Note

Distinguishing the Hitchhiking and Background Selection Models

Hideki Innan*,[1] and Wolfgang Stephan†

*Human Genetics Center, University of Texas Health Science Center, Houston, Texas 77030 and †Section of Evolutionary Biology, Department of Biology II, University of Munich, 80333 Munich, Germany

Manuscript received April 29, 2003
Accepted for publication August 1, 2003

ABSTRACT

A simple method to distinguish hitchhiking and background selection is proposed. It is based on the observation that these models make different predictions about the average level of nucleotide diversity in regions of low recombination. The method is applied to data from *Drosophila melanogaster* and two highly selfing tomato species.

ONE of the signatures of genome-wide selection is the positive correlation between the amount of polymorphism and recombination rate, which was found in Drosophila (Begun and Aquadro 1992), humans (Nachman 2001), the partial selfer *Caenorhabditis elegans* (Cutter and Payseur 2003), and several outcrossing and partially selfing plant species (reviewed in Nordborg and Innan 2002), including wild tomatoes (Stephan and Langley 1998). Hitchhiking (HH) and background selection (BS) are considered as the most important forces causing this positive correlation. Under the HH model, adaptive fixations of strongly favored mutations reduce the level of variation, because such fixations sweep out neutral polymorphisms in the surrounding region while some of them "hitchhike" with the favored mutations (Maynard Smith and Haigh 1974; Kaplan *et al.* 1989). The BS model considers negative (purifying) selection against deleterious mutations as the cause of the reduction of the amount of variation (Charlesworth *et al.* 1993). Negative selection works to eliminate deleterious mutations together with linked neutral variants. Recombination is a very important factor in determining the degree of reduction in the amount of neutral variation in both models. The lower the recombination rate is, the more variation is swept out by a single adaptive fixation. The probability that a neutral polymorphism is eliminated by a single deleterious mutation also increases as the recombination rate decreases. Thus, the two modes of selection may explain the positive correlation between the level of variation and recombination rate. There is no doubt that both selection processes occur. The joint action of the two modes of selection also creates a positive correlation between recombination rate and levels of variation (Kim and Stephan 2000). However, the relative importance of these two models is not well understood and still vigorously debated (reviewed in Andolfatto 2001). In this note, we present an approach to distinguish the two selection models on the basis of data of levels of DNA polymorphism and recombination rates.

Theoretical studies have shown that, in a diploid population (with constant effective size N_e) undergoing recurrent hitchhiking events or background selection, the expected degree of reduction in neutral polymorphism (f) is a function of ρ, the recombination rate per site per generation. That is, the expectation of the amount of variation in a region with a local recombination rate ρ is given by

$$E(\theta) = f\theta_{neu},$$

where $\theta_{neu} = 4N_e\mu$ and μ is the neutral mutation rate per generation. On the basis of the work of Kaplan *et al.* (1989) and Stephan *et al.* (1992), Wiehe and Stephan (1993) obtained a simple formula for f in the HH model,

$$f = \rho/(\rho + a), \quad (1)$$

where a is a parameter that depends on the product of the population selection parameter (population size times selection coefficient) and the rate of sweeps per generation. It should be noted that this equation has several assumptions. First, θ_{neu} and a are constant over the genome. Second, the equation considers only selective sweeps but other types of selection (*e.g.*, negative selection, balancing selection) are neglected. Third, the local recombination rate ρ has a uniform distribution (*i.e.*, recombination hot and cold spots are ignored). Last, multiple concurrent selective sweeps are not allowed, so that (1) does not hold when recombination

[1]*Corresponding author:* Human Genetics Center, School of Public Health, University of Texas Health Science Center, 1200 Hermann Pressler Dr., Houston, TX 77030. E-mail: hideki.innan@uth.tmc.edu

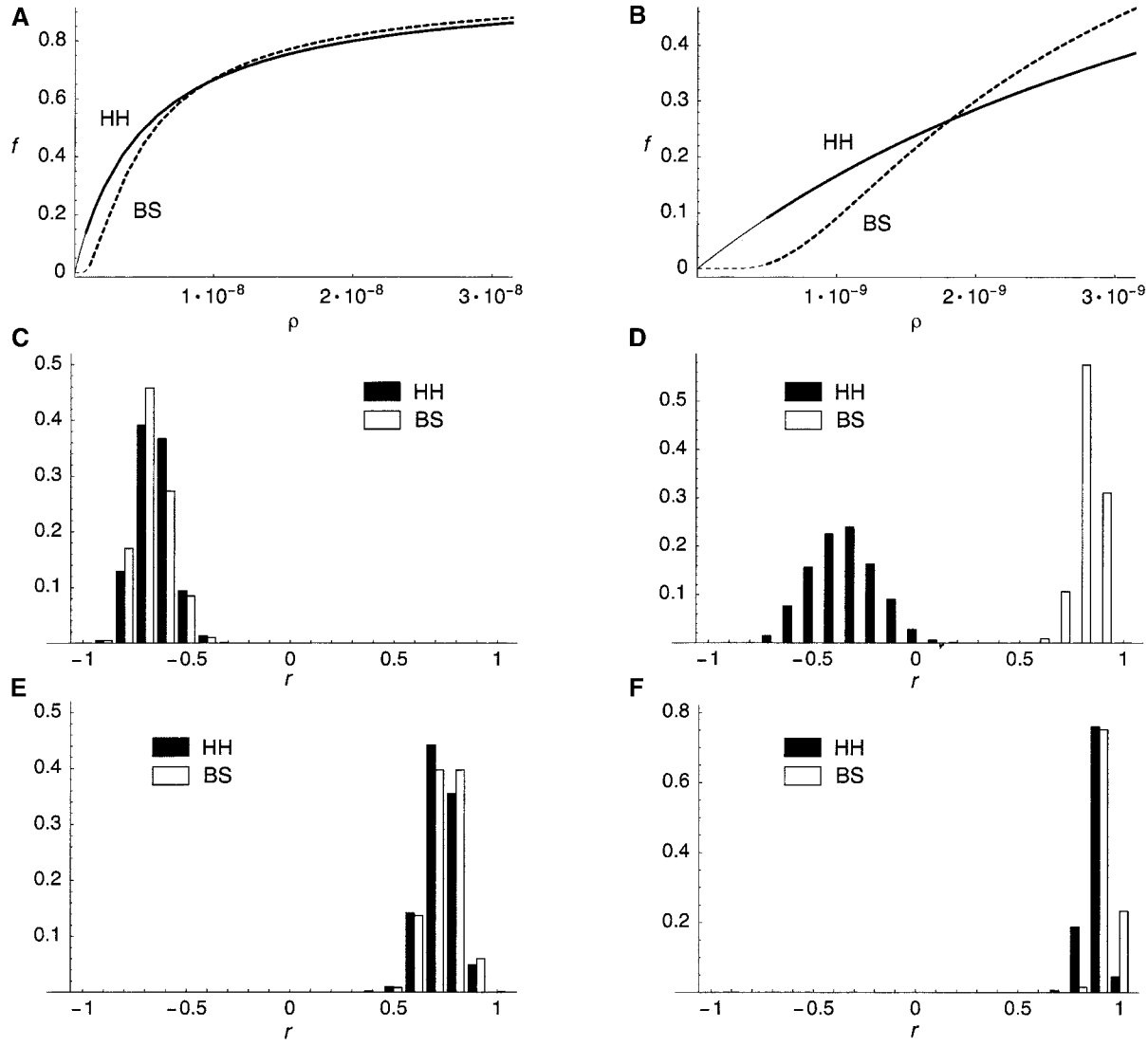

FIGURE 1.—(A) Genetic variation f as a function of the recombination rate ρ for the HH and BS models. The solid curve is for the HH model obtained from (1) with $a = 5 \times 10^{-9}$, and the broken curve is for the BS model obtained from (2) with $u = 4 \times 10^{-9}$. It should be noted that (1) and (2) are not very good approximations when ρ is very small (those parts of f are shown by thin lines). (B) Genetic variation f as a function of the recombination rate ρ for the HH and BS models in regions of low recombination ($\rho \leq 3 \times 10^{-9}$). $a = 5 \times 10^{-9}$ and $u = 2.4 \times 10^{-9}$ are assumed. (C–F) Distributions of the correlation of coefficient r under the HH (solid bars) and BS (open bars) models. See text for details.

rate is very small due to interference among them. KIM and STEPHAN (2003) showed that if the asymptotic value of f, $f_{|\rho=0}$, is ≤0.1, then (1) holds for all ρ values such that $f \geq 0.1$. Furthermore, if $f_{|\rho=0} \geq 0.1$, (1) holds approximately for all ρ values for which $f \geq f_{|\rho=0}$.

Under the BS model and the assumption that recombination rate is not extremely low, f is approximately given by

$$f = \exp(-u/\rho), \qquad (2)$$

where u is the deleterious mutation rate per site per generation (HUDSON and KAPLAN 1995). This equation considers the effect of negative selection alone and also requires assumptions of constant θ_{neu} and u values and a uniform distribution of ρ. It is known that these two equations (1 and 2) produce similar functional relationships for a wide range of ρ. In Figure 1A, the solid curve represents the function of f for the HH model obtained from Equation 1 with $a = 5 \times 10^{-9}$, which is an estimate for *Drosophila melanogaster* (STEPHAN 1995). Figure 1A also shows that we can generate a very similar curve using Equation 2 with $u = 4$ (or 5) $\times 10^{-9}$ (broken line). In the case of HH, f converges to $1 - a/\rho$ when ρ gets large, while for BS f converges to $1 - u/\rho$ for large ρ. That means for $a = u$ the two functions are asymptotically identical. Therefore, given a data set of levels of DNA polymorphism and recombination rates, we can fit both Equations 1 and 2 to the data. For this reason, it is impossible to distinguish the two selection models on the basis of these types of data for large ρ values.

However, a close examination reveals a difference between the two functions in regions of low (but not too low) recombination rates such that both equations are still valid (see above). Figure 1B shows f under the two models with $a = 5 \times 10^{-9}$ for the HH model and $u = 2.4 \times 10^{-9}$ for the BS model. Although the two parameters are chosen to give the same average level of f for $1 \times 10^{-9} \leq \rho \leq 3 \times 10^{-9}$, it is evident that the shapes of the two curves are different. The curve of the HH model is convex in this parameter range while that of the BS model is concave, suggesting that polymorphism data from regions of low recombination might be useful to distinguish the two selection models. Focusing on this difference between the two functions, we propose a simple method with which to distinguish the two selection models. The idea is from STEPHAN (1995), who suggested applying Equation 1 to data by transforming it into the following linear regression formula:

$$E(\theta) = \theta_{neu} - a\frac{E(\theta)}{\rho}. \quad (3)$$

Suppose now that polymorphism data from n independent loci (DNA regions) are available from a single species. Let $\hat{\theta}_i$ be the estimated amount of DNA polymorphism ($\hat{\theta}$) at the ith locus ($i = 1, 2, 3, \ldots, n$). Let ρ_i be the recombination rate (ρ) at the ith locus. We assume that recombination rates are known. Equation 3 indicates that $\hat{\theta}$ has a negative linear correlation with $\hat{\theta}/\rho$. That is, the correlation coefficient between $\hat{\theta}$ and $\hat{\theta}/\rho$ is -1 if $\hat{\theta}_i = \theta_{HHi}$ at all the loci, where θ_{HHi} is the expectation of $\hat{\theta}_i$ under the HH model given ρ_i (obtained from Equation 1). On the other hand, we expect that r is relatively close to $+1$ under the BS model because of the concave behavior of f in regions of low recombination.

In practice, $\hat{\theta}$ is never exactly the same as the theoretical expectation due to genetic drift. Therefore, we investigate the distribution of the correlation coefficient (r) between $\hat{\theta}$ and $\hat{\theta}/\rho$, taking the variance of $\hat{\theta}$ into account. First, the distribution of r is investigated under the HH model assuming $\hat{\theta}_i$ has a normal distribution with mean θ_{HHi} and standard deviation (SD) $k\theta_{HHi}$, where k is a constant value. A computer simulation is carried out in the following way:

1. Determine θ_{neu}, a, and k.
2. Simulate θ for the n loci. θ_i is assumed to be a random variable on the basis of a normal distribution with mean θ_{HHi} and SD $k\theta_{HHi}$. If $\theta_i < 0$, $\theta_i = 0$ is set.
3. Calculate the correlation coefficient, r, between θ and θ/ρ using the simulated θ for the n loci.

Steps 2 and 3 are repeated 10,000 times and the distribution of r is obtained. The procedure to obtain the null distribution of r under the BS model is almost identical to this. That is, θ_i is simulated as a random variable on the basis of a normal distribution with mean θ_{BSi} and SD $k\theta_{BSi}$, where θ_{BSi} is the expectation of θ_i under the BS model given ρ_i according to Equation 2.

Figure 1, C and D, shows the results of the distributions of r. Figure 1, C and E, investigates the distributions when regions of high recombination are studied, and Figure 1, D and F, is for regions of low recombination. a and u in Figure 1, C and E, are the same as in Figure 1A, while in Figure 1, D and F, we use the same a and u as those in Figure 1B. We consider $n = 20$ loci, whose recombination rates are assumed to be $\rho_i = (i + 5) \times 10^{-9}$ so that the range of ρ is 6–25 $\times 10^{-9}$ in Figure 1, C and E. In Figure 1, D and F, ρ_i is assumed to be $(i + 10) \times 10^{-10}$ so that ρ_i ranges from 1.1×10^{-9} to 3×10^{-9}. Figure 1, C and D, studies a case of small k ($k = 0.1$), while $k = 1$ is assumed in Figure 1, E and F.

First, we consider Figure 1D, which shows the distributions of r when regions of low recombination are studied and k is small. The distribution of r under the HH model is nearly symmetrical and the average is -0.33, while r under the BS model has a relatively narrow distribution close to $+1$. This result indicates that r might be a useful summary statistic to distinguish the HH and BS models since the distributions of r are completely different in the two models. However, it should be noted that this method does not work when applied to regions of high recombination. As shown in Figure 1C, the two distributions of r are very similar as expected (discussed above).

The power of this method to distinguish the two models depends on k. Figure 1, E and F, shows the results of the same analysis as those in Figure 1, C and D, respectively, but $k = 1$ is assumed instead of $k = 0.1$. The two distributions of r for the case of low recombination are quite similar, although not identical (Figure 1F). Figure 1E shows the distributions of r when regions of high recombination are investigated. The two distributions are very similar again, except that the means have moved to $\sim$0.7. These results suggest that the two selection models can be best distinguished under the following two conditions: (1) Polymorphism data from regions of low (but not too low) recombination are available; (2) the variances of the estimates of variation are sufficiently small.

Next we discuss how this method may be applied to data. Suppose that we have a data set of estimates of θ from n independent loci and that we know the local recombination rates for the n loci. First, the correlation coefficient between $\hat{\theta}$ and $\hat{\theta}/\rho$, r_{obs}, is calculated, and then r_{obs} is compared with the null distributions of r under the HH and BS models. The procedure described above is modified because we need to estimate a, u, and k from the data. That is, step 1 in the procedure should be replaced by the following two steps:

1a. Determine θ_{neu}.
1b. For the HH model, find a, which gives the best fit of Equation 1 to the data by a least-squares method, which also gives an estimate of k. In a similar way,

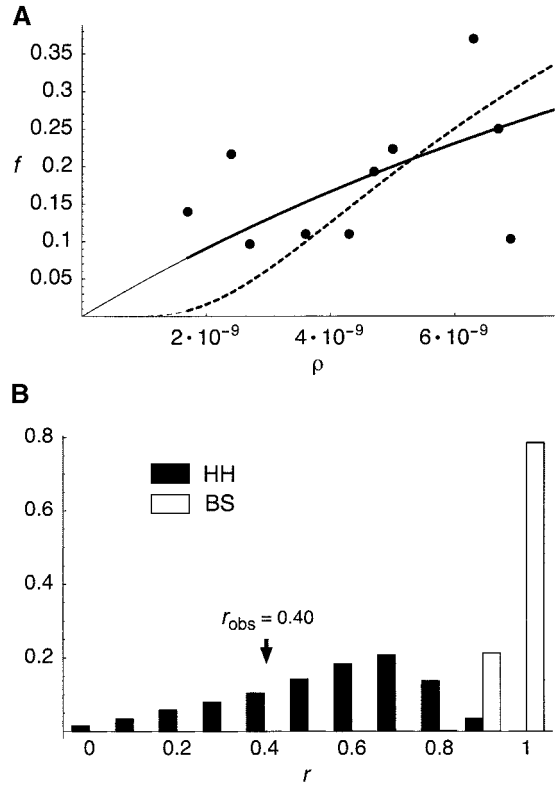

FIGURE 2.—Application of the test to Drosophila.

for the BS model, u is estimated using Equation 2 together with k.

Then, we can follow steps 2 and 3.

We apply this method to the data of *D. melanogaster* from ANDOLFATTO and PRZEWORSKI (2001). We use the 10 X-linked loci with $\rho < 7 \times 10^{-9}$ [*yellow* and *su(s)* are excluded because of too low recombination rates; see Table 2 in ANDOLFATTO and PRZEWORSKI (2001) for details]. Figure 2A shows the observed levels of polymorphism in the 10 loci scaled by θ_{neu}, which is assumed to be 0.03 (*e.g.*, ANDOLFATTO 2001; ANDOLFATTO and PRZEWORSKI 2001). The correlation coefficient of this data set is $r_{obs} = 0.40$. Following the procedure described above, we found the best-fit functions of Equations 1 and 2 when $a = 2 \times 10^{-8}$ and $u = 0.8 \times 10^{-8}$, respectively (solid and dashed lines in Figure 2A). Figure 2B shows the null distributions of r under the HH and BS models with these estimated a and u. The null distribution of the HH model has a relatively wide range and r_{obs} is almost in the middle of the distribution. On the other hand, the BS model predicts a narrow distribution of r around +1 and r_{obs} is too small ($P < 0.0001$). Similar results are obtained when $\theta_{neu} = 0.02$ and 0.04. These results indicate that the observed distribution of the amount of variation in the 10 loci is explained better by a convex function than by a concave one, suggesting that it is very difficult to explain the observation by background selection *alone*. Hitchhiking might be the dominant force creating the pattern of standing polymorphism on the X chromosome of *D. melanogaster*. This is consistent with the conclusion of ANDOLFATTO and PRZEWORSKI (2001), who used a different way to distinguish the two selection models.

Other interesting species to study are partially selfing plants, in which recombination is "effectively" reduced in the whole genome. We use two tomato species, *Lycopersicon pimpinellifolium* and *L. chmielewskii*, whose selfing rate (S) is ~0.9 (RICK 1966, 1983). The polymorphism data are from MILLER and TANKSLEY (1990; also summarized in Table 1 of STEPHAN and LANGLEY 1998). They studied restriction fragment length polymorphism in ~40 loci of nine tomato species including the two highly selfing species mentioned above.

NORDBORG (1997) suggested that population genetics theory for outcrossing species can be easily applied to partially selfing species by rescaling parameters using the inbreeding coefficient, $F = S/(2 - S)$. That is, the recombination rate is decreased to $\bar{\rho} = (1 - F)\rho$, the effective population size decreases to $\bar{N}_e = N_e/(1 + F)$, and selection intensity increases by a factor $1 + F$ when the effect of selection is additive. Then, $\bar{f}$ is defined as the reduction of the amount of polymorphism in comparison with the rescaled neutral expectation, $\bar{\theta}_{neu} = \theta_{neu}/(1 + F)$. Note that the rescaled parameters are represented by a bar. Thus, plugging these rescaling coefficients into (1) and (2), we can study the joint effects of the two mechanisms, selfing and selection, both of which decrease the level of polymorphism.

Then, we apply our method to *L. pimpinellifolium* and *L. chmielewskii*. To avoid the problem that Equations 1 and 2 are invalid when recombination rate is very small, we use only 29 loci of MILLER and TANKSLEY's (1990) data set, excluding loci of very low recombination. $\hat{\theta}$ and ρ for the 29 loci are according to Table 1 in STEPHAN and LANGLEY (1998). $\hat{\theta}$ ranges from 0 to 0.0275 in *L. pimpinellifolium* and from 0 to 0.0143 in *L. chmielewskii*. The recombination rates are rescaled to per site per generation values by multiplying them by a factor of 12.1×10^{-8}. This factor results from the fact that the tomato genome size is ~950 Mb (SHERMAN and STACK 1995; PILLEN *et al.* 1996). The recombination rates for the investigated 29 loci are in the range $1.1 \times 10^{-8} \leq \rho \leq 2.7 \times 10^{-8}$ (or $2 \times 10^{-9} \leq \bar{\rho} \leq 4.9 \times 10^{-9}$). Then, we obtain the correlation coefficient $r_{obs} = 0.945$ and 0.949 for *L. pimpinellifolium* and *L. chmielewskii*, respectively. These very high values of r seem to favor the BS model.

To test this possibility, we investigate null distributions of r under each selection model. We use a relatively wide range of $\bar{\theta}_{neu}$ since it is very difficult to estimate $\bar{\theta}_{neu}$ for highly selfing species in which the level of polymorphism is reduced in the whole genome. The probabilities that r exceeds the observation (r_{obs}) are shown in Table 1. These probabilities are relatively low under the HH model for the two species, suggesting that r_{obs}

TABLE 1

Application to tomatoes

		HH model		BS model	
$\bar{\theta}_{neu}$	$\hat{\theta}/\bar{\theta}_{neu}$[a]	$\bar{f}$[b]	$P(r > r_{obs})$	$\bar{f}$[b]	$P(r > r_{obs})$
		L. pimpinellifolium ($r_{obs} = 0.945$)			
0.02	0–1.38	0.303–0.523	0.015	0.227–0.555	0.035
0.025	0–1.10	0.232–0.431	0.008	0.167–0.491	0.064
0.03	0–0.92	0.197–0.381	0.009	0.111–0.418	0.168
0.04	0–0.69	0.140–0.291	0.020	0.067–0.341	0.362
0.05	0–0.55	0.109–0.235	0.013	0.044–0.290	0.585
0.06	0–0.46	0.089–0.198	0.023	0.033–0.257	0.745
		L. chmielewskii ($r_{obs} = 0.949$)			
0.01	0–1.43	0.197–0.381	0.068	0.123–0.435	0.312
0.015	0–0.95	0.131–0.275	0.073	0.060–0.328	0.594
0.02	0–0.72	0.093–0.206	0.103	0.033–0.257	0.866
0.025	0–0.57	0.075–0.170	0.088	0.022–0.218	0.940

[a] Range of $\hat{\theta}$ from data scaled by $\bar{\theta}_{neu}$ (based on data).
[b] Range of $E(\theta)$ scaled by $\bar{\theta}_{neu}$ predicted by Equations 1 and 2 under the HH and BS models, respectively.

may be too big to be expected under the HH model, especially for *L. pimpinellifolium*. $P(r > r_{obs})$ seems to be quite robust to $\bar{\theta}_{neu}$. Under the BS model, $P(r > r_{obs})$ is relatively sensitive to $\bar{\theta}_{neu}$. This may be because f under the BS model can be either concave or convex depending on $\bar{\theta}_{neu}$. The two models can be distinguished well for $\bar{\theta}_{neu}$ values that generate a concave shape of f. r_{obs} for the two species may be in the acceptable range under the BS model unless a very small $\bar{\theta}_{neu}$ is assumed. For example, $r_{obs} = 0.945$ of *L. pimpinellifolium* could be too big even under the BS model if $\bar{\theta}_{neu} = 0.02$ is assumed, but this small value of $\bar{\theta}_{neu}$ seems to be quite unrealistic because $\hat{\theta}$ can be as large as $1.38 \times \bar{\theta}_{neu}$ in this highly selfing species. Thus, the results of Table 1 seem to suggest that background selection has played a larger role than hitchhiking in shaping genome-wide patterns of variation in the history of these two tomato species.

In this note, we proposed a method to distinguish the HH and BS models. Since the test looks at whether the level of polymorphism is a convex or concave function of the local recombination rate, we should have data from multiple regions in which the recombination rate is low (but not too low). The test is very powerful when the variance of $\hat{\theta}$ is low, indicating that θ should be estimated from sufficiently long regions (such that the variances of $\hat{\theta}$ are reduced due to intragenic recombination). Innan *et al.* (2003) showed that the distribution of $\hat{\theta}$ from 500-kb fragments on human chromosome 21 is very similar to a normal distribution with a quite small SD. The application of our method to Drosophila and tomatoes led to different results. That is, the HH model is preferred in Drosophila while the BS model could better explain the observation in tomatoes. This might be due to the difference in life style and mating system between animals and plants (*e.g.*, Baudry *et al.* 2001). Plant populations (especially highly selfing species) are generally more structured and selective sweeps in such a structured population might not occur as quickly as in a random-mating population (*e.g.*, Cherry and Wakeley 2003; Whitlock 2003).

However, there are some potential problems in the application of our method to the currently available data sets:

1. It was not possible to obtain correct estimates of $\bar{\theta}_{neu}$, especially in the highly selfing tomato species.
2. The variance of $\hat{\theta}$ is relatively large, which decreases the power of the test. Also, our assumption of a normal distribution of $\hat{\theta}$ may not be adequate. These problems could be fixed for Drosophila and other outcrossing species if data from very long regions of low recombination rates are available, together with data from regions of high recombination to estimate θ_{neu}. For highly selfing species, even with such data, it is very difficult to estimate $\bar{\theta}_{neu}$ because the level of polymorphism is reduced in the whole genome.
3. The theory assumes an unstructured population of constant size. The relationship between levels of variation and recombination rate should also be studied in other population models (see also Andolfatto and Przeworski 2001).
4. The theory assumes constant values of a and u in Equations 1 and 2 across the genome. Variation in these parameters could increase the variance of the observed amounts of polymorphism, reducing the power of the test. In such a case, problem 2 becomes more serious.

We thank Yuseob Kim for his stimulating study of the hitchhiking process with interference among adaptive fixations. We also thank

M. Aguadé and two anonymous reviewers for comments and suggestions. H.I. is supported by a fund from University of Texas. W.S. is grateful to the Erwin Schroedinger International Institute for Mathematical Physics in Vienna for support during his stay in winter 2002/2003 and to the Deutsche Forschungsgemeinschaft for funding (STE 325/5-1; Schwerpunktprogramm 1127).

LITERATURE CITED

Andolfatto, P., 2001 Adaptive hitchhiking effects on genome variability. Curr. Opin. Genet. Dev. **11:** 635–641.

Andolfatto, P., and M. Przeworski, 2001 Regions of lower crossing over harbor more rare variants in African populations of *Drosophila melanogaster*. Genetics **158:** 657–665.

Baudry, E., C. Kerdelhué, H. Innan and W. Stephan, 2001 Species and recombination effects on DNA variability in the tomato genus. Genetics **158:** 1725–1735.

Begun, D. J., and C. F. Aquadro, 1992 Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature **356:** 519–520.

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

Cherry, J. L., and J. Wakeley, 2003 A diffusion approximation for selection and drift in a subdivided population. Genetics **163:** 421–428.

Cutter, A. D., and B. A. Payseur, 2003 Selection at linked sites in the partial selfer *Caenorhabditis elegans*. Mol. Biol. Evol. **20:** 665–673.

Hudson, R. R., and N. L. Kaplan, 1995 Deleterious background selection with recombination. Genetics **141:** 1605–1617.

Innan, H., B. Padhukasahasram and M. Nordborg, 2003 The pattern of polymorphism on human chromosome 21. Genome Res. **13:** 1158–1168.

Kaplan, N. L., R. R. Hudson and C. H. Langley, 1989 The "hitchhiking" effect revisited. Genetics **123:** 887–899.

Kim, Y., and W. Stephan, 2000 Joint effects of genetic hitchhiking and background selection on neutral variation. Genetics **155:** 1415–1427.

Kim, Y., and W. Stephan, 2003 Selective sweeps in the presence of interference among partially linked loci. Genetics **164:** 389–398.

Maynard Smith, J., and J. Haigh, 1974 The hitchhiking effect of a favourable gene. Genet. Res. **23:** 23–35.

Miller, J. C., and S. D. Tanksley, 1990 RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. Theor. Appl. Genet. **80:** 437–448.

Nachman, M. W., 2001 Single nucleotide polymorphism and recombination rate in humans. Trends Genet. **17:** 481–485.

Nordborg, M., 1997 Structured coalescent processes on different time scales. Genetics **146:** 1501–1514.

Nordborg, M., and H. Innan, 2002 Molecular population genetics. Curr. Opin. Plant Biol. **5:** 69–73.

Pillen, K., O. Pineda, C. B. Lewis and S. D. Tanksley, 1996 Status of genome mapping tools in the taxon Solanaceae, pp. 281–308 in *Genome Mapping in Plants*, edited by A. H. Paterson and R. G. Landes. Academic Press, Austin, TX.

Rick, C. M., 1966 Some plant-animal relations on the Galapagos islands, pp. 215–224 in *The Galapagos*, edited by R. I. Bowman. University of California Press, Berkeley, CA.

Rick, C. M., 1983 Evolution of mating systems: evidence from allozyme variation, pp. 215–221 in *Genetics: New Frontiers (Proceedings of the XV International Congress on Genetics, New Delhi, December 1983)*, edited by V. L. Chopra, B. C. Joshi, R. P. Sharma and H. C. Bansal. Oxford & IBH Publishing, New Delhi.

Sherman, J. D., and S. M. Stack, 1995 Two-dimensional spreads of synaptonemal complexes from solanaceous plants. VI. High-resolution recombination nodule map for tomato (*Lycopersicon esculentum*). Genetics **141:** 683–708.

Stephan, W., 1995 An improved method for estimating the rate of fixation of favorable mutations based on DNA polymorphism data. Mol. Biol. Evol. **12:** 959–962.

Stephan, W., and C. H. Langley, 1998 DNA polymorphism in Lycopersicon and crossing-over per physical length. Genetics **150:** 1585–1593.

Stephan, W., T. H. E. Wiehe and M. W. Lenz, 1992 The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. Theor. Popul. Biol. **41:** 237–254.

Whitlock, M. C., 2003 Fixation probability and time in subdivided populations. Genetics **164:** 767–779.

Wiehe, T. H. E., and W. Stephan, 1993 Analysis of a genetic hitchhiking model and its application to DNA polymorphism data. Mol. Biol. Evol. **10:** 842–854.

Communicating editor: M. Aguadé