DATA, PART I

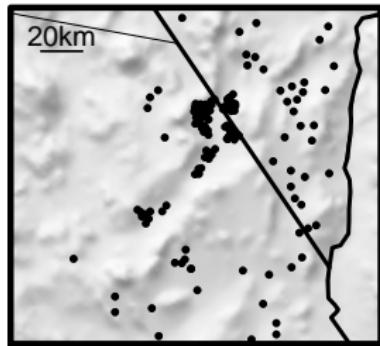
Thanks to **lots of hard work**, we have

- ▶ 83 GIS layers at 30m – *Jannet Vu (UCLA)*
- ▶ tissue samples – *Dick Tracy, Chava Weitzman (UNR), Fran Sandmeier (U. Lindenwood)*
- ▶ 270 tortoises chosen for sequencing – *Evan McCartney-Melstad (UCLA)*

Tortoise Sample Map



Ivanpah Valley



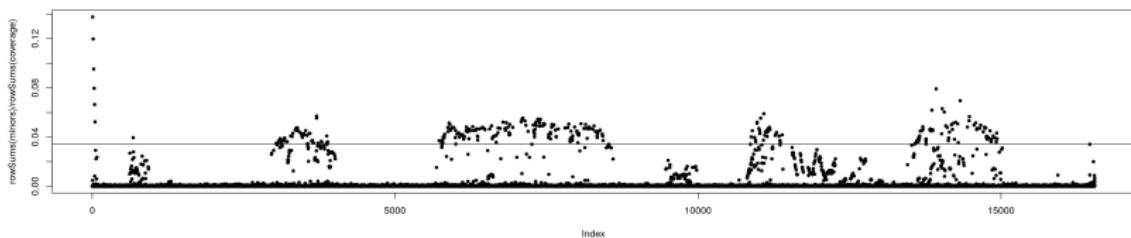
SEQUENCING & BIOINFORMATICS

Evan McCartney-Melstad (UCLA)

- ▶ PCR-free library preparation
- ▶ Illumina sequencing: $> 10^{12}$ bp
- ▶ read QC, trimming, and joining
- ▶ map to draft *G. morokkai* genome: $\approx 1.5 \times$ coverage (1.9 Gb, Kusumi et al, in prep)
- ▶ identify polymorphic **sites** with `angsd` ($\approx 3\%$ with $p < 10^{-6}$)
- ▶ read-based estimation of divergence (using $Q \geq 30$ bases)

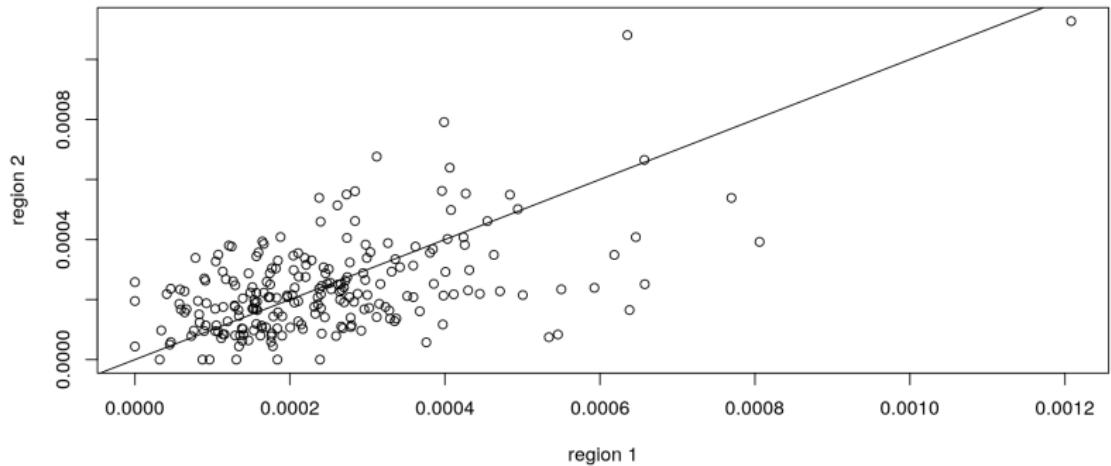
MITOCHONDRIA

- ▶ median coverage: $25\times$
- ▶ numts – regions with consistent “heterozygosity”
- ▶ excluding these, error rate of $(2.22 \pm 0.12) \times 10^{-4}$ consistent with other measures
- ▶ two haplotypes differing by 0.3%



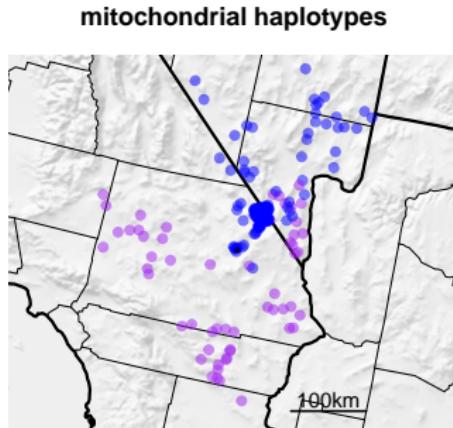
MITOCHONDRIA

- ▶ median coverage: $25\times$
- ▶ numts – regions with consistent “heterozygosity”
- ▶ excluding these, error rate of $(2.22 \pm 0.12) \times 10^{-4}$ consistent with other measures
- ▶ two haplotypes differing by 0.3%



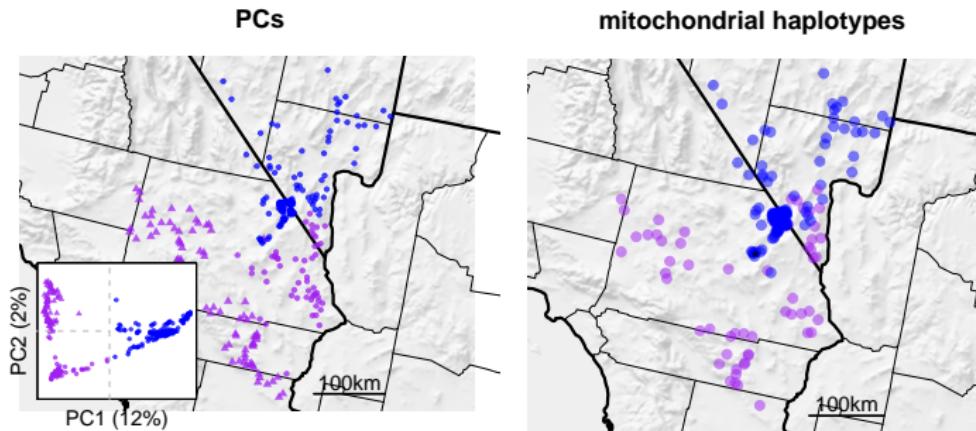
MITOCHONDRIA

- ▶ median coverage: $25\times$
- ▶ numts – regions with consistent “heterozygosity”
- ▶ excluding these, error rate of $(2.22 \pm 0.12) \times 10^{-4}$ consistent with other measures
- ▶ two haplotypes differing by 0.3%



AUTOSOMES: PCA

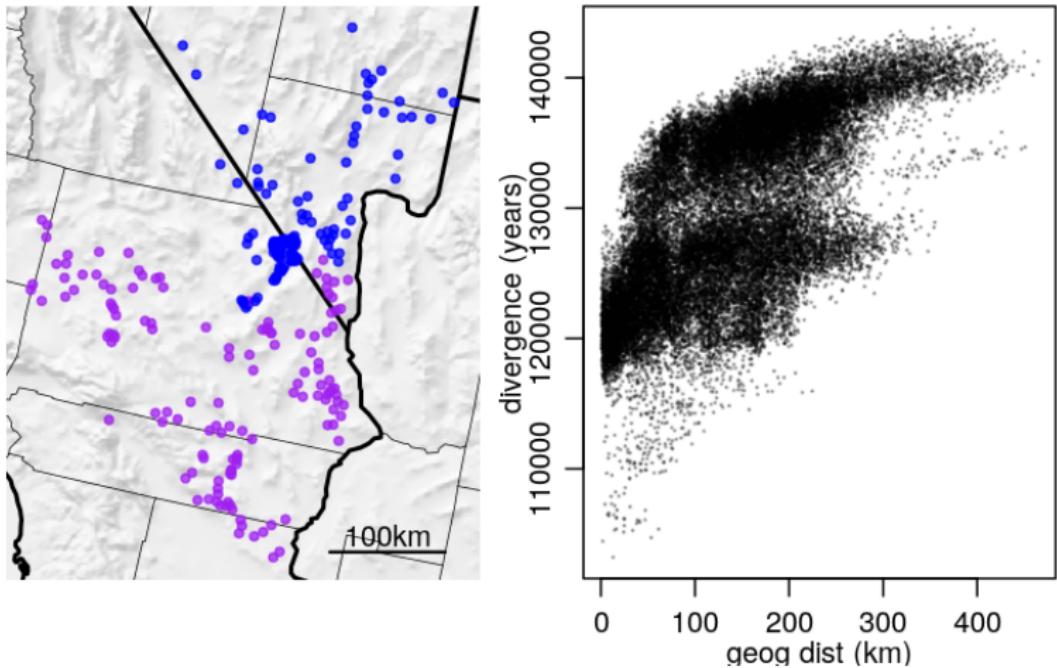
PCA from covariance matrix of **raw allele counts**



CLINAL CONTIGS?

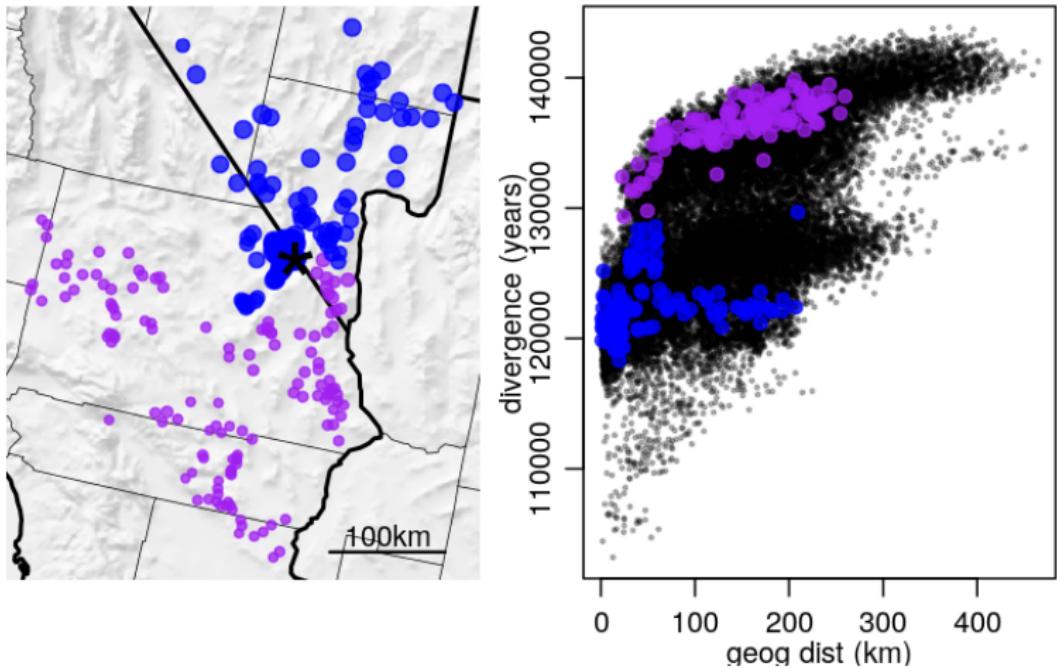
A few contigs show strong correlation with PC1:
XXX

SEQUENCE DIVERGENCE: AUTOSOMES



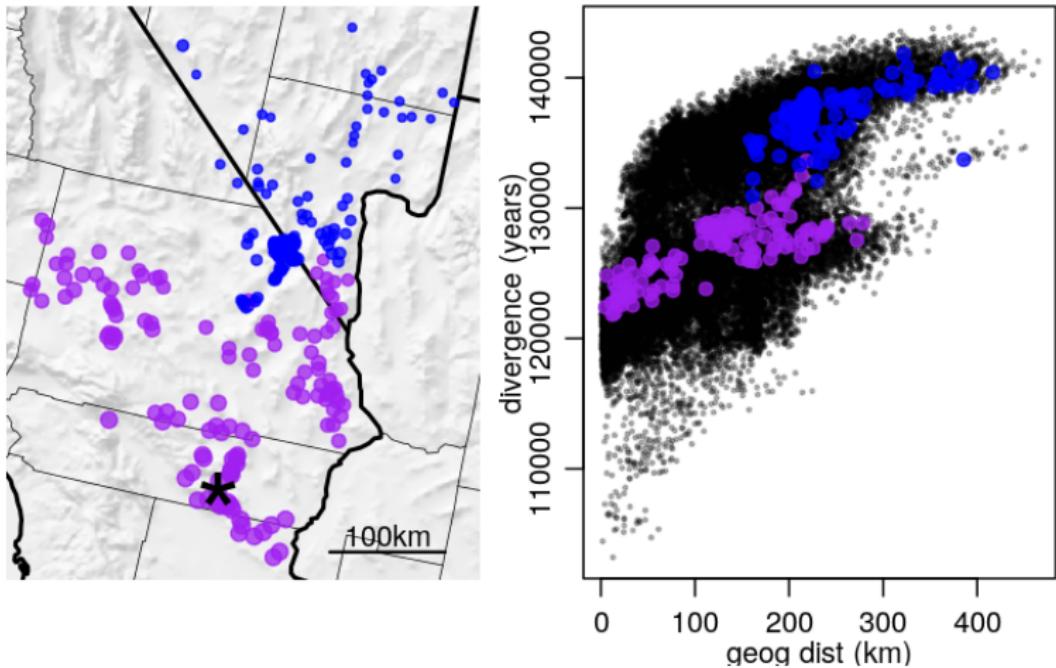
divergence = (mean density of nucleotide differences) $\approx 0.5\%$
= (years) \times (mutation rate)

SEQUENCE DIVERGENCE: AUTOSOMES



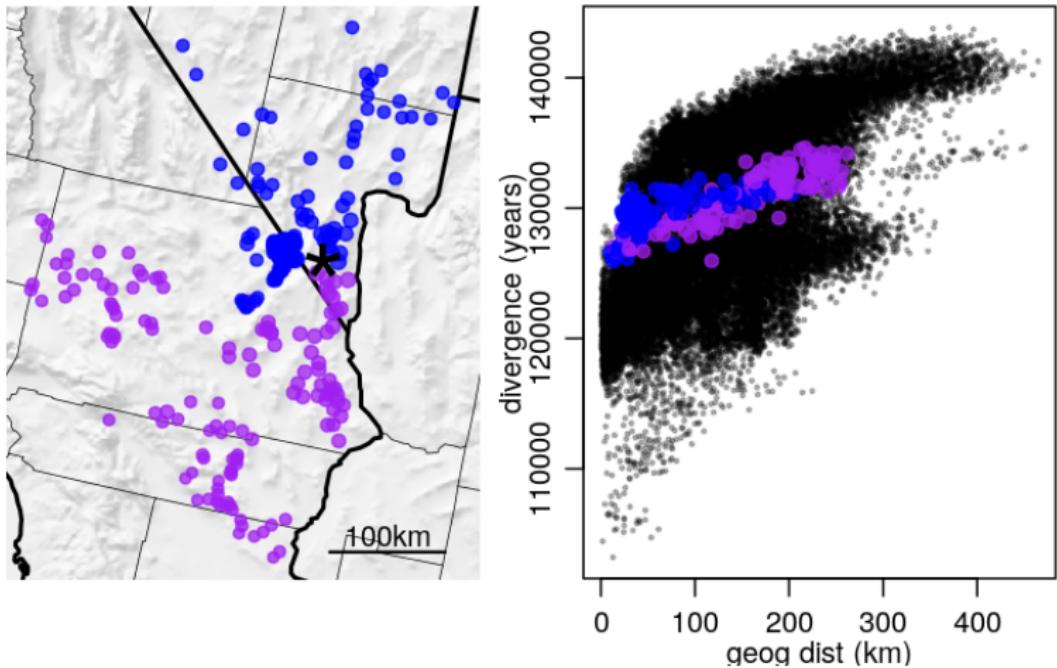
divergence = (mean density of nucleotide differences) $\approx 0.5\%$
= (years) \times (mutation rate)

SEQUENCE DIVERGENCE: AUTOSOMES



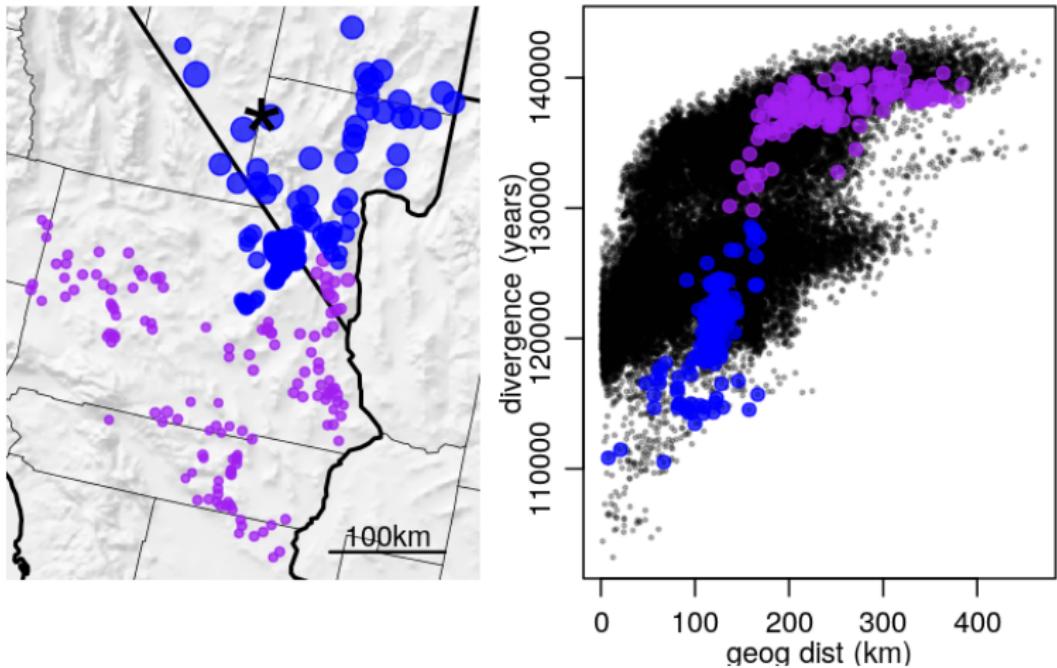
divergence = (mean density of nucleotide differences) $\approx 0.5\%$
= (years) \times (mutation rate)

SEQUENCE DIVERGENCE: AUTOSOMES



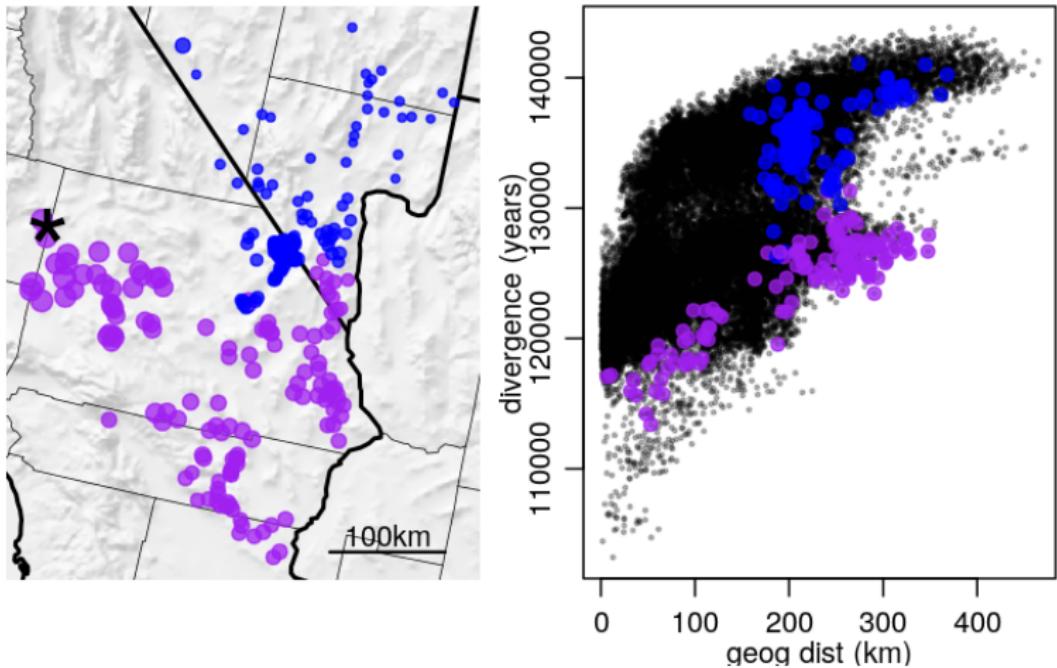
divergence = (mean density of nucleotide differences) $\approx 0.5\%$
= (years) \times (mutation rate)

SEQUENCE DIVERGENCE: AUTOSOMES



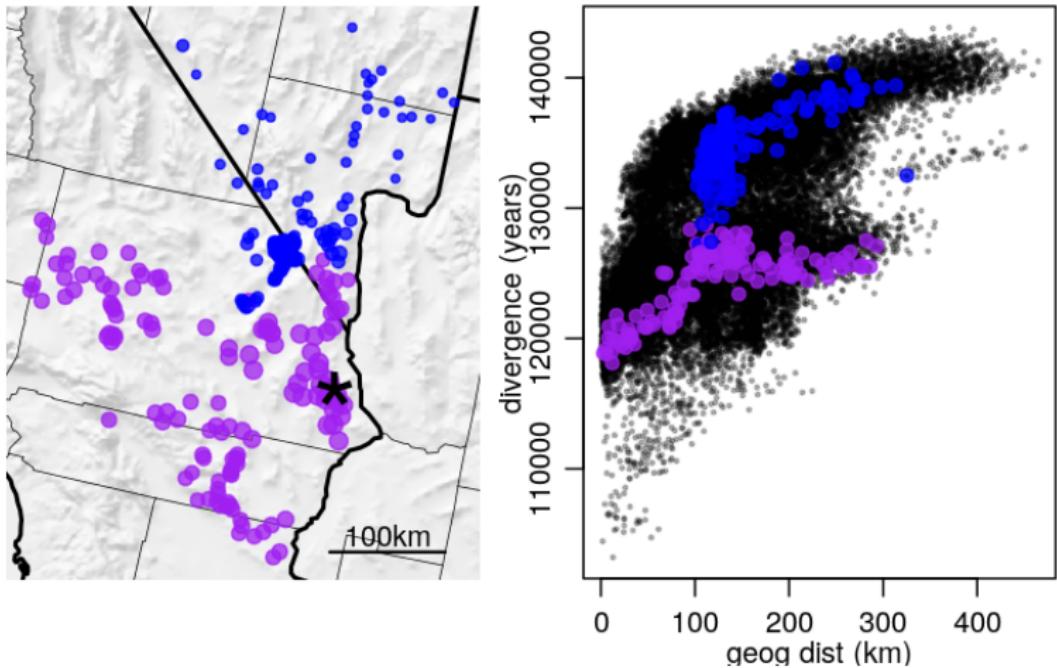
divergence = (mean density of nucleotide differences) $\approx 0.5\%$
= (years) \times (mutation rate)

SEQUENCE DIVERGENCE: AUTOSOMES



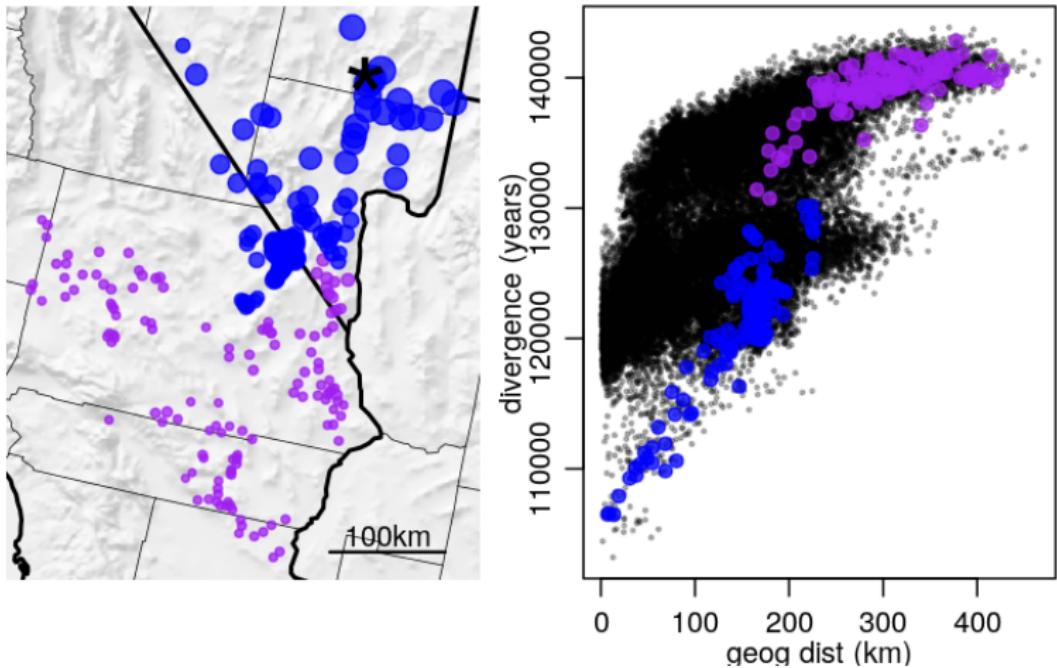
divergence = (mean density of nucleotide differences) $\approx 0.5\%$
= (years) \times (mutation rate)

SEQUENCE DIVERGENCE: AUTOSOMES



divergence = (mean density of nucleotide differences) $\approx 0.5\%$
= (years) \times (mutation rate)

SEQUENCE DIVERGENCE: AUTOSOMES



divergence = (mean density of nucleotide differences) $\approx 0.5\%$
= (years) \times (mutation rate)

FOLLOWING A LINEAGE



A **single lineage** back through time:

- ▶ jump rate is mean age of parent at birth
- ▶ tends to move towards regions that produce more offspring

MODEL FOR LINEAGE MOVEMENT

Recall that we have **landscape variables**, e.g.,

$$g_1(x) = (\text{elevation at } x)$$

$$g_2(x) = (\text{scrub cover at } x)$$

⋮

and define

jump rate at x : $u(x) = 1 / \left(1 + \exp \left(- \sum_{k=1}^n \alpha_k g_k(x) \right) \right)$

habitat quality at x : $\rho(x) = e^\gamma / \left(1 + \exp \left(- \sum_{k=1}^n \beta_k g_k(x) \right) \right)$

Then choose α , β , and γ so that

$$dX_t = \rho(X_t) \nabla u(X_t) dt + \sqrt{\rho(X_t) u(X_t)} dB_t$$

fits the data.

THIS IS AN INVERSE PROBLEM

An **inverse problem** in science is the process of calculating from a set of observations the causal factors that produced them: for example, calculating an image in computer tomography, source reconstructing in acoustics, or calculating the density of the Earth from measurements of its gravity field.

[Inverse problem - Wikipedia, the free encyclopedia](#)

https://en.wikipedia.org/wiki/Inverse_problem Wikipedia ▾

[About this result](#) • [Feedback](#)

THIS IS AN INVERSE PROBLEM

An **inverse problem** in science is the process of calculating from a set of observations the causal factors that produced them: for example, calculating an image in computer tomography, source reconstructing in acoustics, or calculating the density of the Earth from measurements of its gravity field.

[Inverse problem - Wikipedia, the free encyclopedia](#)
https://en.wikipedia.org/wiki/Inverse_problem Wikipedia ▾

[About this result](#) • [Feedback](#)

cue ominous music

Usually, the forwards problem:

history \mapsto what data?

is easy, but the inverse problem:

data \leftarrow what history?

is hard

because many possible histories
map to (nearly) the same data.

Usually, the forwards problem:

history \mapsto what data?

is easy, but the inverse problem:

data \leftarrow what history?

is hard

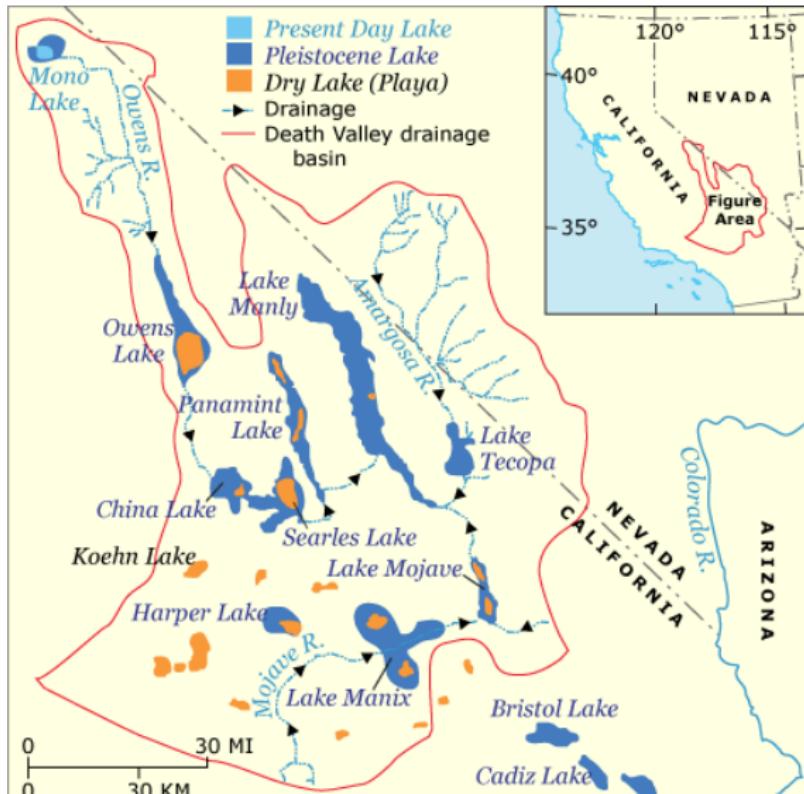
because many possible histories
map to (nearly) the same data.

SLOW AND STEADY

Best-fit model uses only the shape of the range (**flat!**),
but better fit should be possible.

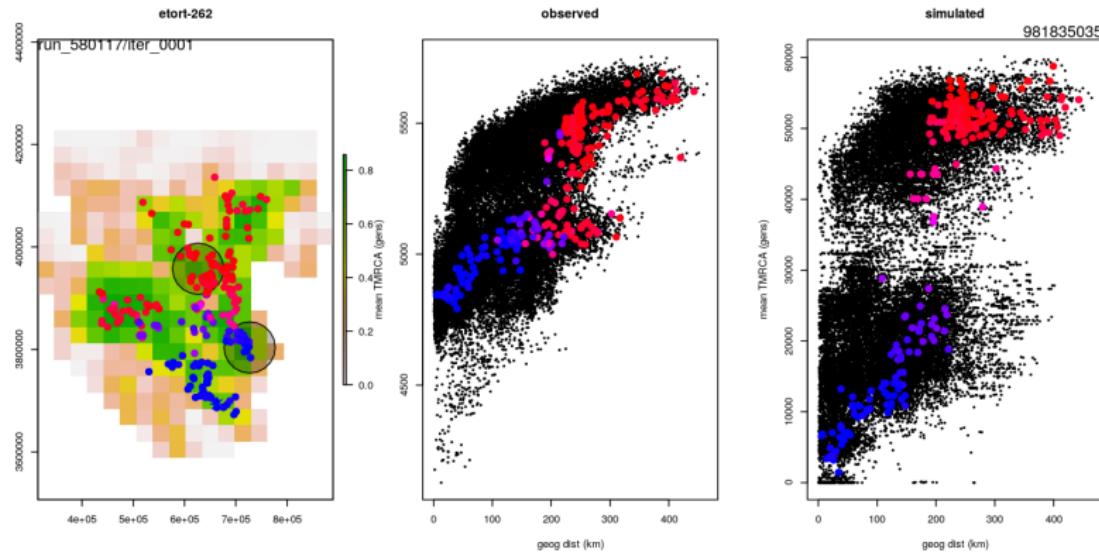
SLOW AND STEADY

Best-fit model uses only the shape of the range (flat!),
but better fit should be possible.



SLOW AND STEADY

Best-fit model uses only the shape of the range (**flat!**),
but better fit should be possible.



WHAT'S NEXT

Population Viability Analysis: (Jaime Ashander)

- ▶ landscape-scale individual-based simulation
- ▶ inform with the literature
- ▶ **but:** juvenile dispersal almost totally unknown
- ▶ fit to genetic data (divergence, local frequency spectra, haplotypes)

Thanks!

Evan McCartney-Melstad



Gideon Bradburd



Brad Shaffer



Erik Lundgren (USC)

Jannet Vu (UCLA)

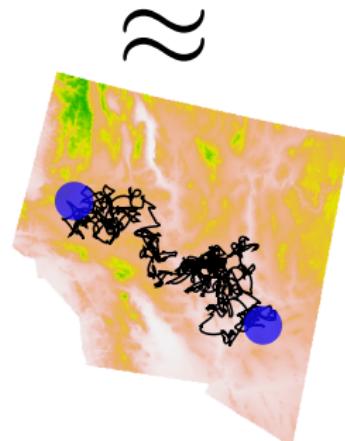
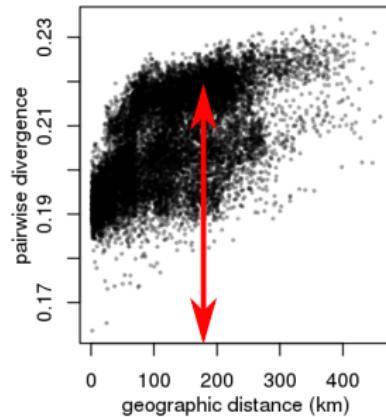
Data: Fran Sandmeier, Chava Weitzman, Dick Tracy

Funding: US F&WS

NSF: ABI // Sloan Foundation

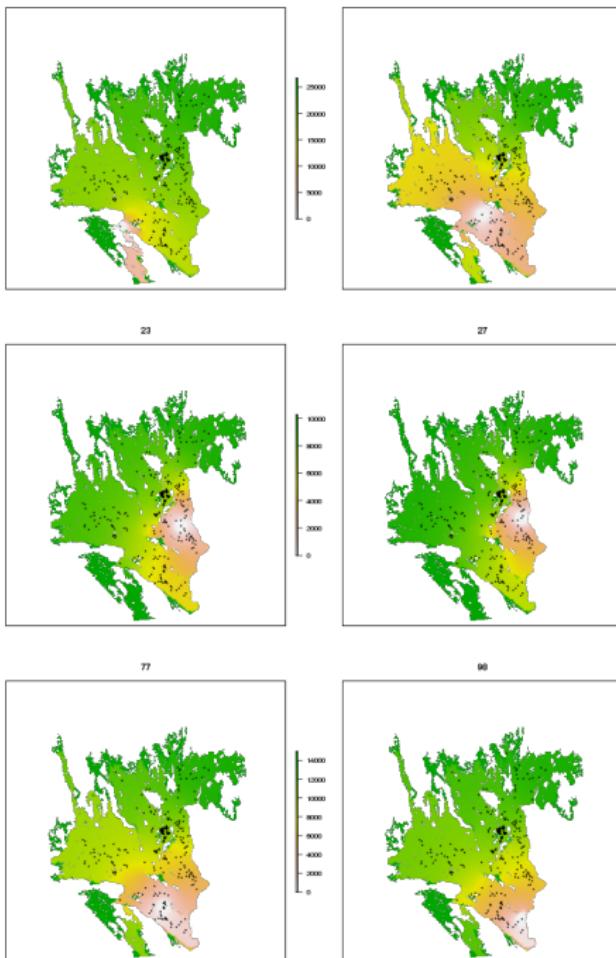
THE REST OF THE METHOD

- ▶ Sequence divergence
≈ mean hitting times of a lineage
- ▶ fit parameters by solving PDE
- ▶ Results: quantitative comparison of different development scenarios.



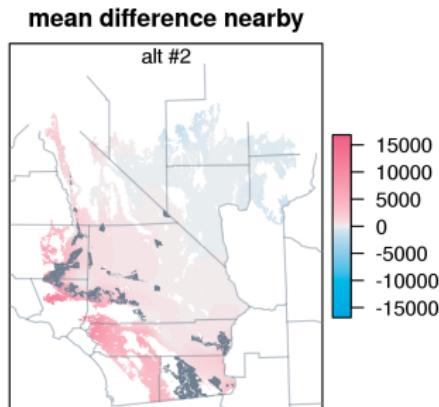
THE REST OF THE METHOD

- ▶ Sequence divergence
≈ mean hitting times of a lineage
- ▶ fit parameters by solving PDE
- ▶ Results: quantitative comparison of different development scenarios.

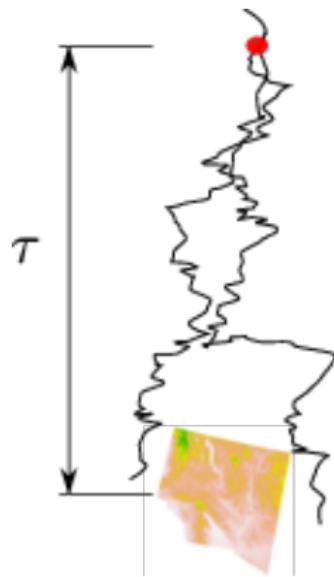


THE REST OF THE METHOD

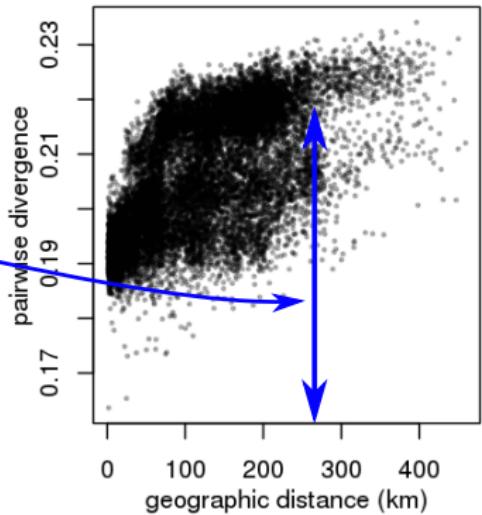
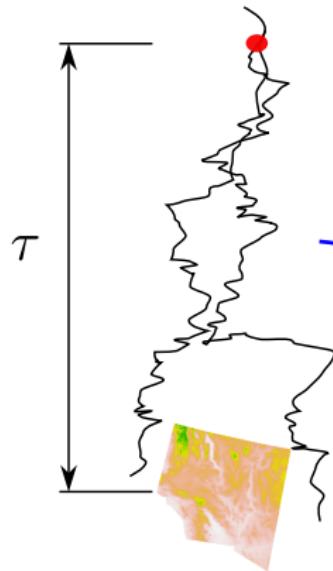
- ▶ Sequence divergence
≈ mean hitting times of a lineage
- ▶ fit parameters by solving PDE
- ▶ **Results:** quantitative comparison of different development scenarios.



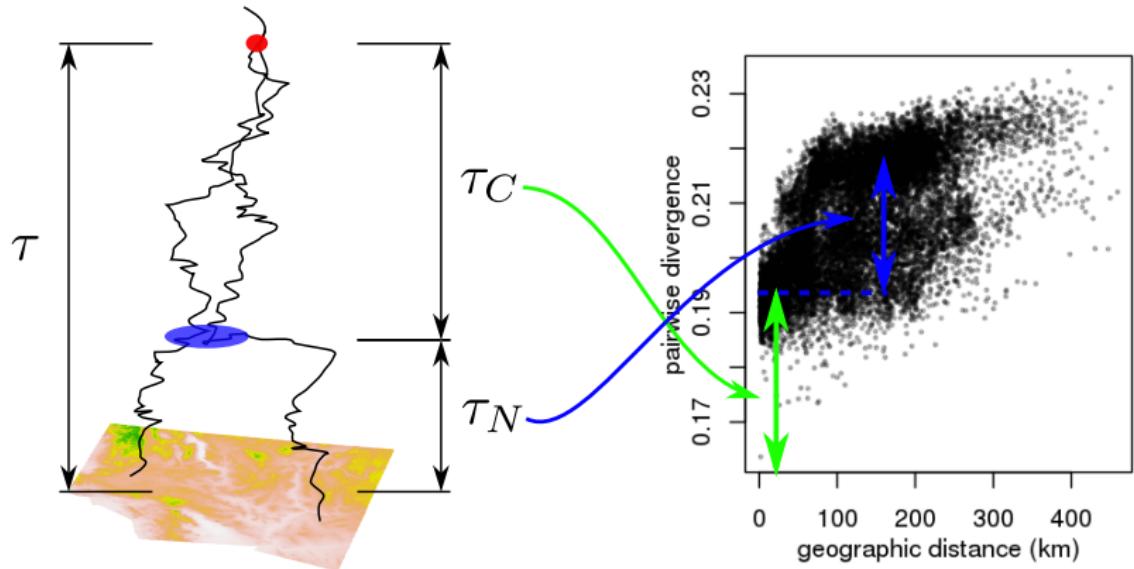
PAIRWISE DIVERGENCE AND HITTING TIMES



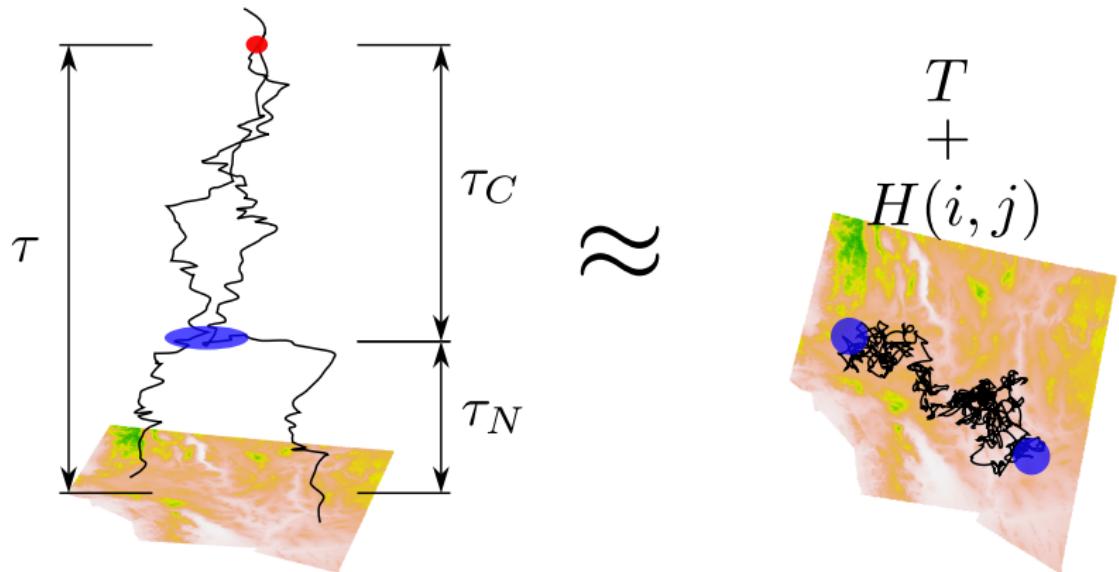
PAIRWISE DIVERGENCE AND HITTING TIMES



PAIRWISE DIVERGENCE AND HITTING TIMES



A MORE TRACTABLE PROBLEM

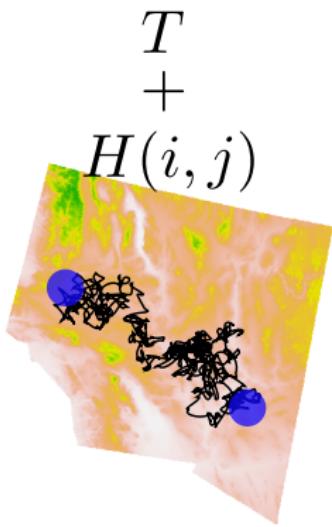


Replace τ_N by commute time
(a.k.a. **resistance distance**, McRae et al)

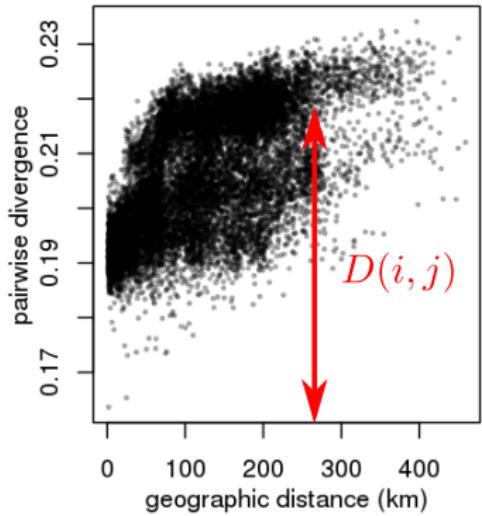
$$N(i, j) = \mathbb{E}[(\text{time to get near } j \text{ started from } i)]$$

$$H(i, j) \approx \frac{N(i, j) + N(j, i)}{2}.$$

A MORE TRACTABLE PROBLEM



??



Find parameters α, β, γ, T to minimize

$$\sum_{ij} |D(i, j) - T - H(i, j)|^2$$

FITTING HITTING TIMES

Parameters α, β, γ determine u and ρ :

$$Gf(x) := \rho(x) \nabla \cdot (u(x) \nabla f(x)).$$

Then

$$h_A(x) := \mathbb{E}[\text{ time for } X_t \text{ to hit } A \text{ from } x],$$

solves

$$Gh_A(x) = -1 \quad \text{for } x \notin A$$

$$h_A(x) = 0 \quad \text{for } x \in A$$

Keywords: multigrid methods for elliptic PDE.

... and we can get derivatives by solving the same sort of equation:

$$G(\partial_\alpha h_A(x)) = -(\partial_\alpha G)h_A(x)$$

$$G(\partial_\alpha^2 h_A(x)) = -(\partial_\alpha^2 G)h_A(x) - 2(\partial_\alpha G)(\partial_\alpha h_A(x))$$

... and use a **trust region algorithm** to optimize.

