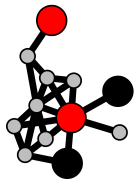


The McDonald-Kreitman Test For Selection: An Extremely Brief Introduction

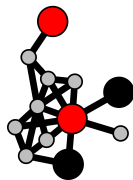
Jake L. Weissman

May 19, 2020



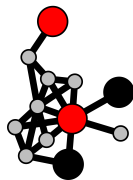
Last Time: dN/dS

- Only needs divergence data
- Unlikely to be > 1 (indicating positive selection) except in extreme cases, not particularly sensitive



This Time: MK-Test

- Needs polymorphism and substitution data
- Much more likely to pick up positive selection



Intuition

- Polymorphism is *recent*
 - Selection hasn't had time to act
 - Strongly positively selected alleles fix so rapidly that you don't see them among polymorphisms
- Therefore polymorphism will not show effects of selection, but substitutions will
- The MK Test compares the fraction of nonsynonymous differences looking at polymorphisms and substitutions respectively, taking polymorphism as the null distribution of nonsynonymous differences (without selection)

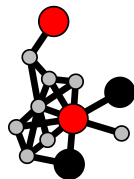


Quantities

The MK-test requires you to calculate four quantities:

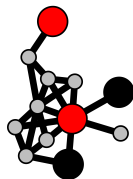
- P_N : Number of nonsynonymous polymorphisms
- P_S : Number of synonymous polymorphisms
- D_N : Number of nonsynonymous fixed difference
- D_S : Number of synonymous fixed differences

	Fixed	Polymorphic
Synonymous	D_S	P_S
Nonsynonymous	D_N	P_N



Typical Data

- 2+ sequences of your gene of interest from your population of interest (to assess polymorphism)
- One (or sometimes more) sequence from an outgroup (to assess fixed differences)



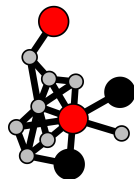
The Test

$$H_0 : \frac{P_N}{P_S} \approx \frac{D_N}{D_S}$$

	Fixed	Polymorphic
Synonymous	D_S	P_S
Nonsynonymous	D_N	P_N

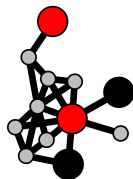
Test for independence:

- Fisher's Exact Test (little data)
- χ^2 (lots of data)



MK-test will often fail to detect positive selection in the presence of weak purifying selection

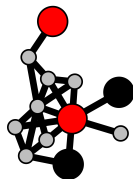
- Several methodological approaches to correct for this (implemented in iMKT package we will use for interactive lesson portion)
- The simplest approach is to just remove low-frequency polymorphisms, since these are more likely to be under weak purifying selection (Fay, Wyckoff and Wu, 2001)



Neutrality Index

$$NI = \frac{\left(\frac{P_N}{D_N}\right)}{\left(\frac{P_S}{D_S}\right)}$$

- Useful for when you want to compare degree of selection across genes



Alpha

$$\alpha = 1 - \frac{D_S P_N}{D_N P_S} = 1 - \text{NI}$$

- Proportion of nonsynonymous substitutions fixed by positive selection (but see Stoletzki and Eyre-Walker, 2011. “Estimation of the neutrality index.” *MBE*)

