Perspective

Prevalence of negative frequency-dependent selection, revealed by incomplete selective sweeps in African populations of *Drosophila melanogaster*

Yuseob Kim*

Department of Life Science and Division of EcoScience, Ewha Womans University, Seoul 03760, Korea

Positive selection on a new beneficial mutation generates a characteristic pattern of DNA sequence polymorphism when it reaches an intermediate allele frequency. On genome sequences of African *Drosophila melanogaster*, we detected such signatures of selection at 37 candidate loci and identified "sweeping haplotypes (SHs)" that are increasing or have increased rapidly in frequency due to hitchhiking. Based on geographic distribution of SH frequencies, we could infer whether selective sweeps occurred starting from *de novo* beneficial mutants under simple constant selective pressure. Single SHs were identified at more than half of loci. However, at many other loci, we observed multiple independent SHs, implying soft selective sweeps due to a high beneficial mutation rate or parallel evolution across space. Interestingly, SH frequencies were intermediate across multiple populations at about a quarter of the loci despite relatively low migration rates inferred between African populations. This invokes a certain form of frequency-dependent selection such as heterozygote advantage. At one locus, we observed a complex pattern of multiple independent that was compatible with recurrent frequency-dependent positive selection on new variants. In conclusion, genomic patterns of positive selection are very diverse, with equal contributions of hard and soft sweeps and a surprisingly large proportion of frequency-dependent selection in *D. melanogaster* populations. [BMB Reports: Perspective 2018; 51(1): 1-2]

Explosive growth of next-generation sequencing data allows us to observe an enormous amount of genetic variation within and between species. Evolutionary biologists are now moving briskly to discover genomic patterns that can reveal species' phylogenetic and demographic history, evidences of both positive and negative Darwinian selection, and origins of important biological processes. It is also revising our basic understanding of evolutionary processes mainly guided by classical theories of population genetics.

One of the most fundamental questions in evolutionary biology is how beneficial mutations arise and spread in natural populations. Beneficial mutations are defined as those that increase carriers' fitness (i.e., expected number of offspring or how well the individual under consideration can survive and reproduce). After a beneficial mutation arises and avoids an initial stochastic loss, its frequency in the population is expected to increase exponentially. Under a simple assumption that the selective advantage of this beneficial allele is constant over time, it will eventually reach fixation (reaching 100% frequency) as described by the standard mathematical theory of positive selection (Diagram 1A). However, such a frequency trajectory has not been demonstrated in natural populations. Furthermore, there are multiple biological scenarios under which the strength of selection may gradually diminish toward zero as the beneficial allele increases to an intermediate frequency, a process termed negative frequency-dependent selection (NFDS). The fitness of a heterozygote with one copy of mutant allele might be higher than that of a mutant homozygote. In addition, a new phenotype might be advantageous only when it is rare under resource competition and other ecological conditions.

We tested whether beneficial alleles in the genome of African *Drosophila melanogaster* would propagate according to the standard theory by observing the distribution of "sweeping haplotypes (SHs)" over different populations. An SH is produced by recent positive selection in a process called genetic hitchhiking or selective sweep (Diagram 1B). When a beneficial allele increases rapidly up to an intermediate frequency, neutral variants at nearby loci on the same chromosome will "hitchhike" along, thus forming a distinct block of homogeneous sequences among randomly sampled
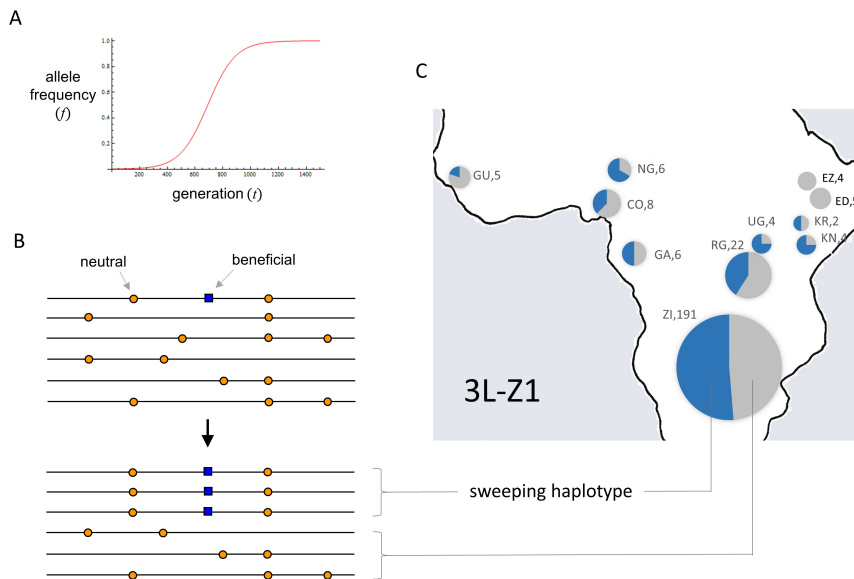
Diagram 1. (A) Allele frequency (*f*) of a beneficial allele with selective advantage (*s*) over time (*t*) is given by the classical equation, $df/dt = sf(1 - f)$. (B) On an interval of chromosome small enough to ignore recombination (crossover) during the exponential growth phase of a beneficial allele, DNA sequences carrying the beneficial allele (blue square) have identical neutral variants (orange circles), thus defining a "sweeping haplotype". (C) Geographical distribution of SH at a candidate locus (named 3L-Z1) under selection. Frequencies of SH at 11 African populations are shown by blue portions of pie charts. Sizes of pie charts reflect sample sizes.

haploid genomes. The frequency of SH in a sample provides a good estimate for the frequency of the putative beneficial allele.

Haploid genome sequences made publicly available by The Drosophila Population Genomics Projects were sampled from eleven populations scattered over Sub-Saharan Africa (Diagram 1C). Genome-wide genetic differentiation among populations indicate that individuals migrate between them at small rates. Under the standard model of constant selection with a small migration rate, if a beneficial allele arises in one population, it will increase to a large frequency by the time it arrives at another population and start to propagate there. However, similar intermediate frequencies of SHs (thus beneficial alleles) as observed in a locus shown in Diagram 1C cannot be explained by the standard model, whereas it is compatible with the model of NFDS in which a beneficial allele stays very long at an intermediate frequency, therefore allowing migrations to equilibrate the pattern of variation among populations. Of 37 loci at which we detected clear SHs, at least eight loci exhibited such evidences of NFDS. Actual proportion of such loci might be even higher because NFDS could not be detected if selective pressures were spatially localized.

We also found that positive selection produced multiple SHs arising from independent mutation events at about half loci. Such results, termed "soft" selective sweeps, are explained by adaptive evolution proceeding in parallel over space or a high rate of mutation to beneficial alleles per locus. In one locus located in chromosome 2R, three to five independent SHs per population, the sum of which however did not reach fixation, were observed. This complex haplotype pattern suggests an interplay between soft sweeps and NFDS.

Overall, this study suggests that diverse modes of positive selection, not anticipated in the classical theory of population genetics, are operating in the African population of *D. melanogaster*. Whether a similar set of selection modes operate in other species including humans needs to be determined in the future. Further investigations are also needed to elucidate the cause of NFDS and the fate of those beneficial alleles, namely whether and how fixations at such loci would eventually occur.

ACKNOWLEDGEMENTS