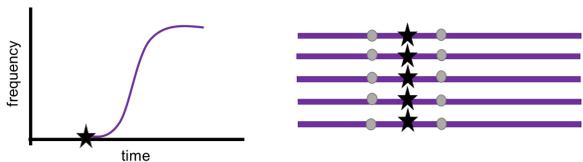


BACKGROUND

Hard sweep: single adaptive mutation rises in frequency

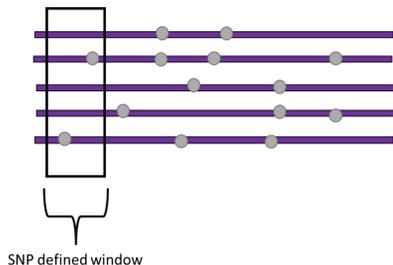


Soft sweep: Multiple adaptive mutations at the same locus sweep through the population

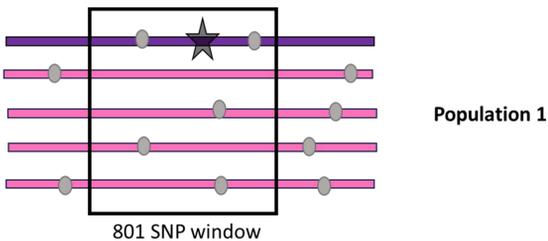


How do we calculate H12?

- H12 scan -> 401 SNP defined windows based on diversity of a single population
- This approach might miss complete hard sweeps because there is a strong dip in diversity around the adaptive site, resulting in few SNPs to define the window with



How do we calculate XP-H12?



Windows defined by Population 1. Populations can be separated either spatially or temporally



APPROACH

- We used SLiM to run partial and complete sweep evolutionary population simulations of 10,000 individuals and collected the data on three timepoints: when the sweep occurred, 100 generations after the sweep, and 300 generations after the sweep.
 - This was done to test how well each of the statistics was able to detect sweeps after the windows had been degraded over time.
- These simulations were also run with varying selection coefficients, 0.005 (weak), 0.05 (moderate), and 0.1 (strong).
- Each of these simulations had 100 final individuals sampled at random and each simulation was run 100 times with a different seed.
- Raw data taken from the simulations was converted and analyzed with H12 and XP-H12.
- A ROC curve was plotted to show the true positive rate and false positive rate to see if the statistics could detect selective sweeps compared to a neutral simulation.

Detection of spatial selective sweeps with haplotype homozygosity statistics

Abstract

Adaptation occurs when a species evolves to become more fit in its environment. Detecting adaptation unveils forces that shape populations, which can be applied to disease or human populations. Adaptation leaves behind selective sweeps—signatures in the genome distinct from neutrality. Recently, the haplotype homozygosity statistic, H12, was developed for detecting hard and soft sweeps (Garud et. al, 2015). In a hard sweep, a single haplotype with a selective advantage rises to frequency in the population, compared to soft sweeps where several haplotypes rise simultaneously. H12 uses SNP defined windows; this statistic has power in numerous evolutionary scenarios, but it may miss complete hard sweeps due to SNP-based windows spanning many base pairs. We propose a new statistic, XP-H12, whereby window size is defined by the diversity of a distinct population, possibly making it easier to detect complete hard sweeps. We tested the ability of the XP-H12 statistic in detecting hard sweeps in several evolutionary scenarios. We find that XP-H12 outperforms H12 in detecting complete sweeps when the selection coefficient is ≥ 0.05 . This research provides insight into the use of XP-H12 in detecting sweeps in populations, highlighting cases where H12 might miss sweeps, and cases where XP-H12 is able to detect such sweeps.

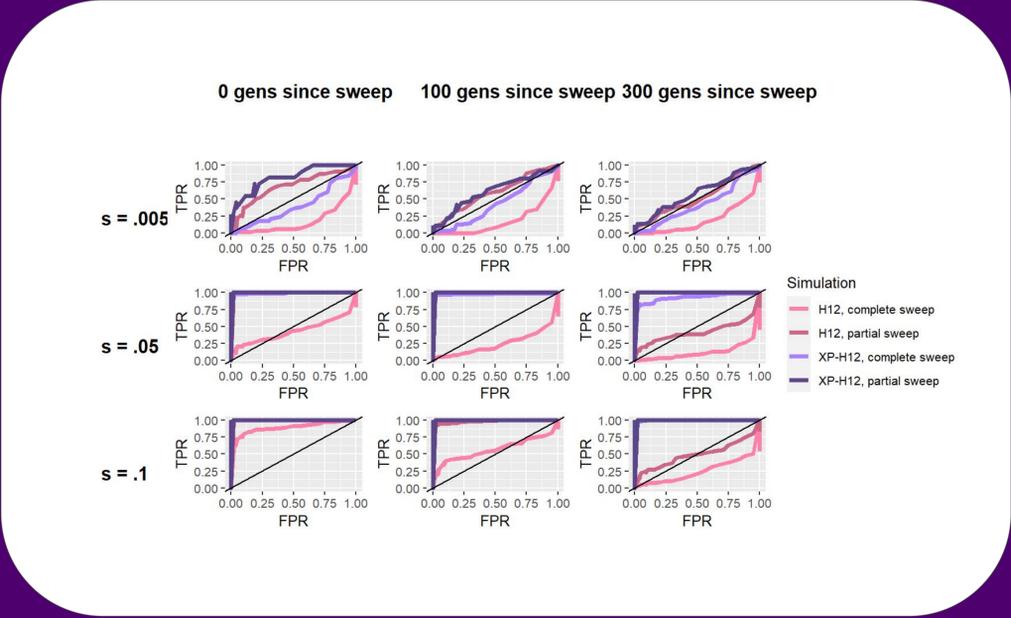


Figure 1.

Authors

Julianna Perez¹, Mariana Harris², Nandita Garud³

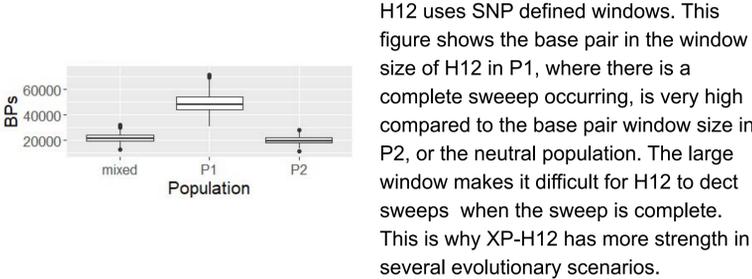
1 BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA
 2 Department of Computational Medicine, UCLA
 3 Department of Ecology and Evolutionary Biology, UCLA

RESULTS

Figure 1. is ROC curves plotted for H12 and XP-H12 under complete and partial sweeps.

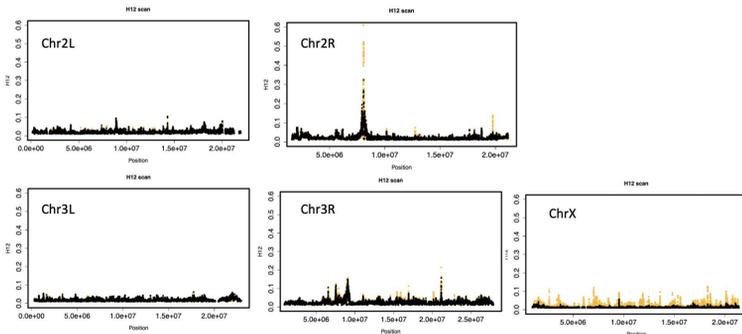
- We found that XP-H12 outperformed H12 in cases of complete sweeps.
- XP-H12 also outperformed H12 after the sweep gets older and the windows had been degraded.
- XP-H12 maintained its strong ability to detect sweeps when the selection coefficient was higher at 0.05 and 0.1 even after 300 generations.
- These promising results show that XP-H12 may be able to detect many sweeps, especially complete sweeps that H12 might have originally missed.

Why does H12 perform worse than XP-H12 in certain scenarios?



H12 uses SNP defined windows. This figure shows the base pair in the window size of H12 in P1, where there is a complete sweep occurring, is very high compared to the base pair window size in P2, or the neutral population. The large window makes it difficult for H12 to detect sweeps when the sweep is complete. This is why XP-H12 has more strength in several evolutionary scenarios.

DISCUSSION



This figure is a scan conducted in *D. melanogaster* data, made by Mariana Harris. The black lines are the sweeps on *Drosophila melanogaster* chromosomes detected by H12. The yellow lines are the sweeps detected by XP-H12. To compute XP-H12 we use a presumably ancestral population from Zambia to look for sweep signatures in a derived population from North America. This scan highlights the fact that XP-H12 likely has more power to detect sweeps than H12 and that it may allow the detection of sweeps completely missed by H12 alone.

XP-H12 is a promising statistic that has the ability to detect sweeps that may have previously gone undetected. Although currently being applied to *Drosophila melanogaster* in PhD student Mariana Harris's work, the applications of this statistic could extend to other evolutionary scenarios. In regards to the current endemic, this statistic might be useful for detecting sweeps in the Sars-CoV 2 virus in order to detect changes in the virus. The strength of this statistic could lead to new discoveries in the sweeps and evolutionary forces at play shaping populations.

Acknowledgements



Special thanks to UCLA BIG summer program and The NSF REU program for funding this research.