Genet. Res., Camb. (2007), **89**, *pp.* 389–390. © 2008 Cambridge University Press
doi:10.1017/S0016672308009580 Printed in the United Kingdom

A hitch-hiking guide to the genome: a commentary on 'The hitch-hiking effect of a favourable gene' by John Maynard Smith and John Haigh

BRIAN CHARLESWORTH*
Institute of Evolutionary Biology, University of Edinburgh, Ashworth Laboratories, King's Building, Edinburgh EH9 3JT, UK

This paper is one of the many seminal contributions to theoretical population genetics by John Maynard Smith, who teamed up with his probability theorist colleague John Haigh after moving to the University of Sussex. As is so often the case, its significance was not recognized at the time, largely because the kind of data needed to apply the results only became available many years later. Although the term hitch-hiking (often spelt hitchhiking) is usually attributed to this paper, it was in fact introduced earlier by Kojima & Schaffer (1967), but their model dealt with the simultaneous increase in frequency of two linked, selected mutations and is now largely forgotten. The catchier term 'selective sweep' is often used as a synonym (Berry *et al*., 1991).

The motivation for the paper was the observation that surveys of allozyme variation show only a weak relation between the mean amount of variability per locus and the number of individuals in a species (Lewontin, 1974). This led Maynard Smith to realize that the fixation in a population of a favourable mutation would cause a loss of variability at closely linked sites, by dragging along with it any neutral variants that were present in the gamete in which the mutation arose. Recombination between the site under selection and the neutral sites reduces this effect, by allowing a flow of variants from the rest of the population into the gametes carrying the favourable mutation while it spreads through the population. Since large populations allow a higher net rate of input of favourable mutations than small populations, we might expect more selective sweeps per genome per generation in large populations, which would counteract the tendency for such populations to harbour more neutral variation (Kimura & Crow, 1964), an idea later taken up by Gillespie (2000).

The idea is easy to grasp, and is equivalent to the concept of periodic selection, introduced by bacterial geneticists, who observed sudden changes in the frequencies of neutral markers in chemostat populations, caused by the spread of favourable mutations (Atwood *et al*., 1951). In this case, there is no recombination between the marker and the targets of selection, and the process is easy to model. Hitch-hiking is probably an important determinant of levels of variability and adaptation in bacterial genomes with low levels of recombination (Berg, 1996).

The challenge is to calculate the magnitude of the hitch-hiking effect on a neutral site in the presence of recombination, as a function of the selection coefficient for the favourable variant, s, and the recombination frequency, r, between the selected site and a neutral site. Maynard Smith and Haigh assumed that two neutral variants, A and a, were segregating in the initial population, with frequencies p and q, respectively. They obtained exact and approximate equations for the final frequency of A, after a favourable mutation arose as a unique event in a gamete carrying A and spread to fixation. The variability in the population can be measured by the diversity statistic $H=2pq$. The expected value of H after a sweep can be determined, as a function of the recombinational distance from the locus under selection.

The results of Maynard Smith and Haigh (1974) have formed the basis for all subsequent work on this problem (Barton, 2000). Their basic conclusion was that the effect of a sweep on variability is dependent on the ratio r/s; this must be substantially less than 1 for there to be much of an effect. Unless selection is strong, there will only be an effect on sites that are very close to the target of selection, or in genome regions where recombination is rare or absent. Maynard Smith and Haigh also made a rather crude attempt to estimate the effects on neutral variability of recurrent selective sweeps at loci scattered around the genome. Subsequent theoretical work has developed predictions for the effects on nucleotide site diversities of recurrent selective sweeps at loci scattered around the genome (Kaplan *et al*., 1989; Stephan, 1995), and for the effects of sweeps on the frequency distributions of single nucleotide polymorphisms (SNPs) under the standard infinite sites model (Braverman *et al*., 1995).

* e-mail: brian.charlesworth@ed.ac.uk

In addition, it is possible to model scenarios in which the favourable mutation does not go to fixation, but remains segregating as a balanced polymorphism: a 'partial sweep' (Hudson *et al*., 1994; Sabeti *et al*., 2002).

This theoretical work laid the foundations for the use of data on molecular polymorphisms as a tool for detecting the signature of selection, from its effects on patterns of variability at neutral sites that happen to be in a part of the genome where a sweep has occurred. An early example of the detection of the signature of a (partial) sweep at the DNA level was provided by the finding of associations between restriction fragment polymorphisms and the haemoglobin S allele of humans, famous for its role in protecting heterozygous carriers against malaria (Kan & Dozy, 1978; Kwiatkowski, 2005). Much interest was later generated by the finding that low-recombination regions of the *Drosophila* genome were associated with low levels of DNA sequence diversity. This was originally interpreted in terms of the effects of selective sweeps (Begun & Aquadro, 1992), but the possibility that selection against deleterious mutations may be a major cause of this pattern (Charlesworth *et al*., 1993) has not yet been excluded. With the advent of genome-wide scans of variability, especially in human populations, elaborate statistical procedures have been developed, with the aim of detecting the effects of partial and complete selective sweeps, and determining which sites are the targets of selection (Sabeti *et al*., 2002; Nielsen, 2005). Recent evidence from *Drosophila* suggests that sweeps in coding sequence may be sufficiently frequent across the genome that the observed level of nucleotide variability is significantly lower than would be expected in their absence (Andolfatto, 2007); however, it seems unlikely that this could account for the pattern of allozyme variation that motivated the original study of hitch-hiking. Instead, it is probable that many allozyme variants are not neutral but are maintained by balancing selection (Eanes, 1999).

References

Andolfatto, P. (2007). Hitchhiking effects of recurrent beneficial amino acid substitutions in the *Drosophila melanogaster* genome. *Genome Research* **18**, 1755–1762.

Atwood, K. C., Schneider, L. K. & Ryan, F. J. (1951). Periodic selection in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the USA* **37**, 146–155.

Barton, N. H. (2000). Genetic hitchhiking. *Philosophical Transactions of the Royal Society, Series B* **355**, 1553–1562.

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

Berg, O. G. (1996). Selection intensity for codon bias and the effective population size of *Escherichia coli*. *Genetics* **142**, 1379–1382.

Berry, A. J., Ajioka, J. W. & Kreitman, M. (1991). Lack of polymorphism on the *Drosophila* fourth chromosome resulting from selection. *Genetics* **129**, 1111–1117.

Braverman, J. M., Hudson, R. R., Kaplan, N. L., Langley, C. H. & Stephan, W. (1995). The hitchiking effect on the site frequency spectrum of DNA polymorphism. *Genetics* **140**, 783–796.

Charlesworth, B., Morgan, M. T. & Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289–1303.

Eanes, W. F. (1999). Analysis of selection on enzyme polymorphisms. *Annual Review of Ecology and Systematics* **30**, 301–326.

Gillespie, J. H. (2000). Genetic drift in an infinite population: the pseudohitchiking model. *Genetics* **155**, 909–919.

Hudson, R. R., Bailey, K., Skarecky, D., Kwiatowski, J. & Ayala, F. J. (1994). Evidence for positive selection in the superoxide dismutase (SOD) region of *Drosophila melanogaster*. *Genetics* **136**, 1329–1340.

Kan, Y. W. & Dozy, A. M. (1978). Polymorphism of DNA sequence adjacent to human beta-globin structural gene: relation to sickle mutation. *Proceedings of the National Academy of Sciences of the USA* **75**, 5631–5635.

Kaplan, N. L., Hudson, R. R. & Langley, C. H. (1989). The 'hitch-hiking' effect revisited. *Genetics* **123**, 887–899.

Kimura, M. & Crow, J. F. (1964). The number of alleles that can be maintained in a finite population. *Genetics* **49**, 725–738.

Kojima, K.-I. & Schaffer, H. E. (1967). Survival process of linked mutant genes. *Evolution* **21**, 518–531.

Kwiatkowski, D. P. (2005). How malaria has affected the human genome and what human genetics can teach us about malaria.. *American Journal of Human Genetics* **77**, 171–192.

Lewontin, R. C. (1974). *The Genetic Basis of Evolutionary Change*. New York: Columbia University Press.

Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Review of Genetics* **39**, 197–218.

Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z. P., Richter, D. J., *et al*. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**, 832–837.

Maynard Smith, J. & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical Research* **23**, 23–35.

Stephan, W. (1995). An improved method for estimating the rate of fixation of favorable mutations based on DNA polymorphism data. *Molecular Biology and Evolution* **12**, 959–962.