Evaluating loci for use in the genetic analysis of population structure

MARK A. BEAUMONT[1] AND RICHARD A. NICHOLS[2]

[1] *Conservation Genetics Group, Institute of Zoology, Regent's Park, London NW1 4RY, U.K.*
[2] *School of Biological Sciences, Queen Mary & Westfield College, Mile End Road, London E1 4NS, U.K.*

SUMMARY

Loci that show unusually low or high levels of genetic differentiation are often assumed to be subject to natural selection. We propose a method for the identification of loci showing such disparities. The differentiation can be quantified using the statistic F_{ST}. For a range of population structures and demographic histories, the distribution of F_{ST} is strongly related to the heterozygosity at a locus.

Outlying values of F_{ST} can be identified in a plot of F_{ST} *vs.* heterozygosity using a null distribution generated by a simple genetic model. We use published data-sets to illustrate the importance of the relationship with heterozygosity. We investigate a number of models of population structure, and demonstrate that the null distribution is robust to a wide range of conditions. In particular, the distribution is robust to differing mutation rates, and therefore different molecular markers, such as allozymes, restriction fragment length polymorphisms (RFLPs) and single strand conformation polymorphisms (SSCPs) can be compared together. We suggest that genetic variation at a discrepant locus, identified under these conditions, is likely to have been influenced by natural selection, either acting on the locus itself or at a closely linked locus.

1. INTRODUCTION

Differences in gene frequency are widely used to draw inferences about those aspects of a species' biology that are not amenable to direct study. Such studies can provide information ranging from migration rates (see Ward *et al.* 1992 for a comparative review) to the history of range expansion (see, for example, Boileau *et al.* 1992). For these purposes, the genetic differentiation between populations is commonly characterized by statistics related to Wright's inbreeding coefficient F_{ST} (Wright 1951).

The ready interpretation of the genetic patterns identified using these techniques can be disrupted by balancing or spatial selection, which can given rise to loci showing atypically low or high levels of differentiation. This paper outlines a method that is able to detect such loci.

The key insight that makes this problem amenable to analysis is that the level of inbreeding experienced at different neutral loci should be the same because of their shared demographic history. Lewontin & Krakauer (1973) proposed a test of neutrality based on this idea. They suggested that the variance of F_{ST} values among loci should be $2\overline{F_{ST}}^2/(n-1)$, where n is the number of subpopulations, and the observed variance could be tested against this using standard variance-ratio tests. This test has been criticized, particularly by Robertson (1975) and Nei & Maruyama (1975), who found that the value of the expected variance given by Lewontin & Krakauer was generally too low, and sensitive to correlations in F_{ST} values among populations.

More recently, Bowcock *et al.* (1991) have considered a number of refinements to the Lewontin–Krakauer test. They treated a large number of restriction fragment length polymorphisms (RFLPs) as independent bi-allelic loci and used simulations to generate the expected distribution of F_{ST} at a range of ancestral allele frequencies. By using this procedure they detected a number of anomalous alleles.

The approach that we have taken is somewhat different to that of Bowcock *et al.* We explicitly consider unlinked or loosely linked regions of DNA that correspond to the classical locus, and which can be validly treated as independent units. Such markers would include RFLP haplotypes in a particular region, allozymes, and single strand conformation polymorphisms (SSCPs). These loci are generally multi-allelic, and we therefore consider the relationship of F_{ST} to heterozygosity rather than gene frequency.

We will first examine the observed distribution of F_{ST} as a function of heterozygosity in two different sets of data. These distributions are compared with that expected in the simplest model of a subdivided population (the symmetrical Island Model, Wright 1951). We investigate how far the expected distribution depends on the assumptions of the model by comparison with a variety of different equilibrium and non-equilibrium models.

(*a*) *Applications*

In this section we first describe the methodology and then the analysis of two published data sets. Tests of significance can be formulated, although they only

apply to the particular model used to generate the distribution of F_{ST}. The analysis is better viewed as a guide to the exploration of the data. With this approach, we argue that we can extend the conclusions of the original authors.

The procedure is applied as follows.

1. F_{ST} is estimated by the statistic $\hat{\beta}$ (Cockerham & Weir 1993)

$$\hat{\beta} = \frac{\hat{f}_0 - \hat{f}_1}{1 - \hat{f}_1}, \quad (1)$$

where $1-\hat{f}_1$ is the average pairwise difference between all possible pairs of genes between each sample, and $1-\hat{f}_0$ is the average pairwise difference between all possible pairs of genes within each sample. Heterozygosity is estimated as $1-\hat{f}_1$.

2. The expected F_{ST} is calculated from the data as the average among loci weighted by their heterozygosity (i.e. $1-\hat{f}_1$), as suggested by Weir & Cockerham (1984). Simulations suggest that this provides an accurate estimator of expected F_{ST}.

3. Coalescent simulations are performed using the Island Model with 100 islands (see, for example, Strobeck 1987). An infinite alleles mutational model is used. Samples, of the same size and number as the data, are simulated, where each sample is taken from a different island. In these models, the population size of each island, N, and the mutation rate, μ, need not be specified separately; genetic variation within and among islands depends only on the scaled mutation rate θ ($= N\mu$, see Appendix). As discussed later in the text, two scaled mutation rates are used to generate loci with a wide range of heterozygosities. F_{ST} and heterozygosity are calculated, and this procedure is repeated a minimum of 5000 times.

4. The distribution of F_{ST} as a function of $1-\hat{f}_1$ is characterized by estimating the quantiles of the distribution. In the examples given below, we choose the 0.025, 0.5 and 0.975 quantiles; the outer two define an envelope in which 95% of the data points are expected to lie.

(*i*) Drosophila *allozyme loci*

Singh & Rhomberg (1987) described the results of a survey of 117 loci in *Drosophila melanogaster* drawn from 15 populations around the world. A total of 61 loci were polymorphic and F_{ST} values were calculated from them. The histogram of F_{ST} values appeared to have a main mode around 0.1. They inferred that loci symmetrically distributed around the mode, about 2/3 of the polymorphic sample, were neutral and reflected an underlying migration rate of two migrants per generation. They suggested that the remaining 20 loci were experiencing selection. In particular, eight loci seemed unusually highly differentiated.

The gene frequency data on which the study was based was kindly provided to us by Professor R. S. Singh. We recalculated heterozygosity and F_{ST} by the methods given above (this results in only very small changes to values published by Singh & Rhomberg). In our preliminary analysis the weighed average of the F_{ST} values was 0.23. We assume that the loci are sufficiently loosely linked that the gene frequencies at each locus can be regarded as independent. The density calculated using this value gave a poor fit to the data, with more points above and below the 0.95 limits than might be expected by chance. Inspection of the data suggested that the poor fit was influenced strongly by eight loci that had high F_{ST} values. With these loci removed, the weighted average F_{ST} was 0.17. Figure 1 plots the expected distribution with this average.

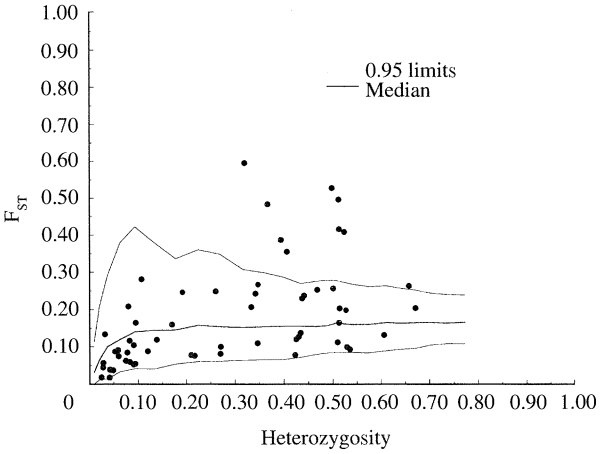

Figure 1. F_{ST} values estimated from 61 allozyme loci in *Drosophila melanogaster* plotted against heterozygosity. The quantiles estimated from an island model with expected $F_{ST} = 0.17$ are also plotted.

It can be seen that the majority of points fall within the 0.95 limits. With $61-8=53$ points we would expect three to lie outside by chance, and in this case one lies just outside. The further simulations, described later, could not find plausible alternative genetic models which generated distributions to include the eight outlying points. Indeed, they lie quite apart from the main body of the distribution. We conclude that genetic variation at these eight loci has been affected by natural selection; most probably geographically varying selection. Although these eight loci were clearly picked out by Singh & Rhomberg, they suggested that a further 12 loci may be affected by selection, for which this analysis finds little support.

(*ii*) *Atlantic cod allozyme and RFLP loci*

Pogson *et al.* (1995) describe the results of a survey of ten allozyme loci and 11 nuclear RFLP loci in six populations of the Atlantic Cod. They sampled approximately 100 fish from each population. They found that the mean F_{ST} value of the RFLP loci was substantially higher than that of the allozyme loci. They suggested that the difference was caused by balancing selection on the allozyme loci, which prevented genetic differentiation.

The RFLP and allozyme data should both conform to an infinite allele model, and therefore there should be no *a priori* reason for expecting the distribution of F_{ST} in the two types of marker to be different from one another. We therefore analyse the data together, as described below.

The paper used Nei's G_{ST} as an estimator of F_{ST}. The data were recalculated using the formulae of Cocker-

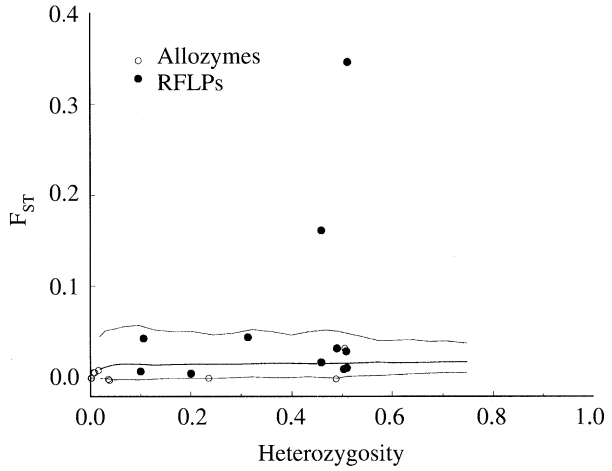

Figure 2. F_{ST} values estimated from ten allozyme and 11 RFLP loci in the Atlantic Cod are plotted against heterozygosity. The quantiles estimated from an island model with expected $F_{ST} = 0.018$ are also plotted.

ham & Weir (1993) to obtain F_{ST} as defined here. In our analysis, the weighted average F_{ST} was 0.06, and, with this value, the fit was generally very poor with many more points above and below the 0.95 limits than would be expected by chance. Inspection of the data suggested that two of the RFLP loci had values of F_{ST} which were substantially greater than might be expected under the model. With these two loci discarded, the average F_{ST} was 0.018.

Figure 2 shows the Atlantic Cod data with the simulated 0.95 limits, using an expected F_{ST} of 0.018, and sample sizes of 200 haploids. It can be seen that, with the exception of the two extreme RFLP outliers, the data fit the model fairly well. A total of four of the allozyme loci lie very close to the origin with heterozygosities that are too low to carry any meaningful information about the expected F_{ST}. The assumption that the distribution of F_{ST} is continuous becomes increasingly untenable with very low heterozygosities, and we therefore discarded loci with heterozygosities less than 2.0/(sample size). The remaining six allozyme loci, with heterozygosities greater than 0.01 have the following estimated two-tailed p-values (in order of increasing heterozygosities): 0.902, 0.080, 0.048, 0.068, 0.030, 0.288. We use Fisher's method to obtain an overall probability of this sample as $\chi^2_{12} = 26.2$ ($p \sim 0.01$). This is only an approximate calculation because the null distribution is estimated from the 5000 sample points, and also the calculation ignores the sampling error in the estimate of expected F_{ST}. Although many of the allozyme $\hat{\beta}$ values are close to zero, it should be noted that $\hat{\beta}$ may take negative values, and the tails of the null distribution are not compressed by the origin. Thus, even having discarded the two outlying RFLP loci, there remains evidence that some of the allozyme loci have unusually low F_{ST} values. Low values of F_{ST} could arise from balancing selection, which keeps alleles at similar frequencies in different populations.

The strongest evidence of selection comes from the two RFLP outliers, each with p-values outside the range of reliable extrapolation. These are in a region where it would be difficult to explain them from any demographic or sampling arguments. The most likely explanation is that geographically varying selection is occurring close to these loci.

2. TESTS OF ROBUSTNESS

We have examined the effects of violations in the assumptions used to generate the expected distributions in the previous examples. Specifically, we investigate the following.

1. The effect of different mutation rates on the expected distribution of F_{ST}.
2. The effect of sample size on this distribution.
3. The effect of equilibrium versus non-equilibrium population structure.
4. The effect of heterogeneity in migration rates or levels of inbreeding among populations.
5. The effect of geographic isolation-by-distance and clumping of sampled populations.

(*a*) *Technical details*

Information on simulation methods and statistical methods is given in the appendix. Below, we give details of specific parameters used in the simulations, the results of which are described in the subsequent section.

(*i*) *Mutation rate*

To investigate the effect of mutation rate on the distribution of F_{ST}, $d = 15$ samples of $n = 50$ haploid individuals were simulated from the Island Model. We estimated the distribution for mutation rates $\theta = 0.01, 0.1, 1.0$ for the higher mutation rate a further 5000 simulations were carried out to generate 5000 sets at the lower heterozygosities (< 0.3), for comparison with the other mutation rates. The migration rate was set at $M = 2.5$ which gives an expected F_{ST} of ~ 0.17 (comparable to the *Drosophila* data described above).

For the remaining results we use simulations in which half the points are generated with $\theta = 0.1$ and half with $\theta = 1.0$ to cover a wide range of heterozygosities.

(*ii*) *Sample size*

To investigate the effect of sample size on the distribution of F_{ST}, we carried out simulations with the parameters given in the previous section, but where the sample size, n, varied from 10 to 100 haploids.

(*iii*) *Equilibrium and non-equilibrium populations*

A model of colonization was used to simulate samples from newly founded populations. Populations are founded simultaneously from a single ancestral population, and then grow exponentially in size. The model is discussed in more detail in Nichols & Beaumont (1996), where the effects of migration from the ancestral population are also considered. In our treatment here, we ignore migration.

The model can be regarded as a special case of the range expansion model of Slatkin (1993), but we also

explicitly consider populations that grow in size. Provided the mutation rate is low, the gene frequency distributions should not be affected by population growth (Griffiths & Tavaré 1994). Interestingly, in exponentially growing populations, the expected value of F_{ST} becomes independent of time. In reality populations do not grow indefinitely, but the time-independent value of F_{ST} is a useful marker-point for the phase transition between the initial dynamics, when the change in F_{ST} depends on population growth rate, and the later dynamics when it depends on population size.

The Appendix gives further details. The parameters used were: growth rate, $r = 0.133$; initial number of founders, $N_0 = 47$; sample size, $n = 50$. This gives an expected F_{ST} of ~ 0.17. Also plotted for comparison is the distribution in the island case, with $n = 50$. All other parameters were the same for both models and are described above. The expected values were calculated directly from recursions given in Nichols & Beaumont (1996), but can also be approximated by

$$F_{ST} = 1 - e^{-\frac{1}{rN_0}},$$

which is accurate for small r.

Because this is a pure drift model, for any N_0 and r we can specify an equivalent stable population of populations size N, sampled at time t, such that $\frac{t}{N} = \frac{1}{rN_0}$. The gene frequency distributions, and hence the distribution and expected value of F_{ST} will be approximately the same.

(iv) Heterogeneous subpopulations

In reality parameters will vary among subpopulations. It is impossible to consider all possible combinations; we therefore restrict ourselves to presenting the extreme case where the expected value of F_{ST} is very different between two groups of subpopulations.

In the island case, 20% of the subpopulations had $M_1 = 0.25$ immigrants per generation and 80% had $M_2 = 10$ immigrants per generation. For the 15 sampled islands, the corresponding numbers are 3 and 12 respectively. In the case of a large number of islands and small mutation rate it can be shown that

$$F_{ST} = \frac{p_1}{1+2M_1} + \frac{p_2}{1+2M_2} = p_1 F^1_{ST} + p_2 F^2_{ST}, \qquad (2)$$

where p_i is the proportion of sampled subpopulations with M_i immigrants per generation and F^i_{ST} is the corresponding homogeneous infinite island F_{ST} with that value of M_i. This result is independent of the migration rates among the unsampled populations (see appendix). With the migration rates given here, this gives $F^1_{ST} = 0.67$, $F^2_{ST} = 0.048$, and an expected F_{ST} of ~ 0.17, and thus the distribution can be compared with the previous examples.

Similarly, in the colonization case $F_{ST} = p_1 F^1_{ST} + p_2 F^2_{ST}$, where F^i_{ST} is the value expected from the set of islands with parameters M_i, r_i, N_{0i}. In the colonization example we chose to vary only the initial number of founders: $N_{01} = 8$, $N_{02} = 175$, $M_1 = M_2 = 0$, $r_1 = r_2 = 0.133$. This gives $F^1_{ST} = 0.67$, $F^2_{ST} = 0.048$ and overall $F_{ST} = \sim 0.17$, as before. Each set of limits was estimated using 15000 simulation points, half of which had a higher mutation rate as described earlier.

(v) Stepping-stone model

We investigated the effects of geographic subdivision on the distribution of F_{ST} by running coalescent-based simulations of the stepping-stone model on a two-dimensional torus. We used a 24×24 torus. The mutational model was the same as in the previous sections with $\theta = 0.1$ and 1. Immigrants had an equal chance of being drawn from any of the four neighbouring squares. We sampled $n = 50$ haploid individuals from 16 of the 576 possible populations. The details of the stepping stone simulations are very similar to those of the island model (migrants are drawn from neighbouring cells rather than at random from all cells). The general method of simulation is given in Slatkin (1993).

We considered three different arrangements of the 16 populations, which gave different observed patterns of isolation-by-distance: (*a*) an even distribution where the 16 populations were separated by five unsampled cells in each of the four directions; (*b*) a clustered distribution (hereinafter denoted 'Cluster 1') where the populations were arranged in four groups of four adjacent cells, spaced by ten unsampled cells in each direction; (*c*) a clustered distribution ('Cluster 2') where 13 cells were arranged as a $1+3+5+3+1$ diamond and the remaining three cells arranged to be maximally apart from each other and the cluster. For these three models we used $M = 5.56$, 5.26 and 4.35 respectively to obtain mean $F_{ST} \sim 0.17$. For comparison, we also simulated 16 samples from the Island model with $d = 576$ islands. For all four models we simulated 10000 samples.

3. RESULTS

(*a*) *Mutation rate*

In figure 3*a* it can be seen that the distribution of F_{ST} depends quite strongly on the observed heterozygosity. The median value of F_{ST} drops off rapidly at heterozygosities less than 0.1. With higher mutation rates, at lower heterozygosities, the distribution is tighter and the median drops off faster.

(*b*) *Sample size*

In figure 3*b* it can be seen that when very small samples are taken from each subpopulation, the distributions are broader. Even moderate sample sizes are surprisingly informative; the distribution for $n = 50$ is virtually indistinguishable from that of $n = 100$.

(*c*) *Non-equilibrium populations*

Although only a specific case has been illustrated here (figure 3*c*), extensive simulations with many different parameter values suggest that the island and colonization models produce indistinguishable distributions when F_{ST} is not too high (less than 0.5).

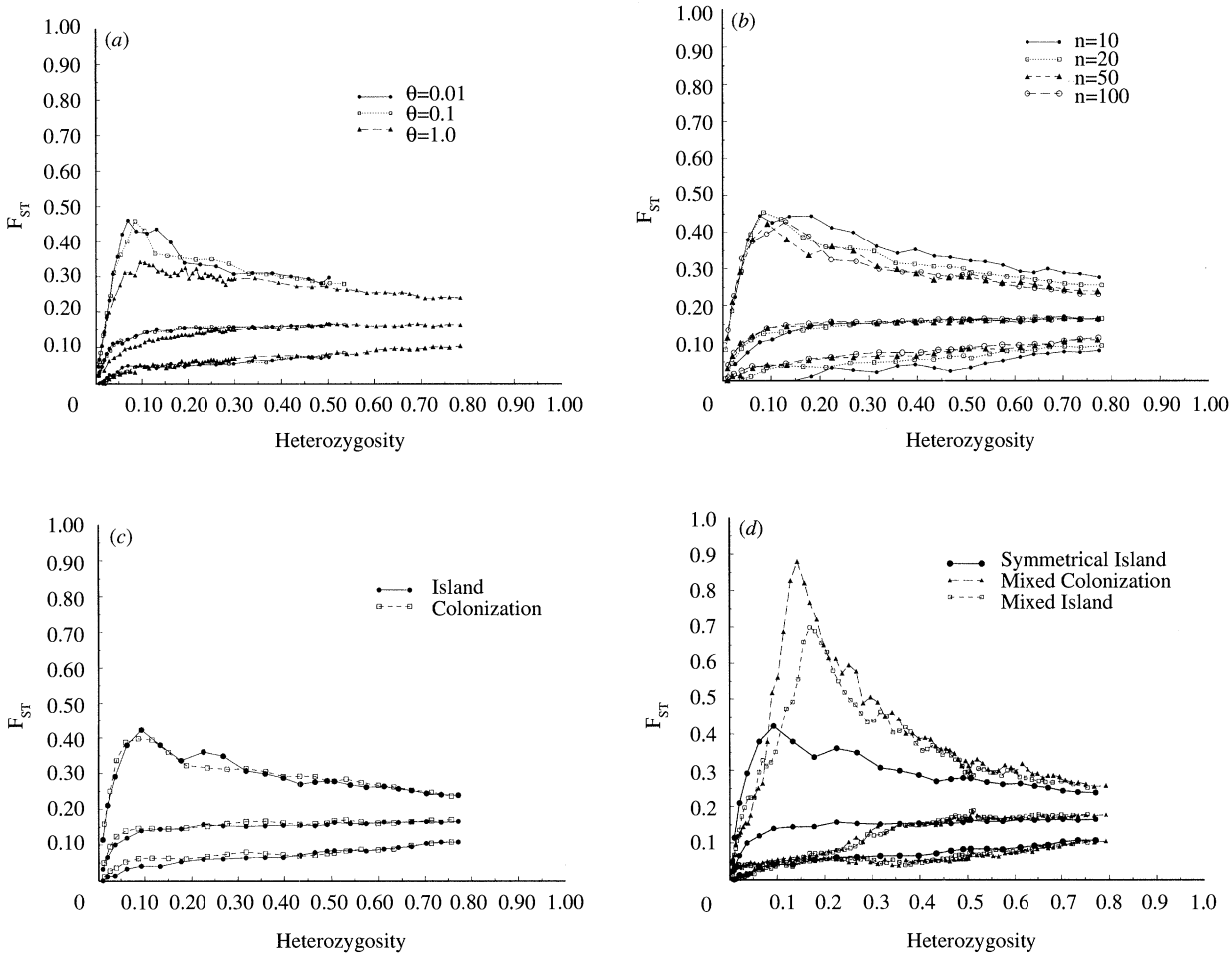

Figure 3. (*a*) The estimated 0.025, 0.5 and 0.975 quantiles of the distribution of F_{ST} against heterozygosity. The quantiles were estimated from simulations from an island model with three different scaled mutation rates. (*b*) The 0.025, 0.5 and 0.975 quantiles estimated from simulations of an island model with four different samples sizes. (*c*) The 0.025, 0.5 and 0.975 quantiles estimated from simulations of an island model and a colonization model. (*d*) The 0.025, 0.5 and 0.975 quantiles estimated from simulations of mixed-parameter island and colonization models, compared with a symmetrical island model.

(*d*) *Heterogeneous subpopulations*

The distribution becomes more skewed in the heterozygosity range 0.1–0.5 for both colonization and for island models (figure 3*d*). An important point to note, however, is that the distribution for heterogeneous models converges to those of the symmetrical island case for observed heterozygosities greater than 0.5.

(*e*) *Stepping-stone model*

In figure 4*a* we plot the pairwise F_{ST} values against cartesian distance for the three models. For comparison we plot the expected values calculated according to the formulae given by Slatkin (1991, 1993). It can be seen that there is evidence of isolation-by-distance in all three models, but is much stronger in the two clustered models where there is a two-fold difference in mean F_{ST} between adjacent pairs in comparison with distant pairs.

When the distribution of F_{ST} against heterozygosity is compared for the three stepping-stone models and the island model, it can be seen that even the quite high level of isolation-by-distance observed in these models does not have a marked effect. Unsurprisingly, the evenly spaced samples, which have the weakest isolation-by-distance appear indistinguishable from the island case. Rather more surprisingly the most clustered model, Cluster 2, with 13 adjacent and three distant samples is also very similar to the island model. Cluster 1 is most dissimilar, with narrower quantile limits at heterozygosities ~ 0.1 and noticeably broader limits at heterozygosities greater than 0.3.

4. DISCUSSION

We set out to investigate the practicality of drawing biological inferences from comparisons of genetic differentiation (as quantified by F_{ST}) at separate loci. This approach must necessarily depend on assumptions about the genetics of the loci in question and on the theoretical distribution of F_{ST}. If the differentiation at each locus is explained by a unique set of processes or if the distribution of F_{ST} is affected by details of population history, the patterns may not be easily deciphered.

The question of consistency among loci can be addressed empirically. In the two examples that we

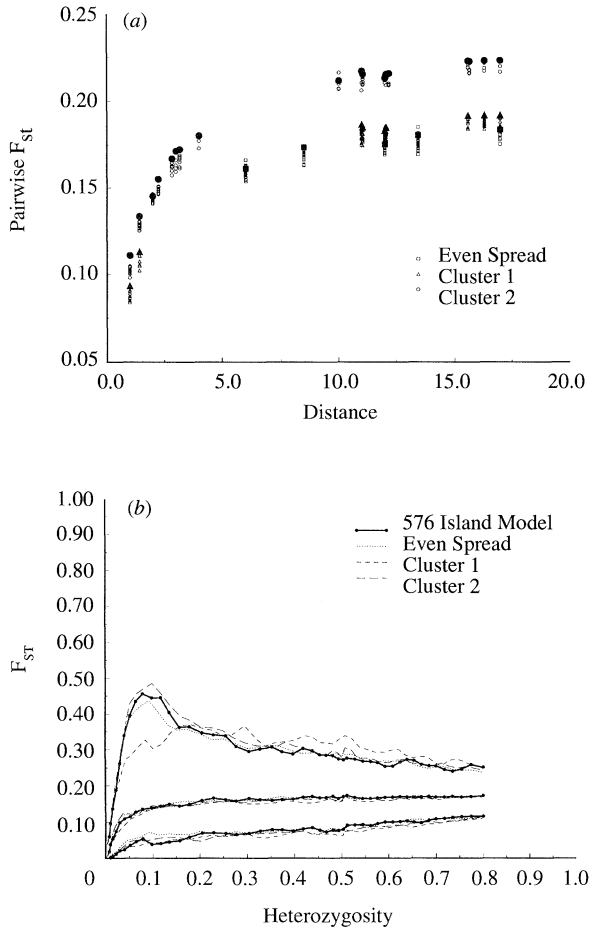

Figure 4. (*a*) Plots of average pairwise F_{ST} against cartesian distance among 16 samples located on a 24×24 torus. A total of three models are compared, and estimates were derived from 5000 simulation in each model. The expected values are denoted by the larger filled symbols. (*b*) The 0.025, 0.5 and 0.975 quantiles estimated from simulations of 16 sampled populations located on a 24×24 torus, and also in a 576-island model. 10000 simulations were done for each model.

presented, a large number did appear to have similar properties and there was a clear disjunction between some outlying loci and this majority. The bulk of the distribution was consistent with that generated by simple models of population structure, which suggests that it requires no complex explanation. The simplest explanation for the outlying loci is that they are subject to selection. The detection of these loci is much enhanced by plotting F_{ST} against heterozygosity (or gene frequency for biallelic data; Bowcock *et al.* 1991). Reassuringly, it would appear that the distribution of F_{ST} as a function of heterozygosity is not strongly affected by the mutation rates typical of many genetic markers.

The result of varying the sample sizes illustrates a point noted previously (see, for example, Slatkin & Barton 1989) that the bulk of the variability of F_{ST} among loci is caused by their independent genealogies and not due to sampling. Thus there appears to be little value in having sample sizes larger than 25 diploids per subpopulation. If one wishes to obtain accurate estimates of F_{ST} it is better to sample more independent loci.

Robertson (1975) along with Nei & Maruyama (1975) rejected Lewontin & Krakauer's attempts to interpret differences in F_{ST} values between loci. Their most fundamental criticism was that the correlation of allele frequencies among populations is itself variable, either because of different degrees of temporal separation among populations (Roberton 1975) or geographic separation (Nei & Maruyama 1975). In the examples they gave, they demonstrated a substantial effect on the variance of F_{ST}. In this paper, we have investigated the effects of variable correlations arising from differences in parameter values among populations and also from geographic differentiation. Although we have been able to demonstrate that these variable correlations can indeed broaden the distribution of F_{ST} among loci, under realistic circumstances we did not induce effects strong enough to undermine the approach. Furthermore, thinning out the data (see below) can effectively ameliorate the broadening.

The strongest effect occurred in the colonization case, where there were two categories of population with a 22-fold difference in the number of founders. Those populations originating from a large number of founders were much less differentiated from each other: the expected F_{ST} is 0.05 as compared with 0.67 for those founded by the smaller number.

Equivalent genetic patterns can arise if a population is fragmented into pieces in two different size categories with negligible gene flow between them. If there is a 22-fold difference in population size, the smaller pieces will diverge faster, and will have reached an F_{ST} of 0.67 when the others have reached an F_{ST} of 0.05. An island model with one category of island 40 times the size of the other (or equivalently 40 times more resistant to immigration) would also show the same pattern of differentiation. Such dramatic differences do occur in vertebrate populations subject to habitat loss and consequent local inbreeding. Despite these large differences in pairwise F_{ST} values it is impressive that all the models considered have similar F_{ST} distributions, especially at high heterozygosities.

Different pairwise F_{ST} values were also generated by different spacings in the stepping stone case. The observed F_{ST} values are close to theoretical values (figure 4*a*) and the model can replicate the results of Slatkin (1993, figure 5*b*). Isolation by distance cannot readily produce the discrepancies in F_{ST} value generated by the previous models: we calculate from Barton's results (appendix in Slatkin 1991) that a torus of side $\sim 10^{27}$ with migration, $M \sim 9.5$ would be required to obtain comparable pairwise values. Our results from a smaller torus showed no discernable effect of clumping samples.

Clearly, it is sensible to avoid sampling from adjacent populations. Likewise it is desirable to avoid sampling repeatedly from populations with correlated allele frequencies such as those derived from large numbers of founders in the colonization model. The 'clusters' need not be geographically adjacent, but can be identified graphically by ordinating the pairwise F_{ST} values. In this case the values are ~ 0.05, 0.36 and 0.67, giving a central cluster and three satellites (not shown).

Mild isolation-by-distance (evident in both the fish and *Drosophila* data) does not appear to affect the method strongly, especially when the number of sampled populations is large. Indeed, larger numbers of samples reduce the variability in F_{ST} among loci, and if taken over a wide geographic area the effects of selection may be more easily detected.

Criticism of the Lewontin–Krakauer test appears to have caused the whole approach to be abandoned; a situation reminiscent of the baby and the bath-water. To rectify this, we suggest the following prescription to identify candidate loci whose allele frequencies may have been affected by natural selection.

1. A large number ($\geqslant 20$) of independent loci should be studied. Preferably, these should have high heterozygosities. Samples (sizes $\geqslant 25$ diploids) should be taken from a large number ($\geqslant 10$) of subpopulations.

2. Values of $\hat{\beta}$ should be calculated for all pairwise comparisons among populations. These should be calculated by weighting each locus by heterozygosity.

3. The resulting distance matrix should be ordinated using either metrical (PCA) or non-metrical (MDS) scaling. Strong clustering can be identified by eye and populations thinned out accordingly. Weak clustering can be ignored.

4. With the remaining populations, do the analysis as applied to the two test data sets described here.

We thank David Balding, Mike Bruford, Kate Byrne, and Michelle Bayes for their comments on the manuscript. This work was supported in part by SERC grant GR/G 11101 to R.A.N. M.A.B. has been supported by a Wellcome Mathematical Biology Training Fellowship, and latterly by the Institute of Zoology. Contact M.Beaumont@ucl.ac.uk or R.A.Nichols@qmw.ac.uk for source code and programs.

REFERENCES

Barbadilla, A., King, L. M. & Lewontin, R. C. 1996 What does electrophoretic variation tell us about protein variation. *Molec. Biol. Evol* **13**, 427–432.

Boileau, M. G., Herbert, P. D. N. & Schwartz, S. S. 1992 Non-equilibrium gene frequency divergence: persistent founder effects in natural populations. *J. Evol. Biol.* **5**, 25–39.

Bowcock, A. M., Kidd, J. R., Mountain, J. L., Hebert, J. M., Carotenuto, L., Kidd, K. K. & Cavalli-Sforza, L. L. 1991 Drift, admixture, and selection in human evolution: a study with DNA polymorphisms. *Proc. natn. Acad. Sci. U.S.A.* **88**, 839–843.

Cockerham, C. C. & Weir, B. S. 1993 Estimation of gene flow from F-statistics. *Evolution* **47**, 855–863.

Crow, J. F. & Aoki, K. 1984 Group selection for a polygenic behavioural trait: estimating the degree of population subdivision. *Proc. natn. Acad. Sci. U.S.A.* **81**, 6073–6077.

Griffiths, R. C. & Tavaré, S. 1994 Sampling theory for neutral alleles in a varying environment. *Phil. Trans. R. Soc. Lond.* B **344**, 403–410.

Hill, I. D., Hill, R. & Holder, R. L. 1976 Fitting Johnson Curves by moments. *Appl. Stat.* **25**, 180–189.

Hudson, R. R. 1990 Gene genealogies and the coalescent process. *Oxf. Surv. Evol. Biol.* **7**, 1–44.

Lewontin, R. C. & Krakauer, J. 1973 Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* **74**, 175–195.

Nei, M. & Maruyama, T. 1975 Lewontin–Krakauer test for neutral genes. *Genetics* **80**, 395.

Nichols, R. A. & Beaumont, M. A. 1996 Is it ancient or modern history that we can read in the genes? In *Aspects of the genesis and maintenance of biological diversity* (ed. M. Hochberg, J. Clobert & R. Barbault), pp. 69–87. Oxford University Press.

Pogson, G. H., Mesa, K. A. & Boutilier, R. G. 1995 Genetic population structure and gene flow in the Atlantic Cod *Gadus morhua*: a comparison of allozyme and nuclear RFLP loci. *Genetics* **139**, 375–385.

Robertson, A. 1975 Remarks on the Lewontin–Krakauer test. *Genetics* **80**, 396.

Singh, R. S. & Rhomberg, L. R. 1987 A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. II. Estimates of heterozygosity and patterns of geographic differentiation. *Genetics* **117**, 255–271.

Slatkin, M. 1993 Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* **47**, 264–279.

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

Slatkin, M. & Barton, N. H. 1989 A comparison of three methods for estimating average levels of gene flow. *Evolution* **43**, 1349–1368.

Strobeck, C. 1987 Average number of nucleotide differences in a sample from a single subpopulation: a test for population subdivision. *Genetics* **117**, 149–153.

Ward, R. D., Skibinski, D. O. F. & Woodwark, M. 1992 Protein heterozygosity, protein structure, and taxonomic differentiation. *Evol. Biol.* **26**, 73–159.

Weir, B. S. & Cockerham, C. C. 1984 Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370.

Wright, S. 1931 Evolution in Mendelian populations. *Genetics* **16**, 97–159.

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

Received 10 June 1996; accepted 6 August 1996

APPENDIX

(*a*) *Simulating samples from the Island Model and Stepping-stone Model*

Samples were simulated using standard coalescent methods (see, for example, Strobeck 1987; Hudson 1990; Slatkin 1993). The basic parameters of the model are: the total number of subpopulations in the metapopulation, d ($= 100$); the number of subpopulations from which a sample is drawn, s; the number in each sample, n and the size of each subpopulation, N. We assume an infinite allele mutation process with rate μ. Some loci have different gene frequency spectra (see, for example, Barbadilla *et al.* 1996) but this should affect the expectation rather than the distribution of F_{ST}. Immigrants into each population are drawn equally from every other subpopulation. Every individual has probability m of having been in a different subpopulation in the previous generation. In the 'asymmetrical' case m varies between subpopulations. One biological interpretation is that some subpopulations are more difficult to reach than others, but each subpopulation contributes equally to the pool of migrants. It is convenient to express mutation and migration in terms of the rescaled rates $\theta = 2dN\mu$, and $M = Nm$.

In the case of the Stepping-stone Model, the implementation is identical to that of the Island Model, with the exception that migrants have an equal chance of being drawn from any of the four neighbouring populations.

(*b*) *Simulating samples from the Colonization Model*

The Colonization Model is described in greater detail in Nichols & Beaumont (1996), where the effect of migration is also considered. The subpopulations are founded by N_0 individuals drawn from an ancestral population of size N. Each subpopulation then grows exponentially with rate r so that at time t it is of size

$$N_t = N_0(1+r)^t. \quad \text{(A 1)}$$

In the 'asymmetrical' case N_0, and r, can vary among colonies.

Nichols and Beaumont show that the value of F_{ST} rapidly approaches a stationary value at a rate that depends on the population growth rate. This stationarity continues for as long as the population grows exponentially. Thereafter, F_{ST} approaches its equilibrium value (i.e. 1, in the absence of migration) at a rate that depends on population size.

For simulating samples, we conservatively chose a time period, $\hat{t}$ sufficiently long that the proportionate change in the expected value of F_{ST} was less than 10^{-6} per generation. Mutations in this short period have negligible effect on F_{ST} and are ignored.

The ancestry of a sample is traced back from time $\hat{t}$, generation by generation. If the probability of coalescence, p_c, is less than 0.2 then with probability p_c two lineages chosen at random are joined. Otherwise each member of the sample is assigned at random to one of $N_t - 1$ parents. Each individual has probability M/N_t of being a migrant. The ancestry of migrants and all lineages tracing back to the t_0 are treated as a random sample from the ancestral population, which is generated by standard coalescent methods with parameters N and μ.

(*c*) F_{ST} *in heterogenous models*

F_{ST} is estimated as

$$\frac{\hat{f}_0 - \hat{f}_1}{1 - \hat{f}_1}. \quad \text{(A 2)}$$

In heterogeneous populations, provided that the expected value of $\hat{f}_1$ is the same for all pairwise comparisons, the expected value of F_{ST} will be a simple average of expected values within groups of similar populations, weighted by their frequency in the sample.

A constant f_1 is expected in the non-equilibrium colonization model (with low mutation rate) because lineages that do not coalesce in the subpopulations must coalesce in the ancestral population. It is also approximately true for the island model when d, the number of islands, is large: two genes in two different islands have approximately a $1/d$ probability of migrating to the same population given that one or both migrate, and therefore on average one or both lineages need to migrate d times before occupying the same subpopulation. Because of these multiple migrations the expected value of f_1 should be similar for all pairs of populations.

This can be obtained directly but tediously for two groups of populations in the Island Model by generalizing the recursions for f_0 and f_1 given in Crow & Aoki (1984) to allow for heterogeneous parameters. Solving the recursions at equilibrium, and letting $\mu \to 0$ and $d \to \infty$ the same result is obtained. The results depend only on the parameters of the sampled populations.

(*d*) *Estimation of the Density of* F_{ST}

Each simulation was used to generate a minimum of 5000 points (pairs of F_{ST} and heterozygosity values). The points were ranked by heterozygosity and grouped into overlapping bins of 400 points centred on every 200th point. Points in the same bin had very similar heterozygosity and the distribution of F_{ST} varied little within bins. The mean, variance, skewness and kurtosis of the 400 F_{ST} values were used to generate a Johnson distribution (Hill *et al.* 1976). The quantiles of this distribution were cross validated by generating further simulated points. They were found to be reliable, and consistently superior to those produced by bivariate kernel density estimation, which over-smoothed the tails of the distribution.