

NEWS AND VIEWS

PERSPECTIVE

The impact of epistatic selection on the genomic traces of selection

SARAH P. OTTO and MICHAEL C. WHITLOCK
*Department of Zoology & Biodiversity Research Centre,
 University of British Columbia, Vancouver, BC, Canada V6T 1Z4*

The rapid accumulation of genomic data has led to an explosion of studies searching for signals of past selection left within DNA sequences. Yet the majority of theoretical studies investigating the traces of selection have assumed a simple form of selection, without interactions among selectively fixed sites. Fitness interactions—'epistasis'—are commonplace, however, and take on a myriad of forms (Whitlock *et al.* 1995; Segrè *et al.* 2005; Phillips 2008). It is thus important to determine how such epistasis would influence selective sweeps. On p. 5018 of this issue, Takahasi (2009) explores the effect of epistasis on genetic variation neighbouring two sites that interact in determining fitness, finding that such epistasis has a dramatic impact on the genetic variability in regions surrounding the interacting sites.

Keywords: epistasis, genetic variability, genomic analyses, selection, sweeps

Received 5 October 2009; revision accepted 14 October 2009

Gene interactions may influence fitness in a variety of ways—from slight quantitative interactions to large-scale changes in the direction of selection (sign epistasis). Takahasi focuses on a particular form of positive epistasis, a synthetic advantageous interaction (Phillips *et al.* 2000), where two alleles A_1 and B_1 are neutral on their own but advantageous when combined. Labelling A_1 as the first allele to arise, its frequency is expected to fluctuate in a neutral fashion until the appearance of B_1 , after which point A_1 becomes selectively favoured. While B_1 is rare, however, the selection coefficient favouring A_1 remains weak, ramping up slowly in proportion to the frequency of B_1 . Similarly, the strength of selection on B_1 is initially weak (depending directly on the frequency of A_1) and rises as the two advantageous alleles spread through the population.

Takahasi (2009) assessed the impact of this gene interaction on surrounding neutral sites by combining forward- and backward-time simulations. The first step was to simulate the neutral drift and then selective spread of the A_1

and B_1 alleles forwards in time. Recording the exact number of A_1 and B_1 alleles at every point in time, the simulations then proceeded backwards in time, from the present, to determine the exact genealogical history for a sample of neutral sites surrounding the **A** and **B** loci. Coalescent events (individuals sharing a common parental allele) would have occurred rapidly during the period when both A_1 and B_1 were common and rising quickly due to selection. Unlike the classic case with a constant selection coefficient, however, any alleles that did not coalesce when A_1 and B_1 were common then followed a much slower coalescent process—almost as slow as if the region were entirely neutral.

The signals of selection left in surrounding sites depended strongly on the time, T , between the appearance of A_1 and B_1 . When T was so large that A_1 was nearly fixed by drift within the population, sites surrounding the **A** locus were hardly affected during the final little boost to fixation of A_1 aided by its beneficial interactions with B_1 , with the result that gene genealogies surrounding the **A** locus were nearly neutral. In contrast, when T was so small that A_1 and B_1 were both simultaneously rare, then both loci exhibited nearly equal signs of selection; for example, relatively low levels of heterozygosity were observed near both loci, given the number of polymorphic sites (negative Tajima's D ; Tajima 1989). In short, with epistasis, the signals of selection left on surrounding DNA sequences spanned the range from completely invisible to indistinguishable from nonepistatic selective sweeps.

The fact that selection increased in strength over time did, however, muddy the picture of selection for intermediate values of T . At sites moderately linked to the **A** locus, there was a decent chance that the genealogical tree had not completely coalesced by the time that B_1 appeared. In such cases, a good deal of polymorphism was preserved by recombination capturing different variants onto the A_1 chromosome before this time. These variants often appeared old and at high frequency among sampled sequences, relative to the number of polymorphic sites, leading to a large number of cases where Tajima's D was positive not negative—a classic sign of balancing selection. Thus, by causing selection coefficients to rise over time, synthetic advantageous gene interactions not only modulated the signal strength of past selection (making Tajima's D more or less negative), but it also reversed the sign of the signal in some samples.

It is important to emphasize that Takahasi's results are driven largely by the increasing strength of selection over time, rather than by the epistatic interactions *per se*. If the environment had changed in such a way that the strength of selection favouring A_1 increased over time in exactly the same way as in the forward-time simulations, the levels of polymorphism surrounding the **A** locus would have been

Correspondence: Sarah P. Otto, Fax: 604 822 2416;
 E-mail: otto@zoology.ubc.ca

the same, even in the absence of the B_1 allele or epistasis. Similar patterns are observed with a recessive beneficial allele, whose selective advantage increases during its spread (Teshima & Przeworski 2006; Teshima *et al.* 2006). It is also true that alleles adapting to a mix of local environments may sometimes be under heterogeneous selection, which may act much like epistasis from the point of view of an allele. Alleles change genetic backgrounds by recombination much like they may change selection regimes over space, and the changing selective context in both cases may result in very similar genetic patterns. Indeed, detecting the form of epistasis—and distinguishing it from temporally or spatially varying selection—may well be impossible once the selected alleles have fixed. Thus, the study by Takahasi does not really provide us with a tool to detect epistasis, rather it asks how tools that we are already using to detect selection might be dulled or sharpened had epistasis been present. Of course, if the selected loci remain polymorphic, then other approaches can be used to detect epistasis, including analyses of linkage disequilibrium and direct fitness assays.

The main limitation of the Takahasi analysis is that only one particular form of epistasis was considered. Nevertheless, the results and explanations provided by Takahasi allow us to speculate about the signatures of selection that would be left in a population fixed for two alleles that exhibit other forms of epistasis. For example, had A_1 been very slightly deleterious in the absence of B_1 (sign epistasis), then A_1 would likely have been segregating at a lower frequency when B_1 appeared than had A_1 been neutral on its own, as if the gap between the appearance of the two mutations, T , had been shorter (see Figures 2–6 of Takahasi). Conversely, had A_1 been very slightly beneficial on its own (so slightly beneficial that selection on this site alone would have been difficult to detect), then A_1 would likely have been segregating at a higher frequency when B_1 appeared than had A_1 been previously neutral, as if T were longer. With even stronger selection favouring A_1 before the appearance of B_1 , the signatures of selection around A_1 would become clearer and clearer, depending on the time course of selection (Fig. 1).

The next stage in this promising work should include integrating over the possible windows of time separating the appearance of epistatically interacting alleles (T), exploring a range of values of selection for or against single alleles. Because the probability of joint fixation depends strongly on the initial allele frequency at the first locus, the distribution of T conditioned on the fixation of both of the beneficial alleles is not uniform. The importance of the patterns that Takahasi has discovered depends on the frequency distribution of T for fixed pairs of alleles and on how this distribution depends on the form of selection acting on the loci, singly and jointly.

Future work is also needed to examine formally the impact of various forms of epistasis on the signature of selection. In addition, other measures of selection that focus on linkage disequilibrium among neutral sites might prove helpful in detecting that epistasis must have been present,

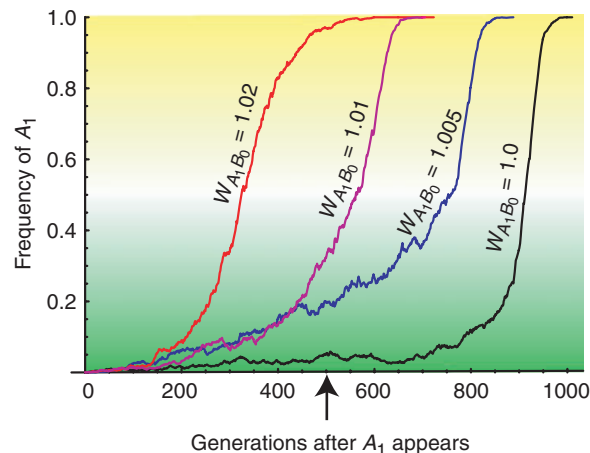


Fig. 1 The spread of allele A_1 is simulated with epistasis between A_1 and an unlinked site B_1 in a haploid population. Measuring fitness relative to the initial resident genotype (A_0B_0), we set the fitness of the favourable A_1B_1 combination to $W_{A_1B_1} = 1.1$ and varied the fitness of the two alleles when they appear by themselves, $W_{A_1B_0} = W_{A_0B_1}$, from 1.02 (red curve), 1.01 (magenta curve), 1.005 (blue curve), to 1 (synthetic advantageous interaction, black curve). Each of these cases results in positive epistasis ($W_{A_1B_1} - W_{A_1B_0}W_{A_0B_1} > 0$), with the strength of epistasis increasing from left to right. The more weakly that A_1 was favoured in the absence of B_1 , the longer it spends at low frequency (green-shaded region). Prolonging the period of time in the green region increases the chance that A_1 recombines onto other genetic backgrounds, which reduces the signals of selection near the **A** locus, and increases the chance that additional A_1 alleles will appear by mutation (ignored here and in the simulations of Takahasi), causing the signals of selection to appear consistent with a soft selective sweep (Pennings & Hermisson 2006). In contrast, when the alleles are favourable on their own, they spread more rapidly, spending relatively less time at low frequency and more time at high frequency (yellow shaded region), hindering the rescue of genetic variation onto the A_1 background and preserving the signals of selection. As in Takahasi's simulations, we illustrate only cases where A_1 remained polymorphic when B_1 appeared (at generation 500, arrow) and where both alleles fixed. The population size was 10 000.

not just temporally varying selection (such as Kelly's 1997 *ZnS*-test or tests based on the number of haplotypes, K , as suggested by Depaulis & Veuille 1998; which have proven more powerful when detecting soft selective sweeps, Pennings & Hermisson 2006). Still, Takahasi has proven that the signals of selection do depend on epistasis. In particular, we must always consider the caveat that a site that appears neutral or weakly selected might in fact have been strongly selected, but only in some genetic contexts.

References

- Depaulis F, Veuille M (1998) Neutrality tests based on the distribution of haplotypes under an infinite-sites model. *Molecular Biology and Evolution*, **15**, 1788–1790.

- Kelly JK (1997) A test of neutrality based on interlocus associations. *Genetics*, **146**, 1179–1206.
- Pennings PS, Hermisson J (2006) Soft sweeps III: the signature of positive selection from recurrent mutation. *PLoS Genetics*, **2**, 1998–2012.
- Phillips PC (2008) Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews Genetics*, **9**, 855–867.
- Phillips PC, Otto SP, Whitlock MC (2000) Beyond the average, the evolutionary importance of gene interactions and variability of epistatic effects. In: *Epistasis and the Evolutionary Process* (eds Wolf JB, Brodie ED III, Wade MJ), pp. 20–38. University Press, Oxford/New York.
- Segrè D, DeLuna A, Church GM, Kishony R (2005) Modular epistasis in yeast metabolism. *Nature Genetics*, **37**, 77–83.
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.
- Takahasi KR (2009) Coalescent under the evolution of coadaptation. *Molecular Ecology*, **18**, 5018–5029.
- Teshima KM, Przeworski M (2006) Directional positive selection on an allele of arbitrary dominance. *Genetics*, **172**, 713–718.
- Teshima KM, Coop G, Przeworski M (2006) How reliable are empirical genomic scans for selective sweeps? *Genome Research*, **16**, 702–712.
- Whitlock MC, Phillips PC, Moore FBG, Tonsor S (1995) Multiple fitness peaks and epistasis. *Annual Review of Ecology and Systematics*, **26**, 601–629.

doi: 10.1111/j.1365-294X.2009.04425.x