Molecular Ecology Notes (2006) **6**, 255–256 doi: 10.1111/j.1471-8286.2005.01082.x

TECHNICAL NOTE

Estimation and adjustment of microsatellite null alleles in nonequilibrium populations

C. VAN OOSTERHOUT, D. WEETMAN and W. F. HUTCHINSON
Department of Biological Sciences, University of Hull, Hull, HU6 7RX, UK

Abstract

Nonamplified (null) alleles are a common feature of microsatellite genotyping and can bias estimates of allele and genotype frequencies, thereby hindering population genetic analyses. The frequency of microsatellite null alleles in diploid populations can be estimated for populations that are in Hardy–Weinberg equilibrium. However, many microsatellite data sets are from nonequilibrium populations, often with known inbreeding coefficients (F) or fixation indices (F_{IS} or F_{ST}). Here, we propose a novel null allele estimator that can be used to estimate the null allele frequency and adjust visible allele frequencies in populations for which independent estimates of F, F_{IS} or F_{ST} are available. The algorithm is currently available as an Excel macro that can be downloaded at no cost from http://www.microchecker.hull.ac.uk/ and will be incorporated into the software MICRO-CHECKER.

Keywords: Hardy–Weinberg deviation, MICRO-CHECKER, microsatellites, nonpanmictic populations, null alleles

Received 15 March 2005; revision received 09 May 2005; accepted 20 May 2005

Traditionally, genetic analyses have been based on assumptions of panmixia, but many natural populations exhibit nonequilibrium conditions (Whitlock & McCauley 1999; Estoup *et al*. 2001). For example, as a result of local breeding structure or admixture, populations may depart from expected Hardy–Weinberg proportions (H–W). Deviations from H–W can also be caused, however, by polymorphism at polymerase chain reaction (PCR) priming sites resulting in the failure of amplification of particular alleles (i.e. null alleles). Null alleles are relatively common in microsatellite analysis (Dakin & Avise 2004) and it is desirable to identify their occurrence and frequency for various population genetic analyses and partition their effects from biological causes of H–W deviation.

Several algorithms have been developed to estimate null allele frequencies, which assume that the population is in Hardy–Weinberg equilibrium (Chakraborty *et al*. 1992; Brookfield 1996). These algorithms are employed by software programs such as MICRO-CHECKER (van Oosterhout *et al*. 2004a). This program can discriminate between H–W deviations caused by null alleles from those caused by inbreeding and Wahlund effects and can identify genotype biases caused by short allele dominance (large allele dropout) and stuttering (van Oosterhout *et al*. 2004a). However, current null allele estimation algorithms are not applicable in populations that show deviations from panmixia.

Here, we propose a calculation for estimating null alleles in populations with known inbreeding coefficients (F) or fixation indices (F_{IS} or F_{ST}), derived from population genetic theory of inbreeding (Crow & Kimura 1970). If p_i is the frequency of allele i, r the frequency of the null allele 0, and F the mean inbreeding coefficient or fixation index, then the expected proportion of homozygotes (p_{ii}) is equal to:

$$p_{ii} = p_i^2(1 - F) + p_iF, \quad (1)$$

and the expected proportion of alleles segregating with a null allele 0 in heterozygous genotypes (p_{i0}):

$$p_{i0} = 2p_ir\,(1 - F). \quad (2)$$

Defining n_2 as the frequency of individuals in the data set, which exhibit a single allelic product, but which may possess either one or two copies of allele i, [i.e. the sum of

Correspondence: C. van Oosterhout, Fax: +44 (0)1482 465458; E-mail: c.van-oosterhout@hull.ac.uk

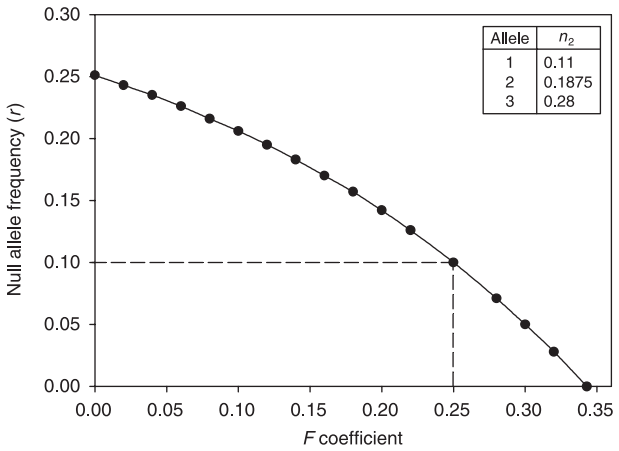

Fig. 1 Relationship between the inbreeding coefficient (F) and null allele frequency (r) in a hypothetical population with an observed frequency of samples that exhibit a single allelic product (n_2), (see insert). In this example, a population with $F = 0.25$ would be estimated to have a null allele frequency $r = 0.10$, as indicated by the drop-lines. Other solutions for r with given F are shown by the curve.

equations (1) and (2)], p_i can be estimated by solving the following quadratic equation:

$$n_2 = p_i^2(1 - F) + p_i(F + 2r - 2Fr). \quad (3)$$

This equation has two solutions:

$$p_i = \{-F - 2r + 2Fr + [(F + 2r - 2Fr)^2 + 4n_2(1 - F)]^{1/2}\}/2(1 - F), \quad (4a)$$

$$p_i = \{-F - 2r + 2Fr - [(F + 2r - 2Fr)^2 + 4n_2(1 - F)]^{1/2}\}/2(1 - F). \quad (4b)$$

Finally, the summed allele frequency over all alleles (including the null allele) at the locus is equal to unity:

$$\sum_i p_i + r = 1 \quad (5)$$

Equations 4a and 4b can be solved, and in conjunction with equation 5, return a single real solution. This allows the null allele frequency to be estimated and amplified allele and genotype frequencies to be adjusted (see figure 1).

The inbreeding coefficient may be obtained from pedigree studies (e.g. van Oosterhout *et al*. 2003, 2004b) or from molecular markers by the Bayesian method of Wilson & Rannala (2003). Similarly, F statistics can be obtained from molecular markers that can be assumed to be free from null alleles such as allozymes and single nucleotide polymorphisms (SNPs). Additionally, a standard outlier analysis, conducted on F_{IS} values obtained from several unlinked microsatellite loci, can be used to separate loci exhibiting null alleles from those reflecting a biological signal of inbreeding.

The algorithm will be incorporated in a future version of MICRO-CHECKER but can currently already be used as an Excel macro. The input required to run the macro consists of the inbreeding coefficient F or the fixation index F_{IS} or F_{ST} (based on an independent set of molecular markers) and the frequency of the putative homozygote allele-size classes (i.e. genotypes exhibiting only one allele).

Availability

The Excel macro including a help file and examples can be downloaded for free from the University of Hull MICRO-CHECKER website http://www.microchecker.hull.ac.uk/.

Acknowledgement

This work was supported by the Natural Environment Research Council, UK (NERC Fellowship NER/I/S/2000/00885 to CVO).

References

Brookfield JFY (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology*, **5**, 453–455.

Chakraborty R, De Andrade M, Daiger SP, Budowle B (1992) Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications. *Annals of Human Genetics*, **56**, 45–57.

Crow JF, Kimura M (1970) *An Introduction to Population Genetics Theory*. Harper & Row, Publishers, New York.

Dakin EE, Avise JC (2004) Microsatellite null alleles in parentage analysis. *Heredity*, **93**, 504–509.

Estoup A, Wilson IJ, Sullivan C, Cornuet J-M, Moritz C (2001) Inferring population history from microsatellite and enzyme data in serially introduced cane toads, *Bufo marinus*. *Genetics*, **159**, 1671–1687.

van Oosterhout C, Trigg RE, Carvalho GR, Magurran AE, Hauser L, Shaw PW (2003) Inbreeding depression and genetic load of sexually selected traits: how the guppy lost its spots. *Journal of Evolutionary Biology*, **16**, 273–281.

van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004a) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

van Oosterhout C, van Heuven MK, Brakefield PM (2004b) On the neutrality of molecular genetic markers: pedigree analysis of genetic variation in fragmented populations. *Molecular Ecology*, **13**, 1025–1034.

Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: F_{ST} doesn't equal $1/(4Nm + 1)$. *Heredity*, **82**, 117–125.

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, **163**, 1177–1191.