TECHNICAL NOTE

Carsten Urban,[1] Ph.D.; Franz Gruber,[1] Ph.D.; Michael Kundi,[2] Ph.D.; Falko G. Falkner,[1] Ph.D.; Friedrich Dorner,[1] Ph.D.; and Thomas Hämmerle,[1] Ph.D.

A Systematic and Quantitative Analysis of PCR Template Contamination

REFERENCE: Urban C, Gruber F, Kundi M, Falkner FG, Dorner F, Hämmerle T. A systematic and quantitative analysis of PCR template contamination. J Forensic Sci 2000;45(6):1307–1311.

ABSTRACT: A quantitative and systematic analysis is provided for ubiquitously present template DNA interfering with the quantification of human DNA by PCR. Two sources contributing to DNA background were identified. The first one is interpreted as DNA present in chemicals and on equipment and the second as caused by operator handling. The amounts were equivalent to 2.5 and 8.9 pg per mL of sample, and the estimated frequencies of contamination were 65 and 35%, respectively, resulting in an effective limit of detection of 17.4 pg/mL. Below this level—named effective laboratory background—a result could not be considered as authentic. Knowledge of these parameters is important for laboratories that analyze minute amounts of human DNA by PCR for purposes such as quantification, typing, and sequencing.

KEYWORDS: forensic science, DNA typing, polymerase chain reaction, template contamination, quantification, background

The application of the polymerase chain reaction (PCR) is hampered by the occurrence of false positive results due to contaminating DNA caused by carryover of PCR products and ubiquitous present template DNA (1). Measures to avoid carryover have been established, for example, separation of pre- and post-PCR laboratory areas and enzymatic or chemical inactivation procedures of PCR products (1–3). However, the influence of contaminating template DNA on the authenticity of a result is still a matter of discussion for diagnostic, paleontological, and forensic applications of PCR (4–7). In particular, the contamination risk of PCR-based applications in forensic sciences has especially been addressed in view of the potential far-reaching consequences (8–11).

Recently, an assay based on the amplification of repetitive DNA by PCR was developed and applied to the quantification of genomic chicken DNA (12). The repetitive sequences of interest were processed together with internal standard molecules and amplified in the presence of fluorescent dye-labeled primers. Thereafter, the PCR products were separated by PAGE and quantitated by measuring laser-induced fluorescence using a commercially available DNA sequencer and appropriate software. The primary results, obtained as copies per mL of sample, were converted into mass units of DNA by applying an experimentally determined conversion factor (12). In continuation of this work, an application of this type of assay was set up and applied to quantitate human genomic DNA. Despite stringent measures to avoid contamination and, therefore, unexpectedly, a relatively high background of template DNA was observed. Thus, the necessity arose to assess the cause, the extent, and the consequences thereof described below.

[1] Baxter, Hyland-Immuno Division, Biomedical Research Center, Uferstr. 15, A-2304 Orth/Donau, Austria.

[2] Institute for Environmental Health, University of Vienna, Kinderspitalgasse 15, 1095 Vienna, Austria.

Received 21 Dec. 1999; and in revised from 27 Dec. 1999; accepted 29 Dec. 1999.

Materials and Methods

Plasmids

The plasmid pAlu+16 is a derivative of the plasmid pCRII (Invitrogen) and harbors a 150 bp long insert consisting of the sequence 5′-CGTAGTGGCG GGCGCCTGTA GTCCTCAGGA GAATGGCTTG AACCCGGAAG CTTGGTACCG AGCTCGGATC CACTAGTAAC CGGGAGGCAG AGCTTGCAGT GAGCTGAGAT CCGGCCACTG CACTCCAGCC TGGGCGACAG AGCGAGACTC-3′ cloned into the multiple cloning site. The plasmid pAlu-20 was derived by deletion of bp46 to 81 of the insert of pAlu+6. These plasmids were purified, linearized with XbaI, diluted to the appropriate working concentration, and used as minus and plus standard. They provide binding sites for the PCR primers Alu-A2/2B 5′-CGTAGTGGCG GGCGCCTGTA GT-3′ and Alu-B2 5′-GAGTCTCGCT CTGTCGCCCA GG-3′, which were designed according to the human Alu repetitive sequence described by Batzer and colleagues (Genbank Acc. No. X55922) (13) and bind at bp155-176 and bp288-267 relative to X55922, respectively. The length of the PCR products derived from pAlu+16, pAlu-20, and wild-type sequences were 150, 114, and 134 bp, respectively.

Extraction of Nucleic Acids, Amplification, and Analysis of PCR Products

For each extraction a sample volume of 400 μL was used. 15 000 and 45 000 molecules of minus and plus standard, respectively, were added in a volume of 6 μL. Amplification was done in the presence of 2.5 m*M* $MgCl_2$ for 22 cycles, with each cycle consisting of 15 s at 95°C, 30 s at 65°C, and 30 s at 72°C. Primer Alu-A2/2B was FAM (5-carboxyfluorescein) labeled.

Genomic DNA

Genomic DNAs were obtained from Clontech, digested with HindIII enzyme, and serially diluted to the appropriate concentration. The commercially available standard of the Quantiblot™ Human DNA Quantitation Kit (Perkin Elmer) was diluted from 2 to 50 pg/mL and analyzed five times. The dilution buffer used was 10 m*M* Tris/HCl pH 8.0, 0.1 m*M* EDTA, and 10 μg/mL fish sperm DNA.

Statistical Analysis

Distribution of DNA concentration for the different experiments were analyzed with respect to deviations from normality by continous Kolmogorov-Smirnov tests. Generally, observed distributions tend to be more closely fitted by log-normal distributions. To investigate the possible presence of contamination of different sources, we adopted the following strategy: The cumulative distribution function was fitted by the weighted sum of two log-normal distributions differing in mean by at least 0.3 (twofold difference in DNA concentration). Parameters were estimated by the sequential quadratic optimization algorithm. The detection limit was set at the 90th percentile of the distribution of negative controls. The 95% confidence limits of the percentile were obtained by non-parametric methods. Further experimental procedures were done as described in Ref 12.

Results

Principle, Linearity, and Specificity of the Assay

In order to develop a procedure for the quantification of human DNA, the Alu repetitive element was chosen and primers were designed according to a published consensus sequence (13). Coamplification was done with known amounts of internal standard molecules that harbored a deletion or an insertion within the region amplified and yielded PCR products different in size. These could be discriminated from the wild-type products by gel electrophoresis under denaturating conditions followed by detection of laser-induced fluorescence. Figure 1 shows a typical chromatogram representing the distribution of the intensity of blue fluorescence versus the size of PCR fragments given in nucleotides (nt). Three prominent peaks with sizes of 114, 150, and 134 nt were detected and identified as the PCR products of the minus and plus standard and wild-type sequences based on the theoretical value defined by the binding sites of the primers on the consensus sequence (13). The amount of wild-type sequences was calculated based on the quantitative fluorescence data and the known amount of standard and was given in copies per mL sample. A relation between detectable copies and mass of DNA was established by analyzing increasing amounts of human genomic DNA and plotting the results (i.e., detectable copies) versus the concentration of genomic DNA as shown in Fig. 2. The slope of the linear range was used to evaluate the conversion factor as 7800. Thus, each primary result was divided by 7800 in order to obtain the amount of DNA in pg. The specificity of the assay was tested by analyzing DNA from different sources. The assay detected human and monkey genomic DNA, but not genomic DNA from hamster, chicken, or mouse even at a concentration of 1 ng/mL (Fig. 3).

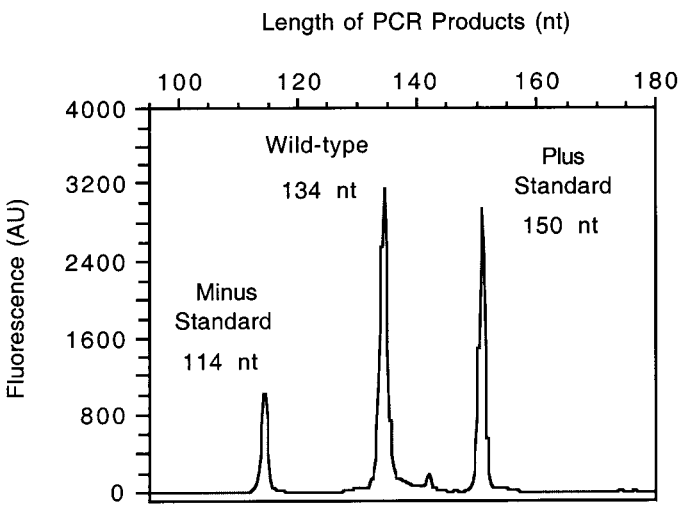

FIG. 1—*Analysis of PCR products obtained after amplification of human DNA. A chromatogram is shown obtained after separation of PCR products by PAGE followed by detection of laser-induced fluorescence. Three prominent peaks were detected having the expected lengths of PCR products of wild-type sequences and of minus and plus standards, which are 114, 150, and 134 nucleotides (nt).*

Persistent Positive Results of a Negative Run Control Indicates Presence of Ubiquitous DNA

Over a period of six months the assay was applied to determine the concentration of genomic DNA in various samples, positive and negative run controls accompanying each run. There was good agreement between the theoretical and experimental values of the concentration of the positive run controls. The theoretical values were 40 and 80 pg/mL, and the experimental values for arithmetrical mean and standard error were 48.2 ± 2.9 and 70.8 ± 3.6 pg/mL, respectively. It should be stressed that for lack of a commonly accepted standard a commercially available specimen containing human DNA was analyzed as described in methods in order to estimate the accuracy of the measurement as defined in the ICH

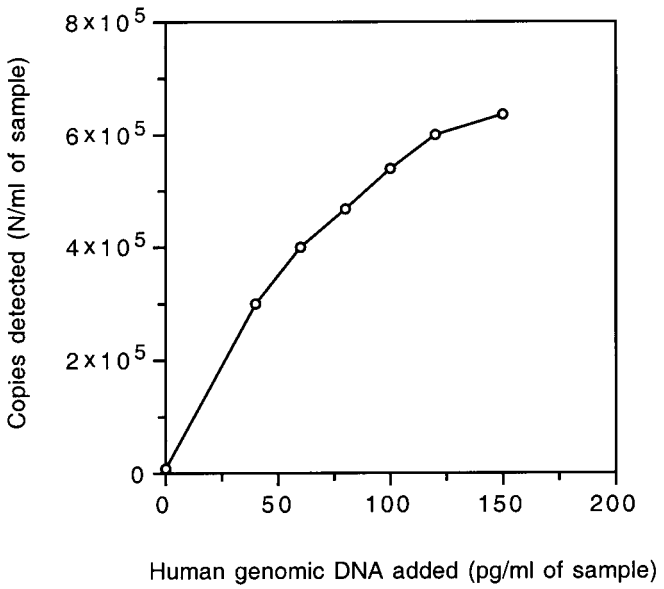

FIG. 2—*Linearity of the assay. Increasing amounts of human genomic DNA were quantitated. The number of copies of detectable Alu repetitive sequences versus the amount of input DNA was plotted. A linear increase was observed up to 100 pg/mL input DNA. Beyond that concentration the plateau phase was reached. From the range of 0 to 100 pg/m the slope was evaluated as 7800 detectable copies per pg DNA. Each data point represents the average of six determinations.*

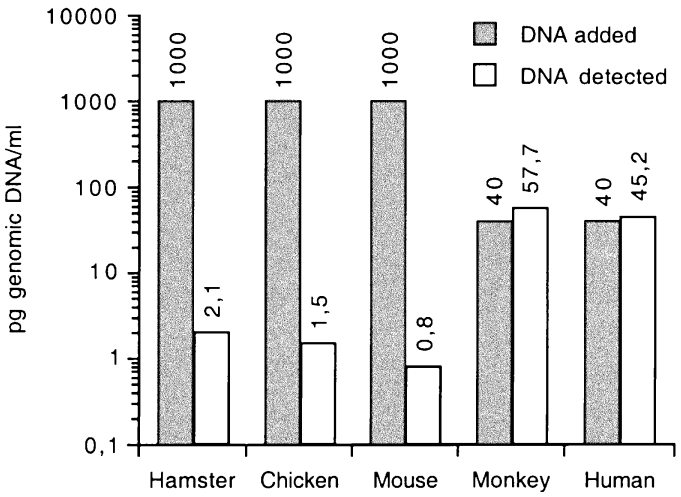

FIG. 3—*Specificity of the assay. Genomic DNA (1 ng/mL) from hamster, chicken, and mouse and genomic DNA (40 pg/mL) from African Green monkey and human sources were analyzed. The figure shows the amount of DNA added and the amount detected for each type of DNA. Monkey and human DNAs were detectable, but no other type of DNA, despite present in 25-fold higher concentration. Note that the results for hamster, chicken, and mouse DNA were presumably caused by ubiquitous present human DNA. The experiment was performed six times.*

guidelines on the validation of analytical procedures (14). The specimen was diluted to 50 pg/mL and processed five times. The results for arithmetical mean value and standard error were 56.3 ± 6.6 pg/mL, indicating an acceptable accuracy. Unexpectedly, the negative run control persistently yielded a positive result. The respective values for arithmetical mean and standard error were 8.9 ± 1.7 pg/mL. The number of data was 90. This type of false positive result was not observed when DNA of chicken embryo fibroblasts was quantitated using an assay based on the same principle (12).

False Positive Results Caused by Human Template DNA

In order to elucidate this observation further, an array of additional negative controls was analyzed. These controls—termed k1, k3, and k4—were done using water instead of sample or extract and did not contain any standard (15). The extraction control k4 is equivalent to the overall negative control except for the omittance of standard. Controls k1 and k3 both are PCR reagent controls but differ with respect to the preparation area. Control k1 was set up in the area exclusively dedicated to the storage of PCR chemicals and the preparation of reaction mixes and should detect contaminations there. Control k3 was prepared in the area where the PCR reaction mixes were combined with aliquots of extracts prior to the amplification process. Thus, control k3 should detect contamination in this area in addition to those detected by control k1. It should be emphasized that the number of handling steps and chemicals used increased from control k1 to k4 and thus the probability of detecting a positive signal. The parameters assessed were the frequency and the intensity of contamination of wild-type sequences and standards.

The frequency of contamination of PCR products of internal standards was zero in all three types of negative controls, while that of the wild-type fragments increased from 51% to 83 and 96% for controls k1, k3, and k4, respectively (Fig. 4*A*). This increase was

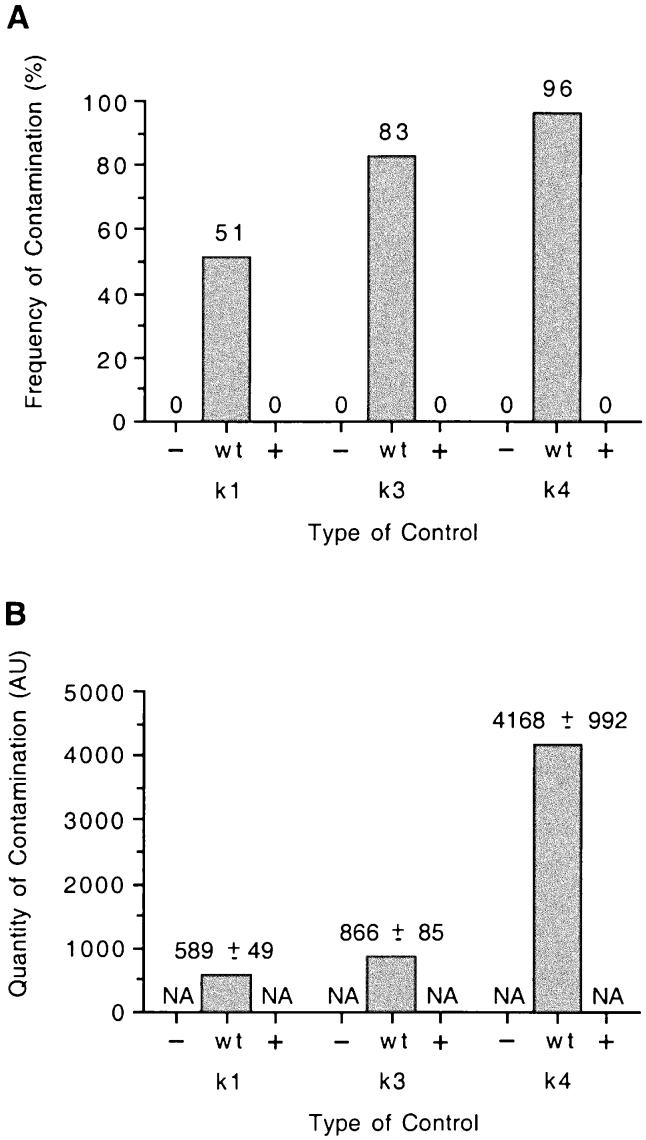

FIG. 4—*Assessment of positive signals observed in negative controls.* (A) *The frequencies of positive signals corresponding to PCR products of wild-type sequences (wt), minus (−) and plus (+) standards were evaluated in percent relative to the total number of determinations (*n = *90). The types of negative controls k1, k3, and k4 are defined in the text. The frequency of contaminations of PCR products of internal standards was zero in all controls, while that of wild-type sequences increased from 51 to 83 to 96%.* (B) *Concomitantly, an increase was observed for the average fluorescence intensity of wild-type PCR fragments. Data given as mean and SEM. NA = not available.*

accompanied by an increase of the mean fluorescence of wild-type products from 589 to 866 and 4168 arbitrary fluorescence units (Fig. 4*B*). Thus, the more handling was done and the more reagents were used, the higher was the probability of detection of a positive wild-type signal and the average intensity of it. This indicated contaminating DNA that accumulated from one or more sources during the course of the procedure. Since only the wild-type signal was detectable, the contamination was exclusively template contamination, i.e., ubiquitous human DNA, but not PCR product carryover, confirming the efficacy of the high degree of anti-contamination measures in the laboratory.

Statistical Analysis of Template DNA Contamination

Concerning both positive run controls, the *d*-values of the continuous Kolmogoroff-Smirnov test for log-normal and normal distribution were 0.0859 and 0.1381, respectively. Therefore, the geometric instead of the arithmetic means were given. These were 45 and 75 pg/mL, respectively, and the scatter factors used as indicators for the precision were 1.5 and 1.4. Note that the theoretical values were 40 and 80 pg/mL. Contrary, the data of the negative run control were derived from two populations whose geometric mean values were estimated as 2.5 and 8.9 pg/mL. The scatter factors were 2.0 and 3.2, and the frequencies of detection were 65 and 35%, respectively. The goodness of the fit was 99.8%. It is reasonable to assume that the two populations were caused by different events. Therefore, they were termed background and accidental contamination, respectively, implying that the latter was caused by operator handling and the former by template DNA present in chemicals and on equipment. It should be mentioned that background contamination was most likely present in all observations but was concealed in cases where accidental contamination was detected. A graphical presentation of the distributions of both positive and the negative run controls—including theoretical values, experimental means, scatter factors, and total number of data—are given in Figs. 5*A* to *C*.

Detection Limit Defined by Template DNA Contamination

With respect to the quantification of a given sample, the data led to the necessity to establish the authenticity of a signal by statistical means, i.e., to exclude the possibility that it was caused by accidental or background contamination. Based on the criteria that a sample had to be determined in duplicate and that the 90th percentile of the distribution of the negative controls was appropriate, the effective limit of detection, termed effective laboratory background of human DNA, was evaluated as 17.4 pg/mL (95% CI: 11.9 to 54.9 pg/mL). Thus, only values higher than this level could statistically be judged as different from background and accidental contamination and be given as the final result.

Discussion

No template contamination problem has been encountered applying this type of assay for the quantification of genomic DNA of chicken embryo fibroblasts cells (12) and Chinese Hamster Ovary cells (unpublished results) with detection limits in the range of 1 to 5 pg/mL. Thus, the contamination problem appears to be specifically linked to human DNA and seems to be an intrinsic phenomenon of an environment inhabited by humans that includes also laboratory area, reagents, and equipment.

The effective laboratory background of 17.4 pg/mL posed by the template contamination is 9.2 times the mass of 1.9 pg DNA, which is the equivalent of one haploid mammalian genome (16). Thus, it corresponds to a few to thousands of copies of the nucleic acid of interest depending on its redundancy in the genome. Mitochondrial sequences are targets for forensic and paleontological applications of PCR (5,7). Assuming that a diploid cell contains 1000 copies of mitochondrial genomes (17), the effective laboratory background can be estimated as 4600 mitochondrial copies per mL, corresponding to about 300 copies per amplification. This estimate is in good agreement with a previously made suggestion of a minimum amount of 100 to 1000 copies per amplification for paleontologial studies using PCR in order to obviate contamination problems (6). Based on the effective laboratory background determined here, 500 to 1000 mitochondrial copies per amplification are recommended.

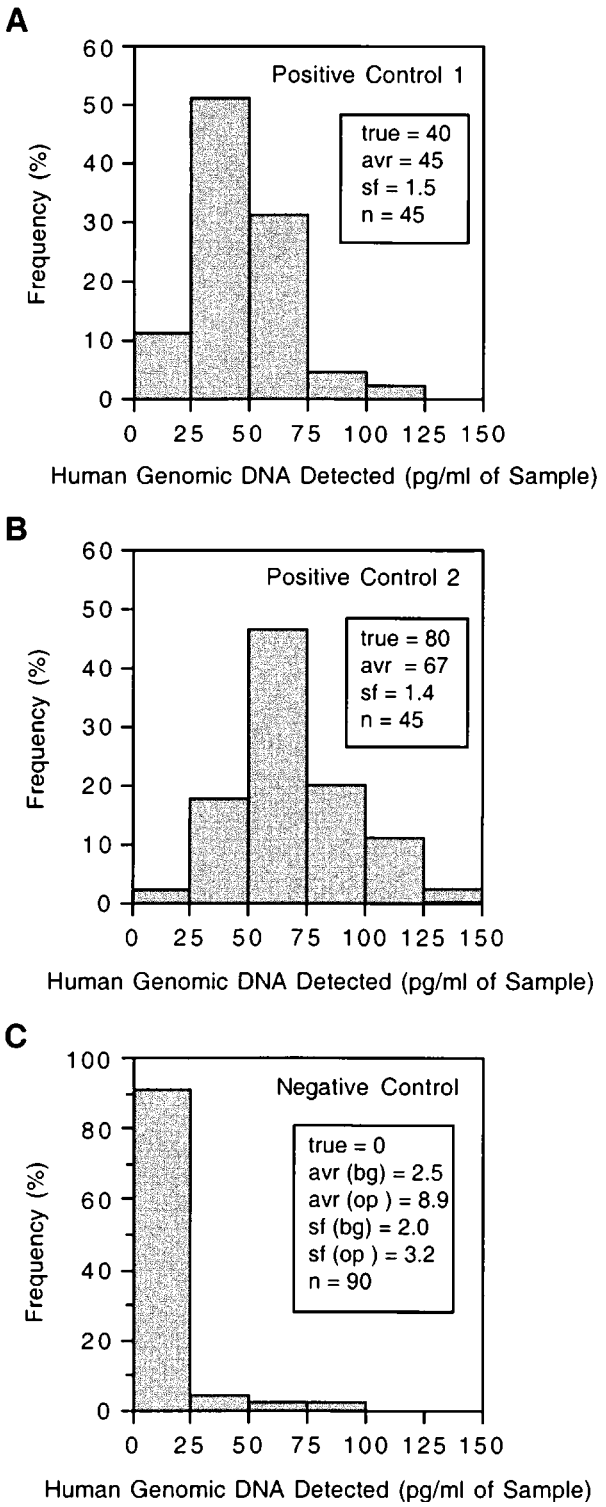

FIG. 5—*Distributions and statistics of data of positive and negative run controls. The X-axis is graduated in intervals, each having a range of 25 pg/mL. The Y-axis gives the frequency of data in the respective interval in percent of total determinations. The inserts provide the values for theoretical (true) and experimental geometric mean (avr), given in pg DNA/mL, the scatter factor (sf), and the number of determinations (*n*). Concerning both positive run controls, the respective data are derived from a single entity, that of the negative run control from two whose geometric mean values were estimated as 2.5 and 8.9 pg/mL.* (A) *Positive Control 1.* (B) *Positive Control 2.* (C) *Negative Control. Note that in the insert of Fig. 5C the values of both subpopulations of the data of the negative run control are given, i.e., background (bg) and accidental-operator-caused contamination (op).*

There might be means and procedures to lower the background observed here, for example, by replacing a particular chemical with another one from a different supplier. However, so far no particular chemical could unequivocally be identified as a predominant source for template DNA. Nevertheless, there might still be a low level of ubiquitous DNA. Also, human operators might still be involved even at an increased level of automatization of the procedure. Therefore, the effective laboratory background should be known for any given PCR-based application for the quantification, typing, and sequencing of human DNA. According to those values, the limit of detection and thus the necessary amount of DNA input per analysis can be adjusted in order to minimize the occurence of false positive results caused by the procedure.

Acknowledgments

We thank C. Hofstädter and N. Makarov for excellent technical assistance and Dr. I. Livey for critical reading of the manuscript.

References

1. Kwok S, Higuchi R. Avoiding false positives with PCR. Nature 1989;339:237–8.
2. Longo MC, Berninger MS, Hartley JL. Use of uracil DNA glycosylase to control carry-over contamination in polymerase chain reactions. Gene 1990;93:125–8.
3. Cimino GD, Metchette KC, Tessman JW, Hearst JE, Isaacs ST. Post-PCR sterilization: a method to control carryover contamination for the polymerase chain reaction. Nucleic Acids Res 1991;19:99–107.
4. Neumaier M, Braun A, Wagener C. Fundamentals of quality assessment of molecular amplification methods in clinical diagnostics. International Federation of Clinical Chemistry Scientific Division Committee on Molecular Biology Techniques. Clin Chem 1998;44:12–26.
5. Audic S, Beraud-Colomb E. Ancient DNA is thirteen years old. Nat Biotechnol 1997;15:855–8.
6. Handt O, Krings M, Ward RH, Paabo S. The retrieval of ancient human DNA sequences. Am J Hum Genet 1996;59:368–76.
7. Butler JM, Levin BC. Forensic applications of mitochondrial DNA. Trends Biotechnol 1998;16:158–62.
8. Toledano T, Quarino L, Leung S, Buffolino P, Baum H, Shaler RC. An assessment of DNA contamination risks in New York City Medical Examiner facilities. J Forensic Sci 1997;42:721–4.
9. Scherczinger CA, Ladd C, Bourke MT, Adamowicz MS, Johannes PM, Scherczinger R, et al. A systematic analysis of PCR contamination. J Forensic Sci 1999;44:1042–5.
10. Wilson MR, Polanskey D, Butler J, DiZinno JA, Replogle J, Budowle B. Extraction, PCR amplification and sequencing of mitochondrial DNA from human hair shafts. Biotechniques 1995;18:662–9.
11. Wilson MR, DiZinno JA, Polanskey D, Replogle J, Budowle B. Validation of mitochondrial DNA sequencing for forensic casework analysis. Int J Legal Med 1995;108:68–74.
12. Himmelspach M, Gruber F, Antoine G, Falkner FG, Dorner F, Hämmerle T. Specific quantitation of genomic DNA in the femtogram range by amplification of repetitive sequences. Anal Biochem 1996;242:240–7.
13. Batzer MA, Kilroy GE, Richard PE, Shaikh TH, Desselle TD, Hoppens CL, et al. Structure and variability of recently inserted Alu family members. Nucleic Acids Res 1990;18:6793–8.
14. FDA. ICH Guideline on Validating Analytical Procedures Finalized #40,709. Washington, DC: FDA, 1995.
15. Hämmerle T, Falkner FG, Dorner F. A sensitive PCR assay system for the quantitation of viral genome equivalents: hepatitis C virus HCV. Arch Virol 1996;141:2103–14.
16. Sambrook J, Fritsch EM, Maniatis T, editors. Molecular Cloning: A Laboratory Manual. 2nd ed. Cold Spring Harbor: Cold Spring Harbor Lab Press, 1989.
17. Bogenhagen D, Clayton DA. The number of mitochondrial deoxyribonucleic acid genomes in mouse L and human HeLa cells. Quantitative isolation of mitochondrial deoxyribonucleic acid. J Biol Chem 1974;249:7991–5.

Additional information and reprint requests:
Thomas Hämmerle, Ph.D.
Baxter
Hyland-Immuno Division
Biomedical Research Center
Uferstr. 15
A-2304 Orth/Donau
Austria
Tel. INT'L +43-1-20100-4306
Fax: INT'L +43-1-20100-4000
E-mail: haemmet@baxter.com