Reports

Incorporation of measurement of DNA integrity into qPCR assays

Michael J. Brisco, Sue Latham, Paul A. Bartley, and Alexander A. Morley
Department of Haematology and Genetic Pathology, Flinders University and Medical Centre, Bedford Park, South Australia, Australia

BioTechniques 49:893-897 (December 2010) doi 10.2144/000113567
Keywords: DNA integrity; DNA damage; qPCR; Poisson distribution

Supplementary material for this article is available at www.BioTechniques.com/article/113567.

Optimal accuracy of quantitative PCR (qPCR) requires correction for integrity of the target sequence. Here we combine the mathematics of the Poisson distribution and exponential amplification to show that the frequency of lesions per base (which prevent PCR amplification) can be derived from the slope of the regression line between cycle threshold (Ct) and amplicon length. We found that the amplifiable fraction (AF) of a target can be determined from this frequency and the target length. Experimental results from this method correlated with both the magnitude of a damaging agent and with other measures of DNA damage. Applying the method to a reference sequence, we determined the values for lesions/base in control samples, as well as in the AFs of the target sequence in qPCR samples collected from leukemic patients. The AFs used to calculate the final qPCR result were generally >0.5, but were <0.2 in a few samples, indicating significant degradation. We conclude that DNA damage is not always predictable; quantifying the DNA integrity of a sample and determining the AF of a specific qPCR target will improve the accuracy of qPCR and aid in the interpretation of negative results.

Quantitative PCR (qPCR) is widely used in biology and medicine to quantify target sequences of interest. The desired endpoint is an absolute or relative measure of the number of target sequences in the sample, but the actual endpoint provided by qPCR is only a measure of the number of intact, amplifiable target sequences. The difference between these two endpoints depends on the extent of damage to the DNA in the sample. Damage may occur in vivo or during sample collection, transport, storage, and processing; its nature and magnitude will usually be unknown when the sample is analyzed. To obtain an accurate qPCR result, the number of amplifiable target sequences needs to be converted to the total number of target sequences, and this requires determination of the proportion of target sequences that are amplifiable (the amplifiable fraction, or AF). In a clinical setting, an accurate result is essential: for example, when a decision to perform a marrow transplant rests on the level of minimal residual disease (MRD), the utmost accuracy is required.

Several different approaches, each with its own assumptions, have been used to assess DNA damage. (*i*) An external reference sample may be used, a reference sequence quantified in it and in the test sample, and either a conclusion made that the test sample is sufficiently intact for analysis, or the qPCR result corrected using the ΔΔ cycle threshold (Ct) method (1). This approach assumes that the AF of the reference sequence in the reference sample is known, and it ignores any difference in length between the reference sequence and the target sequence. (*ii*) The standard curve for qPCR may be constructed using the target sequence in an external standard sample; this assumes that the AF is the same in the standard and test samples. (*iii*) The standard curve may use a standard sequence in the test sample; this assumes the standard and target sequences have the same AF, which is only true if they have the same length.

We have developed a general quantitative approach to the problem of DNA integrity, which avoids the above assumptions. If the lesions in DNA that prevent amplification occur randomly, then the Poisson distribution will describe their number in a given length of DNA. The mean number of lesions per base then provides a simple measure of DNA integrity. The AF of a target sequence can be calculated from this number and the target length.

The nature of many physical and chemical agents that act on DNA suggests that their effect on qPCR targets is approximately random, and most researchers using qPCR to study the effect of various DNA-damaging agents have assumed the Poisson distribution when interpreting differences in the number of a target sequence between damaged DNA and undamaged control DNA (2,3). In terms of experimental data, DNase and UV irradiation were observed to produce random DNA lesions (4,5). Deagle et al. (6) studied highly degraded DNA by qPCR, and observed that the number of target amplicons declined exponentially with target length. The authors suggested that the occurrence of random lesions in DNA impaired PCR amplification and could be described by the Poisson distribution. Colotte et al. (7) developed a similar model in which the occurrence of random lesions was described by the binomial distribution, to which the Poisson distribution approximates, and described how DNA degra-

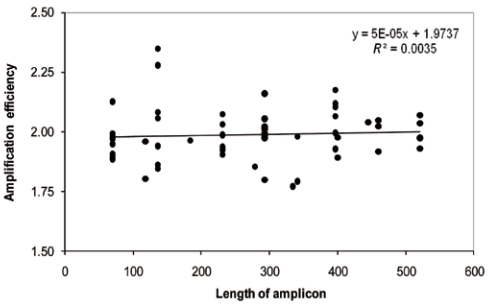

Figure 1. Relationship between amplification efficiency and amplicon length, determined in 11 experiments. The mean amplification efficiency, defined as the proportional increase in the number of amplicons per cycle, was 1.98, and there was no evidence of a decline in efficiency with increasing amplicon length. The degree of experimental variation of a single estimate of amplification efficiency of a single sequence accounts for the finding that 22/60 of the estimates were >2.0.

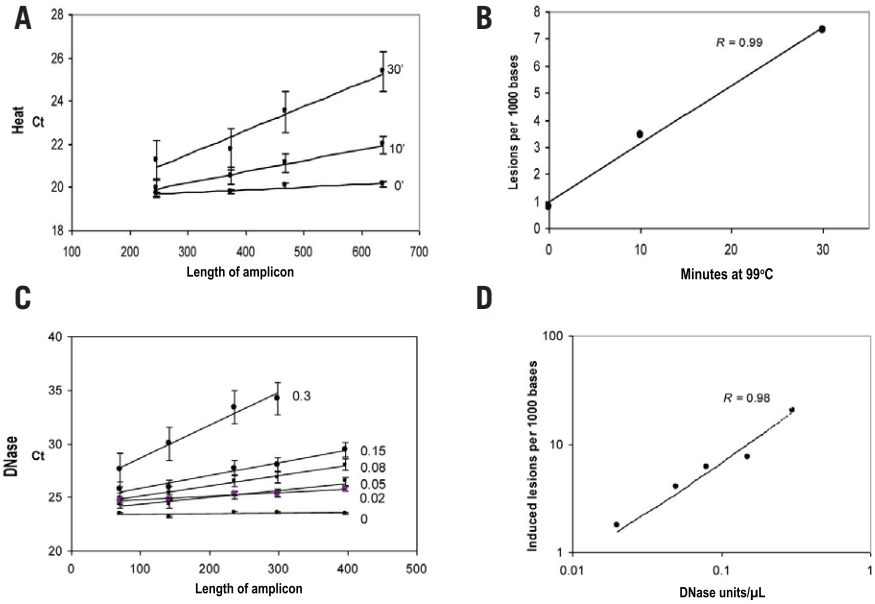

Figure 2. Effect on DNA integrity of heating at 99°C for various times and of various concentrations of DNase for 15 min. (A and C) The data of Ct versus product size, shown as the mean ± 1 SEM, together with the linear regression line. (B and D) The number of lesions per base related to the time of heating or concentration of DNase, together with the linear regression line. Time of heating is shown in minutes and concentration of DNase as units per microliter. The slope of the regression line between lesions/base and DNase concentration was 0.86 and was not significantly different from 1. The relation between lesions/1000 bases and magnitude of the damaging agent was significant ($P = 0.025$ for the effect of heating and $P < 0.005$ for the effect of DNase).

dation could be quantified using this model. However neither group excluded the possibility that a decline of PCR amplification efficiency with increasing target length could have explained their observations, and neither group expressed their model in a form simple enough to enable its practical use with routine qPCR.

For some years, we have used PCR to quantify MRD in human leukemia. Early studies used limiting dilution (digital) PCR, which both quantified MRD and measured the AF of a reference sequence (8,9). The analysis assumed that DNA lesions occur randomly and are described by Poisson statistics. When we changed to real-time qPCR, we continued with these assumptions. In the present study, our aim was to develop a practical method for incorporating measurement of DNA integrity into qPCR assays. We combined the mathematics of the Poisson distribution and of exponential PCR amplification, in order to derive, by qPCR, the frequency of DNA lesions and the AF of a sequence of interest. To provide further experimental evidence for the Poisson hypothesis and determine whether qPCR could determine degradation adequately, we first optimized qPCR conditions to ensure that amplification efficiency was independent of amplicon length over the range of lengths studied. We then used qPCR to quantify DNA damage. Finally, we used the measurement of DNA integrity to improve laboratory measurement by qPCR.

Materials and methods

Mathematical analysis

The following relationships can be derived by assuming that DNA damage occurs randomly, so that the frequency distribution of lesions in a DNA target follows the Poisson distribution and that amplification in real-time qPCR is exponential. The full derivations are given in the Supplementary Materials.

The overall integrity of a DNA sample can be described by r, the mean number of lesions/base in the DNA. If a number of DNA targets each of different length, l, are amplified and the number of amplicons increases with each PCR cycle by a factor of a (the amplification efficiency), then there will be a linear relationship between $Ct \cdot \log_e a$ and l with the slope being r. Use of this relationship is simplified if a reference system is chosen such that a is the same, irrespective of amplicon length, and its value is known. Then, using the slope of the linear relationship between Ct and l,

$$r = slope \cdot \log_e a. \qquad \text{[Eq. 1]}$$

For a particular target, the fraction of sequences that are intact and amplifiable, the AF is given by

$$AF = e^{-lr}. \qquad \text{[Eq. 2]}$$

The AF of a target sequence in a test sample may therefore be obtained by studying a reference sequence in the test sample, determining Ct values for amplicons of different length, using Equation 1 to determine the value of r, and then Equation 2 to determine the AF of the target sequence.

An extension of this approach is to accurately determine, in advance, the AF of a reference sequence in an external reference sample and, in the assay, to amplify a single length of the reference sequence in both the test sample and the external reference sample. The AF of the reference sequence in the test sample can then be determined by applying the ΔΔCt method to the Ct results, and the AF of the target sequence can then be determined from the relationship

$$AF_{tar} = (AF_{ref})^{(l_{tar}/l_{ref})}, \qquad \text{[Eq. 3]}$$

where AF_{tar} is the AF of the target sequence, AF_{ref} is the AF of the reference sequence, and l_{tar} and l_{ref} are the lengths of the target and reference sequences.

Quantitative PCR

Unless otherwise stated, segments of the *N-RAS* gene were amplified by qPCR in a Corbett Rotorgene 2000 Real-Time Cycler (Qiagen, Valencia, CA, USA) or an iCycler iQ real-time instrument (Bio-Rad Laboratories, Hercules, CA, USA) running IQ4 or IQ5 software. Amplification was

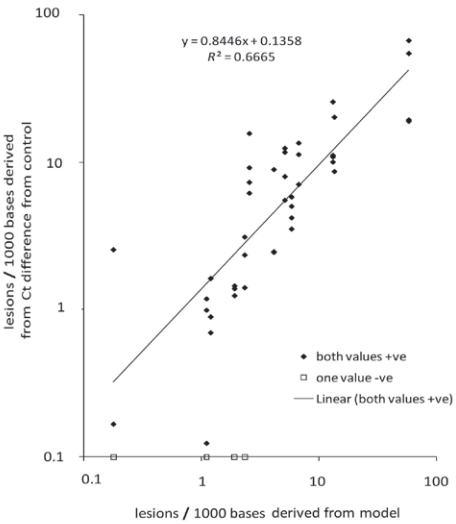

Figure 3. Relationship for samples with various degrees of degradation between lesions/base as calculated from the model and lesions/base as calculated from the Ct difference between the control sample and the degraded sample. DNA was degraded either by DNase or by acid lysis. There were four samples, shown as open squares, for which one of the two estimates was negative; the negative value was converted to a value of 0.1 lesions/1000 bases, but the four samples were not included in the calculation of the regression line.

monitored using a TaqMan hydrolysis probe. Segments of 70–521 bp were amplified using a standard upstream primer and various downstream primers. The sequences for the primers and probe are given in Supplementary Table S1 in the Supplementary Materials.

The PCR protocol was designed for optimal amplification efficiency and involved the use of dTTP rather than dUTP, high concentrations of magnesium and dNTPs, and, importantly, prolonged annealing and extension times. Amplification was in 25 µL containing 50 ng each primer, 1 U Platinum *Taq* DNA polymerase (Invitrogen, Carlsbad, CA, USA), and 2 mM Tris-HCl, pH 8.4, 50 mM KCl, 5 mM $MgCl_2$ and 300 µM each of dATP, dTTP, dCTP, and dGTP. Conditions were 96°C for 120 s, then 50 cycles of 94°C for 15 s, 55°C for 90 s, and 72°C for 60 s. Three or four replicates were used for each data point.

Samples of blood or marrow were collected into EDTA and stored at -80°C. Unless otherwise stated, DNA was extracted using a Qiagen FlexiGene DNA blood kit (Qiagen), and stored at -80°C. For studying genomic DNA, two or three different masses of 2–200 ng were used. DNA was diluted in Maxymum Recovery Tubes (MCT-150-L-C; Axygen, Union City, CA, USA). Amplification efficiency, a, was determined for sequences up to 521 bp, but usually for 70, 136, 231, 293, and 397 bp. Each sequence was amplified in several different masses of DNA, the slope of Ct versus $\log_e$ mass determined, and a calculated as $e^{-1/slope}$.

The value of r in a sample was determined from the slope of the linear regression line relating Ct to amplicon length, when different amplicons were amplified using the same mass of DNA.

DNA degradation

DNA was degraded to various degrees by either heat, DNase digestion, or acid lysis. The methods are described in the Supplementary Materials.

Results and discussion

Amplicon length and amplification efficiency

No relationship between amplicon length and amplification efficiency was evident when amplification efficiency was studied in 11 experiments amplifying sequences of *N-RAS* as described and using DNA from healthy individuals (Figure 1). The mean amplification efficiency/cycle was 1.98 (95% confidence interval = 1.95–2.01) and this value was used for all subsequent studies. A constant value of 0.68 was therefore able to be used for $\log_e a$ in Equation 1 when quantifying DNA integrity.

To determine whether the amplification protocol identified a lack of a relationship between amplicon length and amplicon efficiency for systems other than *N-RAS*, four different regions of the *BCR* gene were studied in a total of seven experiments. In each experiment, the difference between the amplification efficiency of a short sequence (151–175 bp) and that of a longer sequence (294–346 bp) was determined. For the seven experiments, the mean difference in amplification efficiency was 0.0157, with the standard error being 0.0101. This difference was not significant, indicating either that there was no true difference in amplification efficiency or, if there was, that it was negligible.

Induced DNA degradation

The hypothesis that lesions in DNA occur randomly and are described by the Poisson distribution predicts that there should be a linear relationship with a positive slope between Ct and amplicon length, as well as a linear relationship between the magnitude of the DNA damaging agent and the magnitude of the observed effect.

Five experiments studied heat damage, and Figure 2, A and B, shows the results of one. These experiments yielded 19 regression lines between Ct and length. Eighteen had a positive slope ($P = 3.8 \times 10^{-5}$, sign test). For eight, the slope differed significantly from zero: six with $P < 0.05$, one with $P < 0.025$, and one with $P < 0.005$. The experiments yielded five regression lines relating lesions/base and duration of heating. All five had a positive slope: for four, the value differed significantly from zero: $P < 0.05$ for one, $P < 0.025$ for two, and $P < 0.005$ for one.

Three experiments studied DNase, and Figure 2, C and D, shows the results of one. These experiments yielded 13 regression lines for the relationship between Ct and length. All had a positive slope ($P = 1.2 \times 10^{-4}$, sign test). For five, the slope differed significantly from zero: two with $P < 0.01$ and three with $P < 0.05$. The experiments yielded three regression lines relating lesions/base to concentration of DNase; all had a positive slope. For two, the value differed significantly from zero: $P < 0.01$ for one and $P < 0.005$ for one.

We concluded that the results are consistent with both predictions of the

Poisson hypothesis and that the integrity of DNA in a sample can be satisfactorily quantified by performing qPCR to amplify target sequences of different lengths.

Degraded samples: relationship between DNA integrity and experimental results

We tested our model further to determine whether measures of integrity derived from it correlated with measures of integrity derived by other means.

Electrophoresis can be used to assess DNA integrity. Electrophoresis performed under alkaline denaturing conditions on DNA degraded to various extents by DNase or acid lysis showed good correlation between the mean fragment size (calculated as the reciprocal of the number of lesions/base) and the observed fragment size at the point of maximum fluorescence on the gel (see the Supplementary Materials).

When control undamaged DNA is available, the DNA integrity in a damaged sample can be determined from the Ct difference between the damaged and undamaged DNA (2,3). For various samples of control and damaged DNA, we determined the number of lesions/base using the *N-RAS* gene and our model, and also determined the number of lesions/base for four different sequences of the *BCR* gene by using the Ct difference between the control sample and the damaged sample, ΔCt; the amplification efficiency, *a*, of these four sequences, which had previously been determined; the length, *l*, of the sequence; and the formula $\Delta Ct \cdot \log_e a/l$ to calculate lesions/base. The results, shown in Figure 3, show good correlation between these two methods for measurement of DNA integrity.

DNA integrity in control DNA

Values for lesions/1000 bases (*r*), for DNA from hematologically normal individuals are shown in Figure 4. In 23 experiments using carefully collected and processed genomic DNA from four healthy volunteers, the mean ± 1 SE for *r* was 0.13 ± 0.086 (difference from zero, *t* test, $P < 0.1$). In 40 experiments using 20 randomly collected blood samples from healthy volunteers (used as controls in experiments measuring levels of MRD), the mean ± 1 SE for *r* was 0.94 ± 0.21 (difference from zero, $P < 0.001$). For 18 marrow samples collected via thoracotomy from hematologically normal volunteers (used as controls in experiments measuring levels of MRD), the mean ± 1 SE for *r* was was 3.15 ± 0.58 (difference from zero, $P < 0.0001$). Excessive damage to DNA presumably occurred during collection or subsequent processing.

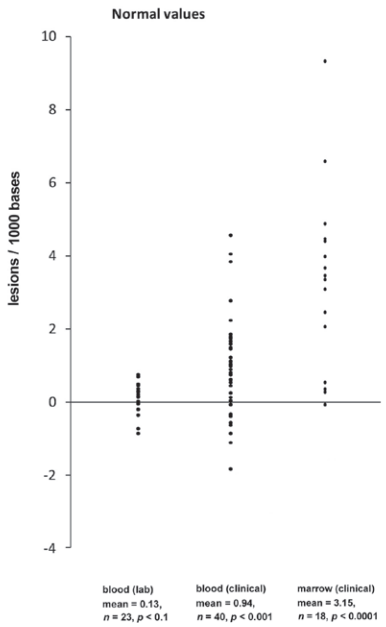

Figure 4. Integrity of normal DNA, expressed as the number of lesions/1000 bases. The clinical samples were control samples collected as part of MRD estimations. The other samples were samples carefully collected and processed in the laboratory. Experimental error is considered to explain those observations that estimate a negative value for lesions/base.

The precision of measuring DNA integrity was determined by studying DNA from a healthy volunteer. One sample was unheated, one was heated at 99°C for 20 min, and one was heated at 99°C for 30 min. Aliquots of each sample were held at -80°C, and one aliquot from each sample was thawed and assayed each day for eight instances. The mean values of *r* for the three samples were 0.37×10^{-3}, 3.7×10^{-3}, and 5.7×10^{-3}, respectively. The mean AF for a 200-bp target was calculated to be 0.93 (0.092), 0.48 (0.038), and 0.32 (0.052), respectively (SD is noted in parentheses).

DNA integrity and quantification of test samples by qPCR

When quantifying a target by qPCR, the amplifiable fraction of the target sequence is required, and it can be determined either by (*i*) amplifying different lengths of a reference sequence and applying Equations 1 and 2 or (*ii*) using an external reference sample for which DNA integrity has been determined, amplifying a single length of a reference sequence in the test and reference samples, and applying Equation 3. In either case, it is important to use an amplification protocol that is the same as or similar to that used in the present study in order to ensure that amplification efficiency is constant, unaffected by sequence length, and preferably close to 2. Use of an external reference sample to measure integrity is the simplest approach for routine use, particularly if multiple samples are being assayed, as only one measurement on the test sample is required. However, the alternative of measuring integrity by amplifying different lengths of the reference sequence in the test sample, although more laborious, does enable measurement of a reference sequence in the test and an external reference sample to be used to provide an independent measure of the mass of DNA being assayed.

Measurement of MRD in leukemia is an example of the use of AF in qPCR. Using patient-specific disease marker sequences and qPCR, 233 independent MRD measurements were performed on 115 marrow samples from 26 children with acute lymphoblastic leukemia (ALL) (10), and 76 independent MRD measurements were performed on 38 blood samples from 24 patients with chronic myeloid leukemia (CML) (11). Each MRD estimation involved performing a standard curve using the marker sequence in the diagnosis marrow, quantifying the marker sequence in the test sample against the standard curve, determining the AF of the marker sequence in the diagnosis (AF_d) and test (AF_t) samples, and multiplying the quantified value for the marker sequence in the test sample by the ratio AF_d/AF_t. The AF was determined either by using three *N-RAS* reference sequences 70, 231, and 397 bp in length (more recently just the 70- and 397-bp sequences), and using Equations 1 and 2, or by using an external reference sample for which the AF of the 70-bp *N-RAS* reference sequence had been previously determined, and using Equation 3. All MRD measurements included appropriate controls (10,11) including spiking controls to exclude PCR inhibition. For the ALL marrow samples, the median AF was 0.81, but the AF was <0.2 in 4% of samples, indicating significant degradation in a few samples; for the CML peripheral blood samples, the median AF was 0.84, and the lowest AF was 0.44; for the CML diagnosis marrow samples, the median AF was 0.84, and the lowest AF was 0.36. The median value of the AF for each of the three groups of samples was significantly less than 1 ($P < 0.001$, Wilcoxon rank-sum test).

In another study of ALL, cells were recovered from marrow smears stored on microscope slides, and the DNA was extracted by the use of proteinase K/phenol/chloroform. The DNA was heavily degraded in most samples, as the AF of the target sequences

ranged between 0.001 and 0.72 with a median of 0.18. Most of these samples were discarded, and alternative samples were obtained.

The results of measurement of DNA integrity in the various samples suggested that DNA integrity is usually well maintained when collection and processing of the sample is well controlled. In this situation, failure to correct for DNA damage would be unlikely to lead to a >2-fold error in the qPCR result, an error that might well be acceptable in practice in many situations. However, the observations on the marrow samples collected at thoracotomy or stored on microscope slides showed that substantial DNA damage can sometimes occur. Furthermore, the results in a minority of the ALL samples showed that significant DNA damage can occasionally be present in samples apparently collected under good conditions.

Even for samples collected under optimal conditions and showing minimal DNA damage, the results suggested that for every ~8000 bases, there is a lesion that prevents polymerase extension. This level of DNA damage is unlikely to be present in vivo and at least most of the observed damage had probably occurred during the process of collection, storage, or processing of the sample. Our unpublished data, involving limiting dilution analysis of denatured and undenatured DNA, have suggested that this damage is mainly due to double-strand lesions, probably double-strand breaks. Electrophoresis of degraded samples likewise suggested that damage was mainly due to double-strand breaks (see the Supplementary Materials).

When the MRD estimation on a sample gave a negative result, the AF of the target in the sample and the mass of DNA used in the assay were used to calculate that value for MRD that would result in one amplifiable target in the mass of DNA analyzed. The negative result was then reported as less than this MRD value. Expression of a negative result in this way improved interpretation, as it quantified the potential limit of detection for that sample and enabled the upper limit for negativity to be defined. This general approach to interpretation of a negative result requires that PCR inhibition not be present.

Acknowledgments

This work was supported by the National Health and Medical Research Council of Australia, the Cancer Council of South Australia, and Monoquant P/L.

Competing interests

A.A.M., M.J.B., and Flinders University hold equity in Monoquant, a company that has a commercial interest in MRD quantification in leukemia. The remaining authors declare no competing interests.

References

1. **Schmittgen, T.D. and K.J. Livak.** 2008. Analyzing real-time PCR data by the comparative C_T method. Nat. Protocols *3*:1101-1108.
2. **Ayala-Torres, S., Y. Chen, T. Svoboda, J. Rosenblatt, and B. Van Houten.** 2000. Analysis of gene-specific DNA damage and repair using quantitative polymerase chain reaction. Methods *22*:135-147.
3. **Santos, J.H., J.N. Meyer, B.S. Mandavilli, and B. Van Houten.** 2006. Quantitative PCR- based measurement of nuclear and mitochondrial DNA damage and repair in mammalian cells. Methods Mol. Biol. *314*:183-199.
4. **Thomas, Jr., C.R.** 1956. The enzymatic degradation of deoxyribose nucleic acid. J. Am. Chem. Soc. *78*:1861-1868.
5. **Govan III, H.L., Y. Valles-Ayoub, and J. Braun.** 1990. Fine-mapping of DNA damage and repair in specific genomic segments. Nucleic Acids Res. *18*:3823-3830.
6. **Deagle, B.E., J.P. Eveson, and S.N. Jarman.** 2006. Quantification of damage in DNA recovered from highly degraded samples—a case study on DNA in faeces. Front. Zool. *3*:11-20.
7. **Colotte, M., V. Couallier, S. Tuffet, and J. Bonnet.** 2009. Simultaneous assessment of average fragment size and amount in minute samples of degraded DNA. Anal. Biochem. *388*:345-347.
8. **Sykes, P.J., S.H. Neoh, M.J. Brisco, E. Hughes, J. Condon, and A.A. Morley.** 1992. Quantitation of targets for the polymerase chain reaction by use of limiting dilution. BioTechniques *13*:444-449.
9. **Brisco, M.J., J. Condon, E. Hughes, S.H. Neoh, P.J. Sykes, R. Seshadri, I. Toogood, K. Waters, et al.** 1994. Outcome prediction in childhood acute lymphoblastic leukaemia by molecular quantification of residual disease at the end of induction. Lancet *343*:196-200.
10. **Morley, A.A., S. Latham, M.J. Brisco, P.J. Sykes, R. Sutton, E. Hughes, V. Wilczek, B. Budgen, et al.** 2009. Sensitive and specific measurement of minimal residual disease in acute lymphoblastic leukaemia. J. Mol. Diagn. *11*:201-210.
11. **Bartley, P.A., D.M. Ross, S. Latham, M.H. Martin-Harris, B. Budgen, V. Wilczek, S. Branford, T.P. Hughes, and A.A. Morley.** Sensitive detection and quantification of minimal residual disease in chronic myeloid leukemia using nested quantitative PCR for *BCR-ABL* DNA. Int. J. Lab. Hematol. *32(6 Pt 1)*:e222-e228.

Received 23 April 2010; accepted 28 October 2010.

Address correspondence to Alec Morley, Haematology and Genetic Pathology, School of Medicine, Flinders University, GPO Box 2100, Adelaide, South Australia 5001. e-mail: alec.morley@flinders.edu.au