Distinguishing modes of convergent adaptation

Kristin Lee, Jeremy Berg, Graham Coop
University of California, Davis

Convergent adaptation

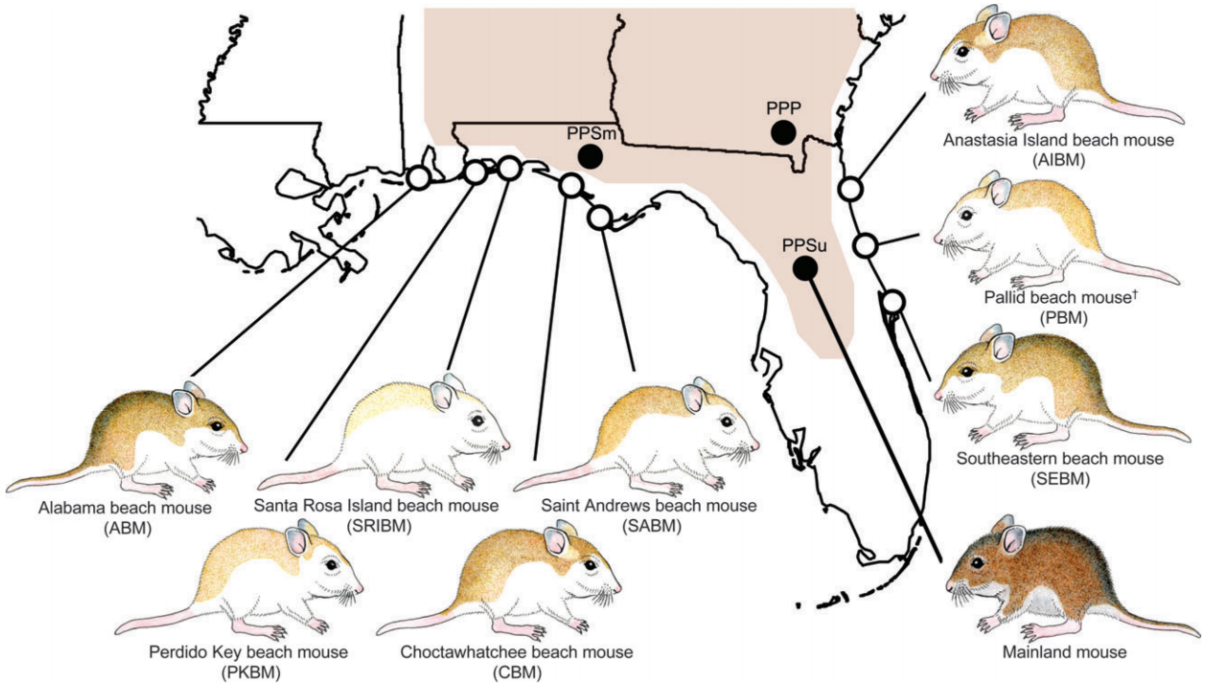
Adaptation across similar environments

Many levels of convergent adaptation:

Phenotypic level across species

http://webecoist.momtastic.com/2012/09/10/arctic-animals-37-photos-of-snow-white-wildlife/

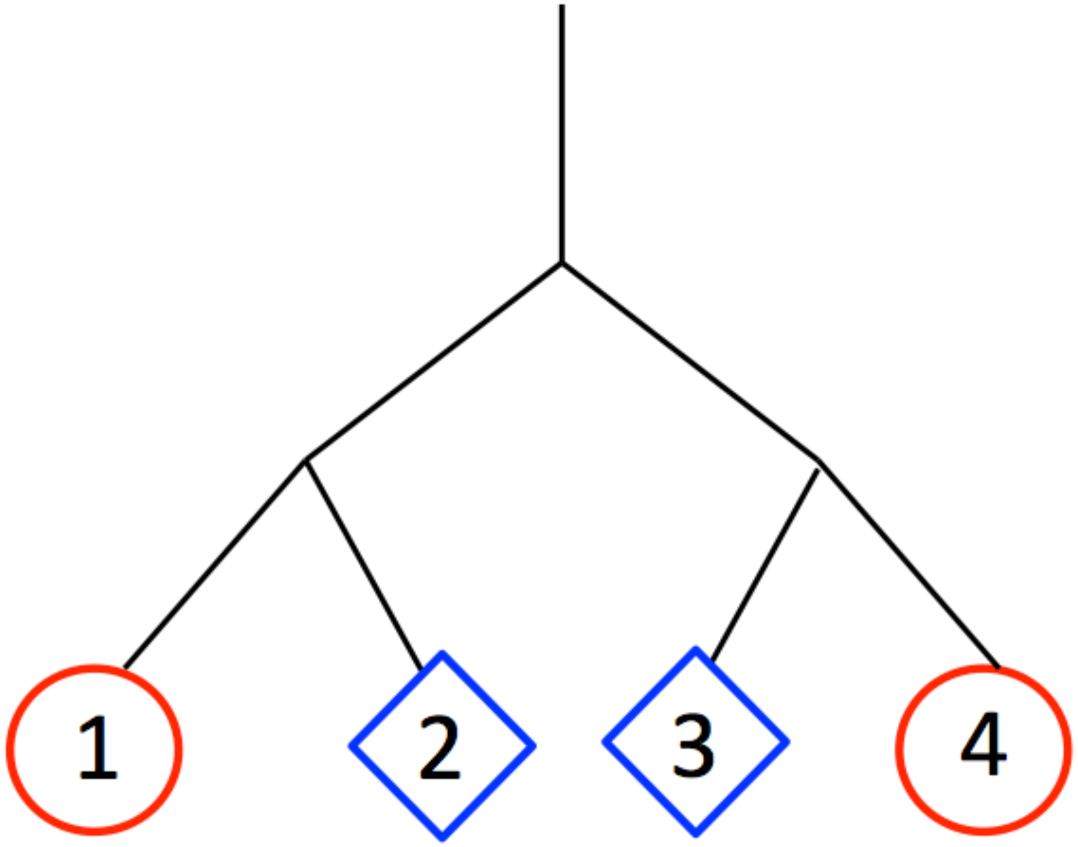
Phenotypic level within species



Steiner et al., *MBE* (2009)

Modes of genotypic convergent adaptation

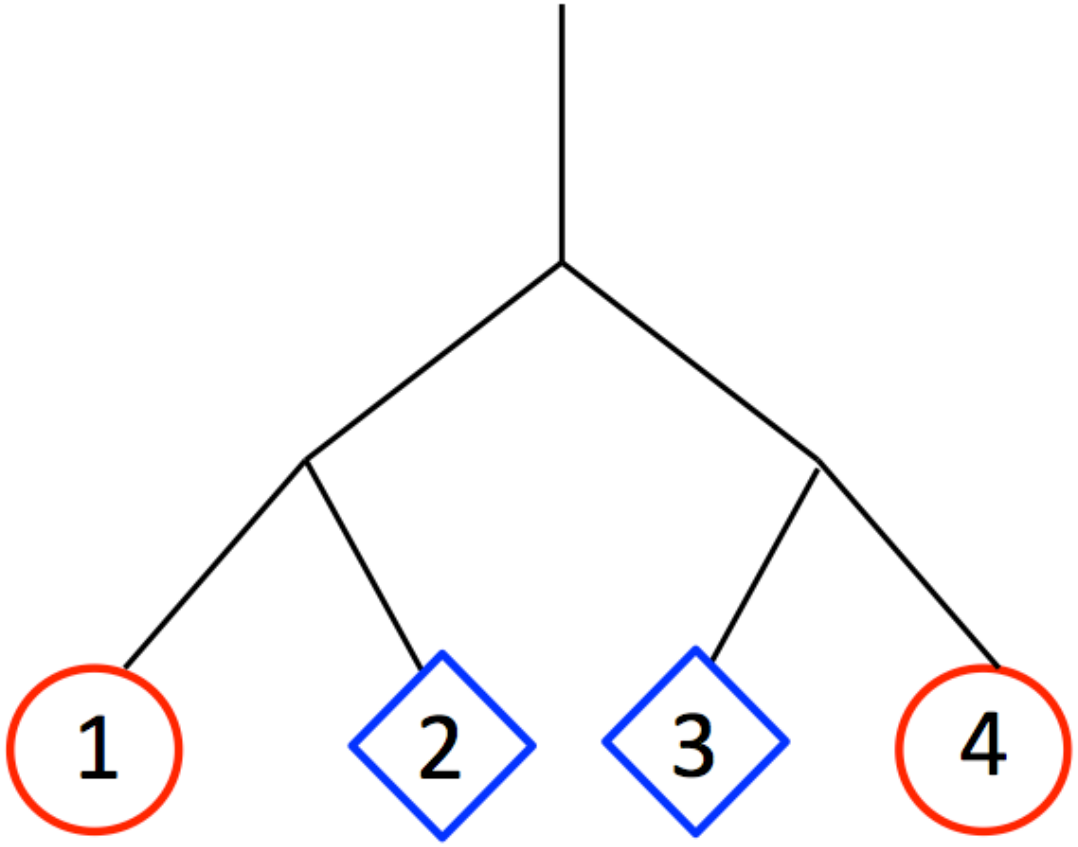
Multiple populations respond to similar environmental pressure using same genes



Modes of genotypic convergent adaptation

Multiple populations respond to similar environmental pressure using same genes

What is the basis of this adaptation?

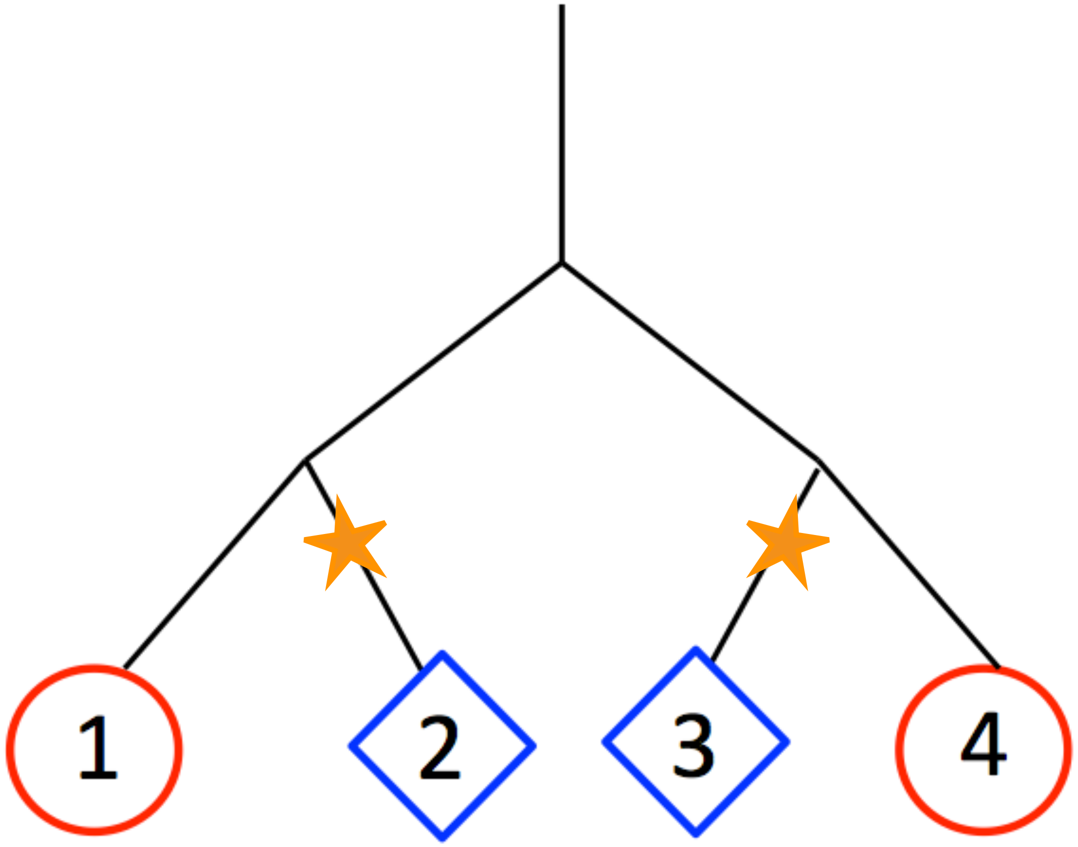


Modes of genotypic convergent adaptation

Multiple populations respond to similar environmental pressure using same genes

What is the basis of this adaptation?

1. Independent mutations in isolated populations

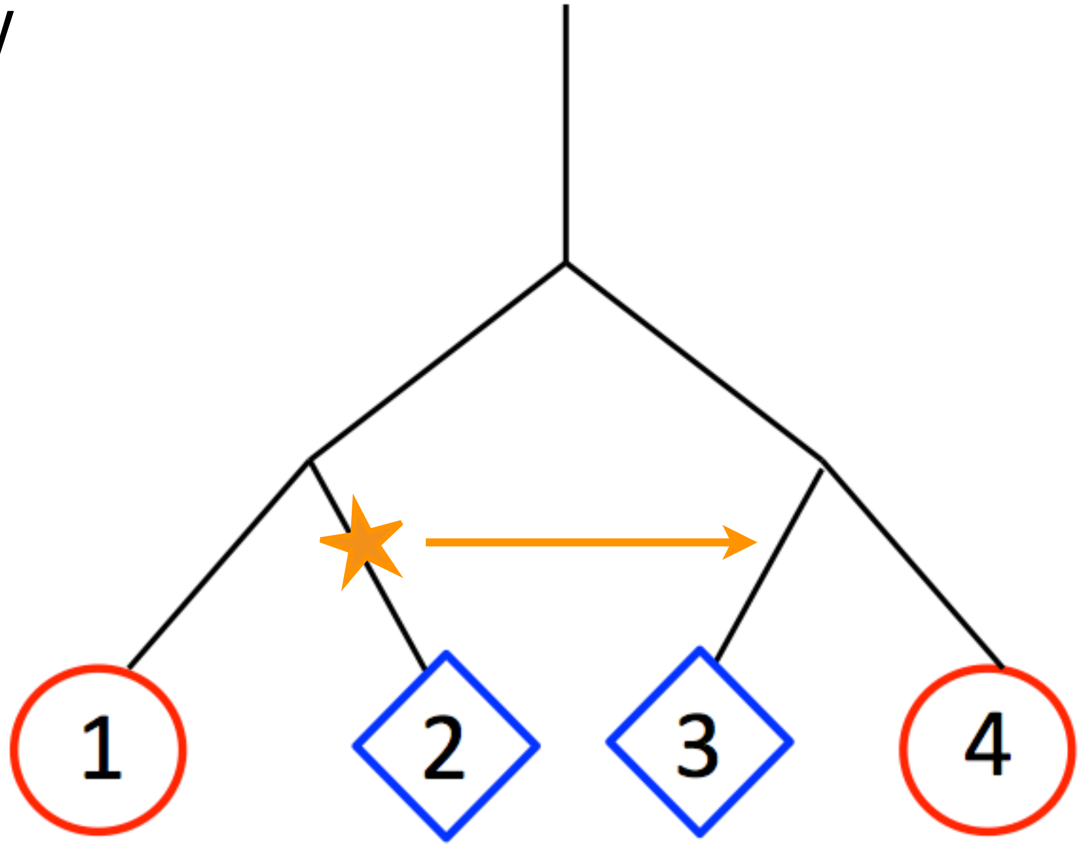


Modes of genotypic convergent adaptation

Multiple populations respond to similar environmental pressure using same genes

What is the basis of this adaptation?

1. Independent mutations in isolated populations
2. Single origin with gene flow

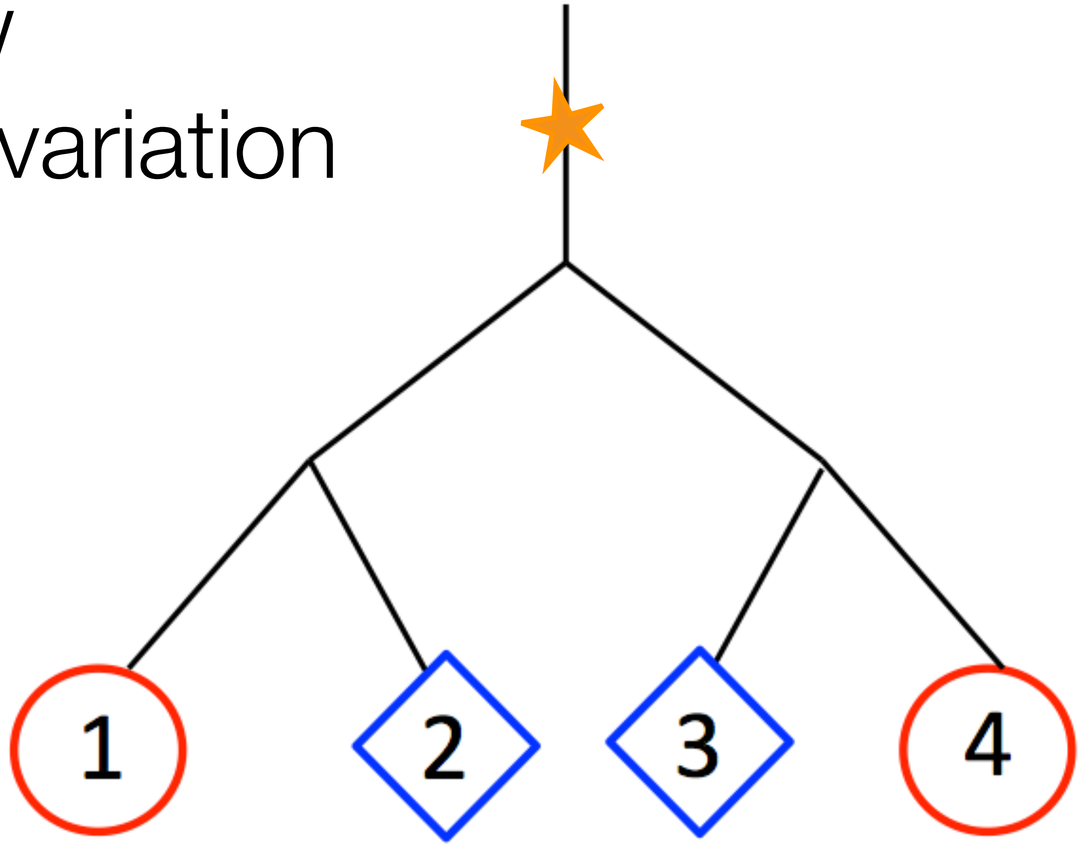


Modes of genotypic convergent adaptation

Multiple populations respond to similar environmental pressure using same genes

What is the basis of this adaptation?

1. Independent mutations in isolated populations
2. Single origin with gene flow
3. Shared ancestral standing variation

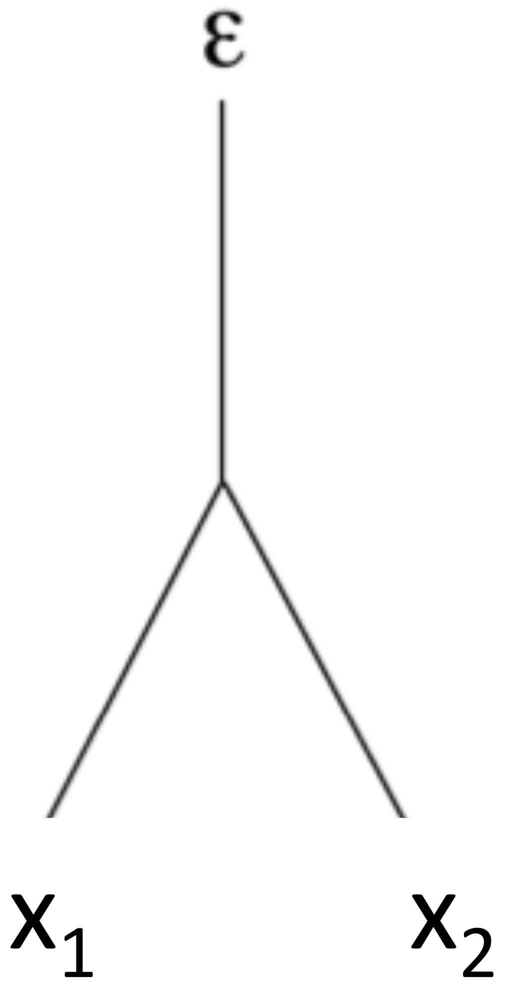


How can we distinguish modes of convergent adaptation using population genomic data?

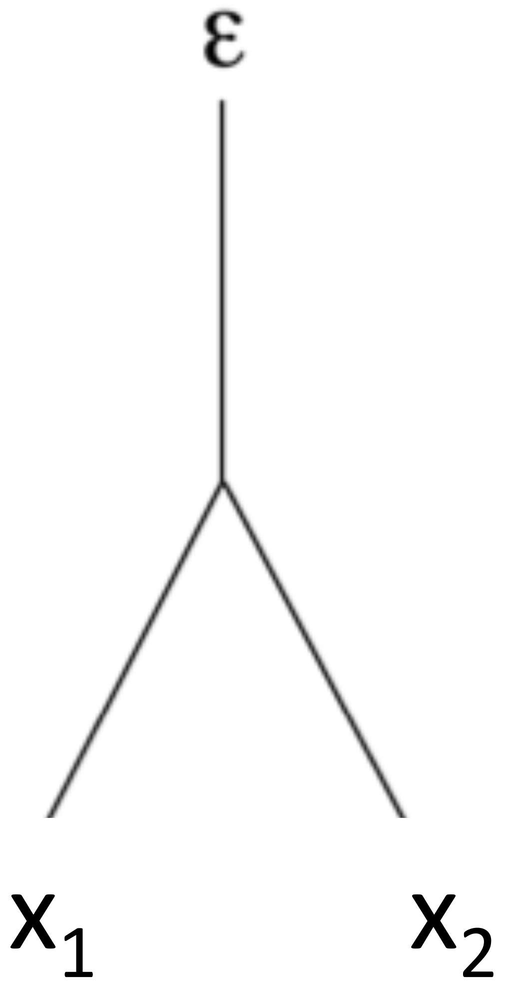
1. Model allele frequencies around selected site
    - Null model
    - Variance within populations experiencing selection
    - Covariance between selected populations

2. Composite-likelihood approach to distinguish between these models

Effect of drift on population allele frequencies

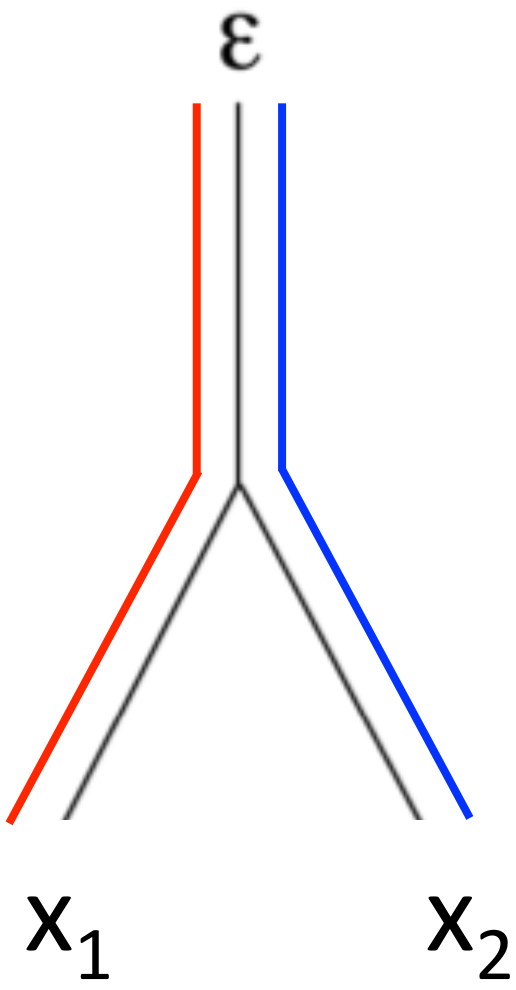


Effect of drift on population allele frequencies



$$\mathbb{E}[\Delta x_1] = \mathbb{E}[\Delta x_2] = 0$$

Effect of drift on population allele frequencies

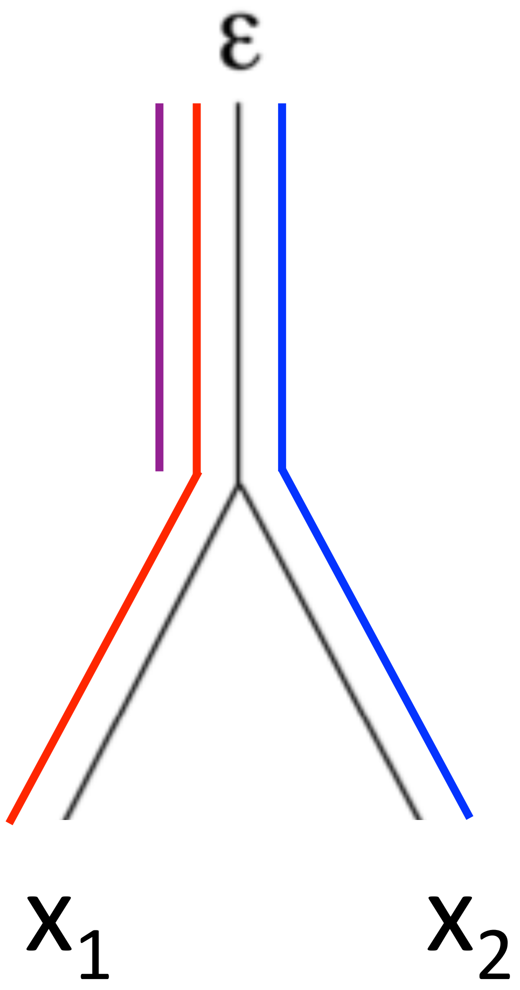


$$\mathbb{E}[\Delta x_1] = \mathbb{E}[\Delta x_2] = 0$$

$$\mathrm{Var}[\Delta x_1] = \epsilon(1-\epsilon)f_{11}$$

$$\mathrm{Var}[\Delta x_1] = \epsilon(1-\epsilon)f_{22}$$

Effect of drift on population allele frequencies



$$\mathbb{E}[\Delta x_1] = \mathbb{E}[\Delta x_2] = 0$$

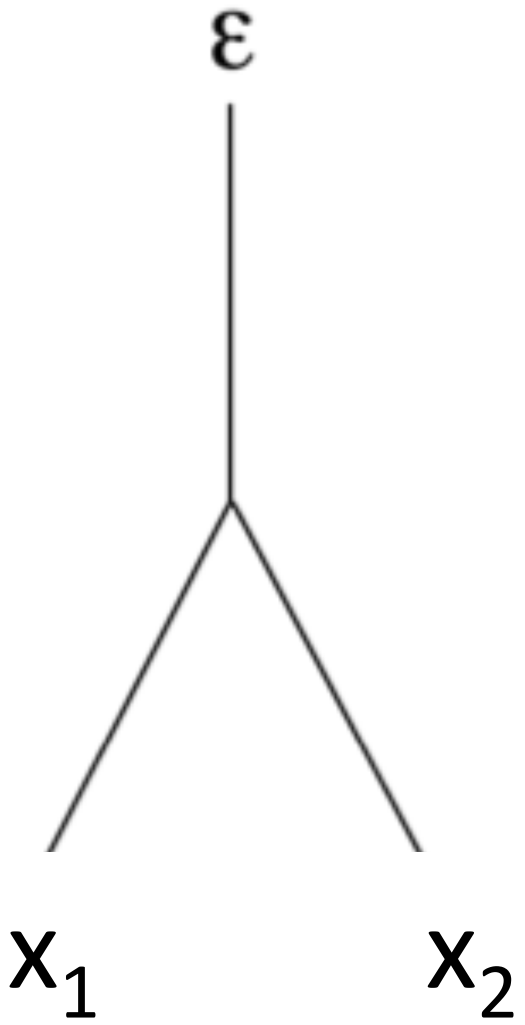
$$\mathrm{Var}[\Delta x_1] = \epsilon(1-\epsilon) f_{11}$$

$$\mathrm{Var}[\Delta x_1] = \epsilon(1-\epsilon) f_{22}$$

$$\mathrm{Cov}[\Delta x_1] = \epsilon(1-\epsilon) f_{12}$$

Coancestry coefficients

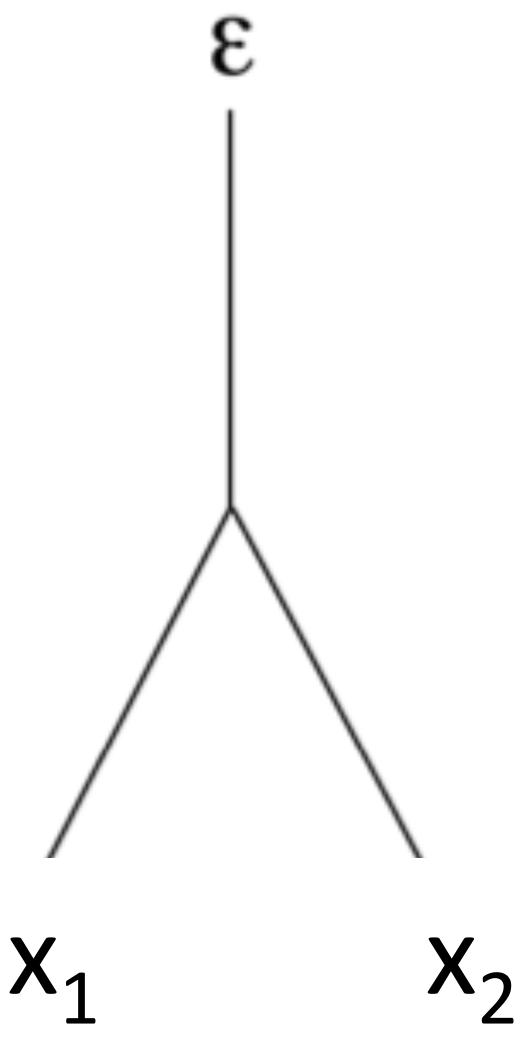
$$f_{ii} = \frac{\mathrm{Var}[x_i]}{\epsilon(1-\epsilon)}$$



Coancestry coefficients

$$f_{ii} = \frac{\mathrm{Var}[x_i]}{\epsilon(1-\epsilon)}$$

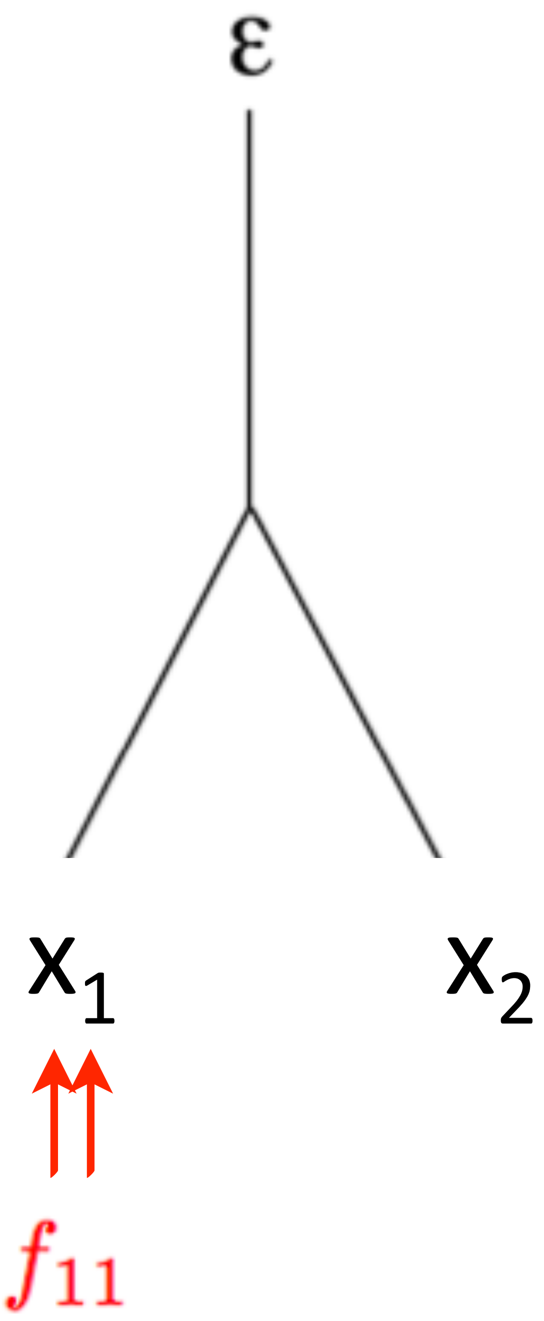
Probability a pair of lineages coalesce before reaching the ancestral population



Coancestry coefficients

$$f_{ii} = \frac{\mathrm{Var}[x_i]}{\epsilon(1-\epsilon)}$$

Probability a pair of lineages coalesce before reaching the ancestral population

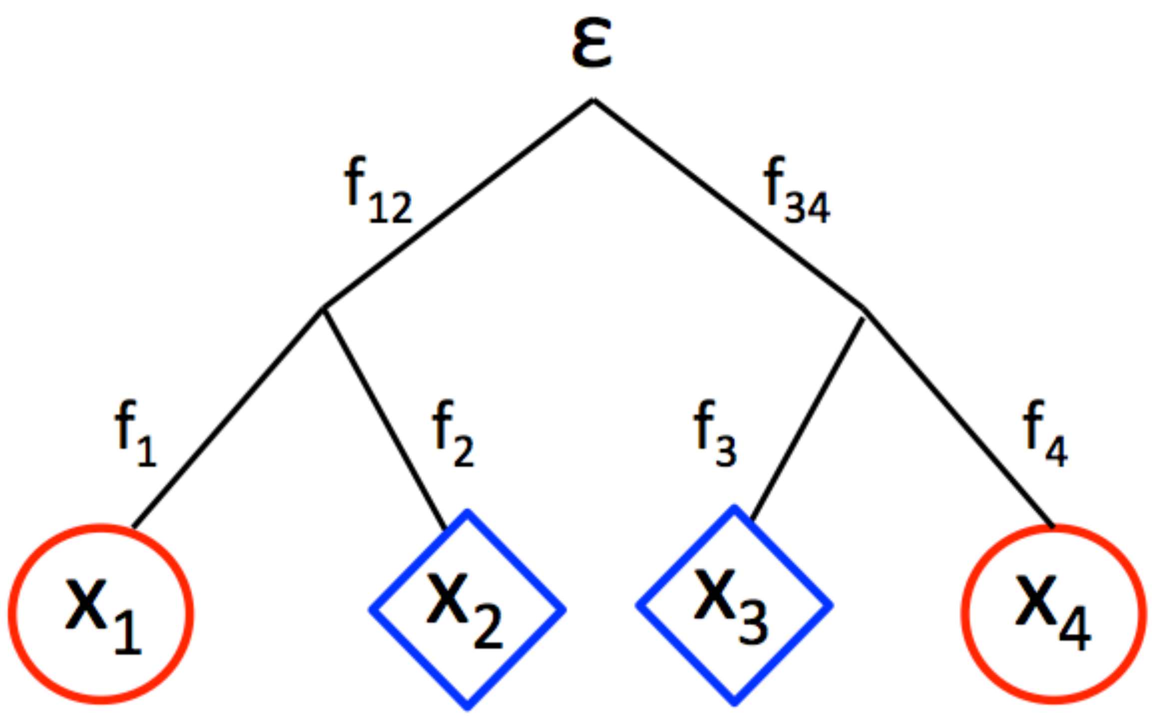


Coancestry coefficients

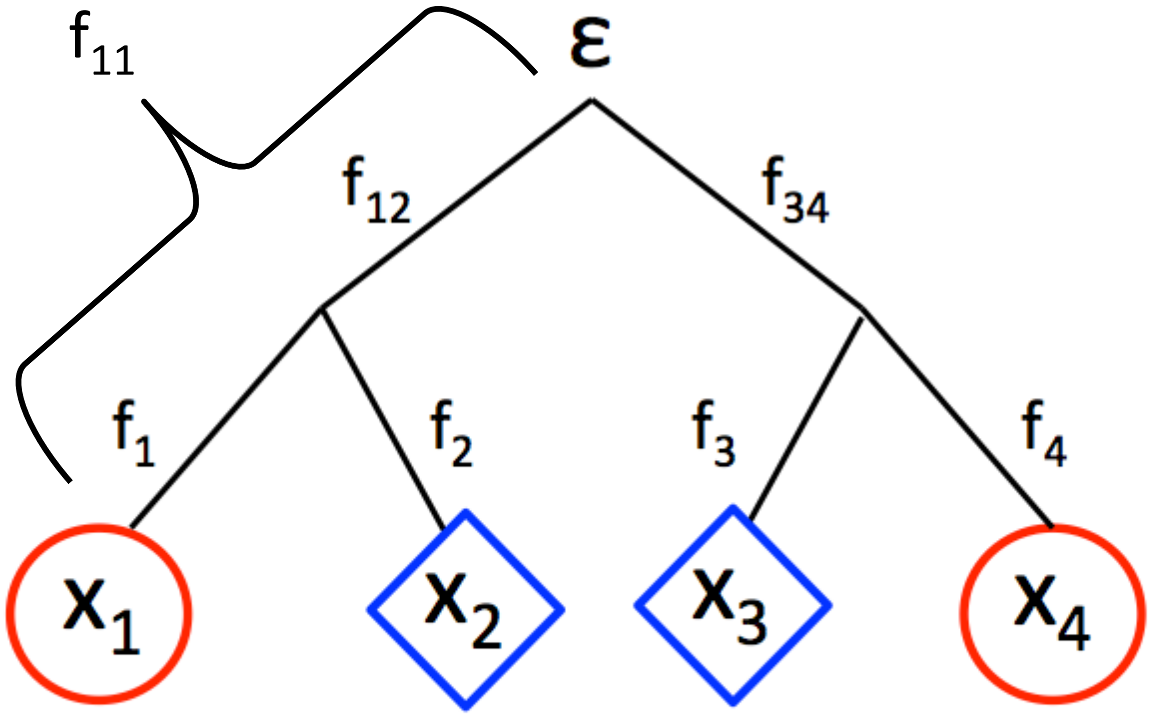
$$f_{ii} = \frac{\mathrm{Var}[x_i]}{\epsilon(1-\epsilon)}$$

Probability a pair of lineages coalesce before reaching the ancestral population

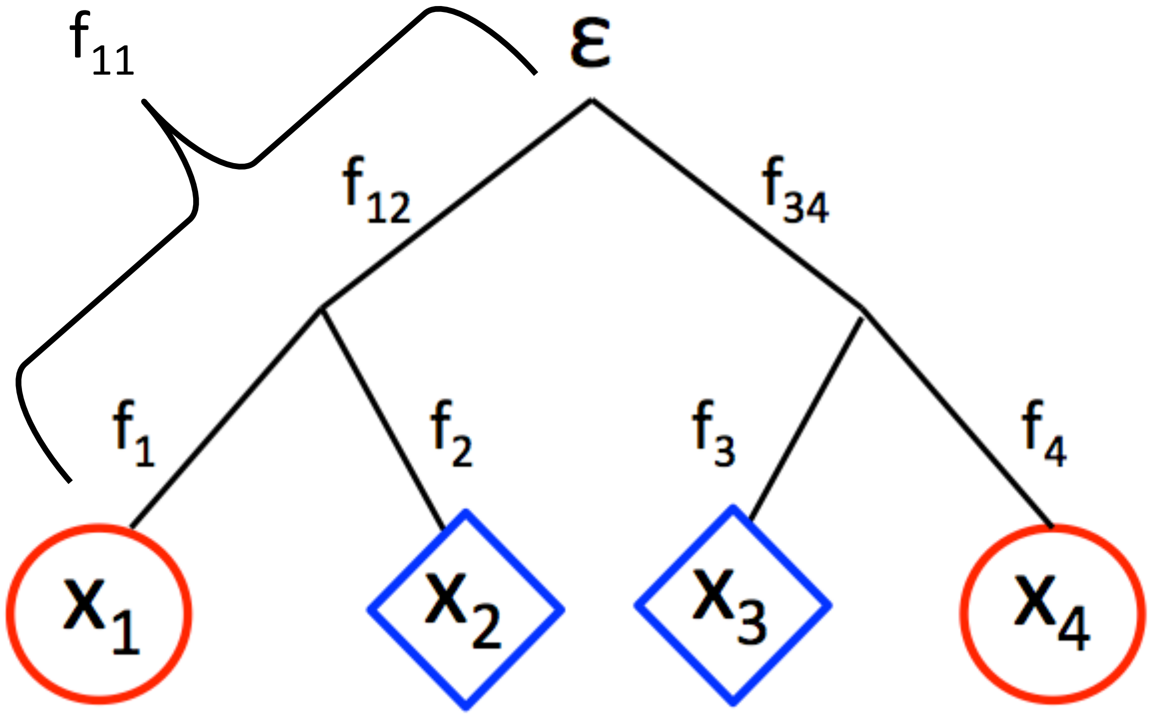
Four population model



Four population model

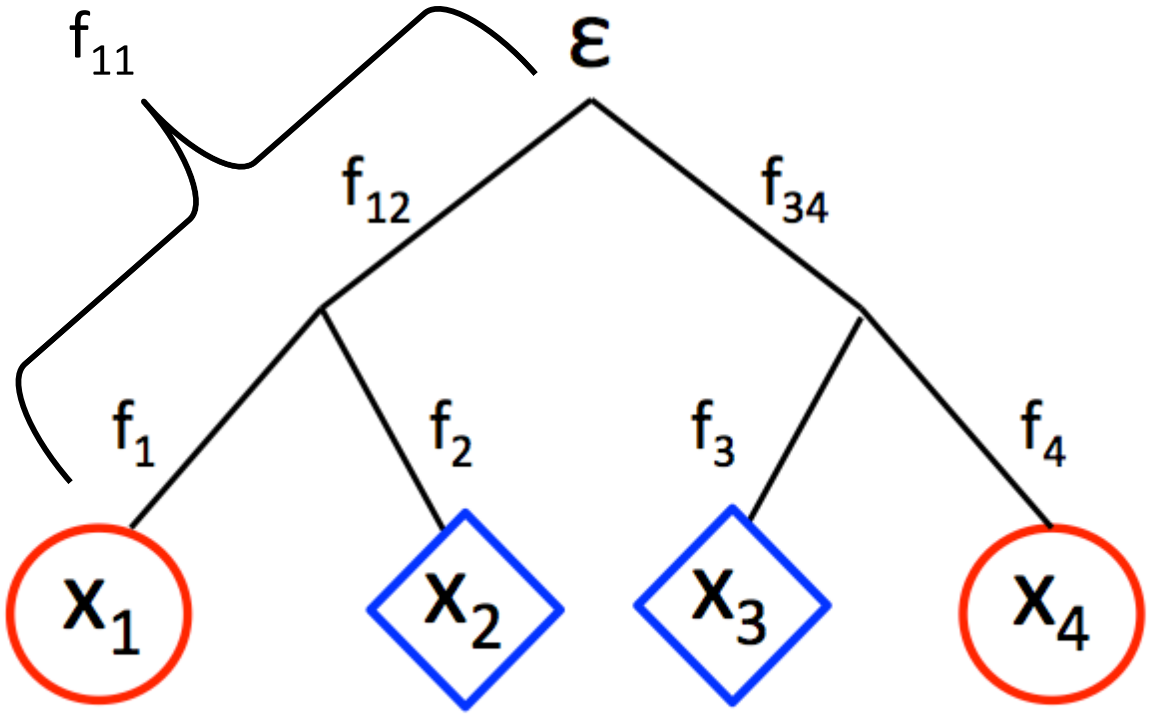


Four population model



$$\mathbf{F} = \begin{bmatrix} f_{11} & f_{12} & 0 & 0 \\ f_{12} & f_{22} & 0 & 0 \\ 0 & 0 & f_{33} & f_{34} \\ 0 & 0 & f_{34} & f_{44} \end{bmatrix}$$

Four population model



$$\mathbf{F} = \begin{bmatrix} f_{11} & f_{12} & 0 & 0 \\ f_{12} & f_{22} & 0 & 0 \\ 0 & 0 & f_{33} & f_{34} \\ 0 & 0 & f_{34} & f_{44} \end{bmatrix}$$

We can model population allele frequencies as multivariate normal

$$\vec{x} \sim \mathcal{N}(\epsilon\vec{1},\, \epsilon(1-\epsilon)\mathbf{F})$$

Cavalli-Sforza et al., 1964; Nicholson et al., 2002;
Weir and Hill, 2002; Coop et al., 2010; Samanta et al., 2009

Effect of hitchhiking

Selection impacts linked neutral allele frequencies

$$\mathbb{E}[\Delta x_1] = 0$$

$$\mathrm{Var}[\Delta x_1] = \epsilon(1 - \epsilon)y^2$$

JH Gillespie, *Genetics* (2000)

Effect of hitchhiking

Selection impacts linked neutral allele frequencies

$$\mathbb{E}[\Delta x_1] = 0$$

$$\mathrm{Var}[\Delta x_1] = \epsilon(1-\epsilon)\enclose{circle}{y^2}$$

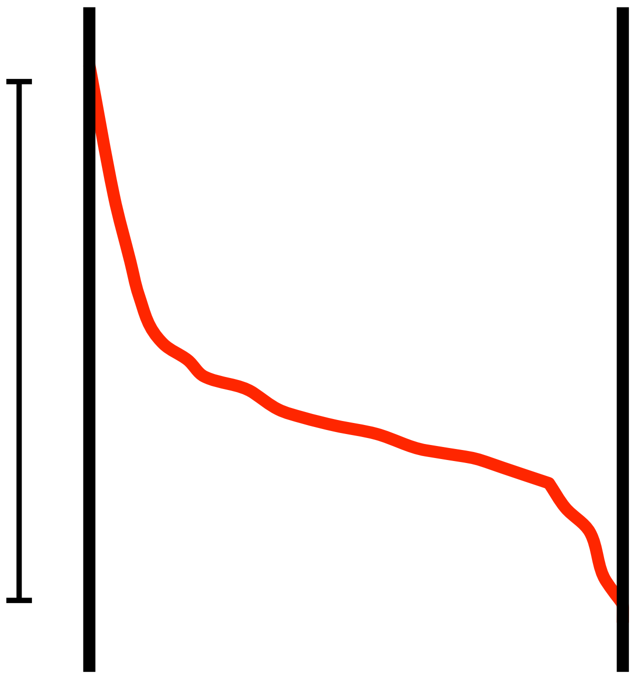
Probability a pair of lineages coalesce during the selective sweep

JH Gillespie, *Genetics* (2000)

Selective sweeps increase probability of coalescing

y^2 = probability two lineages coalesce during the sweep = probability neither recombine off of the beneficial background

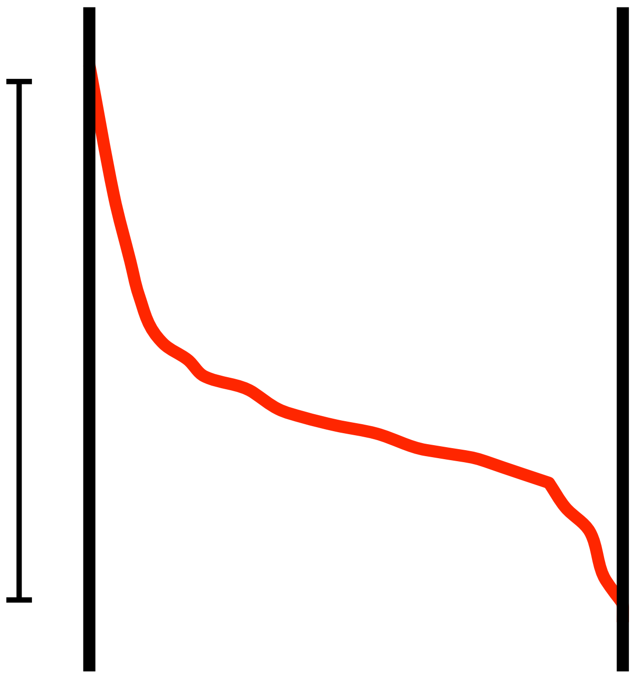
$$t_s \approx \frac{2log(2N_e)}{s}$$



Selective sweeps increase probability of coalescing

y^2 = probability two lineages coalesce during the sweep = probability neither recombine off of the beneficial background

$$t_s \approx \frac{2log(2N_e)}{s}$$

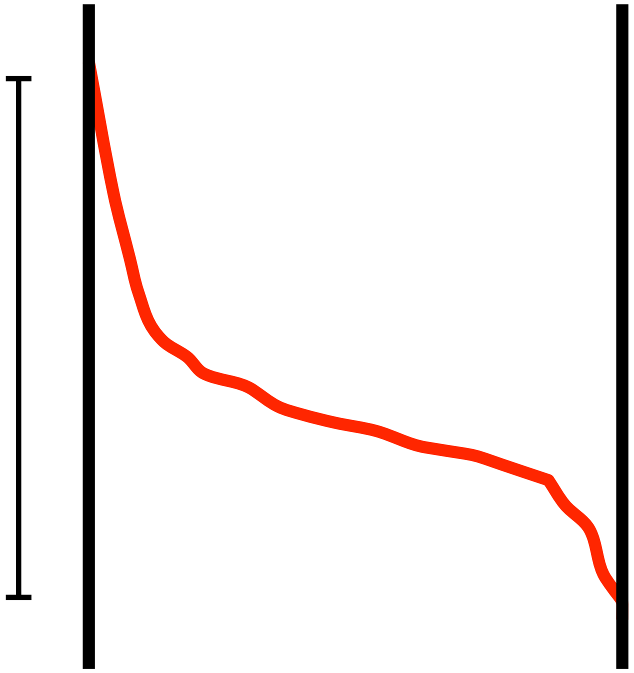


$$y^2 \approx e^{-rt_s}$$

Selective sweeps increase probability of coalescing

y^2 = probability two lineages coalesce during the sweep = probability neither recombine off of the beneficial background

$$t_s \approx \frac{2log(2N_e)}{s}$$



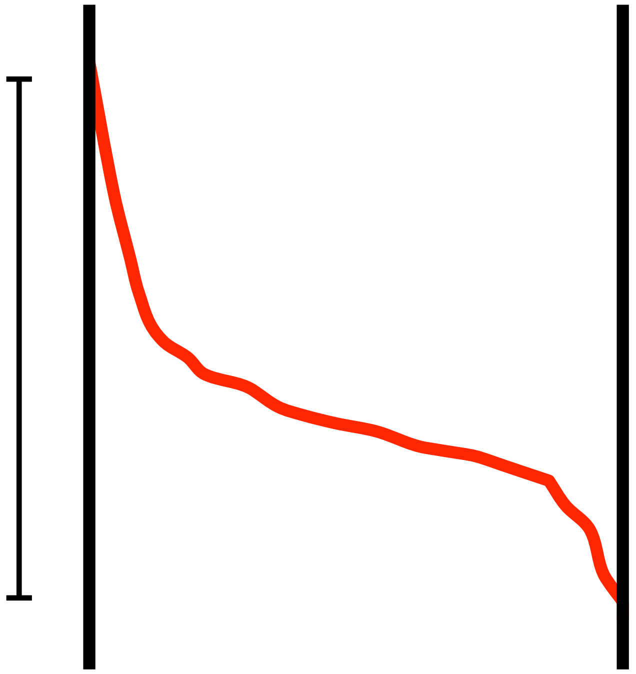
$$y^2 \approx e^{-rt_s}$$

$$\mathrm{Var}[\Delta x_i] \approx \epsilon(1-\epsilon)\left(y^2 + (1-y^2)f_{ii}\right)$$

Selective sweeps increase probability of coalescing

y^2 = probability two lineages coalesce during the sweep = probability neither recombine off of the beneficial background

$$t_s \approx \frac{2log(2N_e)}{s}$$



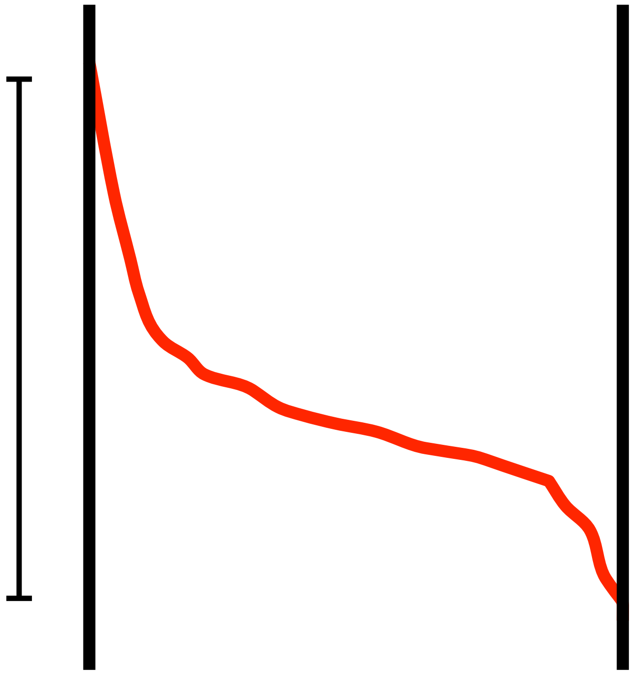
$$y^2 \approx e^{-rt_s}$$

$$\mathrm{Var}[\Delta x_i] \approx \epsilon(1-\epsilon)\left(\underline{y^2} + (1-y^2)f_{ii}\right)$$

Selective sweeps increase probability of coalescing

y^2 = probability two lineages coalesce during the sweep = probability neither recombine off of the beneficial background

$$t_s \approx \frac{2log(2N_e)}{s}$$



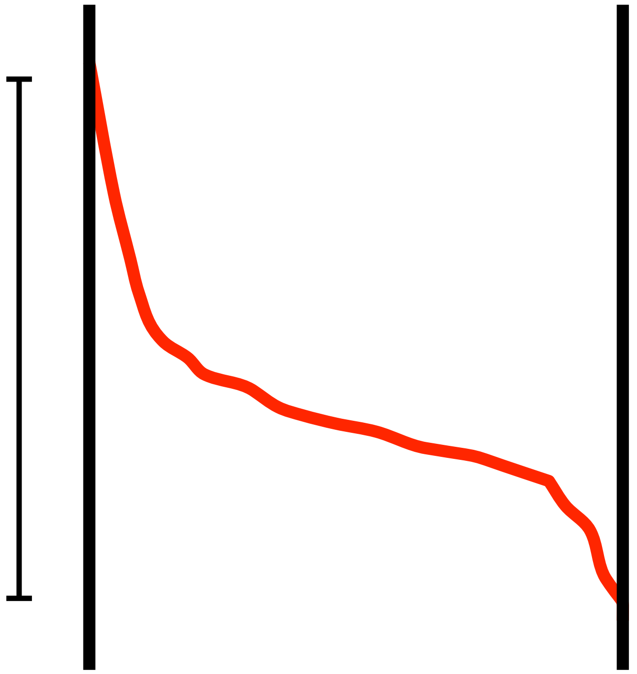
$$y^2 \approx e^{-rt_s}$$

$$\mathrm{Var}[\Delta x_i] \approx \epsilon(1-\epsilon)\left(y^2 + \underline{(1-y^2)}f_{ii}\right)$$

Selective sweeps increase probability of coalescing

y^2 = probability two lineages coalesce during the sweep = probability neither recombine off of the beneficial background

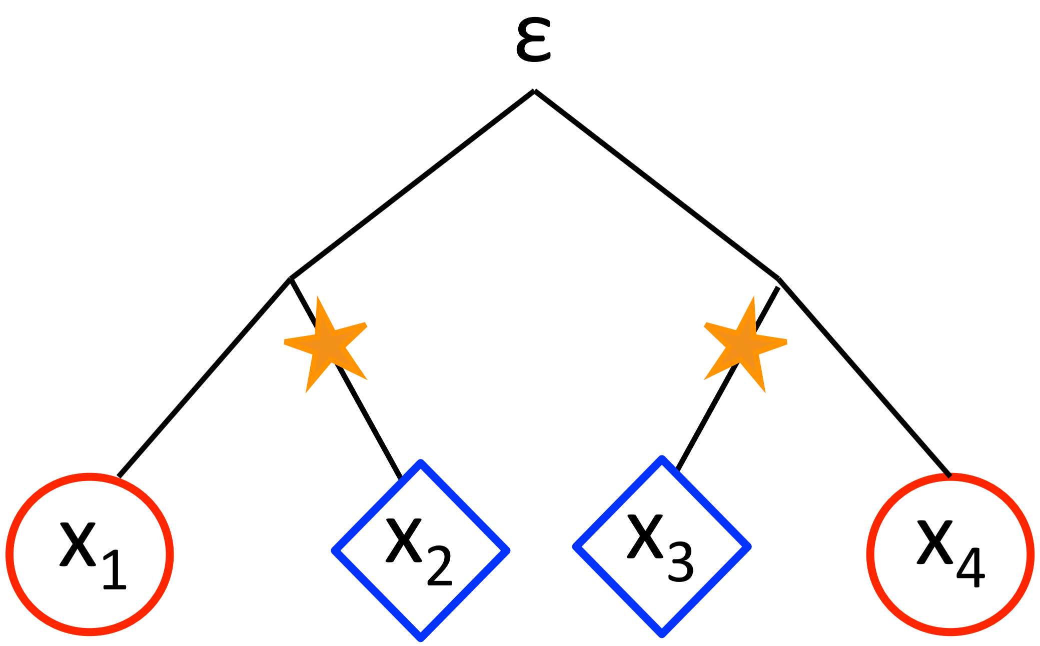
$$t_s \approx \frac{2log(2N_e)}{s}$$



$$y^2 \approx e^{-rt_s}$$

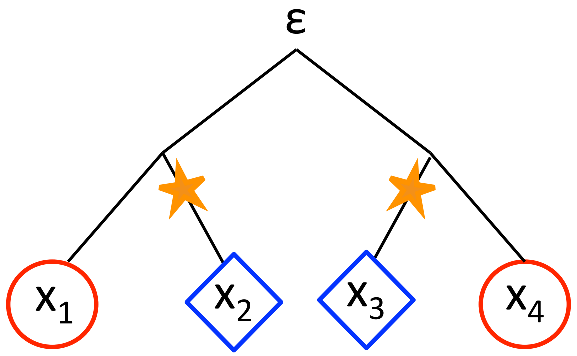
$$\mathrm{Var}[\Delta x_i] \approx \epsilon(1-\epsilon)\left(y^2 + (1-y^2)\underline{f_{ii}}\right)$$

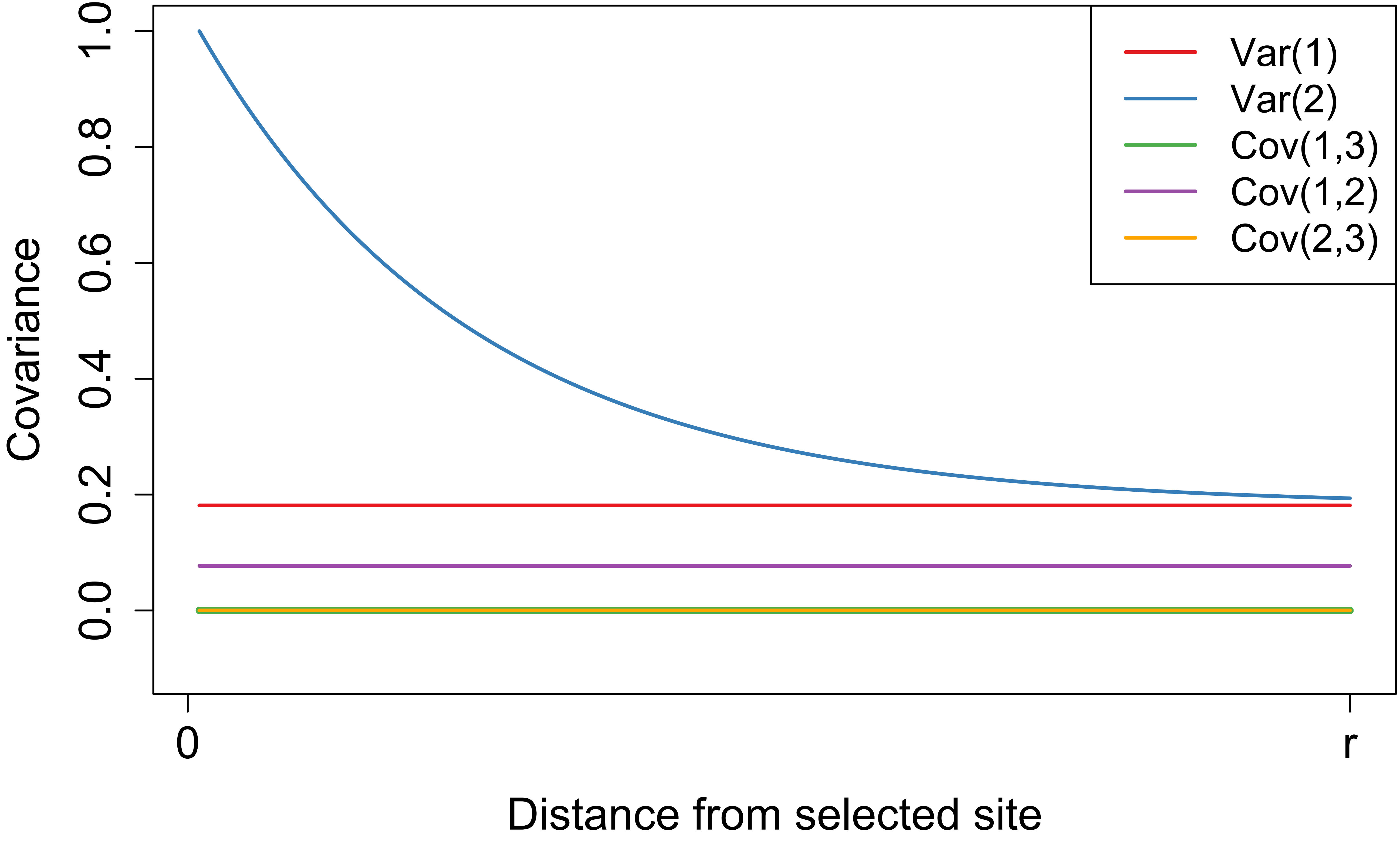
1. Independent sweep model



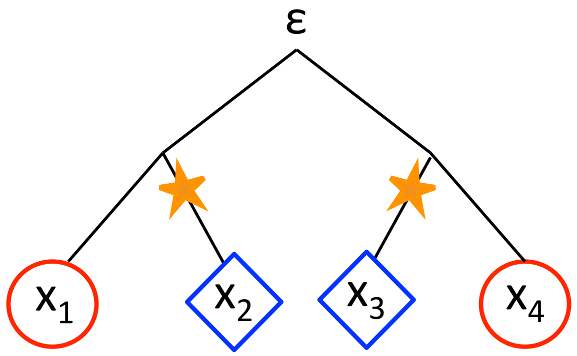
$$\mathbf{F} + \Omega = \begin{bmatrix} f_{11} & f_{12} & 0 & 0 \\ f_{12} & f_{22} + y^2(1 - f_{22}) & 0 & 0 \\ 0 & 0 & f_{33} + y^2(1 - f_{33}) & f_{34} \\ 0 & 0 & f_{34} & f_{44} \end{bmatrix}$$

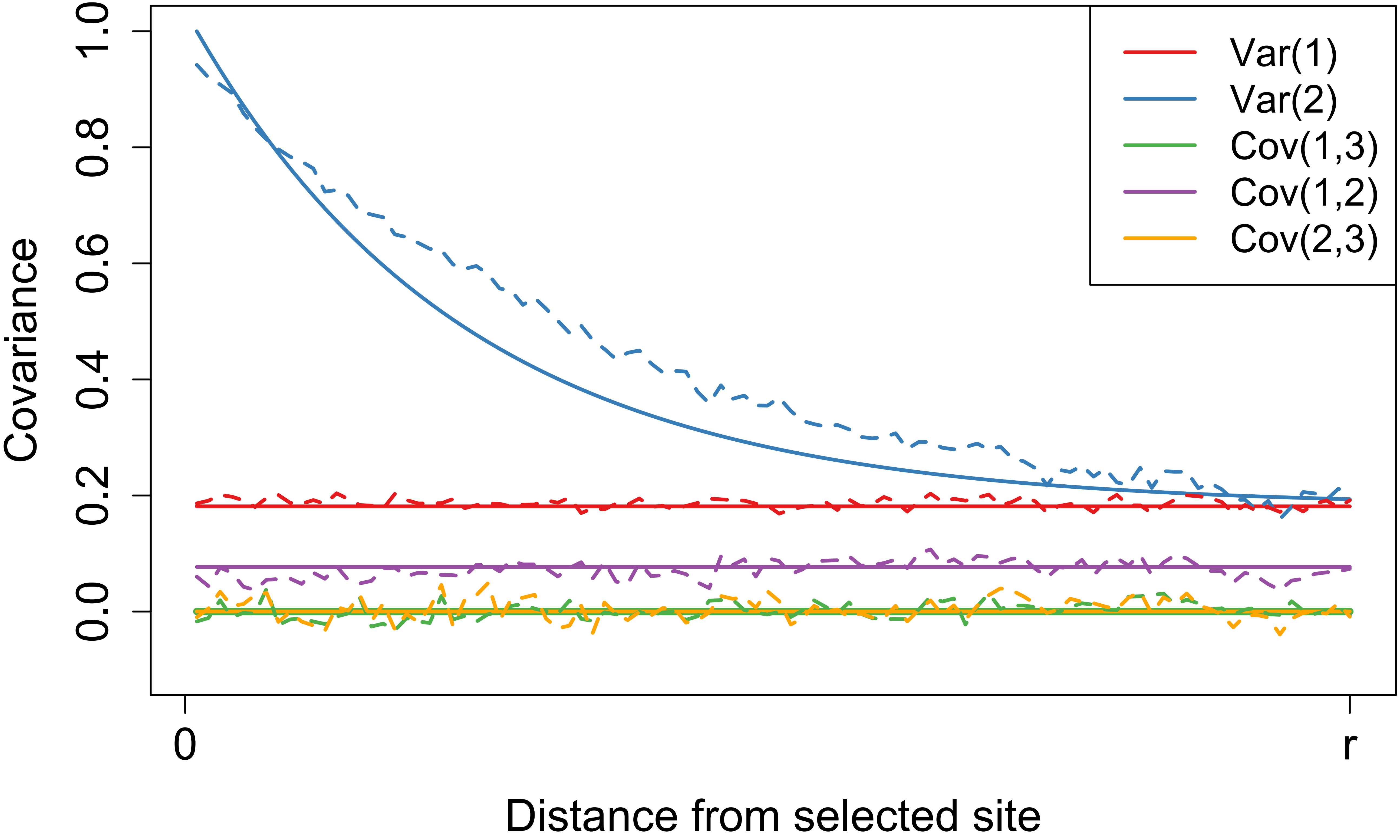
Independent mutation model variances and covariances



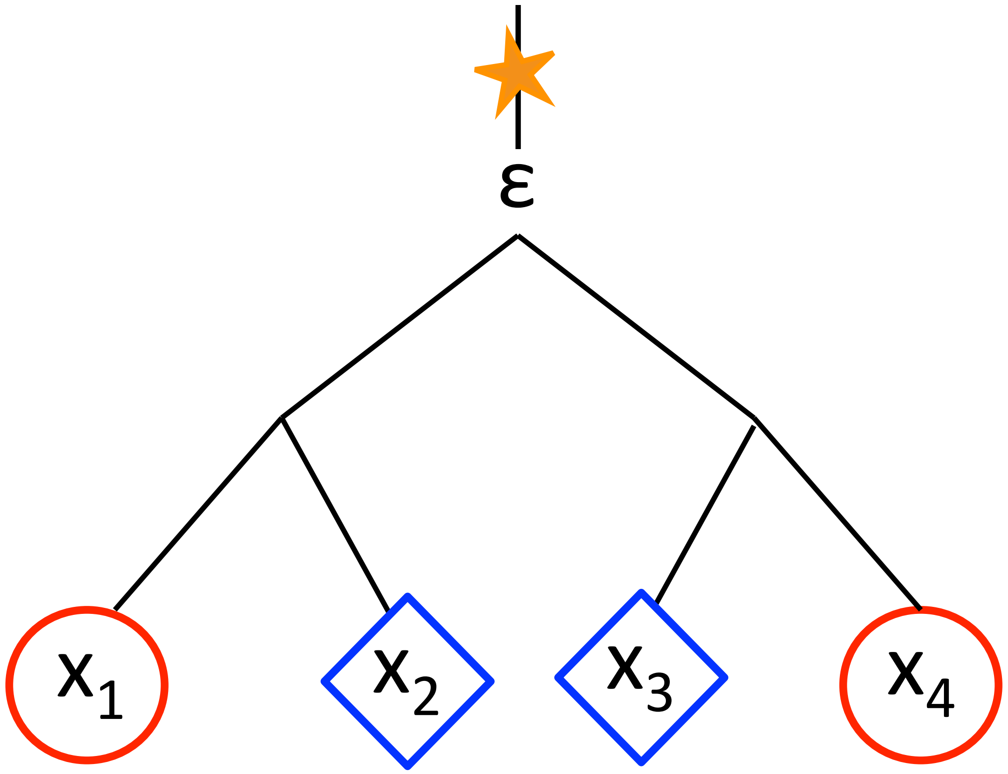


Independent mutation model variances and covariances

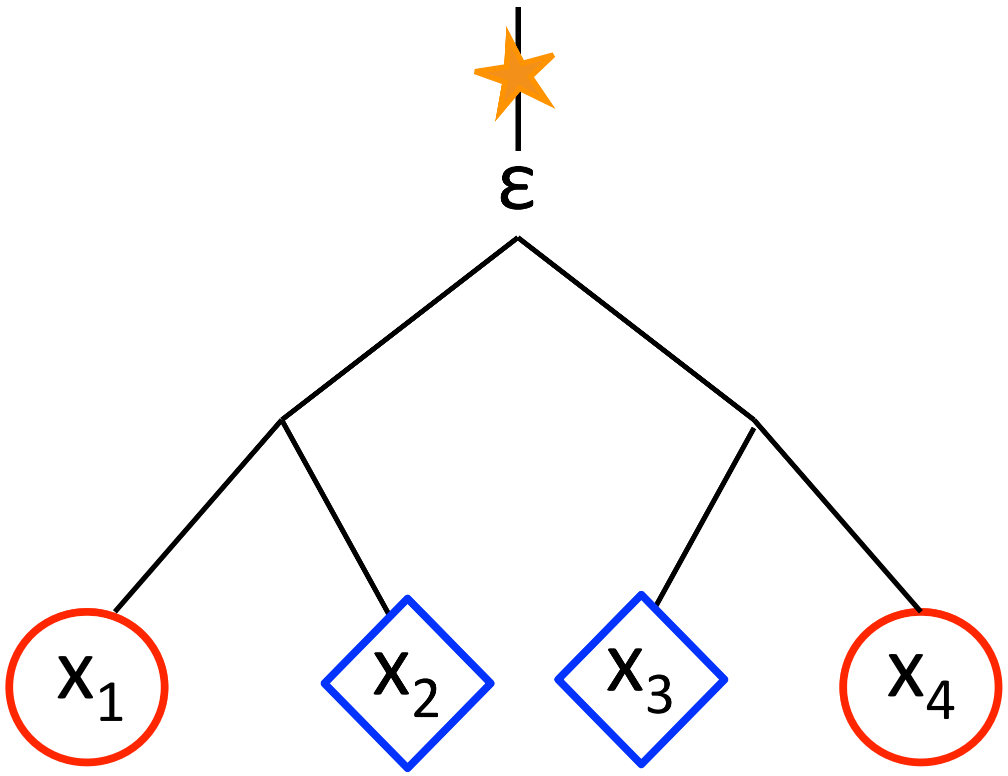


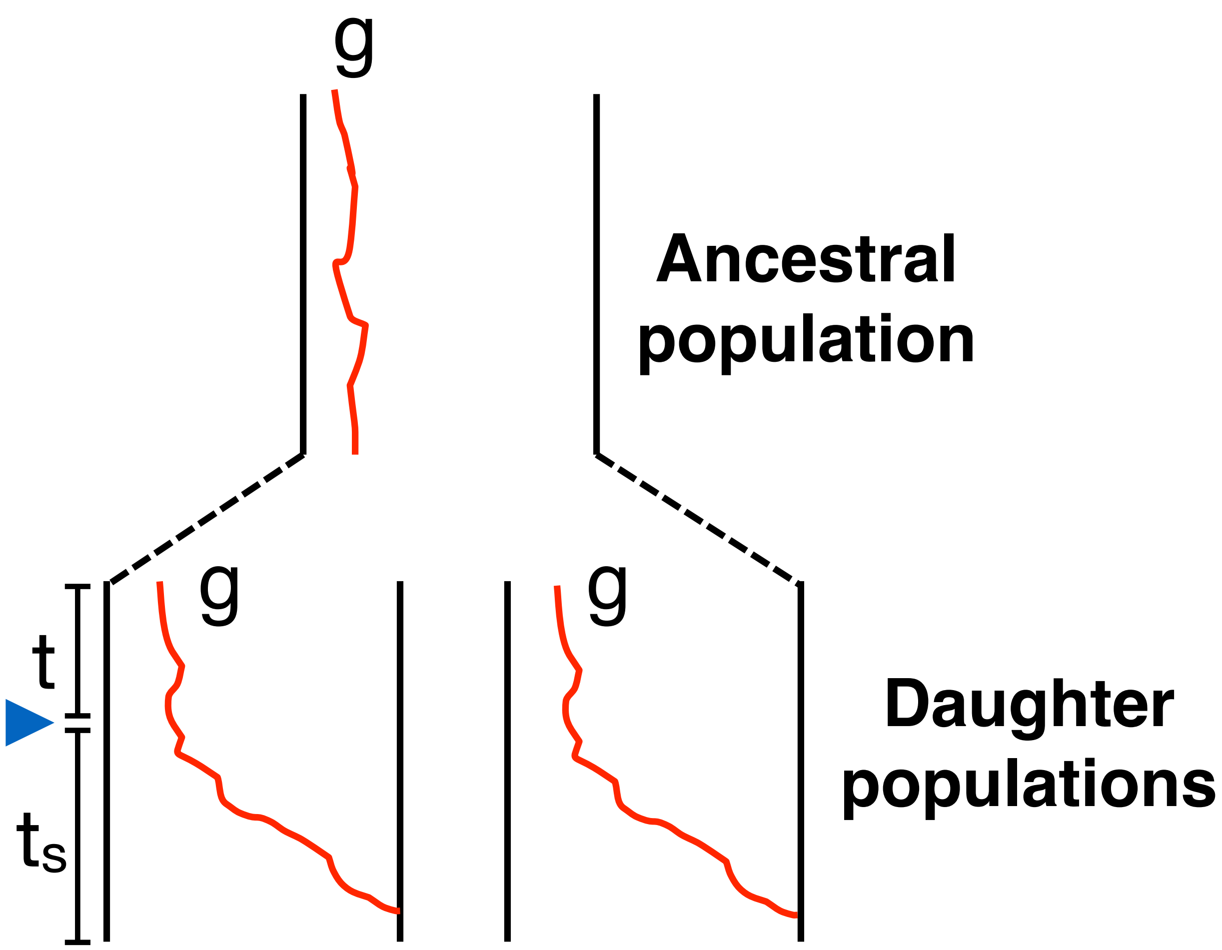


2. Shared ancestral standing variation

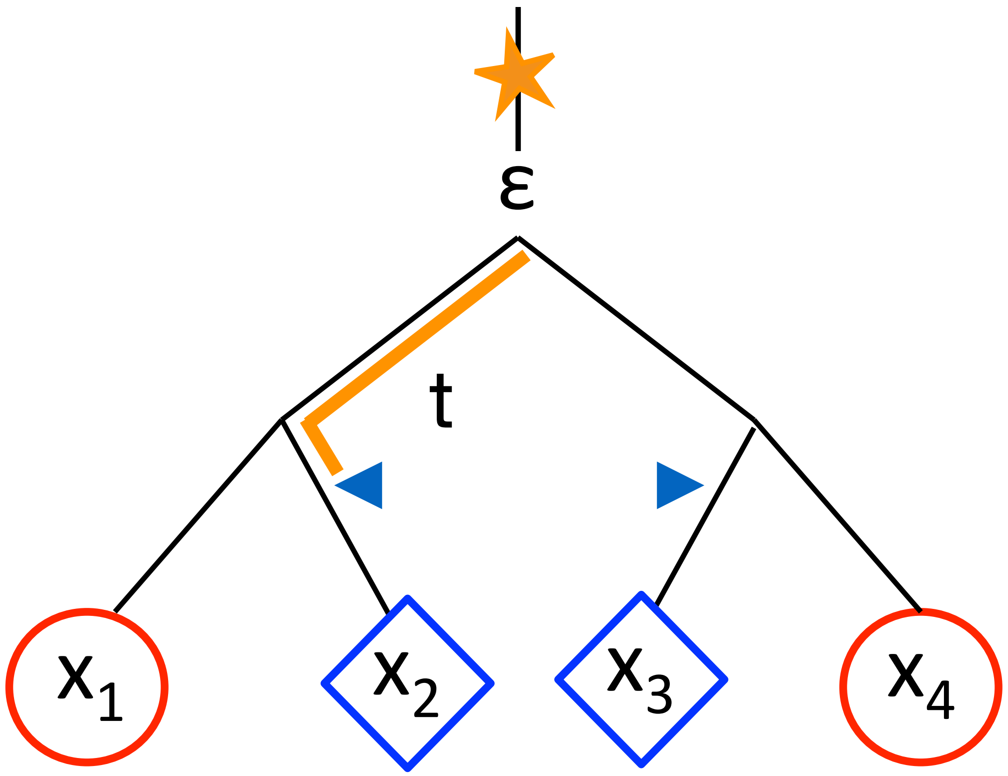


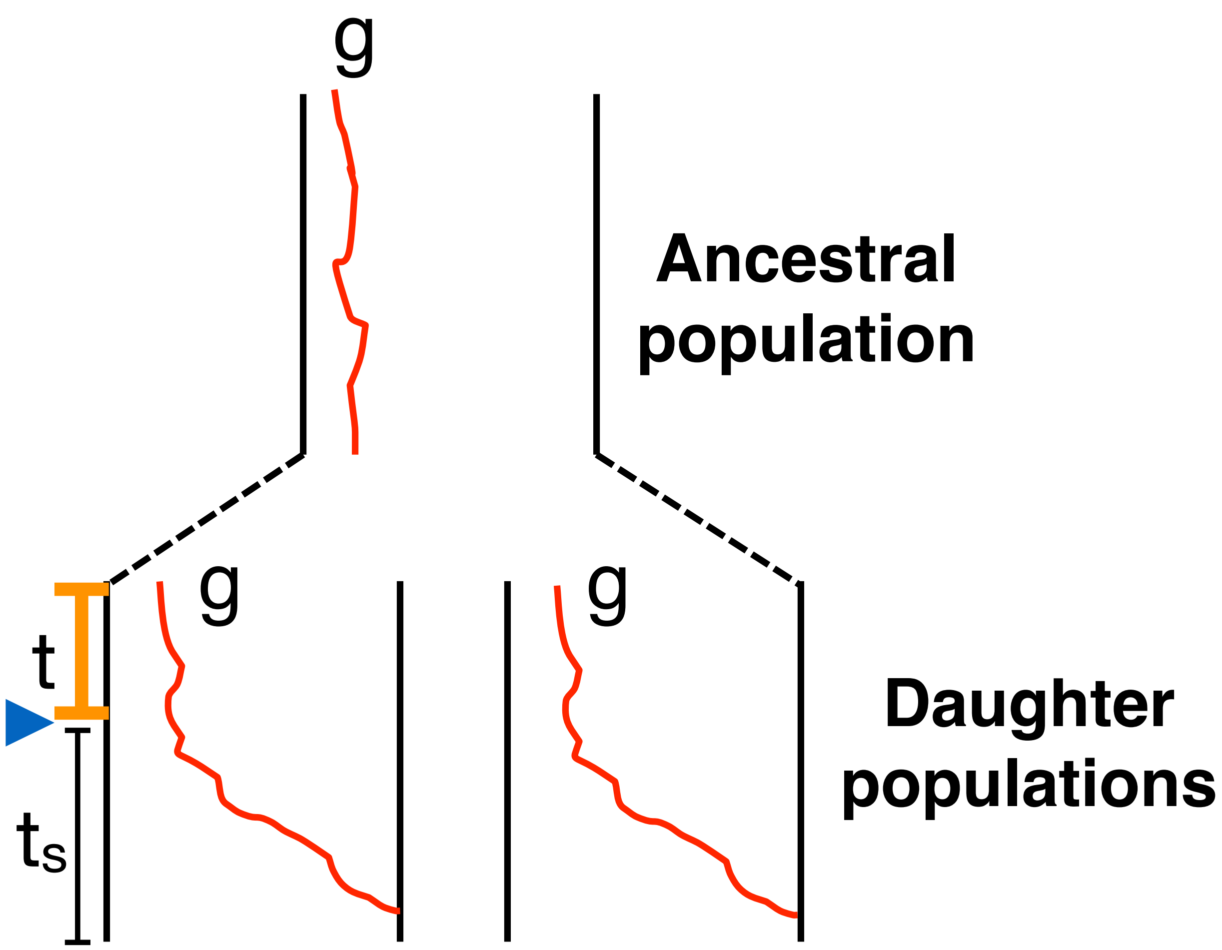
2. Shared ancestral standing variation



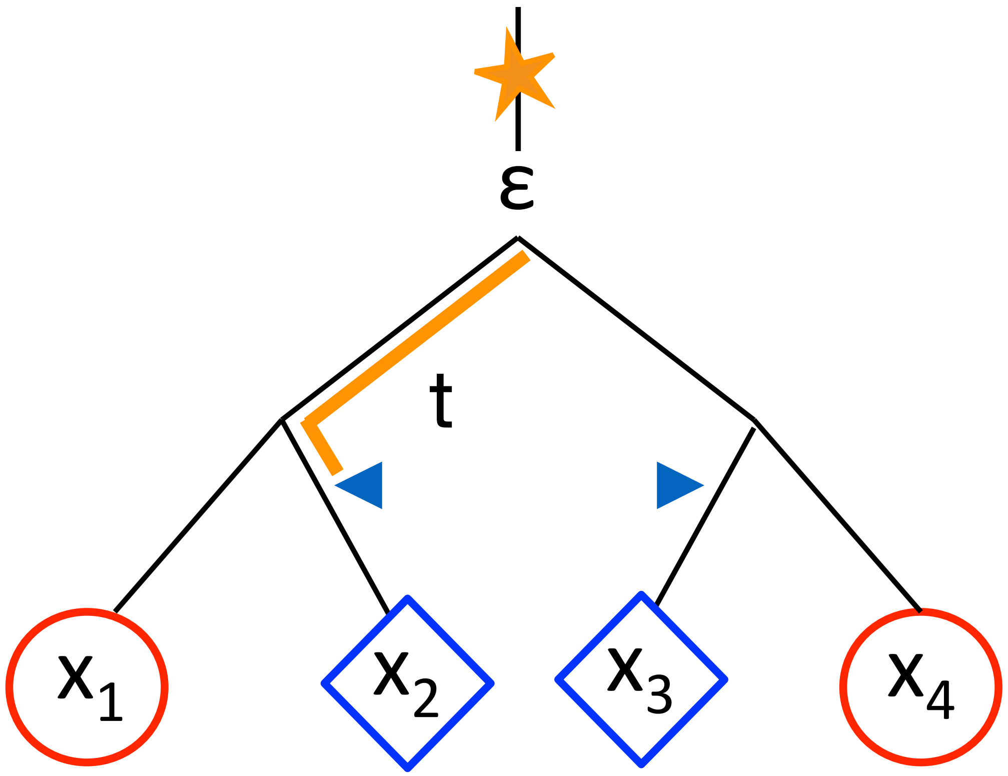


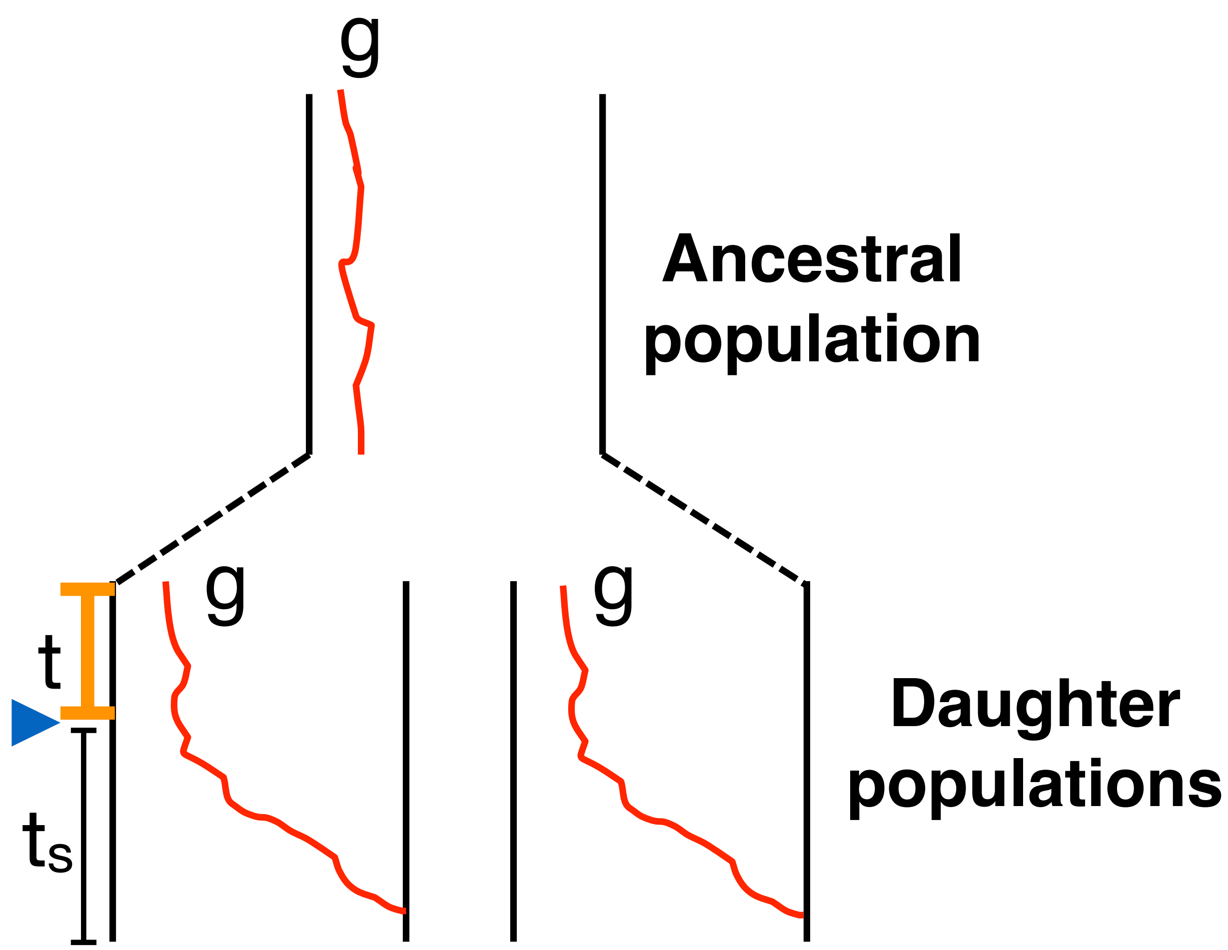
2. Shared ancestral standing variation





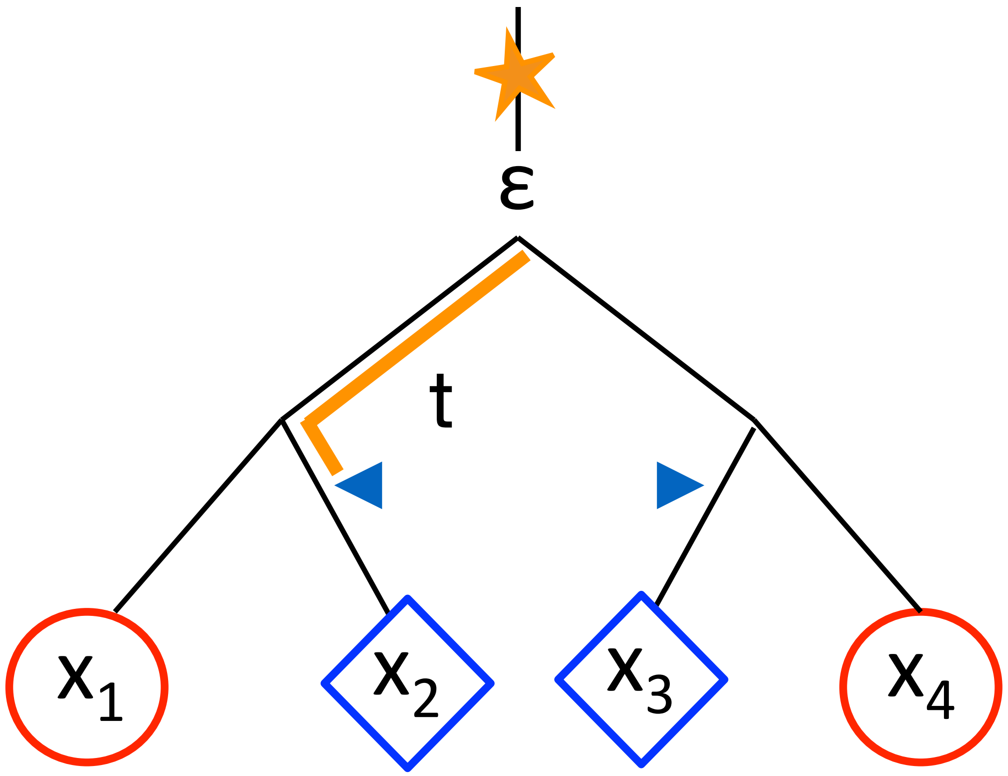
2. Shared ancestral standing variation

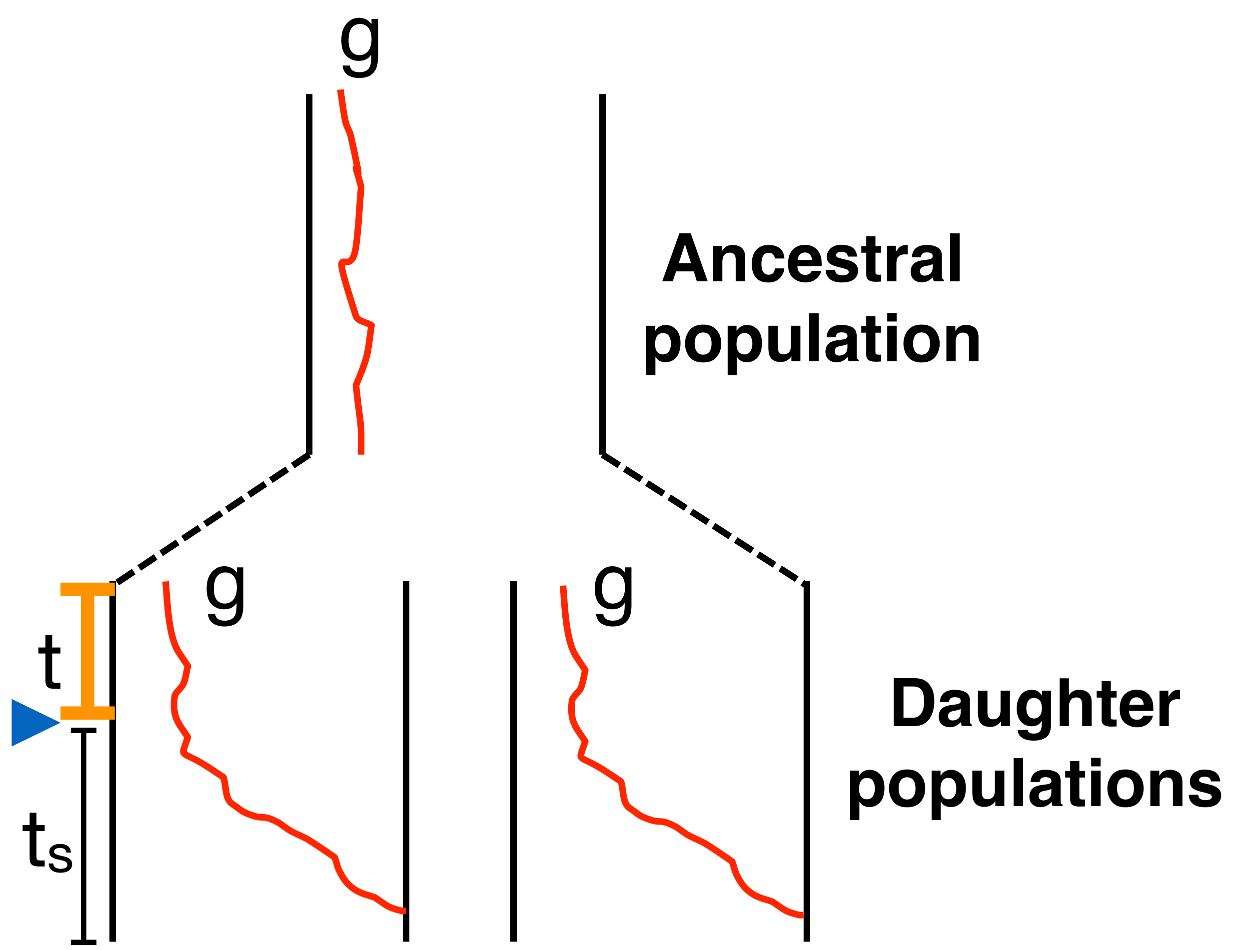




$$\mathrm{Var}[\Delta x_i] \approx \epsilon(1-\epsilon)\big(f_{ii} + y^2 h_1(r,g)(1-f_{ii})\big)$$

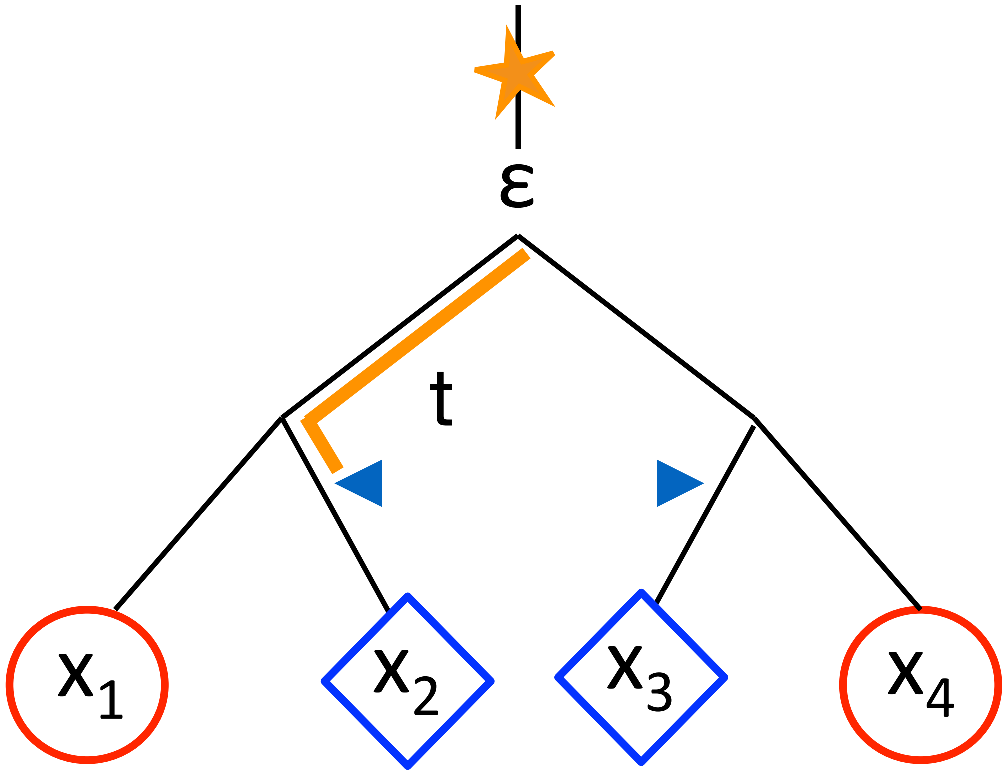
2. Shared ancestral standing variation

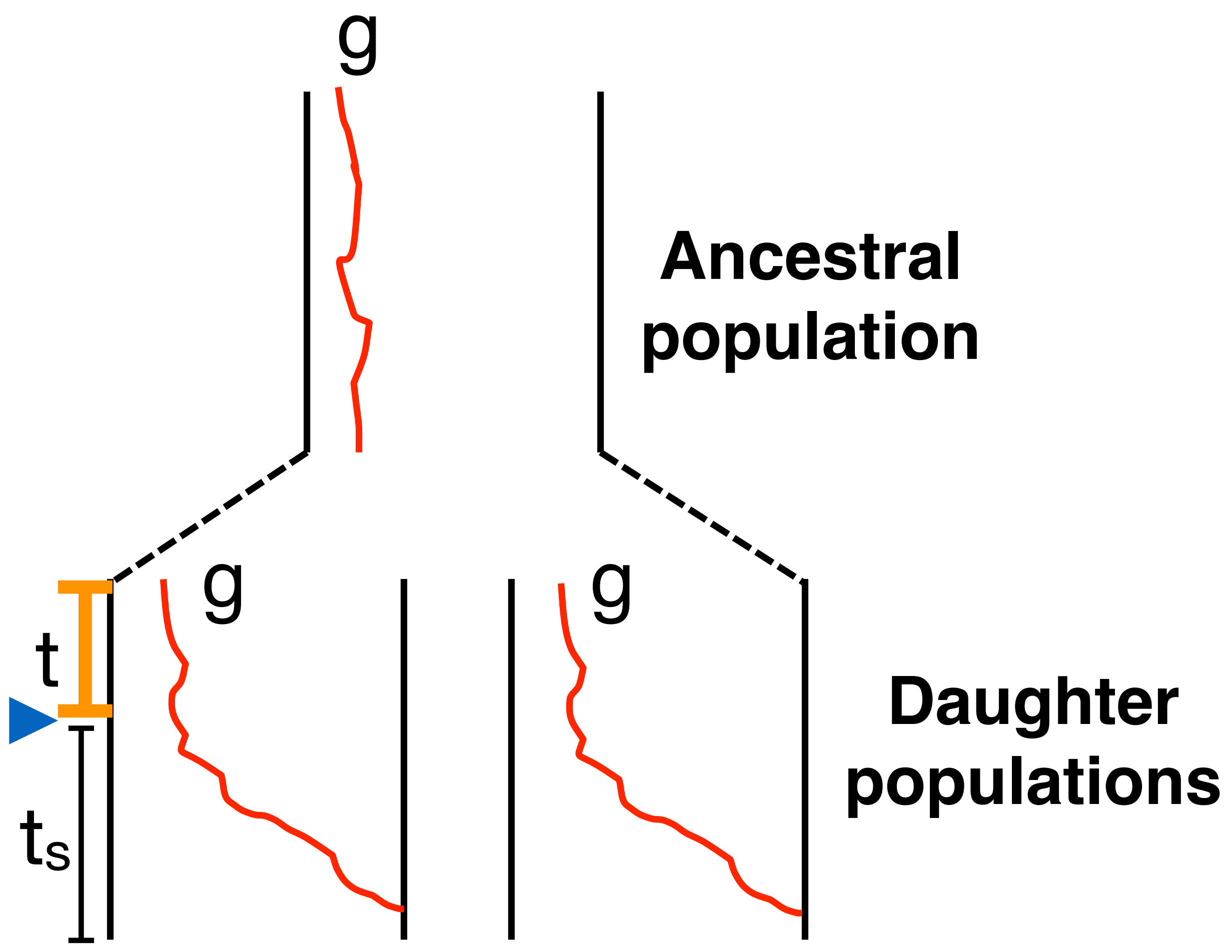




$$\mathrm{Var}[\Delta x_i] \approx \epsilon(1-\epsilon)\Big(f_{ii} + \underline{y^2} h_1(r,g)(1-f_{ii})\Big)$$

2. Shared ancestral standing variation

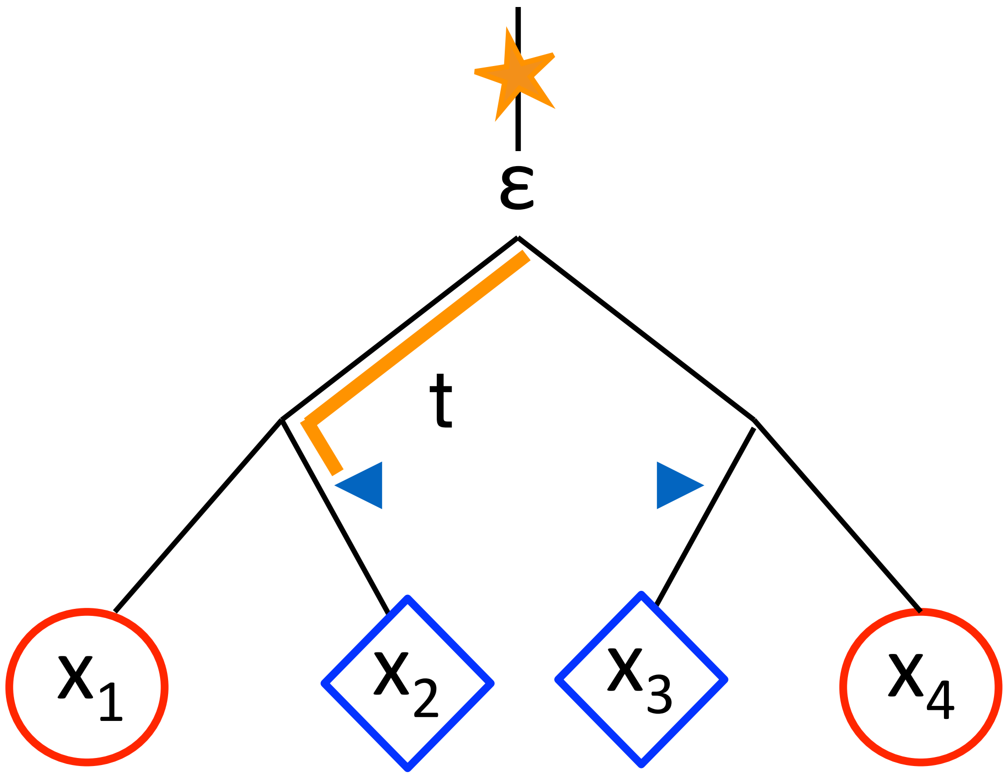


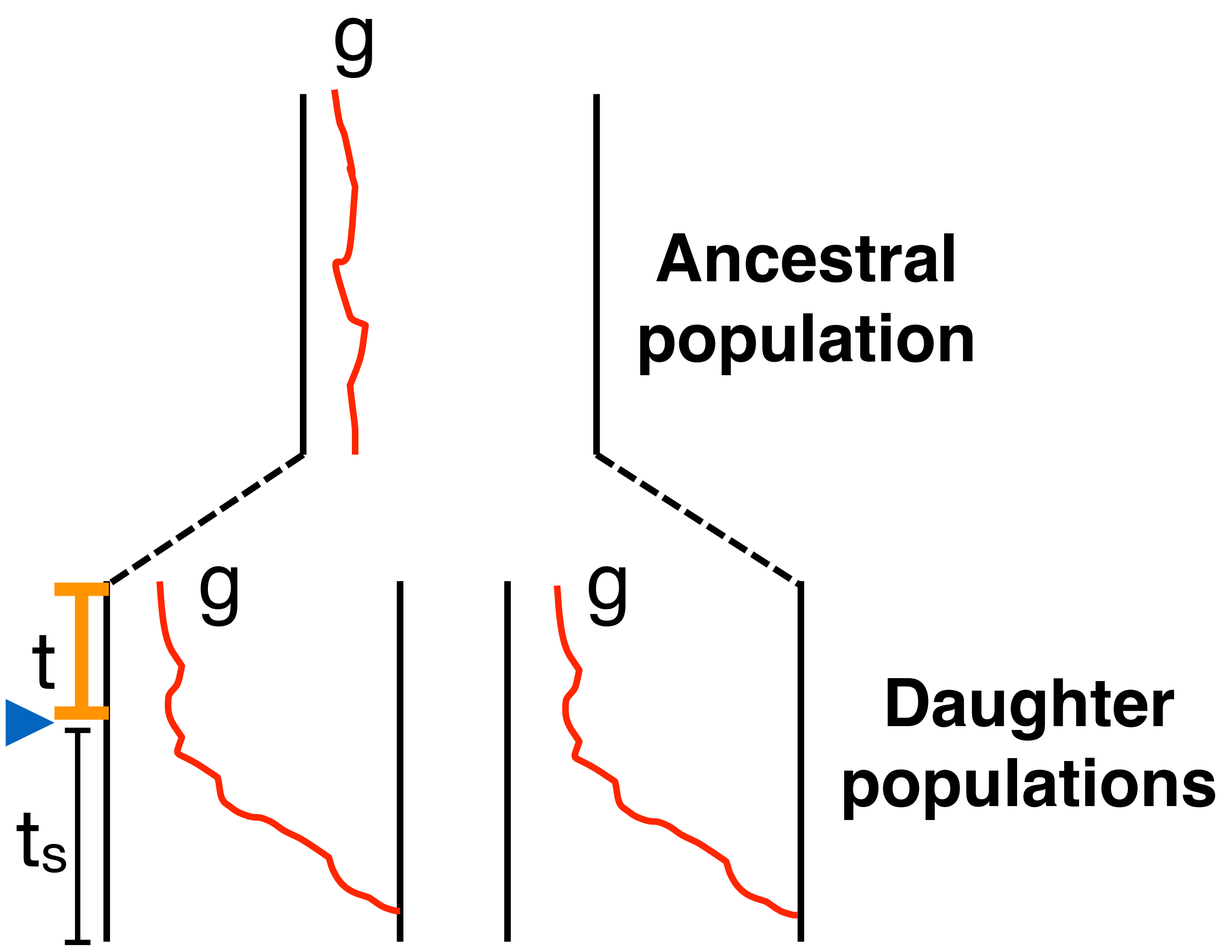


$$\mathrm{Var}[\Delta x_i] \approx \epsilon(1-\epsilon)\Big(f_{ii} + y^2 \underline{h_1(r,g)}(1-f_{ii})\Big)$$

Berg and Coop 2015
Poster 130B

2. Shared ancestral standing variation



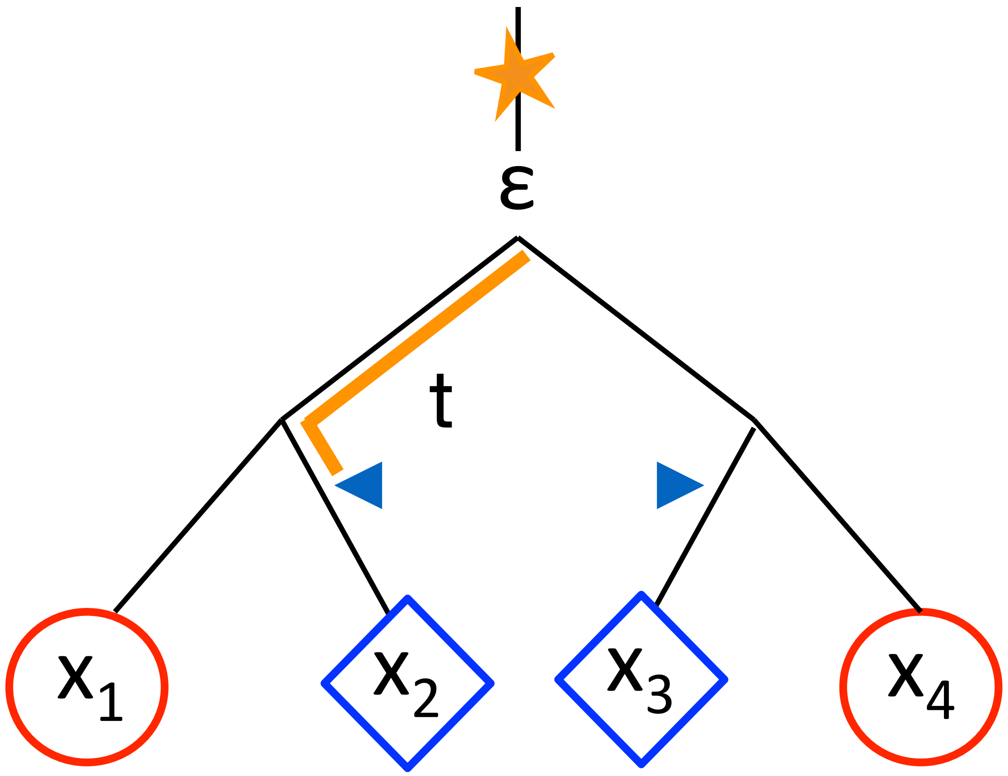


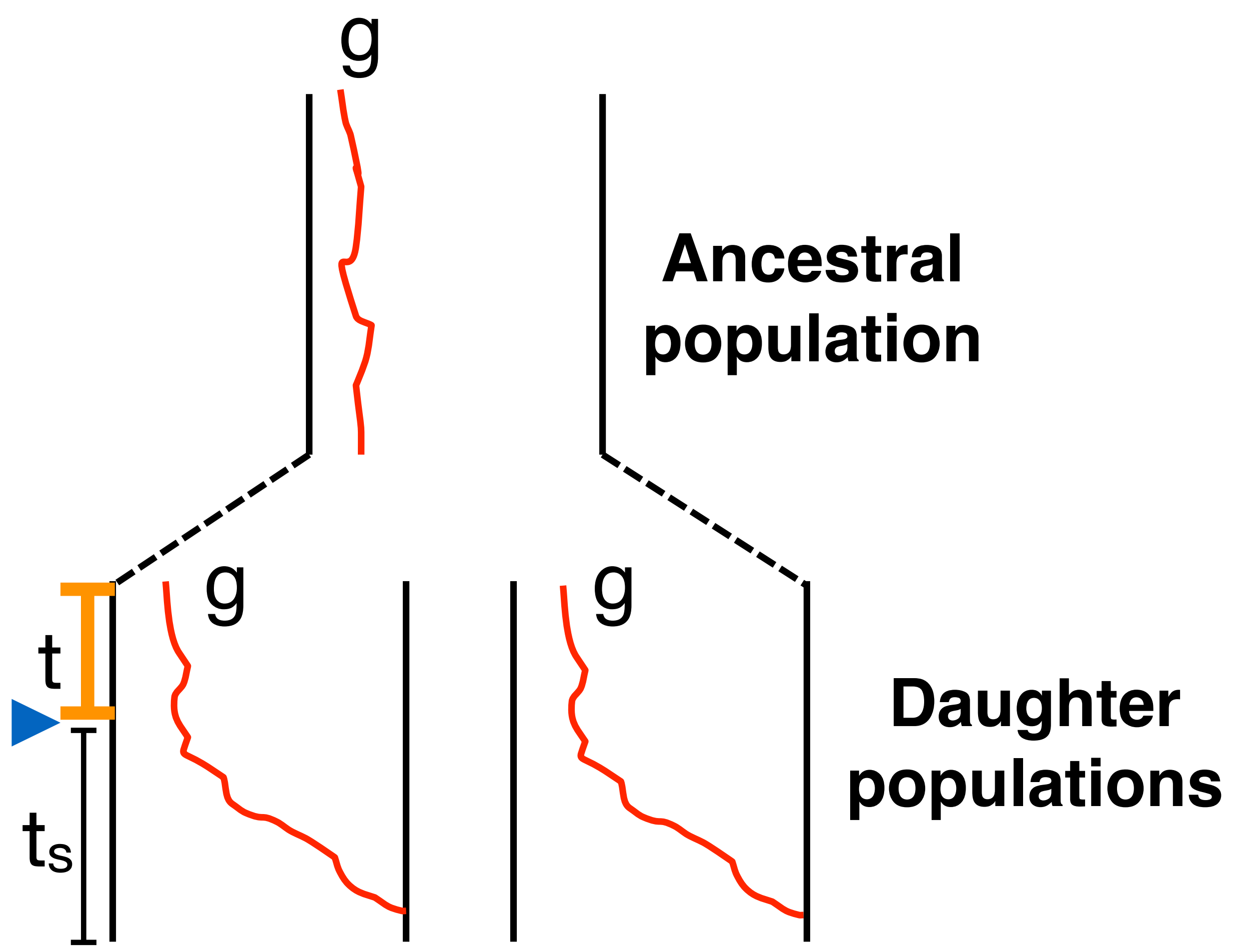
$$\mathrm{Var}[\Delta x_i] \approx \epsilon(1-\epsilon)\big(f_{ii} + y^2 h_1(r,g)(1-f_{ii})\big)$$

$$\mathrm{Cov}[\Delta x_2, \Delta x_3] \approx \epsilon(1-\epsilon)\big(y^2 h_2(r,g,t)\big)$$

Berg and Coop 2015
Poster 130B

2. Shared ancestral standing variation



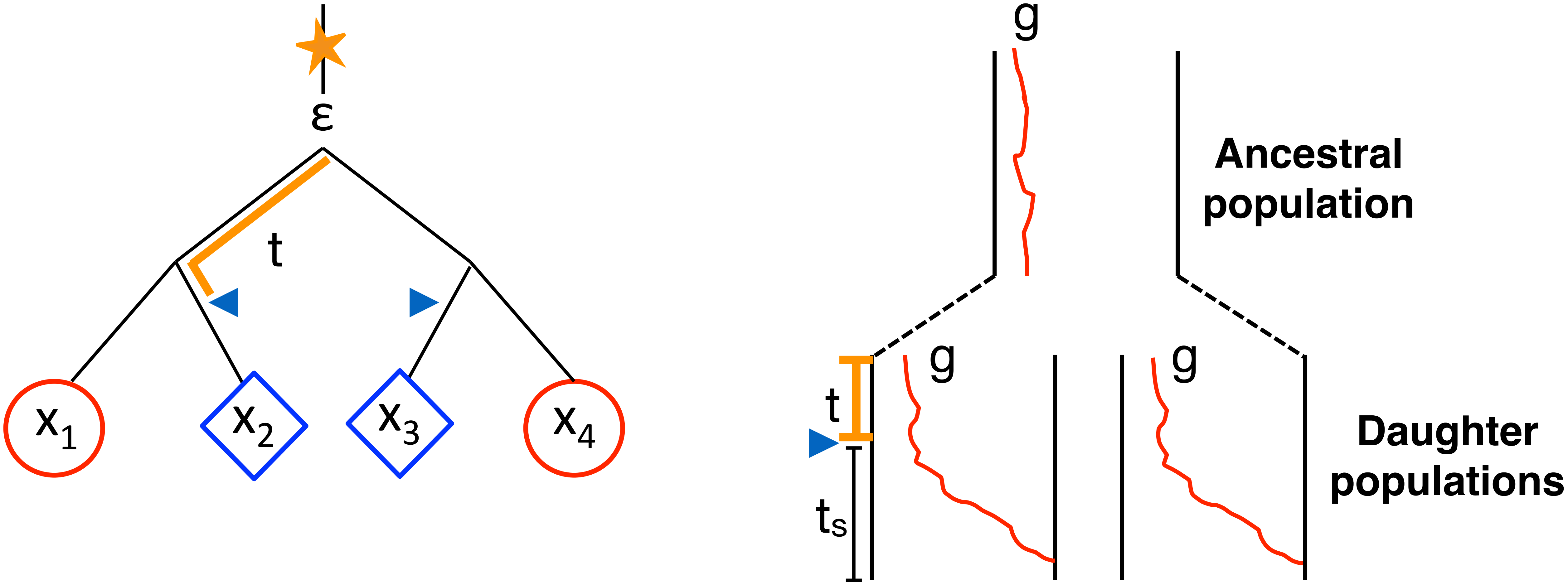


$$\mathrm{Var}[\Delta x_i] \approx \epsilon(1-\epsilon)\big(f_{ii} + y^2 h_1(r,g)(1-f_{ii})\big)$$

$$\mathrm{Cov}[\Delta x_2, \Delta x_3] \approx \epsilon(1-\epsilon)\big(\underline{y^2} h_2(r,g,t)\big)$$

Berg and Coop 2015
Poster 130B

2. Shared ancestral standing variation

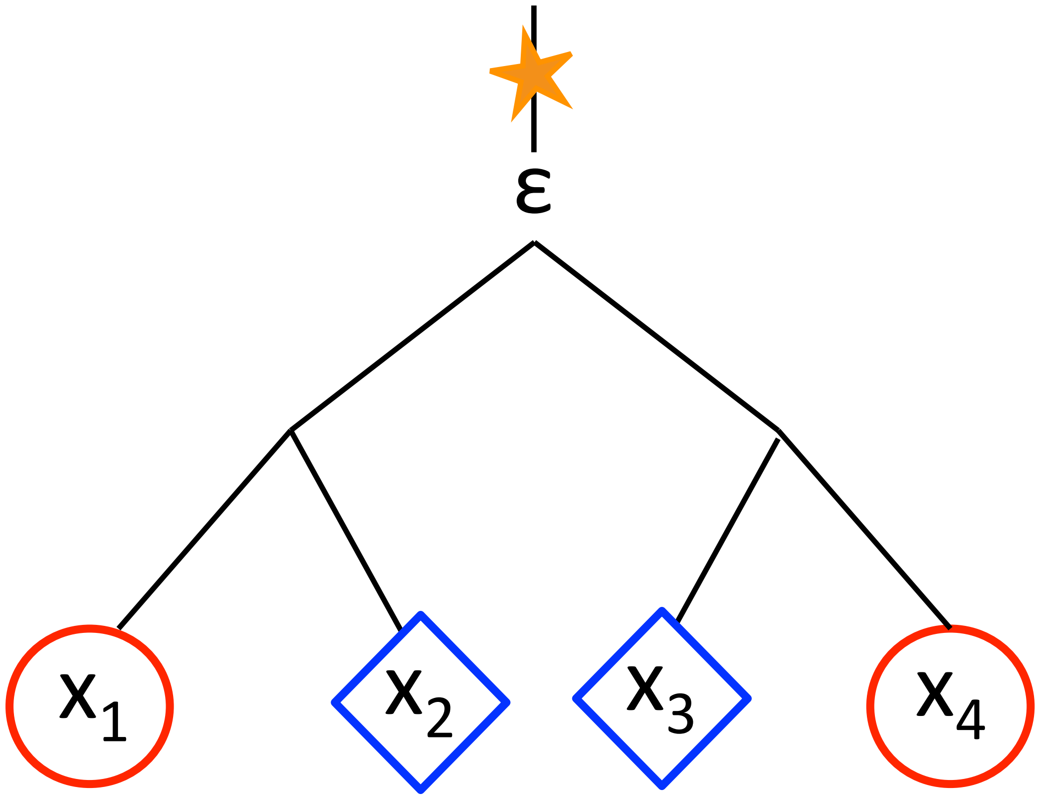


$$\mathrm{Var}[\Delta x_i] \approx \epsilon(1-\epsilon)\big(f_{ii} + y^2 h_1(r,g)(1-f_{ii})\big)$$

$$\mathrm{Cov}[\Delta x_2, \Delta x_3] \approx \epsilon(1-\epsilon)\big(y^2 \underline{h_2(r,g,t)}\big)$$

Berg and Coop 2015
Poster 130B

2. Shared ancestral standing variation

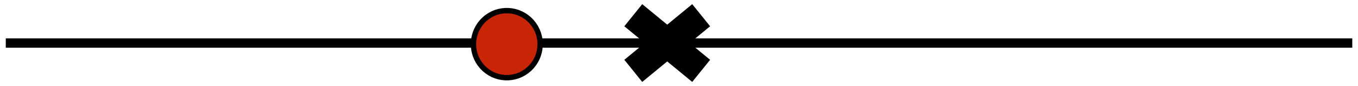


$$\mathbf{F}+\Omega=\begin{bmatrix} f_{11} & f_{12} & 0 & 0 \\ f_{12} & f_{22}+y^2h_1(1-f_{22}) & y^2h_2 & 0 \\ 0 & y^2h_2 & f_{33}+y^2h_1(1-f_{33}) & f_{34} \\ 0 & 0 & f_{34} & f_{44} \end{bmatrix}$$

Inference method

Given location of selected site and model, can calculate probability of observed neutral allele frequencies a given distance from selected site.

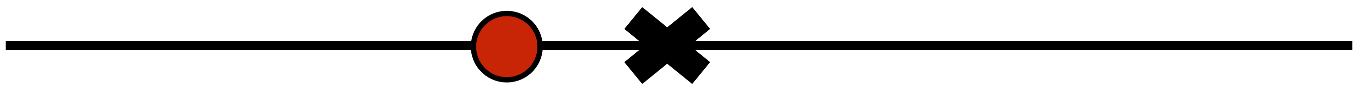
$$P(D_l \mid r_l, \mathbf{F}, \Theta_M, M) \approx \mathcal{N}(D_l \mid \epsilon_l,\ \mathbf{F} + \Omega_M(r_l, \mathbf{F}, \Theta_M))$$



Inference method

Given location of selected site and model, can calculate probability of observed neutral allele frequencies a given distance from selected site.

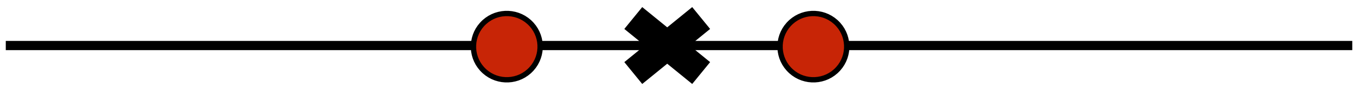
$$P(D_l \mid r_l, \mathbf{F}, \Theta_M, M) \approx \mathcal{N}(D_l \mid \epsilon_l,\, \mathbf{F} + \Omega_M(r_l, \mathbf{F}, \Theta_M))$$



Inference method

Given location of selected site and model, can calculate probability of observed neutral allele frequencies a given distance from selected site.

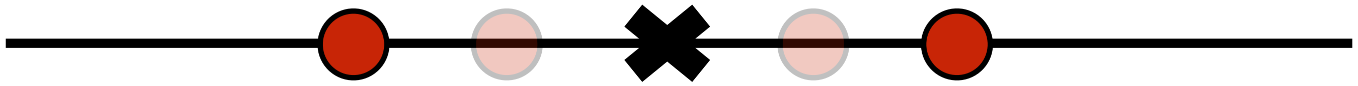
$$P(D_l \mid r_l, \mathbf{F}, \Theta_M, M) \approx \mathcal{N}(D_l \mid \epsilon_l, \ \mathbf{F} + \Omega_M(r_l, \mathbf{F}, \Theta_M))$$



Inference method

Given location of selected site and model, can calculate probability of observed neutral allele frequencies a given distance from selected site.

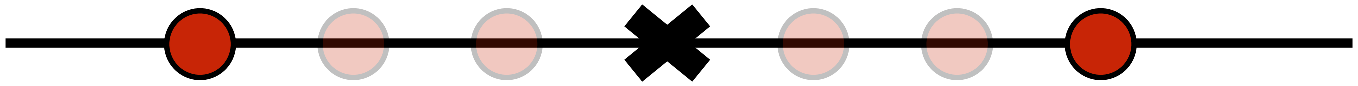
$$P(D_l \mid r_l, \mathbf{F}, \Theta_M, M) \approx \mathcal{N}(D_l \mid \epsilon_l,\, \mathbf{F} + \Omega_M(r_l, \mathbf{F}, \Theta_M))$$



Inference method

Given location of selected site and model, can calculate probability of observed neutral allele frequencies a given distance from selected site.

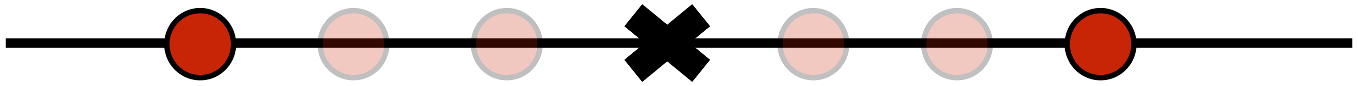
$$P(D_l \mid r_l, \mathbf{F}, \Theta_M, M) \approx \mathcal{N}(D_l \mid \epsilon_l,\, \mathbf{F} + \Omega_M(r_l, \mathbf{F}, \Theta_M))$$



Inference method

Given location of selected site and model, can calculate probability of observed neutral allele frequencies a given distance from selected site.

$$P(D_l \mid r_l, \mathbf{F}, \Theta_M, M) \approx \mathcal{N}(D_l \mid \epsilon_l, \ \mathbf{F} + \Omega_M(r_l, \mathbf{F}, \Theta_M))$$



$$\mathcal{L}(\Theta_M, M; D) = (\prod_{i=1}^{L_{left}} P(D_i \mid r_i, \mathbf{F}, \Theta_M, M))(\prod_{j=1}^{L_{right}} P(D_j \mid r_j, \mathbf{F}, \Theta_M, M))$$

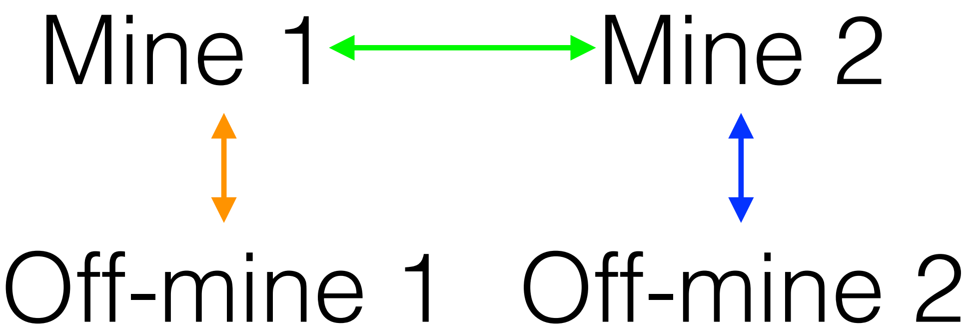
Kim and Stephan, 2002; Nielsen et al., 2005; Chen et al., 2010; Racimo, 2015

Application to data:

Mimulus guttatus adaptation to copper on mine tailings in California

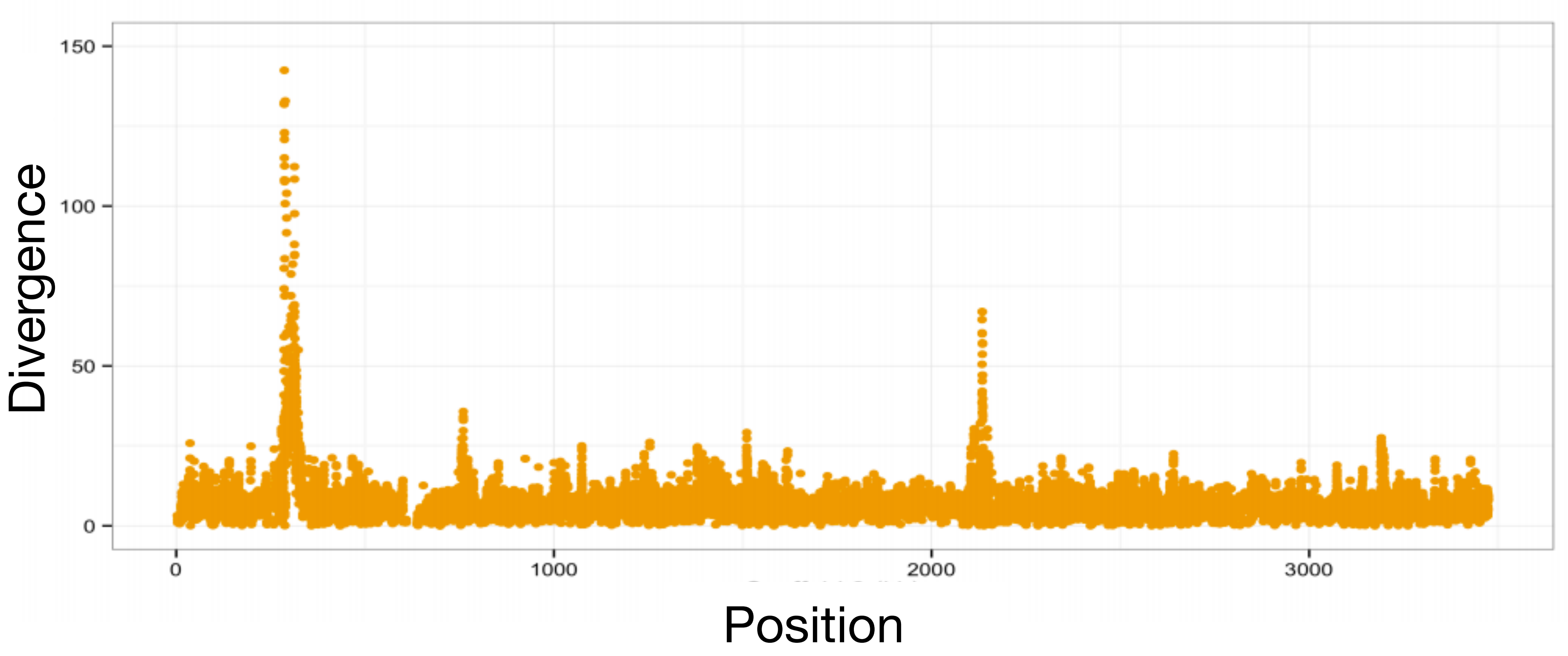
John Willis

Kevin Wright



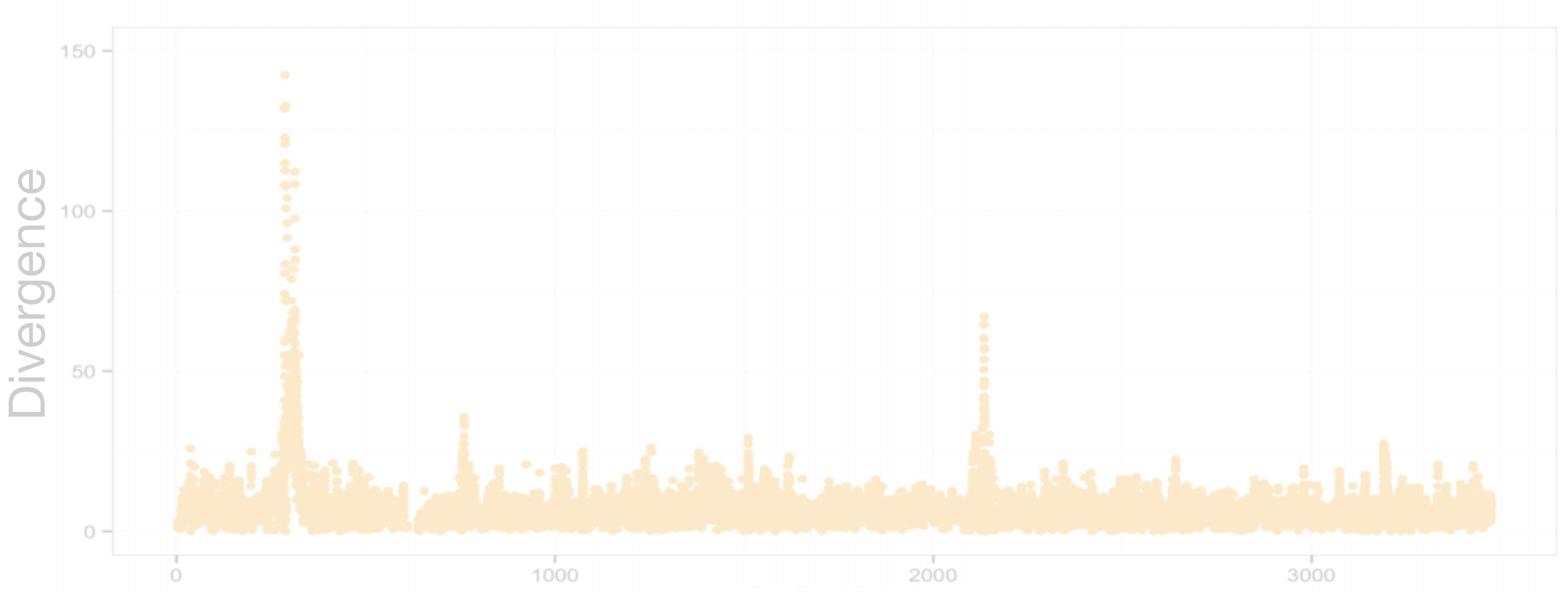
Application to data:

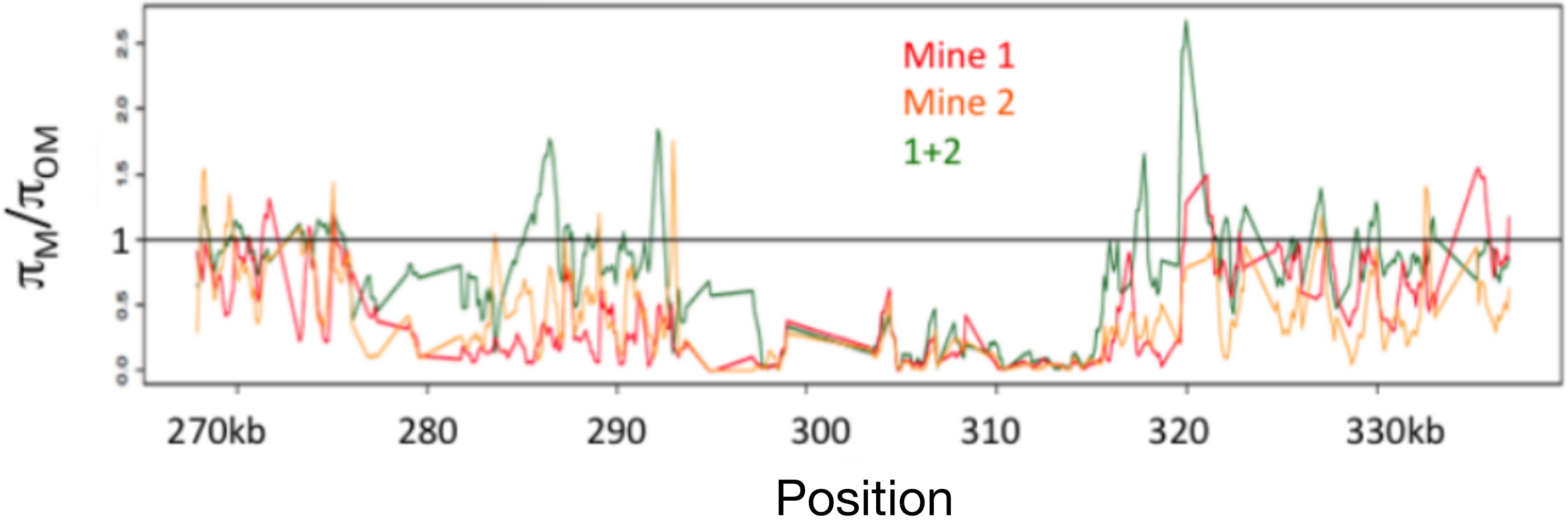
Mimulus guttatus adaptation to copper on mine tailings in California

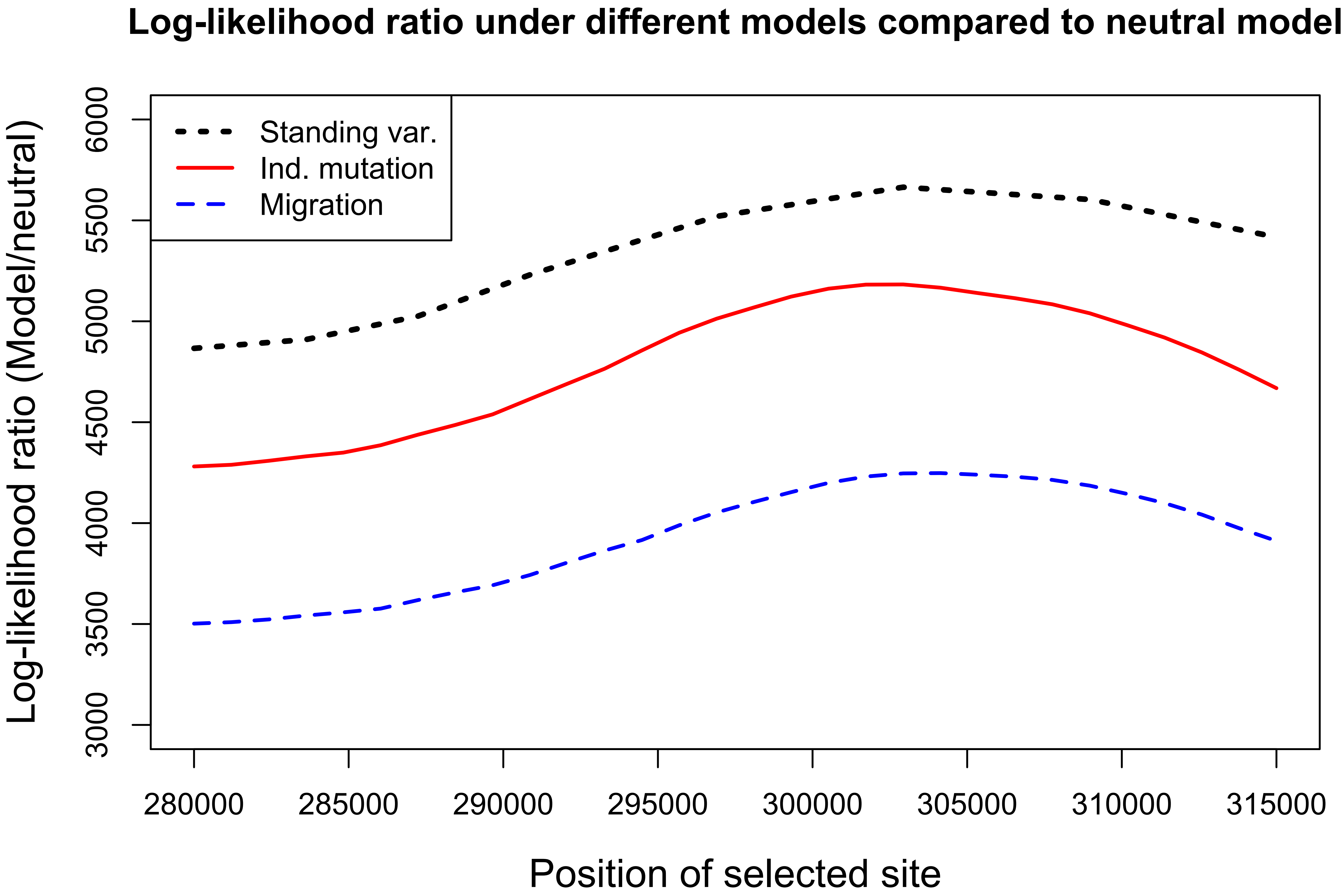


Application to data:

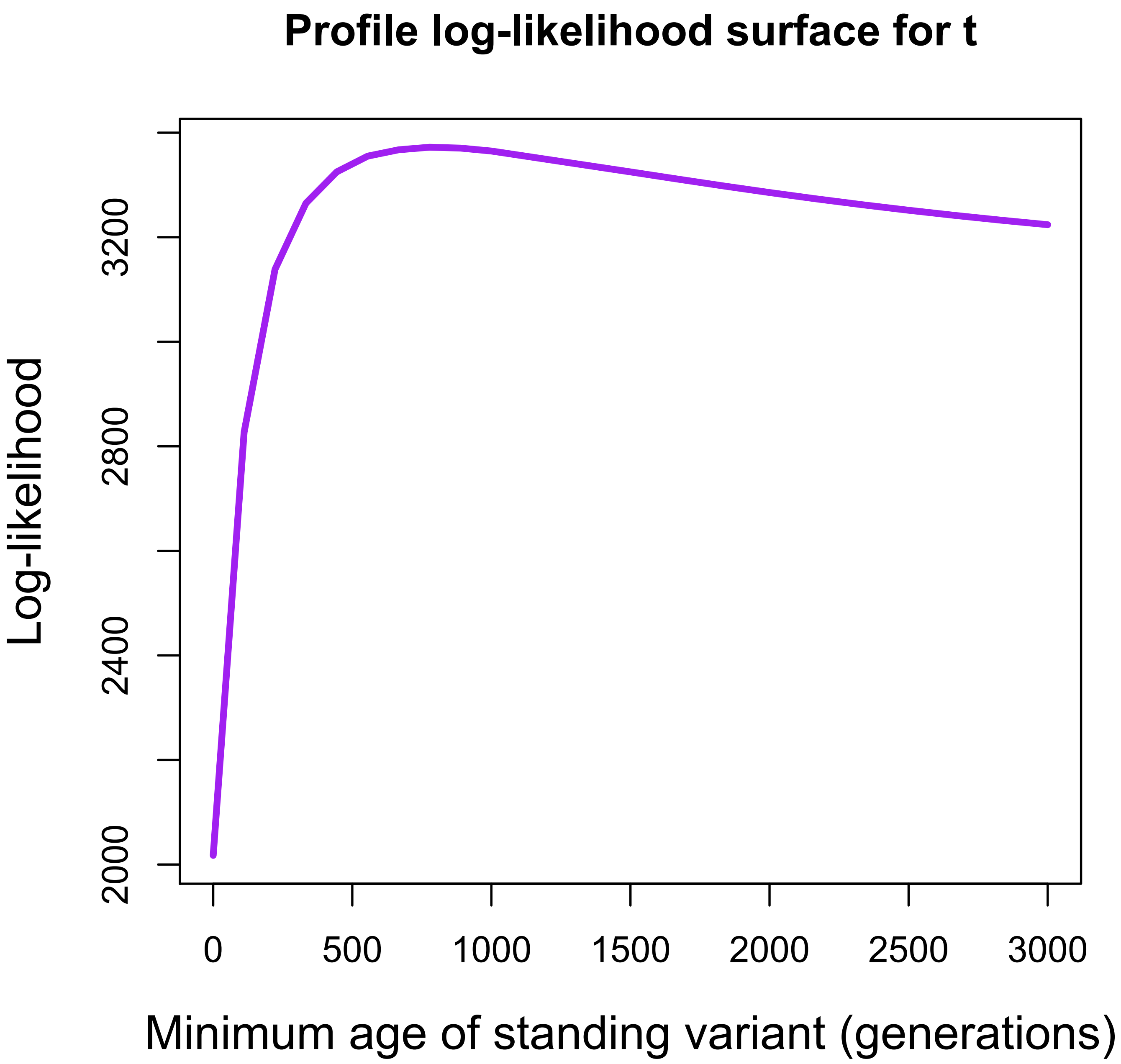
Mimulus guttatus adaptation to copper on mine tailings in California



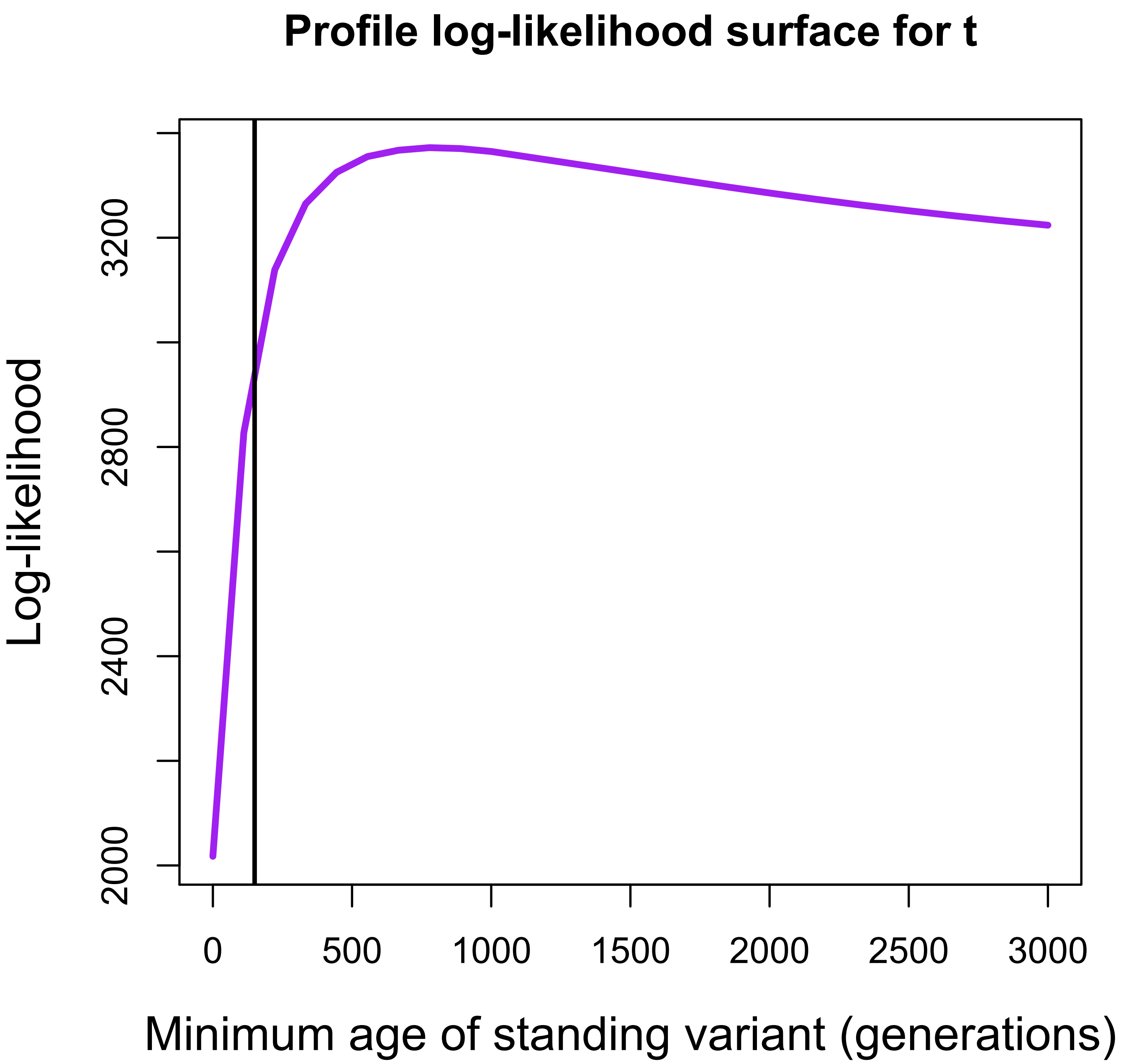



Log-likelihood ratio under different models compared to neutral model
Standing var.
Ind. mutation
Migration
Log-likelihood ratio (Model/neutral)
3000
3500
4000
4500
5000
5500
6000
280000
285000
290000
295000
300000
305000
310000
315000
Position of selected site

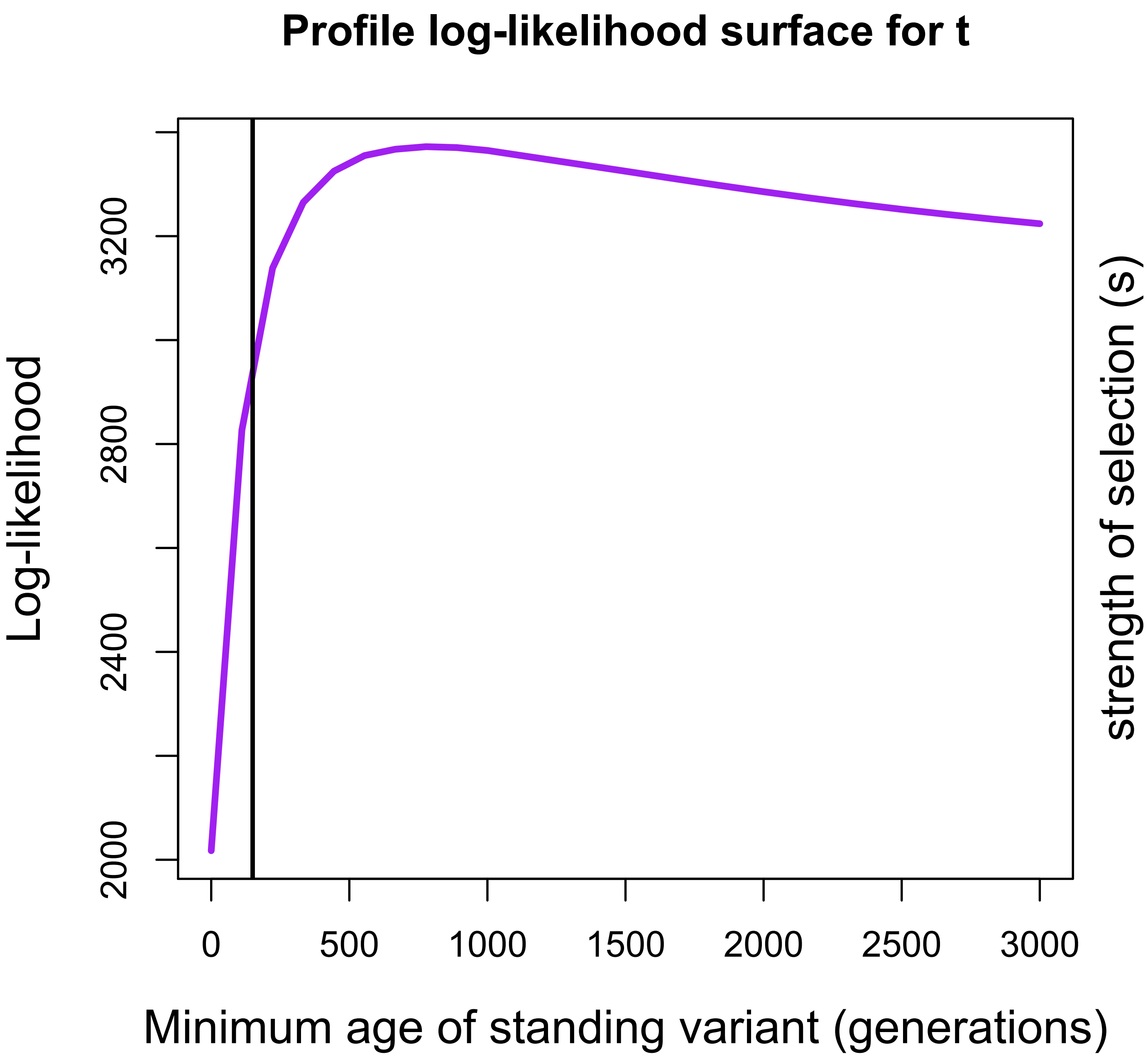
Under standing variation at most likely selected site



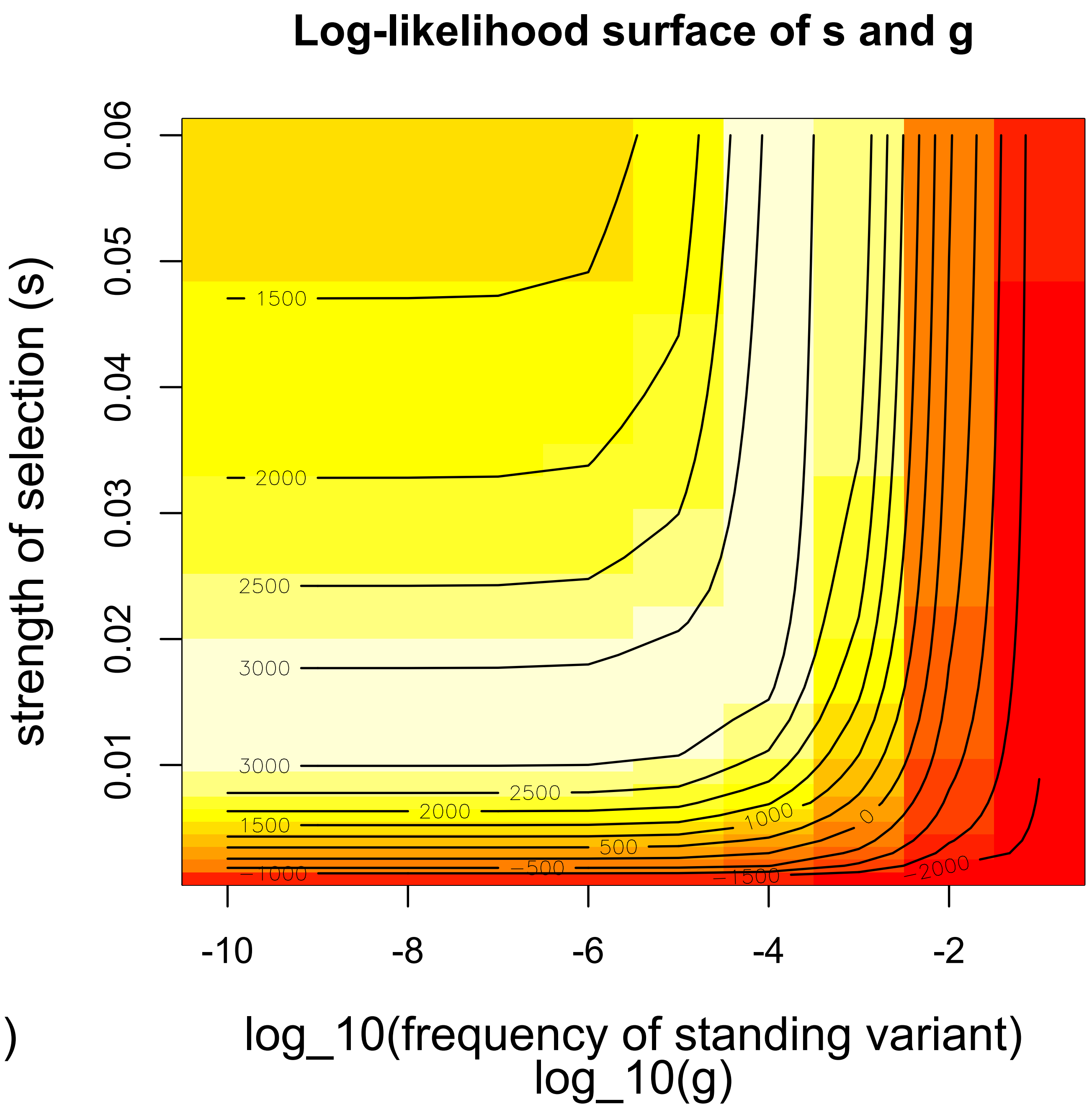
Under standing variation at most likely selected site



Under standing variation at most likely selected site



Profile log-likelihood surface for t



Log-likelihood surface of s and g

Conclusions

1. Can define models for modes of convergent adaptation for multiple populations with arbitrary relationships

2. Applied inference method to Mimulus guttatus copper tolerance data
    - Can apply to a wide range of other datasets

Thank you!

The Coop Lab

Jeremy Berg, Gideon Bradburd, Graham Coop, Alisa Sedghifar, Chenling Xu,
Ivan Juric, Simon Aeschbacher, Vince Buffalo

Application to data:

Mimulus guttatus adaptation to copper on mine tailings in California

GENETIC DIVERSITY IN TOP "SWEEP REGION".