Genetic variation in populations: – Human and Drosophila data suggest that harmful recessive mutations have unexpected effects on variation in low recombination regions

John F.Y. Brookfield

School of Life Sciences

University of Nottingham

University Park

Nottingham NG7 2RD

United Kingdom

Summary

New data are causing the standard model for the effect of selection on linked neutral variation in low recombination regions, combining the effects of background selection and selective sweeps, to be refined to include harmful recessive mutations creating associative overdominance.

Since the origin of population genetics there has been a debate about the causes of genetic variation in populations. The standard neutral model of population genetic variation supposes that there is a constant loss of variation through genetic drift, whereby the sharing of ancestry of individuals within a population results in a phylogenetic tree of alleles with a time depth in generations that increases with the effective size of the population. Superimposed on this statistically predictable model for the sharing of ancestry are neutral mutations, with the result that the predicted level of population genetic variation depends on the product of the effective population size and the neutral mutation rate.

But this neutral, “non-Darwinian”, population genetic vision ignores the selection that operates on DNA sequence variants. Such selection not only influences the changes in frequency of the selected variants themselves, but affects variation at linked sites [1]. One result of selection at linked sites is a

reduction in levels of genetic variation in chromosomal regions with low recombination [2]. This reduction has been thought to be due to a combination of two mechanisms, selective sweeps (SS), created by selectively advantageous variants being spread by natural selection, and background selection (BGS) [3], where it is weakly deleterious mutations that reduce variation, through increasing the relatedness of mutation-free chromosomes. Both of these mechanisms are most effective at reducing variation in flanking neutral variation when the local recombination rate is low. But two new papers [4, 5] highlight an under-appreciated further force, known as associative overdominance (AOD), which can cause an increase in genetic variation in low recombination regions relative to that expected from SS and BGS.

The principle of associative overdominance is illustrated in Figure 1. Here is imagined a stretch of chromosome where there is no recombination (as in the Drosophila 4th chromosome, for example). There are four genes. In this stylised situation, the population consists of three haplotypes, where each haplotype has a harmful recessive mutation in a different gene. An individual that is homozygous for any one of these haplotypes will show the effects of its recessive mutation and will be unfit, whereas any individual with two different haplotypes will carry two recessive mutations at different genes and will be of full fitness. So all three haplotypes will be stably maintained in the population. The stability comes from the result of Hardy-Weinberg principle that rare alleles (or haplotypes) are usually heterozygous, and so the disadvantage to a haplotype with a recessive mutation increases as its frequency increases because it becomes more likely to be homozygous. A fourth gene has no recessive harmful mutation, and I imagine that all variation in this gene is neutral. However, because of the selective maintenance of the three haplotypes, diversity in the neutrally-varying gene is enhanced. This is because the lack of recombination creates association, so-called “linkage disequilibrium”, between variants in the neutral locus and the loci under selection. There are three important features of this model. Firstly, the stable maintenance of haplotypes does not necessarily come from there being any heterozygous advantage (i.e. overdominance for fitness)

at any one of the genetic loci, so this is not the classical model for balancing selection at a single locus, although balancing selection at single loci can also create AOD at linked neutral loci. Secondly, the stable maintenance of multiple haplotypes relies on an absence of the recombination between haplotypes that would create fully wild-type haplotypes, lacking any harmful mutations, which would then be the fittest of all and which would replace the haplotypes bearing harmful mutations. Finally, that there are no wild-type haplotypes, even without recombination, requires these to have been lost, which must have occurred by genetic drift in a finite population. The likely occurrence of AOD thus will depend in a subtle way on the effective population size, the strength of selection against, and the rate of production of, harmful mutations, how recessive these are, and the local level of recombination [6].

Interest in AOD has a long history. In the 1960s, starch gel electrophoresis demonstrated polymorphisms for amino acid substitutions that changed the electric charge of enzyme molecules, which thus had differing electrophoretic mobilities. This engendered a debate whether these polymorphisms were acted on by selection. For those who believed that selection maintained amino acid sequence variation at these loci, one preferred mechanism was heterozygote advantage, i.e. overdominance for fitness. In principle, this can be identified through examining the fitnesses of homozygotes and heterozygotes, but apparent heterozygote advantage at a locus of interest could, in principle, be created by AOD resulting from selection at linked loci. Ohta [7,8] examined quantitatively the parameters that determined the likelihood of AOD operating, specifically in the context of a neutral locus linked to a locus showing true overdominance for fitness [7], or through linkage to loci with an input of harmful recessive mutations [8]. Further evidence thought to support the view that allozyme variants show overdominance for fitness came from positive overall correlations seen between fitness measures and individual heterozygosity at allozyme loci. Here, Pamilo and Pálsson [9] demonstrated, by simulation, that this can arise from AOD created by weakly

harmful recessive mutations in small populations, without selection operating at the allozyme loci themselves.

But is AOD operating in reality and having a detectable effect on levels of variation in low recombination regions? A combination of BGS and SS will have the result that there is low variation in low recombination regions. But this model also predicts a so-called "skew" in the site frequency spectrum (SFS). The SFS is predicted by the standard neutral model mentioned in the first paragraph. If there is BGS and SS one expects, and generally sees, a skew in the distribution towards low frequency variants. However, data from humans and from *Drosophila simulans* [4] show that there is far less skew in the non-recombining regions of the genome than expected (although there is still some relative to the standard neutral model's predictions). Simulations were carried out where the harmful mutations, whose input drives BGS, were made mainly recessive, with the heterozygote for the mutation only showing 20% of the mutation's homozygous effect (as opposed to 50% in the standard BGS model) and where individual mutations differed in the strength of their harmful effects [4]. It was found that for some parameter values, specifically when a proportion of mutations are weakly selected, relative to the reciprocal of the effective population size, AOD appears in no-recombination regions, increasing diversity and decreasing skew.

The key element in creating AOD is that the wild-type haplotype is lost (see Figure 1). Gilbert et al. [5] explore the conditions where this is likely, and find that loss of the wild-type haplotype is enhanced the more mutable loci there are. But selection must not be too strong, or the mutations are purged before they can reach high frequencies. Consideration of AOD thus indicates that there can be enhancements in genetic diversity in low recombination regions. Thus, Gilbert et al. [5] also looked at windows of the human genome which included adjacent low-recombination and higher, medium-recombination regions, and looked at the levels of gene diversity to find cases where the low-recombination region had higher diversity. In this way, they identified 22 regions of the human

genome as candidates for AOD. However, it remains uncertain whether these are cases of AOD or of true balancing selection, which leaves a similar signature in its effects on the SFS.

Clearly, with the avalanche of population genomic data from vast numbers of species, the search for AOD will be another role in which these new datasets can increasingly be used.

References:

1. Hill, W.G., and Robertson, A. (1968) Linkage disequilibrium in finite populations. Theor. Appl. Genetics 38, 226-231.

2. Begun, D., and Aquadro, C.F., (1992) Levels of naturally occurring DNA polymorphism correlate with recombination rate in *Drosophila melanogaster*. Nature, 356, 519-520.

3. Charlesworth, B., Morgan, M.T., and Charlesworth, D. (1993) The effect of deleterious mutations on neutral molecular variation. Genetics, 134:1289-1303.

4. Becher, H., Jackson, B., and Charlesworth, B. (2019) Patterns of genetic variability in genomic regions with low rates of recombination. Current Biology ??

5. Gilbert, K.J., Pouyet, F., Excoffier, L., and Peischl, S. (2019) Transition from background selection to associative overdominance promotes diversity in regions of low recombination. Current Biology ??

6. Zhao, L., and Charlesworth, B., (2016) Resolving the conflict between associative overdominance and background selection. Genetics 203, 1315-1334.

7. Ohta, T., and Kimura, M. (1970) Development of associative overdominance through linkage disequilibrium in finite populations. Genet. Res. Camb. 16, 165-177.

8. Ohta, T. (1971) Associative overdominance caused by linked detrimental mutations. Genet. Res. Camb. 17, 277-286.

9. Pamilo, P., and Pálsson, S. (1998) Associative overdominance, heterozygote advantage and fitness. Heredity 81, 381-389.

Figure 1: A stylised representation of a non-recombining region of a chromosome with four genes, where the population consists of three haplotypes, with harmful recessive mutations in, respectively, genes A, C and D (shown by stars). Variation in gene B is neutral. However, because homozygotes for any haplotype will be of reduced fitness, the three haplotypes will be retained in the population. This will have an impact on the structure of variation in gene B, increasing diversity and reducing skew, through an associative overdominance (AOD) for fitness. The system of AOD could be destroyed if recombination were to create a haplotype with none of the mutations.