Title: A reference for population genetic inference: GC-neutral mutations in short introns

Authors: Claus Vogl and Burcin Yildirim

Abstract: The base composition along chromosomes approximately follows Chargaff's second parity rule, which suggests a symmetric mutation model leading to six classes of polymorphism among the four bases. Of these the two GC-conservative mutation classes are unaffected by GC-biased gene conversion and are expected to have symmetric polymorphic site frequency spectra. For each of the six mutation classes, maximum likelihood estimators of the scaled mutation rate and estimators of divergence can be inferred. The key assumption for both types of estimators is the lack of double mutations, which means for the former small scaled mutation rates and for the latter a short split time. Combinations of these estimators allow for separation of population genetic forces. Joint site frequency spectra of D. melanogaster and D. simulans of the nucleotides 8-30 in the 5'-loop region in short introns of fruitflies of the genus Drosophila approximately fulfill the above assumptions and have previously been used as a neutral reference. Data analyses of the GC-changing mutation classes show the influence of a directional force, likely GC-biased gene conversion, which confounded inference of other non-adaptive forces in earlier studies. On the other hand, the two GC-conservative mutation classes can indeed serve as a reference: as predicted by the neutral theory, the inferred effective population size of the X chromosome is three-fourth that of the autosomes. Furthermore, outside telomeres and centromeres the scaled mutation rate is unaffected by the crossover rate, suggesting little influence of linked selection on variability.