Differences Between Patterns of Gene Flow Inferred from RFLPs and Isozymes in Sea Beet are Consistent across Loci

A. F. Raybould and R. T. Clarke

Centre for Ecology and Hydrology, Winfrith Technology Centre, Dorchester, Dorset DT2 8ZD, UK

ABSTRACT

The distribution of allelic variation at genetic marker loci can be used to estimate parameters such as rates of gene flow and effective population size in populations of crop relatives. These data are useful when designing strategies for the most efficient collection of crop genetic resources. Estimates of gene flow among populations of sea beet (*Beta vulgaris* ssp. *maritima*) differed depending on whether the data were from isozymes or restriction fragment length polymorphisms (RFLPs). The difference was not due to the effect of one or a small subset of 'aberrant' loci within one of the sets of markers. The data suggest that, where practical, sampling strategies should be based on data from more than one type of genetic marker.

Additional Key Words: *Beta vulgaris* spp. *maritima*, dispersal, genetic structure, genetic resources.

The spatial and temporal distribution of allelic variation at genetic marker loci in plant populations can be used to estimate parameters such as rates of gene flow and effective population size. Estimates of these parameters can facilitate conservation of genetic resources in crop relatives (Schoen and Brown, 1993). Models for inferring population dynamics parameters from allele frequency distributions assume that variation is neutral, in other words, the genetic structure of the population at these loci is the result of gene flow and drift. The effect of mutation is assumed to be very small compared with drift and gene flow, except for markers like microsatellites where new alleles are produced by DNA slippage during replication. If the assumption of neutrality is met, each marker locus will give an independent estimate of the combined effects of gene flow and

drift, and data can be combined across loci to give more precise estimates of genetic structure and gene flow. It also follows that estimates of parameters made from different groups of neutral marker loci should not differ statistically.

Genetic markers are frequently used to test whether seed and pollen dispersal is restricted among a group of populations; a phenomenon called isolation by distance (IBD). Slatkin (1993) showed that under a wide range of conditions, IBD could be detected by calculating the regression or correlation coefficient between log (gene flow) and log (distance); a statistically significant negative coefficient indicating IBD. We have shown (Raybould et al., 1996; 1997) that IBD could be detected in a group of sea beet populations in Dorset (UK) when gene flow estimates were based upon restriction fragment length polymorphisms (RFLPs). However, IBD was not detected when isozyme markers were used to estimate gene flow. Furthermore, we have demonstrated that the difference between the isozyme and RFLP regression slopes is statistically significant (Raybould and Clarke, 1999).

Other studies have shown differences between isozymes and other markers when used to describe patterns of gene flow (Karl and Avise, 1992; Pogson et al., 1995). One reason suggested for the differences is that selection is acting on isozyme loci. Because gene flow estimates from several loci are weighted towards more variable loci (Weir and Cockerham, 1984), it is possible that just one 'aberrant' isozyme locus under selection could underlie the difference we detected between RFLPs and isozymes in the sea beet populations. In this paper, we test whether this is the case.

METHODS

Fifty sea beet plants were collected from each of ten populations in Dorset, UK, and scored for variation at seven isozyme and six RFLP loci (see Raybould et al., 1996, for details). For each locus, Weir and Cockerham's (1984) estimator of F_{ST} was calculated for each pair of populations using the program FSTAT (Goudet, 1995) and converted to gene flow (Nm – the number of migrants per population per generation) from the formula $Nm = (1/4F_{ST}) - 0.25$. (Wright, 1951). To test for IBD at each locus, the correlation between $\log_{10}$ (Nm) and $\log_{10}$ (distance) was calculated, and its statistical significance determined with a Mantel randomization test, which allows for the non-independence of pairwise data. Single locus estimates of pairwise F_{ST} values are often undefined or negative. Undefined F_{ST} values are obtained when two populations are fixed for the same allele. Negative F_{ST} values occur when the estimate of the between population mean squares is less than the within population mean squares,

leading to a negative estimate of the between populations component of variance in allele frequency. An undefined or negative F_{ST} value gives a missing value for Nm, so our program for carrying out Mantel tests was modified to cope with this situation. To test for differences in the relationship between Nm and distance between locus X and locus Y, we tested the significance of the correlation coefficient between $\log_{10}(Nm_Y/Nm_X)$ and $\log_{10}$ (distance), where Nm_i is the estimate of Nm from locus i for the same pair of populations (Raybould and Clarke, 1999).

RESULTS

Table 1. Correlation (r) of $\log_{10}$ (gene flow) and $\log_{10}$ (distance) between 10 sea beet populations in Dorset, UK. Separate gene flow estimates were made for 7 isozyme and 6 RFLP loci in the same group of plants. Statistical significance (p) was tested with a one-sided Mantel test.

Locus	r	p
Isozymes		
acph	+0.131	0.8320
est	-0.141	0.1470
got-3	+0.180	0.8670
got-4	-0.108	0.2224
per	+0.111	0.7077
6pgdh	-0.088	0.2618
pgi	+0.150	0.8419
All isozymes	+0.003	0.5039
RFLPs		
L3	-0.465	0.0020
L9	-0.495	0.0036
R1	-0.187	0.1009
R4	-0.398	0.0243
R7	+0.084	0.6779
R13	-0.279	0.0244
All RFLPs	-0.631	0.0007

Table 1 shows the correlation between single locus estimates of $\log_{10}$ (Nm) and $\log_{10}$ (distance). Four of six RFLP loci show significant negative correlations (IBD), whereas none of the correlations for seven isozyme loci was significant. Correlation coefficients for isozymes are significantly greater (i.e. less negative) than those for RFLPs (two-sided Mann-Whitney test $P = 0.007$, $N = 13$).

Table 2 shows the correlation between $\log_{10}(Nm_Y/Nm_X)$ and $\log_{10}$ (distance). This correlation is zero if loci X and Y have the same regression slopes for $\log_{10}$ (Nm) against $\log_{10}$ (distance) (Raybould and Clarke, 1999). Of 21 pairwise comparisons of isozyme loci, only one (*6pgdh* with *got-3*) showed a significant difference in slopes. Similarly, among RFLPs only two of 15 comparisons were significant (*L9/R7*, *R4/R7*). In contrast, of 42 comparisons between isozymes and RFLPs, ten were significant. Pooling the same locus type comparisons (i.e. 3/36), the proportion of significant tests was not significantly greater (at the 5% level) for the between marker comparisons ($\chi^2 = 3.34$, 1 d.f., $P = 0.067$). Nevertheless, the absolute values of the correlations comparing loci of different marker categories were significantly greater than the correlations comparing loci of the same marker type (two-sided Mann-Whitney test $P = 0.047$, N = 78). (The tests are approximate, as the correlations are not totally independent.)

DISCUSSION

The locus-by-locus analysis of the sea beet data showed that the different patterns of gene flow given by RFLPs and isozymes did not result from the large effect of a single 'aberrant' locus. Loci of the same marker type were on average more similar to each other, in terms of the relationship between Nm and distance, than to loci of the other marker type. Four out of six individual RFLP loci indicated significant isolation by distance, whereas for each isozyme locus the correlation between gene flow and distance was not statistically significant. A Mann-Whitney test showed that the rank order of the correlation coefficients was not random with respect to marker category and hence that the correlations for RFLPs were significantly greater than the correlations for isozymes.

The significance of the correlation between $\log_{10}(Nm_Y/Nm_X)$ and log (distance) tests whether the relationship between gene flow and distance is the same for locus X and locus Y. When X was an RFLP locus and Y an isozyme locus, 10 of 42 correlations were significant. However, when loci X and Y were of the same marker type, 3 of 36 correlations were significant. Although this difference seems large, a χ^2 test was not significant at the 5% level (but it was at 7%). Nevertheless, a Mann-Whitney test showed that the absolute values of the correlations when X and Y were different types of marker were greater than the correlations when X and Y were the same type of marker. Again this shows that an isozyme locus on average behaves more similarly to other isozymes than to RFLPs, and RFLPs are more similar to each other than to isozymes. The reasons for these consistent differences between RFLP and isozyme loci are not clear,

Table 2. Correlation between $\log_{10}(Nm_Y/Nm_X)$ and $\log_{10}$(distance) for 10 populations of sea beet, where Nm_i is the estimate of Nm from locus i for the same pair of populations (Raybould and Clarke, 1999). If loci X and Y give the same patterns of gene flow with distance, the correlation between $\log_{10}(Nm_Y/Nm_X)$ and $\log_{10}$ distance is zero. Significance was tested with a two-sided Mantel test.

Locus Y →	acph	*est*	got-3	*got-4*	*per*	6pgdh	pgi	L3	L9	R1	R4	R7
Locus X ↓												
acph												
est	0.230											
got-3	-0.047	-0.244										
got-4	-0.133	-0.010	0.111									
per	0.131	-0.237	-0.050	-0.139								
6pgdh	0.047	-0.147	0.373*	-0.184	-0.155							
pgi	-0.061	-0.289	0.004	-0.207	-0.170	-0.233						
L3	0.244	0.263	0.400*	0.204	0.173	0.320	0.395					
L9	0.330**	0.360**	0.442**	0.205	0.373**	0.206	0.476**	0.199				
R1	0.214	0.053	0.177	-0.037	0.135	0.065	0.340*	-0.202	-0.228			
R4	0.319*	0.398*	-0.169	0.283	0.147	0.273	0.316	0.007	0.007	0.185		
R7	0.047	-0.187	0.105	-0.100	0.123	0.038	0.018	-0.196	-0.400*	-0.129	-0.424*	
R13	0.100	0.189	0.328	0.060	0.123	0.336	0.482**	0.023	-0.225	0.172	-0.263	0.089

however if they are general across loci and populations, estimates of genetic structure in beet based solely on isozymes may be unreliable.

The variation among the patterns of single locus estimates of gene flow with distance raises the question of how loci behave when combined. For example can some of the isozyme loci be combined with RFLP loci to give a stronger relationship between gene flow and distance than is given by RFLPs alone? Although it cannot be considered in detail here, we have found interesting and counterintuitive effects. For example, when the three isozyme loci that give negative correlations with distance are combined to give a multilocus estimate of F_{ST}, the resulting correlation between $\log_{10}$ (Nm) and $\log_{10}$ (distance) is –0.183 (P (r<0) = 0.0787). However, when loci *acph* and *per* are included in the estimate of F_{ST}, the correlation decreases to –0.318 (P (r<0) = 0.0077), even though when used on their own, both *acph* and *per* give positive correlations of Nm and distance. The explanation for this decrease in the multilocus correlation is the subject of current research.

REFERENCES

Goudet, J. 1995. FSTAT V-1.2: a computer program to calculate F-statistics. J. Hered. 86: 485-486.

Karl, S.A. and J.C. Avise. 1992. Balancing selection at allozyme loci in oysters: implications from nuclear RFLPs. Science 256: 100-102.

Pogson, G.H., K.A. Mesa, and R.G. Boutilier. 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.

Raybould, A.F., R.J. Mogg, and R.T. Clarke. 1996. The genetic structure of *Beta vulgaris* ssp. *maritima* (sea beet) populations: RFLPs and isozymes show different patterns of gene flow. Heredity 77: 245-251.

Raybould, A.F., R.J. Mogg, and C.J. Gliddon. 1997. The genetic structure of *Beta vulgaris* ssp. *maritima* (sea beet) populations. II. Differences in gene flow estimated from RFLP and isozyme loci are habitat-specific. Heredity 78: 532-538.

Raybould, A.F. and R.T. Clarke. 1999. Defining and measuring gene flow. In: Gene Flow and Agriculture: Relevance for Transgenic Crops. British Crop Protection Council. pp. 41-48.

Schoen, D.J and A.D.H. Brown 1993. Conservation of allelic richness in wild crop relatives is aided by assessment of genetic markers. Proc. Natl. Acad. Sci. USA 90: 10623-10627.

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

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. Eugenics 15: 323-354.