Molecular Ecology Notes (2006) **6**, 951–957 doi: 10.1111/j.1471-8286.2006.01449.x

REVIEW

Mitigating scoring errors in microsatellite data from wild populations

JENNIFER DEWOODY,* JOHN D. NASON† and VALERIE D. HIPKINS*
**USDA Forest Service, Pacific Southwest Research Station, National Forest Genetics Laboratory, 2480 Carson Road, Placerville, California 95667, USA, †Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, Iowa 50011, USA*

Abstract

Microsatellite data are widely used to test ecological and evolutionary hypotheses in wild populations. In this paper, we consider three typical sources of scoring errors capable of biasing biological conclusions: stuttering, large-allele dropout and null alleles. We describe methods to detect errors and propose conventions to mitigate scoring errors and report error rates in studies of wild populations. Finally, we discuss potential bias in ecological or evolutionary conclusions based on data sets containing these scoring errors.

Keywords: microsatellite, population genetics, scoring errors, SSR

Received 23 February 2006; revision received 25 March 2006; accepted 3 May 2006

Introduction

Microsatellites, or simple sequence repeats (SSRs), are powerful tools commonly used to test a variety of ecological and evolutionary hypotheses in wild populations. Errors in scoring microsatellite data can occur at several steps of the assay, yet conventions in disclosing scoring errors and standard statistical procedures to mitigate these errors are lacking (Bonin *et al*. 2004; Dakin & Avise 2004). Despite the implications of including scoring errors in downstream analyses, both basic researchers (arguably the vast majority of authors in refereed literature) and the management community (Paetkau 2003) continue to use microsatellites without estimating error rates. Fewer still describe the techniques used to mitigate for errors detected (but see Barker 2005), or discuss how errors in the data may bias the conclusions of the study. Studies involving human genetics, as well as those using noninvasive (e.g. fecal or hair) or ancient samples, however, are currently among the few to routinely report error rates (Hoffman & Amos 2005). As more effective analytical tools for detecting genotyping errors continue to be developed, it is time to consider establishing clear and consistent protocols for estimating error rates, reporting these rates in studies of wild populations, and mitigating for errors in downstream analyses and data interpretation. In this paper, we discuss three common microsatellite scoring errors, those due to stuttering, large-allele dropout and null alleles, and consider their impacts on ecological and evolutionary data interpretation. We recommend standard procedures to mitigate these impacts, and consider conventions for reporting potential scoring errors and associated interpretative biases in the literature.

Correspondence: J. DeWoody, Fax: 530-622-2633;
E-mail: jdewoody@fs.fed.us

Three common scoring errors

Taberlet *et al*. (1996), Bonin *et al*. (2004) and Hoffman & Amos (2005) thoroughly discuss the numerous sources of scoring errors (or mistypes, i.e. assigning at least one wrong allele to a genotype) in microsatellite data. Three sources of error, stuttering patterns, large-allele dropout and null alleles, are of particular concern because, unlike errors created by stochastic amplifications, these errors tend to create consistent allelic and genotypic scoring bias that may, in turn, bias data interpretation.

Some loci tend to produce 'stutter' bands due to slipping by *Taq* polymerase, which can make interpreting electrophoretic output difficult (Jones *et al*. 1997; van Oosterhout *et al*. 2004). The magnitude and shape of stuttering patterns varies across loci, with some markers displaying very little stuttering, and others consistently producing two or more stutter peaks. Interpreting patterns at stutter-prone loci becomes particularly difficult in the case of adjacent-allele heterozygotes at loci having a dinucleotide repeat motif. Such stutter can cause these heterozygotes to be scored as

homozygotes for the larger allele. Consistent mistyping of this form will bias allele frequencies towards larger alleles, decrease observed heterozygosity and increase the apparent level of inbreeding at affected loci. It is also possible to mistype a true homozygote as an adjacent-allele heterozygote, but these errors are less likely if the marker has been screened sufficiently, and the shape of a single allele is well known.

Large-allele dropout is another potential scoring error that can bias allele and genotype frequencies. Large-allele dropout results from the preferential amplification of the smaller allele in a heterozygous genotype. As a result, the large allele may have a peak height much shorter than the small allele, and, if template quality is poor, may fail to amplify altogether (Björklund 2005). This phenomenon differs from allelic dropout in low-quality samples (Taberlet *et al*. 1996; Miller *et al*. 2002) in that large-allele dropout is a function of allele size and not stochastic sampling error of template. While investigating the source of errors in noninvasive samples, Buchan *et al*. (2005) found that allelic dropout increased with allele size and that rates of dropout remained high at some loci even with increased DNA concentration, indicating that technical limitations in amplifying large alleles may exist. Large-allele dropout is more prevalent in loci with large differences in allele sizes (Björklund 2005). If undetected, large-allele dropout will cause allele frequencies of shorter alleles to be overestimated and may result in rare large alleles being omitted from the data set altogether. Also, as with mistyping due to stuttering, large-allele dropout will decrease observed heterozygosity and increase the apparent level of inbreeding at affected loci.

Null alleles, the third form of scoring error considered here, are particularly difficult to detect because, by definition, nulls fail to produce a visible product (Dakin & Avise 2004), typically due to mutation at a priming site. When present in a data set, samples that are heterozygous for a null allele will be mistyped as homozygous for the alternate, visible allele, whereas samples that are homozygous for the null allele will appear to have a failed reaction. Similar to the scoring errors described above, null alleles will bias allele frequencies (visible alleles will be overestimated), decrease observed heterozygosity and increase the apparent level of inbreeding.

Compounded over multiple loci, even a small per-locus genotyping error rate can result in relatively large probabilities of a multilocus genotype containing at least one error (Creel *et al*. 2003; Bonin *et al*. 2004; Hoffman & Amos 2005), although error rates are rarely equal across loci, and dropping a single locus may provide a disproportionate decrease in error rate. Nonetheless, failing to account for the effects of these scoring errors may result in misinterpretations of the data when testing ecological and evolutionary hypotheses.

Preventing, mitigating and reporting scoring errors

Stringent protocols to prevent mistypes of any source are likely to be included in studies of human genetics and studies involving noninvasive or other low-quality samples, but rarely in studies involving high-quality samples from wild populations. Many of the procedures used in preventing mistypes are expensive and time-consuming (Tablerlet *et al*. 1996; Ewen *et al*. 2000; Bonin *et al*. 2004), and others are not appropriate for studies of species with poorly described genomes (e.g. when microsatellite primers are transferred across taxa) or when pedigree information is not available (preventing checks for Mendelian inheritance). Given the ubiquity of sources for scoring errors, however, implementation of quality-assurance procedures is warranted in any laboratory to insure reproducible and consistent microsatellite data. Below we discuss potential protocols to prevent, mitigate and report rates of scoring errors due to the three sources described above, and recommend standard procedures to be used in studies of wild populations. The following protocol involves implementing quality assurance procedures at six stages of a study: screening microsatellite loci, reanalysing a subset of samples, scoring the full data set, testing for scoring errors, mitigating for errors in downstream analyses, and reporting error rates.

Screening microsatellite loci

In the spirit of 'an ounce of prevention is worth a pound of cure', effective screening of microsatellite loci prior to data collection may mitigate the potential for errors to compromise the quality of a full data set. Screening a sufficient number of individuals for each marker provides information on peak pattern (amount of stuttering), levels of allelic variation (size range of alleles, a rough indicator for the potential of large-allele dropout), and allows suspect loci (those producing inconsistent or difficult to interpret patterns) to be further tested or omitted. It is difficult to overstate the importance of visual inspection of electrophoretic patterns during marker screening (rather than relying on automated scoring options), since stochastic amplifications, stutter patterns and low-height large alleles (indicative of large-allele dropout) may not be consistently detected in automated scoring outfiles. Indeed, the accuracy of automated allele-call programs during data collection relies on parameter settings determined during the screening process (see Scoring Data). Perhaps the most effective screening study involves progeny arrays, defined here as genotypes from full- or half-sib families and at least one known parent, and offers a powerful tool to detect scoring errors. In general, errors in scoring the parental genotype(s) will be indicated by actual genotypic distributions in the progeny failing to fit expected Mendelian ratios

(Gomes *et al*. 1999), including the observation of unexpected alleles (e.g. a parent that is heterozygous for a null allele may have progeny that appear homozygous for an allele not scored in that parent). Additionally, the potential for large-allele dropout can be tested using 'artificial' heterozygotes created by mixing template of two samples homozygous for different alleles (Wattier *et al*. 1998).

We recommend screening a sufficient number of samples to allow testing for scoring errors using available software (see Testing for Scoring Errors). If scoring errors due to null alleles, stuttering, or large-allele dropout are detected, primers should be redesigned and/or polymerase chain reaction (PCR) conditions optimized to reduce the overall error rate. Stuttering bands may be reduced by varying the reaction conditions (e.g. including additives such as formamide, bovine serum albumin, or dimethyl sulphoxide), redesigning the reverse primer (Brownstein *et al*. 1996), or adjusting the PCR programme by using touchdown or hot start techniques, reducing the number of cycles, or maintaining a stringent annealing temperature. Since different loci may benefit from opposite treatments (e.g. high or low concentrations of magnesium; short or long extension times), systematic testing of each variable will likely be necessary to reduce the number of stutter peaks. In addition, stuttering patterns may be minimized, and large-allele dropout prevented by using high-fidelity *Taq* or PCR kits designed especially for microsatellite analysis (e.g. Multiplex PCR Kit from QIAGEN). Successfully multiplexing microsatellite loci and reducing the need for re-amplifications of error-prone loci can reduce the total number of amplifications needed for a data set, compensating for the increased cost of these products. Ideally, loci containing null alleles, high error rates, or peak patterns that are difficult to interpret are dropped from the analysis. In reality, limited resources, time, or insufficient screening may prevent the redesigning of primers (Dakin & Avise 2004), or the goal of the research (e.g. fine-scale genome mapping) may require that suboptimal loci be included in the data set (Ewen *et al*. 2000). In these cases, accurately detecting scoring errors, and adjusting downstream analyses for their presence, as described below, is necessary in order to effectively test ecological or evolutionary hypotheses.

Reanalysing a subset of samples

Reanalysing (i.e. independent re-amplification and scoring of a sample) some or all samples in a study provides the opportunity to identify and quantify scoring errors. The most comprehensive approach to resolve a multilocus genotype is the 'multiple tube' approach (Taberlet *et al*. 1996), where the same sample is amplified multiple times for each marker, with the expectation that allelic dropout, false alleles and other amplification-based errors will be mitigated by comparing multiple products to obtain the true genotype. This method allows for the rate of mistypes to be estimated over all samples, and is most commonly used in studies involving noninvasive or ancient DNA samples. The cost of multiple replicates required to produce a genotype of certain reliability may be unrealistic for many laboratories, however, when sample sizes are large.

An alternative to the multiple-tube approach is to regenotype a subset of samples, which allows an error rate to be estimated without the cost of re-amplifying every sample (Bonin *et al*. 2004; Hoffman & Amos 2005). How many and which specific samples to include depends on the goal of the reanalyses. A random set of blind samples may be reanalysed for all loci in order to estimate an error rate at each locus and over the entire study (Bonin *et al*. 2004). Alternatively, only those loci prone to error may need to be re-amplified (Paetkau 2003; Hoffman & Amos 2005).

Given high-quality template DNA, we recommend reanalysing a random 10% of samples at all loci in order to provide basic estimates of error rates (see Reporting Error Rates). In addition, any samples found to contain suspect peak patterns (e.g. indicative of adjacent-allele heterozygotes at a stutter-prone locus) should be reanalysed in order to confirm the genotype at that locus. Template DNA of low quality or low concentration may require further reanalysis in order to determine the true multilocus genotype of each sample.

Scoring data

Sufficient screening of loci should reveal the expected range of allele sizes, characteristic peak patterns and potential scoring problems for each marker locus. Given this knowledge, we recommend combining automated allele calling with visual inspection of each sample. This process, while seemingly redundant, provides a balance between the efficiency and consistency of automated allele-calling software, and the accuracy provided by human inspection in detecting novel alleles outside the expected size range of a locus, stochastic amplifications within the size range, and potential mistypes due to stutter patterns or large-allele dropout.

Several programs are available to automate scoring of microsatellite data. GENOTYPER and GENEMAPPER (Applied Biosystems, Inc.) are capable of scoring alleles, and although GENOTYPER is more efficient for visual inspection of individual samples, GENEMAPPER may be more efficient when studying well-characterized loci in high-quality samples. TRUEALLELE (Cybergenetics, Inc.) is designed for allele calling in high-throughput systems. DECODE-GT (Pálsson *et al*. 1999) complements TRUEALLELE by quality-checking allele calls, reducing the amount of manual inspection required. It should be noted that automated allele calling has the potential to introduce additional scoring errors if the loci

are not sufficiently described during the screening process. Such errors are likely when parameters on automated scoring programs do not match the characteristics of a particular locus, especially when insufficiently stringent in allele calling (Pálsson *et al*. 1999) or binning (Ghosh *et al*. 1997). Inconsistent binning of alleles (required to convert the raw decimal data into integers) may result in scoring errors, especially if reaction conditions cause inconsistent nontemplate nucleotide additions by *Taq* polymerase (Smith *et al*. 1995). Although automated binning of data, such as that provided by MS EXCEL, will provide consistency across analyses, the program ALLELOGRAM (Manaster 2002) graphs each allele and bin, allowing visual inspection of data for binning errors.

Testing for scoring errors

Once all samples are scored at each locus, the resulting genotypic data set should be tested for scoring errors. Statistical tools are available to detect potential scoring errors from intact genotypes on a per-locus basis. Most operate by testing for heterozygote deficiencies, which can be indicative of genotyping errors (Paetkau 2003), particularly when identified in only a few loci (Gomes *et al*. 1999). Demographic or mating system processes such as a Wahlund effect or inbreeding are expected to result in excess homozygosity at all loci, whereas errors due to stuttering, large-allele dropout and null alleles should affect only a subset of loci.

Ewen *et al*. (2000) describes two methods to check for errors: concordance checking and Mendelian-inheritance checking. Although neither method will detect all types of errors, the Mendelian-inheritance checking, which requires pedigree information, is more robust and recommended for genotyping studies, whereas the concordance checking technique is recommended for fine-scale mapping studies (Ewen *et al*. 2000). Miller *et al*. (2002) developed a maximum-likelihood method to assess genotype reliability and help focus re-amplification efforts to error-prone loci. This method only detects errors due to allelic dropout, but has the potential to reduce the number of re-amplifications of low-quality samples from seven (Taberlet *et al*. 1996) to two replicates without severely decreasing the reliability of the resulting data (Miller *et al*. 2002). Large-allele dropout may be detected analytically by regressing allele-specific F_{IS} statistics on allele size, since the deficit of large allele heterozygotes should result in an excess of short allele homozygotes, and thus a negative slope (Wattier *et al*. 1998).

Some software has the capability to test for null alleles in addition to completing other analyses. In addition to a suite of individual- and population-level analyses, GDA (Lewis & Zaykin 2001) provides per-locus tests for Hardy–Weinberg equilibrium, deviations from which may indicate scoring errors. Thomas (2005) provides a concise summary of error-checking programs for genotype data sets of known pedigree. CERVUS (Marshall *et al*. 1998), a program for parentage analysis, can detect null alleles and estimate adjusted allele frequencies for progeny arrays.

Other programs, such as DROPOUT (Mckelvey & Schwartz 2005), which checks for mistypes in capture-mark-recapture studies, are more specialized. GIMLET (Valière 2002) and EXCEL MICROSATELLITE TOOLKIT (Park 2001), check for identical genotypes, and can be used to estimate error rates from reanalysed quality-assurance samples. MICRO-CHECKER (van Oosterhout *et al*. 2004) is designed to test for not only null alleles, but also to distinguish between scoring errors due to large-allele dropout and those due to stuttering, although it can be cumbersome for data sets containing large numbers of loci or populations.

As no single method will identify all types of errors, the most effective protocol will use multiple tests to detect scoring errors (Hoffman & Amos 2005). We recommend a two-level approach to detect errors: comparing individual genotypes for each reanalysed sample, and testing for scoring errors over the entire data set using at least one statistical program. When mistypes are identified in reanalysed samples, the potential cause of each scoring error should be determined by comparing electrophoretic files. Other samples having similar electrophoretic patterns to those mistyped should be reanalysed in order to confirm the genotype at the problem locus (or loci).

Mitigating errors in downstream analyses

If scoring errors are detected, it may be possible to significantly decrease error rates by reanalysing or dropping one or a few problematic loci. In addition, demographic or evolutionary processes should be considered as potential sources of heterozygote deficiencies if scoring errors are detected in most or all loci. Scoring errors are expected to affect only a few loci in well-screened data sets. If error-prone loci are included in hypothesis tests, downstream data analyses must be adjusted in order to prevent biases in biological conclusions.

Typically, errors due to stuttering are best mitigated by re-amplifying and re-scoring samples displaying potentially problematic peak patterns (e.g. indicative of adjacent-allele heterozygotes). Sufficient screening of stutter-prone loci and comparison of patterns among reanalysed samples may allow near-neighbour heterozygotes to be distinguished from homozygous peak patterns. In this case, careful and consistent manual scoring of electrophoretic patterns may reduce the number of scoring errors in the final data set.

Errors due to large-allele dropout can be mitigated by re-amplifying all samples homozygous for short alleles, although this approach can become expensive and time-consuming for some data sets. We recommend adjusting

allele frequencies at loci found to contain large-allele dropout following the method of Björklund (2005). In this procedure, the large allele is considered recessive to the short allele in heterozygotes. The frequency of the large allele is estimated from the number of homozygotes observed in the data set, assuming that all alleles at the locus are in Hardy–Weinberg equilibrium, and the frequency of other alleles are then adjusted to account for the change in the 'recessive' (i.e. large) allele (Björklund 2005).

If null alleles are detected, adjusted allele frequencies can be estimated (assuming Hardy–Weinberg equilibrium) for cases where all samples have at least one visible allele (Chakraborty *et al*. 1992), samples lacking a visible allele are thought to be artefacts (Brookfield 1996), or where null homozygotes are present in the data (Brookfield 1996). MICRO-CHECKER (van Oosterhout *et al*. 2004) provides adjusted allele frequencies based on each potential case. Estimates of adjusted allele frequencies can also be estimated for populations in Hardy–Weinberg disequilibrium when an accurate fixation index, independent of the affected locus, is available (van Oosterhout *et al*. 2006).

In order to minimize erroneous conclusions due to the presence of scoring errors, tests of ecological or evolutionary hypotheses should use the appropriate adjusted allele frequencies, and not the raw genotypic data, when possible. Although some group or population-level estimates require genotypic data, such as tests for recent population bottlenecks (Cornuet & Luikart 1996), others, including hierarchical *F*-statistics of populations or regions (excluding F_{IS} and F_{IT}), genetic distances, and isolation-by-distance measures, can be estimated from allele frequency data. The use of adjusted allele frequencies to calculate these estimates is hindered, however, by a lack of software accepting allele frequencies as input. As examples, GDA (Lewis & Zaykin 2001) does not accept allele frequency data, ARLEQUIN (Schneider *et al*. 2000) accepts allele frequency data only on a single-locus basis, BIOSYS (Swofford & Selander 1989) accepts allele frequency data but is limited in the number of alleles it can accept, and SGS (Degen *et al*. 2001) accepts allele frequency data but estimates only a limited number of parameters. The development of software that accept allele frequencies when appropriate will facilitate the reporting of scoring errors and allow more accurate hypothesis testing using microsatellite data from wild populations.

Typically, parameters estimated for individual population samples require genotypic data, and cannot be estimated from allele frequency data. These parameters include individual inbreeding or relatedness coefficients, spatial autocorrelation statistics, assignment of parentage, individual estimates of fertility and individual estimates of admixture and population assignment. Care must be taken to discuss the potential bias in biological conclusions based on data known to contain scoring errors (see below).

Reporting error rates

Several methods have been proposed to report scoring errors. Creel *et al*. (2003) estimate 'errors per allele tested' as the number of errors detected divided by the number of cases where an error could have been detected. This distinction measures only those cases where an error will result in an incorrect genotype, and may result in error rates that appear higher than those calculated over all alleles or genotypes. Ewen *et al*. (2000) in contrast, reports error rates as the number of genotypes containing an error, and Bonin *et al*. (2004) report error rates per allele and per genotype. Reporting errors in four statistics (errors per allele and per reaction, summarized for each locus and over all loci) will facilitate the comparison of error rates among studies (Hoffman & Amos 2005). Additionally, error rates should be used to evaluate the quality of the data, much as statistical tests are required to support conclusions (Bonin *et al*. 2004).

We recommend reporting the four statistics proposed by Hoffman & Amos (2005): errors per allele and per reaction, summarized for each locus and over all loci. These values are easily estimated from the random subset of samples reanalysed for quality assurance. Consistent reporting of these values will allow more objective review of error rates observed across studies, and will provide quantitative measures that can be used to infer the potential effects of error rates on the resulting biological conclusions. In addition, the results of each statistical test for scoring errors should be provided, and if errors are indicated, the data used in each downstream analysis (genotypic or adjusted allele frequency) should be clearly indicated, along with the method of adjustment, if appropriate. In addition, both raw and adjusted allele frequencies should be reported when adjusting for a scoring error (Björklund 2005).

Potential interpretive bias in biological conclusions

The procedures described above are designed to minimize the potential for scoring errors to affect biological conclusions based on microsatellite data. Conducting downstream statistical analyses on data sets containing scoring errors may lead to a misinterpretation of the data and erroneous ecological or evolutionary conclusions. The robustness of a statistical analysis to scoring errors varies with the hypothesis being tested. Estimates of population size are sensitive to genotyping errors (Mckelvey & Schwartz 2004). A consensus has not been reached regarding the effect of scoring errors on parentage analyses. Dakin & Avise (2004) concluded that the typical frequency of null alleles (less than 0.2) is 'unlikely to introduce serious bias into parentage' (p. 509). Others note that mistypes can be the basis to erroneously exclude parents, however, and show

that allowing up to three mismatches between genotypes can increase the rate of successfully assigning parentage (Marshall *et al*. 1998; Vandeputte *et al*. 2006). Most estimates of population differentiation, in contrast, are based on allele frequencies and not individual genotypes, and reasonable error rates (< 2% of genotypes mistyped) are unlikely to seriously bias these results (Bonin *et al*. 2004). Errors that produce an apparent excess of homozygotes, including the three scoring errors considered here, can bias estimates of some demographic parameters, in particular overestimating within-population inbreeding (Gomes *et al*. 1999).

It should be kept in mind, however, that programs to detect mistyping are not infallible and may falsely identify scoring errors with the rate of false positives depending on the algorithm employed. Failing to mitigate for falsely detected scoring errors is not expected to affect biological conclusions to the same extent as including error-laden data in downstream tests, since these procedures (e.g. using adjusted allele frequencies) typically force the affected locus into Hardy–Weinberg equilibrium, making downstream tests of biological hypotheses more conservative.

Despite the potential for scoring errors in microsatellite data to bias ecological and evolutionary conclusions, these markers are fast becoming a preferred tool for population genetics studies. In order to produce accurate analyses of wild populations, researchers must be familiar with the behaviour of each microsatellite locus and carefully consider potential errors in the resulting data set. Testing for, mitigating and reporting scoring errors due to stuttering, large-allele dropout and null alleles will help insure that studies appropriately harness the power of these markers.

References

Barker JS (2005) Population structure and host-plant specialization in two *Scaptodrosophila* flower-breeding species. *Heredity*, **94**, 129–138.

Björklund M (2005) A method for adjusting allele frequencies in the case of microstellite allele drop-out. *Molecular Ecology Notes*, **5**, 676–679.

Bonin A, Bellemain E, Bronken Eidesen P, Pompanon F, Brochmann C, Taberlet P (2004) How to track and assess genotyping errors in population genetic studies. *Molecular Ecology*, **13**, 3261–3273.

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

Brownstein MJ, Carpten JD, Smith JR (1996) Modulation of non-templated nucleotide addition by *Taq* DNA polymerase: primer modifications that facilitate genotyping. *BioTechniques*, **20**, 1004–1010.

Buchan JC, Archie EA, Van Horn RC, Moss CJ, Alberts SC (2005) Locus effects and sources of error in noninvasive genotyping. *Molecular Ecology Notes*, **5**, 680–683.

Chakraborty R, De Andrade M, Diager 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.

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, **144**, 2001–2014.

Creel S, Spong G, Sands JL *et al.* (2003) Population size estimation in yellowstone wolves with error-prone noninvasive microsatellite genotypes. *Molecular Ecology*, **12**, 2003–2009.

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

Degen B, Petit R, Kremer A (2001) SGS: SPATIAL GENETIC SOFTWARE: A computer program for analysis of spatial genetic and phenotypic structures of individuals and populations. *Journal of Heredity*, **92**, 447–448.

Ewen KR, Bahlo M, Treloar SA *et al.* (2000) Identification and analysis of error types in high-throughput genotyping. *American Journal of Human Genetics*, **67**, 727–736.

Ghosh S, Karanjawala ZE, Hauser ER *et al*. Fusion Study Group (1997) Methods for precise sizing, automated binning of alleles, and reduction of error rates in large-scale genotyping using fluorescently labeled dinucleotide markers. *Genome Research*, **7**, 165–178.

Gomes I, Collins A, Lonjou C *et al*. (1999) Hardy–Weinberg quality control. *Annals of Human Genetics*, **63**, 535–538.

Hoffman JI, Amos W (2005) Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. *Molecular Ecology*, **14**, 599–612.

Jones CJ, Edwards KJ, Castaglione S *et al.* (1997) Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Molecular Breeding*, **3**, 381–390.

Lewis PO, Zaykin D (2001) GENETIC DATA ANALYSIS: computer program for the analysis of allelic data. Free program distributed by the authors over the internet from: http://lewis.eeb.uconn.edu/lewishome/software.html.

Manaster CJ (2002) ALLELOGRAM: a program for normalizing and binning microsatellite genotypes. Available from the author through the website http://s92417348.onlinehome.us/software/allelogram/.

Marshall TC, Slate J, Kruuk LE, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology*, **7**, 639–655.

Mckelvey K, Schwartz MK (2004) Providing reliable and accurate genetic mark-recapture in a cost-effective manner. *Journal of Wildlife Management*, **68**, 453–456.

Mckelvey K, Schwartz MK (2005) DROPOUT: a program to identify problem loci and samples for noninvasive genetic samples in a capture-mark-recapture framework. *Molecular Ecology Notes*, **5**, 716–718.

Miller CR, Joyce P, Waits LP (2002) Assessing allelic dropout and genotype reliability using maximum likelihood. *Genetics*, **160**, 357–366.

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

van Oosterhout C, Weetman D, Hutchinson WF (2006) Estimation and adjustment of microsatellite null alleles in nonequilibrium populations. *Molecular Ecology Notes*, **6**, 255–256.

Paetkau D (2003) An empirical exploration of data quality in DNA-based population inventories. *Molecular Ecology*, **12**, 1357–1387.

Pálsson B, Pálsson F, Perlin M, Gudbjartsson H, Stefánsson K, Gulcher J (1999) Using quality measures to facilitate allele

calling in high-throughput genotyping. *Genome Research*, **9**, 1002–1012.

Park SDE (2001) *Trypanotolerance in West African cattle and the population genetic effects of selection*. PhD Thesis, University of Dublin.

Schneider S, Roessli D, Excoffier L (2000) ARLEQUIN: *A software for population genetics data analysis*. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Geneva, Switzerland.

Smith JR, Carpten JD, Brownstein MJ *et al.* (1995) Approach to genotyping errors caused by nontemplated nucleotide addition by *Taq* DNA polymerase. *Genome Research*, **5**, 312–317.

Swofford DL, Selander RB (1989) BIOSYS-*1: A computer program for the analysis of allelic variation in population genetics and biochemical systematics, release 1.7 edition*. Illinois Natural History Survey, Champaign, Illinois.

Taberlet P, Griffin S, Goossens B *et al.* (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research*, **24**, 3189–1394.

Thomas A (2005) GMCHECK: Bayesian error checking for pedigree genotypes and phenotypes. *Bioinformatics*, **21**, 3187–3188.

Valière N (2002) GIMLET: a computer program for analysing genetic individual identification data. *Molecular Ecology Notes*, **2**, 377–379.

Vandeputte M, Mauger S, Dupont-Nivet M (2006) An evaluation of allowing for mismatches as a way to manage genotyping errors in parentage assignment by exclusion. *Molecular Ecology Notes*, **6**, 265–267.

Wattier R, Engel CR, Saumitou-Laprade P, Valero M (1998) Short allele dominance as a source of heterozygote deficiency at microsatellite loci: experimental evidence at the dinucleotide locus Gv1CT in *Gracilaria gracilis* (Rhodophyta). *Molecular Ecology*, **7**, 1569–1573.