

# Adaptive hitchhiking effects on genome variability

Peter Andolfatto

The continuing deluge of nucleotide polymorphism data is providing insights into the role of adaptation in shaping genome-wide patterns of variability and molecular evolution. Population genetic models in which linkage and selection interact (i.e. hitchhiking) predict that selection can leave 'footprints' in closely linked genomic regions. New analytical approaches show promise for distinguishing the signature of adaptation from that of several non-adaptive alternatives. Accounting for the effects of population structure and history poses a challenge for future investigations.

## Addresses

Institute of Cell, Animal and Population Biology, University of Edinburgh, Ashworth Labs, King's Buildings, Edinburgh, EH9 3JT, UK; e-mail: Peter.Andolfatto@ed.ac.uk

Current Opinion in Genetics & Development 2001, 11:635–641

0959-437X/01/\$ – see front matter

© 2001 Elsevier Science Ltd. All rights reserved.

## Introduction

Most nucleotide variability observed in natural populations may be neutral, or nearly so [1]. Even so, natural selection can play an important role in shaping this variability through the effects of genetic linkage. In particular, neutral variants linked to a strongly favoured mutation can 'hitchhike' to fixation in the population, while other variants are lost [2]. More generally, because selection affects closely linked neutral variability, one can use the latter to draw inferences about the frequency and mode of adaptive evolution at genes of interest.

What is the evidence that selection is having an impact on genome-wide nucleotide variability? The genomes of sexual organisms often exhibit considerable heterogeneity in rates of meiotic crossing over. In *Drosophila melanogaster*, which is the best-studied species at present, the local crossing over rate is a strong predictor of levels of nucleotide variability (reviewed in [3]); regions of reduced crossing over harbour reduced variability. A similar trend is observed in other species of *Drosophila*, as well as a variety of other organisms, including humans (reviewed in [4–6]).

The positive correlation between nucleotide variability and the local rate of recombination may be explained by repeated episodes of hitchhiking (or selective sweeps) caused by the fixation of newly arising advantageous mutations [7,8]. The average size of a hitchhiked region depends on the ratio of  $s$  to  $r$ , where  $s$  is the selective advantage of the favourable mutation and  $r$  is the local rate of recombination. Thus, if selection intensities are similar across the genome, loci in regions of low recombination will be affected by more selective sweeps per unit time, and thus are more likely to be sampled shortly after a hitchhiking event [7].

An alternative (non-adaptive) explanation for the correlation, however, is purifying selection against strongly deleterious mutations across the genome, hereafter referred to as 'background selection' [9]. In this model, a neutral variant will persist in the population only if it arises on a deleterious mutation-free chromosome (or segment of chromosome) or recombines onto one [9–12]. If selection coefficients and deleterious mutation rates are the same in different genome regions, the rate of recombination will determine the effect of background selection on linked variability.

Thus, the correlation between recombination rate and variability — though compelling evidence for an interaction between selection and linkage — is not a unique prediction of adaptive hitchhiking. Although processes like adaptive hitchhiking and the 'negative hitchhiking' effect caused by background selection undoubtedly occur, there is considerable uncertainty about the rates of occurrence of both advantageous and deleterious mutations and the distribution of selection coefficients associated with these mutations. As a result, a debate has developed concerning the relative importance of these modes of selection in shaping patterns of genome variability. In this review, I focus on the more recent evidence for adaptive hitchhiking.

## Distinguishing between background selection and hitchhiking using the frequency spectrum of polymorphisms

As both adaptive hitchhiking and background selection (as well as other forms of hitchhiking) predict reduced variability in regions of low recombination [13,14], recent studies have focussed on other features of the data to try to distinguish between alternative models. One example is the distribution of nucleotide polymorphism frequencies in a population sample, hereafter referred to as the 'frequency spectrum'. An excess of low-frequency polymorphisms (a negative skew in the frequency spectrum) is expected at neutral sites closely linked to recent fixations of advantageous mutations [15]. This effect should be more pronounced in regions of lower recombination where hitchhiking effects are stronger [15,16]. In contrast, a marked skew in the frequency spectrum is not generally expected under the background selection model [10,11]. Langley *et al.* [17\*] have recently exploited this distinction to interpret nucleotide polymorphism data collected at two genes located in a region of reduced crossing over in *D. melanogaster*. They found marked skews toward low-frequency polymorphisms in most population samples, particularly in older (European) and ancestral (African) populations, consistent with the adaptive hitchhiking model. Similarly, M Przeworski and I [18\*] have investigated the relationship between the frequency spectrum and local rates of crossing-over for 29 loci in African (predominantly Zimbabwe) populations of *D. melanogaster*. We observed a

trend toward lower polymorphism frequencies in regions of reduced crossing-over, a pattern consistent with adaptive hitchhiking but unlikely under background selection.

Although these data are consistent with adaptive hitchhiking, it is difficult to completely rule out more complex deleterious mutation models. In particular, models of weakly deleterious mutations may also produce a skew in the frequency spectrum [10,19<sup>\*</sup>]. Such weakly deleterious mutations may be caused by a fraction of amino-acid replacement mutations, small insertion–deletion mutations and transposable elements [13,20]. Weakly selected deleterious mutations will contribute little to the reduction of linked neutral variants unless the recombination rate is very low relative to the mutation rate; however, a bimodal distribution of deleterious mutation effects, with a large class of weakly deleterious mutations (e.g. [13]) may help to account for the frequency spectrum data from *D. melanogaster* [18<sup>\*</sup>].

Fay and Wu [21<sup>\*</sup>] have recently proposed a new statistical test of neutrality that focuses on a predicted effect of adaptive hitchhiking that is distinct from expectations under weakly deleterious background selection models: an excess of high-frequency derived mutations. The latter can be identified by comparisons with closely related species in order to infer the ancestral state of each polymorphic mutation. As this signature of selective sweeps is very short-lived ([22<sup>\*</sup>]; M Przeworski, personal communication), this test is probably best applied to loci where selection is *a priori* thought to have occurred recently. Two recent examples from *D. melanogaster* may include *desat2* [23], a gene underlying population-specific differences in cuticular hydrocarbon pheromones, and *Acp26Aa* [21], a protein found in seminal fluid.

Although weakly deleterious mutation and adaptive hitchhiking models make partially overlapping predictions in regions of reduced recombination, of the two, only selective sweeps can account for departures from a ‘neutral panmictic population’ model in regions of high recombination. Yet, there are an increasing number of reports (in various organisms) of extreme diversity reductions and negative skews in the frequency spectrum at specific loci in regions of high recombination [24,25<sup>\*</sup>,26,27]. There has also been a recent burst of studies showing unusually strong linkage disequilibrium patterns (or ‘haplotype structure’) at loci in high-recombination regions [28–40]. The latter may be consistent with selective sweep models if selection was strong and recent ([30]; M Przeworski, personal communication).

### Comparing levels of variability on different chromosomes

Another approach to distinguishing between hitchhiking and background selection has been to compare patterns of variability on different chromosomes. For example, in a recent study of the Y chromosome of *D. melanogaster* and *D. simulans*, Zurovcova and Eanes [41] demonstrated that nucleotide variability is markedly reduced. As genes, and

thus targets for deleterious mutations, on this chromosome were thought to be few, the authors argue that background selection is an unlikely explanation. Carvalho *et al.* [42], however, have since demonstrated that the Y chromosome in *D. melanogaster* harbours more genes than was previously thought. Thus, the ability of deleterious mutation models to account for the Y chromosome is still an open question. With the recent completion of several genome projects, we should soon see the integration of detailed information on gene density into linkage-selection models.

Recessive mutations are more visible to selection on the sex chromosomes relative to autosomes. Aquadro *et al.* [3] have suggested that if mutations are recessive, and all other factors are equal, adaptive hitchhiking models predict lower diversity on the X chromosome relative to the autosomes, whereas the background selection model predicts the opposite pattern [13]. Although this prediction has yet to be quantified, Begun and Whitley [43<sup>\*</sup>] have recently demonstrated that X-linked nucleotide variability is significantly lower than autosomal variability in *D. simulans*, even after correcting for the sex-specific inheritance of these chromosomes. Background selection is an unlikely explanation for this observation [13]. However, this pattern may not be a general prediction of adaptive hitchhiking models either. Orr and Betancourt [44<sup>\*</sup>] have recently investigated models in which advantageous mutations were previously deleterious, and thus held at low frequencies by a balance between mutation and purifying selection. Under this model, the probability of fixation of advantageous mutations is largely independent of their degree of recessivity. This implies that the prediction of lower diversity on the X as a result of hitchhiking is specific to the case in which most advantageous mutations are newly arising, and may not apply to models under which selection pressure on mutations changes over time [14,44<sup>\*</sup>,45].

### The problem of population structure and history

In addition to natural selection, patterns of variability within species can also be shaped by demographic history, that is, the structure and size of populations over time. Humans [6] and *D. melanogaster* [46] are thought to have African origins and, perhaps like most species, are likely to have complicated population histories and structures. In both species, multi-population surveys of nucleotide variability are beginning to reveal differences in patterns of variability between African (ancestral) and other (derived) populations [6,17<sup>\*</sup>,18<sup>\*</sup>,29,47–51]. Uncertainty about the impact of demographic history on genome-wide patterns of variability poses a difficult problem when interpreting data in the context of models of selection (as above).

For example, like selection [3], departures from assumptions like constant population size, random mating and equal male and female reproductive success can affect patterns of variability in a chromosome-specific manner (e.g. [52,53]). In this light, it is noteworthy that the reduced variability observed at X-linked relative to autosomal genes

in predominantly non-African (derived) populations of *D. simulans* [43<sup>\*</sup>] is less apparent in African (ancestral) populations [48]. This may not be unexpected if most hitchhiking has been confined to non-African populations (i.e. local adaptation). X-linked loci in the non-African *D. simulans* data [43<sup>\*</sup>], however, do not show a greater skew towards low-frequency polymorphisms — as expected under increased hitchhiking — than do autosomal loci. In fact, the opposite pattern is observed [54]. These features of the data suggest demographic departures from model assumptions or a more complicated model of hitchhiking (or both, see below).

Demographic perturbations can also increase statistical noise from locus to locus relative to expectations under panmictic, constant population size null models. In the absence of demographic information, this will limit the interpretation of various proposed statistical tests of neutrality at specific loci (e.g. see [21<sup>\*</sup>,30,31,55] and others reviewed in [56]). In principle, demography and selection can be distinguished by considering data from multiple loci; demography will have a similar effect on the whole genome, whereas selection will have locus-specific effects. Thus, genome-wide patterns favour demographic explanations over hitchhiking. By this line of reasoning, the evidence for hitchhiking based on excess haplotype structure in *D. melanogaster* and *D. simulans* is made less convincing by the generality of this pattern observed in the two species ([57<sup>\*</sup>,58]; J Wall, P Andolfatto, M Przeworski, unpublished data). *D. simulans*, in particular, exhibits a marked paucity of haplotypes at many independent loci [28,32,34,35,37,39,43<sup>\*</sup>,59,60].

Considerably more information can be mined from multi-locus data where samples have been drawn identically from a population for each locus. Sophisticated statistical approaches have recently been employed to distinguish between population bottlenecks and hitchhiking in such samples [61<sup>\*</sup>; Wall, this issue [pp 647–651]). Galtier *et al.* [61<sup>\*</sup>] employ a ‘full likelihood’ statistical approach to rule out a population bottleneck as the cause for differences in variability of patterns observed at three loci from a single population of *D. melanogaster*. However, in some cases, demographic effects on variability patterns can also be identified. We have recently used a ‘summary likelihood’ method (J Wall, P Andolfatto, M Przeworski, unpublished data; see also Wall, this issue, pp 647–651) to show that a recent bottleneck is a more likely explanation than recurrent hitchhiking for chromosome-specific variability patterns in a *D. simulans* population [43<sup>\*</sup>]. These approaches appear to be promising ways to untangle the effects of background selection, adaptive hitchhiking and demography.

Selection models almost always assume that populations are panmictic. In addition to generating patterns that mimic selection in neutral populations, demography can also alter expectations in the presence of selection. For instance, geographically localised selection (i.e. local adaptation) can

increase differentiation between populations at closely linked neutral sites [62,63]. Indeed, much of the evidence for adaptive hitchhiking in human populations is observed at genes that underlie locally adapted traits such as resistance to malaria [40,64<sup>\*</sup>], lactose tolerance [65] and skin melanisation [66]. However, adaptation need not be local to increase population differentiation at linked neutral sites. Two recent studies have demonstrated that globally advantageous mutations can also increase population differentiation at linked neutral sites when migration is limited between sub-populations of a species [67,68].

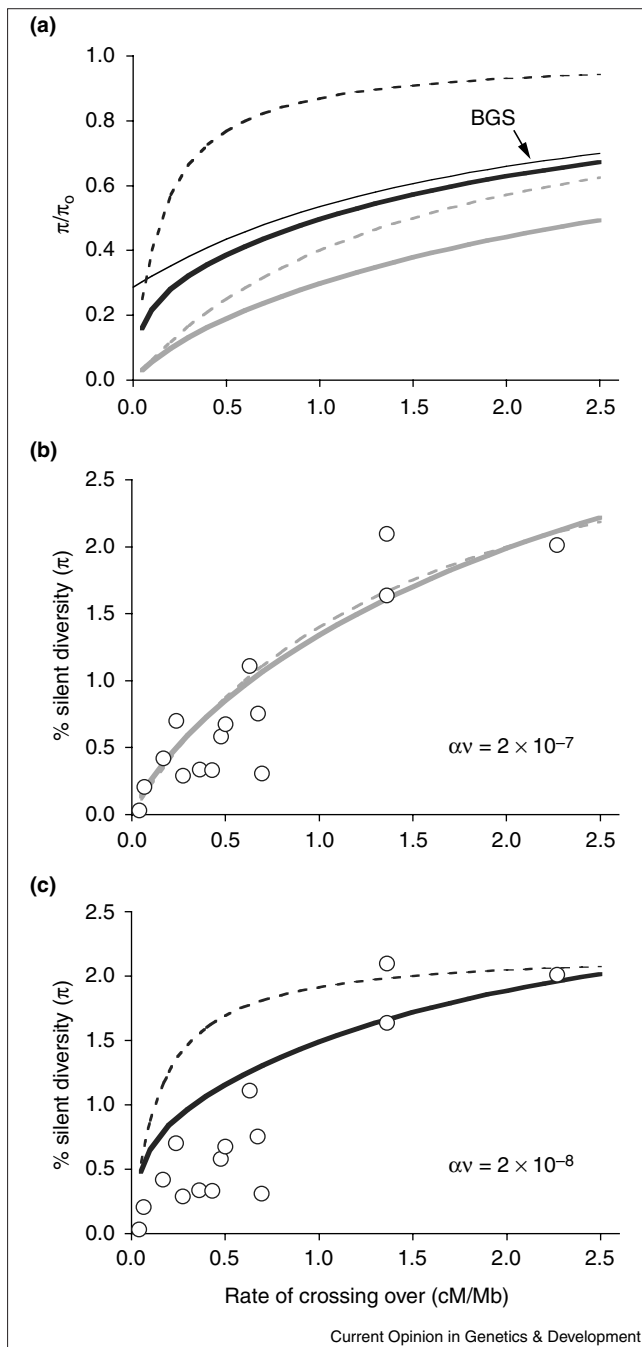
For *D. melanogaster* and *D. simulans*, both historically tropical species, it is possible that the recent colonisation of temperate habitats was accompanied by increased adaptive hitchhiking. In this case, several key assumptions of simple hitchhiking models — on which many of our intuitions are based — may be violated. First, selection would be episodic rather than constant over time. Second, selection on many traits may be operating on variants already segregating in the population (i.e. standing variation) rather than on newly arising variants. Third, selection would be geographically localised rather than species-wide. Departures from simple hitchhiking model predictions may account for the unexpectedly positive skew in the frequency spectrum on the X chromosome of *D. simulans* [43<sup>\*</sup>,54]. They may also explain why the skew toward low-frequency polymorphisms in regions of reduced crossing over [18<sup>\*</sup>] was not apparent in similar data collected from a North American population of *D. melanogaster* [3]. Multi-population variability surveys and extensions of simple hitchhiking models that relax some common assumptions [14,44<sup>\*</sup>,45] may provide new insights.

### Rates of advantageous mutations and interference among positively and negatively selected mutations

Although their relative importance is uncertain, processes like background selection and hitchhiking most likely operate in concert. Selection at one locus will limit the efficiency of selection at linked loci [69]. If deleterious mutations have the potential to impede adaptation, it makes sense to model background selection and hitchhiking together. Kim and Stephan [22<sup>\*</sup>] present a model of the joint effects of these forces on linked neutral variation. If background selection is caused by strongly deleterious selection, it can be approximated as a local reduction in population size [9,11,12,22<sup>\*</sup>]. An approximate equation is provided [22<sup>\*</sup>] that predicts the effects of hitchhiking in the presence of background selection. Figure 1 shows a fit of this model (equation 6 of [22<sup>\*</sup>]) to X chromosome data from *D. melanogaster* [18<sup>\*</sup>].

First, consider the fit of the recurrent hitchhiking model in the absence of background selection. A reasonably good fit to the data is obtained for  $\alpha w$  on the order of  $10^{-7}$  (Figure 1b), where  $\alpha$  is a measure of the intensity of selection —  $2Ns$ , where  $N$  is the population size and  $s$  is the

Figure 1



A recurrent hitchhiking model with background selection. The background selection parameters were arbitrarily chosen to give moderate reductions in variability, as a more extreme model need not invoke hitchhiking to account for the data [13]. **(a)** The reduction in variability ( $\pi$ ) relative to the neutral model expectation ( $\pi_0$ ) for background selection (BGS, thin black line), and recurrent hitchhiking using equation 6 of [22] with  $\alpha v$  set to  $2 \times 10^{-8}$  (dark grey lines) and  $2 \times 10^{-7}$  (light grey lines). The reduction due to hitchhiking alone is given in thick dashed lines; hitchhiking with background selection with thick solid lines. **(b)** Fit of *D. melanogaster* X chromosome data of [18] as (open circles) to the above hitchhiking model (as above,  $\alpha v = 2 \times 10^{-7}$ ) in the absence (dashed line,  $\pi_0 = 3.5\%$ ) and in the presence (solid line,  $\pi_0 = 4.5\%$ ) of background selection. **(c)** Fit of the hitchhiking model (as above,  $\alpha v = 2 \times 10^{-8}$ ) to the same data in the absence (dashed line,  $\pi_0 = 2.2\%$ ) and in the presence (solid line,  $\pi_0 = 3.0\%$ ) of background selection.

selection coefficient—and  $v$  is the rate of sweeps per generation. A similar value was obtained by Stephan [70] using data from Aquadro *et al.* [3]. Interestingly, with this  $\alpha v$ , moderate levels of background selection make little difference to the fit of the hitchhiking model to variability levels (Figures 1a and 1b), but simply imply a larger  $\pi_0$  (expected variability in the absence of hitchhiking effects). With an  $\alpha v$  on the order of  $10^{-8}$  or smaller (see Figure 1a,c), the combined background selection and hitchhiking models appear to fit the data better than either process modelled alone.

Can advantageous mutations alone account for the correlation between diversity and recombination in *D. melanogaster*? Assuming an  $\alpha v = 10^{-7}$  (see Figure 1) a large population size ( $2N = 10^7$ ) and strong selection ( $s = 0.01$ ), the advantageous mutation rate ( $\sim v/2s$ ) will be on the order of  $10^{-10}$ ; only two orders of magnitude smaller than several estimates of the neutral mutation rate (reviewed in [57]). A smaller average  $s$  would require a higher advantageous mutation rate to achieve the same mean effect on diversity. Thus, it appears that advantageous mutations would have to be both extremely common *and* strongly selected to account for the correlation between diversity and recombination alone. In humans, this requirement is even more extreme as the (historical) population size is several orders of magnitude smaller; the advantageous mutation rate would have to be on the order of the neutral mutation rate (cf. [71]) or higher. As only a small fraction of newly arising mutations are expected to be strongly advantageous in a given environment, this seems highly unlikely. As a result, adaptive hitchhiking may only be a minor factor contributing to the correlation between diversity and recombination in humans. In *Drosophila*, hitchhiking probably shares a role with background selection in accounting for this pattern.

In addition to 'variability levels', other aspects of the data, such as the frequency spectrum, can inform us about the relative importance of adaptive hitchhiking. The frequency spectrum can be summarised with the statistic Tajima's  $D$  [55]. Under the standard neutral model, the expectation of  $D$  is close to zero;  $D$  will be negative when there is an excess of low frequency polymorphisms, as expected under adaptive hitchhiking models [15,16]. As an illustration, let us assume that variability is reduced 10-fold below the neutral expectation ( $\pi_0$ ) in regions of high crossing over (silent variability for  $r > 1$  cM/Mb is  $\sim 2\%$  per site in *D. melanogaster*, Figure 1). In recurrent hitchhiking simulations (program courtesy of M Przeworski), the expected  $D$  [55] under this model is about  $-1$  in a sample of 500 neutral sites from 10 chromosomes, indicating a sharp skew towards low-frequency polymorphisms. For a 2.5-fold reduction in variability (as expected for  $\alpha v = 2 \times 10^{-7}$ , Figure 1a and b), the average  $D$  is  $\sim -0.7$ . In comparable *D. melanogaster* data [18], the average Tajima's  $D$  in regions of high crossing over on the X chromosome is  $-0.07$ , whereas the expectation under strict neutrality (i.e. setting  $\alpha v = 0$ ) is  $-0.08$ . A tentative conclusion is that

recurrent hitchhiking, and other modes of selection that produce a marked negative skew in the frequency spectrum, may have a minor impact on average variability over most of the highly recombining portion of the *D. melanogaster* genome. Still, adaptive hitchhiking may be frequent enough to account for the skew toward rare polymorphisms observed in regions of reduced crossing over [18\*].

## Conclusions

Our understanding of the interaction of selection and linkage and its effects on patterns of genome-wide variability is improving, but uncertainty about the demographic history of a species poses two problems. On one hand, the effects of demography on patterns on nucleotide variability can be conflated with those of hitchhiking. On the other, demographic structure will modify our expectations under selection relative to panmictic population models.

If we are to continue making progress in understanding the impact of hitchhiking on genome variability, it will be crucial to understand more about the demographic structure and history of the populations being studied. Multi-population variability surveys are beginning to reveal the extent to which geographically distant populations differ in various aspects of nucleotide variability patterns. It will be important to focus part of these surveys on ancestral populations as they comprise the source for all other populations. Recent multi-locus statistical approaches show promise for untangling the effects of demography and selection.

Under the neutral model, nucleotide diversity at neutral sites is directly proportional to  $N$  [1]. It has been suggested that a steady stream of advantageous mutations, through the effects of linkage, can uncouple the relationship between population size and levels of nucleotide diversity, even in regions of 'high' recombination [2,16]. Here, I have argued that hitchhiking models that predict a marked skew in the frequency spectrum are probably having a minor impact on variability levels in high crossing-over regions of *D. melanogaster*. This, in turn, limits the ability of such models to account for diversity reductions in regions of reduced recombination in this species (see Figure 1). Adaptive hitchhiking in humans may not be an important factor producing the observed correlation between variability and crossing-over rate. Yet the signature of adaptive hitchhiking is apparent in the vicinity of a number locally adapted traits in humans. Future investigations should consider not just genome-wide variability levels but also the frequency spectrum, patterns of linkage disequilibrium (see Wall, this issue [pp 647–651]) and the geographic distribution of these features of the data.

## Acknowledgements

Thanks to B Charlesworth, I Gordo, M Przeworski, J Wall and S Wright for helpful discussions and comments on the manuscript. Special thanks to M Przeworski for the hitchhiking simulation program and sharing unpublished results. This work was supported by a European Molecular Biology Organisation postdoctoral fellowship.

## References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
  - of outstanding interest
1. Kimura M: *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press; 1983.
  2. Maynard Smith J, Haigh J: **The hitch-hiking effect of a favourable gene**. *Genet Res* 1974, **23**:23-35.
  3. Aquadro CF, Begun DJ, Kindahl EC: **Selection, recombination, and DNA polymorphism in *Drosophila***. In *Non-Neutral Evolution: Theories and Molecular Data*. Edited by Golding B. Chapman and Hall; 1994:46-56.
  4. Aquadro CF: **Insights into the evolutionary process from patterns of DNA sequence variability**. *Curr Opin Genet Dev* 1997, **7**:835-840.
  5. Charlesworth D, Charlesworth B: **Sequence variation: looking for effects of genetic linkage**. *Curr Biol* 1998, **8**:R658-R661.
  6. Przeworski M, Hudson RR, Di Rienzo A: **Adjusting the focus on human variation**. *Trends Genet* 2000, **16**:296-302.
  7. Kaplan NL, Hudson RR, Langley CH: **The 'hitchhiking effect' revisited**. *Genetics* 1989, **123**:887-899.
  8. Wiehe TH, Stephan W: **Analysis of a genetic hitchhiking model, and its application to DNA polymorphism data from *Drosophila melanogaster***. *Mol Biol Evol* 1993, **10**:842-854.
  9. Charlesworth B, Morgan MT, Charlesworth D: **The effect of deleterious mutations on neutral molecular variation**. *Genetics* 1993, **134**:1289-1303.
  10. Charlesworth D, Charlesworth B, Morgan MT: **The pattern of neutral molecular variation under the background selection model**. *Genetics* 1995, **141**:1619-1632.
  11. Hudson RR, Kaplan NL: **Gene trees with background selection**. In *Non-neutral Evolution: Theories and Molecular Data*. Edited by Golding B. Chapman and Hall; 1994:140-153.
  12. Hudson RR, Kaplan NL: **Deleterious background selection with recombination**. *Genetics* 1995, **141**:1605-1617.
  13. Charlesworth B: **Background selection and patterns of genetic diversity in *Drosophila melanogaster***. *Genet Res* 1996, **68**:131-149.
  14. Gillespie JH: **Junk ain't what junk does: neutral alleles in a selected context**. *Gene* 1997, **205**:291-299.
  15. Braverman JM, Hudson RR, Kaplan NL, Langley CH, Stephan W: **The hitchhiking effect on the site frequency spectrum of DNA polymorphisms**. *Genetics* 1995, **140**:783-796.
  16. Gillespie JH: **Genetic drift in an infinite population. The pseudohitchhiking model**. *Genetics* 2000, **155**:909-919.
  17. Langley CH, Lazzaro BP, Phillips W, Heikinen E, Braverman J: **Linkage disequilibria and the site frequency spectra in the *su(s)* and *su(w<sup>a</sup>)* regions of the *Drosophila melanogaster* X chromosome**. *Genetics* 2000, **156**:1837-1852.  
A multi-population survey of nucleotide variation at two genes in a region of reduced crossing over. Provides evidence that an alternative mechanism of recombination (i.e. gene conversion) may not be reduced in regions of reduced crossing-over.
  18. Andolfatto P, Przeworski M: **Regions of lower crossing over harbor more rare variants in African *Drosophila melanogaster***. *Genetics* 2001, **158**:657-665.  
In African population samples of *D. melanogaster*, a summary of the frequency spectrum is shown to be positively correlated with the local rate of crossing-over, as expected under adaptive hitchhiking models.
  19. Tachida H: **Molecular evolution in a multisite nearly neutral mutation model**. *J Mol Evol* 2000, **50**:69-81.  
A model of weakly deleterious selection at many linked sites. Weakly deleterious selection can produce a negative skew in the frequency spectrum of linked neutral polymorphisms.
  20. Akashi H: **Within- and between-species DNA sequence variation and the 'footprint' of natural selection**. *Gene* 1999, **238**:39-51.

21. Fay JC, Wu CI: **Hitchhiking under positive Darwinian selection.** *Genetics* 2000, **155**:1405-1413.  
Immediately after a selective sweep, an excess of high frequency *derived* polymorphisms is expected. This prediction is used to formulate a new statistical test of neutrality.
22. Kim Y, Stephan W: **Joint effects of genetic hitchhiking and background selection on neutral variation.** *Genetics* 2000, **155**:1415-1427.  
The first study of the joint effects of adaptive hitchhiking and background selection on linked neutral variants. An approximate formula is provided to describe the joint effect of these forces on variability levels.
23. Takahashi A, Tsaur SC, Coyne JA, Wu CI: **The nucleotide changes governing cuticular hydrocarbon variation and their evolution in *Drosophila melanogaster*.** *Proc Natl Acad Sci USA* 2001, **98**:3920-3925.
24. Depaulis F, Brazier L, Mousset S, Turbe A, Veuille M: **Selective sweep near the *In(2L)t* inversion breakpoint in an African population of *Drosophila melanogaster*.** *Genet Res* 2000, **76**:149-158.
25. Yi S, Charlesworth B: **A selective sweep associated with a recent gene transposition in *Drosophila miranda*.** *Genetics* 2000, **154**:1753-1763.  
The authors provide evidence for a selective sweep associated with the recent transposition of a gene from the partner of a degenerating neo-Y chromosome onto a different chromosome. Selection for dosage compensation for the loss of activity of the neo-Y chromosomal allele is implicated, suggesting that such duplication may provide a novel way of evolving dosage compensation.
26. Karn RC, Nachman MW: **Reduced nucleotide variability at an androgen-binding protein locus (*Abpa*) in house mice: evidence for positive natural selection.** *Mol Biol Evol* 1999, **16**:1192-1197.
27. Nachman MW, Crowell SL: **Contrasting evolutionary histories of two introns of the duchenne muscular dystrophy gene, *dmd*, in humans.** *Genetics* 2000, **155**:1855-1864.
28. Begun DJ, Aquadro CF: **Evolutionary inferences from DNA variation at the 6-Phosphogluconate dehydrogenase locus in natural populations of *Drosophila*: selection and geographic differentiation.** *Genetics* 1994, **136**:155-171.
29. Begun DJ, Aquadro CF: **Molecular variation at the *vermillion* locus in geographically diverse populations of *Drosophila melanogaster* and *D. simulans*.** *Genetics* 1995, **140**:1019-1032.
30. Hudson RR, Bailey K, Skarecky D, Kwiatowski J, Ayala FJ: **Evidence for positive selection in the *superoxide dismutase* (*Sod*) region of *Drosophila melanogaster*.** *Genetics* 1994, **136**:1329-1340.
31. Andolfatto P, Wall JD, Kreitman M: **Unusual haplotype structure at the proximal breakpoint of *In(2L)t* in a natural population of *Drosophila melanogaster*.** *Genetics* 1999, **153**:1297-1311.
32. Labate JA, Biermann CH, Eanes WF: **Nucleotide variation at the *run* locus in *Drosophila melanogaster* and *Drosophila simulans*.** *Mol Biol Evol* 1999, **16**:724-731.
33. Depaulis F, Brazier L, Veuille M: **Selective sweep at the *Drosophila melanogaster* *Suppressor of Hairless* locus and its association with the *In(2L)t* inversion polymorphism.** *Genetics* 1999, **152**:1017-1024.
34. Hamblin MT, Veuille M: **Population structure among African and derived populations of *Drosophila simulans*: evidence for ancient subdivision and recent admixture.** *Genetics* 1999, **153**:305-317.
35. Andolfatto P, Kreitman M: **Molecular variation at the *In(2L)t* proximal breakpoint site in natural populations of *Drosophila melanogaster* and *D. simulans*.** *Genetics* 2000, **154**:1681-1691.
36. Ford MJ: **Effects of natural selection on patterns of DNA sequence variation at the *transferrin*, *somatolactin*, and *p53* genes within and among chinook salmon (*Oncorhynchus tshawytscha*) populations.** *Mol Ecol* 2000, **9**:843-855.
37. Duvernell DD, Eanes WF: **Contrasting molecular population genetics of four hexokinases in *Drosophila melanogaster*, *D. simulans* and *D. yakuba*.** *Genetics* 2000, **156**:1191-1201.
38. Pogson GH: **Nucleotide polymorphism and natural selection at the *panthophysin* (*PanI*) locus in the Atlantic cod, *Gadus morhua* (L.).** *Genetics* 2001, **157**:317-330.
39. Rozas J, Gulland M, Blandin G, Aguadé M: **DNA variation at the *rp49* gene region of *Drosophila simulans*: evolutionary inferences from an unusual haplotype structure.** *Genetics* 2001, **158**:1147-1155.
40. Tishkoff SA, Varkonyi R, Cahinhinan N, Abbes S, Argyropoulos G, Destro-Bisol G, Drousiotou A, Dangerfield B, Lefranc G, Loiselet J *et al.*: **Haplotype diversity and linkage disequilibrium at human *G6pd*: recent origin of alleles that confer malarial resistance.** *Science* 2001, **293**:455-462.
41. Zurovcova M, Eanes WF: **Lack of nucleotide polymorphism in the Y-linked sperm flagellar dynein gene *Dhc-Yh3* of *Drosophila melanogaster* and *D. simulans*.** *Genetics* 1999, **153**:1709-1715.
42. Carvalho AB, Lazzaro BP, Clark AG: **Y chromosomal fertility factors *kl-2* and *kl-3* of *Drosophila melanogaster* encode dynein heavy chain polypeptides.** *Proc Natl Acad Sci USA* 2000, **97**:13239-13244.
43. Begun DJ, Whitley P: **Reduced X-linked nucleotide polymorphism in *Drosophila simulans*.** *Proc Natl Acad Sci USA* 2000, **97**:5960-5965.  
Evidence for reduced variability on the X chromosome relative to autosomes in *D. simulans*. Background selection and a lower mutation rate on the X can be ruled out as explanations. Models of positive selection, including those which posit recessivity of advantageous mutations, are invoked as possible explanations.
44. Orr HA, Betancourt AJ: **Haldane's sieve and adaptation from the standing genetic variation.** *Genetics* 2001, **157**:875-884.  
The authors show that when advantageous variants are chosen from standing variation, their probability of fixation is roughly independent of their degree of recessivity. Thus, the prediction of faster rates of adaptation on the X chromosome only applies to newly arising variants.
45. Barton NH: **Genetic hitchhiking.** *Philos Trans R Soc Lond B Biol Sci* 2000, **355**:1553-1562.
46. David JR, Capi P: **Genetic variation of *Drosophila melanogaster* natural populations.** *Trends Genet* 1988, **4**:106-111.
47. Begun DJ, Aquadro CF: **African and North American populations of *Drosophila melanogaster* are very different at the DNA level.** *Nature* 1993, **365**:548-550.
48. Andolfatto P: **Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans*.** *Mol Biol Evol* 2001, **18**:279-290.
49. Wall JD, Przeworski M: **When did the human population size start increasing?** *Genetics* 2000, **155**:1865-1874.
50. Tishkoff SA, Dietsch E, Speed W, Pakstis AJ, Kidd JR, Cheung K, Bonne-Tamir B, Santachiara-Benerecetti AS, Moral P, Krings M: **Global patterns of linkage disequilibrium at the *CD4* locus and modern human origins.** *Science* 1996, **271**:1380-1387.
51. Frisse L, Hudson RR, Bartoszewicz A, Wall JD, Donfack J, Di Rienzo A: **Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels.** *Am J Hum Genet* 2001, **69**:831-843.
52. Fay JC, Wu CI: **A human population bottleneck can account for the discordance between patterns of mitochondrial versus nuclear DNA variation.** *Mol Biol Evol* 1999, **16**:1003-1005.
53. Charlesworth B: **The effect of life-history and mode of inheritance on neutral genetic variability.** *Genet Res* 2001, **77**:153-166.
54. Begun DJ: **The frequency distribution of nucleotide variation in *Drosophila simulans*.** *Mol Biol Evol* 2001, **18**:1343-1352.
55. Tajima F: **Statistical method for testing the neutral mutation hypothesis by DNA polymorphism.** *Genetics* 1989, **123**:585-595.
56. Kreitman M: **Methods to detect selection in populations with applications to the human.** *Annu Rev Genom Hum Genet* 2000, **1**:539-559.
57. Andolfatto P, Przeworski M: **A genome-wide departure from the standard neutral model in natural populations of *Drosophila*.** *Genetics* 2000, **156**:257-268.  
Most loci sampled from (predominantly non-African) populations of *D. melanogaster* and *D. simulans* exhibit greater linkage disequilibrium than expected under neutral panmictic equilibrium models. The generality of this pattern implicates demography as an important determinant of linkage disequilibrium levels in these two species.
58. Teeter K, Naeemuddin M, Gasperini R, Zimmerman E, White KP, Hoskins R, Gibson G: **Haplotype dimorphism in a SNP collection from *Drosophila melanogaster*.** *J Exp Zool* 2000, **288**:63-75.
59. Begun DJ, Whitley P: **Adaptive evolution of relish, a *Drosophila* *NF-kappaB/IkappaB* protein.** *Genetics* 2000, **154**:1231-1238.

60. Hasson E, Wang IN, Zeng LW, Kreitman M, Eanes WF: **Nucleotide variation in the triosephosphate isomerase (*Tpi*) locus of *Drosophila melanogaster* and *Drosophila simulans*.** *Mol Biol Evol* 1998, **15**:756-769.
61. Galtier N, Depaulis F, Barton NH: **Detecting bottlenecks and selective sweeps from DNA sequence polymorphism.** *Genetics* 2000, **155**:981-987.
- A full-likelihood method is applied to multi-locus data to distinguish between a bottleneck model and a selective sweep model (treated here as nested models). The authors conclude that the selective sweep model is a better explanation for variability patterns observed at three loci in a single West African population of *D. melanogaster*.
62. Charlesworth B, Nordborg M, Charlesworth D: **The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations.** *Genet Res* 1997, **70**:155-174.
63. Stephan W, Xing L, Kirby DA, Braverman JM: **A test of the background selection hypothesis based on nucleotide data from *Drosophila ananassae*.** *Proc Natl Acad Sci USA* 1998, **95**:5649-5654.
64. Hamblin MT, Di Rienzo A: **Detection of the signature of natural selection in humans: evidence from the *Duffy* blood group locus.** *Am J Hum Genet* 2000, **66**:1669-1679.
- Evidence for local adaptive hitchhiking at a locus conferring resistance to the malaria parasite, *Plasmodium vivax*. The paper contains a good discussion about how uncertainty about human demography complicates the interpretation of the data.
65. Hollox EJ, Poulter M, Zvarik M, Ferak V, Krause A, Jenkins T, Saha N, Kozlov AI, Swallow DM: **Lactase haplotype diversity in the Old World.** *Am J Hum Genet* 2001, **68**:160-172.
66. Rana BK, Hewett-Emmett D, Jin L, Chang BH, Sambuughin N, Lin M, Watkins S, Bamshad M, Jorde LB, Ramsay M *et al.*: **High polymorphism at the human *melanocortin 1 receptor* locus.** *Genetics* 1999, **151**:1547-1557.
67. Slatkin M, Wiehe T: **Genetic hitch-hiking in a subdivided population.** *Genet Res* 1998, **71**:155-160.
68. Majewski J, Cohan FM: **Adapt globally, act locally: the effect of selective sweeps on bacterial sequence diversity.** *Genetics* 1999, **152**:1459-1474.
69. Hill WG, Robertson A: **The effect of linkage on limits to artificial selection.** *Genet Res* 1966, **8**:269-294.
70. Stephan W: **An improved method for estimating the rate of fixation of favorable mutations based on DNA polymorphism data.** *Mol Biol Evol* 1995, **12**:959-962.
71. Nachman MW, Crowell SL: **Estimate of the mutation rate per nucleotide in humans.** *Genetics* 2000, **156**:297-304.