

INVITED REVIEW

How to track and assess genotyping errors in population genetics studies

A. BONIN,* E. BELLEMAIN,*† P. BRONKEN EIDENSEN,‡ F. POMPANON,* C. BROCHMANN‡ and P. TABERLET*

*Laboratoire d'Ecologie Alpine, CNRS-UMR 5553, Université Joseph Fourier, BP 53, 38041 Grenoble Cedex 09, France, †Department of Ecology and Natural Resource Management, Agricultural University of Norway, Box 5003, NO-1432 Ås, Norway, ‡National Centre for Biosystematics, Natural History Museums and Botanical Garden, University of Oslo, PO Box 1172 Blindern, NO-0318 Oslo, Norway

Abstract

Genotyping errors occur when the genotype determined after molecular analysis does not correspond to the real genotype of the individual under consideration. Virtually every genetic data set includes some erroneous genotypes, but genotyping errors remain a taboo subject in population genetics, even though they might greatly bias the final conclusions, especially for studies based on individual identification. Here, we consider four case studies representing a large variety of population genetics investigations differing in their sampling strategies (noninvasive or traditional), in the type of organism studied (plant or animal) and the molecular markers used [microsatellites or amplified fragment length polymorphisms (AFLPs)]. In these data sets, the estimated genotyping error rate ranges from 0.8% for microsatellite loci from bear tissues to 2.6% for AFLP loci from dwarf birch leaves. Main sources of errors were allelic dropouts for microsatellites and differences in peak intensities for AFLPs, but in both cases human factors were non-negligible error generators. Therefore, tracking genotyping errors and identifying their causes are necessary to clean up the data sets and validate the final results according to the precision required. In addition, we propose the outline of a protocol designed to limit and quantify genotyping errors at each step of the genotyping process. In particular, we recommend (i) several efficient precautions to prevent contaminations and technical artefacts; (ii) systematic use of blind samples and automation; (iii) experience and rigor for laboratory work and scoring; and (iv) systematic reporting of the error rate in population genetics studies.

Keywords: AFLP, automation, blind samples, genotyping errors, human factors, microsatellites, scoring process

Received 25 May 2004; revision received 21 August 2004; accepted 21 August 2004

Introduction

In the past decade, new challenges in population genetics and technical improvements in DNA fingerprinting have led to the production and analysis of larger molecular data sets for a wide range of organisms (Waits *et al.* 2000; Akey *et al.* 2002; Kauer *et al.* 2003; Segovia-Lerma *et al.* 2003). Unfortunately, as a logical corollary of this major breakthrough, the number of genotyping errors is expected to

increase with the size of data sets (Sobel *et al.* 2002) and those errors often go undetected because they are generally unobtrusive.

Genotyping errors occur when the genotype determined after molecular analysis does not correspond to the real genotype of the individual under consideration. However, this real genotype is inaccessible directly and has to be assessed with molecular analyses. In practice, genotyping errors are thus defined as the differences observed between two or more molecular genotypes obtained independently from the same sample. Eradicating genotyping errors cannot be achieved, mainly because molecular

Correspondence: A. Bonin. Fax: +33 0 4 76 51 42 79; E-mail: aurelie.bonin@ujf-grenoble.fr

assays and manual sample handling are not 100% reliable. Genotyping errors can be generated at every step of the genotyping process (sampling, DNA extraction, molecular analysis, scoring, data analysis) and by a variety of factors (chance, human causes, technical artefacts).

Virtually every data set obtained by genotyping contains some errors; consequently, they should not be overlooked as they might greatly bias the final results. For instance, in mapping studies, undetected error rates as low as 1% can lead to incorrect map orders and inflation of map lengths (Buetow 1991; Lincoln & Lander 1992; Hackett & Broadfoot 2003). In population genetics surveys, genotyping errors affect both the allele frequency estimates and the accurate discrimination of different genotypes. False estimates of allele frequency can create an artificial excess of homozygotes (Taberlet *et al.* 1996; Gagneux *et al.* 1997a), a false departure from Hardy–Weinberg equilibrium (Xu *et al.* 2002), an overestimation of inbreeding (Gomes *et al.* 1999; Taberlet *et al.* 1999) or unreliable inferences about population substructures (Miller *et al.* 2002). Erroneous genotypes can distort population size estimates (Creel *et al.* 2003; McKelvey & Schwartz 2003), individual identification (Taberlet & Luikart 1999; Paetkau 2003) and parentage analysis (Miller *et al.* 2002). For instance, when investigating the reproductive behaviour of chimpanzee in a Tai Forest community, Gagneux *et al.* (1997b) concluded, on the basis of molecular typings, that half the offspring had an extra-group father, as they did not display any allele inherited from an intragroup father. In contradiction with several other studies (Constable *et al.* 2001; Vigilant *et al.* 2001), these results were actually due to allelic dropouts (Constable *et al.* 2001), i.e. amplifications of only one of the two alleles for heterozygous individuals, producing false homozygotes (Taberlet *et al.* 1996).

Specific protocols for limiting the extent of such errors and for estimating their rate of occurrence are standard practice in only some genetics research areas, including noninvasive sampling (see Paetkau 2003 for a review), ancient DNA analysis (Hofreiter *et al.* 2001; Yoder & Delefosse 2002) or the single nucleotide polymorphism (SNP) technology (Wang *et al.* 1998; Kennedy *et al.* 2003). In contrast, in most other areas of molecular ecology, genotyping errors are generally not given much importance and they are sometimes completely ignored. To illustrate this fact, we inspected all the studies published in *Molecular Ecology* in 2003 that deal with microsatellite or amplified fragment length polymorphisms (AFLP) data. Only about 6% of the 125 papers using microsatellite markers clearly mentioned an error rate value, or at least reported the percentages of allelic dropouts or false allele amplifications. Concerning AFLP data, the error rate was estimated in only two of the 14 studies. This does not necessarily mean that researchers do not evaluate the reproducibility of their typings, as many studies indicate precautions

to limit genotyping errors. However, without any published indications about error rates, one might suspect the reliability of some typings, especially in case of curious genotypes or controversial results.

In this context, this work has three main objectives. First, we aim to demonstrate that erroneous genotypes are a reality that needs to be faced. For this purpose, we use four data sets from our laboratory to identify genotyping errors sources and quantify the error rate: two sets of brown bear (*Ursus arctos*) microsatellite data obtained either by non-invasive methods or tissue sampling, and AFLP data for common frog (*Rana temporaria*) and dwarf birch (*Betula nana*). It is important to note that three of the four data sets involve standard tissue-derived samples, not samples suspected to be problematic. Second, we would like to widen the debate on typing errors to population genetics, and point out the benefits of considering this issue when making use of molecular markers for the purpose of genotyping. Third, we outline a general protocol designed especially to deal with these typing errors before the genotyping procedure, by taking precautions to limit them, and subsequently, by estimating the error rate and validating the results.

Materials and methods

General strategy and definition of the error rate

To illustrate how to track and assess genotyping errors, we chose four case studies representing a wide range of population genetics investigations (parentage analysis, population size estimation, detection of loci under selection and phylogeographical studies). These case studies are complementary as they differ in their sampling strategies (noninvasive or tissue sampling), in the type of organism studied (plant or animal) and the molecular markers used (microsatellites or AFLPs).

Unless specified, the error rate was calculated at the allelic level. For microsatellites, it was estimated as the ratio between observed number of allelic differences and total number of allelic comparisons. For AFLPs, as only the phenotype (band presence or absence) is accessible, the error rate was estimated as the ratio between observed number of phenotypic differences and total number of phenotypic comparisons.

*Case study 1: microsatellites from brown bear (*Ursus arctos*) tissues*

Context of the study and sampling. This study aimed to investigate the mating system of the brown bear in south-central Sweden by determining parentage of cubs using genetic tools (Swenson *et al.* 1998; Bellemain *et al.* unpublished data). Tissue samples obtained from live and hunter-killed

bears were stored in 95% ethanol after collection until DNA extraction.

DNA extractions and typing. DNA extractions from tissues were carried out using the QIAamp Tissue Kit (Qiagen) involving overnight digestion with proteinase K. Tubes without tissue samples were treated in the same way in order to check for contaminations. Microsatellites primers described in Paetkau & Strobeck (1994), Paetkau *et al.* (1995) and Taberlet *et al.* (1997) were used to amplify DNA with a polymerase chain reaction (PCR). Eighteen microsatellite loci (G1A, G1D, G10B, G10C, G10L, G10P, G10X, G10H, G10O, G10J, Mu05, Mu10, Mu15, Mu23, Mu50, Mu51, Mu59, Mu61) were amplified to obtain a multilocus genotype for each sample following the protocol described in Waits *et al.* (2000) and loaded on an ABI Prism 3100 DNA sequencer (Perkin-Elmer). The gels were analysed using GENESCAN® Analysis 2 and GENOTYPER® 1.1 software packages (Applied Biosystems).

Error rate checking. The genetic database included 977 individual Scandinavian brown bears genotyped between 1996 and 2003 by two different investigators (Waits *et al.* 2000; Bellemain, unpublished data). Thirty-four samples were chosen randomly from this database to be genotyped again blindly following the above protocol. The genotypes obtained were compared to the previous ones, and the number of allelic mismatches was counted.

Case study 2: microsatellites from brown bear (Ursus arctos) faeces

Context of the study and sampling. This study aimed at estimating brown bear population size from noninvasive genetic sampling in south-central Sweden. The 1904 collected faecal samples were conserved in 95% ethanol until DNA extraction (Bellemain *et al.* 2004).

DNA extractions and typing. DNA extractions were performed using the Qiamp DNA Stool origin kit (Qiagen) developed especially for this type of material and the manufacturer's extraction protocol was followed. All extractions were carried out in a room dedicated to hair and faeces processing. Tubes without faeces were included in the process as negative controls in order to check for contaminations. Six microsatellite loci (Mu10, Mu23, Mu50, Mu51, Mu59, G10L) were amplified four times following the multiplex preamplification method (Pigott *et al.* 2004) and the protocol described in Bellemain & Taberlet (2004). Samples were typed as heterozygous at one locus if both alleles appeared at least twice among the four replicates, and they were typed as homozygous if all the profiles showed four identical homozygous profiles. If neither of those cases occurred, the alleles were treated as missing

data. The samples were loaded on an ABI Prism 3100 DNA sequencer and analysed using GENEMAPPER® 3.0 (Applied Biosystems).

Genotyping error rate checking. The error rate was assessed in two ways. First we numbered the allelic differences between genotypes obtained from the same sample but from different DNA extractions (processed 8 months apart). Forty-eight already successfully genotyped samples were chosen randomly to be re-extracted blindly, reamplified another four times and analysed. Second, we counted the allelic differences between the genotypes obtained for each of the four replicates to the 'consensus' genotype for 96 samples (~5%) chosen randomly among the successfully genotyped samples. Note that these four replicates were obtained from the same DNA extraction. In both cases, the inconsistencies between the typings were classified as due to 'allelic dropout' or to 'false allele or contamination'.

Case study 3: AFLPs from common frog (Rana temporaria) tissues

Context of the study and sampling. The purpose of this study was to detect loci under selection reflecting local adaptive divergence between lowland and mountain populations of common frog (Bonin *et al.* unpublished data). F_{ST} values between these two groups are expected to be 'outlier' for loci under diversifying selection, i.e. lower or higher than expected based on neutral evolution (Beaumont & Nichols 1996). Therefore, AFLP markers were chosen because these provide a large coverage of the genome. In total, 192 samples consisting of adult frog fingers and live tadpoles were collected in six plain or mountain populations of the North French Alps. Frog fingers were stored in silica gel and tadpoles were kept alive until DNA extraction.

DNA extractions and typing. Total DNA was extracted using the DNeasy Tissue Kit (Qiagen) following the manufacturer's instructions. The AFLP procedure was modified slightly from Ajmone-Marsan *et al.* (1997). Genomic DNA (400 ng) was digested first with 5 U of *TaqI* at 65 °C for 2 h, and second with 5 U of *EcoRI* at 37 °C for 2 h. Double-stranded *TaqI* and *EcoRI* adapters were then ligated to restriction fragments for 3 h at 37 °C. Five different primer pairs (Table 1) were used for the selective amplification following the PCR program given in Gaudet *et al.* (2000). Selective products were analysed on an ABI Prism 3100 DNA sequencer and AFLP patterns were visualized with GENESCAN® Analysis 3.7 (Applied Biosystems) and GENOGAPHER 1.6.0 (Benham *et al.* 1999; <http://hordeum.oscs.montana.edu/genographer/>). AFLP profiles were scored according to the absence/presence of peaks, i.e. as dominant markers. A drop in intensity is generally clearly distinguishable between high intensity peaks

Primer pair	Sequence primer <i>EcoRI</i>	Sequence primer <i>TaqI</i>
<i>Rana temporaria</i>		
1	5'-GACTGCGTACCAATTC ACG -3'	5'-GATGAGTCCTGAGCGAA AAAC -3'
2	5'-GACTGCGTACCAATTC ACG -3'	5'-GATGAGTCCTGAGCGAA ACA -3'
3	5'-GACTGCGTACCAATTC ACT -3'	5'-GATGAGTCCTGAGCGAA AAAC -3'
4	5'-GACTGCGTACCAATTC ACT -3'	5'-GATGAGTCCTGAGCGAA ACT -3'
5	5'-GACTGCGTACCAATTC ACT -3'	5'-GATGAGTCCTGAGCGAA CAA -3'
Primer pair	Sequence primer <i>EcoRI</i>	Sequence primer <i>MseI</i>
<i>Betula nana</i>		
6	5'-GACTGCGTACCAATTC AGT -3'	5'-GATGAGTCCTGAGTAA CTC -3'
7	5'-GACTGCGTACCAATTC ATG -3'	5'-GATGAGTCCTGAGTAA CAC -3'
8	5'-GACTGCGTACCAATTC ATG -3'	5'-GATGAGTCCTGAGTAA CAA -3'

Table 1 Primer pairs used for the AFLP procedure. Selective bases are shown in bold type

(corresponding to selective amplification) and low intensity peaks (corresponding to nonselective amplification, i.e. background noise). This discontinuity was scrutinized to set up the scoring threshold, and it was usually situated around 10% of the highest peak's intensity for the marker under consideration.

Before error checking, several precautions were taken to ensure the reliability of the genotypes. First, negative controls were run at each step of the genotyping process to check for exogenous contaminations. Second, samples with poor DNA quality, i.e. with fragmented genomic DNA on agarose gel, were not genotyped. Third, reproducibility was monitored using intra- and interrun replicates. Fourth, samples with odd profiles, i.e. where most of the peaks are not observed in other profiles, were run a second time. If the genotyping was still doubtful, these samples were discarded.

Genotyping error rate checking. First, in order to estimate the genotyping error rate, 24 blind DNA samples were chosen randomly in the pool of already extracted samples, run and analysed as ordinary samples by the same person. It was not possible to proceed from independent extractions, because the whole tissue samples had been used for the first extraction. Matrices obtained for the blind and the original samples were then compared to calculate the error rate and the main error sources were identified. Second, the raw profiles for primer pairs 1, 2 and 3 (Table 1) were scored by two different people, with no scoring method decided a priori in order to evaluate the amount of subjectivity entering into the scoring process. Levels of genetic differentiation and outlier loci determined with the two resulting data sets were then compared. For this purpose, assuming Hardy–Weinberg equilibrium, F_{ST} values between mountain and plain populations were calculated independently for each marker in each data set and the correlation between the two resulting F_{ST} values was tested.

Case study 4: AFLPs from dwarf birch (*Betula nana*)

Context of the study and sampling. This study aimed to analyse the postglacial phylogeography of *Betula nana* (Bronken Eidesen *et al.* unpublished data). Leaves of *B. nana* were sampled from 240 plants in 25 populations throughout the distribution range, with emphasis on the North Atlantic region. Samples were stored immediately in silica gel and for 19 populations, one plant was sampled randomly twice, to be used as a blind replicate throughout the analyses.

DNA extractions and typing. Total DNA was extracted using DNeasy Plant Mini Kit or DNeasy 96 Plant Kit (Qiagen) according to the manufacturer's instructions. The AFLP protocol was carried out following Gaudeul *et al.* (2000) using three selective primer combinations (Table 1). The scoring procedure and the precautions taken to ensure reliability of the genotypes were the same ones as described above for *R. temporaria*.

Genotyping error rate checking. First, we estimated the technical difference rate before scoring, i.e. the rate of artefactual differences created during laboratory work. Differences in peak intensities due to irregular PCR efficiencies, slight shifts between two homologous peaks occurring during the migration, restriction anomalies or parasite peaks caused by contamination are common examples of these technical differences. Their immediate consequence is that two profiles obtained independently for the same individual are not always exactly identical. To estimate the technical difference rate, the duplicated AFLP profiles were compared for differences between all fragments using GENESCAN® Analysis in the range from 50 to 500 base pairs (bp). The technical difference rate was calculated as the number of differences per profile divided by the total number of fragments per profile.

Second, we estimated the error rate after scoring by comparing the 1/0 matrices obtained for the duplicated

samples. Differences detected here could be due either to the technical work, and/or to the subjectivity introduced during the scoring process, for instance when evaluating whether a fragment with low intensity should be scored as present or not.

Third, the raw profiles were scored independently by two different people (hereafter referred to as double scoring). To compare the biological contents of the two resulting data sets, two independent principal coordinate (PCO) analyses were run for each scoring, and the correlation between the values obtained for the first axis in both analyses was tested.

Results

Case study 1: microsatellites from brown bear tissues

In total, 1209 alleles were compared, i.e. 34 samples typed for 36 alleles, except for 15 alleles that could be typed for only one extract but not the other. Ten genotyping errors were identified, including six allelic dropouts and four false alleles or contaminations. This gave an error rate of 0.8%. Six (17.6%) multilocus genotypes contained at least one error among the 34 genotypes checked.

Case study 2: microsatellites from brown bear faeces

Genotyping error rate between consensus genotypes, i.e. genotypes obtained from two independent extractions from the same faecal sample. We obtained genotypes for 47 of the 48 re-extracted samples. One microsatellite (Mu10) failed to amplify for 24 DNA extracts from the second extraction. We also disregarded six nonamplifying alleles that could be typed for one extract but not the other. In total, 10 allelic differences were identified of the 510 comparisons, giving a genotyping error rate of 2.0%. Actually, the erroneous alleles all belonged to the same multilocus genotype among the 47 genotypes checked (2.1% of erroneous genotypes).

Genotyping error rate between replicates from the same DNA extract. We compared the genotypes of each of the four replicates to the consensus genotype for each of the 96 samples for six microsatellite loci. Thirty-four non-amplifying alleles were disregarded from the comparison. Overall, 28 allelic differences were found of 2270 alleles examined, and we obtained a genotyping error rate of 1.2%. Allelic dropout accounted for 89% of the inconsistencies, and false alleles or contaminations for 11%.

Case study 3: AFLP from frog tissues

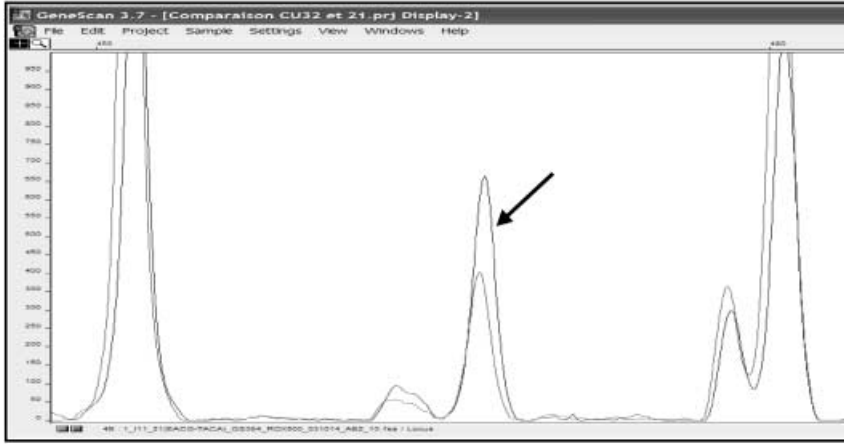
A total of 222 polymorphic bands were chosen as markers among the five primer combinations and 192 individuals were screened for these markers. One blind sample was a negative control (no DNA) and showed no amplification. The remaining 23 blind samples were all assigned correctly to the corresponding individual after amplification and blind scoring. However, 174 differences were observed of 5106 phenotypic comparisons, giving an error rate of 3.4%. It appeared that differences were concentrated on seven markers and one blind sample, whose profile was particularly weak. After excluding these seven markers and this individual from the calculation, the error rate between blind samples and corresponding individuals was 2.0% (96 differences of 4730 phenotypic comparisons) and 20 of 23 multilocus genotypes contained at least one error (86.6%). Differences between two profiles were due mainly to disparities in peak fluorescence intensities (63%, Fig. 1a) and a slight shift between two peaks (21%, Fig. 1b). Peak appearance or disappearance accounted only for 9% of the differences, and were observed especially for the weakest profiles, so they could be explained by stochastic anomalies in the amplification.

Results of the double scoring by two different experimenters are summarized in Table 2. The two people tended not to choose the same markers on a given profile: only 44% of the chosen markers were found in both data

Table 2 Double scoring of raw profiles for *Rana temporaria* and *Betula nana*. The percentages indicated were calculated in relation to the total number of markers for each primer pair

Organism	Primer pair	Total no. of markers	% scored by person 1	% scored by person 2	% scored by both people	% markers scored by person 1 only	% markers scored by person 2 only
<i>Rana temporaria</i>	1	61	77.05	73.77	50.82	26.23	22.95
	2	73	86.30	60.27	46.57	39.72	13.70
	3	71	76.06	60.56	36.62	39.44	23.94
	Mean	68.33	79.80	64.87	44.67	35.13	20.20
<i>Betula nana</i>	6	34	82.35	67.65	50.00	32.35	17.65
	7	47	82.98	55.32	38.30	44.68	17.02
	8	33	66.67	57.57	24.24	42.42	33.33
	Mean	38.00	77.33	62.22	37.51	39.82	22.67

(a)



(b)

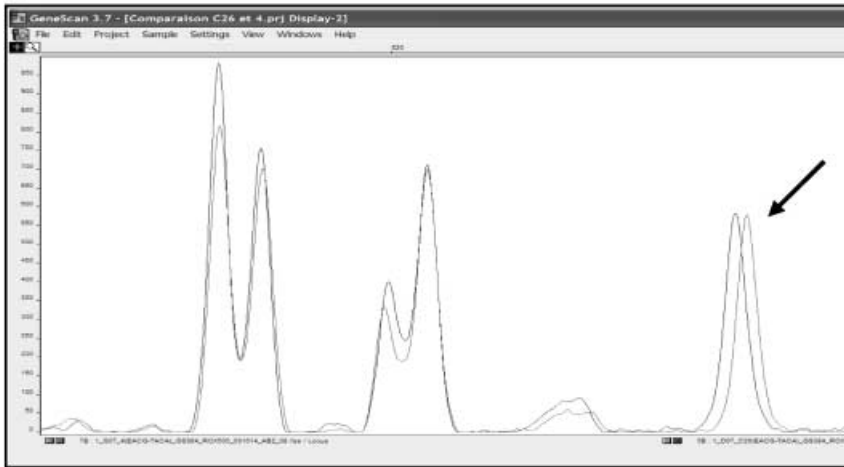


Fig. 1 Main error sources in AFLP genotyping. The two overlapping profiles were obtained independently from the same individual (species: *Rana temporaria*). The black arrow indicates the peaks under consideration. (a) Differences in peak intensities. (b) Slight shift between two peaks.

sets (91 markers of 205 in total). The genotypes based on these 91 markers were compared for 171 individuals. Four markers were scored totally differently, and were excluded from the calculation. The difference between the two independent scorings was 2.1%, i.e. 308 differences of 14 877 comparisons. The F_{ST} values between plain and mountain populations were determined in each data set and compared in Fig. 2. After normalization of the F_{ST} values for the 91 markers in common, the Pearson correlation coefficient was 0.94 ($P = 10^{-4}$). When considering 71 markers chosen by one scorer only and scored afterwards by the second one, the Pearson correlation coefficient was 0.92 ($P = 10^{-4}$).

Case study 4: AFLPs from dwarf birch

The DNA extraction turned out to be a problematic step in this study, and several samples were discarded due to too

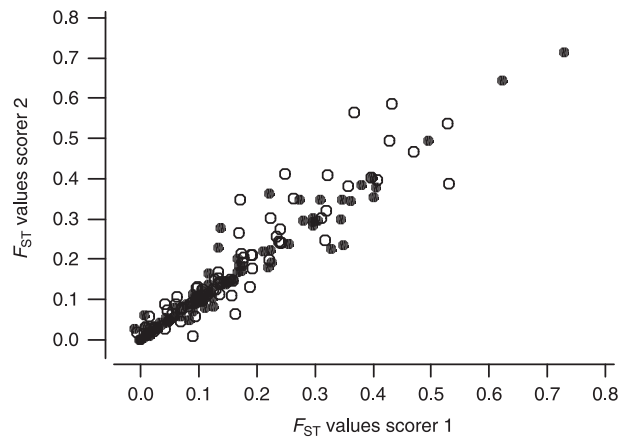


Fig. 2 Comparative F_{ST} values for the two different scorings of the total *Rana temporaria* data set. Each point represents one marker present in both data sets (closed circles) or chosen by one scorer first and then rescored by the second scorer (open circles).

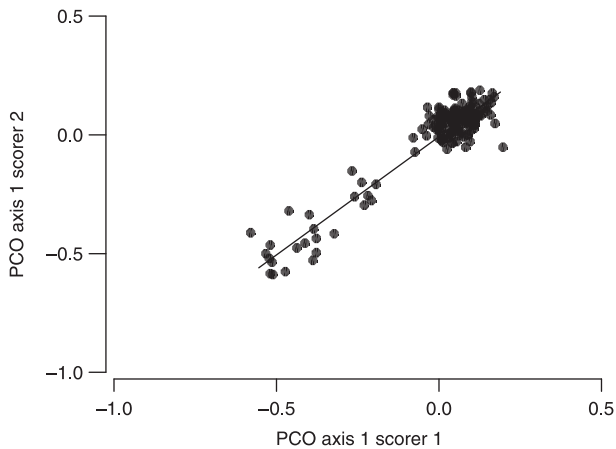


Fig. 3 Comparative values obtained for the first axis in the independent principal coordinate analyses executed on the two different scorings of the total *Betula nana* data set (174 individuals). Person 1 scored 83 markers, person 2 scored 69 markers. Each point represents one individual. The equation of the regression line is not significantly different from $y = x$ (Student's test, d.f. = 172, $\alpha = 0.05$, $t = 0.96$ for the slope and $t = 0.10$ for the constant).

low DNA concentration. In total, 83 polymorphic markers were scored for the three primer combinations, and 174 plants were genotyped. Eleven duplicated samples worked successfully for all primer combinations and were used to calculate the technical difference and the error rate. When comparing the profiles of duplicated samples, the average number of differences per profile was 3.1, and the average number of fragments per profile was 241.1 (combined for all three primer combinations). Thus, the technical difference rate was 1.3%.

After scoring, the genotyping error rate between the blind samples and the corresponding individuals was 2.6%, i.e. 25 differences of 968 phenotypic comparisons. The differences were not distributed evenly among the comparisons: the comparisons with an enhanced number of technical differences also had more genotyping errors. These errors were due mainly to intensity differences between profiles.

Results of the double scoring of the same profiles are summarized in Table 2. As for *R. temporaria*, the two scorings were not based exactly on the same polymorphic markers, as only about 38% of the chosen markers were in common, i.e. 43 markers of 114. The genotypes based on these 43 markers were compared for all 174 plants, giving a rate of difference of 1.5% between the two scorings, i.e. 112 differences of 7482 comparisons. Although they did not include exactly the same sets of markers, the two scorings displayed the same main patterns in the PCO analyses. The values obtained for the first PCO axis in each data set are showed in Fig. 3. The Pearson correlation coefficient was 0.93 ($P = 10^{-4}$).

Discussion

Causes of genotyping errors

Genotyping errors can be generated during all steps of the genotyping process by causes that may be multiple, cryptic and unpredictable. In the literature only technical causes are well documented, because they are easier to identify. They include, for example, amplification artefacts (Koonjul *et al.* 1999; Rodriguez *et al.* 2001), biochemical anomalies (Polisky *et al.* 1975; Smith *et al.* 1995), electrophoresis discrepancies (Fernando *et al.* 2001), laboratory temperature variation (Davison & Chiba 2003), material and protocol used (Delmotte *et al.* 2001; Papa *et al.* 2004) or template DNA quality or quantity (Goossens *et al.* 1998; Matthes *et al.* 1998; Bradley & Vigilant 2002). Errors due to technical causes are in general related inversely to the quality of reagents and equipment, and to the organization of the laboratory in different rooms to avoid contaminations. On the other hand human factors, such as lack of care or skill, thoughtlessness and subjectivity are largely ignored in the literature as potential causes of errors, with a few exceptions (Ewen *et al.* 2000; Paetkau 2003). However, our results showed that human factors should not be underestimated as they accounted for many of the typing errors in our data sets, especially when combined with technical causes. Sample swaps, pipetting errors or confusion in the data entry after scoring are common errors due to human factors when dealing with many individuals. Due to the infinite potential sources of genotyping errors, it is impossible to identify and eradicate them all. In order to minimize the occurrence of errors, one should instead try to focus on the most probable causes.

Genotyping errors in our data sets

Table 3 gives an overview of all our results. In our two microsatellite data sets, the genotyping error rate varied depending on the type of sample and assessed error. For the tissue samples, we calculated a genotyping error rate of 0.8%. For the faecal samples, the error rate among replicates from the same DNA extract approximated 1.2% and the error rate between consensus genotypes from independent extractions was 2.0%. The latter errors were probably caused by a sample mix-up because all inconsistent alleles occurred in one genotype and corresponded to the alleles of another sample.

Allelic dropouts were found to be non-negligible both in tissue and faecal DNA samples. This error category is linked intimately to the quality/quantity of the DNA contained in the sample and is considered mainly in noninvasive studies, because of usually low quantity and degraded DNA in samples such as faeces, hairs or feathers (Taberlet *et al.* 1996; Goossens *et al.* 1998; Taberlet & Luikart 1999).

Table 3 Overview of the results

Molecular markers used	Microsatellites		AFLPs	
Species	<i>Ursus arctos</i>		<i>Rana temporaria</i>	<i>Betula nana</i>
Sampling	Tissue	Faeces	Tissue	Tissue
Genotyping error rate between alleles	0.8%	2% (between consensus genotypes) 1.2% (between replicates)	2.0%	2.6%
Frequency of erroneous multilocus genotypes	25.1% (theoretically) 17.6% (in practice)	21.5% (theoretically) 2.1% (in practice)	98.8% (theoretically) 86.6% (in practice)	
Main error sources	Allelic dropouts (+ human error)		Differences in intensity between profiles Human error (scoring subjectivity)	
Technical difference rate				1.3%
Difference rate between independent scorings			2.1% (44% markers in common)	1.5% (38% markers in common)

However, our results show that accurately amplifying DNA from tissue samples can be challenging as well, even if such technical difficulties are probably under-represented in the literature (Jeffery *et al.* 2001).

In the study using microsatellites from brown bear faeces we used the multitube approach, based on the replication of the DNA amplifications up to eight times to determine a consensus genotype (Taberlet *et al.* 1999). This approach is used widely in noninvasive studies to deal with the allelic dropout problem but has been criticized for its financial cost. Several alternative strategies have been developed in order to avoid too many amplifications (Miller *et al.* 2002; McKelvey & Schwartz 2003; Paetkau 2003). Each of them has proved its efficiency in detecting allelic dropouts in some particular cases. However, they are sometimes difficult to put into practice as they limit automation and may consequently increase the risks of errors due to human handling. Moreover, these alternative methods depend on strong assumptions such as low error rates (Mowat & Paetkau 2002), absence of other error types (Miller *et al.* 2002) or sufficient number of amplified loci (McKelvey & Schwartz 2003). As a result, they cannot be generalized systematically. Therefore, when starting a noninvasive study, one should first determine the best approach in order to achieve accuracy of the genotyping.

In our two AFLP data sets, the genotyping error rates (2.0% for *R. temporaria* and 2.6% for *B. nana*) were consistent with the results of previous reproducibility tests reporting AFLP error rates below 5% in plants (Jones *et al.* 1997; Hansen *et al.* 1999) as well as in animals (Ajmone-Marsan *et al.* 1997; Ajmone-Marsan *et al.* 2001; Bagley *et al.* 2001). Compared to other multilocus genotyping methods, such as RAPD, the AFLP procedure is highly reproducible mainly because of stringent hybridization conditions (Vos

et al. 1995), but some classical genotyping error sources are well documented in the literature. The protocol is known to be sensitive to contamination, especially from bacterial or fungi exogenous DNA (Savelkoul *et al.* 1999; Dyer & Leonard 2000), because amplification is not taxon-specific. PCR inhibition or restriction artefacts have also been reported previously (Polisky *et al.* 1975; Koonjul *et al.* 1999), as well as comigration of nonhomologous fragments (O'Hanlon & Peakall 2000; Vekemans *et al.* 2002). Vos *et al.* (1995) reported that reproducibility could be affected if the selective primers have more than two selective bases compared to the preselective primers, because of possible nonspecific annealing during the amplification. Our experience also shows that dealing with many samples or populations tends to increase the error rate (Bronken Eidsen *et al.* unpublished data). Here, we first emphasized what we called the 'technical differences', i.e. differences due the fact that laboratories conditions can be changeable. Our results show that they are rare (1.3% of differences between profiles for *B. nana*) and can be assessed simply by comparing profiles from replicate samples. Some of these technical differences are due to the appearance of low-intensity peaks that would not be chosen as markers anyway, or variations in the peak intensities. As a consequence, only some of the technical differences result in genotyping errors. However, we noticed that typing errors often come from a technical difference combined with a misreading during the scoring process. In particular, 63% of the genotyping errors in *R. temporaria* were actually disparities in the peak fluorescence intensity misinterpreted as real differences. In order to limit the extent of typing errors in AFLP, technical differences should thus be avoided as much as possible, especially disparities in peak intensities. This can be achieved by careful standardization of the amount of DNA used in the global procedure. According to our experience,

limiting the technical differences is also greatly facilitated by normalization of the profiles from different runs with software packages such as GENOGRAPHER.

We also examined the differences encountered between scorings from different experimenters. To our knowledge, estimates of this kind of differences have never been reported in the literature for AFLPs. Our results show that independent scoring of each profile by two people can lead to differences in both the set of markers chosen and the genotypes for the markers chosen by both scorers. Considering the choice of markers, we found that less than half the peaks that one scorer categorized as markers were also chosen by a second scorer. This major discrepancy in marker choice indicates that even if clear rules and written procedures are established, this step relies mainly on subjectivity and experience. Considering only the markers chosen by both scorers, the difference rates were comparable for *B. nana* (1.5%) and *R. temporaria* (2.0%). None the less, the effect of the scoring differences on the final conclusion depends on whether the biological conclusions are derived by averaging over the whole data set or by comparing one subset of data with another. For *B. nana*, the whole data set was used to investigate the phylogeography of the species. Even if the two data sets originating from the two scorings were not perfectly consistent and did not include the same sets of markers, the extracted biological inferences were similar (Fig. 3). This suggests that in at least some cases, genotyping errors do not bias the results but just add some noise to the biological signal and that all the markers express the same information. On the other hand, for *R. temporaria*, the analysis considered each marker independently from the others. F_{ST} values were statistically comparable for the markers present in both data sets (Fig. 2) but a distortion can none the less be observed for markers chosen by only one scorer, especially for high F_{ST} values indicative of markers likely to be outliers. In marker-specific approaches such as outlier loci detection, double scoring of AFLP profiles can be used to limit the bias linked to scoring subjectivity.

In short, we showed that the scoring process is the most error-prone step in the AFLP procedure, generating most of the errors, because it relies on subjective decisions. None the less, following several straightforward recommendations can minimize the amount of subjectivity introduced in the scoring. Before setting the scoring threshold, we suggest looking for a drop in intensity among the peaks corresponding to the marker under consideration. A clear discontinuity indicates the frontier between nonselective (i.e. background) and selective amplifications and using different software packages is helpful to scrutinize this drop in intensity, as the AFLP profiles are presented in different but complementary ways. Moreover, software packages such as GENOGRAPHER allow a convenient semi-automated scoring, which improves the overall data reli-

ability (Papa *et al.* 2004), even if manual checking must still be performed.

It is important to note that even a low percentage of genotyping errors can lead to a high percentage of incorrect multilocus genotypes (having at least one error). Assuming a stochastic and independent distribution of errors, the probability P of encountering at least one error in a multilocus genotype (n loci with an error rate of r_i for locus i) can be estimated with the following formula:

$$P = 1 - \prod_{i=1}^n (1 - r_i)^x$$

where $x = 2n$ for codominant markers (microsatellites) and $x = n$ for dominant markers (AFLPs). [erratum: see below](#)

In our microsatellite tissue database, considering 18 loci in a genotype, a 0.8% genotyping error rate implies a 25.1% theoretical probability of encountering at least one error per genotype. For the faecal microsatellite database, considering six loci per genotype and an error rate of 2%, this probability is equal to 21.5%. In our AFLP data sets, this theoretical probability rises up to 98.8% for *R. temporaria* (222 markers). Actually, the frequency of erroneous genotypes found in our data sets was lower than this theoretical probability, because of a nonindependent distribution of errors (several errors in the same genotype). Increasing the number of markers tends to amplify the background noise due to genotyping errors (Waits & Leberg 2000; McKelvey & Schwartz 2003), so a compromise should be found between the number of markers necessary to obtain reliable results and the tolerated error rate.

Recommendations for limiting and identifying genotyping errors

When examining population genetics papers, we mainly found (i) no quantitative estimation of errors and/or (ii) protocols to reduce genotyping errors that are sporadic and designed primarily to address only the specific type of error that is expected. There is a clear lack of guidelines to account for the reality of genotyping errors. Therefore, in Table 4, we propose a comprehensive list of actions to ensure a minimal error rate and a good estimate of that rate. We would like to emphasize here several crucial points on this list. First, performing blind samples is strongly recommended, because it is the most rigorous way to estimate the error rate. Blind samples go through the entire genotyping process and therefore accumulate the effects of all potential error sources. We are aware that performing blind samples is expensive and time-consuming; however, as errors often are rare throughout data sets, an accurate estimation of their rate requires a sufficient number of blind replicates (5–10% of the samples). Second, automation of laboratory work has proved to be a reliable way to limit genotyping errors as genetic data sets

Table 4 List of recommendations to track and assess genotyping errors (before and after the genotyping process, respectively)

Step	Recommendations
Whole process	
Before and after	Perform blind samples
Sampling	
Before	Use a standard protocol for labelling and conserving samples Use an updated database Carry out the sampling in good field conditions (e.g. weather, temperatures, etc.)
DNA extraction	
Before	Perform negative control to monitor contaminations Follow a rigorous protocol (automation) Extract DNA in a different room if problematic samples (low quality/quantity DNA)
DNA treatment/amplification	
Before	Perform negative control to monitor contaminations Perform a pilot study (check Mendelian transmission and reproducibility) Acquire a good knowledge of the marker and its technical limitations Follow a rigorous protocol (automation) Discard samples with amplification/repeatability problems Include a previously typed sample in each amplification as a reference
After	Replicate amplifications (for both noninvasive and ordinary samples) Perform independent repeatability tests Confirm the amplifications in a different laboratory in case of doubtful results
Scoring	
Before	Favour automated scoring and check data by hand Obtain a good experience with the marker scoring prior to any genotyping session Discard low quality DNA samples Use reference samples to control scoring
After	Cross-read the data sets
Analysis	
Before	Eliminate suspicious markers (clean the data)
After	Compare genetic data with all other available data (field data, geographical data or literature data) Quantify the overall genotyping error rate and decide if it is low enough for the intended purpose Include the genotyping error rate in the analysis programs, if possible Discard markers whose frequency is in the range of the error rate Consider genotyping errors as a possible cause of Hardy–Weinberg or linkage disequilibrium

tend to become larger and larger, increasing the error risks linked to the handling of a huge number of samples, such as intercontaminations, sample swaps or confusion in the data entry. Finally, as discussed previously, human factors turn out to have a major impact on the extent of genotyping errors. Consequently, experience and rigor in the laboratory work is necessary before any extensive genotyping is undertaken.

In some particular cases, the genetic data can be confronted with additional information, allowing a better detection of genotyping errors. It is, for instance, possible to check the accuracy of the genotyping results in the field (Bellemain *et al.* 2004). Genetic data can also be compared with pedigree data to confirm Mendelian transmission of alleles between parents and offspring. This approach has experienced considerable success in mapping studies, where statistical tools have been developed to track such

Mendelian inconsistencies (Ewen *et al.* 2000; Sobel *et al.* 2002). However, it has been demonstrated that more than 25% of typing errors can be Mendelian-compatible (Douglas *et al.* 2002), which makes them harder to detect.

What to do with the genotyping error rate?

As a first step, the error rate can be used to detect odd genotypes or unreliable markers (markers that are unstable or difficult to score) and to clean up the data. It helps to determine a second and final error rate which, ideally, is lower. In the *R. temporaria* data set, the first error rate calculated with blind samples (3.4%) allowed us to discard seven untrustworthy markers of 222 markers in total, and to obtain a smaller error rate (2.0%). In a second step, the final error rate should be regarded as a means to evaluate the data quality. Mentioning the error rate should

not discredit the final conclusion drawn from the data. On the contrary, it should be considered as a proof of data quality and thus could definitively promote the use of those data in other studies. Results presented without any indication of statistical significance are generally not considered acceptable and in the same way, stating the genotyping error rate in population genetics studies should become a convention such as the *P*-value in statistical tests. In case of an unusually high error rate, this value can also give a useful clue about technical difficulties encountered and this would benefit the scientific community greatly.

Once the error rate has been estimated, a decision has to be made whether the results can be considered trustworthy. For this purpose, the error rate value must be first placed into its context. It is logical that the smaller the study level, the lower should be the error rate. Studies requiring individual identification (estimation of population sizes, parentage analyses) are particularly affected by high genotyping error rates (Paetkau 2003). On the other hand, when working at a population level (e.g. genetic diversity assessments, population structure investigations), trying to reach 0.1% error rate instead of 2% may require a great deal of time and cost to finally provide little benefit in terms of increased precision and reduced bias.

Statistical aspects of genotyping errors

In this work, statistical aspects of genotyping errors will not be discussed in detail, as they would require much more consideration than we can give here (see Gordon *et al.* 2002; Sobel *et al.* 2002 for a review). None the less, we would like to insist on some important challenges concerning the statistical treatment of genotyping errors. One of them is the modelling of how genotyping errors occur (Akey *et al.* 2001; Mitchell *et al.* 2003). Genotyping errors can indeed arise stochastically, or be concentrated on specific alleles or loci. For example, some microsatellite loci are particularly prone to allelic dropouts (Constable *et al.* 2001; Jeffery *et al.* 2001; Creel *et al.* 2003). The mistyped allele can also have an equal probability of being mistaken for any other allele, or a higher probability of being mistaken for a specific allele. This occurs, for example, when errors are due to a systematic contamination by the same allele.

Another important challenge regarding statistical aspects of genotyping errors is the development of methods integrating error models and rates to evaluate the impact of such errors on the final inferences. This has been investigated mainly with programs dealing with noninvasive data (Valière 2002; Valière *et al.* 2002) and pedigree data (Cercueil *et al.* 2002; Duchesne *et al.* 2002; Sobel *et al.* 2002; Wang 2004), on which genotyping errors can have particularly tremendous effects. Nevertheless, it must be realized that for parentage analyses exclusion-based methods are still used, although they assume error-free data sets

which are largely unrealistic. On the other hand, there is a critical need for population genetics software packages incorporating error models and rates as entry parameters and reporting a confidence interval for the results. Simulation programs generating data sets with a particular genotyping error model would also help to evaluate the bias they can introduce in the results. Unfortunately, such programs are rare, except for linkage studies (Akey *et al.* 2001; Hackett & Broadfoot 2003; Mitchell *et al.* 2003). Statistical analyses of the effects of typing errors, although a neglected research area, deserve more attention and we hope this study will promote an interest in this kind of research.

Acknowledgements

We would like to thank Claude Miaud, Olivier Marquis, Inger Greve Alsos, all collectors from the ARKTØK-Project and the Scandinavian Brown Bear Project for providing the samples and Ludovic Gielly, Christian Miquel and Delphine Rioux for technical assistance. We are grateful to Robin Waples, Jon Swenson and three anonymous referees for useful comments on an earlier version of the manuscript. A. Bonin was funded by the French Ministry of Research. The studies on *Ursus Arctos* were part of the Scandinavian Brown Bear Project and the study on *Betula nana* was funded by grant no. 150322/720 to C. Brochmann from the Research Council of Norway.

References

- Ajmone-Marsan P, Negrini R, Crepaldi P *et al.* (2001) Assessing genetic diversity in Italian goat populations using AFLP markers. *Animal Genetics*, **32**, 281–288.
- Ajmone-Marsan P, Valentini A, Cassandro M *et al.* (1997) AFLP markers for DNA fingerprinting in cattle. *Animal Genetics*, **28**, 418–426.
- Akey JM, Zhang K, Xiong M, Doris P, Jin L (2001) The effect that genotyping errors have on the robustness of common linkage-disequilibrium measures. *American Journal of Human Genetics*, **68**, 1447–1456.
- Akey JM, Zhang G, Zhang K, Jin L, Shriver MD (2002) Interrogating a high-density SNP map for signatures of natural selection. *Genome Research*, **12**, 1805–1814.
- Bagley MJ, Anderson SL, May B (2001) Choice of methodology for assessing genetic impacts of environmental stressors: polymorphism and reproducibility of RAPD and AFLP fingerprints. *Ecotoxicology*, **10**, 239–244.
- Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London*, **263**, 1619–1626.
- Bellemain E, Swenson JE, Tallmon DA, Brunberg S, Taberlet P (2004) Estimating population size of elusive animals using DNA from hunter-collected feces: comparing four methods for brown bears. *Conservation Biology*, in press.
- Bellemain E, Taberlet P (2004) Improved non invasive genotyping method: application to brown bear (*Ursus arctos*) faeces. *Molecular Ecology Notes*, **4**, 519–522.
- Benham J, Jeung JU, Jasieniuk M, Kanazin V, Blake T (1999) GENOGRAPHER: a graphical tool for automated fluorescent AFLP and microsatellite analysis. *Journal of Agricultural Genomics*, **4**, 399.

- Bradley BJ, Vigilant L (2002) False alleles derived from microbial DNA pose a potential source of error in microsatellite genotyping of DNA from faeces. *Molecular Ecology Notes*, **2**, 602–605.
- Buetow KH (1991) Influence of aberrant observations on high-resolution linkage analysis outcomes. *American Journal of Human Genetics*, **49**, 985–994.
- Cercueil A, Bellemain E, Manel S (2002) PARENTE: computer program for parentage analysis. *Journal of Heredity*, **93**, 458–459.
- Constable JL, Ashley MV, Goodall J, Pusey AE (2001) Noninvasive paternity assignment in Gombe chimpanzees. *Molecular Ecology*, **10**, 1279–1300.
- 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.
- Davison A, Chiba S (2003) Laboratory temperature variation is a previously unrecognized source of genotyping error during capillary electrophoresis. *Molecular Ecology Notes*, **3**, 321–323.
- Delmotte F, Leterme N, Simon JC (2001) Microsatellite allele sizing: difference between automated capillary electrophoresis and manual technique. *Biotechniques*, **31**, 810, 814–816, 818.
- Douglas JA, Skol AD, Boehnke M (2002) Probability of detection of genotyping errors and mutations as inheritance inconsistencies in nuclear-family data. *American Journal of Human Genetics*, **70**, 487–495.
- Duchesne P, Godbout MH, Bernatchez L (2002) PAPA (package for the analysis of parental allocation): a computer program for simulated and real parental allocation. *Molecular Ecology Notes*, **2**, 191–193.
- Dyer AT, Leonard KJ (2000) Contamination, error, and nonspecific molecular tools. *Phytopathology*, **90**, 565–567.
- 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.
- Fernando P, Evans BJ, Morales JC, Melnick DJ (2001) Electrophoresis artifacts — a previously unrecognized cause of error in microsatellite analysis. *Molecular Ecology Notes*, **1**, 325–328.
- Gagneux P, Woodruff DS, Boesch C (1997a) Furtive mating in female chimpanzees. *Nature*, **387**, 358–359.
- Gagneux P, Boesch C, Woodruff DS (1997b) Microsatellite scoring errors associated with noninvasive genotyping based on nuclear DNA amplified from shed hair. *Molecular Ecology*, **6**, 861–868.
- Gaudeul M, Taberlet P, Till-Bottraud I (2000) Genetic diversity in an endangered alpine plant, *Eryngium alpinum* L. (Apiaceae), inferred from amplified fragment length polymorphism markers. *Molecular Ecology*, **9**, 1625–1637.
- Gomes I, Collins A, Lonjou C *et al.* (1999) Hardy–Weinberg quality control. *Annals of Human Genetics*, **63**, 535–538.
- Goossens B, Waits LP, Taberlet P (1998) Plucked hair samples as a source of DNA: reliability of dinucleotide microsatellite genotyping. *Molecular Ecology*, **7**, 1237–1241.
- Gordon D, Finch SJ, Nothnagel M, Ott J (2002) Power and sample size calculations for case–control genetic association tests when errors are present: application to single nucleotide polymorphisms. *Human Heredity*, **54**, 22–33.
- Hackett CA, Broadfoot LB (2003) Effects of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps. *Heredity*, **90**, 33–38.
- Hansen M, Kraft T, Christiansson M, Nilsson NO (1999) Evaluation of AFLP in Beta. *Theoretical and Applied Genetics*, **98**, 845–852.
- Hofreiter M, Serre D, Poinar HN, Kuch M, Paabo S (2001) Ancient DNA. *Nature Reviews Genetics*, **2**, 353–359.
- Jeffery KJ, Keller LF, Arcese P, Bruford MW (2001) The development of microsatellite loci in the song sparrow, *Melospiza melodia* (Aves), and genotyping errors associated with good quality DNA. *Molecular Ecology Notes*, **1**, 11–13.
- 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.
- Kauer M, Dieringer D, Schlotterer C (2003) A microsatellite variability screen for positive selection associated with the ‘Out of Africa’ habitat expansion of *Drosophila melanogaster*. *Genetics*, **165**, 1137–1148.
- Kennedy GC, Matsuzaki H, Dong S *et al.* (2003) Large-scale genotyping of complex DNA. *Nature Biotechnology*, **21**, 1233–1237.
- Koonjul PK, Brandt WF, Farrant JM, Lindsey GG (1999) Inclusion of polyvinylpyrrolidone in the polymerase chain reaction reverses the inhibitory effects of polyphenolic contamination of RNA. *Nucleic Acids Research*, **27**, 915–916.
- Lincoln SE, Lander ES (1992) Systematic detection of errors in genetic linkage data. *Genomics*, **14**, 604–610.
- Matthes MC, Daly A, Edwards KJ (1998) Amplified length polymorphism (AFLP). In: *Molecular Tools for Screening Biodiversity: Plants and Animals* (eds Karp A, Isaac PG, Ingram DS), pp. 183–192. Chapman & Hall, London.
- McKelvey KS, Schwartz MK (2004) Genetic errors associated with population estimation using non-invasive molecular tagging: problems and new solutions. *Journal of Wildlife Management*, **68**, 439–448.
- Miller CR, Joyce P, Waits LP (2002) Assessing allelic dropout and genotype reliability using maximum likelihood. *Genetics*, **160**, 357–366.
- Mitchell AA, Cutler DJ, Chakravarti A (2003) Undetected genotyping errors cause apparent overtransmission of common alleles in the transmission/disequilibrium test. *American Journal of Human Genetics*, **72**, 598–610.
- Mowat G, Paetkau D (2002) Estimating marten *Martes americana* population size using hair capture and genetic tagging. *Wildlife Biology*, **8**, 201–209.
- O’Hanlon PC, Peakall R (2000) A simple method for the detection of size homoplasy among amplified fragment length polymorphism fragments. *Molecular Ecology*, **9**, 815–816.
- Paetkau D (2003) An empirical exploration of data quality in DNA-based population inventories. *Molecular Ecology*, **12**, 1375–1387.
- Paetkau D, Calvert W, Stirling I, Strobeck C (1995) Microsatellite analysis structure of population structure in Canadian polar bears. *Molecular Ecology*, **4**, 347–354.
- Paetkau D, Strobeck C (1994) Microsatellite analysis of genetic variation in black bear populations. *Molecular Ecology*, **3**, 489–495.
- Papa R, Troglio M, Ajmone-Marsan P, Nonnis Marzano F (2004) An improved protocol for the production of AFLP markers in complex genomes by means of capillary electrophoresis. *Journal of Animal Breeding and Genetics*, in press.
- Pigott M, Bellemain E, Taberlet P, Taylor A (2004) A multiplex pre-amplification method that significantly improves microsatellite amplification and error rates for faecal DNA in limiting conditions. *Conservation Genetics*, **5**, 417–420.
- Polisky B, Greene P, Garfin DE *et al.* (1975) Specificity of substrate recognition by the EcoRI restriction endonuclease. *Proceedings of the National Academy of Sciences USA*, **72**, 3310–3314.
- Rodriguez S, Visedo G, Zapata C (2001) Detection of errors in dinucleotide repeat typing by nondenaturing electrophoresis. *Electrophoresis*, **22**, 2656–2664.

- Savelkoul PH, Aarts HJ, de Haas J *et al.* (1999) Amplified-fragment length polymorphism analysis: the state of an art. *Journal of Clinical Microbiology*, **37**, 3083–3091.
- Segovia-Lerma A, Cantrell RG, Conway JM, Ray IM (2003) AFLP-based assessment of genetic diversity among nine alfalfa germplasms using bulk DNA templates. *Genome*, **46**, 51–58.
- 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.
- Sobel E, Papp JC, Lange K (2002) Detection and integration of genotyping errors in statistical genetics. *American Journal of Human Genetics*, **70**, 496–508.
- Swenson JE, Sandegren F, Bjärvall A, Wabakken P (1998) Living with success: research needs for an expanding brown bear population. *Ursus*, **10**, 17–23.
- Taberlet P, Camarra JJ, Griffin S *et al.* (1997) Noninvasive genetic tracking of the endangered Pyrenean brown bear population. *Molecular Ecology*, **6**, 869–876.
- Taberlet P, Griffin S, Goossens B *et al.* (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research*, **24**, 3189–3194.
- Taberlet P, Luikart G (1999) Non-invasive genetic sampling and individual identification. *Biological Journal of the Linnean Society*, **68**, 41–55.
- Taberlet P, Waits LP, Luikart G (1999) Noninvasive genetic sampling: look before you leap. *Trends in Ecology and Evolution*, **14**, 323–327.
- Valière N (2002) GIMLET: a computer program for analysing genetic individual identification data. *Molecular Ecology Notes*, **2**, 377–379.
- Valière N, Berthier P, Mouchiroud D, Pontier D (2002) GEMINI: software for testing the effects of genotyping errors and multi-tubes approach for individual identification. *Molecular Ecology Notes*, **2**, 83–86.
- Vekemans X, Beauwens T, Lemaire M, Roldan-Ruiz I (2002) Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Molecular Ecology*, **11**, 139–151.
- Vigilant L, Hofreiter M, Siedel H, Boesch C (2001) Paternity and relatedness in wild chimpanzee communities. *Proceedings of the National Academy of Sciences USA*, **98**, 12890–12895.
- Vos P, Hogers R, Bleeker M *et al.* (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, **23**, 4407–4414.
- Waits JL, Leberg PL (2000) Biases associated with population estimation using molecular tagging. *Animal Conservation*, **3**, 191–199.
- Waits L, Taberlet P, Swenson JE, Sandegren F, Franzen R (2000) Nuclear DNA microsatellite analysis of genetic diversity and gene flow in the Scandinavian brown bear *Ursus arctos*. *Molecular Ecology*, **9**, 421–431.
- Wang J (2004) Sibship reconstruction from genetic data with typing errors. *Genetics*, **166**, 1963–1979.
- Wang DG, Fan JB, Siao CJ *et al.* (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science*, **280**, 1077–1082.
- Xu J, Turner A, Little J, Bleecker ER, Meyers DA (2002) Positive results in association studies are associated with departure from Hardy–Weinberg equilibrium: hint for genotyping error? *Human Genetics*, **111**, 573–574.
- Yoder AD, Delefosse T (2002) The rise and fall and rise of ancient DNA studies. In: *Ancient DNA*, pp. 9–14. McGraw-Hill/Yearbook of Science and Technology, New York.

AB, EB, PBE and FP, CB, PT are PhD students and senior scientists, respectively. The two laboratories involved have a long-term concern in the production of high quality molecular genetic data. The Laboratoire d'Ecologie Alpine (Grenoble, France) is indeed a leading laboratory in the development of non-invasive molecular techniques and the National Center for Biosystematics (Oslo, Norway) has a comprehensive understanding of molecular markers (AFLPs, RAPDs) used in population genetics and phylogeographic studies. On the basis of their background and experience, both laboratories would like to promote a molecular data quality control in the population genetics field.
