

GENOTYPING ERRORS: CAUSES, CONSEQUENCES AND SOLUTIONS

François Pompanon, Aurélie Bonin, Eva Bellemain and Pierre Taberlet

Abstract | Although genotyping errors affect most data and can markedly influence the biological conclusions of a study, they are too often neglected. Errors have various causes, but their occurrence and effect can be limited by considering these causes in the production and analysis of the data. Procedures that have been developed for dealing with errors in linkage studies, forensic analyses and non-invasive genotyping should be applied more broadly to any genetic study. We propose a protocol for estimating error rates and recommend that these measures be systemically reported to attest the reliability of published genotyping studies.

PATERNITY EXCLUSION

The elimination of a male as the potential father of a given offspring, owing to incompatibility between the multilocus genotypes of the two individuals concerned.

NON-INVASIVE GENOTYPING

Genotyping from samples that are collected without capturing the animal (such as hair or faeces).

AMPLIFIED FRAGMENT-LENGTH POLYMORPHISMS

A PCR-based DNA fingerprinting technique that reveals polymorphisms in restriction-enzyme recognition sites by generating dozens of dominant marker bands.

Laboratoire d'Ecologie Alpine, UMR CNRS 5553, Université Joseph Fourier, B.P. 53, 38041 Grenoble Cedex 9, France.

Correspondence to F.P. e-mail: francois.pompanon@ujf-grenoble.fr

doi:10.1038/nrg1707
Published online
27 September 2005

In 1997 a genotyping study revealed a striking new model for chimpanzee mating behaviour by indicating that half the offspring of a community were sired by males from outside the group¹. It soon turned out that this conclusion² resulted from genotyping errors that led to erroneous PATERNITY EXCLUSION. This is just one example of the serious effect that such errors can have on important biological issues.

A genotyping error occurs when the observed genotype of an individual does not correspond to the true genotype³. Although genotyping errors occur in all but the smallest data sets that are generated in genetic studies, they have almost exclusively been recognized in linkage analyses in humans, in forensic analysis and in NON-INVASIVE GENOTYPING. They were considered in these research areas because independent approaches pointed out the inconsistencies of some genotypes with other evidence, such as known pedigrees in linkage analysis. In 1976 Thompson was one of the first to note that mismatches in pedigree data could result from laboratory errors⁴.

An error can be detected as a discrepancy between the genotype of an experimental sample and a known genotype that has been inferred, for example, from replicate genotyping (multiple-tube approach)⁵. An error can also be spotted if the experimental genotype is incompatible with reliable independent evidence, such as pedigree data. However, few studies in population genetics and evolution quantify the rate of genotyping

error^{3,6} that would ensure the reliability of the inferred biological conclusion. Moreover, there is no consensus strategy or strict standard for limiting or quantifying the occurrence of the main types of error⁷. A bibliographic survey indicates that an increasing number of researchers are aware of this difficulty (FIG. 1), but that the effect of genotyping errors still remains neglected. This is an important issue if we consider that all studies in which errors were checked reported a non-negligible error rate (from 0.2% to more than 15% per locus), and that a rate of between 0.5% and 1% is usual in many laboratories^{8,9}. Even higher error rates are known to occur in studies that involve DNA of poor quality or quantity^{10,11}. A realistic estimate of the proportion of SNPs in public databases that were not confirmed by subsequent studies is 16%. This is the result of sequencing errors or low allele frequencies¹². The fact that an error rate as low as 0.5–1.0% has the potential to obscure medically important findings¹³ reinforces the need to confront this issue.

The aim of this paper is to examine the causes and consequences of genotyping errors and to give recommendations to limit their occurrence and their effect on the resulting biological message. All types of molecular marker are prone to genotyping error, including sequence data¹⁴, but here we focus on the most commonly used markers¹⁵: AMPLIFIED FRAGMENT-LENGTH POLYMORPHISMS (AFLPs), MICROSATELLITES and SNPs (TABLE 1). We do not consider DNA sequence

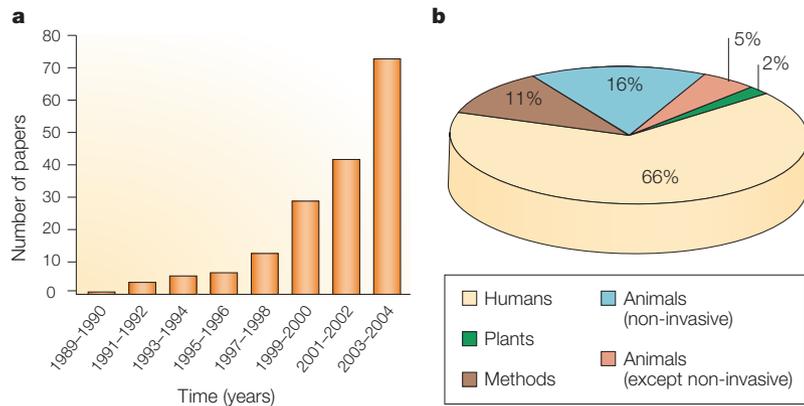


Figure 1 | The recent increase in the number of papers that deal with genotyping errors. **a** | The trend in the number of papers that mention genotyping errors since 1989. **b** | Their repartition according to subject area. The figure represents the result of a search on *ISI Web of Science* in April 2005, with keywords: genotyp* error* OR allelic dropout OR false allele*. Subject categories are as follows. Humans: papers that deal with human genetics, including some that are clearly methodological. Animals (non-invasive): papers that deal with non-invasive genetic sampling in animals. Animals (except non-invasive): papers that deal with animals, without reference to non-invasive genetic sampling. Plants: papers that deal with plants. Methods: methodological papers, without any obvious reference to the type of data (human, animal or plant).

errors that have a bearing on specific methods that are related to phylogenetic analysis^{4,16}. However, the main principles enounced here apply to all genetic markers, including DNA sequences.

Causes of genotyping errors

An extensive survey of the literature and our own experience show that genotyping errors result from diverse, complex and sometimes cryptic origins. When an error is detected, the first difficulty is to clearly identify its cause, so that the experimental protocols can

be improved to reduce error rates. Grouping errors into discrete categories according to their causes is challenging because different causes sometimes interact to generate an error. For clarity, we propose to group them into four categories: errors that are linked to the DNA sequence itself, errors that are due to the low quality or quantity of the DNA, biochemical artefacts and human factors. Below we develop one example for each category. A more extensive survey is given in TABLE 2.

Variation in DNA sequence. An error that is linked to the DNA sequence can be generated by a mutation close to a marker, if this flanking sequence is involved in the marker-detection process. In microsatellite studies, the most common error of this type is the occurrence of null alleles^{17,18}. This corresponds to the failure to amplify an allele due to a mutation that is located on the complementary sequence of one of the primers, which prevents efficient amplification. Usually, only substitutions close to the 3' end of the primer or insertions or deletions cause problems. An insertion or deletion that is close to a microsatellite marker can also generate *SIZE HOMOPLASMY*, which leads to the scoring of two different alleles as a single one.

Low quantity or quality of DNA. Low DNA quantity and/or quality are known to promote genotyping errors. A low number of target DNA molecules in an extract results from either extreme dilution of the DNA or from degradation, which leaves only a few intact molecules. Both these conditions favour *ALLELIC DROPOUTS* and *FALSE ALLELES*¹⁰. They also markedly increase the risk of contamination¹⁰, because contaminant molecules have a higher probability of being amplified when the number of template DNA molecules is low.

MICROSATELLITE

A class of repetitive DNA that is made up of repeats that are 2-8 nucleotides in length. They can be highly polymorphic and are frequently used as molecular markers in population genetics studies.

SIZE HOMOPLASMY

The generation of alleles that are the same size which are not the result of common ancestry (not homologous), but arose independently in different ancestors by parallel or convergent mutations.

ALLELIC DROPOUTS

The stochastic non-amplification of an allele; that is, amplification of only one of the two alleles present at a heterozygous locus.

FALSE ALLELE

An allele-like artefact that is generated by PCR.

Table 1 | **Diversity of molecular markers in genetic studies**

Extracted information	Common use	Research area
Microsatellites – multiallelic, codominant and highly variable		
Single-locus genotypes	Association studies	Medicine; agronomy
Multilocus genotypes	Individual identification; kinship studies; assignment tests	Population genetics; population biology; forensic investigations
Allele frequencies	Population structure assessment; population size and gene-flow estimation	Population genetics; population or conservation biology; evolutionary biology
SNPs – biallelic, usually codominant and highly abundant across genomes		
Multilocus genotypes	Individual identification	Population genetics; population biology
Allele frequencies	Estimation of genetic diversity	Population genetics; population or conservation biology; evolutionary biology
Pattern of segregation in crosses; linkage disequilibrium between markers	QTL or gene mapping; linkage studies	Medicine; evolutionary biology; agronomy; population genomics
AFLPs – biallelic, dominant and highly abundant across genomes		
Multilocus genotypes	Individual identification	Population genetics; population biology
Allele frequencies	Estimation of genetic diversity; population structure assessment	Phylogeography; population genetics; evolutionary biology
Pattern of segregation in crosses; linkage disequilibrium between markers	QTL or gene mapping; linkage studies	Medicine; evolutionary biology; agronomy

Table 2 | **Classification of errors according to their main cause**

Cause of error	Mechanism of error occurrence	Consequence of the error for the genotype	Reference(s)
Interactions between DNA molecules			
DNA sequence flanking the marker	No amplification (or less efficient amplification) because of a mutation in the target primer sequence	Null allele	18,36,85,86
DNA sequence flanking the marker	Insertion or deletion in the amplified fragment	Size homoplasmy of different alleles	87
DNA sequence flanking the marker	In heterozygous individuals, preferential amplification of one allele when its denaturation is favoured (because of low GC content)	Allelic dropout	85
Sample quality			
Low quality or quantity of DNA	In heterozygous individuals, amplification of only one allele	Allelic dropout	2,10,88
Low quality or quantity of DNA	In heterozygous individuals, preferential amplification of the shorter allele	Short allele dominance (preferential long allele dropout)	85,89,90,91
Contamination of the DNA extract	Amplification of a contaminant allele	Mistaken allele	92
Low extract quality	No restriction (or less restriction) or amplification that is due to inhibitors	Allelic dropout	2,23,91,93,94
Biochemical artefacts and equipment			
Low quality reagents	No restriction (or less restriction) or amplification that is due to inhibitors	Allelic dropout; mistaken allele	NR
Low quality reagents	Poor fragment labelling and detection	Allelic dropout; mistaken allele	NR
Poor equipment precision or reliability	Examples include stochastic pipetting, evaporation during PCR, and poor fluorescent label detection	Allelic dropout; mistaken allele	NR
<i>Taq</i> polymerase errors	Slippage in the first steps of the PCR	False allele	10,95,96,97
<i>Taq</i> polymerase errors	Incomplete addition of extra adenine residues at the 3' end of amplified fragments	False allele	19,20
Lack of specificity	Amplification of non-specific products that is due to annealing of the primer to another locus	Mistaken allele	91
Lack of specificity	Non-specific restriction reactions	Mistaken allele	98
Electrophoresis artefact	Inconsistency of allele size between different experiments, devices or studies (for example, capillary versus manual electrophoresis or fluorescence versus radioactive detection)	Size homoplasmy of different alleles; mistaken allele	22,69,99,100
Electrophoresis artefact	Distortion of the allele size by factors that alter the migration (for example, temperature or high concentration of PCR products)	Size homoplasmy of different alleles; mistaken allele	101,102
Human factor			
Sample manipulation	Confusion between samples (for example, mislabelling or tube mixing)	Mistaken allele(s)	41
Experimental error	Contamination with an exogenous DNA or cross-contamination between samples	Mistaken allele(s)	103,104,105
Experimental error	Use of an inappropriate protocol (for example, reactant not added; incorrect T_m , primers or concentration of reactants)	Allelic dropout; mistaken allele(s)	NR
Data handling	Misreading of the profile or misidentification of the fluorescence peak	Mistaken allele	3,6,41,104
Data handling	Miscopying or confusion of the genotypes in the database	Mistaken allele	6,106,107
Data handling	Computing data (for example, bug in the database or analysis program)	Mistaken allele	NR

AFLP, amplified fragment-length polymorphism; false allele, allele-like PCR-generated artefact¹⁰; null allele, a non-amplifying allele that is due to a mutation in the primer target sequence¹⁷; allelic dropout, the stochastic non-amplification of an allele, that is, amplification of only one of the two alleles present at a heterozygous locus²; mistaken allele, an allele that does not correspond to the true allele, excluding the null allele, allelic dropout and false allele. NR, not reported in the literature to our knowledge, but widely recognized; T_m , melting temperature.

Biochemical artefacts. At the end of the elongation step of a PCR, the *Taq* polymerase has a tendency to add a non-templated nucleotide (usually an adenine) to the 3' end of the newly synthesized strand^{19,20}. This '+A artefact' is common, and creates an artefactual band or peak on the readout gel or trace, respectively. The relative proportions of the true fragment and the +A artefactual fragment are very sensitive to the sequence of the 5' end of the primer used in the genotyping assay, but also to PCR conditions and to the long elongation times that promote the +A artefact. In such a context, this biochemical artefact represents an important cause of genotyping error.

Human error. Unexpectedly, in the few studies designed to analyse the precise causes of genotyping error, the main cause was related to human factors. In their impressive study on microsatellite genotyping errors used in paternal exclusion in the Antarctic fur seal, Hoffman and Amos attributed 80.0%, 10.7%, and 6.7% of the errors to scoring, data input and allelic dropouts, in corresponding order⁶. The remaining 2.7% probably resulted from sample mix-up, pipetting error or contamination. This means that human factors were responsible for about 93% of the errors in this study. Admittedly, part of the error detected in this work resulted from the manual scoring of autoradiographs, a practice that is going out of use. However, scoring errors might also be an important issue in the automated and semi-automated scoring of fluorescence profiles^{3,21}. For example, human subjectivity during manual scoring represented the main source of discrepancy between the AFLP data sets that were generated by independent scorers who were using the same electropherograms³. This means that the expertise and standards of the researcher have a bearing on the selection of AFLP loci. ALLELE CALLING has also been identified as a potential problem in SNP studies²². Therefore, among the various causes of error, allele calling might be the most important difficulty. Obviously, the risk of human scoring error strongly depends on the quality of the data.

Quantifying genotyping errors

The most common metric for quantifying genotyping errors is the error rate per locus, but several other estimates are commonly used, such as the error rate per PCR, per allele or per multilocus genotype. All these metrics measure the proportion of mismatches between REPLICATED GENOTYPES and implicitly involve the comparison to a reference genotype (BOX 1).

Because errors are not randomly distributed across PCRs, alleles or loci, the link between these metrics is not straightforward. In a study of brown bears that involved 18 microsatellites, a 0.8% error rate per locus should theoretically have given a multilocus genotype error rate of 25.1%. However, in practice the multilocus error rate was only 17.6% because errors did not occur independently³. This emphasizes the need for using common metrics to allow comparison between studies. But different metrics are not equally appropriate across different studies. When measuring the error rate

per locus, allelic dropouts are less likely to be detected at homozygous loci (a heterozygous locus affected by allelic dropout and a *bona fide* homozygous locus will both appear as a single band or peak), and therefore rates are not comparable between loci or populations that vary in heterozygosity²³. Calculating the error rates per PCR and per allele makes no sense for AFLP studies because a single PCR generates many dominant alleles, each one characterized by the presence of a single fragment. The error rate per multilocus genotype is meaningful for individual identification, population assignment, kinship studies and census studies because it reflects the reliability of the genotypes obtained. This estimate increases with the number of loci²⁴, and a relatively low error rate per locus can generate a high error rate per multilocus genotype, which might not be compatible with the scientific question. The error rate for a particular allele or locus provides complementary information, and helps to identify error-prone loci that can be removed from the study to increase its reliability. For example, Bonin *et al.* showed, for AFLP data, that the mean error rate per locus can drop from 3.4% to 2.0% by removing the 7 (out of 222) polymorphic loci that had the greatest error rate³. Usually some loci are more error-prone than others. For microsatellite markers, it has been demonstrated that the number of errors is directly correlated with the size of the PCR product⁶.

To summarize, the most universal metric is the error rate per locus. It gives an idea of the reliability of the laboratory protocol and of the experimental procedure, allowing comparisons to be made between studies and different types of marker.

However, the true error rate might be higher than the estimated rate. In SNP studies, the error-detection rate that has been estimated using trio designs (that is, involving the genotypes of the mother, father and offspring) and has been based on Mendelian inheritance does not exceed 61% of existing errors^{25,26}. This difference between the true and the estimated error rates is mainly due to errors that are 'Mendelian compatible' (that is, errors that produce genotypes that are consistent with Mendelian inheritance among relatives). Unfortunately, many studies that are based on pedigree data are only checked for compatibility with Mendelian inheritance. Ewen *et al.* called for a more realistic approach to identifying other types of error, and highlighted the need for a consensus strategy not only based on Mendelian verification, but also on complementary methods such as duplicate samples and independent allele calling⁷. In addition, even within a pedigree, true genotypes might not be Mendelian compatible if a mutation has occurred, as it might when large data sets are being analysed. Moreover, by nature some errors are almost undetectable. An example of this would be when two identical genotypes result from different mutations, as might occur when an insertion or deletion at different sites along a DNA sequence generates size homoplasy among PCR products (TABLE 2). Finally, it is always difficult and sometimes impossible to distinguish between errors, mutations and rare alleles in population studies.

ALLELE CALLING

The determination of an allele from an electropherogram or a fluorescent profile.

REPLICATED GENOTYPES

Genotypes that are produced from different (preferentially independent) samples from the same individual.

Box 1 | **Quantifying error rates**

Different estimates, based on replicates within a data set, have been defined to quantify error rates. Some metrics have been proposed for specific errors such as allelic dropouts or false alleles⁸⁰. More global metrics, which take into account all types of detectable genotyping error, are also commonly used, although they have never been explicitly defined.

In this box we indicate how to estimate error rates at the allelic, locus, multilocus and reaction levels. First, a reference genotype must be defined as the genotype that minimizes the number of errors in comparisons between replicates. Several reference genotypes might exist. If only two replicates are carried out and give contradictory genotypes, either one or the other can be considered to be the reference. The calculation of error rates is based on the number of mismatches between the reference genotype and the replicates. Here we consider a case where n individual single-locus genotypes have been replicated t times. For diploid individuals, $2nt$ alleles and nt loci are typed and can be compared with the reference.

The following formulae are valid for codominant markers, but can be extended to dominant markers such as amplified fragment-length polymorphisms (AFLPs) by considering phenotypic mismatches instead of allelic mismatches, and phenotypes (the presence or absence of a fragment) instead of genotypes.

Mean error rate per allele

The mean allelic error rate, e_a , is the ratio between m_a , the number of allelic mismatches, and $2nt$, the number of replicated alleles.

$$e_a = \frac{m_a}{2nt} \tag{1}$$

Mean error rate per locus

The mean error rate per locus, e_l , is the ratio between m_l , the number of single-locus genotypes including at least one allelic mismatch, and nt , the number of replicated single-locus genotypes.

$$e_l = \frac{m_l}{nt} \tag{2}$$

This metric can also be estimated for each particular locus, to help identify error-prone loci.

Error rate per multilocus genotype

The observed error rate per multilocus genotype, e_{obs} , is the ratio between m_g , the number of multilocus genotypes including at least one allelic mismatch, and nt , the number of replicated multilocus genotypes.

$$e_{obs} = \frac{m_g}{nt} \tag{3}$$

If genotyping errors occur independently among l loci (which is unlikely), the error rate per multilocus genotype, e_{ind} , is deduced from the single-locus error rate, e_i , at each locus, i :

$$e_{ind} = 1 - \prod_{i=1}^l (1 - e_i) \tag{4}$$

Error rate per reaction

The error rate per reaction, e_r , is the ratio between m_l , the number of single-locus genotypes including at least one allelic mismatch, and r , the total number of reactions.

$$e_r = \frac{m_l}{r} \tag{5}$$

This metric is equivalent to the mean error rate per locus when the PCR reaction involves one locus or to the multilocus error rate when all loci are amplified in a single multiplex reaction.

The following table shows the estimation of the error rates per allele and per locus, for four replicates ($t = 4$) of three individuals ($n = 3$).

Individual	Allele	Replicate 1	Replicate 2	Replicate 3	Replicate 4	Reference genotype	Error rate per allele (mean = 1/4)	Error rate per locus (mean = 5/12)
1	1	A	A	B	A	A	3/8	2/4
	2	A	B	C	A	A		
2	1	A	B	B	A	A or B	2/8	2/4
	2	B	B	B	B	B		
3	1	A	A	A	A	A	1/8	1/4
	2	C	C	B	C	C		

Box 2 | **Genotyping errors and their effects: a case study**

Case study (goal and methods)

In this study, bear (*Ursus arctos*) tissues were genotyped to establish pedigrees and study sexually selected infanticide⁸¹. Eighteen microsatellite loci^{11,82} were amplified following the protocol described in Waits *et al.*⁸³. The genotyping error rate in the data set was calculated by blind replication of ~3.5% of the amplifications (34 of the 977 samples) and was estimated to be 0.8% (the mean error rate per locus), due to allelic dropouts and mistaken alleles³.

Microsatellite genotyping and error consequences

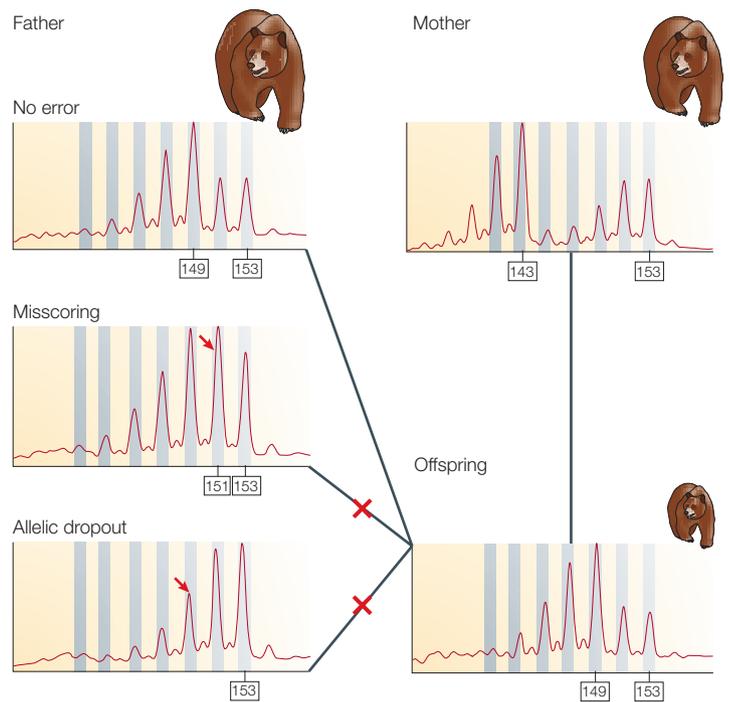
The scoring of a microsatellite allele depends on the profile of this microsatellite, and requires strict rules to be defined in advance. Typical microsatellite profiles are characterized by a succession of peaks that have growing intensity due to STUTTER BANDS¹⁰. The figure illustrates two types of genotyping error that are likely to induce false paternity and consequently bias the biological conclusions. In this hypothetical example, let us assume that the male is the real father of the offspring (see figure). The real male genotype is 149–153 ('no error'): the offspring inherited allele 153 from his mother and allele 149 from his father. In the other two cases the male would incorrectly be excluded as the father of the offspring, either because of a scoring error (misscoring) or because of allelic dropout in the male genotype. Allele 149 is mistyped as allele 151 in the first case (probably owing to overlapping peaks between the two alleles) and is missing in the second case (allelic dropout). Those errors have arisen when typing DNA that is extracted from tissues (that is, DNA which is presumed to be of good quality).

Dealing with errors, and some recommendations

Generally, the number of typed loci should be chosen as a compromise between the probability of identity⁸² and the probability of error. The more loci that are typed, the lower the probability of identity and the higher the multilocus genotyping error rate²⁴. In our study, with 18 typed loci, the probability of identity was low even for relatives (3.1×10^{-17} ; for siblings

2.4×10^{-7}) — allowing unambiguous distinction between individuals — and the error rate was 0.8% per locus. This error rate per locus indicates the occurrence of one error in every four or five multilocus genotypes. Because an incompatibility in the comparison of the trio (offspring–mother–father) genotypes was highly probable, we allowed one genetic mismatch in the parentage analysis, using the software PARENTE⁸⁴.

More generally, to avoid this type of incorrect paternity exclusion we recommend allowing one or several genetic mismatches in the parentage analysis. This number will depend on the probability of identity (which in turn depends on the allele frequencies in the population) and on the calculated error rate.



Consequences of genotyping errors

Linkage and association studies. Erroneous genotypes might markedly affect linkage and association studies by masking the true segregation of alleles. The effect on the results is measured by experimental or simulation studies and can be serious for even low rates of error (for example, < 3%)²⁷.

For example, in linkage studies genotyping errors can affect HAPLOTYPE frequencies²⁸ and eventually lead to an inflation of genetic map lengths^{29–31}. Error rates as low as 3% can have serious effects on linkage-disequilibrium analysis²⁷, and a 1% error rate can generate a loss of 53–58% of the linkage information for

a trait locus³². However, modest error rates might be tolerable in situations that do not involve rare alleles, as in QTL studies¹³.

In association studies, because recombination is rare, errors mostly affect non-recombinant genotypes, which are then erroneously interpreted as being the result of recombination. Errors, therefore, decrease the power for detecting associations^{8,13,33,34}.

The importance of the experimental design also needs to be emphasized as it can generate errors that are not randomly distributed across phenotypes (these are known as differential errors). This can occur when cases and controls are genotyped in different

STUTTER BANDS
Artefacts that occur during the PCR amplification of microsatellites.

HAPLOTYPE
The combination of alleles found at neighbouring loci on a single chromosome or haploid DNA molecule.

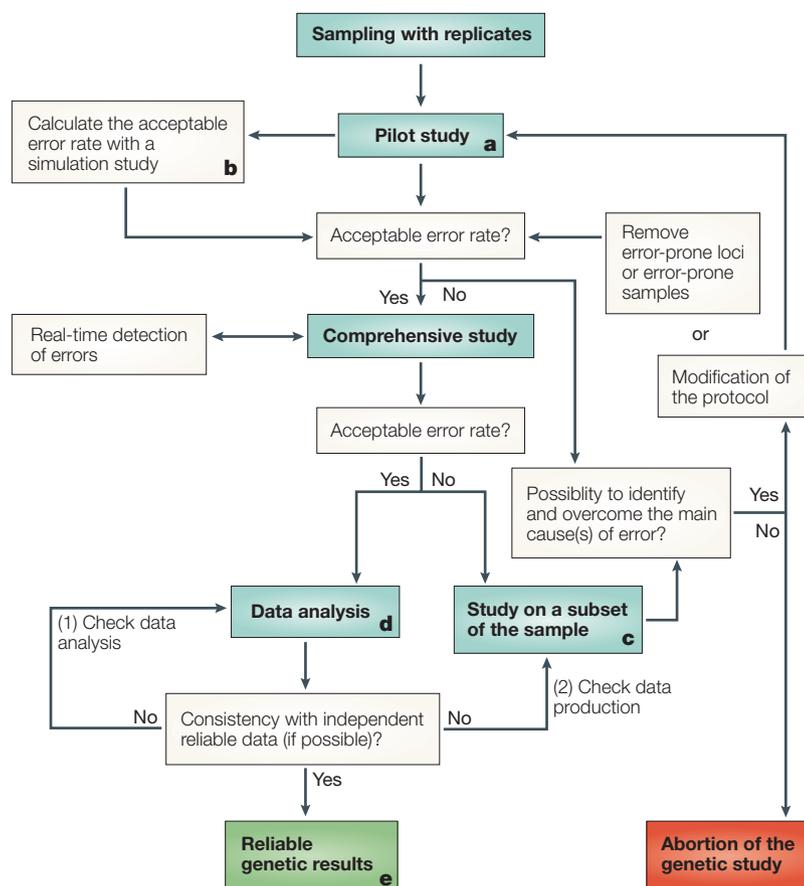


Figure 2 | Flow chart that shows the important steps in a genotyping process for limiting the occurrence and effect of genotyping errors. The steps that end with a superscript letter (a–e) should be qualified as follows: **a** | The goal is to estimate the error rate associated with the samples, the method and the protocol used. This is done by replicating a sufficient number of samples. **b** | Deciding on an acceptable error rate depends on the error rate, the purpose of the genetic study, the genotyping method used, the ability to detect eventual errors and the cost in terms of money and time. **c** | The control study aims to find the cause of errors that did not exist in the pilot study. **d** | The calculated error rate must be considered in the data analysis. **e** | The results should be published with a reliability index that is based on the error rate measured.

assays during the investigation of the genetic basis of a disease³⁵. Differential and non-differential errors can have opposite consequences on the rate at which false positives are detected in statistical tests of association.

Individual identification. Genotyping errors can strongly affect individual identification studies that are based on multilocus genotypes by erroneously increasing the number of genotypes that are observed in a population sample. In census studies of rare or elusive species, the population size can be estimated on the basis of the genotypes that are identified from non-invasive samples (such as hair or faeces) that are collected in the field. In this context, genotyping errors can lead to a serious overestimate of population size^{24,36}. A 200% overestimate of population size was found with a 5% error rate per locus when using 7 to 10 loci for genotype identification²³. Such an overestimate obviously increases with the number of loci and with the number of samples per genotype²⁴.

Genotyping errors also have an effect on parentage analysis, as they can generate incorrect paternity or maternity exclusion^{2,36–38}(BOX 2). Such information on population size and structure is required in a field such as conservation biology, because inaccurate estimates that are caused by genotyping errors can result in incorrect decisions being made in population management. In forensic DNA analyses, a false multilocus genotype can prevent the identification of a corpse or lead to erroneous identification (or exoneration) of criminal offenders³⁹.

Population genetic studies. Most population genetics studies that take genotyping error into account use non-invasive samples, which are error-prone because of the low quality and/or quantity of DNA. However, it has been demonstrated that even with high-quality DNA the error rate might not be negligible. Measurements on DNA tissue extracts from Antarctic fur seals⁶, as well as from brown bears³, detected an error rate of up to 0.8% per microsatellite locus (BOX 2). The effect of genotyping error remains largely unknown in this field, because few studies have dealt with this topic until now^{3,6,37,38}. Genotyping errors might lead to incorrect allele identification or incorrect allele frequencies, resulting in incorrect F_{ST} ESTIMATES, false migration rates, or false detection of selection or POPULATION BOTTLENECKS. Analyses that are based on allele frequencies will not be as affected by errors as those that are based on individual identification⁵ (for example, parentage analysis), but they will be sensitive to sampling effects.

In population genetics, the effect of scoring differences might seem to be less than in other kinds of study. This has been demonstrated by an AFLP data set that was scored by two scientists. The two scorers shared only 38% of the marker loci, but the same biological conclusions about the genetic structure of the population were extracted from the data³. In this study, the robustness of the inferred biological message to scoring differences was certainly due to the redundancy of the information contained in the large number of AFLP markers (more than 200 polymorphic loci were screened in total). However, population genomics studies that follow selected markers among several hundred markers⁴⁰ would be sensitive to the effect of genotyping error, especially if the errors were population-specific. There is a great need for studies on the effect of genotyping error in this emerging field.

How to limit genotyping errors and their effect

Unarguably the worst situation is to realize at the end of a study that the data were not reliable owing to genotyping errors, and that the data set is not retrievable. Such situations are almost never reported in the literature, but their occurrence is probably not rare. Therefore, it is important to take into account the possibility of genotyping errors when designing an experimental protocol.

F_{ST} ESTIMATES

Statistics that were first defined by Sewall Wright to describe the genetic structure at different hierarchical levels (individuals, subpopulations and total populations).

POPULATION BOTTLENECK

A marked reduction in population size that often results in the loss of genetic variation and more frequent matings among closely related individuals.

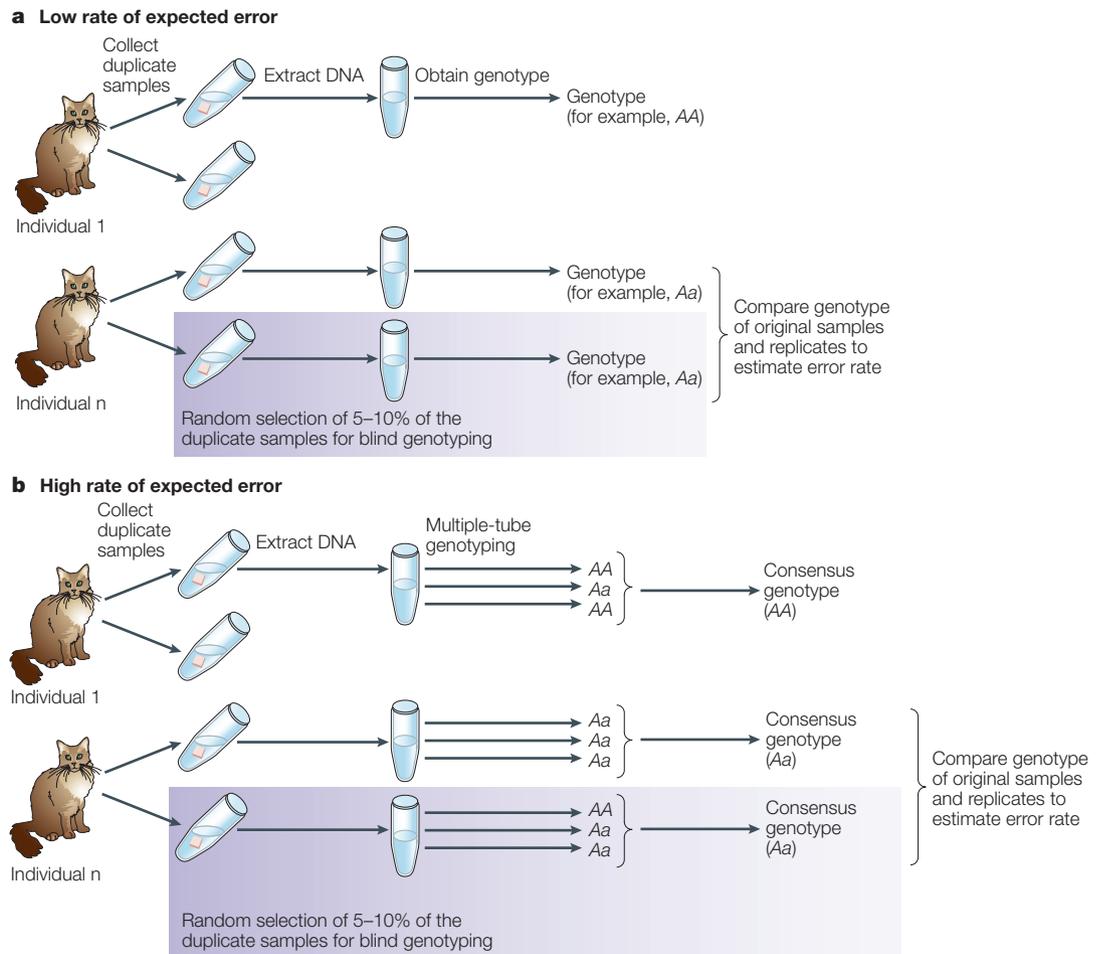


Figure 3 | **The use of blind replicates to estimate the error rate. a** | When the error rate is expected to be relatively low (as in most of the studies that use tissues as a source of DNA), a further 10% of blind experiments (starting from the sample, and not from the DNA extract) should be carried out. **b** | When the error rate is expected to be high, as in non-invasive studies, the multiple-tube approach (using the same DNA extract for the replicates) should also be used.

This strategy involves demonstrating, through an appropriate procedure, that the data produced and the results obtained are reliable. The diversity of case studies, causes of error and laboratory contexts makes it impossible to propose a universal and simple procedure; the possible solutions for limiting the occurrence and effect of genotyping errors are therefore case-specific. The optimal strategy is determined by several factors, such as the biological question, the tolerable error rate, the sampling possibilities, the equipment and technical skills that are locally available, the financial support and time constraints. Nevertheless, some general guidelines can be proposed to design the optimal procedure that can be adapted to a particular case.

General recommendations. The first step is to check that the genotyping experiments that are necessary to reach the scientific goal are realistic according to the sample quality and the technical skills that are available. Poor sample quality and limited technical skills⁴¹ obviously influence the error rate. The second step involves carrying out a pilot study that is designed first to evaluate

the theoretical error rate that is compatible with the data analysis, and second to estimate the real error rate on the basis of the analysis of a subset of the samples (FIG. 2). Finally, it is important to be aware of potential problems throughout the experimental procedure, even after a successful pilot study, from sampling to data analysis. Therefore, quality controls should be carried out in real time during each step and each batch of experiments. They should also be able to detect as many types of error as possible. For example, highly reproducible errors such as null alleles cannot be detected by replicating the genotyping assays and so require HARDY-WEINBERG TESTS or inheritance studies. On the contrary, stochastic allelic dropouts might not be detected by Hardy-Weinberg tests, but by replicating the genotyping assays.

Control procedures are costly and time consuming. Therefore, the effort for reducing the error rate must be adapted to the predictable effect of the genotyping errors. Because genotyping errors can be generated even with high-quality standards³, and because they cannot all be detected⁴², efforts must be directed towards limiting both their production and their subsequent effect.

HARDY-WEINBERG TEST
A test that assesses whether the frequency of each diploid genotype at a locus equals that expected from the random union of alleles.

Limiting the production of errors during genotyping.

Given that human factors are the main issue during genotype production^{3,6}, the most efficient approach is to concentrate first on minimizing human error. This can be achieved in different ways. First, only well-trained bench scientists or technicians should be involved, as suggested by quality-assurance standards for forensic DNA-testing laboratories⁴³. Second, only standardized and validated procedures should be used⁴³. Third, human manipulation should be reduced as much as possible, according to the automation possibilities, from all handling and pipetting steps up to allele scoring^{21,44}. However, software packages are not yet sophisticated enough to prevent scoring errors. Semi-automated scoring followed by human visual inspection seems to be the most reliable procedure⁴⁵.

Limiting genotyping errors during laboratory experiments requires the systematic use of an appropriate number of positive and negative controls, but also requires the implementation of replicates for real-time error detection and error-rate estimation. As positive and negative controls are widely used, we will only focus on replicates.

In every situation, even with high-quality DNA, replicating 5–10% of the samples has been recommended^{3,46}, but the amount can vary according to the goal of the study and the potential effect of errors. As far as possible, these replicates have to be carried out blind and independently. This involves implementing the process blind from the beginning of the experiment, by carrying out a systematic duplication of the samples during sample collection (FIG. 3a). Such a procedure will not only allow the detection of all laboratory errors, but will also pick up handling errors at any stage of the analysis. Moreover, comparing blind samples and original experiments will produce a fair estimate of the error rate.

When genotyping errors are highly probable (for example, in non-invasive studies that involve poor-quality DNA extracts), blind replicates are still necessary but are not sufficient. The systematic replication of each genotyping assay (that is, a multiple-tube approach^{10,47}) is required to define the consensus genotypes (FIG. 3b). Because the cost that is associated with the systematic use of many replicates is far from negligible, many attempts have been made to reduce this number using a MAXIMUM LIKELIHOOD APPROACH⁴⁸, or even to bypass the replication steps if the error rate is low enough to be monitored by alternative approaches⁴⁹. There is a trade-off between the cost of the experiments and the reliability of the genotypes; one role of the pilot study is to determine the optimal number of replicates required.

In some cases, errors can also be detected by replicating the genotyping process by using a different technology⁵⁰, such as sequencing, that is associated with lower error rates than standard genotyping technologies. This procedure allows genotyping error rates to be determined directly, without assumptions about independence of measures or an underlying model for

the errors. However, this approach is hardly applicable to some types of marker, such as AFLPs, and might not be suitable to detect some types of error (for example, allelic dropouts and null alleles).

Cleaning the data set after genotyping. Even if all erroneous genotypes that are detected during the experiments are removed, and eventually corrected after re-genotyping, some undetected errors will remain in the data set. Some of these can still be detected or inferred by looking at the concordance of the scored genotypes with independent data³. For example, where pedigrees are known, checking for Mendelian inheritance can detect most of the remaining errors in linkage analyses^{51,52}. However, the problem still persists if non-genetic data are not reliable⁵³.

The power of detecting errors by consistency with independent data can influence the strategy for limiting errors. It might be more efficient to re-type erroneous genotypes by checking for consistency than by running many blind replicates. This is the case in experiments that involve the SNP typing of parents and offspring, in which it is possible to inspect Mendelian inheritance^{54,55}. Testing a Hardy–Weinberg equilibrium is commonly used to check the quality of the data, under the assumption that a high error rate generates disequilibrium^{56–59}. However, many other causes lead to disequilibrium, including selection, inbreeding and POPULATION ADMIXTURE⁶⁰. Moreover, only a few types of error cause disequilibrium, such as null alleles and allelic dropouts. Therefore, there is still a need for other controls and replicates for detecting errors that are compatible with Mendelian inheritance and a Hardy–Weinberg equilibrium.

Several computer programs specifically designed to detect potential errors are now available (TABLE 3). Most of them check for Mendelian consistency and/or a Hardy–Weinberg equilibrium, and are commonly used for pedigree analyses and linkage studies. Some others have been developed to track errors that are compatible with Mendelian inheritance or a Hardy–Weinberg equilibrium. For example, some detect a spurious excess of recombinants in linkage studies and others focus on inconsistencies between replicates.

However, removing errors might not reduce bias, depending on the number and kind of errors detected and the bias each one creates. For example, when correcting Mendelian-incompatible genotypes by re-typing or removing families in which they occur, the undetected errors can produce an excess of false positives for some family-based association tests⁶¹. This problem has been addressed by developing an appropriate LIKELIHOOD RATIO TEST that is based on a general genotype error model⁶¹. In general, taking into account the occurrence of errors in the analysis is crucial, especially for large or error-prone data sets.

Accounting for errors during data analysis. The overall objective of this review is to help researchers to realize that they have to deal with genotyping errors by setting

MAXIMUM LIKELIHOOD APPROACH

A statistical approach that is used to make inferences about the combination of parameter values that gives the greatest probability of obtaining the observed data.

POPULATION ADMIXTURE

A process that leads to a composite gene pool in which at least some individuals come from more than one population.

LIKELIHOOD RATIO TEST

A method for hypothesis testing. The maximum of the likelihood that the data fit a full model of the data is compared with the maximum of the likelihood that the data fit a restricted model and the likelihood ratio (LR) test statistic is computed. If the LR is significant, the full model provides a better fit to the data than does the restricted model.

Table 3 | **The main software programs that account for genotyping errors**

Program	Principle	Field of application	References	URLs
Evaluation of the effects of errors				
GEMINI	Carries out simulation studies	PG/D	108	http://pbil.univ-lyon1.fr/software/Gemini/gemini.htm
PAWE	Calculates asymptotic power and sample size in biallelic loci	L/Q	109,110	http://linkage.rockefeller.edu/pawe
Detection and/or calculation of genotyping error rate				
PREST	Checks for Mendelian-inconsistent errors (only)	P	111	http://fisher.utstat.toronto.edu/sun/Software/Prest
Pedcheck	Checks for Mendelian-inconsistent errors (only)	P, L/Q	112	http://watson.hgen.pitt.edu/register/docs/pedcheck.html
PedManager	Checks for Mendelian-inconsistent errors (only)	P, L/Q	still under development	http://www.broad.mit.edu/ftp/distribution/software/pedmanager
MENDEL	Checks for Mendelian-inconsistent and Mendelian-consistent errors (for example, spurious excess of recombinants)	P, L/Q, PG/D	55	http://www.genetics.ucla.edu/software
SIMWALK	Checks for Mendelian-inconsistent and Mendelian-consistent errors (for example, spurious excess of recombinants)	P, L/Q	65	http://www.genetics.ucla.edu/software
Genocheck	Checks for Mendelian-inconsistent and Mendelian-consistent errors (for example, spurious excess of recombinants)	L/Q	113	http://softlib.rice.edu/geno.html
R/QTL	Checks for Mendelian-inconsistent and Mendelian-consistent errors (for example, spurious excess of recombinants)	L/Q	114	http://www.biostat.jhsph.edu/~kbroman/qtl
CERVUS	Checks for Mendelian-inconsistent errors and Hardy–Weinberg equilibrium	P	37	http://helios.bto.ed.ac.uk/evolgen/cervus/cervus.html
GIMLET	Checks for consistency between repeats	PG/D	115	http://pbil.univ-lyon1.fr/software/Gimlet/gimlet.htm
RelioType	Checks for consistency between repeats	PG/D	48	http://www.cnr.uidaho.edu/lecg/pubs_and_software.htm
Micro-checker	Checks for Hardy–Weinberg equilibrium	PG/D	107	http://www.microchecker.hull.ac.uk
DROPOUT	Calculates distribution of pairwise difference between genotypes	PG/D	116	http://www.fs.fed.us/rm/wildlife/genetics
Analysis of data sets that contain errors				
PARENTE	Allows allelic mismatches in the analysis	P, PG/D	84	http://www2.ujf-grenoble.fr/leca/membres/manel.html
PAPA	Allows an underlying error model	P	67	http://www.bio.ulaval.ca/louisbernatchez/downloads_fr.htm
PseudoMarker	Allows an underlying error model	L/Q	8	http://www.helsinki.fi/~tsjuntun/pseudomarker
TDtae	Calculates maximum likelihood estimates of genotyping error rates and tests statistical inference of association	L/Q	52,54	ftp://linkage.rockefeller.edu/software/tdtae2
LRTae	Calculates maximum likelihood estimates of genotyping error rates and tests statistical inference of association	L/Q	68	ftp://linkage.rockefeller.edu/softare/lrtae

A more complete list of software packages that deal with genotyping errors can be found in general websites (see Online links box). The list of programs that are described in this table is not exhaustive. In particular, software programs that are designed for very specific purposes are not included. L/Q, linkage or QTL studies; P, pedigree analysis; PG/D, population genetics or demography.

SHORT-ALLELE DOMINANCE
The preferential PCR amplification of the shorter allele from a heterozygote individual. This is equivalent to a long-allele dropout.

up a strategy that is appropriate for their own particular situation. Following ‘ready-to-use’ protocols is dangerous because they do not allow the error-detection strategy to be adapted to the study in question, and can therefore lead to an inefficient management of errors. Even for a given study and a particular type of error, the way to deal with errors will vary according to several parameters. For example, consider the simple case in which F_{ST} values among populations are estimated from a microsatellite data set in which allelic dropout is the main cause of error. Allelic dropout can occur

stochastically (when alleles do not differ much in size) or not (SHORT-ALLELE DOMINANCE, when allele sizes are sufficiently different). The generated bias is even more important because the distribution of alleles is different between populations. Therefore, the effect on the resulting F_{ST} values, and the tolerable error rate, depend on the real differentiation between populations. The acceptable error rate depends on many parameters, even for a precise topic such as F_{ST} estimation. The only way to estimate this is by comparing the results of a pilot study with those from simulations (see FIG. 2).

PROBABILITY OF IDENTITY

The overall probability that two individuals drawn at random from a given population share identical genotypes at all typed loci.

EFFECTIVE POPULATION SIZE

The size of the ideal population in which the effects of random drift would be the same as those seen in the actual population.

DIRECTED-ERROR MODEL

A model postulating that there is a greater probability for a particular allele to be consistently incorrectly genotyped.

Such pilot studies are the best way to assess the situation and decide how to handle the genotyping error issue in practice.

However, there are several ways of accounting for error rate in the analysis, and of minimizing their effect. Attention must first be paid to choosing statistics that are robust to genotyping errors. For example, Akey *et al.* showed that among four common estimates of linkage disequilibrium, two were less sensitive to genotyping error, although there were exceptions depending on haplotype frequencies²⁷. Consequently, choosing the more robust measure is not straightforward. Theoretical studies and simulations are needed to quantify the robustness to genotyping error of a wide variety of other population genetic estimates (for example, F_{ST} , migration rate, linkage disequilibrium, PROBABILITY OF IDENTITY and EFFECTIVE POPULATION SIZE). The effect of error on these estimates remains to be investigated. For example, errors that occur stochastically are expected to increase the migration rate and decrease the F_{ST} value among populations; however, to our knowledge, no study has ever dealt with this topic. A further possibility is to use tests that, because of their statistical power, are robust to the occurrence of genotyping errors^{62,63}. Errors can also be dealt with by allowing a certain number of inconsistencies (considered to result from errors) to occur between genotypes. This is the case in parentage studies³⁶ or in individual identification from non-invasive samples (for example, population-size estimates²³). When a mismatch occurs, the difficulty is to estimate whether it comes from a genotyping error or has a biological cause. The estimation of the error rate within the data set is crucial to estimate the relative influences of these two causes.

It is also valuable to use methods that calculate the likelihood of obtained genotypes or pedigrees using a model of error occurrence, such as a uniform or empirical distribution of errors^{48,64,65}. The field of linkage analysis has made the greatest effort to take genotyping errors into account during analysis^{27,52,66}. Thorough studies are still necessary to apply such approaches to other fields that use genetic tools and to develop programs that allow the analysis of data sets that contain errors^{52,54,67,68} (TABLE 3).

Finally, analysing data can be all the more complicated because more than one type of error can affect a study, and each error type has a different effect on the result. For example, errors that follow a stochastic-error model have less severe effects on linkage-disequilibrium estimates than errors that follow a DIRECTED-ERROR MODEL²⁷. Therefore, different errors might be taken into account separately in the analysis by giving them different weights, to avoid skewing the results⁶⁶.

Towards quality processes for genotyping

In every scientific discipline the reliability of the conclusions strongly depends on the quality of the data. For geneticists, genotyping errors can affect results^{2,13,23,24}. We propose that the protocol that is used for minimizing the occurrence of errors, the methods for

error detection and the estimated error rate should be provided for each study (FIG. 2). With this information it will be possible to assign a quality index to each genotype, allowing the scientific community to provide a critical assessment when unexpected results are published. Quality standards, such as the rules imposed by the FBI for forensic DNA analysis⁴³, should be promoted even outside the forensic area.

An increasing number of studies, often in the context of international programmes, generate enormous data sets that cannot be produced in a single laboratory^{40,65}. Therefore, the reproducibility of genotyping becomes increasingly important^{69–71}. Even for markers that are known to be robust (SNPs, microsatellites, AFLPs), differences appear between laboratories, and over time within the same laboratory⁷⁰. These complications have led to initiatives such as the European Molecular Genetics Quality Network, which was established in 1996 to spread quality assurance across Europe and harmonize national activities⁷².

The trend towards quality standards in genetics is not restricted to genotyping. Expression studies that use microarray experiments are known to be error-prone, and the scientific community reacted by designing strict standards: the ‘Minimum Information About a Microarray Experiment’ (MIAME) document⁷³ comprises a checklist to ensure that data are made publicly available in a format that allows unambiguous interpretation and potential verification of the conclusion. It includes several steps for verifying experimental design, sample preparation and data measurement.

We have been aware of genotyping errors since the beginning of molecular genetics. Their consequences for statistical genetics were pointed out in 1957 (REF. 74), and null alleles in blood groups have been recognized since 1938 (REF. 75). Errors too often remained neglected and, given their marked effect on some studies, it is clear that they merit more attention. Recently, many papers have dealt with genotyping errors, and it seems that the scientific community has begun to realize their importance. The fields of ancient DNA^{76,77} and gene expression^{78,79} suffered a crisis of confidence, with a series of erroneous papers published in leading journals. As a result, these two scientific communities set up strict standards to promote data quality, which solved the crisis. In population genetics, the situation is different because only a few erroneous papers have been published, and therefore this community has not been given such a strong incentive to establish strict standards. Another explanation for the delay in establishing strict standards might be related to the complexity of the problems. The wide range of questions, molecular markers and data-analysis methods has prevented simple solutions from being devised. Because of the recent awareness about the occurrence of genotyping errors and their potential effect, we predict that increasing attention will be paid to these difficulties when designing experimental protocols and publishing results.

1. Gagneux, P., Woodruff, D. S. & Boesch, C. Furtive mating in female chimpanzees. *Nature* **387**, 358–359 (1997).
2. Gagneux, P., Boesch, C. & Woodruff, D. S. Microsatellite scoring errors associated with noninvasive genotyping based on nuclear DNA amplified from shed hair. *Mol. Ecol.* **6**, 861–868 (1997).
This paper deals with genotyping errors in non-invasive studies and is the first one to mention 'allelic dropout'.
3. Bonin, A. *et al.* How to track and assess genotyping errors in population genetics studies. *Mol. Ecol.* **13**, 3261–3273 (2004).
An extensive study of the causes and consequences of genotyping errors on AFLP and microsatellite data, with practical recommendations for limiting error occurrence and effect.
4. Thompson, E. A. A paradox of genealogical inference. *Adv. Appl. Probab.* **8**, 648–650 (1976).
5. Taberlet, P., Waits, L. P. & Luikart, G. Noninvasive genetic sampling: look before you leap. *Trends Ecol. Evol.* **14**, 323–327 (1999).
This article focuses on the processes for limiting the occurrence of genotyping errors in non-invasive studies, highlighting the role of pilot studies.
6. Hoffman, J. I. & Amos, W. Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. *Mol. Ecol.* **14**, 599–612 (2005).
A careful examination of the causes of genotyping errors on microsatellite data, showing the importance of human factors.
7. Ewen, K. R. *et al.* Identification and analysis of error types in high-throughput genotyping. *Am. J. Hum. Genet.* **67**, 727–736 (2000).
8. Göring, H. H. H. & Terwilliger, J. D. Linkage analysis in the presence of errors II: Marker-locus genotyping errors modeled with hypercomplex recombination fractions. *Am. J. Hum. Genet.* **66**, 1107–1118 (2000).
9. Brzustowicz, L. M. *et al.* Molecular and statistical approaches to the detection and correction of errors in genotype databases. *Am. J. Hum. Genet.* **53**, 1137–1145 (1993).
10. Taberlet, P. *et al.* Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Res.* **24**, 3189–3194 (1996).
This study shows the difficulty for producing reliable genotype data that is due to the occurrence of false alleles and false homozygotes (that is, allelic dropout).
11. Taberlet, P. *et al.* Noninvasive genetic tracking of the endangered Pyrenean brown bear population. *Mol. Ecol.* **6**, 869–876 (1997).
12. Mitchell, A. A., Zwick, M. E., Chakravarti, A. & Cutler, D. J. Discrepancies in dbSNP confirmation rates and allele frequency distributions from varying genotyping error rates and patterns. *Bioinformatics* **20**, 1022–1032 (2004).
13. Abecasis, G. R., Cherny, S. S. & Cardon, L. R. The impact of genotyping error on family-based analysis of quantitative traits. *Eur. J. Hum. Genet.* **9**, 130–134 (2001).
14. Yao, Y.-G., Bravi, C. M. & Bandelt, H.-J. A call for mtDNA data quality control in forensic science. *Forensic Sci. Int.* **141**, 1–6 (2004).
15. Schlotterer, C. The evolution of molecular markers — just a matter of fashion? *Nature Rev. Genet.* **5**, 63–69 (2004).
16. Bandelt, H. J., Lahermo, P., Richards, M. & Macaulay, V. The fingerprint of phantom mutations in mitochondrial DNA data. *Am. J. Hum. Genet.* **71**, 1150–1160 (2002).
17. Callen, D. F. *et al.* Incidence and origin of 'null' alleles in the (AC)_n microsatellite markers. *Am. J. Hum. Genet.* **52**, 922–927 (1993).
The first study to report the occurrence of non-amplifying microsatellite alleles (that is, null alleles).
18. Paetkau, D. & Strobeck, C. The molecular basis and evolutionary history of a microsatellite null allele in bears. *Mol. Ecol.* **4**, 519–520 (1995).
19. Brownstein, M. J., Carpten, J. D. & Smith, J. R. Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. *BioTechniques* **20**, 1004–1010 (1996).
20. Magnuson, V. L. *et al.* Substrate nucleotide-determined non-templates addition of adenine by Taq DNA polymerase: implications for PCR-based genotyping and cloning. *BioTechniques* **21**, 700–709 (1996).
21. Li, J. L. *et al.* Toward high-throughput genotyping: dynamic and automatic software for manipulating large-scale genotype data using fluorescently labeled dinucleotide markers. *Genome Res.* **11**, 1304–1314 (2001).
22. Ghosh, S. *et al.* Methods for precise sizing, automated binning of alleles, and reduction of error rates in large-scale genotyping using fluorescently labeled dinucleotide markers. *Genome Res.* **7**, 165–178 (1997).
23. Creel, S. *et al.* Population size estimation in Yellowstone wolves with error-prone noninvasive microsatellite genotypes. *Mol. Ecol.* **12**, 2003–2009 (2003).
24. Waits, J. L. & Leberg, P. L. Biases associated with population estimation using molecular tagging. *Anim. Conserv.* **3**, 191–199 (2000).
25. Gordon, D., Heath, S. C. & Ott, J. True pedigree errors more frequent than apparent errors for single nucleotide polymorphisms. *Hum. Hered.* **49**, 65–70 (1999).
26. Geller, F. & Ziegler, A. Detection rates for genotyping errors in SNPs using the trio design. *Hum. Hered.* **54**, 111–117 (2002).
27. Akey, J. M., Zhang, K., Xiong, M. M., Doris, P. & Jin, L. The effect that genotyping errors have on the robustness of common linkage-disequilibrium measures. *Am. J. Hum. Genet.* **68**, 1447–1456 (2001).
A study that investigates the effects of genotyping error on estimates of linkage disequilibrium, and shows that the robustness of the estimates depends on allelic frequencies and assumed error models.
28. Kirk, K. M. & Cardon, L. R. The impact of genotyping error on haplotype reconstruction and frequency estimation. *Eur. J. Hum. Genet.* **10**, 616–622 (2002).
29. Hackett, C. A. & Broadfott, L. B. Effects of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps. *Heredity* **90**, 33–38 (2003).
30. Buetow, K. H. Influence of aberrant observations on high-resolution linkage analysis outcomes. *Am. J. Hum. Genet.* **49**, 985–994 (1991).
31. Goldstein, D. R., Zhao, H. Y. & Speed, T. P. The effects of genotyping errors and interference on estimation of genetic distance. *Hum. Hered.* **47**, 86–100 (1997).
32. Douglas, J. A., Boehnke, M. & Lange, K. A multipoint method for detecting genotyping errors and mutations in sibling-pair linkage data. *Am. J. Hum. Genet.* **66**, 1287–1297 (2000).
33. Terwilliger, J. D., Weeks, D. E. & Ott, J. Laboratory errors in the reading of marker alleles cause massive reductions in LOD score and lead to gross overestimation of the recombination fraction. *Am. J. Hum. Genet.* **47**, A201 (1990).
34. Gordon, D., Matisse, T. C., Heath, S. C. & Ott, J. Power loss for multiallelic transmission/disequilibrium test when errors introduced: GAW11 simulated data. *Genet. Epidemiol.* **17**, S587–S592 (1999).
35. Rebbeck, T. R. *et al.* SNPs, haplotypes, and cancer: applications in molecular epidemiology. *Cancer Epidemiol. Biomarkers Prev.* **13**, 681–687 (2004).
36. Pemberton, J. M., Slate, J., Bancroft, D. R. & Barrett, J. A. Non-amplifying alleles at microsatellite loci: a caution for parentage and population studies. *Mol. Ecol.* **4**, 249–252 (1995).
37. Marshall, T. C., Slate, J., Kruuk, L. E. B. & Pemberton, J. M. Statistical confidence for likelihood-based paternity inference in natural populations. *Mol. Ecol.* **7**, 639–655 (1998).
38. Dakin, E. E. & Avise, J. C. Microsatellite null alleles in parentage analysis. *Heredity* **93**, 504–509 (2004).
39. Weiser Eastale, P. & Eastale, S. The forensic use of DNA profiling. *Trends Issues Crime Crim. Justice* **26**, 1–8 (1990).
40. Luikart, G., England, P., Tallmon, D., Jordan, S. & Taberlet, P. The power and promise of population genomics: from genotyping to genome typing. *Nature Rev. Genet.* **4**, 981–994 (2003).
41. Paetkau, D. An empirical exploration of data quality in DNA-based population inventories. *Mol. Ecol.* **12**, 1375–1387 (2003).
A review of the various approaches that were designed to probe the reliability of data in non-invasive studies on bears.
42. Douglas, J. A., Skol, A. D. & Boehnke, M. Probability of detection of genotyping errors and mutations as inheritance inconsistencies in nuclear-family data. *Am. J. Hum. Genet.* **70**, 487–495 (2002).
43. Butler, J. M. *Forensic DNA Typing: Biology and Technology Behind STR Markers* (Academic Press, San Diego, 2001).
44. Perlin, M. W., Lancia, G. & Ng, S. K. Toward fully automated genotyping: genotyping microsatellite markers by deconvolution. *Am. J. Hum. Genet.* **57**, 1199–1210 (1995).
45. Papa, R., Troggio, M., Ajmone-Marsan, P. & Nonnis Marzano, F. An improved protocol for the production of AFLP markers in complex genomes by means of capillary electrophoresis. *J. Anim. Breed. Genet.* **122**, 62–68 (2005).
46. Millikan, R. The changing face of epidemiology in the genomics era. *Epidemiology* **13**, 472–480 (2002).
47. Navidi, W., Arnheim, N. & Waterman, M. S. A multiple-tubes approach for accurate genotyping of very small DNA samples by using PCR: statistical considerations. *Am. J. Hum. Genet.* **50**, 347–359 (1992).
48. Miller, C. R., Joyce, P. & Waits, L. P. Assessing allelic dropout and genotyping reliability using maximum likelihood. *Genetics* **160**, 357–366 (2002).
49. Paetkau, D., Calvert, W., Stirling, I. & Strobeck, C. Microsatellite analysis of population structure in Canadian polar bears. *Mol. Ecol.* **4**, 347–354 (1995).
50. Tenenbein, A. A double sampling scheme for estimating from misclassified multinomial data with applications to sampling inspection. *Technometrics* **14**, 187–202 (1972).
51. Stringham, H. M. & Boehnke, M. Identifying marker typing incompatibilities in linkage analysis. *Am. J. Hum. Genet.* **59**, 946–950 (1996).
52. Gordon, D., Heath, S. C., Liu, X. & Ott, J. A transmission/disequilibrium test that allows for genotyping errors in the analysis of single-nucleotide polymorphism data. *Am. J. Hum. Genet.* **69**, 371–380 (2001).
53. Boehnke, M. & Cox, N. J. Accurate inference of relationships in sib-pair linkage studies. *Am. J. Hum. Genet.* **61**, 423–429 (1997).
54. Gordon, D. *et al.* A transmission disequilibrium test for general pedigrees that is robust to the presence of random genotyping errors and any number of untyped parents. *Eur. J. Hum. Genet.* **12**, 752–761 (2004).
55. Lange, K. *et al.* Mendel version 4.0: a complete package for the exact genetic analysis of discrete traits in pedigree and population data sets. *Am. J. Hum. Genet.* **69**, A1886 (2001).
56. Chakraborty, R., De Andrade, M., Daiger, S. P. & Budowle, B. Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications. *Ann. Hum. Genet.* **56**, 45–57 (1992).
57. Gomes, I. *et al.* Hardy-Weinberg quality control. *Ann. Hum. Genet.* **63**, 535–538 (1999).
58. Xu, J., Turner, A., Little, J., Bleecker, E. R. & Meyers, D. A. Positive results in association studies are associated with departure from Hardy-Weinberg equilibrium: hint for genotyping error? *Hum. Genet.* **111**, 573–574 (2002).
59. Hosking, L. *et al.* Detection of genotyping errors by Hardy-Weinberg equilibrium testing. *Eur. J. Hum. Genet.* **12**, 395–399 (2004).
60. Morton, N. E. & Collins, A. E. Statistical and genetic aspects of quality control for DNA identification. *Electrophoresis* **16**, 1670–1677 (1995).
61. Morris, R. W. & Kaplan, N. L. Testing for association with a case-parents design in the presence of genotyping errors. *Genet. Epidemiol.* **26**, 142–154 (2004).
62. Kang, S. J., Gordon, D. & Finch, S. J. What SNP genotyping errors are most costly for genetic association studies? *Genet. Epidemiol.* **26**, 132–141 (2004).
63. Zou, G. H. & Zhao, H. Y. The impact of errors in individual genotyping and DNA pooling on association studies. *Genet. Epidemiol.* **26**, 1–10 (2004).
64. Rice, K. M. & Holmans, P. Allowing for genotyping error in analysis of unmatched case-control studies. *Ann. Hum. Genet.* **67**, 165–174 (2003).
65. Sobel, E., Papp, J. C. & Lange, K. Detection and integration of genotyping errors in statistical genetics. *Am. J. Hum. Genet.* **70**, 496–508 (2002).
A nice reference study that illustrates several possibilities for integrating genotyping errors in statistical analyses in human pedigree studies.
66. Wang, J. L. Siblings reconstruction from genetic data with typing error. *Genetics* **166**, 1963–1979 (2004).
67. Duchesne, P., Gobout, M.-H. & Bernatchez, L. PAPA (Package for the Analysis of Parental Allocation): a computer program for simulated and real parental allocation. *Mol. Ecol. Notes* **2**, 191–194 (2002).
68. Gordon, D. *et al.* Increasing power for tests of genetic association in the presence of phenotype and/or genotype error by use of double-sampling. *Stat. Appl. Genet. Mol. Biol.* **3**, a26 (2004).
69. Weeks, D. E., Conley, Y. P., Ferrell, R. E., Mah, T. S. & Gorin, M. B. A tale of two genotypes: consistency between two high-throughput genotyping centers. *Genome Res.* **12**, 430–435 (2002).
70. Jones, C. J. *et al.* Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Mol. Breed.* **3**, 381–390 (1997).
71. Dequeker, E. & Cassiman, J. J. Evaluation of CFTR gene mutation testing methods in 136 diagnostic laboratories: report of a large European external quality assessment. *Eur. J. Hum. Genet.* **6**, 165–175 (1998).
72. Muller, C. R. Quality control in mutation analysis: the European Molecular Genetics Quality Network (EMQN). *Eur. J. Pediatr.* **160**, 464–467 (2001).
73. Brazma, A. *et al.* Minimum information about a microarray experiment (MIAME) — toward standards for microarray data. *Nature Genet.* **29**, 365–371 (2001).

74. Smith, C. A. B. Counting methods in genetic statistics. *Ann. Hum. Genet.* **21**, 254–276 (1957).
75. Stevens, W. L. Estimation of blood-group gene frequencies. *Ann. Eugen. Lond.* **8**, 362–375 (1938).
76. Zischler, H. *et al.* Detecting dinosaur DNA. *Science* **268**, 1192–1193 (1995).
77. Austin, J. J., Ross, A. J., Smith, A. B., Fortey, R. A. & Thomas, R. H. Problems of reproducibility — does geologically ancient DNA survive in amber-preserved insects? *Proc. R. Soc. Lond. B* **264**, 467–474 (1997).
78. Quackenbush, J. Computational analysis of microarray data. *Nature Rev. Genet.* **2**, 418–427 (2001).
79. Aach, J., Rindone, W. & Church, G. M. Systematic management and analysis of yeast gene expression data. *Genome Res.* **10**, 431–445 (2000).
80. Broquet, T. & Petit, E. Quantifying genotyping errors in noninvasive population genetics. *Mol. Ecol.* **13**, 3601–3608 (2004).
- A critical analysis of the various methods available to estimate allelic dropout rates and false allele rates in protocols designed for non-invasive studies.**
81. Bellemain, E., Swenson, J. E. & Taberlet, P. Mating strategies in relation to sexually selected infanticide in a nonsocial carnivore: the brown bear. *Ethology* **111**, 1–14 (2005).
82. Paetkau, D. & Strobeck, C. Microsatellite analysis of genetic variation in black bear populations. *Mol. Ecol.* **3**, 489–495 (1994).
83. Waits, L. P., Taberlet, P., Swenson, J. E. & Sandegren, F. Nuclear DNA microsatellite analysis of genetic diversity and gene flow in the Scandinavian brown bear (*Ursus arctos*). *Mol. Ecol.* **9**, 421–431 (2000).
84. Cercueil, A., Bellemain, E. & Manel, S. PARENTE: computer program for parentage analysis. *J. Hered.* **93**, 458–459 (2003).
85. Walsh, P. S., Erlich, H. A. & Higuchi, R. Preferential PCR amplification of alleles: mechanisms and solutions. *PCR Methods Appl.* **1**, 241–250 (1992).
86. Shaw, P. W., Pierce, G. J. & Boyle, P. R. Subtle population structuring within a highly vagile marine invertebrate, the veined squid *Loligo forbesi*, demonstrated with microsatellite DNA markers. *Mol. Ecol.* **8**, 407–417 (1999).
87. Vekemans, X., Beauwens, T., Lemaire, M. & Roldan-Ruiz, I. Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Mol. Ecol.* **11**, 139–151 (2002).
88. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362**, 709–715 (1993).
89. Wattier, R., Engel, C. R., Saumitou-Laprade, P. & Valera, M. Short allele dominance as a source of heterozygote deficiency at microsatellite loci: experimental evidence at the dinucleotide locus Gv1CT in *Gracilaria gracilis* (Rhodophyta). *Mol. Ecol.* **7**, 1569–1573 (1998).
90. Martinez, J. G. & Burke, T. Microsatellite typing of sperm trapped in the perivitelline layers of avian eggs. *J. Avian Biol.* **34**, 20–24 (2003).
91. Kohn, M. H. & Wayne, R. K. Facts from feces revisited. *Trends Ecol. Evol.* **12**, 223–227 (1997).
- A review on non-invasive DNA analyses from faeces, with valuable technical notes on the sources of error.**
92. Valière, N. & Taberlet, P. Urine collected in the field as a source of DNA for species and individual identification. *Mol. Ecol.* **9**, 2150–2152 (2000).
93. Uchihi, R., Tamaki, K., Kojima, T., Yamamoto, T. & Katsumata, Y. Deoxyribonucleic acid (DNA) typing of human leukocyte antigen (HLA)-DQA1 from single hairs in Japanese. *J. Forensic Sci.* **37**, 853–859 (1992).
94. Koonjul, P. K., Brandt, W. F., Farrant, J. M. & Lindsey, G. G. Inclusion of polyvinylpyrrolidone in the polymerase chain reaction reverses the inhibitory effects of polyphenolic contamination of RNA. *Nucleic Acids Res.* **27**, 915–916 (1999).
95. Foucault, F., Praz, F., Jaulin, C. & Amor-Gueret, M. Experimental limits of PCR analysis of (CA)_n repeats alterations. *Trends Genet.* **12**, 450–452 (1996).
96. Parsons, K. M. Reliable microsatellite genotyping of dolphin DNA from faeces. *Mol. Ecol. Notes* **1**, 341–344 (2001).
97. Shinde, D., Lai, Y., Sun, F. & Arnheim, N. Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)_n and (AT)_n microsatellites. *Nucleic Acids Res.* **31**, 974–980 (2003).
98. Polisky, B. *et al.* Specificity of substrate recognition by the EcoRI restriction endonuclease. *Proc. Natl Acad. Sci USA* **72**, 3310–3314 (1975).
99. Haberl, M. & Tautz, D. Comparative allele sizing can produce inaccurate allele size differences microsatellites. *Mol. Ecol.* **8**, 1347–1350 (1999).
100. Delmotte, F., Leteime, N. & Simon, J.-C. Microsatellite allele sizing: difference between automated capillary electrophoresis and manual technique. *BioTechniques* **31**, 810–818 (2001).
101. Fernando, P., Evans, B. J., Morales, J. C. & Melnick, D. J. Electrophoresis artefacts — a previously unrecognized cause of error in microsatellite analysis. *Mol. Ecol. Notes* **1**, 325–328 (2001).
102. Davison, A. & Chiba, S. Laboratory temperature variation is a previously unrecognized source of genotyping error during capillary electrophoresis. *Mol. Ecol. Notes* **3**, 321–323 (2003).
103. Gerloff, U. *et al.* Amplification of hypervariable simple sequence repeats (microsatellites) from excremental DNA of wild bonobos (*Pan paniscus*). *Mol. Ecol.* **4**, 515–518 (1995).
104. Constable, J. L., Ashley, M. V., Goodall, J. & Pusey, A. E. Noninvasive paternity assignment in Gombé chimpanzees. *Mol. Ecol.* **10**, 1279–1300 (2001).
105. Matsuzaki, H. *et al.* Parallel genotyping of over 10,000 SNPs using a one-primer assay on a high-density oligonucleotide array. *Genome Res.* **14**, 414–425 (2004).
106. Ekstrom, C. T. Detecting low-quality markers using map expanders. *Genet. Epidemiol.* **25**, 214–224 (2003).
107. van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M. & Shipley, P. Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes* **4**, 535–538 (2004).
108. Valière, N., Berthier, P., Mouchiroud, D. & Pontier, D. GEMNI: software for testing the effects of genotyping errors and multitubes approach for individual identification. *Mol. Ecol. Notes* **2**, 83–86 (2002).
109. Gordon, D., Finch, S. J., Nothnagel, M. & Ott, J. Power and sample size calculations for case-control genetic association tests when errors are present: application to single nucleotide polymorphisms. *Hum. Hered.* **54**, 22–23 (2002).
110. Gordon, D., Levenstien, M. A., Finch, S. J. & Ott, J. Errors and linkage disequilibrium interact multiplicatively when computing sample sizes for genetic case-control association studies. *Pac. Symp. Biocomput.*, 490–501 (2003).
111. McPeck, M. S. & Sun, L. Statistical tests for detection of misspecified relationships by use of genome-screen data. *Am. J. Hum. Genet.* **66**, 1076–1094 (2000).
112. O'Connell, J. R. & Weeks, D. E. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am. J. Hum. Genet.* **63**, 259–266 (1998).
113. Ehm, M. G., Cottingham, R. W. Jr & Kimmel, M. Error detection in genetic linkage data using likelihood based methods. *J. Biol. Syst.* **3**, 13–25 (1995).
114. Broman, K. W., Wu, H., Sen, S. & Churchill, G. A. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**, 889–890 (2003).
115. Valière, N. GIMLET: a computer program for analysing genetic individual identification data. *Mol. Ecol. Notes* **2**, 377–379 (2002).
116. McKelvey, K. S. & Schwartz, M. K. DROPOUT: a program to identify problem loci and samples for noninvasive genetic samples in a capture-mark-recapture framework. *Mol. Ecol. Notes* **5**, 716–718 (2005).

Acknowledgements

The authors are grateful to G. Luikart for fruitful discussions and comments on the manuscript and to the persons from the Scandinavian Brown Bear Research Project who provided the bear samples. They thank three anonymous reviewers for providing references and helpful comments.

Competing interests statement

The authors declare no competing financial interests.

Online links

FURTHER INFORMATION

An alphabetical list of genetic analysis software:

<http://linkage.rockefeller.edu/soft/list2.html>

DNA Advisory Board Quality Assurance Standards for Forensic DNA Testing Laboratories:

<http://www.cstl.nist.gov/biotech/strbase/dabqas.htm#quality%20assurance%20standards>

European Molecular Genetics Quality Network: <http://www.emqn.org/emqn.php>

ISI Web of Science: <http://wok.mimas.ac.uk>

Minimum Information About a Microarray Experiment:

<http://www.mged.org/Workgroups/MIAME/miame.html>

PARENTE: http://www2.ujf-grenoble.fr/leca/membres/manel_a.html

Programs useful for detecting genotyping and pedigree errors:

<http://www2.qimr.edu.au/davidD/Course/part6.html>

UCLA Human Genetics Software Distribution: <http://www.genetics.ucla.edu/software>

Access to this interactive links box is free online.