

Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA

P. Sean Walsh^{*,†}, Nicola J. Fildes[†] and Rebecca Reynolds

Human Identity Group, Roche Molecular Systems Inc., Alameda, CA 94501, USA

Received February 8, 1996; Revised and Accepted May 24, 1996

ABSTRACT

The PCR amplification of tetranucleotide short tandem repeat (STR) loci typically produces a minor product band 4 bp shorter than the corresponding main allele band; this is referred to as the stutter band. Sequence analysis of the main and stutter bands for two sample alleles of the STR locus vWA reveals that the stutter band lacks one repeat unit relative to the main allele. Sequencing results also indicate that the number and location of the different 4 bp repeat units vary between samples containing a typical versus low proportion of stutter product. The results also suggest that the proportion of stutter product relative to the main allele increases as the number of uninterrupted core repeat units increases. The sequence analysis and results obtained using various DNA polymerases appear to support the slipped strand displacement model as a potential explanation for how these stutter products are generated.

INTRODUCTION

The analysis of short tandem repeat (STR) loci by PCR methods has proven to be informative for a variety of applications because of the high degree of repeat number polymorphism at many of the loci in the human population (1). Dinucleotide repeat loci (for example, CA repeats) are very abundant and are routinely used as markers for use in linkage studies and for building genetic maps (2). Trinucleotide and tetranucleotide repeat loci are also used for such studies (3). Recently, tetranucleotide repeat loci have been studied for use as forensic markers (4,5). Many of these tetranucleotide repeat loci are highly polymorphic, with heterozygosity values for some loci >0.90 (6).

PCR-amplified STR loci are analyzed by gel electrophoresis and visualized using silver or fluorescent stains (7,8) or using primers labeled with a fluorescent tag (9). Capillary electrophoresis can also be used. Amplified alleles are visualized as bands on gels or are represented by peaks on electropherograms. All of these analytical methods detect a particular type of PCR product that is produced by amplification of STR loci. This product, which differs in size from the main allele by multiples of the repeat unit size, has been referred to in the literature as a

shadow band, DNA polymerase slippage product or stutter band (10–12). In the case of dinucleotide repeat loci, the most prevalent stutter band is generally two bases shorter than the main allele band, with additional stutter bands four and six bases shorter also visible (13). Because of the multiband pattern for each allele, interpretation of dinucleotide repeat loci is sometimes complicated, particularly for DNA samples that are mixtures from two or more individuals or when two alleles from a single individual are close in size (10).

PCR amplification results from tetranucleotide repeat loci are easier to interpret because only a single stutter band is typically observed, in a position four bases shorter than each allele band, and the intensity of the stutter band is generally <10% of the main band (4). Given the simpler interpretation, tetranucleotide repeat loci have been studied extensively as potential forensic DNA markers. Nonetheless, the stutter bands that are amplified from tetranucleotide repeat loci can still complicate the interpretation of mixed samples, which are common in forensic DNA analysis. Specifically, because the stutter bands are the same size as actual allele bands, it is not always possible to identify a faint band in a mixed sample as a real allele band or a stutter band if its position is four bases shorter than a more intense main allele band.

Because of the importance of stutter bands in sample interpretation, the experiments described in this paper aim to address the identity and properties of the stutter bands in one tetranucleotide repeat marker, vWA (14). The difference between the various vWA alleles is generally in the number of 4 bp repeat units they contain. Following amplification and separation by gel or capillary electrophoresis, alleles are distinguished on the basis of the length of the PCR products. In this paper, DNA sequences of allele bands and their corresponding stutter band from two samples were determined. In addition, DNA sequences of six allele bands of the same size were determined. Four of these samples have virtually undetectable stutter bands for this particular size vWA allele, while the other two samples have a more typical amount of stutter band for the same size allele. The DNA sequencing results revealed that all vWA alleles contain three repeat unit sequences (TCTA, TCTG and TCCA) that can differ in number and relative position within the repeat region. The most common repeat motif in the variable expansion region of a STR locus is referred to as the core repeat sequence. Comparison of the DNA sequences obtained from the main and stutter bands confirm that the two stutter bands have one fewer

* To whom correspondence should be addressed

† Present address: Human Identification Group, Perkin-Elmer Applied Biosystems Division, Foster City, CA 94404, USA

core repeat unit (TCTA) than the corresponding main allele bands. The sequencing results also suggest that the proportion of stutter band that is produced during the PCR is highest for alleles that have long stretches of uninterrupted core repeat sequence. Additionally, ABI 373A DNA Sequencer analysis of several vWA alleles of different lengths revealed that the proportion of stutter band generally increases with allele length. Likely mechanisms for stutter band formation and ideas for improving the interpretation of stutter bands are discussed.

MATERIALS AND METHODS

DNA amplification

Samples were amplified in PCR mix containing 50 mM KCl, 10 mM Tris-HCl, pH 8.3, 1.5 mM MgCl₂, 200 μM each of dATP, dCTP, dGTP and dTTP, 1.5 U AmpliTaq[®] DNA polymerase and 15 pmol each primer. One primer was labeled with a 5' FAM dye for those samples analyzed on the ABI 373A DNA Sequencer. Cycling was performed in a Perkin-Elmer DNA Thermal Cycler 480 (Norwalk, CT), using GeneAmp thin-walled reaction tubes, set to heat at 94°C for 60 s (denature), incubate at 60°C for 60 s (anneal) and incubate at 72°C for 60 s (extend) by the 'step-cycle' program. After 25–29 cycles, the samples were incubated for an additional 30 min at 60°C. Additional PCR enzymes were also tested, according to the manufacturer's instructions for the PCR conditions. The enzymes tested were as follows: the Stoffel fragment of *Taq* DNA polymerase, *Thermus thermophilus* (*Tth*), AmpliTaq DNA polymerase CS, AmpliTaq DNA polymerase FS (Perkin-Elmer), *Pyrococcus furiosus* (*Pfu*) (Stratagene, La Jolla, CA) and *Thermus brokianus* (*Tbr*) (Amresco, Solon, OH).

Band isolation

Isolation of PCR product bands prior to DNA sequence analysis was necessary for two reasons: (i) separation of stutter bands from main bands; (ii) separation of one allele from the other allele in a heterozygous individual. The first step of band isolation was to amplify the genomic DNA samples (10–20 ng) for 29 cycles, as described above, using unlabeled vWA primers. The PCR products (2 μl) were then heat denatured and run at a constant 40 W for 2.5 h on a 6.5% acrylamide gel containing 7.5 M urea. The PCR product bands were then detected by silver staining the gel directly on the glass plate. The bands of interest (stutter bands or main bands) were then carefully scraped off the plate using a small pipette tip (e. g. P20) and swirled into 100 μl TE buffer in a 0.5 ml microcentrifuge tube. This tube containing the small scrap of acrylamide gel was then heated at 85°C for 10 min to elute the DNA from the gel. Two microliters of this eluate were then added to vWA PCR Mix and amplified for 25 cycles. The PCR products from this amplification were then run again on a 6.5% acrylamide gel containing 7.5 M urea to verify that the band of interest had been successfully amplified and isolated. Finally, these PCR products were subjected to two rounds of Centricon[®]-100 microdialysis (Amicon, Beverly, MA) in TE buffer to remove primers and other PCR Mix components. The dialyzed PCR products were then used as DNA templates for the sequencing reactions.

DNA sequence analysis

Sequencing was performed using the Perkin-Elmer Ampli-Cycle[™] Sequencing Kit, which is based on cycle sequencing using a modified form of *Taq* DNA polymerase lacking 5'→3' exonuclease activity. [α -³³P]dATP (2.5 μCi) was used for labeling and detection. The sequencing reaction was performed as described in the kit instructions, using 25 cycles for the sequencing reaction and 4 pmol one primer. Completed sequencing reactions were run on a 6.0% acrylamide gel containing 8.3 M urea, run at a constant 70 W for 80 min. Film exposure times ranged from 1 to 2 days.

Sample analysis and band intensity quantitation

Samples were prepared for analysis by mixing 1.5 μl PCR product with 4 μl formamide and 0.5 μl GeneScan[™] 500 Rox labeled internal lane standard. This mixture was heat denatured at 95°C for 2–3 min and then quickly cooled on ice. After 5 min on ice, a 4 μl sample was then loaded on a 6.5% acrylamide (19:1 acrylamide:bis-acrylamide) gel containing 7.5 M urea. The well-to-read distance was 12 cm. The gel buffer was 1.2× TBE and the running buffer was 0.6× TBE. Electrophoresis (1200 V for 4 h) and fluorescence detection were performed on the ABI 373A DNA Sequencer. The data were analyzed with GeneScan[™] 672 software.

RESULTS

DNA sequence analysis of main allele and stutter products

DNA from two individuals possessing the 174 bp vWA allele was amplified as described above. Following gel electrophoresis, the 174 bp main band and the corresponding 170 bp stutter band from each individual were isolated and sequenced (Fig. 1). In this figure, only the repeat unit regions are shown; the flanking sequences were identical to previously reported vWA sequences (15).

All four of the isolated products contain the following 4 bp repeat unit sequences: TCTA, TCTG and TCCA. The TCTA sequence is the most frequent and is considered to be the core repeat unit for the vWA locus. To aid comparison of the main allele and stutter sequences, all non-core repeat units in Figure 1 are shaded. However, it should be noted that by convention the last two repeat units shown in Figure 1 are not counted as part of the total number of repeat units in vWA alleles (i.e. the main alleles for samples NF and RH contain 16 repeat units by convention, instead of 18).

For both samples (NF and RH), the stutter product has one less core repeat unit (TCTA) than the corresponding 174 bp allele. The sequences of the NF and RH alleles demonstrate that there can be sequence variation between alleles that are the same length. The RH 174 bp allele has a TCTG substituted for TCTA at the fourth repeat unit relative to the NF main allele. The substitution of a TCTG for a TCTA repeat at this position is common among vWA alleles (15). The TCTG substitution was also observed in the RH stutter product.

Sequence analysis of alleles with typical and reduced intensity stutter peaks

One of the advantages of using the 373 DNA Sequencer is that relative quantities of PCR products can be determined. In particular, the amount of stutter product amplified relative to the

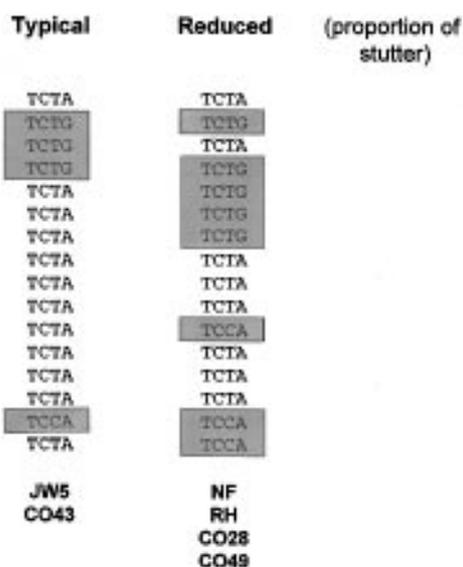


Figure 3. DNA sequence analysis of vWA alleles that yield typical and reduced amounts of stutter product. DNA samples from six individuals containing a vWA allele of the same length were amplified. The 166 bp PCR product from each sample was isolated and sequenced as described in Materials and Methods. The non-core repeat unit sequences have been shaded. Two of the individuals have identical allele sequences as indicated; the other four individuals have a different sequence. For samples JW5 and C043, the proportion of stutter product was ~4% of the main allele, which is typical for alleles that are in the size range of 166 bp. Reduced indicates that the proportion of stutter product was significantly <4% of the main allele (i.e. no stutter product was observed in these four samples). Refer to Figure 2 for additional details.

samples have a higher percent stutter than the shorter alleles. An example of this is shown in the electropherograms for samples JW5 and C043 in Figure 2. Subsequent experiments analyzing the proportion of stutter from over 100 population samples has confirmed this trend that percent stutter increases with allele length. Also, the percent stutter was observed to be quite reproducible for a given allele length; the standard deviation in percent stutter was observed to be 0.6% across all alleles (except for the 166 bp allele, where the percent stutter is either 4% or undetectable, depending on the repeat structure of the allele, as described above). An example of the increase in percent stutter with allele length is summarized for alleles from six samples in Table 1, where the range of stutter is 3.5% for the shortest allele and up to 8.7% for the longest allele.

Table 1. Proportion of stutter peak observed for seven vWA alleles

vWA allele length (bp)	Percent stutter
158	<3.9
166	4.7
170	5.6
174	6.3
178	7.5
182	8.1
194	8.7

Note that the proportion of stutter peak can be overestimated for alleles that are 8 bp longer than the other allele in the same sample

(see the long alleles for samples RH and NF in Fig. 2). The overestimation is due to the fact that the stutter peak resides on the 'shoulder' of the peak for the shorter allele. It might be possible to minimize this effect by running longer gels or implementing other gel condition modifications that increase peak sharpness or resolution. None of the alleles described in Table 1 have stutter peaks that are near the shoulder of a shorter allele, so the percent stutter that is noted in the table should be quite accurate.

Effect of different enzymes on proportion of stutter

Our laboratory has compared several thermostable DNA polymerases with AmpliTaq DNA polymerase in an attempt to identify an enzyme that gives a lower proportion of stutter peak following amplification of tetranucleotide repeat loci. Four of the enzymes tested have different exonuclease activities compared with AmpliTaq DNA polymerase, which has no 3'→5' exonuclease (proofreading) activity, but does have 5'→3' exonuclease activity. Experiments were performed by amplifying several DNA samples with AmpliTaq and also with one of the alternative enzymes, in the same thermal cycling run. PCR products were then analyzed together on a 373A gel, as described earlier. The proportion of stutter product observed for at least six alleles of differing length was then compared between the two enzymes. Table 2 summarizes the DNA polymerases tested, their exonuclease activities and the degree of stutter product relative to *Taq* polymerase. None of the enzymes resulted in an improvement in the proportion of stutter peak; in most cases, the proportion of stutter product for a particular allele was within 1% of the proportion observed for *Taq* polymerase. However, as shown in Figure 4, amplification with Stoffel fragment resulted in a 3-fold increase in the proportion of stutter peak compared with AmpliTaq. Stoffel fragment is a 61 kDa modified form of recombinant AmpliTaq[®] DNA polymerase, from which the N-terminal 289 amino acids have been deleted.

Table 2. Proportion of stutter product amplified by various DNA polymerases

Enzyme	5'→3' exo	3'→5' exo	Proportion of stutter compared with <i>Taq</i>
<i>Taq</i> ^a	Yes	No	
<i>Tbr</i> ^b	Yes	No	same as <i>Taq</i>
<i>Tth</i> ^c	Yes	No	same as <i>Taq</i>
<i>Pfu</i> ^d	Yes	Yes	same as <i>Taq</i>
AmpliTaq, CS	No	No	same as <i>Taq</i>
Stoffel ^e	No	No	3-fold increase
AmpliTaq, FS	No	No	same as <i>Taq</i>

^aAmpliTaq DNA polymerase.

^b*Thermus brokianus*.

^c*Thermus thermophilus*.

^d*Pyrococcus furiosus*.

^eStoffel fragment of AmpliTaq DNA polymerase.

DISCUSSION

The results presented in this paper reveal several interesting features of the stutter product that is produced in the amplification of tetranucleotide repeat sequences. Sequencing results of the

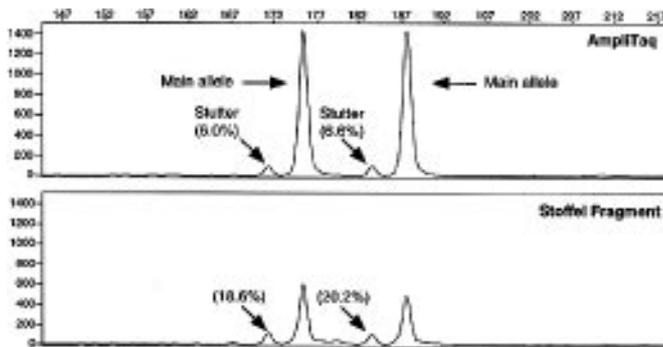


Figure 4. Comparison of proportion of stutter product generated using AmpliTaq polymerase and Stoffel fragment. DNA was amplified using 1.5 U AmpliTaq DNA polymerase (top electropherogram) or 3 U Stoffel fragment (lower electropherogram). The amplified products were analyzed on the 373A DNA Sequencer. The stutter peaks were observed to be 6% of the main allele peak height in amplifications containing AmpliTaq polymerase and increased to 20% in the amplifications containing Stoffel fragment.

vWA main allele bands and stutter bands indicate that the stutter band simply lacks one core repeat unit (TCTA) relative to the main band. A similar sequencing approach was also used in our laboratory to investigate the stutter bands in another tetranucleotide repeat locus, ACTBP2 (core repeat unit AAAG). The stutter bands at the ACTBP2 locus were also found to lack a single core repeat unit relative to the main band (data not shown). Note that the missing repeat unit is not necessarily from a particular position in the core repeat array; the repeat unit may be deleted from all positions equally. At least two other laboratories have reported that stutter bands at a dinucleotide repeat locus differ from the main allele by a deleted CA core repeat unit (12,13). Thus, the stutter bands at many short tandem repeat loci simply contain one fewer repeat units than the main allele band.

A mechanism known as slipped strand mispairing has been proposed to explain the stutter bands that result from amplification of dinucleotide repeat loci (12,16,17). According to this proposal, the template strand and extending strand can breath apart during synthesis through the repeat region, perhaps when/if the DNA polymerase has fallen off during the PCR (11). A single repeat unit can then loop out in the template strand before the two strands re-anneal, as shown in Figure 5. The result is that the newly extended strand will have one fewer repeat unit than the template strand when synthesis is complete. Interestingly, amplification using the Stoffel fragment of *Taq* actually increased the proportion of stutter over that seen with AmpliTaq DNA polymerase. Stoffel fragment has a processivity of ~5–10 bases, whereas the processivity of *Taq* is 50–60 bases (18). This observation is consistent with the slipped strand mispairing model, since the lower processivity would allow more opportunity for breathing of the DNA strands during PCR. Other researchers have predicted that a thermostable DNA polymerase with high processivity would greatly reduce stutter bands in the amplification of dinucleotide repeat sequences (13).

It is also conceivable in slipped strand mispairing that the extending strand could loop out, thus resulting in a newly extended strand having an additional repeat unit relative to the template strand. Stutter bands longer than the main allele have been observed at low levels with some dinucleotide repeat loci,

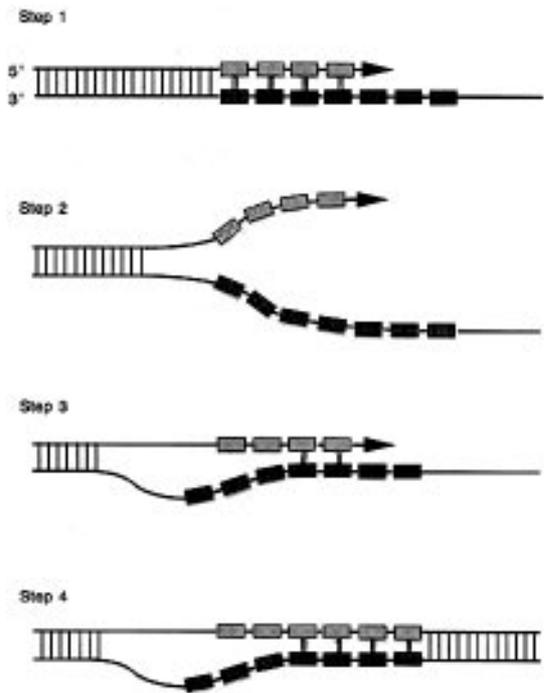


Figure 5. Slipped strand mispairing model. Diagrammed in this figure is a proposed mechanism to explain the formation of stutter bands one repeat unit shorter than the main allele band. The template strand (main allele) in this figure contains seven repeat units, as represented by the dark gray shaded boxes. In Step 1, the DNA polymerase has extended through four repeat units, represented by the light gray boxes. If/when the polymerase falls off the extending strand, the template and extending strand can breath apart, as shown in Step 2. When the two strands re-anneal in Step 3 the template strand has looped out and the extending strand aligns out-of-register by one repeat unit. Thus the polymerase can only add an additional two repeat units, instead of the correct three repeat units. Step 4 demonstrates that at the completion of synthesis the newly extended strand contains only six repeat units, while the template strand has seven repeat units.

although the most prevalent stutter bands are shorter than the main allele (J. Ziegler, personal communication). A speculative explanation for the relative lack of longer stutter bands is that the DNA polymerase may associate with the stretch of DNA at the 3'-end of the extending strand when the strands are unpaired (after breathing); this association may then somehow inhibit loop out formation in the extending strand. In the case of vWA and several other tetranucleotide repeat loci, stutter bands longer than the main allele have not been observed (4,5).

Results also indicate that stutter band formation is elevated in samples that have long stretches of core repeat sequences. The vWA sequencing results presented in this paper indicate that alleles in which the core repeat sequence (TCTA) is interrupted by a related 4 bp sequence have much lower stutter bands compared with alleles with 10 core repeat units in a row. Considering the slipped strand mispairing model, it is possible that the template strand has more of an opportunity and/or tendency to loop out when long stretches of core repeats are present. Also, the strand alignment that exists when loop out does occur in an *interrupted* core repeat stretch is less likely to position the 3'-end of the extending strand across from a complementary base; the polymerase then would not complete synthesis of what would otherwise become the shortened, stutter strand. Results

also show that the longer vWA alleles, which likely contain the longest stretches of the TCTA core repeat, have a higher proportion of stutter peaks compared with shorter alleles (15). Interestingly, these results may relate to observations from other laboratories that normal alleles at the fragile X locus have single base interruptions in the core triplet repeat sequence, whereas alleles with long stretches of perfect repeat are less stable *in vivo* and more likely to change in size (19). Also, other laboratories have indicated that alleles with greater numbers of repeats are more likely to exhibit instability in certain types of tumors (20). Indeed, repeated sequences in general have been described in the literature as 'Slippery DNA' (21). These results suggest that one might be wise to choose loci whose alleles contain overall fewer repeat units (for the longest alleles) or with interrupted repeat stretches if the goal is to select loci with low stutter.

No methods for eliminating the stutter band in tetranucleotide repeat loci have yet been identified. However, the proportion of stutter band is quite reproducible for each of the different alleles and so general rules for interpretation can be invoked. The main difficulty posed by stutter bands is in the interpretation of mixed samples; faint bands in a position four bases shorter than intense bands could be either stutter bands or real alleles from the minor component in a mixture. Interpretation is particularly difficult when the PCR products are detected by silver staining, for example when no quantitative analysis is performed on the band intensities. A potential improvement to the interpretation, therefore, is to detect the PCR products by a quantitative method, such as fluorescent detection on the 373A DNA Sequencer. With quantitative detection, the proportion of stutter band for each allele at a given locus can be determined in unmixed samples, as was done in the experiments presented here for some of the alleles at the vWA locus. For example, take the case of a hypothetical vWA allele, where the stutter band is determined to be a maximum of 5% of the main allele band. Given this information, it may then be reasonable to conclude that bands detected in this 'stutter' position that are, for instance, 15% of the main band can be identified as real allele bands (as the minor component of a mixture), as opposed to stutter bands. This approach can easily be implemented by using the 'filtering' algorithm available in ABI Genotyper™ software. The confidence that is given to such an interpretation would be related to the amount of experience and data that are available for the different STR alleles.

In conclusion, the results reported here provide useful information about the identity and properties of the stutter bands in one tetranucleotide repeat locus, vWA. Sequencing results indicate that the stutter band lacks one core repeat unit relative to the main band. This result is consistent with the mechanism that has been proposed to account for stutter bands at dinucleotide repeat loci,

slipped strand mispairing. Perhaps the most interesting finding was the sequencing result from a particular vWA allele with low stutter, showing that the TCTA core repeat unit in this allele sequence is interrupted, unlike the other vWA alleles, which have many more consecutive TCTA repeat units. Indeed, the results indicate that the proportion of stutter generally increases with alleles having longer core repeat unit regions. Until a method is available to eliminate stutter band formation during PCR, interpretation of results (particularly from mixed DNA samples) can likely be improved by careful characterization and quantitation of stutter bands for all alleles at the STR loci of interest.

ACKNOWLEDGEMENTS

The authors would like to thank David Gelfand and Rick Abramson for valuable discussions regarding the work in this manuscript and Jeanette Wallin for supplying supporting information.

REFERENCES

- Edwards, A., Hammond, H.A., Lin, J., Caskey, C.T. and Chakraborty, R. (1992) *Genomics*, **12**, 241–253.
- Fornage, M., Chan, L., Siest, G. and Boerwinkle, E. (1992) *Genomics*, **12**, 63–68.
- Edwards, A., Civitello, A., Hammond, H.A. and Caskey, C.T. (1991) *Am. J. Hum. Genet.*, **49**, 746–756.
- Urquhart, A., Oldroyd, N.J., Kimpton, C.P. and Gill, P. (1995) *BioTechniques*, **18**, 116–121.
- Fregeau, C.J. and Fourney, R.M. (1993) *BioTechniques*, **15**, 100–119.
- Kimpton, C.P., Gill, P., Walton, A., Urquhart, A., Millican, E.S. and Adams, M. (1993) *PCR Methods Applicat.*, **3**, 13–22.
- Bassam, B.J., Caetano-Anolles, G. and Gresshoff, P.M. (1991) *Anal. Biochem.*, **196**, 80–83.
- Hays, S.R., Yue, S., Wemmer, D.E., Quesada, M.A., Haugland, R.P., Mathies, R.A. and Glazer, A.N. (1992) *Nucleic Acids Res.*, **20**, 2803–2812.
- Sullivan, K.M., Pope, S., Gill, P. and Robertson, J. (1992) *PCR Methods Appl.*, **2**, 34–40.
- Litt, M., Hauge, X. and Sharma, V. (1993) *BioTechniques*, **15**, 280–284.
- Schlotterer, C. and Tautz, D. (1992) *Nucleic Acids Res.*, **20**, 211–215.
- Hauge, X.Y. and Litt, M. (1993) *Hum. Mol. Genet.*, **2**, 411–415.
- Murray, V., Monchawin, C. and England, P.R. (1993) *Nucleic Acids Res.*, **21**, 2395–2398.
- Kimpton, C., Walton, A. and Gill, P. (1992) *Hum. Mol. Genet.*, **1**, 287.
- Moller, A., Meyer, B. and Brinkmann, B. (1994) *Int. J. Legal Med.*, **106**, 319–323.
- Strand, M., Prolla, T.A., Liskay, R.M. and Petes, T.D. (1993) *Nature*, **365**, 274–276.
- Levinson, G. and Gutman, G.A. (1987) *Mol. Biol. Evol.*, **4**, 203–221.
- Lawyer, F., Stoffel, S., Saiki, R., Chang, S., Landre, P., Abramson, R. and Gelfand, D. (1993) *PCR Methods Appl.*, **2**, 275–287.
- Hirst, M.C., Grewal, P.K. and Davies, K.E. (1994) *Hum. Mol. Genet.*, **3**, 1553–1560.
- Wooster, R. *et al.* (1994) *Nature Genet.*, **6**, 152–156.
- Kunkel, T.A. (1993) *Nature*, **365**, 207–208.