The genetic structure of sea beet (*Beta vulgaris* ssp. *maritima*) populations. III. Detection of isolation by distance at microsatellite loci

A. F. RAYBOULD*, R. J. MOGG, C. ALDAM, C. J. GLIDDON†, R. S. THORPE† & R. T. CLARKE

Institute of Terrestrial Ecology, Furzebrook Research Station, Wareham, Dorset BH20 5AS and †School of Biological Sciences, University of Wales, Bangor, Gwynedd LL57 5UW, U.K.

Variation at three microsatellite loci was investigated in five populations of sea beet on the Dorset coast. Genetic variation among the populations was estimated using two different methods. Estimates of F_{ST} were obtained using the program FSTAT, and estimates of R_{ST} were obtained using the program AMOVA. These different estimates of genetic distance were used to test for isolation by distance with partial matrix correspondence tests and were compared with patterns of variation in F_{ST} estimates from isozymes and single-copy nuclear restriction fragment length polymorphisms (RFLPs). F_{ST} from isozymes, F_{ST} from RFLPs and R_{ST} from microsatellites all showed evidence of isolation by distance, whereas F_{ST} from microsatellites showed strong similarities within populations that were not related to distance. These results suggest that, compared with isozymes and RFLPs, mutation rates at microsatellite loci are high in these populations. They also suggest that stepwise mutation processes are occurring at these loci. The data provide valuable evidence that all three types of marker can detect similar patterns of population structure. The results are discussed with respect to estimates of genetic distance at microsatellite loci in human populations.

Keywords: *Beta vulgaris*, isolation by distance, isozymes, microsatellites, RFLPs, stepwise mutation.

Introduction

Robust estimation of genetic structure and gene flow requires information from many loci, in order that 'badly behaved loci' may be identified and excluded from the analysis (Slatkin, 1985; Goudet *et al.*, 1994). It may also be necessary to use different classes of genetic marker. For example, Karl & Avise (1992) and Raybould *et al.* (1996a, 1997) have shown that single-copy nuclear restriction fragment length polymorphisms (SCN RFLPs) can give estimates of gene flow that are very different from those obtained by isozymes, possibly because of some form of selection at isozyme loci.

Molecular biology techniques provide plant population geneticists with an almost limitless source of highly polymorphic neutral markers with which to estimate gene flow. Since the mid-1980s, SCN RFLPs have proved very valuable to geneticists producing genetic maps of crop species (Helentjaris & Burr, 1989). Nevertheless, despite providing a potentially large number of very polymorphic co-

dominant loci, they have been little used in the study of natural plant populations. Reasons for this may be connected with the availability of probes, problems of isolating good quality DNA from wild plants, the use of radioisotopes, high expense, time constraints and the relatively high degree of skill needed to use RFLPs successfully (e.g. Rafalski & Tingey, 1993). More recently, microsatellites (also called simple sequence repeats, SSRs) have been introduced into many areas of genetics that might benefit from the use of RFLPs (Jarne & Lagoda, 1996). They seem to have advantages over RFLPs in terms of shorter development and processing times, no necessity for radioisotopes or high-quality DNA and generally require much less complex experimental protocols, while retaining the potential of RFLPs to deliver a large number of highly polymorphic codominant loci. However, if primer sequences are not available, start-up costs can be high (Rafalski & Tingey, 1993).

Microsatellites are beginning to fulfil their potential in areas such as genome mapping (e.g. Szewc-McFadden *et al.*, 1996) and the study of relatedness

*Correspondence. E-mail: a.raybould@ite.ac.uk

127

in social animals (e.g. Seppa & Gertsch, 1996). They should also prove very suitable for testing models of population genetic structure. A rigorous test of the suitability of microsatellite variation requires a system in which genetic structure has been described independently of microsatellite variation. The cliff-top populations of *Beta vulgaris* ssp. *maritima* (sea beet) in Dorset provide such a system. Gene flow (Nm), derived from F_{ST} estimates, among these populations has been analysed using SCN RFLPs and isozymes (Raybould *et al.*, 1996a, 1997). Regressions of log Nm and log distance between all pairs of populations (Slatkin, 1993) indicate that isolation by distance operates in these populations. In this paper, we examine whether microsatellite variation can also be used to detect isolation by distance in these populations.

There is much debate over how F_{ST} will perform as a measure of genetic distance at microsatellite loci, which tend to have high mutation rates (e.g. 10^{-3} per generation in humans; Weber & Wong, 1993) and usually mutate by gain or loss of one or two repeats per mutation event. Slatkin (1995) has shown that these properties fail to satisfy the assumptions that allow demographic parameters (such as Nm) to be inferred from F_{ST}. These assumptions are that mutation rates are low and that mutation obeys an infinite alleles model (or k-alleles model with k tending to infinity). If the stepwise mutation model with high mutation rates describes microsatellite variation better than the infinite alleles model with low mutation rates, parameters of genetic structure need to allow for the difference in allele size (number of repeats). Slatkin (1995) described such a parameter, R_{ST}, which is analogous to F_{ST}. R_{ST} is defined by variances of allele sizes such that

$$R_{ST} = (S - S_w)/S,$$

where S_w is twice the average of the estimated variances in allele size (number of repeats) between alleles within each subpopulation and S is twice the estimated variance in allele size in the population as a whole (Slatkin, 1995).

The properties of parameters describing genetic variation based on the different mutation models have been examined by Slatkin (1995), Michalakis & Excoffier (1996) and Rousset (1996), and theoretical situations in which the estimators of the different parameters would be expected to characterize actual genetic distance best have been described. However, there is no consensus at present as to whether R_{ST} or similar estimators are more appropriate than F_{ST} for the analysis of microsatellite data from real popula-

tions (Jarne & Lagoda, 1996). Here, we use both Weir & Cockerham's (1984) F_{ST} estimator and Michalakis & Excoffier's (1996) estimator of R_{ST} in regressions of genetic and geographical distance to examine whether microsatellites can detect evidence of isolation by distance in populations in which this process is known to occur, and whether the different estimators vary in their ability to detect isolation by distance.

Materials and methods

Microsatellite analysis

Microsatellite analysis was carried out on plants from five cliff-top populations surveyed previously for isozyme and RFLP variation (50 plants per population; Raybould *et al.*, 1996a, 1997). DNA was extracted from all remaining 230 tissue samples using the method of Edwards *et al.* (1991). Microsatellite sequences were amplified from 2.5 μL of (unquantified) DNA using the primers described by Mörchen *et al.* (1996). Amplifications were performed in 25 μL samples with one unit of *Taq* polymerase, 5 pmol of each primer, 5 μmol of each nucleotide and 1.2 per cent formamide. Denaturation was for 1.5 min at 94°C followed by 30 cycles of 2 min at the annealing temperature, 1.5 min at 72°C and 20 s at 94°C. The amplification was completed with 2 min at the annealing temperature and 5 min at 72°C. The annealing temperature was calculated as the melting temperature of the primer (T_m) minus 5°C, where $T_m = 2$ (A + T bp) + 4 (C + G bp). Amplification products were separated by electrophoresis on 10 per cent acrylamide gels followed by ethidium bromide staining (Mörchen *et al.*, 1996).

Tests for isolation by distance

For the analysis of genetic structure, populations were subdivided into patches approximating to randomly mating units (Goudet *et al.*, 1994; Raybould *et al.*, 1996b). Because of lack of tissue and some failures in amplification (see below), complete microsatellite data were available for 204 plants only (AMOVA, the program for calculating R_{ST} estimates, requires plants to have data at all loci). For consistency, plants without complete microsatellite data were removed from the RFLP and isozyme data sets. Patches with fewer than five plants remaining were pooled with the closest neighbouring patch. This gave 21 patches. Genetic distances were then estimated between all pairs of patches. Genetic distances were calculated as both F_{ST} (for RFLPs, isozymes and microsatellites), using

Weir & Cockerham's (1984) estimator as computed by the program FSTAT (Goudet, 1995) and Michalakis & Excoffier's (1996) estimator of R_{ST} (for microsatellites only) as computed by the program AMOVA (Excoffier et al., 1992). Pooling of patches may introduce biases caused by substructure within the new groupings, leading to underestimation of the absolute values of F_{ST} and R_{ST}, although this should have minimal effects on the detection of processes such as isolation by distance (Raybould et al., 1997).

The pairwise genetic distances, F_{ST} and R_{ST}, were classified into those between patches in the same population (coded 1 in a new variable, 'population membership', distinguishing intra- from interpopulation comparisons of patches) and those between patches in different populations (coded 0). The separate effects of population membership and distance were examined by carrying out partial regressions and partial matrix correspondence tests (Manly, 1991; Thorpe & Baez, 1993; Thorpe & Malhotra, 1996) of genetic distance on geographical distance, population membership and the interaction between distance and population membership (= distance multiplied by either 0 or 1), respectively. The interaction term tests for uniformity of the effect of distance within and between populations. The same procedures were adopted with log-transformed distances. For each group of regression coefficients (transformed/untransformed distances), significance levels were set by sequential Bonferroni procedures (Rice, 1989).

Results

In the Dorset material, three of the four microsatellite primer pairs described by Mörchen et al. (1996) were found to give polymorphism interpretable as variation at single loci. Bvm1, which was monomorphic in French material, was polymorphic in the Dorset populations, but the variation could not be interpreted in any simple fashion. Table 1 shows the allele frequencies at the interpretable loci. The number of alleles per locus is lower in the Dorset sample, even though it has more plants, probably because it was collected over a smaller area. It can be seen from Table 1 that fewer than 230 phenotypes were obtained for each microsatellite locus. We believe that failure of samples to amplify was caused by factors other than null alleles (Pemberton et al., 1995) because, in most cases, a sample either amplified with all or none of the primer pairs.

Table 2 summarizes the results of the partial regressions. The results show clearly that F_{ST} and R_{ST}

estimators behave differently for microsatellites in the sea beet populations. There is a very strong population membership effect when genetic distance between patches is estimated as F_{ST}, whereas population membership is not a significant factor in the R_{ST} estimates between patches. The partial regressions of R_{ST} behave similarly to F_{ST} for isozymes and RFLPs. With log-transformed distances, the estimates of F_{ST} for microsatellites give the most negative partial regression with population membership, but no coefficient is significant.

The tests for isolation by distance are more clear-cut with the log-transformed data. F_{ST} at isozyme and RFLP loci show isolation by distance, confirming the results of Raybould et al. (1997) using Slatkin's (1993) method of detecting isolation by distance. The partial regression coefficient of R_{ST} estimates and log distance is significant at the 5 per cent level for a single test and marginally significant (<10 per cent) after table-wide testing. F_{ST} at microsatellite loci, however, fails to detect isolation by distance in these populations.

For all estimators, with both transformed and untransformed distances, no interaction term is significant. This suggests that the rate of change of genetic distance with geographical distance is the same within and between these populations.

Discussion

Our results show that isolation by distance can be detected in the sea beet populations by microsatellite loci, but that the choice of genetic distance estimator is important. When genetic distances at microsatellite loci are estimated by F_{ST}, patches of plants within the same population are significantly more similar to each other than those in different populations when the effect of distance is removed. However, there is no significant effect of population membership (with distance effects removed) for isozymes and RFLPs. This suggests that mutation rates at these microsatellites are higher than rates of gene flow, whereas gene flow rates are higher than mutation rates at isozyme and RFLP loci. This is consistent with observations of high mutation rates at microsatellite loci in humans (Weber & Wong, 1993; Dib et al., 1996) and mice (Dallas, 1992). The number of alleles per microsatellite locus (four to five) is higher than that found at RFLP and isozyme loci in the same populations (c. three; Raybould et al., 1996a), although the number of alleles is much smaller than that found by Mörchen et al. (1996).

If R_{ST} is used as the genetic distance estimator, a pattern of genetic structure emerges that is very

similar to isozymes and RFLPs. With R_{ST}, the significant population effect is removed. Using transformed distance data, isolation by distance can be detected with R_{ST} at a significance level of 7 per cent after table-wide testing. Thus, providing a suitable estimator is chosen and suitable transformations of distance are carried out, a significant relationship between genetic distance and geographical distance can be found at three different sets of loci. This is a valuable result as it gives us confidence in both the robustness of the marker systems and the analytical procedures.

It should not be inferred from these data that a genetic distance estimator based on a stepwise mutation model will always be more appropriate than an estimator such as F_{ST}. For example, Perez Lezaun *et al.* (1997) have recently shown that neighbour-joining trees based on human microsatellite varia-

Table 1 Microsatellite allele frequencies in Dorset populations of sea beet

Locus	Allele size (bp)	Old Harry	Durlston	Seacombe	Winspit	St. Aldhelm's Head	Wild beets in France*
Bvm2	99	0.1389	0.3026	0.3085	0.04335	0.0500	0.12
	102						0.03
	111						0.24
	114						0.09
	117	0.0556	0.0526	0.2609	0.0217		
	120	0.6528	0.4474	0.4022	0.6196	0.7700	
	123						0.06
	132						0.03
	135						0.23
	138						0.03
	147	0.1528	0.1974	0.1848	0.3152	0.1800	0.06
	>800						
		$n = 36$	$n = 38$	$n = 47$	$n = 50$	$n = 50$	$n = 17$
Bvm3	96						0.03
	100						0.28
	104	0.2639				0.1700	
	108						0.12
	110	0.6528	0.0957	0.7979	0.6600	0.6800	0.19
	112						0.16
	114	0.0556	0.8830	0.2021	0.3200	0.1300	
	118						0.03
	120						0.13
	122	0.0278			0.0200	0.0100	0.06
	128		0.0213			0.0100	
		$n = 36$	$n = 47$	$n = 47$	$n = 50$	$n = 50$	$n = 17$
Bvm4	55						0.03
	63						0.22
	65						0.10
	69						0.03
	79		0.0139				
	81						0.03
	83		0.5417	0.1522	0.1064	0.0100	0.06
	85	0.4054	0.0566	0.2609	0.1915	0.5100	0.06
	89	0.2568	0.0417	0.4022	0.4574	0.1000	0.06
	93						0.03
	95	0.3378	0.2222	0.1848	0.2447	0.3800	0.25
	97						0.13
	103		0.1250				
		$n = 37$	$n = 36$	$n = 46$	$n = 47$	$n = 50$	$n = 17$

*Mörchen *et al.* (1996)

tion are different depending on the choice of genetic distance estimator. However, in this case, the trees based on estimates of F_{ST} are in closer agreement with archaeological and other genetic data than are trees based on estimates of R_{ST}. These results, and our results from sea beet, imply that, if possible, microsatellite data should be analysed using estimators of parameters based on both stepwise mutation and infinite alleles models. If significant differences arise as a result of these different forms of analysis, only the presence of independent information may allow a decision to be made as to which method is more appropriate. Rousset (1996) has pointed out

that, for the mutation process to have a significant effect on F-statistics, the mutation rate must be high enough for two or more mutation events to occur in different populations since the time of common ancestry. When population studies are made using hypervariable loci, which may be assumed to have high mutation rates, theory predicts that populations separated by relatively large geographical distances (and having relatively low migration rates between them) will show significant effects of mutation, whereas populations with higher migration (geographically closer) will not. Such data sets may need to be analysed using F_{ST} for closer populations

Table 2 Partial regression coefficients and partial matrix correspondence test probabilities for partial regressions of genetic distance estimators and geographical distance, population membership and distance × membership interaction for approximately randomly mating patches of sea beet in five Dorset populations. Figures in bold are regression coefficients with untransformed distances; figures in parentheses are regression coefficients with log-transformed distances. Significance levels set by sequential Bonferroni methods (Rice, 1989) for transformed and untransformed estimates. NS, significance level >0.1

Marker type	Estimator	Regression coefficient (b) (distance)	Probability b not >0	Significance level
Isozymes	F_{ST}	**0.0136**	**0.0014**	**0.0056**
		(0.145)	(0.0010)	(0.0040)
RFLPs	F_{ST}	**0.0077**	**0.0490**	**NS**
		(0.116)	(0.0010)	(0.0040)
Microsatellites	F_{ST}	**−0.0045**	**0.8544**	**NS**
		(0.0379)	(0.2120)	(NS)
Microsatellites	R_{ST}	**0.0052**	**0.0529**	**NS**
		(0.0731)	(0.0350)	(0.0700)

Marker type	Estimator	Regression coefficient (b) (population)	Probability b not <0	Significance level
Isozymes	F_{ST}	**−0.0089**	**0.4206**	**NS**
		(−0.0003)	(0.5034)	(NS)
RFLPs	F_{ST}	**−0.0135**	**0.3918**	**NS**
		(0.0728)	(0.7106)	(NS)
Microsatellites	F_{ST}	**−0.2050**	**<0.0001**	**<0.0004**
		(−0.136)	(0.1242)	(NS)
Microsatellites	R_{ST}	**−0.0521**	**0.1309**	**NS**
		(0.0967)	(0.8036)	(NS)

Marker type	Estimator	Regression coefficient (b) (interaction)	Probability b = 0	Significance level
Isozymes	F_{ST}	**0.0668**	**0.8907**	**NS**
		(−0.1550)	(0.0665)	(NS)
RFLPs	F_{ST}	**0.0033**	**0.9964**	**NS**
		(−0.0745)	(0.4703)	(NS)
Microsatellites	F_{ST}	**0.1640**	**0.7687**	**NS**
		(−0.0314)	(0.7377)	(NS)
Microsatellites	R_{ST}	**0.5540**	**0.3099**	**NS**
		(0.0015)	(0.9871)	(NS)

and using R_{ST} for those further apart in order to dissect the effects of isolation by distance from those of mutation.

Acknowledgements

References

DALLAS, J. F. 1992. Estimation of microsatellite mutation rates in recombinant inbred strains of mouse *Mamm. Genome*, **3**, 452–456.

DIB, C., FAURE, S., FIZAMES, C., SAMSON, D., DROUOT, N., VIGNAL, A. *ET AL*. 1996. A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature*, **380**, 152–154.

EDWARDS, K., JOHNSTONE, C. AND THOMPSON, C. 1991. A simple and rapid method for preparation of plant genomic DNA for PCR analysis. *Nucl. Acids Res.*, **19**, 1349.

EXCOFFIER, L., SMOUSE, P. E. AND QUATTRO, J. M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

GOUDET, J. 1995. FSTAT V-1.2. A computer program to calculate F-statistics. *J. Hered.*, **86**, 485–486.

GOUDET, J., DE MEEÜS, T., DAY, A. J. AND GLIDDON, C. J. 1994. The different levels of population structuring of the dogwhelk, *Nucella lapillus*, along the south Devon coast. In: Beaumont, A. R. (ed.) *Genetics and Evolution of Aquatic Organisms*, pp. 81–95. Chapman & Hall, London.

HELENTJARIS, T. AND BURR, B. 1989. *Development and Application of Molecular Markers to Problems in Plant Genetics*. Cold Spring Harbor Laboratory Press, New York.

JARNE, P. AND LAGODA, P. J. L. 1996. Microsatellites, from molecules to populations and back. *Trends Ecol. Evol.*, **11**, 424–429.

KARL, S. A. AND AVISE, J. C. 1992. Balancing selection at allozyme loci in oysters: implications from nuclear RFLPs. *Science*, **256**, 100–102.

MANLY, B. F. J. 1991. *The Statistics of Natural Selection*. Chapman and Hall, London.

MICHALAKIS, Y. AND EXCOFFIER, L. 1996. A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics*, **142**, 1061–1064.

MÖRCHEN, M., CUGUEN, J., MICHAELIS, G., HÄNNI, C. AND SAUMIOU-LAPRADE, P. 1996. Abundance and length polymorphism of microsatellite repeats in *Beta vulgaris* L. *Theor. Appl. Genet.*, **92**, 326–333.

PEMBERTON, J. M., SLATE, J., BANCROFT, D. R. AND BARRETT, J. A. 1995. Nonamplifying alleles at microsatellite loci: a caution for parentage and population studies. *Mol. Ecol.*, **4**, 249–252.

PEREZ LEZAUN, A., CALAFELL, F., MATEU, E., COMAS, D., RUIZ PACHECO, R. AND BERTRANPETIT, J. 1997. Microsatellite variation and modern humans. *Hum. Genet.*, **99**, 1–7.

RAFALSKI, J. A. AND TINGEY, S. V. 1993. Genetic diagnostics in plant breeding: RAPDs, microsatellites and machines. *Trends Genet.*, **9**, 275–280.

RAYBOULD, A. F., MOGG, R. J. AND CLARKE, R. T. 1996a. The genetic structure of sea beet (*Beta vulgaris* ssp. *maritima*) populations. RFLPs and isozymes show different patterns of gene flow. *Heredity*, **77**, 245–250.

RAYBOULD, A. F., GOUDET, J., MOGG, R. J., GLIDDON, C. J. AND GRAY, A. J. 1996b. Genetic structure of a linear population of sea beet (*Beta vulgaris* ssp. *maritima*) revealed by isozyme and RFLP analysis. *Heredity*, **76**, 111–117.

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

RICE, W. B. 1989. Analyzing tables of statistical tests. *Evolution*, **43**, 223–5.

ROUSSET, F. 1996. Equilibrium values of measures of population subdivision for stepwise mutation processes. *Genetics*, **142**, 1357–1362.

SEPPA, P. AND GERTSCH, P. 1996. Genetic relatedness in the ant *Campanotus herculeanus* — a comparison of estimates from allozyme and microsatellite markers. *Insect Soc.*, **43**, 235–246.

SLATKIN, M. 1985. Gene flow in natural populations. *Ann. Rev. Ecol. Syst.*, **16**, 393–430.

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

SLATKIN, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–462.

SZEWC-McFADDEN, A. K., KRESOVICH, S., BLIEK, S. M., MITCHELL, S. E. AND McFERSON, J. R. 1996. Identification of polymorphic, conserved simple sequence repeats (SSRs) in cultivated *Brassica* species. *Theor. Appl. Genet.*, **93**, 534–538.

THORPE, R. S. AND BAEZ, M. 1993. Geographic variation in scalation of the lizard *Gallotia stehlini* within the island of Gran Canaria. *Biol. J. Linn. Soc.*, **48**, 75–87.

THORPE, R. S. AND MALHOTRA, A. 1996. Molecular and morphological evolution within small islands. *Phil. Trans. R. Soc. B*, **351**, 815–822.

WEBER, J. L. AND WONG, C. 1993. Mutation of human short tandem repeats. *Hum. Mol. Genet.*, **2**, 1123–1128.

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