Measures of Divergence Between Populations and the Effect of Forces that Reduce Variability

Brian Charlesworth[1]

Department of Ecology and Evolution, University of Chicago

Wright's F_{ST} and related statistics are often used to measure the extent of divergence among populations of the same species relative to the net genetic diversity within the species. This paper compares several definitions of F_{ST} which are relevant to DNA sequence data, and shows that these must be used with care when estimating migration parameters. It is also pointed out that F_{ST} is strongly influenced by the level of within-population diversity. In situations where factors such as selection on closely linked sites are expected to have stronger effects on within-population diversity at some loci than at others, differences among loci can result entirely from differences in within-population diversities. It is shown that several published cases of differences in F_{ST} among regions of high and low recombination in *Drosophila* may be caused in this way. For the purpose of comparisons of levels of between-population differences among loci or species which are subject to different intensities of forces that reduce variability within local populations, absolute measures of divergence between populations should be used in preference to relative measures such as F_{ST}.

Introduction

Two recent studies of DNA variation within and between populations of *Drosophila* have reported much higher levels of between-population differentiation for genes in regions of low recombination than for genes in regions with normal recombination rates, as measured by F_{ST} (Wright 1951) or related statistics (Begun and Aquadro 1993, 1995; Stephan 1994). Similarly, there are much higher between-semispecies F_{ST} values for loci on the X chromosome of *Drosophila athabasca* than for autosomal loci (Ford and Aquadro 1996). Within-population variability is reduced in all three cases at the loci which show high F_{ST} values. This has led to the suggestion that these patterns are caused either by recent selection for locally favored alleles at the loci under study, or by selective sweeps of locally favored alleles at nearby loci (Begun and Aquadro 1993, 1995; Stephan 1994).

The purposes of this paper are (1) to discuss different measures of between-population differentiation that may be applied to data on nucleotide site variation; (2) to show that the patterns reported by Stephan (1994) and Begun and Aquadro (1993, 1995) probably do not require local selective sweeps but can be explained by any process that reduces within-population diversity; (3) to discuss various scenarios that might explain the observed patterns; and (4) to point out that measures of the relative amounts of between-population and total diversity, such as F_{ST}, are not necessarily appropriate if we wish to compare loci with very different levels of within-population variation.

[1] Present address: Institute for Cell, Animal, and Population Biology, University of Edinburgh, Edinburgh, U.K.

Key words: *Drosophila,* F_{ST}, population differentiation, population size, selection, linkage.

Address for correspondence and reprints: Institute for Cell, Animal, and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, United Kingdom. E-mail: brian.charlesworth@ed.ac.uk.

Methods

Measures of Genetic Diversity Within and Between Populations

Measures of population differentiation can be computed using either data on allele frequencies, as in the case of allozyme or microsatelllite data, or data on variation at the nucleotide site level. There is little difference between these genetic levels for descriptive purposes, although the evolutionary interpretation of the results may differ because of differences in the underlying genetic mechanisms generating variability. I will concentrate here on nucleotide site diversity values, because these were used in the above *Drosophila* studies, and because a simple evolutionary model (the infinite-sites model [Kimura 1971]) is plausible as an underlying mechanism. For simplicity, I will describe the measures in terms of population parameters rather than sample statistics. A diversity measure is then defined as the probability that two alleles with a defined origin differ at a random nucleotide site. In practice, an estimate of a diversity measure is obtained from the mean number of differences between a pair of alleles, normalizing by the numbers of bases in the sequence (Nei 1987, chapter 10).

Pairwise measures of diversity per nucleotide site for a set of *n* populations of the same species can be partitioned into the total diversity (π_T), calculated by pooling the set of populations and averaging over pairs of distinct alleles sampled randomly from the set, and the mean within-population diversity (π_S) (cf. Nei 1973; Holsinger and Mason-Gamer 1996). In general, the contribution of a population to the pool should be weighted by its size, but this is usually unknown, so either sample sizes are used as surrogates for population size, or all populations are weighted equally. Let the weight attached to population *i* be w_i, (the weights are assumed to sum to one), let the diversity for a pair of alleles drawn from population *i* be π_{ii}, and let the diversity between alleles drawn from two different populations *i* and *j* be π_{ij}. We have

$$\pi_T = \sum_i w_i^2 \pi_{ii} + 2 \sum_{i<j} w_i w_j \pi_{ij} \qquad (1a)$$

$$\pi_S = \frac{\sum_i w_i^2 \pi_{ii}}{\sum_i w_i^2}. \qquad (1b)$$

In addition, the mean divergence between a pair of alleles sampled from two different populations is

$$\pi_B = \frac{\sum_{i<j} w_i w_j \pi_{ij}}{\sum_{i<j} w_i w_j}. \qquad (1c)$$

Following Nei (1973), the between-population component of diversity, correcting for within-population diversity, is given by

$$\pi_{T-S} = \pi_T - \pi_S. \qquad (1d)$$

Under the infinite-sites model, expected neutral nucleotide site diversities for pairs of alleles are proportional to the corresponding expected coalescence times, since, at most, one new mutation per site is assumed to have originated since the origin of the genealogy of the set of alleles (Hudson 1990). The expectations for the diversity measures π_S and π_B in the above expressions are thus proportional to the mean coalescence times of pairs of alleles sampled from the same population ($\bar{t}_0$) or from different populations ($\bar{t}_1$), respectively (Slatkin 1991, 1993). The difference between these coalescent times reflects the increased mean time to coalescence caused by the time it takes for an allele to migrate to another population, where it can coalesce with a resident allele. Similarly, π_{T-S} is proportional to the difference in coalescence times between a pair of alleles sampled randomly from the set of populations as a whole and a pair of alleles sampled from a randomly chosen population. It thus provides a measure of the absolute degree of differentiation among populations (Nei 1973).

We can also write π_B as $\pi_S + \pi_D$, where π_D is proportional to $\bar{t}_1 - \bar{t}_0$ and may therefore be viewed as another appropriate absolute measure of between-population divergence (it corresponds to $\bar{D}_m$ of Nei [1973]). We thus have

$$\pi_T = \left(\sum_i w_i^2\right)\pi_S + 2\left(\sum_{i<j} w_i w_j\right)\pi_B$$

$$= \pi_S + 2\left(\sum_{i<j} w_i w_j\right)\pi_D \qquad (2a)$$

$$\pi_{T-S} = 2\left(\sum_{i<j} w_i w_j\right)\pi_D. \qquad (2b)$$

If equal weights of $1/n$ are attached to each population, these reduce to

$$\pi_T = \pi_S + \frac{(n-1)}{n}\pi_D \qquad (2c)$$

and

$$\pi_{T-S} = \frac{(n-1)}{n}\pi_D. \qquad (2d)$$

The amount of between-population divergence relative to overall diversity is conventionally measured by Wright's (1951) F_{ST}, originally defined in terms of the correlation between two random gametes sampled from the same local population, assigning a zero correlation to two gametes sampled randomly from the species as a whole. In the context of genetic diversity measures, at least three definitions of F_{ST} have been proposed. The first is that suggested by Weir and Cockerham (1984), Lynch and Crease (1990), and Hudson, Slatkin, and Maddison (1992):

$$F_{ST} = \frac{\pi_D}{\pi_S + \pi_D}. \qquad (3a)$$

The second is that proposed by Slatkin (1993):

$$F_{ST} = \frac{\pi_D}{\pi_S + \pi_B}$$

$$= \frac{\pi_D}{2\pi_S + \pi_D}. \qquad (3b)$$

The third is that of Hudson, Boos, and Kaplan (1992), which is equivalent to Nei's (1973) G_{ST}:

$$F_{ST} = \frac{\pi_{T-S}}{\pi_T}$$

$$= \frac{2\left(\sum_{i \neq j} w_i w_j\right)\pi_D}{\pi_S + 2\left(\sum_{i \neq j} w_i w_j\right)\pi_D}. \qquad (3c)$$

If the weights for each population are not grossly unequal, we can write

$$w_i = \frac{1}{n} + \epsilon_i,$$

where ϵ_i is of order $1/n$. Under this condition, it is easily seen that as n increases, definitions (3a) and (3c) converge to the same value.

Definition (3a) seems to be the most natural definition of the between-population component of diversity as a proportion of the sum of the within- and between-population components of diversity, although π_S and π_D do not sum to π_T, as noted by Nei (1973). For the case of two populations, definitions (3b) and (3c) are equivalent, but definition (3b) does not converge to a natural definition of the relative amount of between-population diversity as n increases. However, in the context in which it was originally proposed, it was applied to a sample of two populations from a species with a large value of n. In this case, equation (3b) provides an estimator of F_{ST} as defined by equation (3c) (Slatkin 1993).

Using standard results on $\bar{t}_0$ and $\bar{t}_1$ (Slatkin 1991, 1993), a neutral model of coalescence in an island model, with a species composed of n populations of equal breeding size N and migration rate m such that Nm is of order 1, gives the theoretical prediction for equation (3a) as

$$F_{ST} = \cfrac{1}{4Nm\cfrac{n}{(n-1)} + 1}. \qquad (4a)$$

Definition (3b) gives

$$F_{ST} = \cfrac{1}{8Nm\cfrac{n}{(n-1)} + 1}, \qquad (4b)$$

and definition (3c) gives

$$F_{ST} = \cfrac{1}{4Nm\cfrac{n^2}{(n-1)^2} + 1}. \qquad (4c)$$

Equations (4a) and (4c) converge on Wright's (1951) classic result for the island model as the number of populations in the species becomes very large, but, in general, the three equations yield different relations between F_{ST} and Nm. In most applications, lack of knowledge of the true patterns of migration among populations has led investigators to assume an island model with large n. This is not necessarily appropriate. For instance, an isolation-by-distance model may be more appropriate (Slatkin 1993), or the species may be divided into two partially isolated populations, within which there is effectively no local differentiation (e.g., Ford and Aquadro 1996). Even if the assumption of an island model with large n is justified, the different estimators of F_{ST} depend in different ways on the sample sizes and numbers of populations sampled (Cockerham and Weir 1993). Care must therefore be used regarding the estimation of F_{ST} if it is to be used to estimate the scaled migration parameter $M = Nm$, as is frequently done.

Results

I now discuss the interpretation of these measures in relation to the *Drosophila* studies mentioned above, and the utility of relative versus absolute measures of divergence between populations.

Divergence Between *Drosophila* Populations in Regions of Low Recombination

The six-cutter restriction map data of Stephan and Langley (1989) and Stephan and Mitchell (1992) on divergence between an Indian and Burmese pair of populations of *Drosophila ananassae* were reanalyzed by Stephan (1994). For the *f* locus, which is in a region of normal recombination on the X chromosome, he estimated that $Nm = 0.72$ by equating F_{ST} to the value given by equation (4a) with infinite n (W. Stephan, personal communication). This corresponds to an F_{ST} of 0.26. The estimate of π_S (with equal weights for each population) is 0.0072. From equation (3a), we thus have $\pi_D = \pi_S F_{ST}/(1 - F_{ST}) = 0.0025$. For the *fw* locus, which is in a centromeric region with reduced recombination, there is no variation within either population, and there are two fixed restriction site differences between populations (i.e., $F_{ST} = 1$). Using the method of Nei and Tajima (1983) for converting restriction site dif-

ferences into nucleotide site differences and taking into account the fact that 43 restriction sites were used in this study of *fw*, we have $\pi_B = \pi_D = 0.0040$. This is only 1.6 times the estimate of π_D for *f* and is unlikely to be statistically significantly different, given the high sampling and stochastic evolutionary variances of these statistics (Lynch and Crease 1990; Wakeley 1996). The corresponding ratio of F_{ST} values is 3.8. A large component of the difference in F_{ST}, if not all of it, is therefore attributable to the different levels of within-population variability.

Begun and Aquadro (1993) present similar four-cutter data on three X-chromosomal loci of *Drosophila melanogaster* in regions of low recombination (*y, ac,* and *su(f)*) and four loci in regions of normal recombination (*Pgd, G-6pd, v,* and *w*), contrasting U.S. and Zimbabwe populations. Their data show that the mean values of π_S for the low- and high-recombination genes are 0.0010 and 0.0043, respectively, consistent with the well-known reduction in within-population variation in regions of low recombination (Aquadro, Begun, and Kindahl 1994). The corresponding mean values of F_{ST} (calculated using eq. 3a) are 0.57 and 0.29, respectively. Recalculation of the π_D values for each locus gives the result that the mean π_D for the low-recombination genes is 0.0014, and the value for the high-recombination genes is 0.0018. The difference in F_{ST} between the two sets of loci is in the opposite direction to the difference between the estimates of π_D, and is due entirely to differences in within-population diversity values.

Divergence Between Populations of *D. athabasca* at X-Linked and Autosomal Loci

Ford and Aquadro (1996) report consistently higher F_{ST} values for comparisons of pairs of semispecies of *D. athabasca* at X-linked loci than at autosomal loci (means of 0.74 and 0.20, respectively). In this case, there is a clear difference between the mean π_D values for the X chromosome and the autosomes: 0.0048 and 0.00150, respectively (a ratio of 3.2, compared with 3.9 for F_{ST}). This suggests that there may well be a real difference between the X-linked and autosomal loci, which is not simply attributable to the reduced within-population variation for the X chromosome loci (mean $\pi_S = 0.0017$ and 0.0058 for X-linked and autosomal loci, respectively). This is consistent with the results of the statistical test (the DF_{ST}) of the differences in F_{ST} among X-linked and autosomal loci performed by Ford and Aquadro (1996). It seems likely that tests of this nature may more efficiently be carried out using π_D to derive test statistics than F_{ST}, since π_D does not involve a ratio. This would require sequence comparisons with closely related species to correct for differences in mutation rates among loci on the lines of the HKA test (Hudson, Kreitman, and Aguadé 1987).

Possible Causes of Differences in Levels of Divergence

I now turn to the question of how these patterns may be explained. In cases where the differences in F_{ST} are accounted for entirely by differences in within-pop-

ulation diversity values between loci, the problem simply reduces to explaining how variation within local populations can be reduced without simultaneously reducing variation within the whole set of populations. In the *D. ananassae* case, Stephan (1994) argued that there is a low probability of obtaining fixed differences at *fw* on a purely neutral model with $Nm = 0.72$, based on consideration of the probability that a pair of sampled genes from the same population have a most recent common ancestor within that population. On this basis, he concluded that local selective sweeps must have removed variation within each population and caused fixations of different variants in the two populations. While this argument is, of course, correct, its interpretation must take into account the possibility that a gene in a region of reduced recombination is likely to experience reduced effective population size N_e as a result of selection against deleterious mutations (background selection) as well as selective sweeps (Charlesworth, Morgan, and Charlesworth 1993). Fluctuating selection pressures within populations may also contribute to reduced variability in regions of reduced recombination (Gillespie 1994; Barton 1995).

Such effects of selection at linked loci will also cause a reduction in coalescence times within local populations, whereas they have little or no effect on the excess time required for the coalescence of genes sampled from different populations; i.e., π_{T-S} or π_D is expected to be left unaffected by such reductions in N_e, whereas π_S is multiplied by a factor $\alpha < 1$, whose magnitude depends on the processes involved (Charlesworth, Nordborg, and Charlesworth 1997; Nordborg 1997). N in equations (4) is therefore replaced by αN. This obviously leads to an inflation of F_{ST}, purely as a result of reduced within-population variation. Clearly, if there is no variation within samples from a population because of low N_e, any differences between samples from different populations must be fixed differences, but this does not imply that an agent such as local selective sweeps has caused an increase in the degree of population differentiation. There is thus no need to appeal to local selective sweeps to explain the *D. ananassae* case or the *D. melanogaster* data of Begun and Aquadro (1993, 1995), although, of course, they are not ruled out. More extensive data on additional populations suggest, however, that reduced variability in regions of low recombination cannot account for all of the patterns of population differention observed in *D. anannasae* (W. Stephan, personal communication).

On the same principle, at statistical equilibrium under mutation, drift, and a steady rate of selective sweeps (Kaplan, Hudson, and Langley 1989; Stephan, Wiehe, and Lenz 1992), local selective sweeps are not expected to increase the absolute degree of neutral differentiation between populations, since they do not alter the rate of neutral nucleotide site substitutions along a genealogy of given length (Birky and Walsh 1988). A recent pair of local selective sweeps in a pair of populations that initially have normal levels of variability and divergence may, however, cause a transient increase in divergence, analogous to that associated with a population bottleneck (Chakraborty and Nei 1977; Templeton 1980), due to the fixation of different variants that were initially rare within each population, but which come to distinguish them by being hitchhiked up to high frequencies. The maximum expected level of between-population diversity that can be achieved by this process is obviously equal to π_S. This is the scenario favored by Ford and Aquadro (1996) to explain their data on *D. athabasca*. It is interesting to note that their mean π_D for sex-linked loci is similar to three quarters of the mean π_S for autosomal loci (0.0048 vs. 0.0043), as might be expected if variation at the sex-linked loci were originally at the level expected for a population with a 1:1 breeding sex ratio.

Another possibility, suggested by Hilton, Kliman, and Hey (1994), is that selection against variants introduced from one population into another may inhibit population differentiation at linked neutral sites. This type of scenario is, in fact, the only way in which a steady-state local selective sweep process could produce enhanced values of π_D or π_{T-S} at statistical equilibrium and raise the value of π_D over the equilibrium level of π_S for loci not subject to repeated sweeps. Migration of the newly fixed allele into a new population is prevented by its selective disadvantage, thereby reducing gene flow at linked sites compared to regions which are not subject to sweeps. It is not necessary to appeal to repeated sweeps to produce an elevation of π_D, since a locus that maintains constant allelic differences because of differing selection pressures in local populations or that experiences heterozygote inferiority (Barton 1979; Bengtsson 1985; Barton and Bengtsson 1986) is sufficient to produce a large increase in π_D at closely linked neutral sites (Charlesworth, Nordborg, and Charlesworth 1997). Of course, an independent explanation of reduced variation, such as background selection, is then needed to explain reduced variation. In the case of *D. athabasca,* it is possible that fitness disadvantages of this kind associated either with X-linked genes or the X-linked inversions that distinguish the populations might play a role of this kind (Ford and Aquadro 1996). These effects of local selection at linked loci mean that estimates of Nm are strongly biased; such bias is morely likely to be encountered in genomic regions with restricted recombination or in highly inbreeding species (Charlesworth, Nordborg, and Charlesworth 1997).

One final possibility that should be considered is that many of the molecular variants under study are not neutral but are subject to weak counterselection. Under these conditions, a sufficiently large reduction in N_e due to selection at linked sites might cause an acceleration of the rates of fixation of variants within different local populations, as well as a reduction in variation within populations (Birky and Walsh 1988; Charlesworth 1994). The increased number of amino acid substitutions observed between species in the *melanogaster* subgroup in two genes located in regions of reduced recombination is consistent with this scenario (Hilton, Kliman, and Hey 1994). While this scenario has not yet been explored quantitatively in the context of a set of populations connected by gene flow, it nevertheless seems

worth considering as a potential contributor to enhanced values of π_D associated with reductions in π_S.

The Use of Absolute and Relative Measures of Divergence

Relative measures of between-population divergence, such as F_{ST}, are inherently dependent on the extent of within-population diversity (Nei 1973, 1987, p. 190). Indeed, for loci with very high levels of diversity such as microsatellites, F_{ST} is a poor measure of between-population divergence even in the absence of forces that affect diversity, since F_{ST} is necessarily low even if absolute divergence is high (T. Nagylaki, personal communication). As shown above, factors which affect within-population diversity at otherwise comparable loci can cause substantial among-locus differences in these relative measures in the absence of any differences in absolute measures of between-population diversity. A similar problem exists when comparing species with different levels of inbreeding; inbreeding can reduce within-population diversity, and hence inflate F_{ST}, even if absolute levels of between-population divergence are similar (Charlesworth, Nordborg, and Charlesworth 1997). The use of absolute measures of divergence is thus necessary when comparing genomic regions with different levels of recombination or species with different breeding systems. As mentioned above, tests of significance will then require adjustment for possible differences in mutation rates, by the use of comparisons with related species (Hudson, Kreitman, and Aguadé 1987).

Acknowledgments

LITERATURE CITED

AQUADRO, C. F., D. J. BEGUN, and E. C. KINDAHL. 1994. Selection, recombination, and DNA polymorphism in *Drosophila*. Pp. 46–56 *in* B. GOLDING, ed. Non-neutral evolution: theories and molecular data. Chapman and Hall, London.

BARTON, N. H. 1979. Gene flow past a cline. Heredity **43**:333–339.

———. 1995. Linkage and the limits to natural selection. Genetics **140**:821–841.

BARTON, N. H., and B. O. BENGTSSON. 1986. The barrier to genetic exchange between hybridising populations. Heredity **56**:357–376.

BEGUN, D. J., and C. F. AQUADRO. 1993. African and North American populations of Drosophila melanogaster are very different at the DNA level. Nature **365**:548–550.

———. 1995. Evolution at the tip and base of the X chromosome in an African population of *Drosophila melanogaster*. Mol. Biol. Evol. **12**:382–390.

BENGTSSON, B. O. 1985. The flow of genes through a genetic barrier. Pp. 31–42 *in* P. J. GREENWOOD, P. H. HARVEY, and M. SLATKIN, eds. Evolution. Essays in honour of John Maynard Smith. Cambridge University Press, Cambridge, England.

BIRKY, C. W., and J. B. WALSH. 1988. Effects of linkage on rates of molecular evolution. Proc. Natl. Acad. Sci. USA **85**:6414–6418.

CHAKRABORTY, R., and M. NEI. 1977. Bottleneck effects on average heterozygosity and genetic distance with the stepwise mutation model. Evolution **31**:347–356.

CHARLESWORTH, B. 1994. The effect of background selection against deleterious alleles on weakly selected, linked variants. Genet. Res. **63**:213–228.

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

CHARLESWORTH, B., M. NORDBORG, and D. CHARLESWORTH. 1997. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genet. Res. **70**:155–174.

COCKERHAM, C. C., and B. S. WEIR. 1993. Estimation of gene flow from *F*-statistics. Evolution **47**:855–863.

FORD, M. J., and C. F. AQUADRO. 1996. Selection on X-linked genes during speciation in the *Drosophila athabasca* complex. Genetics **144**:689–703.

GILLESPIE, J. H. 1994. Alternatives to the neutral theory. Pp. 1–17 *in* B. GOLDING, ed. Non-neutral evolution: theories and molecular data. Chapman and Hall, London.

HILTON, H., R. M. KLIMAN, and J. HEY. 1994. Using hitchhiking genes to study adaptation and divergence during speciation within the *Drosophila melanogaster* species group. Evolution **48**:1900–1913.

HOLSINGER, K. E., and R. J. MASON-GAMER. 1996. Hierarchical analysis of nucleotide diversity in geographically structured populations. Genetics **142**:629–639.

HUDSON, R. R. 1990. Gene genealogies and the coalescent process. Oxf. Surv. Evol. Biol. **7**:1–45.

HUDSON, R. R., D. D. BOOS, and N. L. KAPLAN. 1992. A statistical test for detecting geographic subdivision. Mol. Biol. Evol. **9**:138–151.

HUDSON, R. R., M. KREITMAN, and M. AGUADÉ. 1987. A test of neutral molecular evolution based on nucleotide data. Genetics **116**:153–159.

HUDSON, R. R., M. SLATKIN, and W. P. MADDISON. 1992. Estimation of levels of gene flow from population data. Genetics **132**:583–589.

KAPLAN, N. L., R. R. HUDSON, and C. H. LANGLEY. 1989. The "hitchhiking" effect revisited. Genetics **123**:887–899.

KIMURA, M. 1971. Theoretical foundations of population genetics at the molecular level. Theor. Popul. Biol. **2**:174–208.

LYNCH, M., and T. J. CREASE. 1990. The analysis of population survey data on DNA sequence variation. Mol. Biol. Evol. **7**:377–394.

NEI, M. 1973. Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. USA **70**:3321–3323.

———. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

NEI, M., and F. TAJIMA. 1983. DNA polymorphism detectable by restriction endonucleases. Genetics **97**:145–163.

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

SLATKIN, M. 1991. Inbreeding coefficients and coalescence times. Genet. Res. **58**:167–175.

———. 1993. Isolation by distance in equilibrium and non-equilibrium populations. Evolution **47**:264–279.

STEPHAN, W. 1994. Nucleotide sequence variation in *Drosophila ananassae*. Pp. 57–66 *in* B. GOLDING, ed. Non-neutral

evolution: theories and molecular data. Chapman and Hall, London.

STEPHAN, W., and C. H. LANGLEY. 1989. Molecular genetic variation in the centromeric region of the X chromosome in three *Drosophila ananassae* populations. I. Contrasts between the *vermilion* and *forked* loci. Genetics **121**:89–99.

STEPHAN, W., and S. J. MITCHELL. 1992. Reduced levels of DNA polymorphism and fixed between-population differences in the centromeric region of *Drosophila ananassae*. Genetics **132**:1039–1045.

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.

TEMPLETON, A. R. 1980. Modes of speciation and inferences based on genetic distances. Evolution **34**:719–729.

WAKELEY, J. 1996. The variance of pairwise differences in two populations with migration. Theor. Popul. Biol. **49**:39–57.

WEIR, B. S., and C. C. COCKERHAM. 1984. Estimating *F*-statistics for the analysis of population structure. Evolution **38**: 1358–1370.

WRIGHT, S. 1951. The genetical structure of populations. Ann. Eugen. **15**:323–354.