

# Hitch-hiking and polygenic adaptation

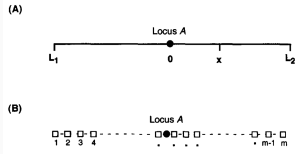
---

Kevin Thornton

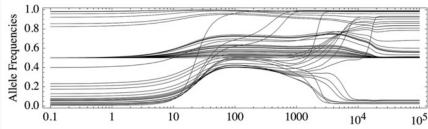
Ecology and Evolutionary Biology, UC Irvine

# Linked selection vs. fates of selected mutations

Hudson & Kaplan, 1995



De Vladar & Barton, 2014

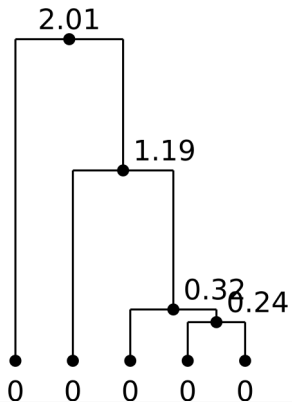


# Modeling traditions

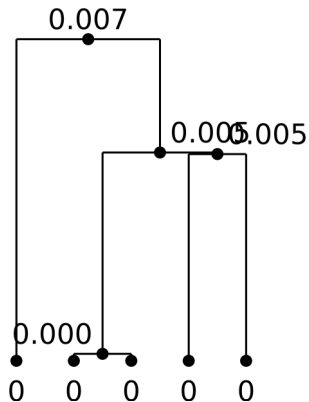
Population genetics	Evol. quantitative genetics
Fixed effect sizes	Variable effect sizes
Single selected site	Many sites
Directional sel'n	Stabilizing selection
Partial linkage	LE or QLE

# Tree structures

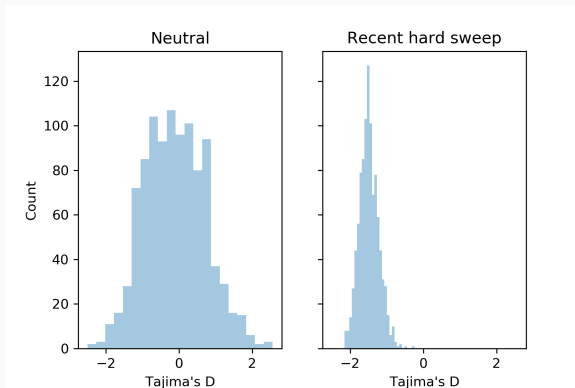
Neutral



Recent hard sweep



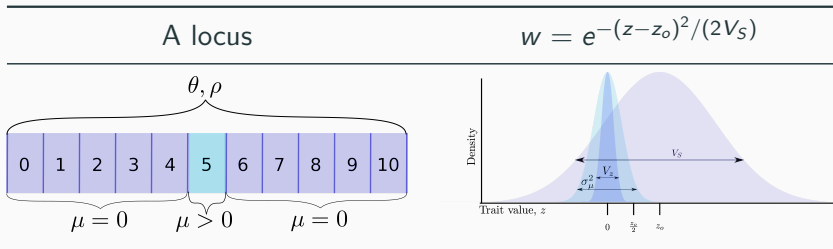
# Patterns reflect the tree structures



## Linked selection during polygenic adaptation

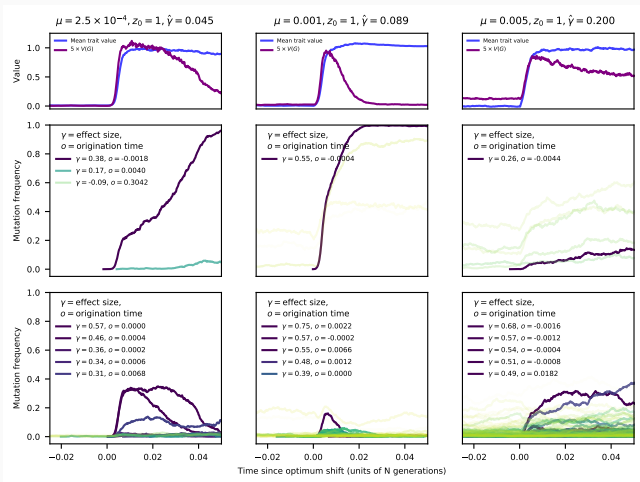
- Use forward simulations
- fwdpy11 is a Python package
- Uses a C++ back-end (Thornton, 2014, Genetics)

# Simulation scheme



- 10 unlinked loci,  $\theta = \rho = 1,000$  per locus
- Additive mutations arise at rate  $\mu$ ,  $\Theta = 4N\mu$
- Two thetas cannot possibly be confusing.
- $N = 5,000$  diploids
- Evolve under GSS for  $10N$  generations with optimal trait value of 0
- Shift optimal trait value to  $z_o > 0$  and evolve for  $10N$  more generations

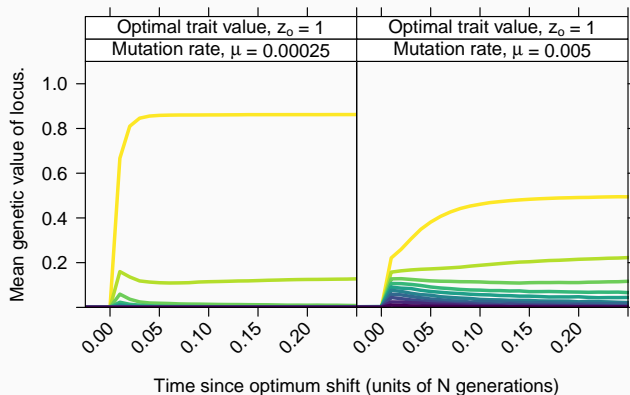
# Adaptation occurs *rapidly* and *before* fixation



**Figure 1:** Large optimum shift,  $z_0 = 1$  with  $V_S = 1$ .

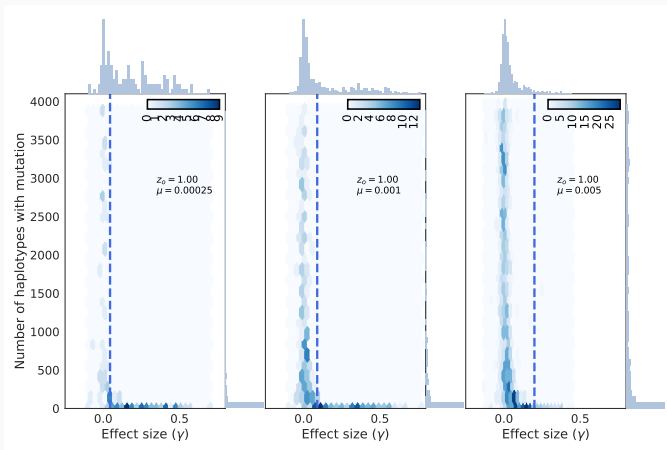


## Contributions of different loci



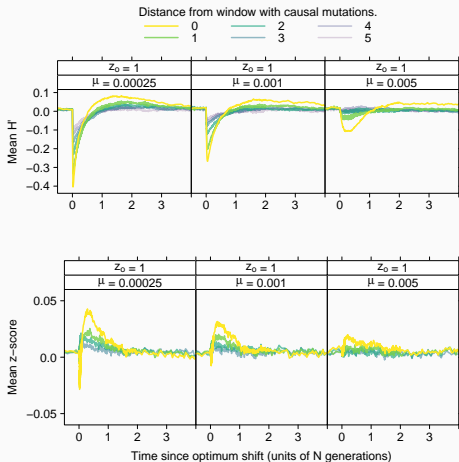
**Figure 2:** Mean trait value *per locus*, colored by rank.

## Sweeps from SGV start out rare



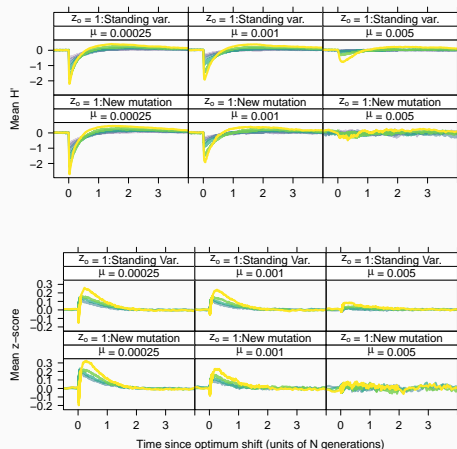
This predicts “hard” sweep signals due to sweeps from large-effect SGV.

# Temporal and spatial patterns of “selection signals”



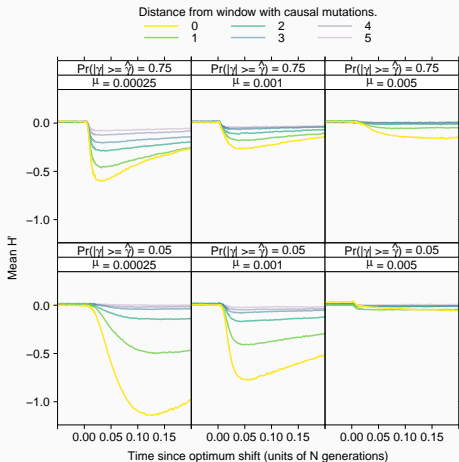
**Figure 3:** Mean statistic per window over time for a large optimum shift.  $z$  scores are for the  $nS_L$  statistic (Ferrer-Admetlla et al. (2014), MBE

# Similar patterns for new mutations vs SVG



**Figure 4:** Same data, but conditioning on fixations of large effect

# Mutational variance matters



**Figure 5:** Choose  $\sigma_\mu$  so that probability of a large-effect mutation is constant. Time scale is determined by  $\delta q$  of fixations.

# Implications

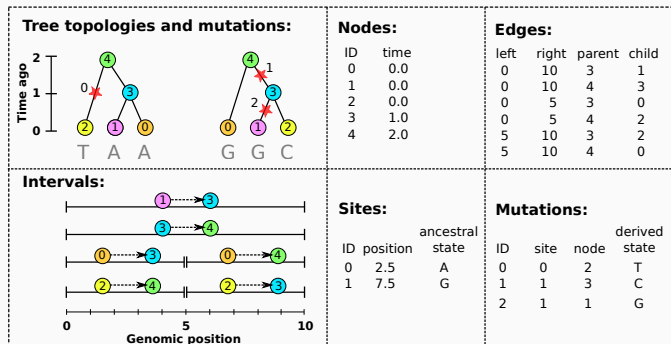
- Patterns unique to “soft sweeps” are not generated by this model!
  - We are using supervised machine learning (Schrider/Kern) to further investigate this.
- Hitch-hiking signals *decrease* as  $\Theta$  *increases*
  - Keep in mind that our “tests” are usually designed to detect hard sweeps

Data not shown:

- Small optimum shifts leave less dramatic patterns

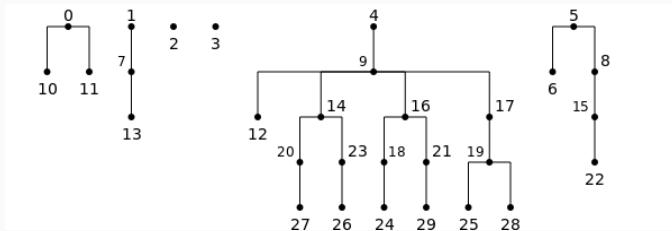
# Tree sequences: representing genetic data using tables

Kelleher, et al. 2016. PLoS Computational Biology  
a.k.a "The msprime paper"



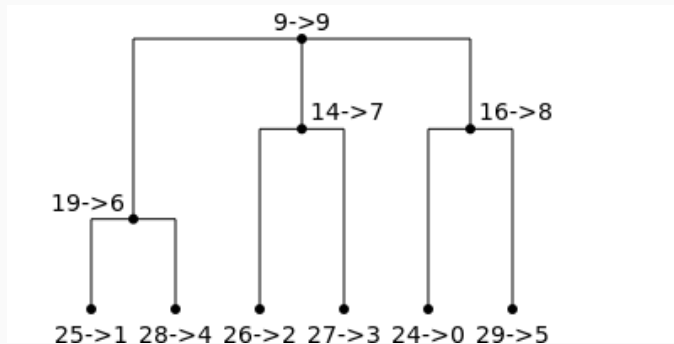
# Tree sequence simplification...

Kelleher, et al. 2018. PLoS Computational Biology

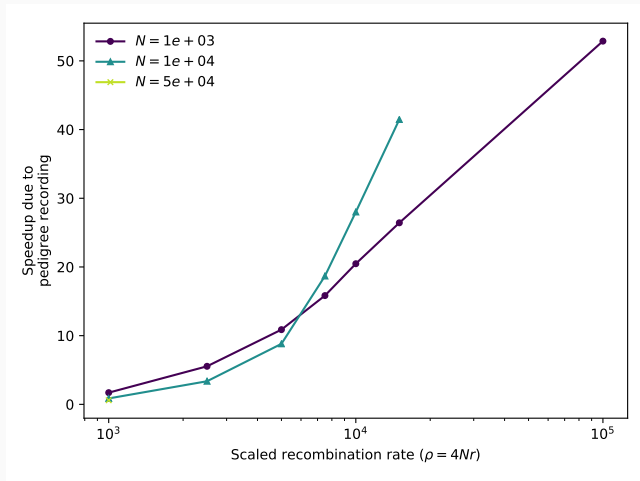




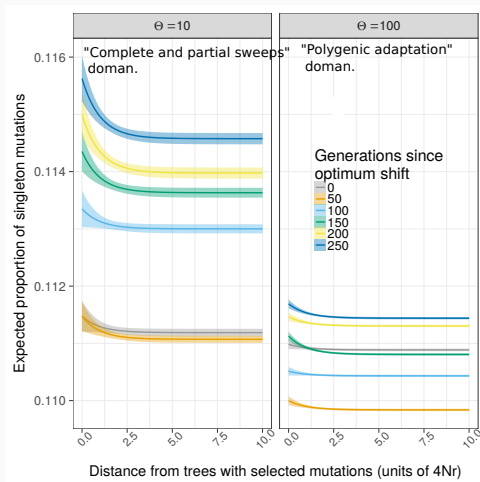
...can be done in FAST linear time...



...and give a huge performance boost...



## ... allowing chromosome-scale simulations in large $N$



**Figure 6:**  $N = 2 \times 10^5$  diploids,  $\rho = 10^5$  ( $\approx 100\text{MB}$  in humans),  $\gamma \sim N(0, 0.25)$ ,  $V_S = 1$ . Analysis based on  $n = 3,000$  diploids.

## Facilitates better testing

- Methods for detecting polygenic adaptation of continuous traits shouldn't be evaluated with simulations of strong sweeps.
- Methods assuming linkage equilibrium need to be tested using simulations involving partial linkage
- etc.

- fwdpy11: <https://fwdpy11.readthedocs.org>
- msprime: <https://msprime.readthedocs.org>
- Tree sequence tutorials: <https://tskit-dev.github.io/tutorials/>
- The tree sequence toolkit: <https://github.com/tskit-dev/tskit> (“almost ready”)

# Thanks

- David Lawrie
- Khoi Hyunh
- Jaleal Sanjak
- Tony Long
- Jerome Kelleher, Jaime Ashander, Peter Ralph
- NIH for funding
- UCI HPCC for computing support