Effects of Intra-gene Fitness Interactions

on the Benefit of Sexual Recombination

Richard A. Watson[1], Daniel M. Weinreich[2], John Wakeley[2]

[1.] Natural Systems group, School of Electronics and Computer Science, Southampton University, UK.

[2.] Dept. of Organismic and Evolutionary Biology, Harvard University, USA.

Abstract

Whereas spontaneous point mutation operates on nucleotides individually, sexual recombination manipulates the set of nucleotides within an allele as an essentially particulate unit. In principle, these two different scales of variation enable selection to follow fitness gradients in two different spaces: in nucleotide sequence space, and allele sequence space, respectively. Epistasis for fitness at these two scales, between nucleotides and between genes, may be qualitatively different and significantly influence the advantage of mutation-based and recombination-based evolutionary trajectories, respectively. We examine scenarios where the genetic sequence within a gene strongly influences the fitness effect of a mutation in that gene, whereas epistatic interactions between sites in different genes are weak or absent. We find that, in cases where beneficial alleles of a gene differ from one another at several nucleotide sites, sexual populations can exhibit enormous benefit over asexual populations: not only discovering fit genotypes faster than asexual populations, but also discovering high-fitness genotypes that are effectively not evolvable in asexual populations.

Popular models for the benefit of recombination (e.g. 1, 2, 3, 4) assume that the alleles of a gene are mutational neighbours and thus the space of variations possible by recombination of alleles are merely a subset of those available by mutation. However, when the alleles of genes differ at several sites, the substitution of an allele by recombination can create new haplotypes that are not immediate neighbours of extant haplotypes under mutational variation. The significance of genetic variation at these two different scales, combinations of sites and combinations

of alleles, will be affected by the nature of epistatic fitness interactions between units at these two scales. That is, natural selection acting on allelic variation enabled by recombination is affected by epistasis between alleles but it is unable to respond to the details of fitness interactions among individual nucleotides within each gene since combinations of nucleotides vary as a group under this variation mechanism. In contrast, natural selection acting on mutational variation alone will follow fitness gradients in nucleotide sequence space, and an asexual population will not be able to respond directly to fitness gradients in allele sequence space since it cannot vary units of genetic material at this scale.

We examine scenarios where epistatic fitness dependencies between units at these two scales, between nucleotides and between genes, are qualitatively different, and examine the impact that this has on the benefit of sex. Specifically, we assume that the fitness effect of a mutation is strongly influenced by the genetic sequence of the gene that it occurs in but relatively weakly influenced by the genetic sequence of other genes. For example, where the DNA sequence of a gene determines the shape and function of the protein that it codes for, the proper functioning of that gene and hence the fitness impact of mutations therein may be strongly dependent on detailed interdependencies between the nucleotides in that sequence. In contrast, we assume that epistasis between sites in different genes is relatively weak. In fact, in contrast to previous models for the interaction of sex and epistasis (e.g. 4), the benefit of sex that we study here does not require *any* epistasis between genes. If we may refer to the fitness dependencies between individual nucleotides *within* a gene as 'intra-gene epistasis', and to the, more widely studied, fitness dependencies *between* genes as 'inter-gene epistasis' then our model assumes that intra-gene epistasis is stronger than inter-gene epistasis.

In the manner of Kim and Orr (5), we study the adaptation of sexuals and asexuals in a DNA sequence-based model of evolution following Gillespie's (6) mutational landscape model. As per Kim and Orr, we study a large but finite population that experiences a sudden change in environment, whereby a finite number of sites then admit beneficial mutations. These mutations have differing selective effect and occur recurrently. Kim and Orr found that under these conditions the Fisher/Muller model for the elimination of competition among beneficial mutations (7) is weak, and that sexual and asexual populations behave similarly. However, their model assumes that mutation and recombination produce variation on the same physical scale, conflating nucleotide sites and recombinant loci, whereas we model loci that each contain many tightly linked sites with strong epistasis among those sites.

One way to understand the absence of a significant effect in Kim and Orr's substitution model is that, as per Fisher and Muller's assumptions, it is assumed that the fitness landscape is single-peaked. Thus, although asexuals may be slower to make fitness improvements, they will soon enough find the same fitness peak that sexuals find. If however, the fitness landscape is multi-peaked, the story can be very different – even when the number of beneficial mutations is finite and mutations occur recurrently. In finite populations with stochastic effects of sampling and of mutation (8), the particular local fitness peak a population finds in a multi-peaked fitness surface may vary. In our model the fitness of the peak that is likely to be found by asexual populations will be lower in fitness than that found by sexual populations. This produces a sustained advantage to sexuals. Moreover, because in a multi-peaked landscape mutations that are beneficial in some genetic backgrounds are deleterious in others (9), it is not necessarily the case that a given mutation of large benefit

discovered in a sexual population is also available to an asexual individual – this creates greater variance, as well as sustained variance, in the fitnesses of individuals, and increased benefit to sexual populations.

Of course, the significance of multiple peaks in fitness landscapes has been debated at length (10). But this debate has been about local optima in allele sequence space created by epistasis between genes. Here we assume free recombination between genes precluding selection on combinations of alleles, thus agreeing with Fisher that changes in allele frequency will respond only to the additive component of the fitness effect of alleles (1). Inter-gene epistasis is therefore not of consequence to this model. The possibility of multiple peaks in the intra-gene DNA sequence landscape is a different issue entirely. Evolutionary trajectories driven by mutational change will be constrained by epistasis in sequence-space (8), but nonetheless selection on alleles, that is, selection on the particular combination of nucleotides within a gene, is maintained in sexual populations by virtue of their tight physical linkage.

The basic intuition of our model is thus as follows. In asexual populations, individuals will accumulate beneficial mutations, ultimately finding some local peak in DNA sequence space. At this point, all beneficial mutations are exhausted within all genes. For a given gene, some individuals may have found better alleles than other individuals, and in general any one individual will have a set of alleles conferring a range of different fitnesses, some alleles better than average and some below average in comparison to the rest of the population. An individual that has a good combination of alleles will be selected for, but an individual that has all the best alleles will not be present in the population. A further fitness improvement would require a specific multi-point mutation that jumps to a higher fitness genotype (or an evolutionary

trajectory that moves counter to selective gradients to traverse the fitness valley between local optima). In a sexual population, as per the asexual population, some individuals may find better alleles than other individuals, for a given gene. However, unlike the asexual population, the sexual population can simply select for the best alleles that have arisen in the population, independently of the backgrounds they arose in, as per Fisher and Muller's intuition. The difference is that, although this model and Fisher/Muller's original model both describe an advantage for sexuals derived from the independent selection of beneficial alleles, in the Fisher/Muller model beneficial alleles are always available assuming recurrent mutation, but here some superior alleles will be selectively inaccessible from other inferior alleles and are thus effectively permanently unavailable to asexual populations.

Using this model we can show a dramatic and sustained advantage to sexual populations. Some limitations of the model are apparent, however. First, in panmictic populations, all individuals tend to find the same fitness peak even though multiple peaks exist and thus recombination of the resultant converged population has negligible effect. In a subdivided population, where higher genetic diversity is maintained, the effect is shown strongly. Second, as mentioned, if the alleles that are manipulated by recombination differ from one another by only single nucleotide substitutions, then any allele substitution enabled by recombination could also be enabled by a point mutation. Again, subdivision may strongly influence how likely it is that alleles differ at multiple sites.

Whereas previous models have shown a distinction in the average fitness of genotypes in the population (4) or the speed of accumulation of beneficial mutations (3), we believe this is the first biologically plausible model to show, given these

caveats, a scenario where fit genotypes are evolvable in sexual populations that are

unreachable in asexual populations (11, 12).

1. Fisher, R.A. (1930) *The Genetical Theory of Natural Selection*. Oxford University Press, Oxford

2. Muller, H.J. (1932) *Am. Nat.* 66, 118–138

3. Hill, W.G., and Robertson, A. (1966) *Genet. Res.* 8, 269–294

4. Kondrashov, A.S. (1988) *Nature* 336, 435-440

5. Kim, Y. and Orr, H.A. (2005) *Genetics*, 171:3, 1377-1386

6. Gillespie, J.H. (1984) *Evolution* 38, 1116–1129

7. Otto, S.P., and Lenormand, T. (2002) *Nat. Rev. Genet.* 3, 252–261

8. Otto, S.P., and Barton, N.H. (2001) *Evolution* 55, 1921–1931

9. Weinreich, D.M., Watson, R.A., and Chao, L. (2005) *Evolution* 59(6), 1165-1174

10. Wright, S. (1967) *Proceedings National Academy of Sciences* 58: 165–179

11. Watson, R.A. (2006) *Compositional Evolution*, MIT Press

12. Jansen, T., and Watson, R.A. (2006) "A Building-Block Royal Road for Crossover", *Proceedings of Parallel Problem Solving From Nature (PPSN IX)*, submitted