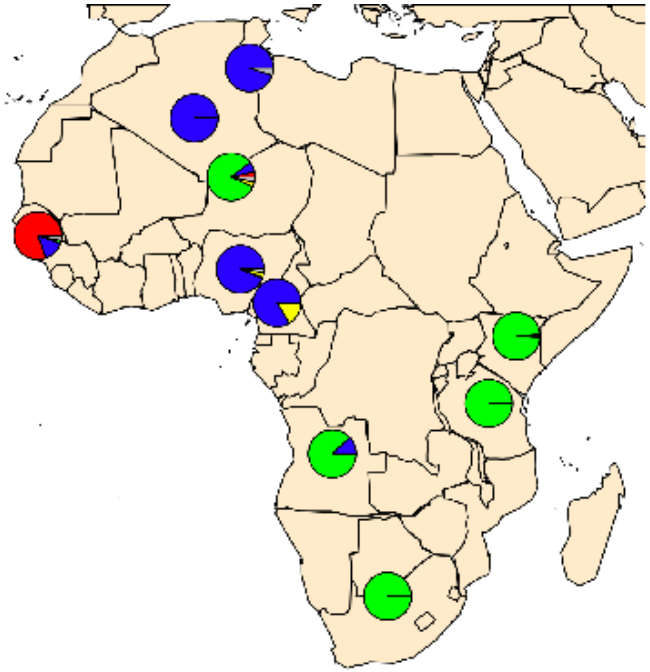
Tools for Understanding the Geography of Adaptation

Peter Ralph

USC Computational Biology and Bioinformatics

Simons Institute
May 1st, 2014
work with Graham Coop

MOTIVATION: HEMOGLOBINOPATHIES

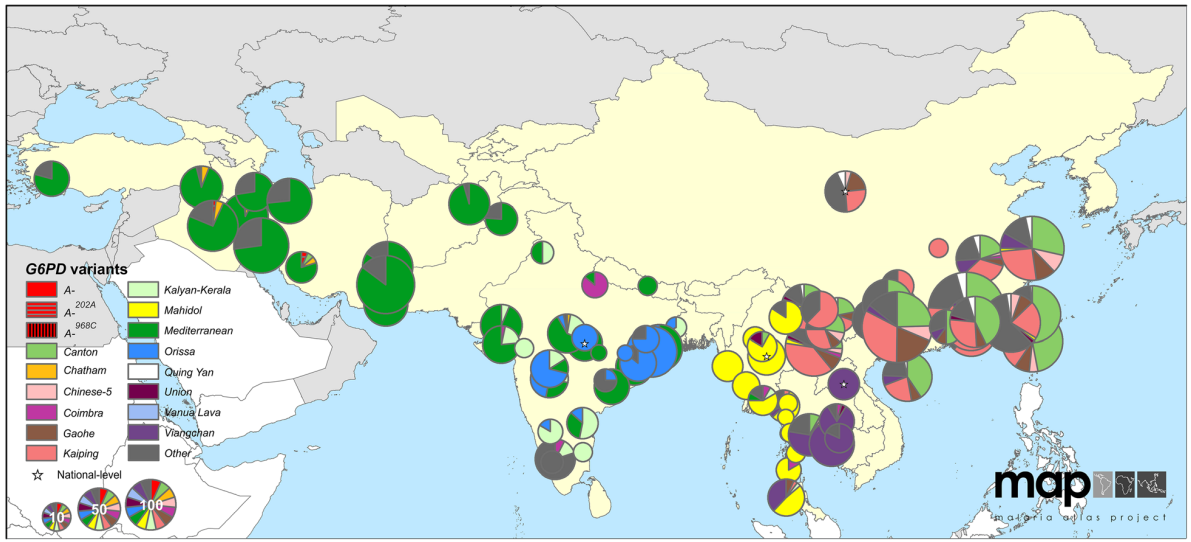


(from Flint et al 1998)

Human sickle-cell allele (HbS): (Currat et al 2002)

- Single base substitution
- provide protection against malaria (but deleterious in homozygotes)

Motivation: Hemoglobinopathies



(Howes et al 2013)

Human G6PD variants:

- over 130 G6PD deficiency alleles; 34 variants at high frequency
- provide protection against malaria but increases risk of anemia
- Estimated ages 40-400 generations (various)

MOTIVATION: A NICE STORY ABOUT MICE.

mice: AH Harris

- Dark-pigmented mammals and reptiles on volcanic outcrops in the Southwest. (Dice, Benson 1936)
- "Dark" allele beneficial on outcrops, deleterious elsewhere.
- MC1R: basis is shared between species but not between populations (Nachman, Hoekstra)

MOTIVATION: A NICE STORY ABOUT MICE.

image: Hoekstra

- Dark-pigmented mammals and reptiles on volcanic outcrops in the Southwest. (Dice, Benson 1936)
- "Dark" allele beneficial on outcrops, deleterious elsewhere.
- MC1R: basis is shared between species but not between populations (Nachman, Hoekstra)

IMPLICATIONS

The geography of adaptation

i.e. how adaptations are shared or not across the landscape tells us about

- local adaptation: what do fitness landscapes look like?
- constraint: many possible solutions or not?
- speciation: how fast can distinct adaptations accumulate?

Main summarizing question:

What is the geographic resolution of adaptation?

Also: When is there sharing of solutions?
Every region its own solutions?
How can we tell which is happening?

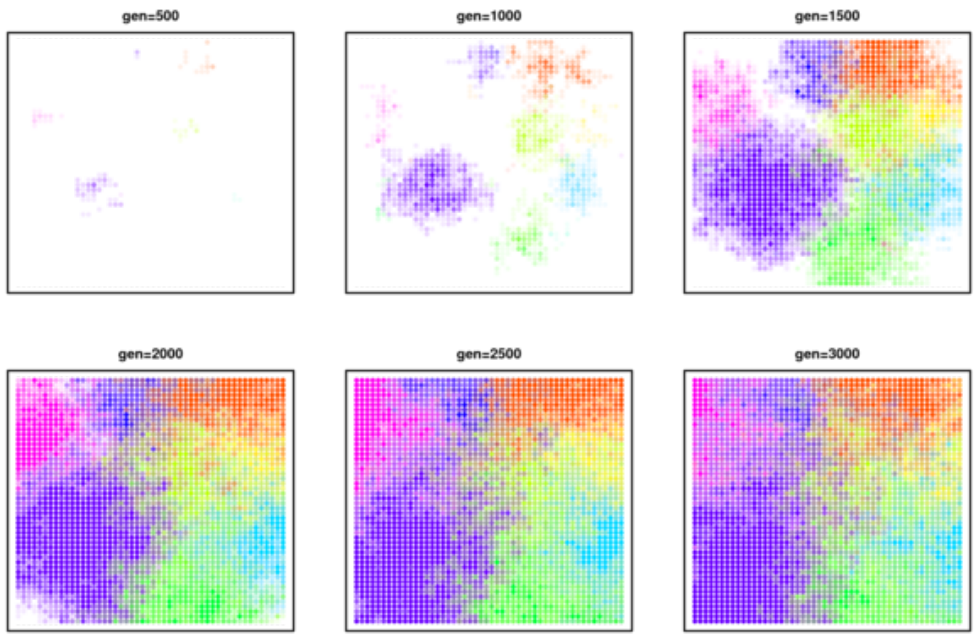
THE REST OF THE TALK

- Homogeneous landscape (quickly?)
- Patchy landscape: transients

note motivation from: Pennings & Hermisson, Soft Sweeps

First: homogeneous landscape

- Continuous species range with constant population density
- Selective pressure is geographically uniform, changes from deleterious $(1 - s_d)$ to beneficial $(1 + s_b)$
- Selected mutations are selectively equivalent (e.g. same base pair or on same pathway) and so *exclude* each other



Dynamics

Some alleles are present as standing variation

- in small clusters with effective density $\approx \rho\mu/s_d$.

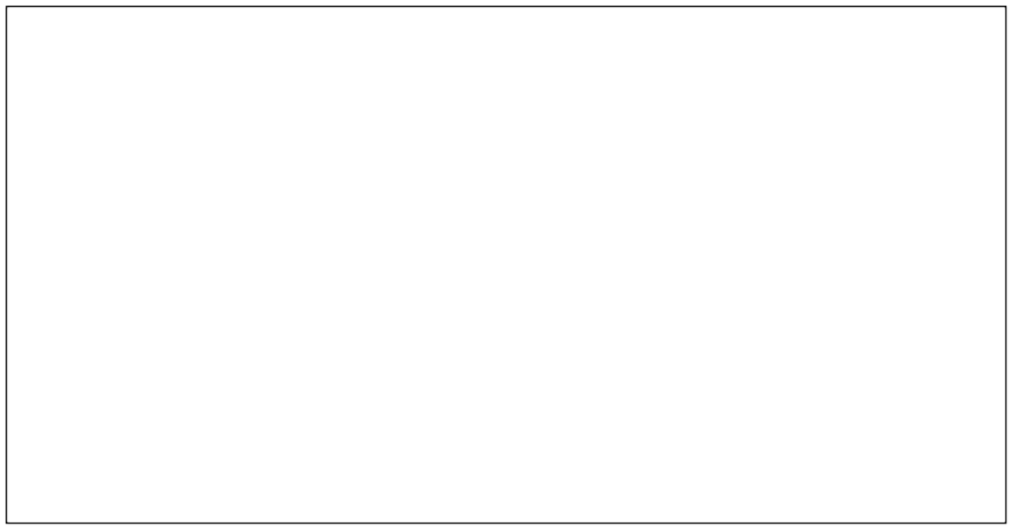
Beneficial (s_b) alleles:

- fix locally with probability $\approx 2s_b/\xi^2$, and if they do:
- spread radially with speed $\approx \sigma\sqrt{s_b}$ (Fisher; KPP 1937)

Mutations as a Point Process

Combining these,

- Standing, and new mutations as Poisson processes
- thinned by chance of local fixation
- then spreading outward at constant speed
- and excluding further mutations.

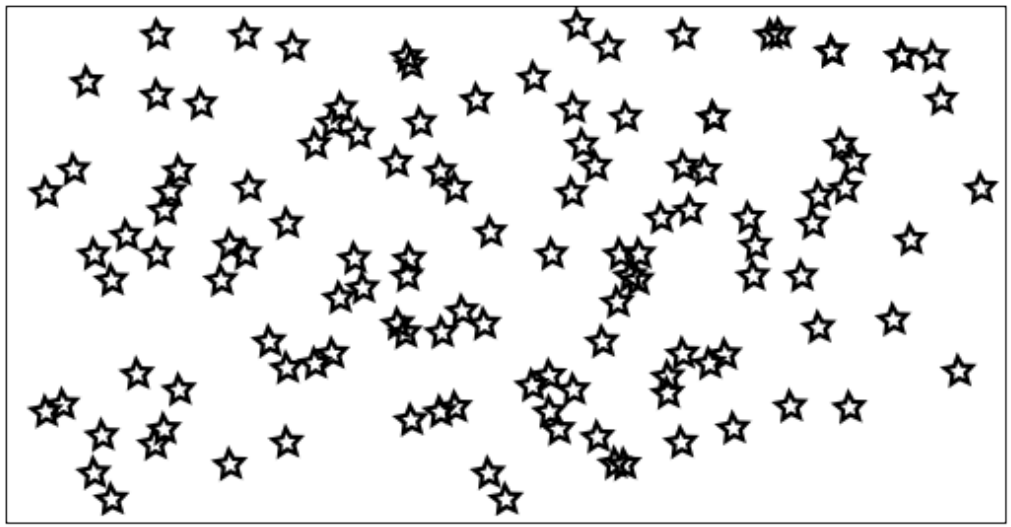


s = selection coefficient
ρ = pop density
μ = mutation rate
σ = SD dispersal distance
ξ = SD # of offspring

Mutations as a Point Process

Combining these,

- Standing, and new mutations as Poisson processes
- thinned by chance of local fixation
- then spreading outward at constant speed
- and excluding further mutations.

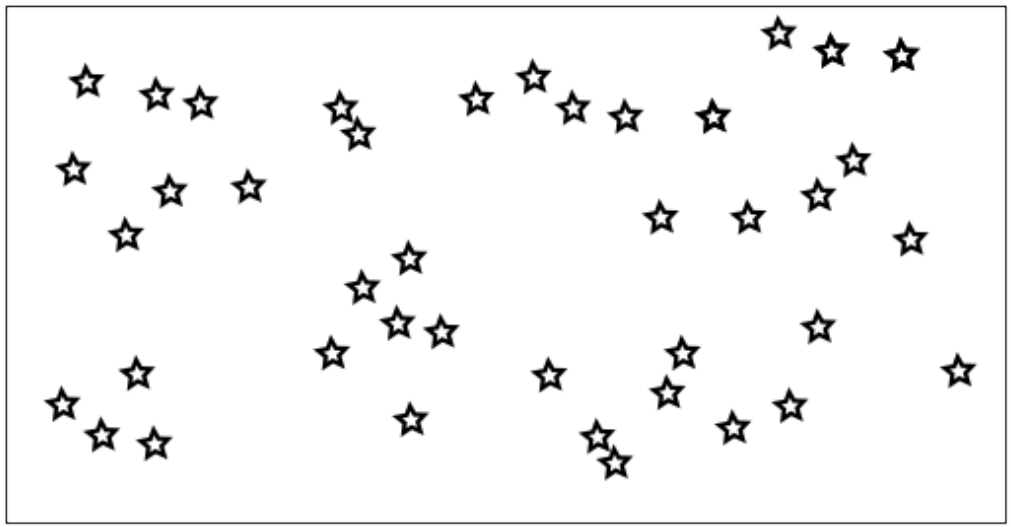


s = selection coefficient
ρ = pop density
μ = mutation rate
σ = SD dispersal distance
ξ = SD # of offspring

Mutations as a Point Process

Combining these,

- Standing, and new mutations as Poisson processes
- thinned by chance of local fixation
- then spreading outward at constant speed
- and excluding further mutations.

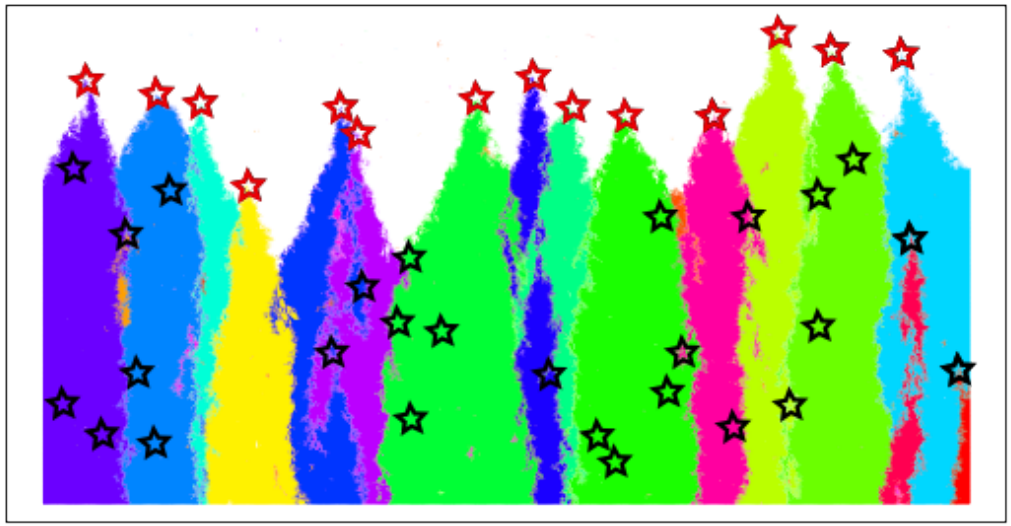


s = selection coefficient
ρ = pop density
μ = mutation rate
σ = SD dispersal distance
ξ = SD # of offspring

Mutations as a Point Process

Combining these,

- Standing, and new mutations as Poisson processes
- thinned by chance of local fixation
- then spreading outward at constant speed
- and excluding further mutations.

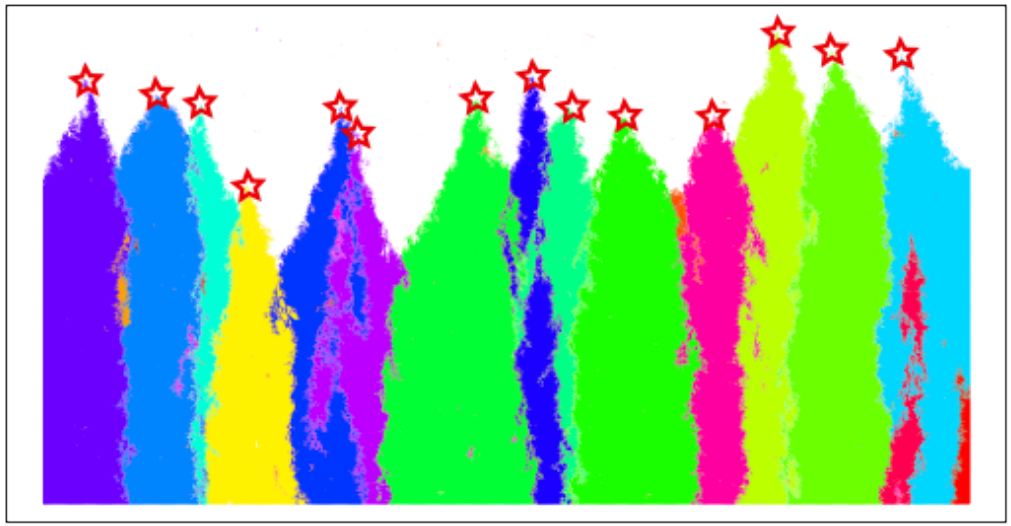


s = selection coefficient
ρ = pop density
μ = mutation rate
σ = SD dispersal distance
ξ = SD # of offspring

Mutations as a Point Process

Combining these,

- Standing, and new mutations as Poisson processes
- thinned by chance of local fixation
- then spreading outward at constant speed
- and excluding further mutations.



s = selection coefficient
ρ = pop density
μ = mutation rate
σ = SD dispersal distance
ξ = SD # of offspring

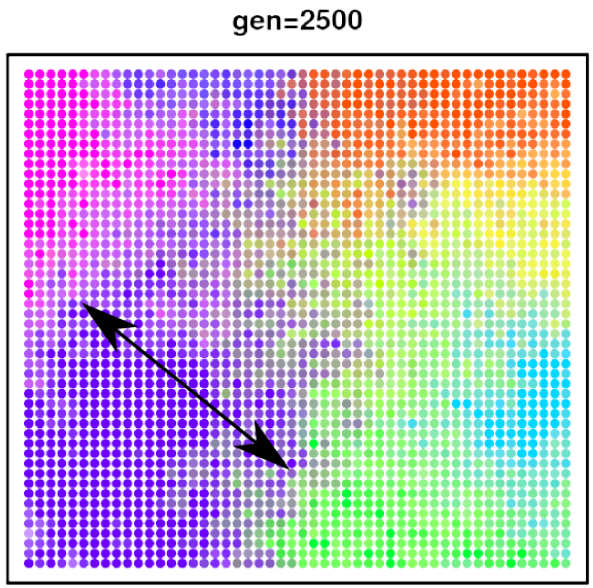
RULE OF THUMB: A CHARACTERISTIC LENGTH

Spatial properties mainly determined by a characteristic length, solving:

$$\lambda_0 \pi \chi^2 + \lambda \pi \chi^3 / v = 1,$$

χ is diameter of space-time cone in which expected to find one successful mutation.

λ_0 = density of standing mutations
λ = flux of new mutations
v = speed of wave
s_b = fitness advantage, $t > 0$
s_d = fitness disadvantage, $t < 0$
ρ = pop density
μ = mutation rate
σ = SD dispersal distance
ξ = SD # of offspring
π = 3.1415 . . .

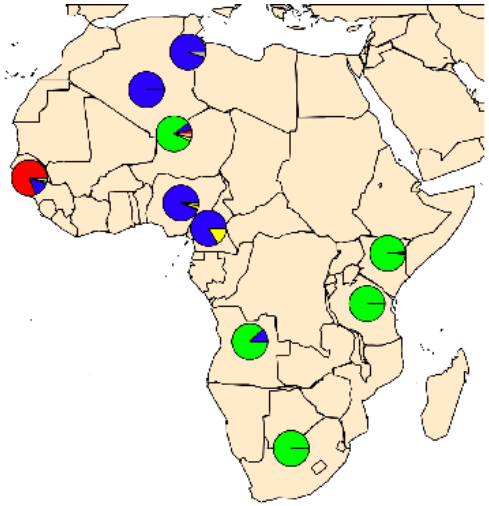


With no standing variation:

$$\chi = \left(\frac{\sigma \xi^2}{\rho \mu \sqrt{2s}\, \pi} \right)^{1/3}$$

Also: mean time until adaptation; proportion from standing variation; size of sampled cluster, etc.

SICKLE-CELL ALLELE (HBS)



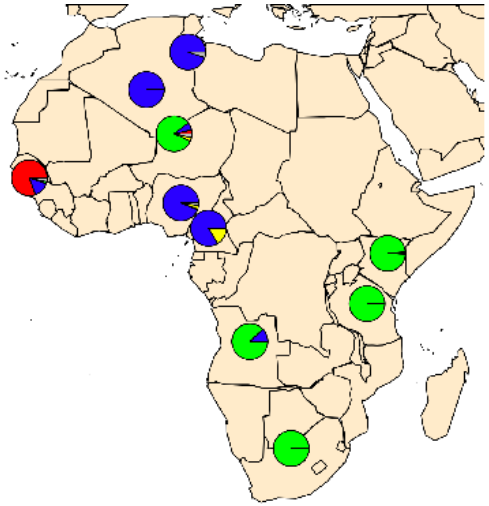
(from Flint et al 1998)

Human sickle-cell allele (HbS): (Currat et al 2002)

- Single base substitution: $\mu = 10^{-8}$
- Balancing selection: no problem
- Say, $\sigma = 50$ km and $\rho = 2$ people$/km^2$ and $s_d = .05$ or $= .5$.

	estimated	weaker	stronger	
s_b	0.15	0.15	0.15	
s_d	–	0.05	0.5	
σ	10–100	50	50	km
μ	10^{-8}	10^{-8}	10^{-8}	
ρ	2	2	2	km^{-2}
χ	1000	980	1120	km
age	10–70	63	77	gen
z_0	–	52%	6%	

SICKLE-CELL ALLELE (HBS)



(from Flint et al 1998)

Observed pattern:

- Haplotype pattern on scale of 1000 km
- Estimated age 10-70 generations

  (Currat et al; Modiano et al)

	estimated	weaker	stronger	
s_b	0.15	0.15	0.15	
s_d	–	0.05	0.5	
σ	10–100	50	50	km
μ	10^{-8}	10^{-8}	10^{-8}	
ρ	2	2	2	km^{-2}
χ	1000	980	1120	km
age	10–70	63	77	gen
z_0	–	52%	6%	

Sickle-cell allele (HbS)

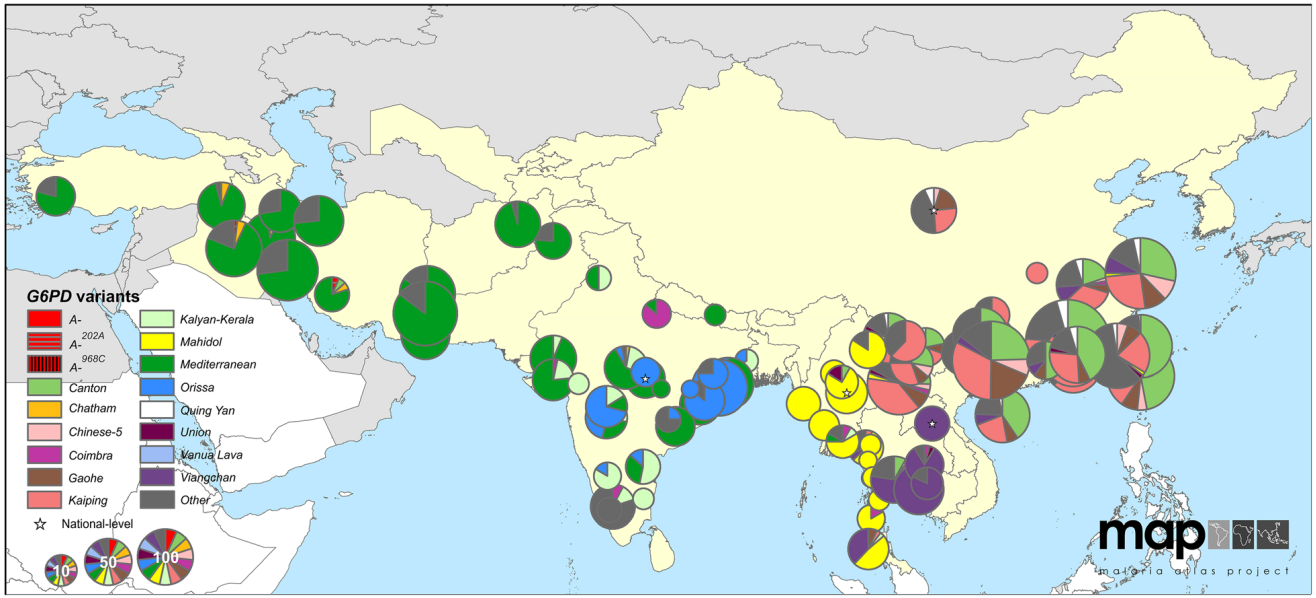
s_b = fitness advantage, $t > 0$
s_d = fitness disadvantage, $t < 0$
ρ = pop density
μ = mutation rate
σ = SD dispersal distance
χ = characteristic length
z_0 = proportion from standing

We compute:

- characteristic length $\chi \approx 1000$ km
- mean adaptation time $\mathbb{E}[\tau] \approx 70$ generations
- proportion from standing variation z_0 significant
- time until pattern is erased $R^2/\sigma^2 \approx 6,400$ generations

	estimated	weaker	stronger	
s_b	0.15	0.15	0.15	
s_d	–	0.05	0.5	
σ	10–100	50	50	km
μ	10^{-8}	10^{-8}	10^{-8}	
ρ	2	2	2	km^{-2}
χ	1000	980	1120	km
age	10–70	63	77	gen
z_0	–	52%	6%	

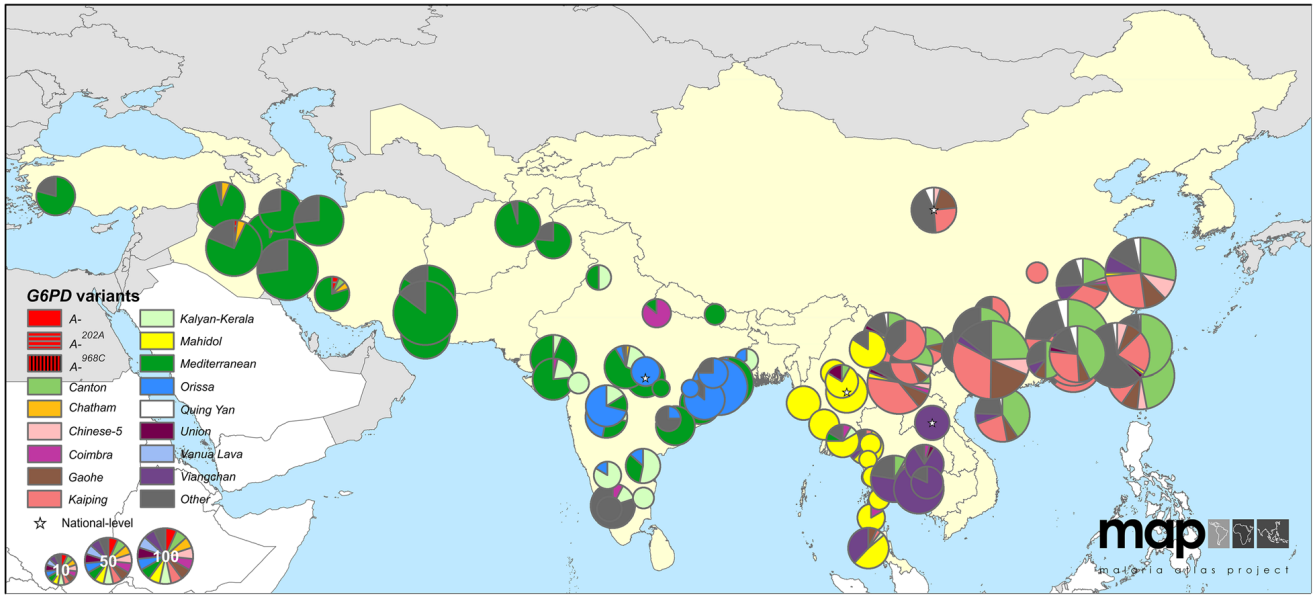
EXAMPLE: G6PD



Human G6PD variants: *(Howes et al 2013)*

- 34 variants across $4,000,000$ km^2: $\chi \approx 350$ km
- Estimated ages 40-400 generations (various)
- Estimated $s = .25$ (Slatkin et al 2008)
  or $s = .04$ (Tishkoff et al 2001)
- 150 coding bases: $\mu = 150 \times 10^{-8}$
- Say, $\sigma = 50$ km and $\rho = 2$ people$/km^2$

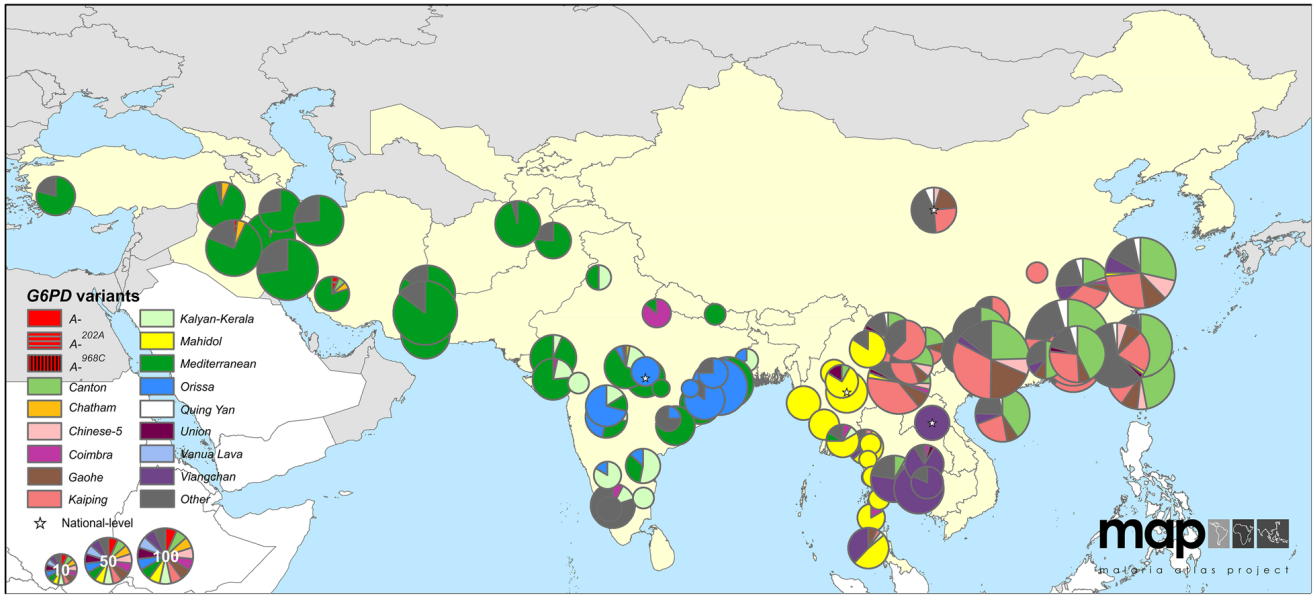
EXAMPLE: G6PD



Human G6PD variants: *(Howes et al 2013)*

	estimated	weaker	stronger	
s_b	0.25 or 0.04	0.25	0.04	
s_d	0.1	–	–	
σ	50	–	–	km
μ	1.5×10^{-6}	–	–	
ρ	2	–	–	km^{-2}
χ	350	144	254	km
age or $\mathbb{E}[\tau]$	40–400	7	32	gen
z_0	–	83%	51%	

EXAMPLE: G6PD



Human G6PD variants:

(Howes et al 2013)

	estimated	weaker	stronger	
s_b	0.25 or 0.04	0.25	0.04	
s_d	0.1	–	–	
σ	50	–	–	km
μ	1.5×10^{-6}	–	–	
ρ	2	–	–	km^{-2}
χ	350	144	254	km
age or $\mathbb{E}[\tau]$	40–400	7	32	gen
z_0	–	83%	51%	

Second case: Patchy environment

Focal allele is beneficial in patches; deleterious between:

A = patch size

R = patch separation

s_b = on-patch selective advantage

$-s_d$ = between-patch selective cost

ρ = pop density

μ = mutation rate

σ = SD dispersal distance

ξ = SD # of offspring

What is the time scale of:

- appearance of new mutations in unadapted patches?
- transit of mutations between patches?

What does the latter look like?

ADAPTATION BY NEW MUTATION

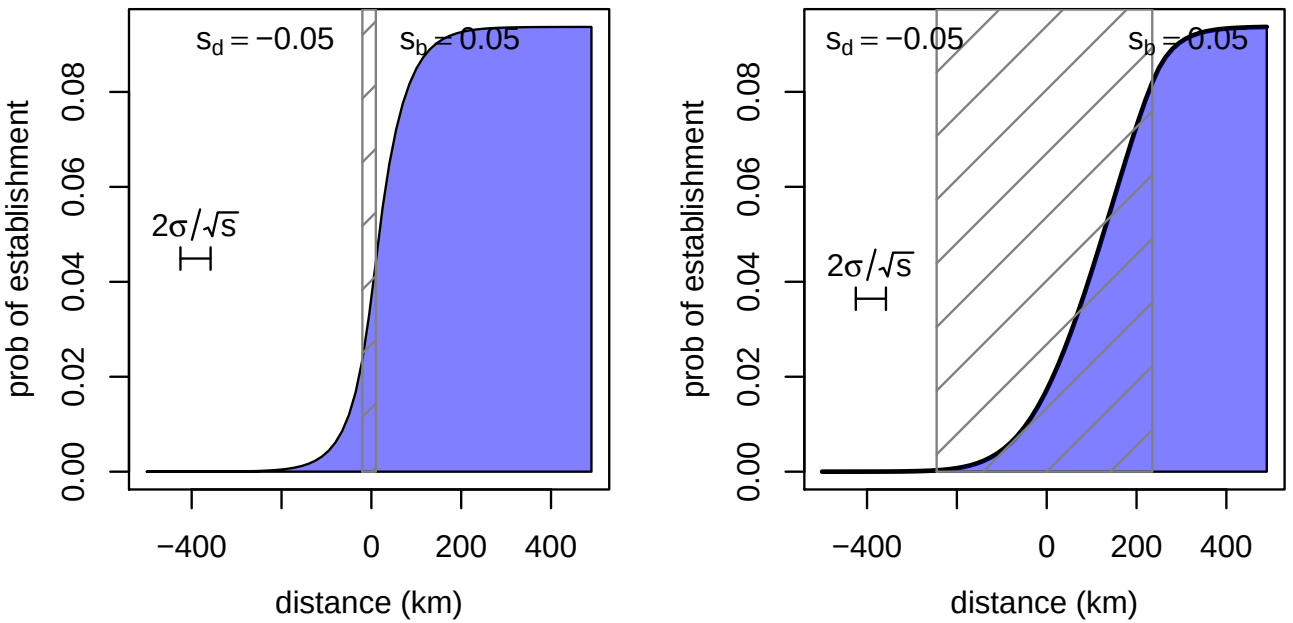
Naively: a new mutant in a patch establishes with probability

$$p_{\text{estab}} = \frac{2s_b}{\xi^2}$$

so rate of influx is

$$\lambda_{\text{mut}} = A\rho\mu\frac{2s_b}{\xi^2} \quad \text{per generation.}$$

Actually, probability of establishment at distance r is $\simeq \exp(-r\sqrt{s_d}/\sigma)$ (Barton 1987) . . . naive calculation does pretty good:



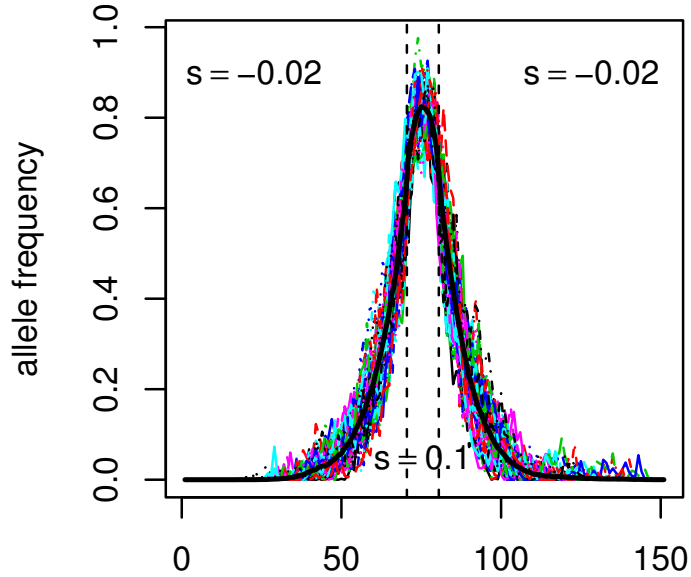
MIGRATION: TRANSITING FAMILIES

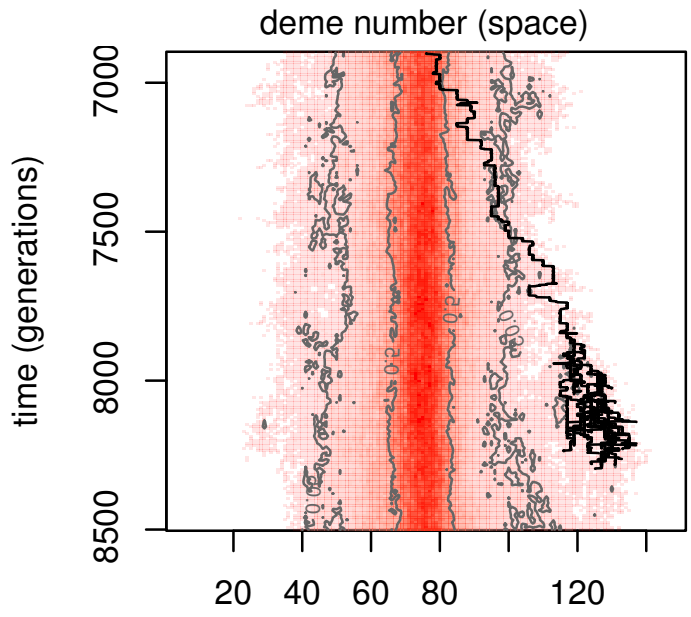
Suppose a mutation has fixed locally in one patch.
How long until it reaches another at distance R?

At migration-selection balance: frequency at distance r is

$$q(r) \approx \frac{1}{2}\left(\frac{r\sqrt{2s_d}}{\sigma}\right)^{-\frac{d-1}{2}} \exp\left\{-\frac{r\sqrt{2s_d}}{\sigma}\right\}.$$

This deterministic "equilibrium" is composed of rare long-distance migrant families.





FAMILY DECOMPOSITION

Draw a circle at distance r_0 from the original patch.

Definition: Any two individuals outside the patch that share an ancestor who lived outside r_0 are in the same "family".

In the continuum limit, these families are subcritical branching processes, killed on hitting a patch, with inhomogeneous branching rate.

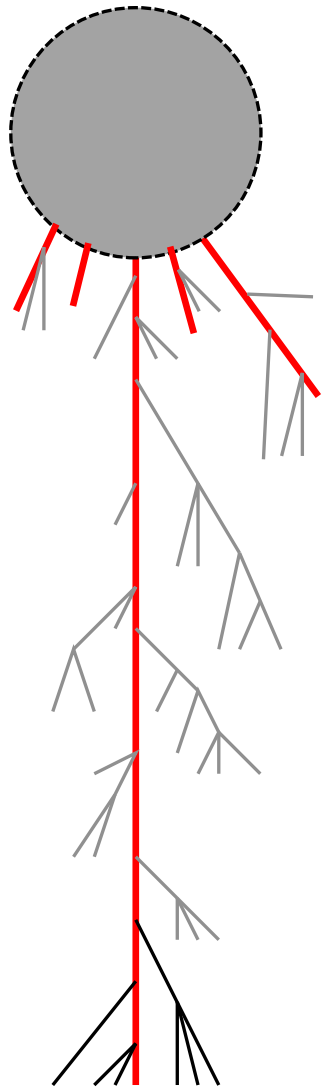
For large r_0, branching rate (nearly) homogeneous $1 - s_d$.

(at least those making it to the new patch)

Modeling transiting families

General idea:

- Between patches, transiting families die out:
  growth rate $-s_d < 0$ ("subcritical").
- Chance that one lives for t generations is $\sim e^{-s_d t}$
- In the (rare) event it does, looks like a single "trunk"
  with transient "branches" (Geiger 1999)
- Trunk moves as a random walk.



THE EFFECTIVE MIGRATION RATE

Rate at which mutations transit between patches is

$\lambda_{\text{mig}}(R) =$ (outflux of families)
$\times$ (prob family establishes in patch at R).

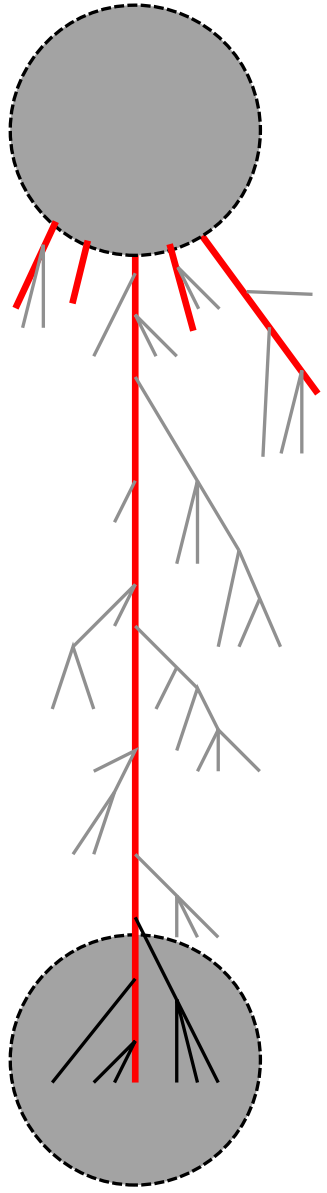
On the other hand, without the new patch,

$q(R) =$ (outflux of families)
$\times$ (occupation time of a family near R).

Each is the probability that the family hits the patch, multiplied by

- the probability it establishes, or
- its occupation time

given it gets there.



MIGRATION AND MUTATION COMPARED

Constant is $\approx \rho \times 2s_b/\xi^2 \times 1/s_d$, so

$$\lambda_{\text{mig}} \approx \frac{\rho s_b}{\xi^2 s_d} \left(\frac{R\sqrt{2s_d}}{\sigma} \right)^{-\frac{d-1}{2}} \exp\left(-\frac{R\sqrt{2s_d}}{\sigma} \right).$$

The relevant time scales of each are $T_{\text{mig}} = 1/\lambda_{\text{mig}}$ and $T_{\text{mut}} = 1/\lambda_{\text{mut}}$.

$$\lambda_{\text{mut}} = A\rho\mu 2s_b/\xi^2, \qquad \text{so}$$

$\lambda_{\text{mut}} \approx \lambda_{\text{mig}}$ if (in 1D)

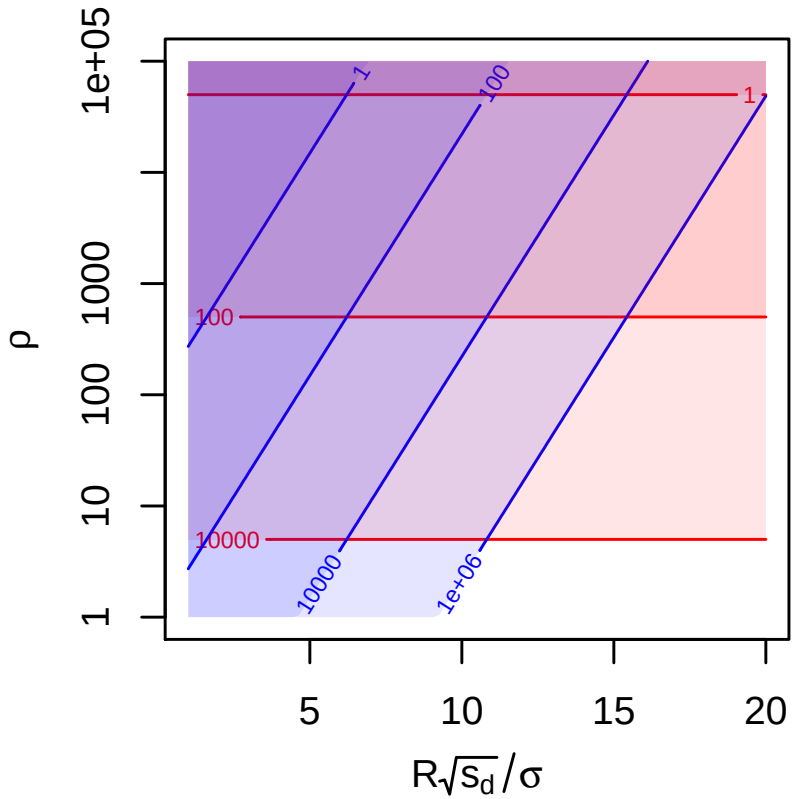
$$A\mu \approx (1/s_d) \exp(-R\sqrt{2s_d}/\sigma).$$

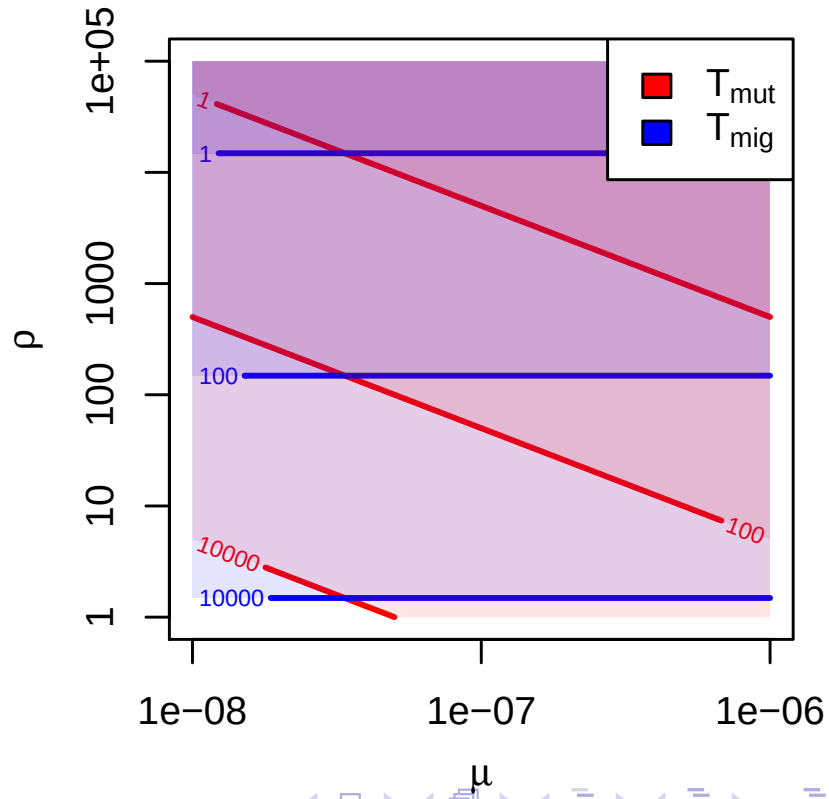
Migration and mutation compared

Constant is $\approx \rho \times 2s_b/\xi^2 \times 1/s_d$, so

$$\lambda_{\text{mig}} \approx \frac{\rho s_b}{\xi^2 s_d} \left(\frac{R\sqrt{2s_d}}{\sigma} \right)^{-\frac{d-1}{2}} \exp\left(-\frac{R\sqrt{2s_d}}{\sigma} \right).$$

The relevant time scales of each are $T_{\text{mig}} = 1/\lambda_{\text{mig}}$ and $T_{\text{mut}} = 1/\lambda_{\text{mut}}$.





CAN WE DETECT ALLELES SHARED BY MIGRATION?

Shared alleles will share a haplotype; how long?

- Haplotype whittled down by recombination between patches:
- if transit between patches takes t generations,
- shared haplotype will be $\approx$ Exponential with mean $1/t$.

The trunk: $\approx$ Brownian, killed at rate s_d.

Let τ be the hitting time of the new patch, at distance r.

If length of shared haplotype is L, then in 1–D:

$$\mathbb{P}\{L > \ell\} = \mathbb{E}\left[e^{-\ell\tau} | \tau < \infty\right] = \exp\left(-\frac{R}{\sigma}\left(\sqrt{2(\ell + s_d)} - \sqrt{2s_d}\right)\right).$$

LENGTH OF SHARED HAPLOTYPE

The probability the haplotype is of length $> \ell$ is

$$\mathbb{P}\{L > \ell\} = \mathbb{E}\left[e^{-\ell\tau}\right] = \exp\left\{-(R/\sigma)\left(\sqrt{2(\ell + s_d)} - \sqrt{2s_d}\right)\right\}$$

For large R:

- Mean transit time of allele is $\mathbb{E}[\tau] \approx R/(\sigma\sqrt{2s_d})$.
- Haplotype length is $\mathbb{E}[L] \approx \sigma\sqrt{2s_d}/R$.
- $L \stackrel{d}{=} (Y + \sqrt{2s_d})^2 - 2s_d$, with $Y \sim$ Exponential$(R\sqrt{2}/\sigma)$.
- CLT for τ as $R \to \infty$:

$$\tau \approx \frac{R}{\sigma\sqrt{2s_d}} + \sqrt{\frac{R}{\sigma(2s_d)^{3/2}}}Z, \qquad Z \sim N(0,1).$$

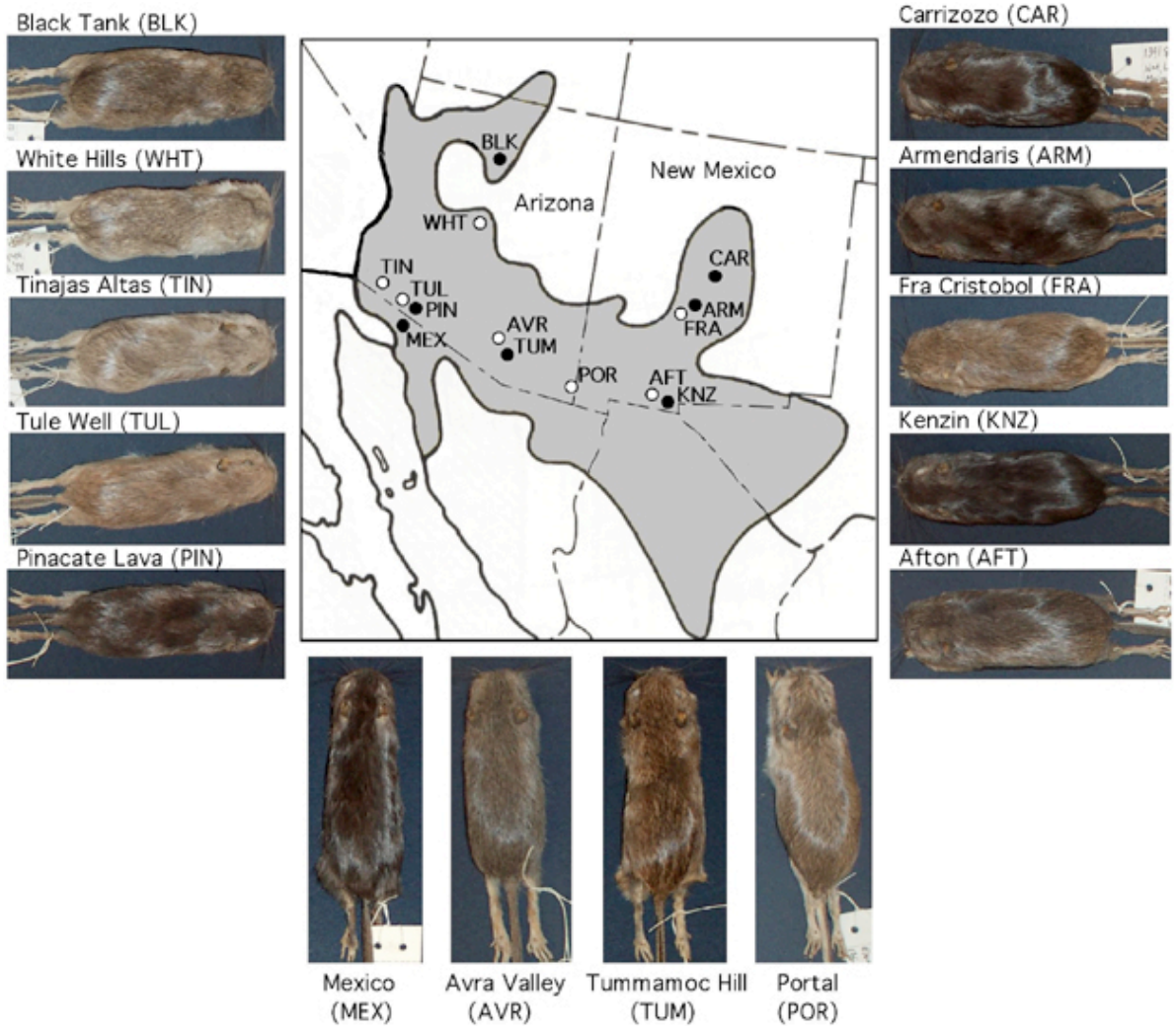
Application to Rock Pocket Mice

Back to *Chaetodipus intermedius* (Dice, Benson, Nachman, Hoekstra, etc)



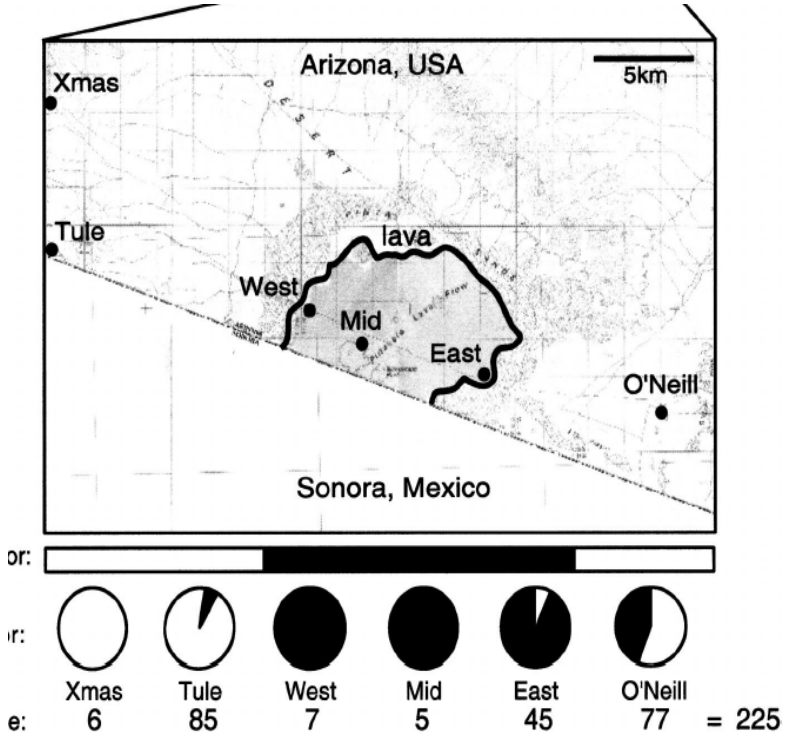

Application to Rock Pocket Mice

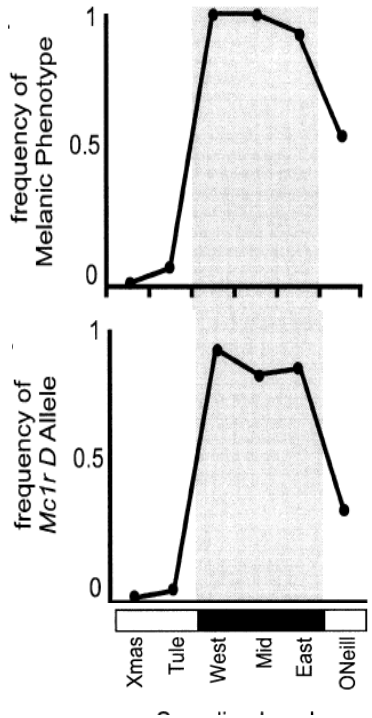
Back to *Chaetodipus intermedius* (Dice, Benson, Nachman, Hoekstra, etc)



Application to Rock Pocket Mice

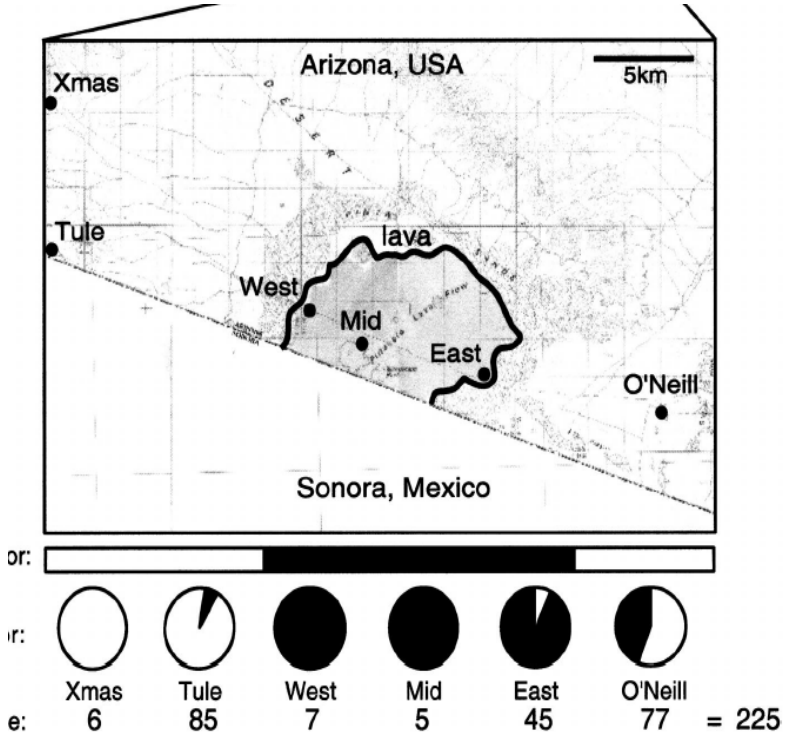
On one flow (Pinacate) an allele of MC1R is responsible for much of the change to a dark pelage (Nachman et al 2003).
Hoekstra et al (2004): further study of nearby flows.

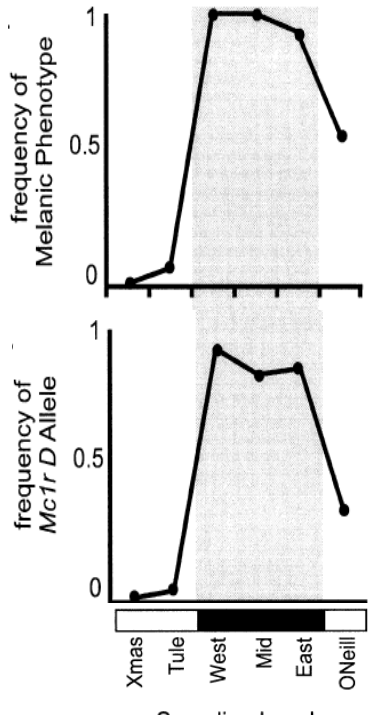




Application to Rock Pocket Mice

- Hoekstra et al (2004) estimated $0.03 < s_b < 0.3$.
- Cline width suggests $3km < \sigma/\sqrt{s_d} < 30km$
- Another species suggests $\sigma = 0.28km$
- . . . so maybe $10^{-4} < s_d < 0.01$?

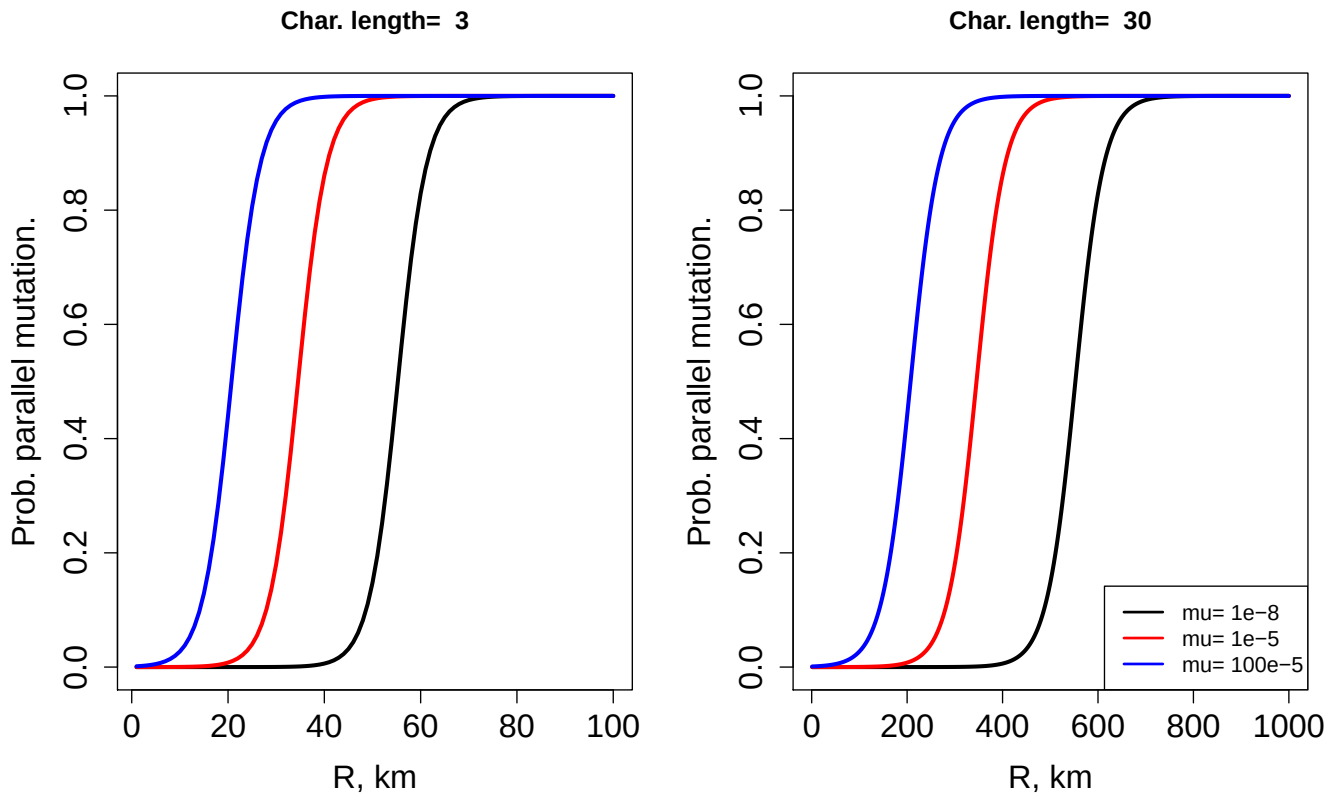




Application to Rock Pocket Mice

Probability of parallel adaptation, i.e. adaptation by mutation before mutation, is

$$\frac{\lambda_{\text{mut}}}{\lambda_{\text{mut}} + \lambda_{\text{mig}}} = \frac{2A\mu s_b}{2A\mu s_b + \exp(-\sqrt{2s_d}R/\sigma)}$$



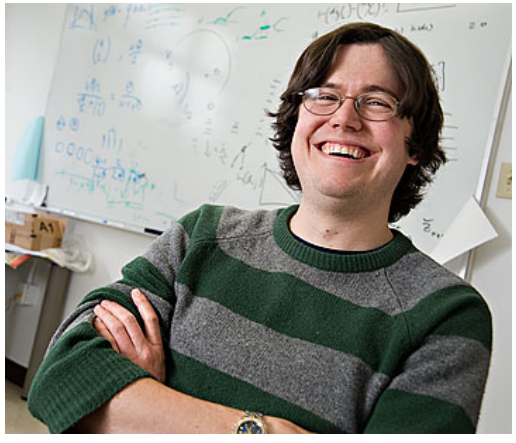
. . . and size of shared haplotype $\approx$ 56Kb.

Conclusions

- Order-of-magnitude estimates and Poisson process calculations get not unreasonable answers
- Some aspects of geometry are insensitive to parameters.
- We can describe well the initial dynamics; what happens next?
- Need to better model interactions (e.g. G6PD).

thanks

gideon bradburd



graham coop

yaniv brandvain

Michael Nachman

WITHIN A PATCH

In a discrete deme, large population model, lineages of the adapted allele moves from x to y at rate

$$q(x \leftarrow y) = \rho p(y)(1+s(y))m(y \to x).$$

If migration is nearest-neighbor,

the stationary distribution is

$$\pi(x) = p(x)(1 + s(x)).$$

Diffusion limit for motion of the

lineage is

$$dX_t = \pi'(X_t)dt + dB_t.$$