Exploring genome-wide signals of selection against gene flow

Simon Aeschbacher

with
Graham Coop

saeschbacher@ucdavis.edu

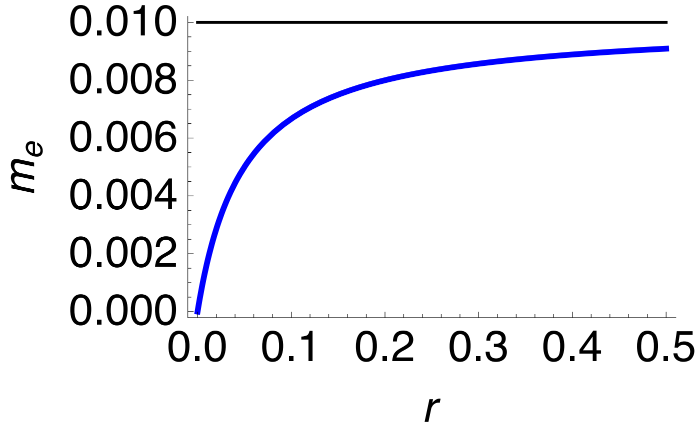
Motivation

- Spatially heterogeneous selection
  - Certain alleles maladaptive in particular environments or genomic backgrounds
  - Local adaptation in the face of (strong) gene flow
- Effective reduction in gene flow
  - At closely linked neutral sites (e.g. Petry 1983)
  - At other selected sites (e.g. Bürger & Akerman 2011)
- Heterogeneity in divergence along the genome
  - Teeter et al. (2008), Carneiro et al. (2009), Geraldes et al. (2011); review: Nachman & Payseur (2012)

Effective migration rate

- Petry (1983), Bengtsson (1985), and many more:
  - Effective reduction in gene flow due to selection against maladapted alleles

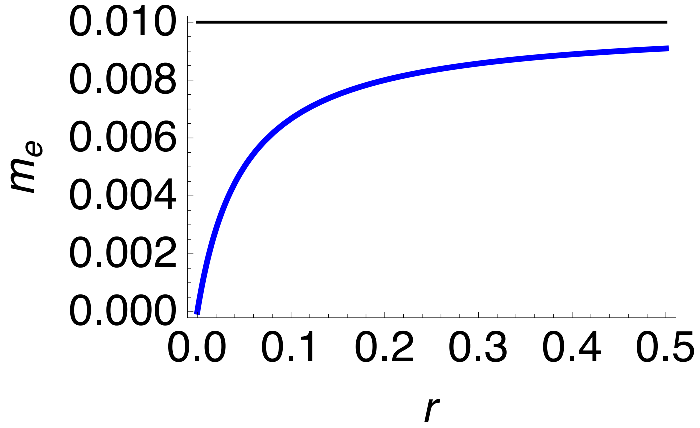
$$m_e \approx m\left(1+\frac{a}{r}\right)^{-1}$$



Effective migration rate

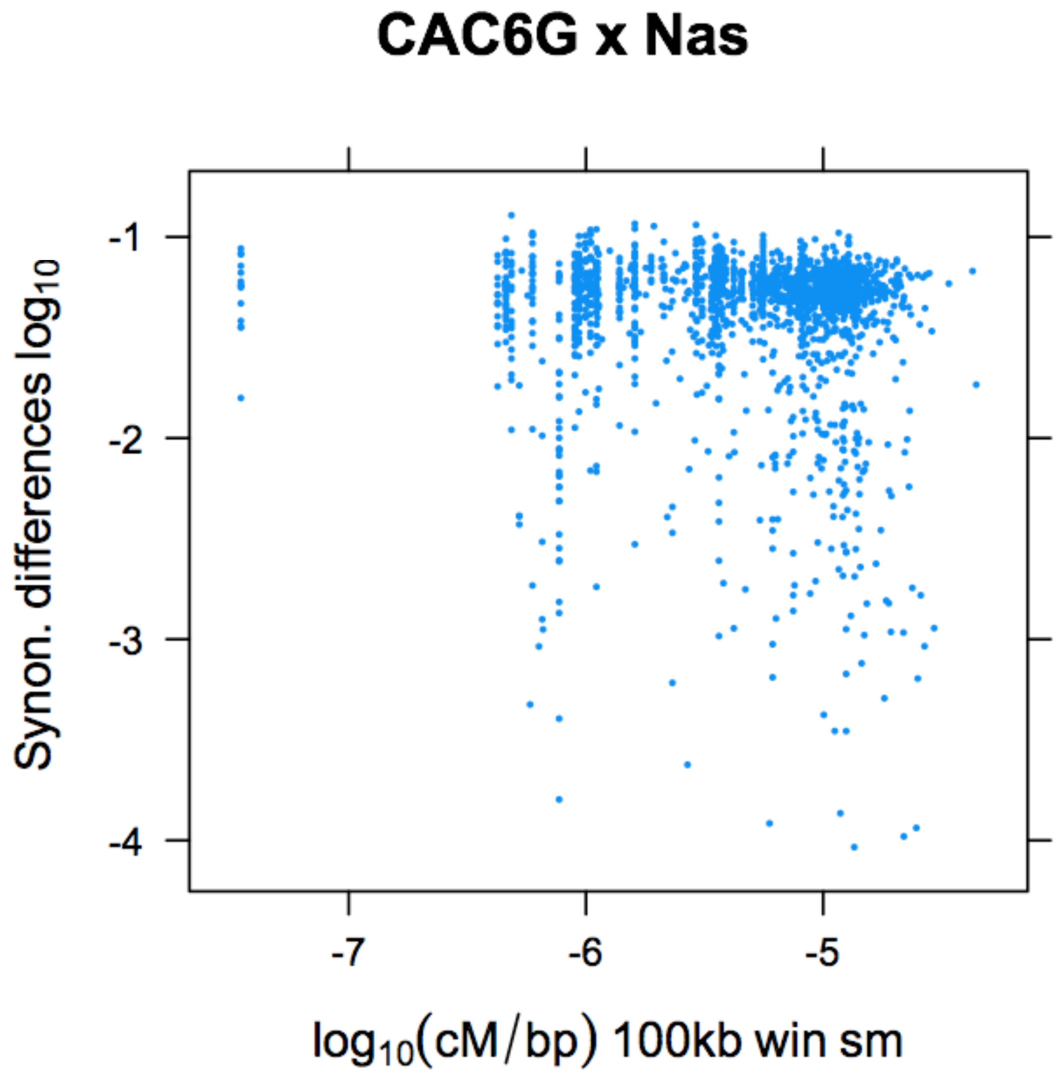
- Petry (1983), Bengtsson (1985), and many more:
  - Effective reduction in gene flow due to selection against maladapted alleles

$$m_e \approx m \left(1 + \frac{a}{r}\right)^{-1}$$



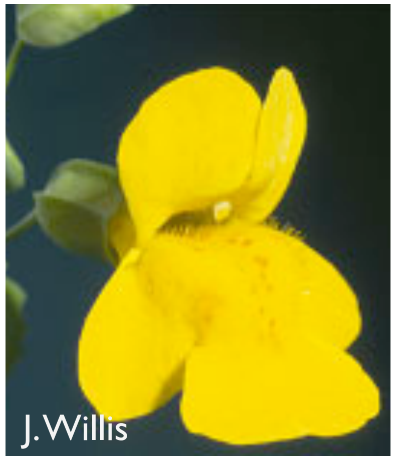
- Nordborg (1997), Charlesworth et al. (1997):
  - Effect on pairwise coalescent times within and between populations of linkage to a migration–selection polymorphism
  - Advocating use of absolute divergence rather than F_{ST}

Example 1: Introgression of *Mimulus nasutus* into *M. guttatus*



rho = -0.0496; p = 0.0343

M. nasutus
Nas



M. guttatus
CAC6G

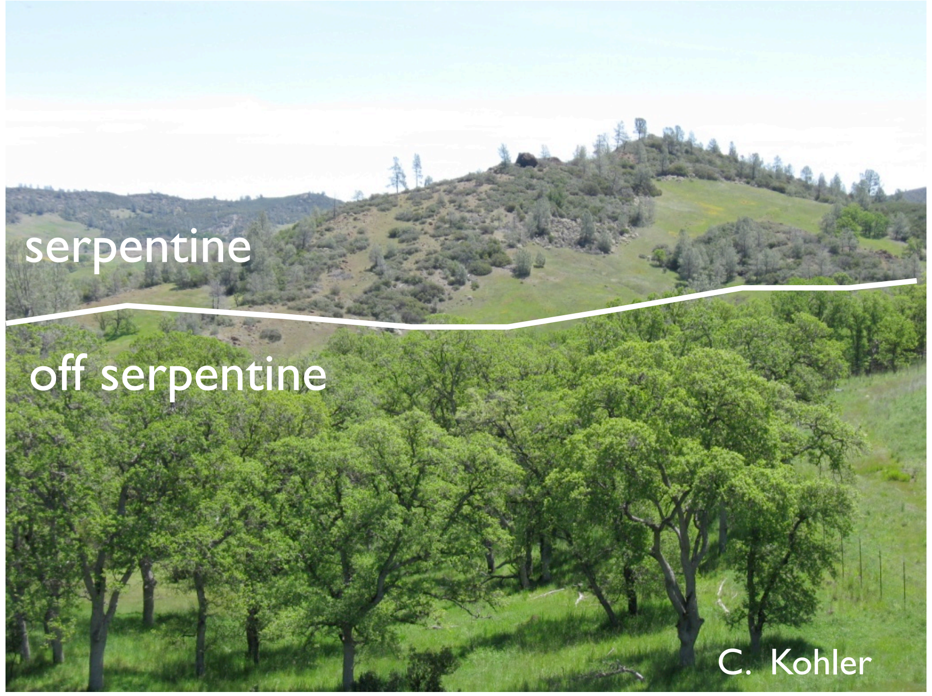
Evidence for selection against gene flow into the 'wrong' genomic background?

Brandvain, Kenney et al. 2014 bioRxiv http://dx.doi.org/10.1101/000109 | Photos: http://www.nsf.gov [06/22/2014; modified]

Example 2: Local adaptation to serpentine soil in *M. guttatus*


S. Aeschbacher

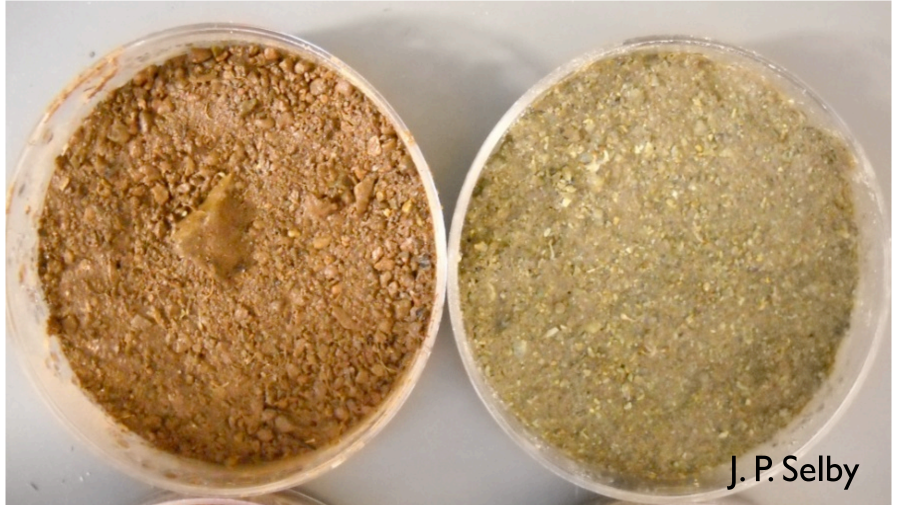

S. Aeschbacher



C. Kohler


J. P. Selby

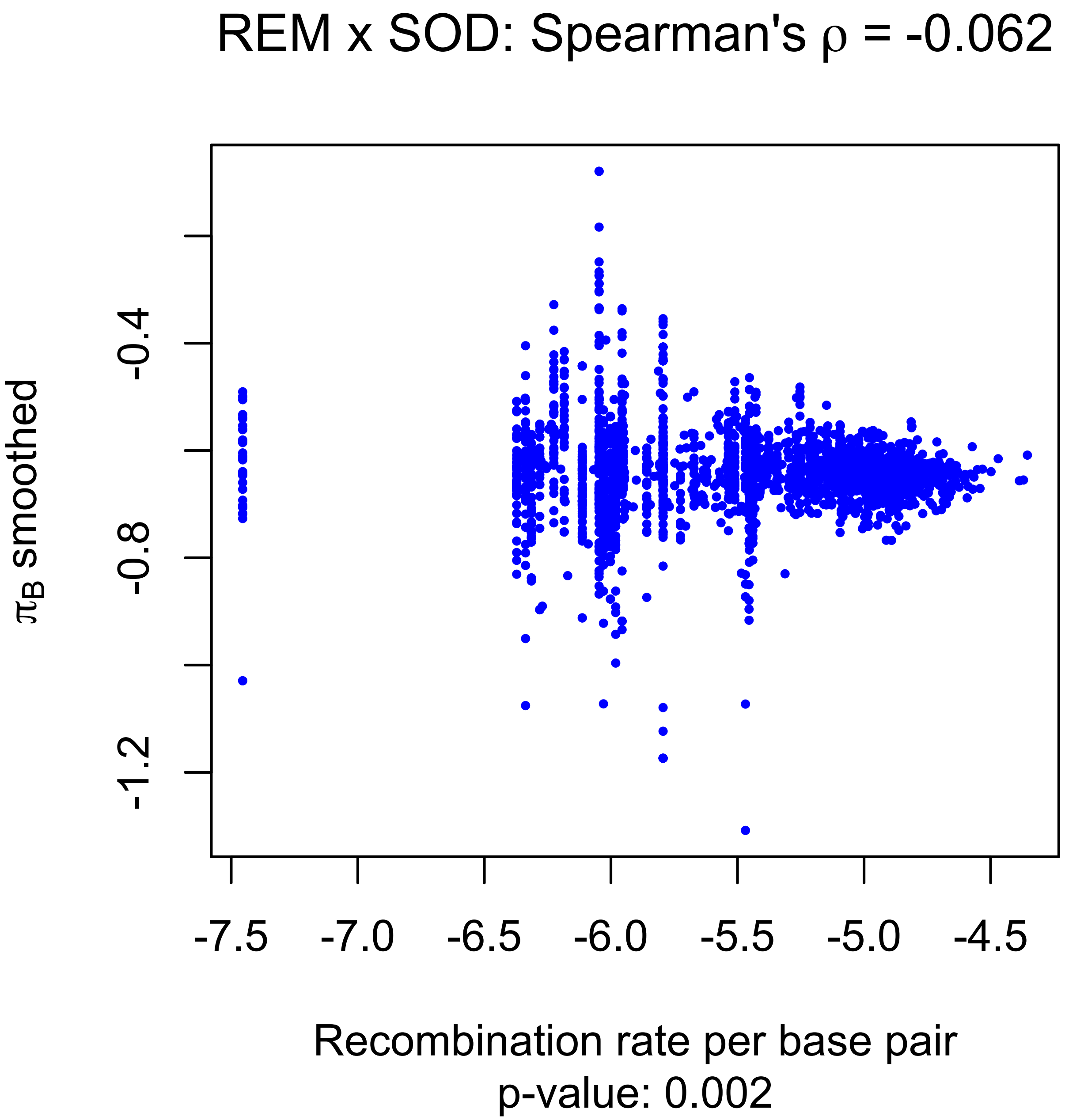
Serpentine seedlings
on
serpentine soil


J. P. Selby

Non-serpentine seedlings
on
serpentine soil

Jessica P. Selby (personal communication)

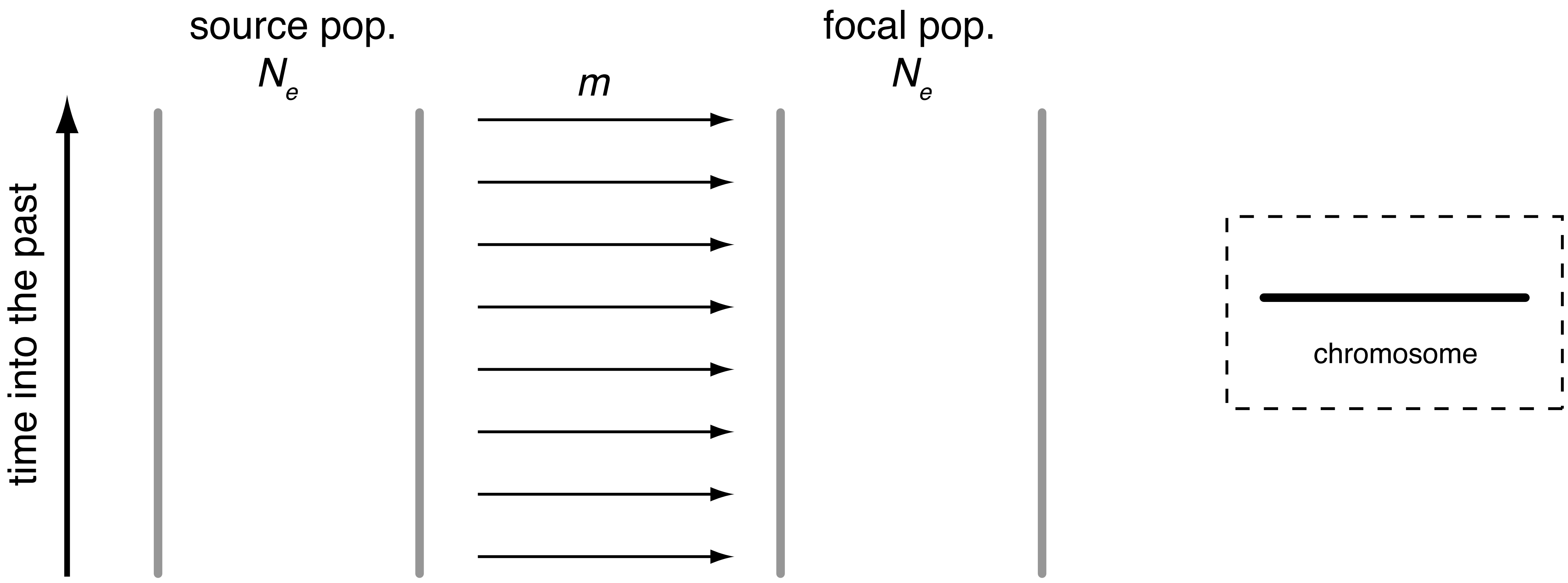
Maladaptive gene flow into serpentine patches in *M. guttatus*?



Evidence for **selection against** maladapted **immigrants from non-serpentine patches**?

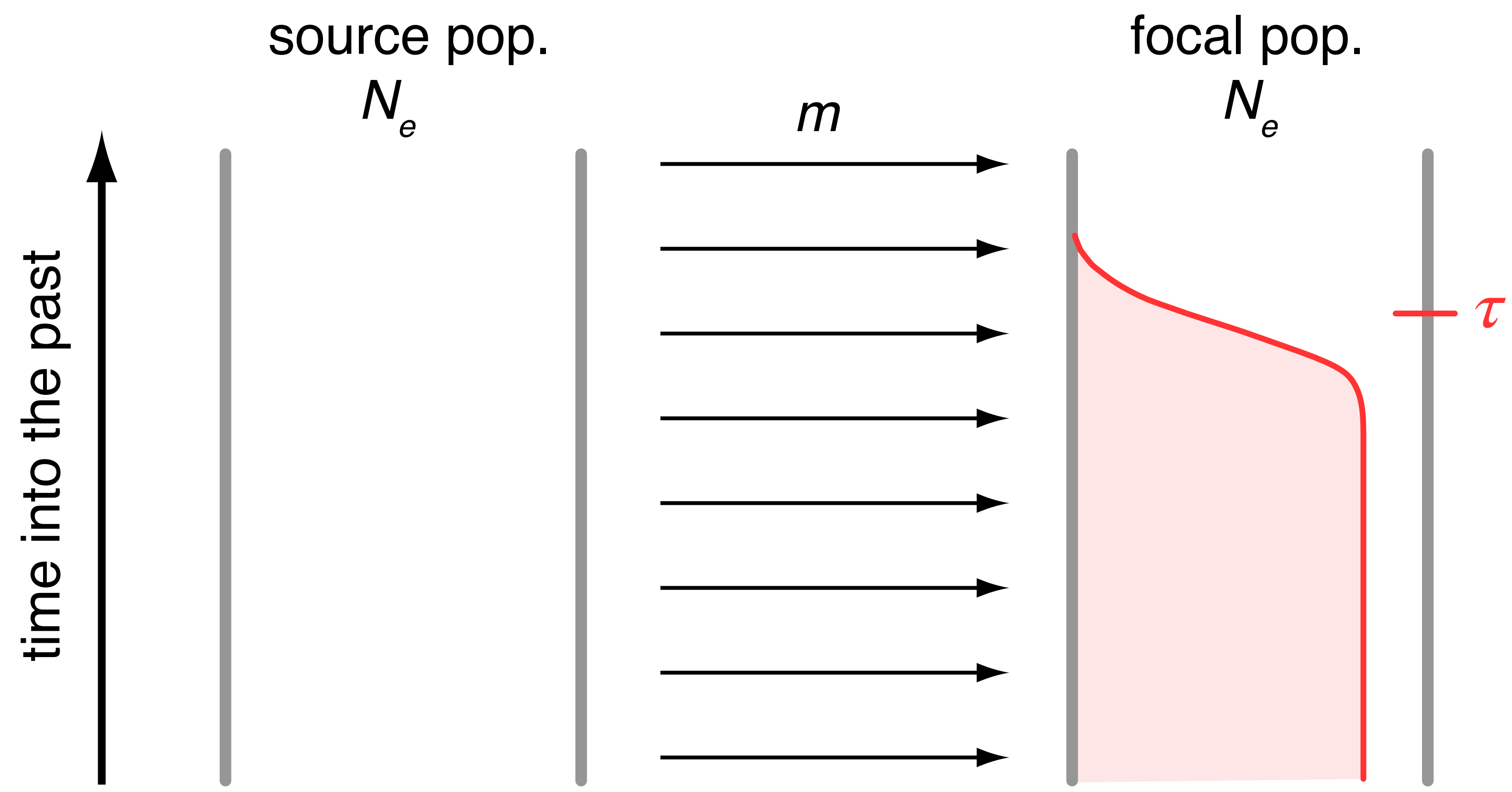
Data obtained from Willis Lab (Duke Univ.); www.phytozome.org, www.mimulusevolution.org

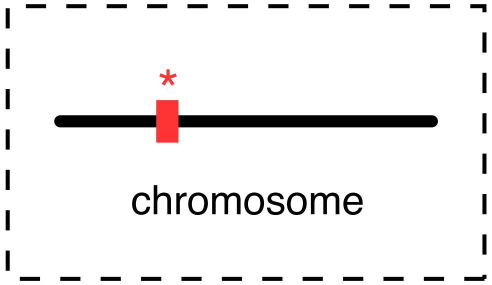
Between-population coalescent time at sites linked to selection



Two populations. One-way migration.

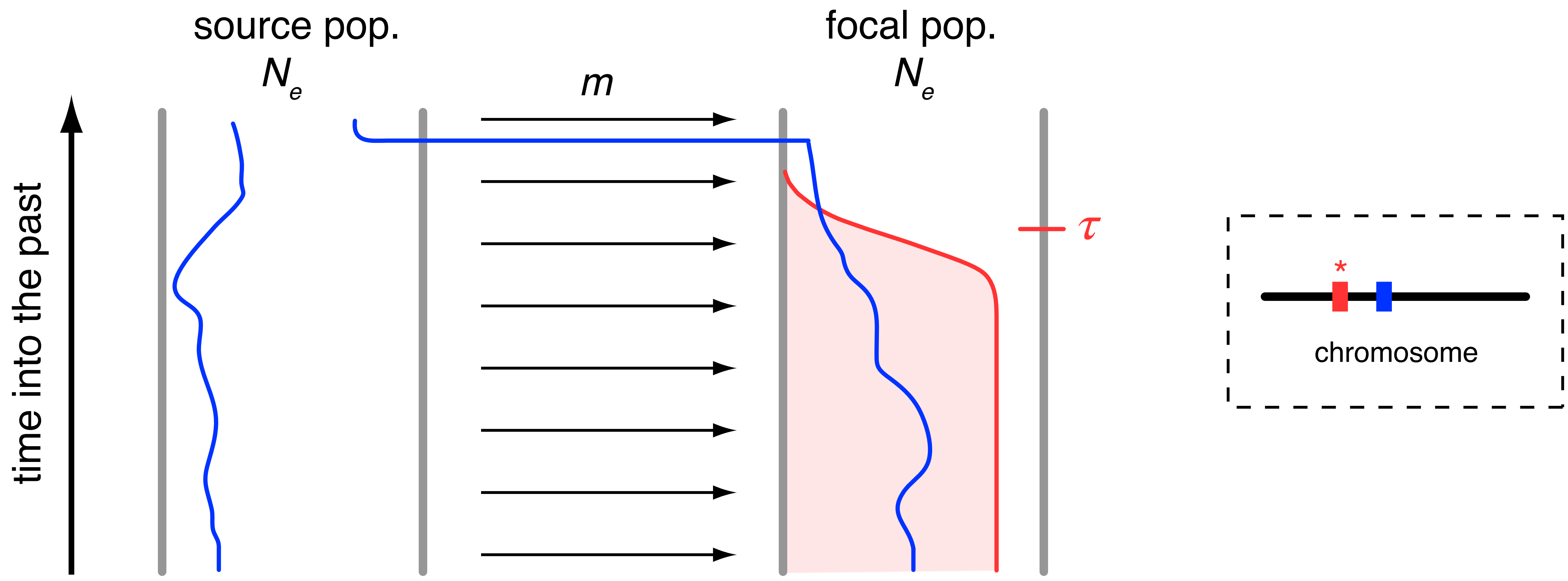
Between-population coalescent time at sites linked to selection





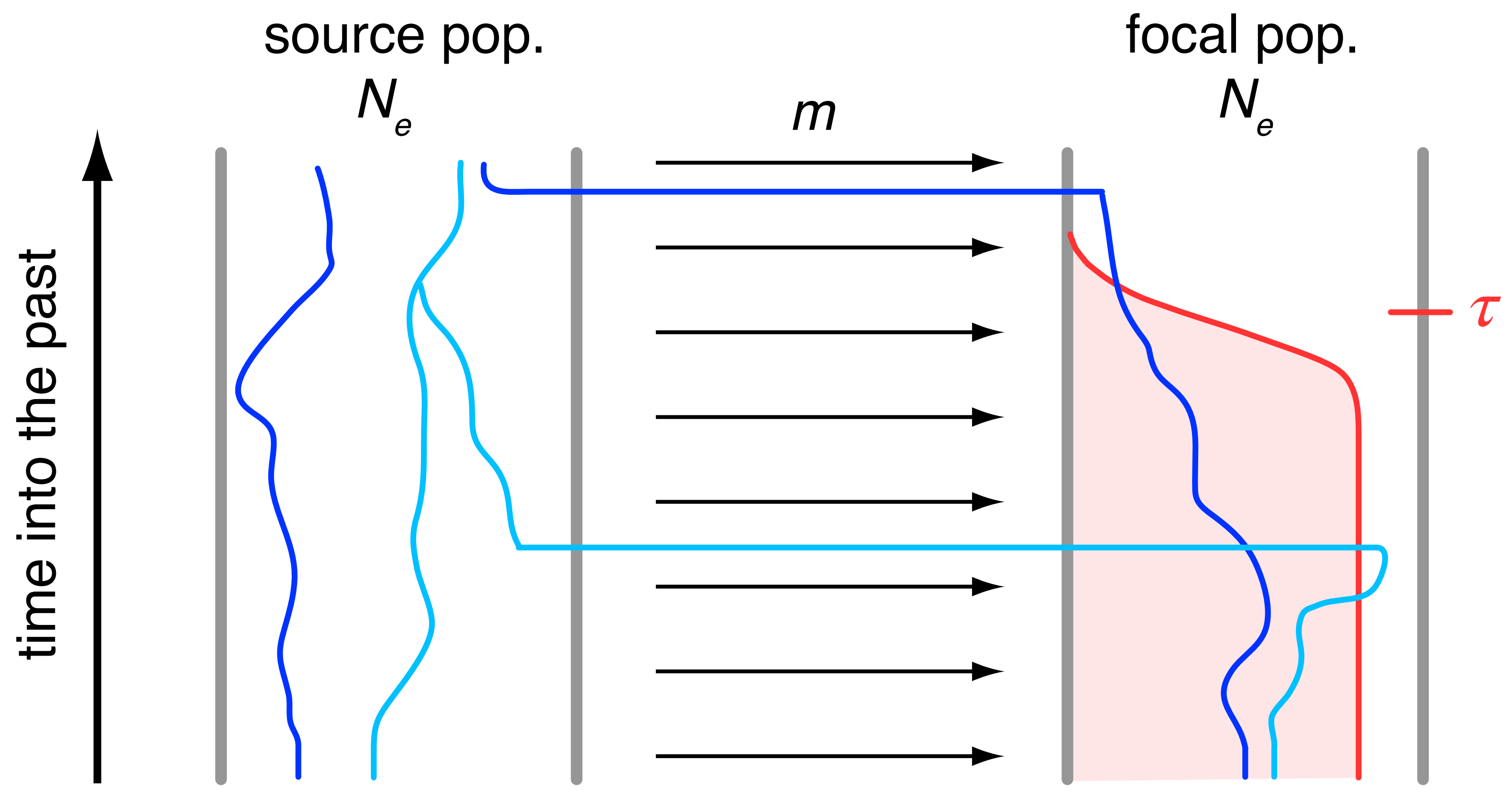
Locally beneficial mutation arose at time τ

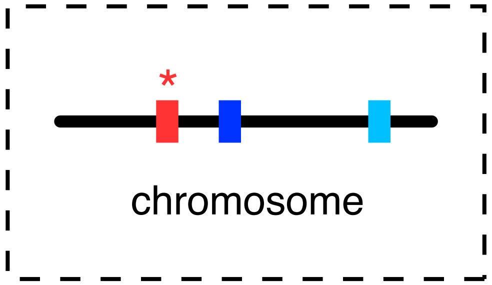
Between-population coalescent time at sites linked to selection



Trajectory of a closely linked pair of lineages

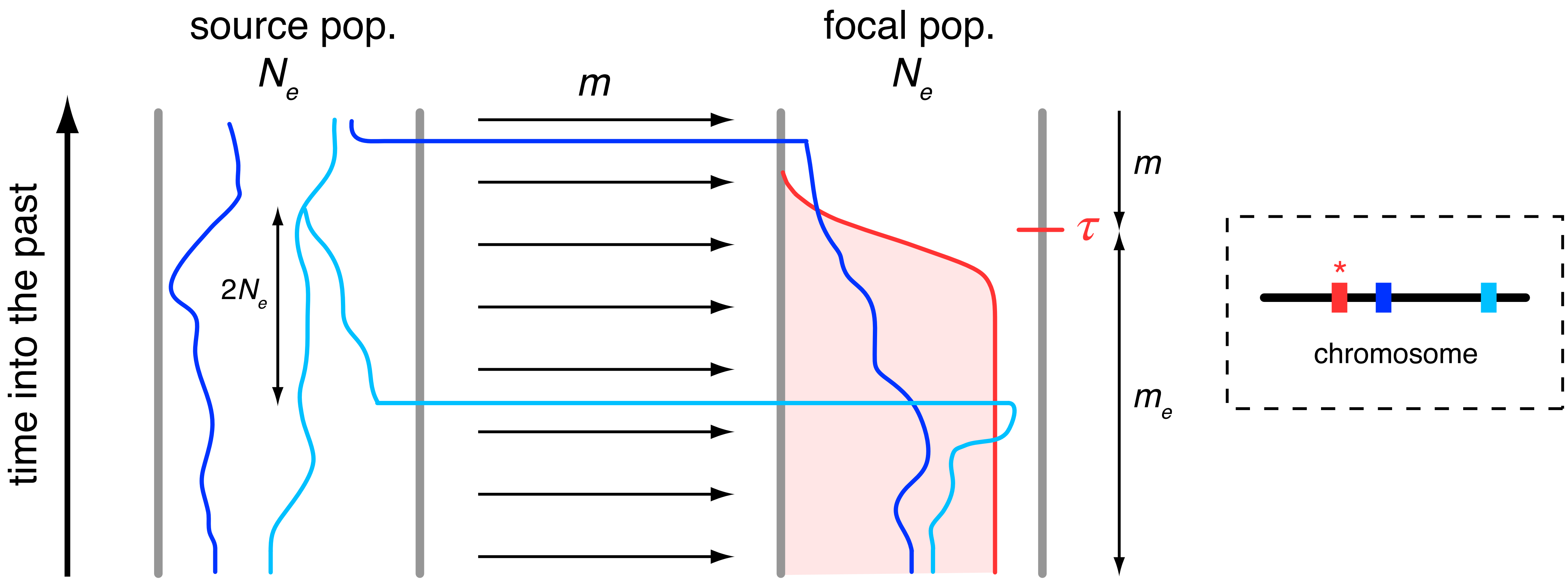
Between-population coalescent time at sites linked to selection





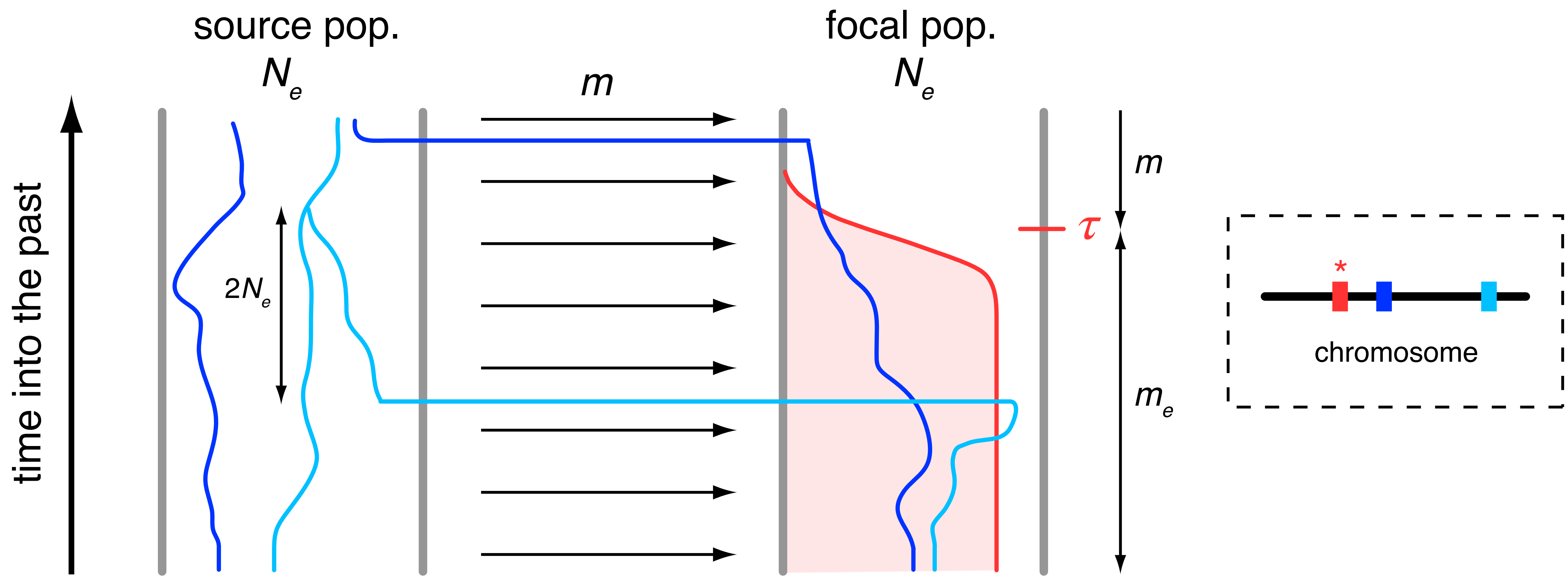
Trajectory of a loosely linked pair of lineages

Between-population coalescent time at sites linked to selection



$$\mathbb{E}[T_\mathrm{B}] = \int_0^\tau (t + 2N_e)\, m_e e^{-m_e t} dt + e^{-m_e t} \int_0^\infty (\tau + s + 2N_e) m e^{-ms} ds$$

Between-population coalescent time at sites linked to selection



$$\mathbb{E}[T_{\mathrm{B}}] = \int_0^{\tau} (t + 2N_e)\, m_e e^{-m_e t} dt + e^{-m_e t} \int_0^{\infty} (\tau + s + 2N_e) m e^{-ms} ds$$

A genomic perspective: borrowing strength across multiple sites

$$\begin{aligned}\mathbb{E}_L\big[\mathbb{E}\left[T_{\mathrm{B}}\right]\big] &= \int_1^{\Lambda} \mathbb{E}[T_{\mathrm{B}}] f_L(l;\nu_{\mathrm{BP}})dl \\ &= g(N_e, m, a, \nu_{\mathrm{BP}}, r_{\mathrm{BP}})\end{aligned}$$

A genomic perspective: borrowing strength across multiple sites

Distribution of distances
between neutral and selected site

$$\begin{aligned}\mathbb{E}_L\left[\mathbb{E}\left[T_{\mathrm{B}}\right]\right] &= \int_1^{\Lambda} \mathbb{E}[T_{\mathrm{B}}]\boxed{f_L(l;\nu_{\mathrm{BP}})}dl \\ &= g(N_e, m, a, \nu_{\mathrm{BP}}, \boxed{r_{\mathrm{BP}}})\end{aligned}$$

Selection coefficient

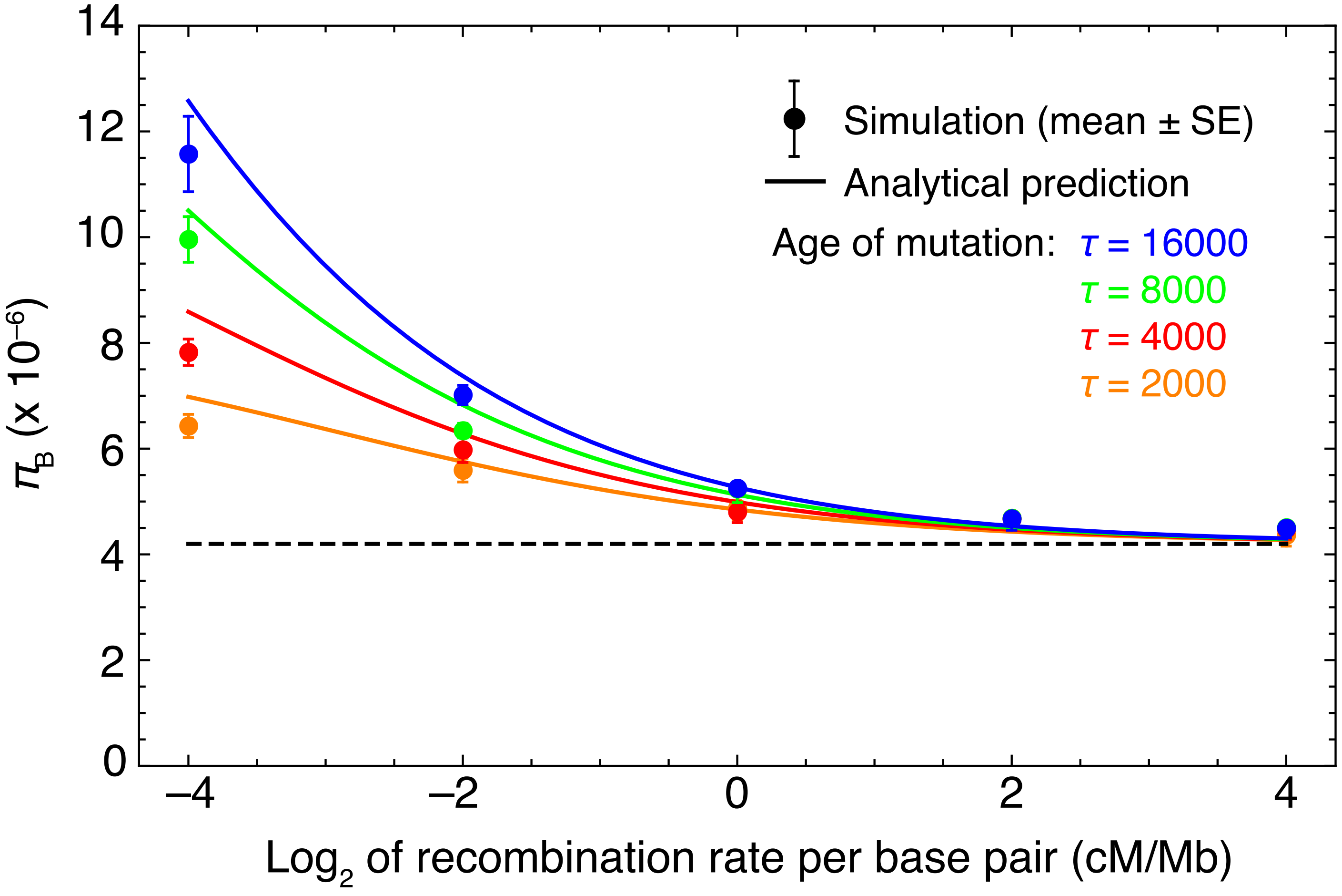
Per-base pair
recombination rate

Rate at which selected
loci occur along the genome

$$\mathbb{E}_L[\pi_{\mathrm{B}}] = 2u\mathbb{E}_L\left[\mathbb{E}[T_{\mathrm{B}}]\right]$$

Analytical prediction versus individual-based simulations

- Simulated 5 *independent* chromosomes (20 Mb), each with a given recombination rate using SLiM (Messer 2013; modified)
- One migration–selection polymorphism per chromosome
- 10 replications



$$N_e = 1000$$

$$m = 0.01$$

$$a = 0.1$$

$$\pi_{\mathrm{B}} = 2N_e + \frac{1}{m}$$

More than one selected locus – BUT forget ages of mutations

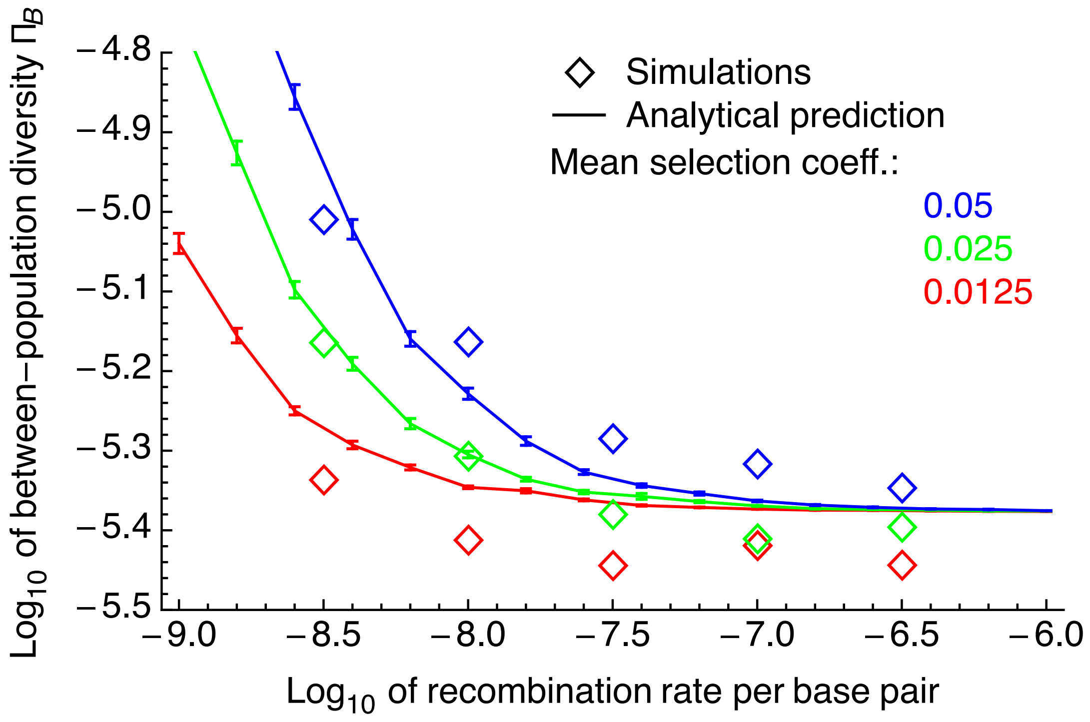
- Effective migration rate at a neutral site linked to multiple migration–selection polymorphisms (Aeschbacher & Bürger 2014)

$$m_e \approx m \left[\prod_{i=1}^{I} \left(1 + \frac{a_i}{\sum_{k=1}^{i-1} a_k + r_i} \right)^{-1} \right] \times \left[\prod_{j=1}^{J} \left(1 + \frac{b_j}{\sum_{k=1}^{j-1} b_k + r_j} \right)^{-1} \right]$$

More than one selected locus – BUT forget ages of mutations

- Effective migration rate at a neutral site linked to multiple migration–selection polymorphisms (Aeschbacher & Bürger 2014)

$$m_e \approx m \left[\prod_{i=1}^{I} \left(1 + \frac{a_i}{\sum_{k=1}^{i-1} a_k + r_i} \right)^{-1} \right] \times \left[\prod_{j=1}^{J} \left(1 + \frac{b_j}{\sum_{k=1}^{j-1} b_k + r_j} \right)^{-1} \right]$$



$$\mathbb{E}[\pi_{\mathrm{B}}] = 2u \left(2N_e + \frac{1}{m_e} \right)$$

- 10 selected loci per chromosome
- Chromosomes of length 100 Mb
- 5 recombination strata
- 24 replicates
- Mutation rate of 10^{-9}
- SLiM (Messer 2013; modified)

Multilocus m_e: exponential decay as a function of a compound parameter

$$m_e \approx m e^{2\eta\{\ln(K)+\gamma\}}$$

Multilocus m_e: exponential decay as a function of a compound parameter

of migration–selection polymorphisms

Euler's constant ~ 0.58

$$m_e \approx m e^{2\eta\{\ln(K)+\gamma\}}$$

Multilocus m_e: exponential decay as a function of a compound parameter

\# of migration–selection polymorphisms

Euler’s constant ~ 0.58

$$m_e \approx m e^{2\eta\{\ln(K)+\gamma\}}$$

$$\boxed{\eta = \frac{\bar{a}\nu_{\text{BP}}}{r_{\text{BP}}}}$$

Amount of selection against gene flow per unit of recombination

Multilocus m_e: exponential decay as a function of a compound parameter

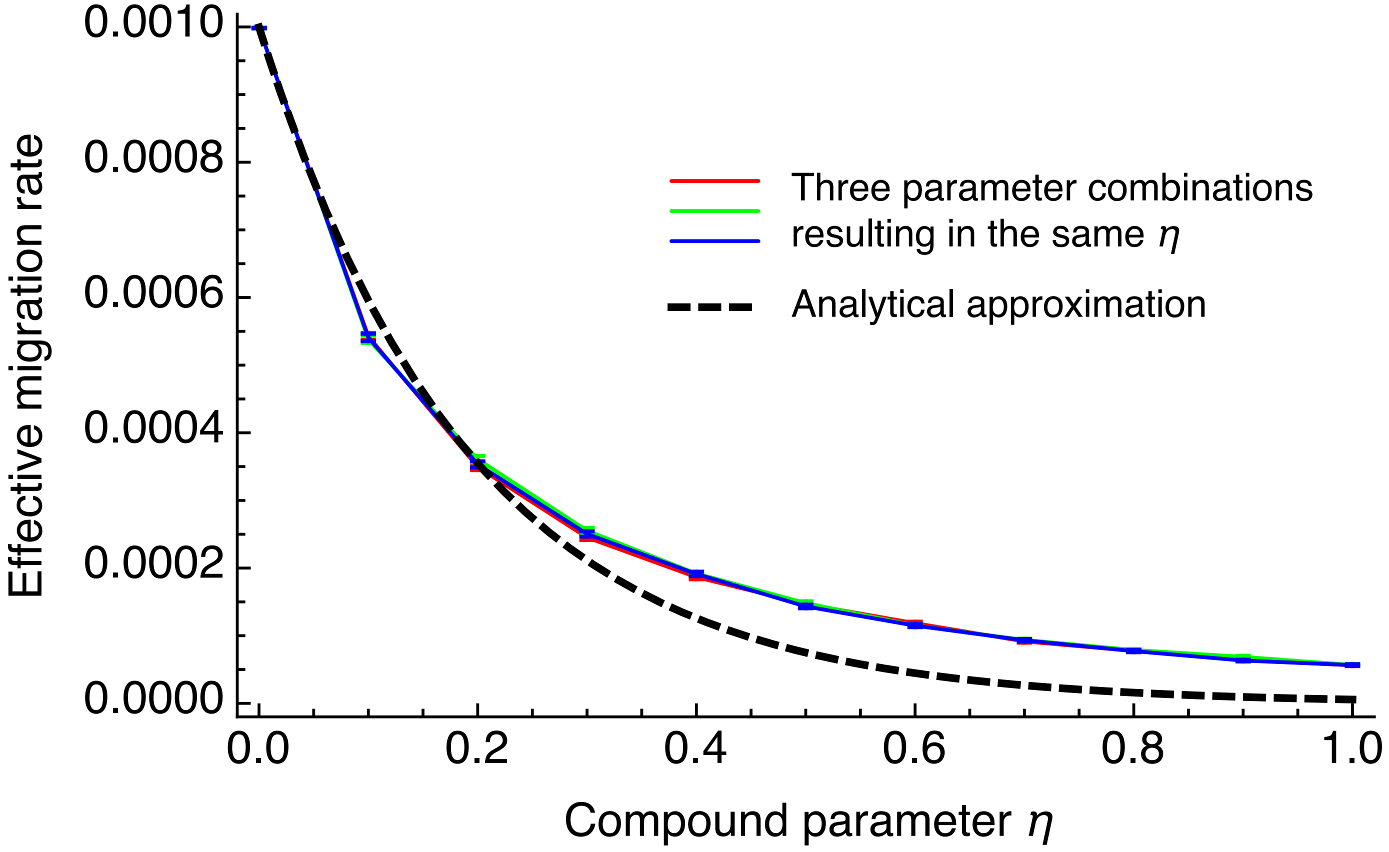
of migration–selection polymorphisms

Euler's constant ~ 0.58

$$m_e \approx m e^{2\eta\{\ln(K)+\gamma\}}$$

$$\boxed{\eta = \frac{\bar{a}\nu_{\mathrm{BP}}}{r_{\mathrm{BP}}}}$$

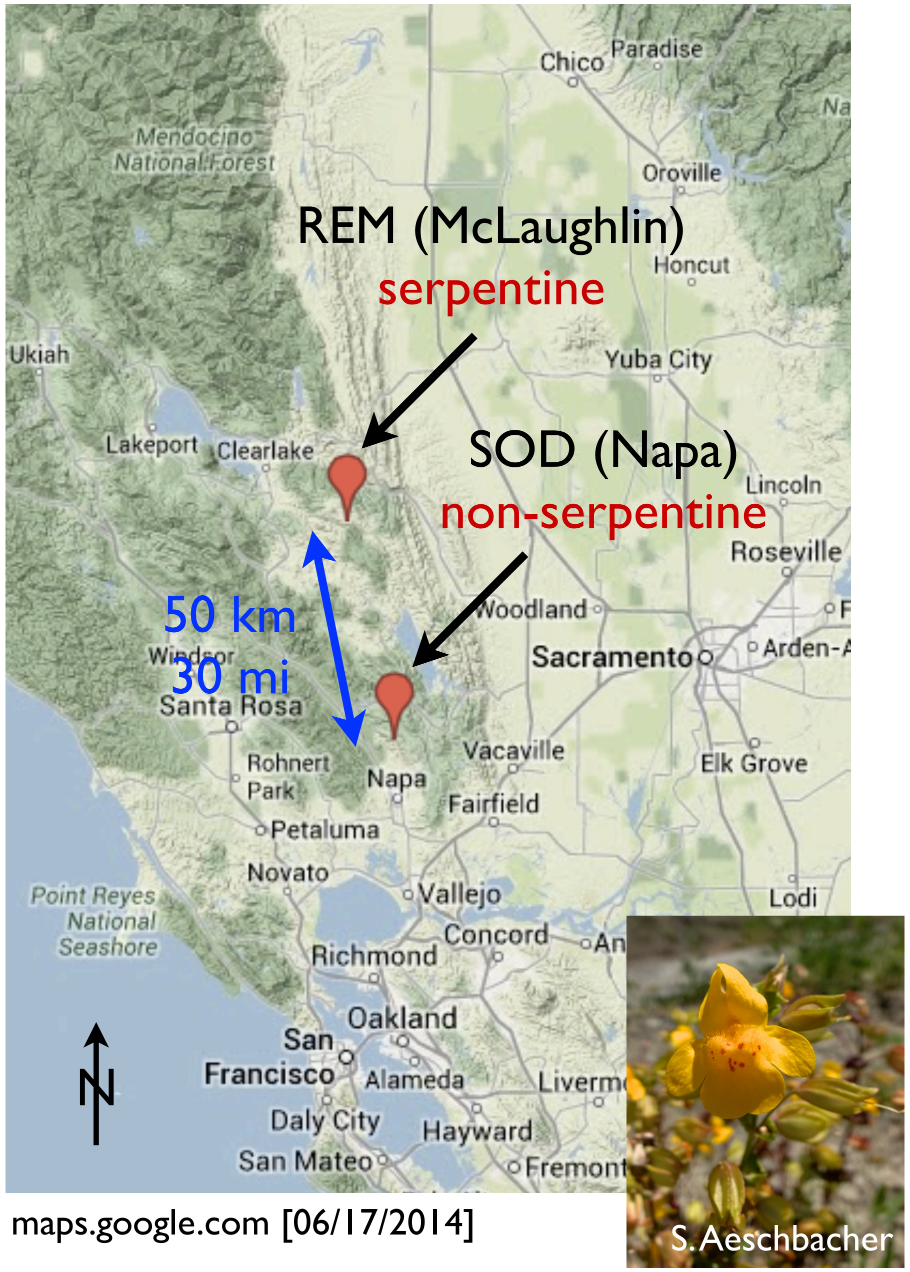
Amount of selection against gene flow per unit of recombination



$$K = 100$$

$$m = 0.001$$

Application: *M. guttatus* on and off serpentine soil



maps.google.com [06/17/2014]

S.Aeschbacher

- Genome-wide sequence data from Willis Lab (Duke Univ.)
- Aligned to *M. guttatus* reference v. 2.0
- Called SNPs, recorded informative sites
- Filtered out low coverage, tested for constant Ti/Tv ratio
- Physical linkage map (Lex Flagel, Yaniv Brandvain)
- Binned data into windows of 100 kb

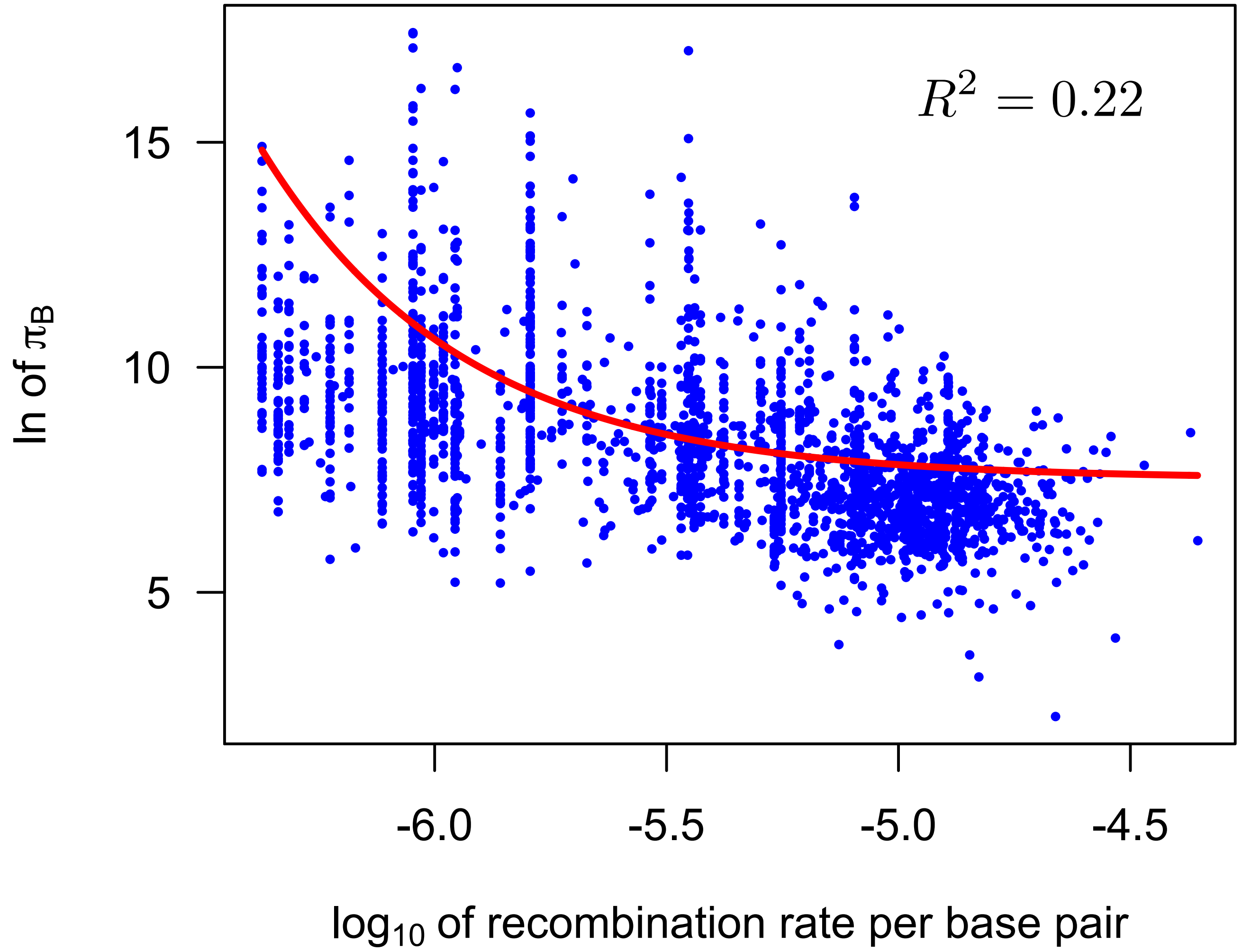
www.phytozome.org | www.mimulusevolution.org

Application: Least-squares fit to *M. guttatus* genomic data

$$
\begin{aligned}
\ln\left(\frac{\hat{\pi}_{\mathrm{B}}}{2u} - 2\hat{N}_e\right) &= -\ln(m) + 2\eta\left\{\ln(K) + \gamma\right\} \\
&= \boxed{-\ln(m)} + \frac{1}{r_{\mathrm{BP}}}\boxed{2\nu_{\mathrm{BP}}\bar{a}\left\{\ln(K) + \gamma\right\}} \\
&= a + \frac{1}{r_{\mathrm{BP}}}b
\end{aligned}
$$

Application: Least-squares fit to *M. guttatus* genomic data

Omitting very low per-base pair recombination rates ($<10^{-7.5}$):



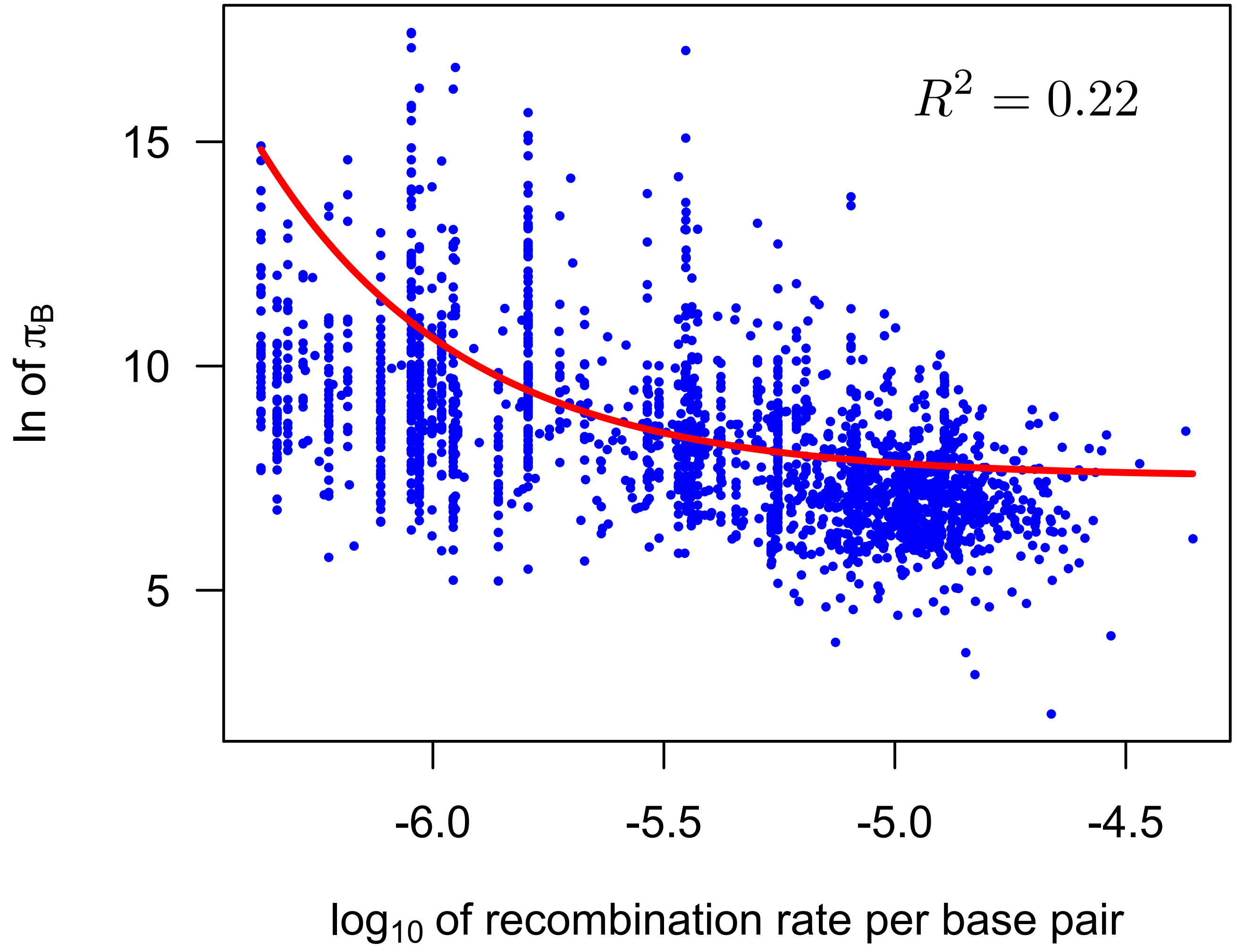
$$-\ln m = 7.53 \pm 0.047$$

$$2\nu_{\mathrm{BP}}\bar{a}\left\{\ln(K) + \gamma\right\} = 1.55 \times 10^{-6} \pm 6.3 \times 10^{-8}$$

$$m \approx 5.4 \times 10^{-4}$$

Application: Least-squares fit to *M. guttatus* genomic data

Omitting very low per-base pair recombination rates ($<10^{-7.5}$) and assuming $u = 10^{-8}$:



$$-\ln m = 7.53 \pm 0.047$$

$$2\nu_{\mathrm{BP}}\bar{a}\,\{\ln(K) + \gamma\} = 1.55 \times 10^{-6} \pm 6.3 \times 10^{-8}$$

$$m \approx 5.4 \times 10^{-4}$$

Example:

$$K = 200$$

$$\text{Average dist.} = 100\text{ kb}$$

$$\bar{a} \approx 0.01$$

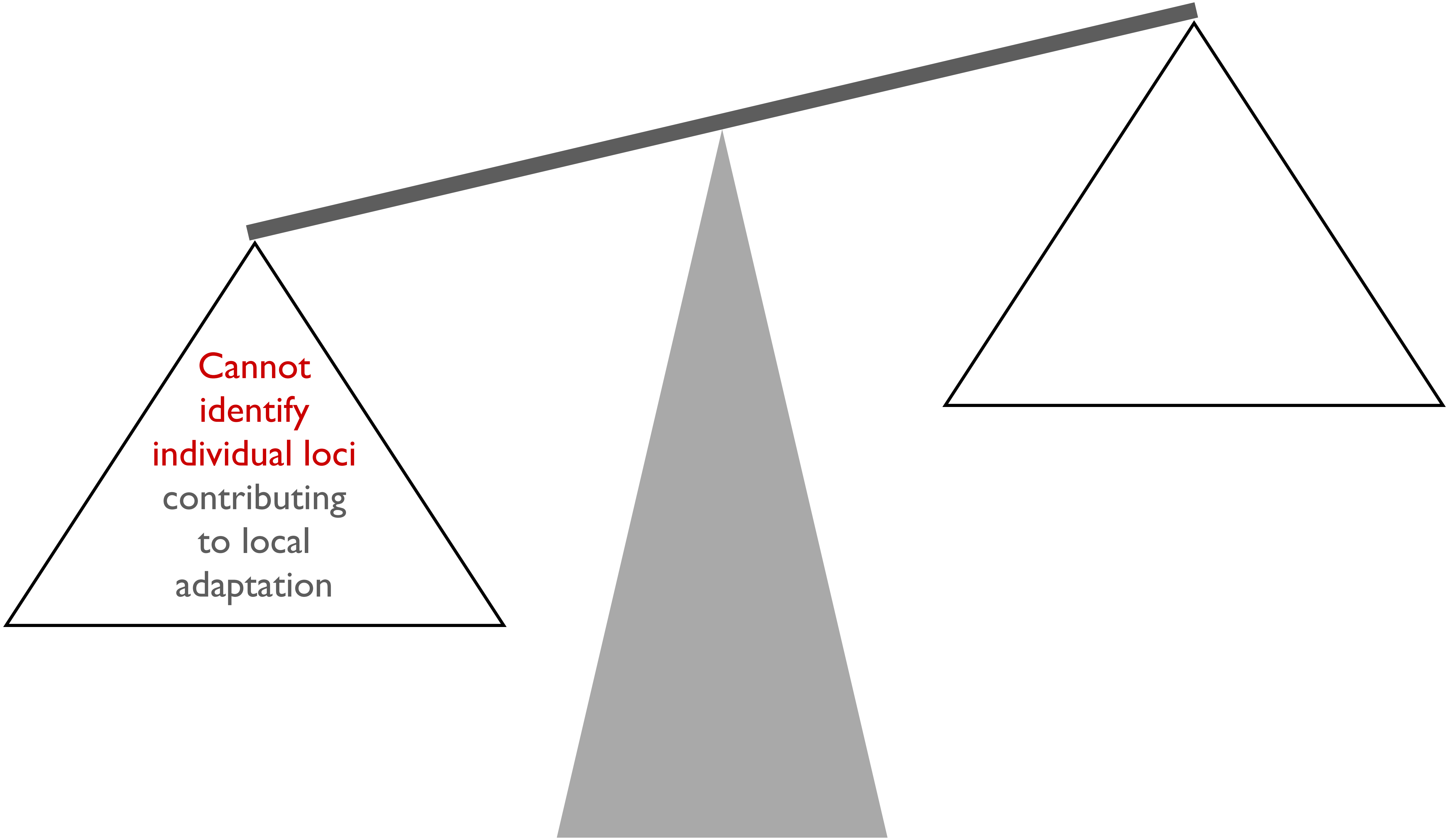
Summary

- Selection against gene flow produces a unique genomic pattern
  - Negative correlation between 'absolute' genetic divergence and local recombination rate
- Identified a compound parameter that, together with the length of the genome, quantifies the per-base pair effect of selection against gene flow

$$\eta = \bar{a}\nu_{\mathrm{BP}}/r_{\mathrm{BP}}$$

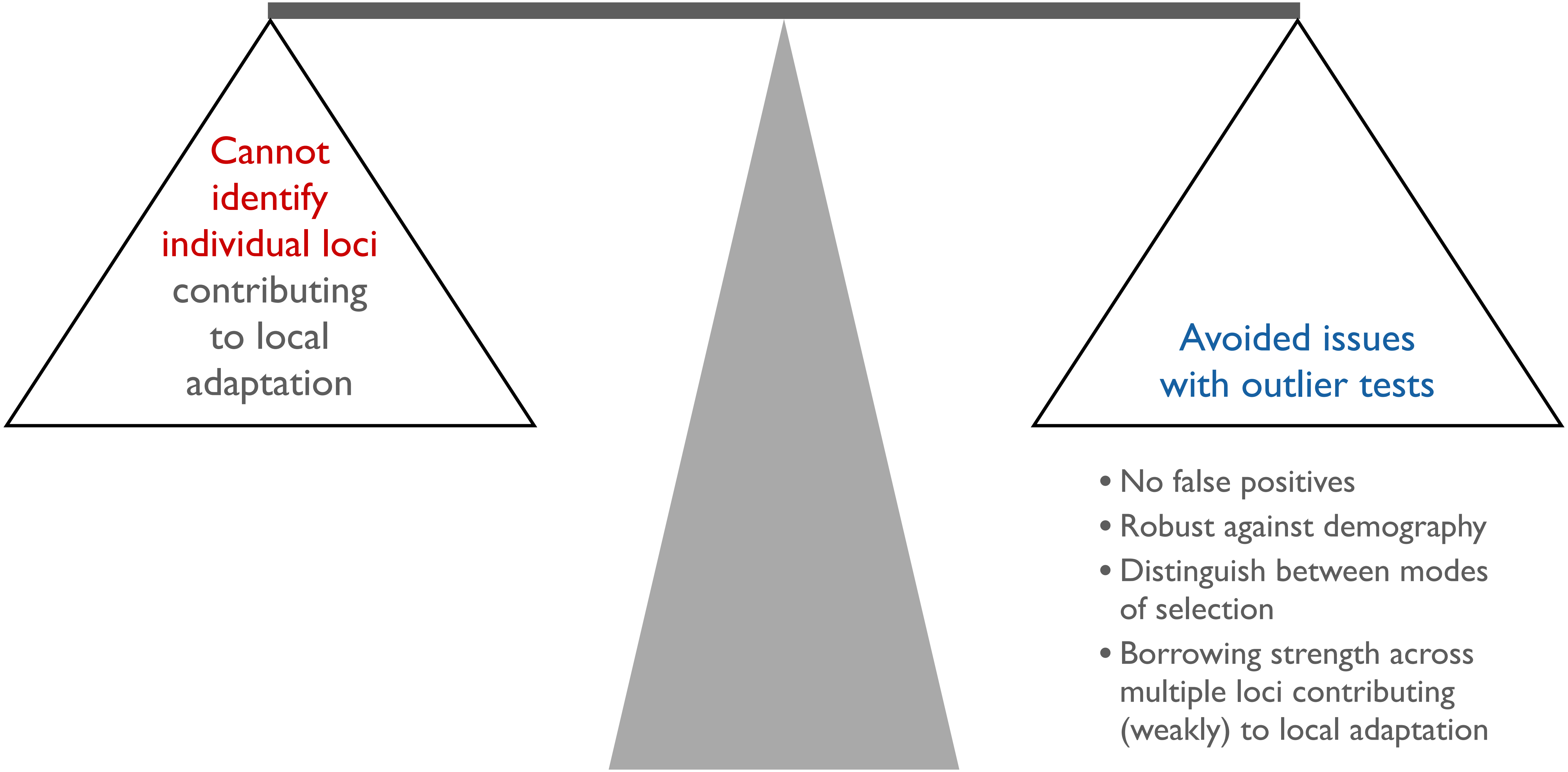
- Quantified the aggregate strength of selection against migration to serpentine patches in *M. guttatus*

Trade-off



Cannot identify individual loci contributing to local adaptation

Trade-off



Cannot identify individual loci contributing to local adaptation

Avoided issues with outlier tests

- No false positives
- Robust against demography
- Distinguish between modes of selection
- Borrowing strength across multiple loci contributing (weakly) to local adaptation

Outlook

- Allow for migration in both directions
- Account for interplay among loci under selection
- Incorporate recent local sweeps and background selection
- Analyse more pairs of *Mimulus* populations on and off serpentine

Thank you!

Photo: D. Begun

Coop Lab, UC Davis:

- Graham Coop
- Chenling Xu

Willis Lab, Duke Univ.:

- Jessica P. Selby
- John Willis

Various:

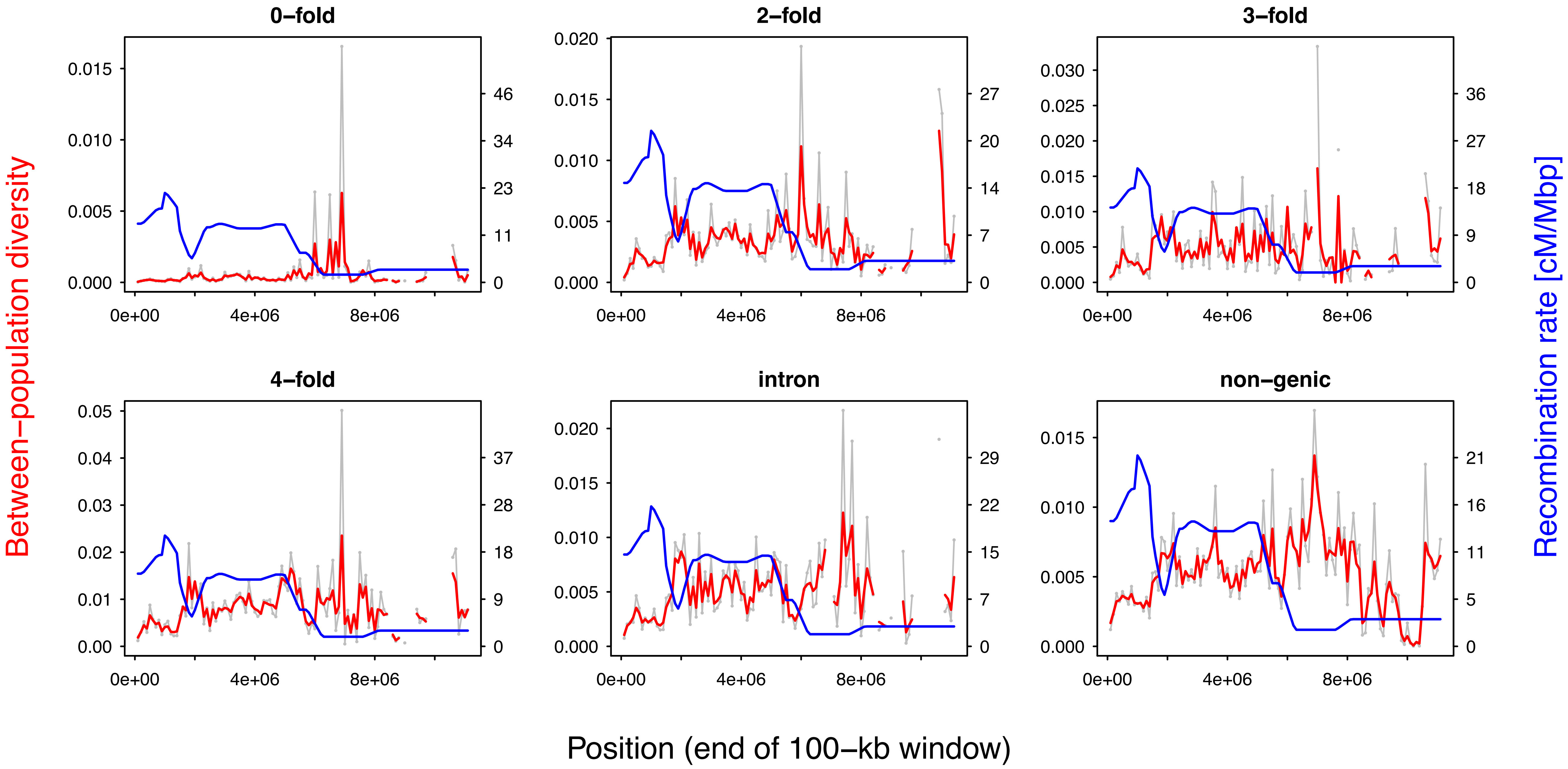
- Yaniv Brandvain (Univ. of Minnesota)
- Lex Flagel, Uffe Hellsten
- Kevin Wright (Harvard University)
- Michael Nachman, Megan Phifer–Rixey (UC Berkeley)

UCDAVIS
UNIVERSITY OF CALIFORNIA

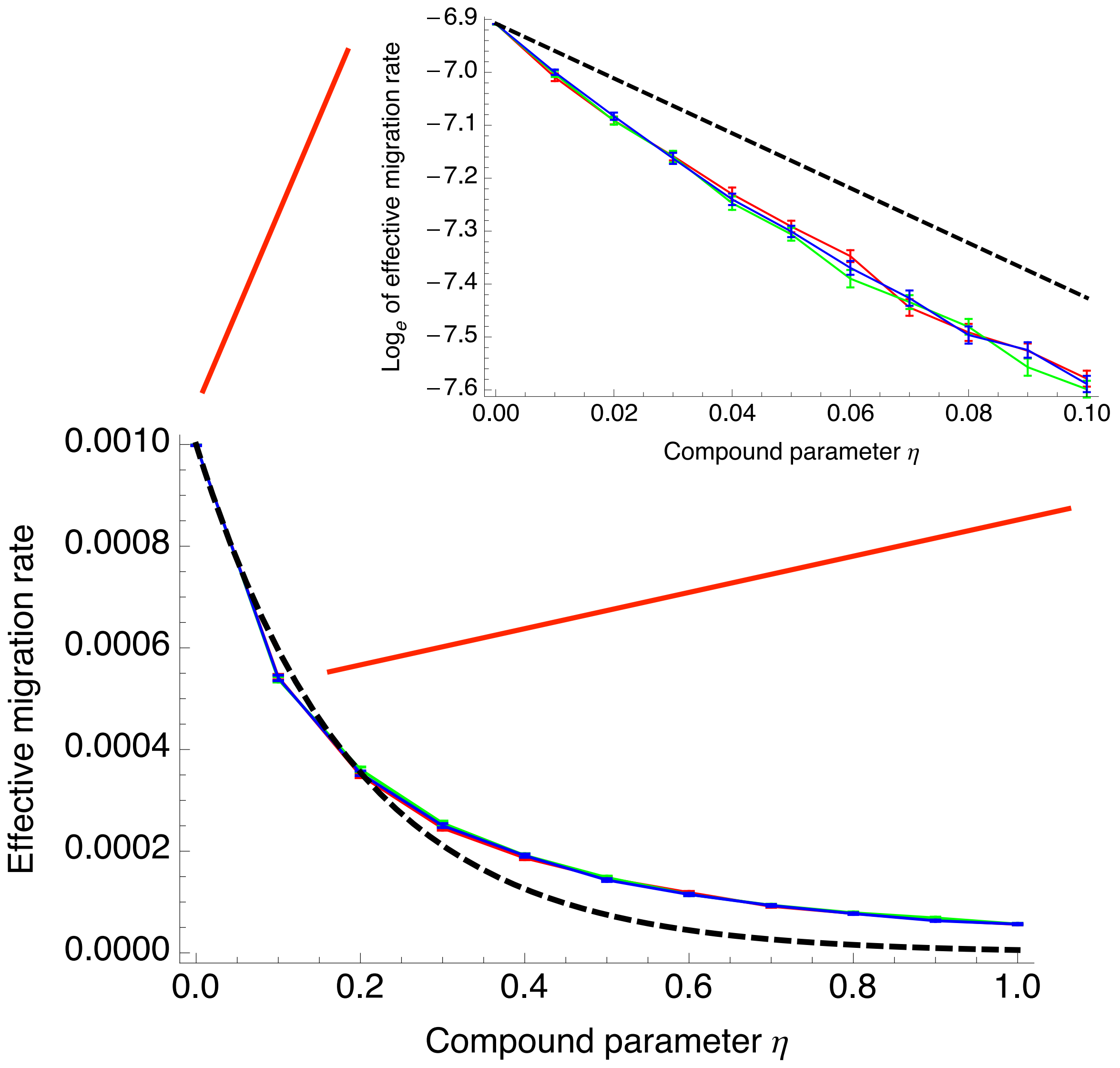
NSF

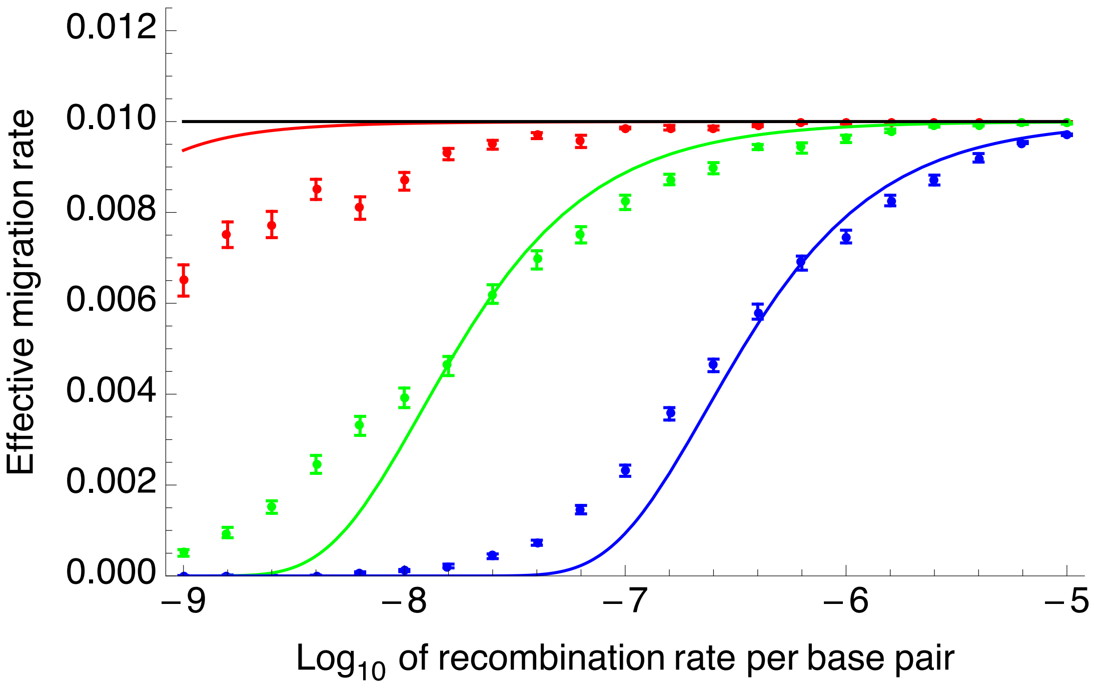
saeschbacher@ucdavis.edu | www.saeschbacher.ch

Divergence and recombination



Confirming the existence of a compound parameter





Why is F_{ST} problematic?

- [Formula of Fst: ratio of two measures of diversity]
- [Illustration: simulation results under local adaptation and background selection, yielding the same pattern of Fst against rec. rate]
- [Illustration: show how π_B distinguishes between the two scenarios]
- [I might not have time for this long version, but just mention it at the end of the previous slide.]

Application: Least-squares fit to *M. guttatus* genomic data

$$
\begin{aligned}
\ln\left(\frac{\hat{\pi}_{\mathrm{B}}}{2u} - 2\hat{N}_e\right) &= -\ln(m) + 2\eta\left\{\ln(K) + \gamma\right\} \\
&= -\ln(m) + \frac{1}{r_{\mathrm{BP}}} 2\nu_{\mathrm{BP}}\bar{a}\left\{\ln(K) + \gamma\right\} \\
&= a + \frac{1}{r_{\mathrm{BP}}} b
\end{aligned}
$$