Nucleic Acids Research, Vol. 19, No. 20 ***5613–5617***

Human Alu subfamilies and their methylation revealed by blot hybridization

Carl W.Schmid
Departments of Genetics and Chemistry, University of California, Davis, CA 95616, USA

Received July 16, 1991; Revised and Accepted September 12, 1991

ABSTRACT

By a simple direct blot hybridization strategy, the existence of human Alu family subfamilies is confirmed. Using consensus restriction cleavage sites, individual bands can be resolved from genomic human DNA digests corresponding to three distinct Alu subfamilies. Digestion with methylation sensitive and insensitive restriction enzymes shows that the numerous CpG residues in the youngest Alu subfamilies are largely methylated *in vivo*, suggesting a model for the transcriptional regulation of Alu repeats.

INTRODUCTION

By sequence analysis using a variety of criteria, five groups of investigators have proposed that Alu repeats can be divided into subfamilies having different consensus sequences and different times of appearance in the human lineage (1–5). RNA secondary structure has been proposed to be one constraint on the evolution of these subfamily sequences (6). The nomenclature and exact identity of these subfamilies is not fully resolved. Here, the youngest group of Alus is called the 'PV' subfamily, the next older group is named the 'precise' subfamily and an even larger and older group is the 'major' subfamily (2, 7, 8). The PV subfamily is identical to the HS subfamily identified by the Deininger group. Using oligonucleotide hybridization probes directed toward diagnostic mutations of the putative PV subfamily, this and P. Deininger's laboratory have provided direct evidence that the PV subfamily includes members that share five proposed diagnostic mutations relative to the precise subfamily, has recently expanded in the human lineage, and corresponds to one or more transcriptionally and transpositionally active loci (7–10). The transpositional activity of the precise subfamily is heterogeneous; many of its members certainly predate the divergence of human and ape (7, 8, 9), whereas others are dimorphic for their presence or absence in the human genome. There are on the order of one thousand PV members in the human genome and perhaps as many as one hundred thousand precise members (7, 9). Most of the nearly one million Alus present in the human genome belong to the older major subfamily which is neither transcriptionally or transpositionally active.

Another noteworthy feature of the PV subfamily is that its members are nearly exact matches to their consensus sequence, unlike the older major subfamily Alus which typically differ by about 10%, (ca. 28 positions) from their 281 nt consensus sequence (8, 9). The nearly perfect fidelity of restriction cleavage sites in the PV subfamily, as well as in the somewhat more divergent precise subfamily, suggests the experimental possibility of releasing these subfamily Alus as single bands in restriction digest of human DNA. This possibility is realized in results presented here.

CpG dinucleotides are very frequent in the young Alu subfamilies. There are 24 CpG dinucleotides in the 281 nt PV consensus sequence and 25 CpG dinucleotides in the precise consensus sequence (8, 10). The frequency of CpG's in these Alu repeats is about nine times greater than the frequency of CpG's (ca. 1%) in human DNA (11). Several investigators have noted that CpG's within Alu repeats exhibit a high transition mutation frequency of C to T or equivalently G to A (3–5). This high mutation rate is good indirect evidence that these CpG's are normally modified *in vivo* to give 5 methyl cytosine (12). Exploiting the previous suggestion of possibly resolving the PV and precise subfamilies from human DNA as single bands in appropriate restriction digests, their *in vivo* methylation at CpG containing sites could be tested by digestion with methylation sensitive and insensitive restriction enzymes. The Discussion addresses the possible significance of extensive methylation of Alu repeats.

METHODS

Human spleen DNA was used in all experiments. Restriction digestions were performed using conditions and buffers supplied by the manufacturers. Genomic digestions were performed using an equivalent amount of DNA, 6 μg, for all lanes. However, MspI and HpaII digestions involved an intervening phenol extraction and ethanol precipitation to change the restriction buffer prior to double digestion with HinfI. Thus, these two lanes are slightly underloaded compared to other digestions (Figure 1A, 2A, 3A).

A HaeIII-HinfI double digestion of 200 μg of spleen DNA was size fractionated by gel electrophoresis to isolate the 225 bp PV/precise Alu subfamily restriction fragment. Equivalent amounts of this size fractionated DNA, 1 μg, were digested as indicated with a third restriction enzyme (Figure 1B, 2B, 3B).

All electrophoresis was performed on 1.5% agarose gels in 1×TBE (13). A λ PstI digest was used as the size marker in all experiments. Gels were blotted onto 0.2 μ pore size nitrocellulose. The oligonucleotide GM2, 5'-ATCGAG-ACCATCCCGGCTAAAA-3', which spans positions 76 to 97, was kinase labelled with P-32 and used as a specific hybridization

probe of the precise and PV subfamilies by adjusting the washing stringency to 55°C and 67°C in 5×SSPE, respectively (7). The Alu subclone, Blur 2, was labeled by random priming and used as a non-discriminating hybridization probe to human Alu repeats by washing at 60°C in 1×SSC.

RESULTS

Consensus restriction sites and experimental outline

Consensus restriction sites for the Alu PV, precise and major subfamilies are compiled in Table I. Because PV members show little divergence from their consensus sequence, nearly all PV members are expected to share the sites listed in Table I (8). Precise subfamily members are about 95% identical so that a tetranucleotide restriction cleavage site is probably present in 80% of the members of this subfamily (2). Most of the more divergent major subfamily Alus would probably not share combinations of these consensus sites, an expectation demonstrated below.

The logic of my experiment is that at a 55°C washing temperature (Figure 1) the oligonucleotide GM2 hybridizes to both the PV and precise subfamilies but not to Alus belonging to the major subfamily (7). After washing at 67°C (Figure 2), this oligonucleotide hybridizes almost exclusively to the PV subfamily. Under non-stringent hybridization conditions Blur 2 hybridizes with members of all Alu subfamilies (Figure 3). A diagnostic BstYI site present at position 233 in the PV subfamily and absent from the precise subfamily provides an internal control to discriminate between hybridization to the closely related PV and precise subfamilies (7, 8, Table I). It should also be noted that several enzymes, HaeIII, HpaII, MspI and TaqI cut at sites within the region (positions 76 and 97) spanned by the hybridization probe, GM2. Cleavage at these sites within the GM2 complement would decrease or eliminate hybridization.

Hybridization to the PV and precise Alu subfamilies

As predicted by Table I, digestion of human DNA with both HaeIII and HinfI produces a major ~225 nt band (lane 2, Figure 1A) when probed with GM2 and washed at 55°C. Digestion with HinfI alone does not release any band (lane 3, Figure 1A) whereas a minor amount of an Alu size band is detected by the HaeIII digestion (lane 1, Figure 1A). Figure 1A is intentionally overexposed to reveal minor non-consensus bands for comparison with the more specific probing of the PV subfamily (Figure 2) and with the completely non-specific Alu probing using the Blur clone (Figure 3). A slightly longer non-consensus band is present as a minor component in the HaeIII-HinfI double digest (lane 2). Loss of the consensus HaeIII site at position 45 and retention of the HaeIII site at position 1 would account for this minor non-consensus fragment (Table I).

A MspI-HinfI double digest (lane 4, Figure 1A) eliminates hybridization of the GM2 probe at 55°C. Numerous MspI sites in the PV and precise subfamily consensus sequences (Table I) predict fragments too small to adhere to nitrocellulose under these blot hybridization conditions. The hybridization probe, GM2, spans the MspI site at position 88 in the PV subfamily so that cleavage at this site would also eliminate PV subfamily hybridization. In contrast, a HinfI double digestion with HpaII, which is methylation sensitive, does not digest the PV/precise subfamilies (lane 5, Figure 2A). The minor band present in this lane suggests that a small fraction of HpaII sites are unmethylated and subject to digestion. Consistent with the methylation of C in CpG dinucleotides, neither HhaI, a methylation sensitive enzyme, or HhaI in combination with HinfI (lanes 8 and 9) reveal any consensus bands.

TaqI digestion alone does not produce any bands (lane 7) but a HinfI-TaqI double digest produces an expected size fragment (lane 6) that is slightly shorter than the HaeIII-HinfI band observed in lane 2. The intensity of this TaqI-HinfI band product is significantly less than that of comparable digests. TaqI cleavage at position 77 removes two complementary nucleotides, an A and a T, from the region spanned by GM2 (position 76 to 98).

Table I. Alu Subfamily Restriction Sites

Enzyme	Position	Subfamily Consensus		
		PV	Precise	Major
HaeIII	1	+	+	+
	45	+	+	+
	89	−	−	+
MspI/HpaII	3	+	+	+
	88	+	−	−
	137	+	+	+
	205	+	+	+
HhaI/CfoI	7	+	+	+
	153	+	+	−
	237	−	+	−
ThaI	8	+	+	+
	236	−	+	−
TaqI	77	+	+	+
HinfI	192	−	−	+
	271	+	+	+
BstYI	232	+	−	−
AspI	271	+	+	+

Consensus restriction sites for Alu subfamilies are taken from References 7 and 8. The position indicates the first nucleotide in the site. The presence and absence of the site within a particular consensus sequence is indicated by the plus and minus signs, respectively. The hybridization probe GM2 spans positions 76 to 97.

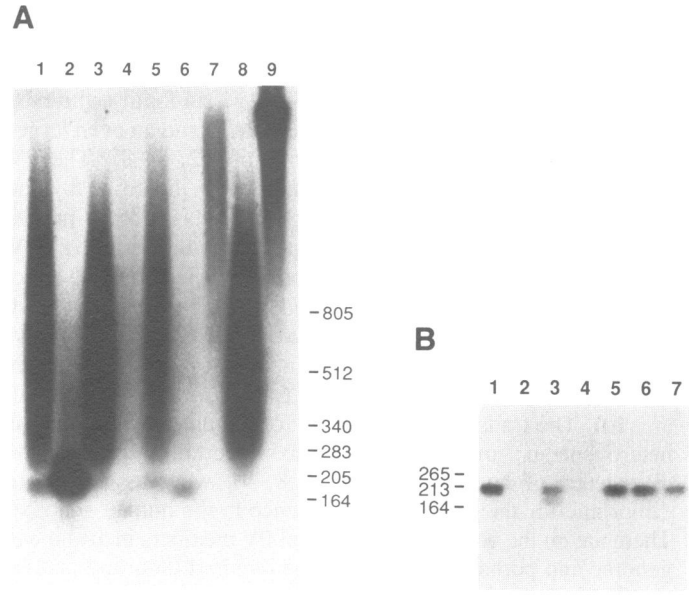

Figure 1. Hybridization with the oligonucleotide GM2 and final washings were performed at 55°C in 5×SSPE. *Part A*: Human genomic DNA digests are: lane *1 HaeIII*, lane *2 HaeIII-HinfI*, lane *3 HinfI*, lane *4 MspI-HinfI*, lane *5* HpaII-HinfI, lane *6 TaqI-HinfI*, lane *7 TaqI*, lane *8 HhaI-HinfI* and lane *9 HhaI*. The exposure time was fourteen hours. *Part B:* A size selected HaeIII-HinfI fraction was digested with a third enzyme as follows: lane *1 uncut*, lane *2 TaqI*, lane *3 BstYI*, lane *4 MspI*, lane *5 HpaII*, lane *6 CfoI*, lane *7 HhaI*. Exposure time was three hours.

The predictably lower temperature of the resulting duplex plausibly accounts for the significantly lower hybridization intensity of the predicted length HinfI-TaqI band.

The HaeIII-HinfI double digestion product (Figure 1A, lane 2) is easily size selected by gel electrophoresis, resulting in the product shown in lane 1 of Figure 1B. Individual lanes in Figure 1B compare identical amounts of this product. TaqI digestion (lane 2) again nearly eliminates the signal which, as mentioned above, is consistent with TaqI digestion within the GM2 complement. BstYI digestion (lane 3, Figure 1B) is diagnostic for the PV subfamily (Table I), thus the doublet in this lane demonstrates that the PV and precise subfamilies hybridize to approximately equal intensities under these conditions. MspI digestion (lane 4, Figure 1B) nearly erases the hybridization signal whereas HpaII digestion (lane 5, Figure 1B) results in little change in hybridization intensity suggesting that HpaII/MspI sites are largely methylated. Neither CfoI nor HhaI noticeably affect this hybridization signal (lanes 6 and 7, Figure 1B). Again, since both of these isoschizomers are methylation sensitive, CpG dinucleotides must be extensively modified by methylation in the PV and precise Alu subfamily members.

Preferential hybridization to the PV Alu subfamily

The results of Figure 1 are confirmed and extended by more stringent washing conditions examined in Figure 2. The diagnostic BstYI digestion in lane 3, Figure 2B, shows that under these washing conditions, hybridization is primarily due to PV subfamily members, whereas the hybridization signals described previously for Figure 1 are contributed about equally by the PV and precise subfamilies (lane 3, Figure 1B). These more stringent washing conditions greatly simplify the interpretation of the hybridization results. Unlike the complications previously ascribed to non-consensus restriction sites, neither HaeIII nor HinfI alone (lanes 1 and 3, Figure 2A) release a specific band. A HaeIII-HinfI double digest (lane 2, Figure 2A) produces the predicted band with a negligible non-consensus background hybridization. Similarly, a MspI-HinfI double digestion eliminates the hybridization signal (lane 4) but an HpaII-HinfI double digestion does not (lane 5). This confirms the previous interpretation that CpG dinucleotides in the PV subfamily members are largely methylated. Methylation sensitive enzymes such as HhaI (lanes 8 and 9, Figure 2A) do not release the PV subfamily. As discussed above in conjunction with Figure 1, TaqI digestion (lanes 6 and 7, Figure 2A and lane 2, Figure 2B) eliminates PV hybridization. The washing temperature in this experiment selects for an exact match of GM2 with its complement (7); cleavage with TaqI at position 77 would reduce this duplex melting temperature by 4°C.

Digestion of the size selected HaeIII-HinfI band with methylation insensitive (MspI, lane 4, Figure 2B) and methylation sensitive enzymes (HpaII, lane 5) again shows that these restriction sites are extensively methylated. PV subfamily hybridization to the HaeIII-HinfI band is also unaffected by digestion with two other methylation sensitive enzymes CfoI and HhaI (lanes 6 and 7, Figure 2B).

Non-subfamily specific Alu hybridization

Blur 2 hybridization under non-stringent conditions approximates the smear of the entire DNA mass with some specific differences (Figure 3A). Neither HaeIII nor HinfI digestion alone releases a noticeable band against the high background (lanes 1 and 3, Figure 3A). However, three low molecular weight bands are barely discernible above the non-specific smear of DNA in a HaeIII-HinfI double digest (lane 2, Figure 3A). The two larger of these three bands correspond to the doublet previously described in conjunction with the PV/precise subfamily probing (Figure 1A). The lowest molecular weight band corresponds to a restriction fragment that is peculiar to the major subfamily. The existence of this band is predicted by a major subfamily consensus HaeIII site at position 89 which is absent in the PV

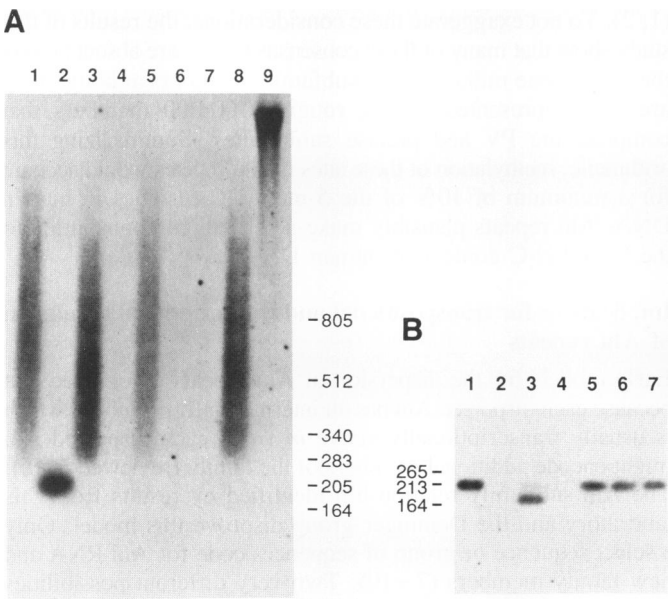

Figure 2. Following autoradiography resulting in the data of Figures 1A and 1B, the blots were washed twice at 67°C for 5 min in 5×SSPE. The exposure time was 72 hrs for both panels A and B.

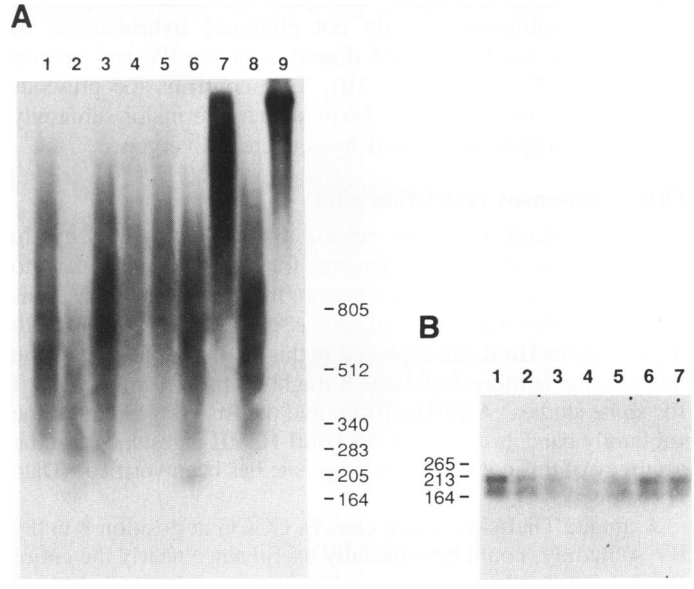

Figure 3. The blots used in Figure 1 and 2 were stripped by washing in water with 0.5% SDS at 67°C and then hybridized overnight with Blur 2 at 60°C in 3×SSC followed by washing in 1×SSC at 60°C. The exposure time for both panels A and B was three hrs.

and precise subfamilies (Table I, 2, 7). Confirming this interpretation, the size selected HaeIII-HinfI digest is revealed as a major subfamily doublet when probed by Blur 2 (all lanes, Figure 3B). It should also be noted that this major subfamily consensus HaeIII site is internal to the oligonucleotide hybridization probe, GM2. Thus, not surprisingly, this lower molecular weight doublet band was entirely undetected in probing with GM2 (Figures 1B, 2B).

The outstanding difference between the major subfamily (Figure 3A), the precise subfamily (Figure 1A) and, even more especially, the PV subfamily (Figure 2A) is the poor conservation of consensus restriction sites among the major subfamily membership. For example, GM2 probes essentially a single band in a HaeIII-HinfI digest (lane 2, Figures 1A, 2A) but most major subfamily Alus do not share these consensus sites (lane 2, Figure 3A). A special example of poorly conserved consensus restriction sites in the major subfamily are the four HpaII/MspI sites in the PV/precise subfamily (lanes 4 and 5, Figure 3A). Unlike results described for Figures 1 and 2, MspI and HpaII digests of the major subfamily are virtually indistinguishable. This is direct evidence that these restriction sites are largely inactivated by mutations in the case of the older, more divergent major subfamily, whereas the previous results of Figures 1 and 2 indicated that these sites are both highly conserved and extensively methylated in the PV and precise subfamilies. Matera *et al.* specifically noted that, unlike previous consensus sequences which are ill defined averages, the precise and more especially the PV subfamily consensus sequences approximate exact sequence representations (8). These restriction digestions confirm that interpretation.

TaqI-HinfI double digestion of the major subfamily (lane 6, Figure 3A) produces the predicted consensus band but also produces a smear of non-consensus DNA fragment lengths. TaqI digestion alone does not reveal any band (lane 7, Figure 3A). HhaI-HinfI double digestion also does not result in the consensus band length (lane 8, Figure 3A).

TaqI (lane 2, Figure 3B), BstYI (lane 3), and MspI digestion (lane 4) all result in a slightly decreased hybridization signal for the major subfamily but do not eliminate hybridization as observed for TaqI and MspI digestion of the PV and precise subfamilies (Figure 1B and 2B). This confirms the previous interpretation that consensus MspI sites in the major subfamily have been largely inactivated by sequence divergence.

Other consensus restriction sites

The new found ability to release the PV and precise Alu subfamilies as restriction fragments for blot analysis is likely to be technically useful. For this reason, the use of other consensus restriction sites was explored. An AspI site is coincident with the consensus HinfI site exploited in this study (Table I). Because this enzyme is more fastidious it might be preferred over HinfI for some studies. AspI-HaeIII double digests result in the same subfamily band as described for HinfI-HaeIII digests, so that the accuracy of this consensus cleavage site has been verified. (Data not shown).

A unique ThaI site, which cleaves CGCG at position 8 in the PV subfamily, could be especially useful since nearly the entire PV Alu subfamily sequence could be released by a ThaI-HinfI double digest. Despite repeated trials, I have not detected this predicted product. I take this negative result as further evidence for the extensive methylation of CpG dinucleotides within Alu repeats.

DISCUSSION

PV/precise Alu consensus sequences are nearly exact representations

A working assumption of our model and that of the Deininger laboratory for Alu evolution is that the PV and precise subfamilies are young, and as such their members have little sequence divergence unlike the members of older subfamilies (7–10). This assumption is consistent with sequence data bank analyses and has been supported by sequence analysis of a few cloned PV and precise Alu subfamily members (1–5, 8, 10). Restriction digests presented here demonstrate that the PV and precise Alu subfamily consensus sequences approximate exact sequences of the individual members rather than an ill defined average, as is the case for the older and more numerous major Alu subfamily members.

Alu transposition imports many CpG dinucleotides

As stated in the Introduction, CpG residues are about nine fold over-represented in PV/precise Alu repeats relative to total human DNA. The inactivity of methylation sensitive restriction enzymes, CfoI, HpaII, ThaI and HhaI, is consistent with the *in vivo* methylation of CpG dinucleotides, a conclusion that is also consistent with the known high transition mutation rate at these sites within Alu repeats (3–5). While only a subset of the 24 CpGs in PV Alus are assayed by these particular enzymes, I assume that for the purpose of discussion they are not unique in their methylation state. Young L1 repeats, like the young Alus investigated here, have relatively more CpG dinucleotides than older members of these families (15).

The maximum possible effect of methylation of Alu CpG's should be noted. A little less than 1% of human DNA depending on the tissue source consists of 5-methyl-C residues, corresponding to about 50 million 5-methyl-C's in the human haploid genome of 5 billion nucleotides (14 and references therein). There are about one million Alu repeats per haploid genome having a consensus sequence that includes about 25 CpG residues or equivalently 50 Cs that are potential methylation sites (1, 2). To not exaggerate these considerations, the results of this study show that many of these consensus CpG's are absent among the nearby one million major subfamily members although they are well represented among roughly 100,000 members that comprise the PV and precise subfamilies. Summarizing this arithmetic, methylation of these sites in Alu repeats would account for a minimum of 10% of the 5-methyl-C residues in human DNA. Alu repeats plausibly make a substantial contribution to the 5-methyl-C content of human DNA.

Implications for transpositional and transcriptional regulation of Alu repeats

Early models for the dispersion of Alu repeats recognized that because each dispersed Alu has an internal PolIII promoter, which is usually transcriptionally active *in vitro*, each dispersed Alu might encode additional members of the family (reviewed in 16). The Alu subfamily relationship identified by results from this laboratory and the Deininger group disprove this model. Only a select sequence or group of sequences code for Alu RNA and new family members (7–10). Two very different possibilities might account for this selectivity. There may be only one or a few functional loci, perhaps requiring necessary *cis*-acting flanking sequences, that are actively transcribed (8). Alternatively, only the most recently dispersed family members

might be transcriptionally and transpositionally active, corresponding to a relay of founder genes in which the older Alus are selectively inactivated (10). One difficulty with this relay model has been to account for the selective inactivation of older Alu repeats relative to more recently transposed members.

The present results do not distinguish between these two models but suggest a plausible mechanism for regulating the transcriptional activity of dispersed Alus and for inactivating older Alus in a relay of founder genes. Methylation of CpG sites in tRNA genes represses their *in vivo* transcriptional activity by RNA polymerase III (17). The extensive methylation of PV and precise Alu subfamilies observed here presumably could reversibly repress their *in vivo* transcriptional activity. However, the high transition frequency of 5-mC to T, which has been implied by sequence analysis of Alu repeats (4, 6) and confirmed by the inactivation of consensus MspI sites in this study, could preferentially and irreversibly inactivate older Alu members by mutating essential promoter elements so that only recently inserted members are potentially, transcriptionally competent. Naturally, other factors such as the chromatin context and methylation pattern imposed by the surrounding DNA sequences could also selectively determine the transcriptional activity of Alus which are otherwise potentially, transcriptionally competent. This simple model is consistent with all available evidence, including the remarkable sequence differences of the most recently transposed Alu repeats which implies the existence of many distinct source genes (8).

ACKNOWLEDGEMENTS

This research was supported by USPHS grant GM21346 and by the Agricultural Experiment Station, University of California, Davis, CA 95616. I thank C. Rubin, M. Hintz and U. Hellmann and especially E. Leeflang for their critical advice.

REFERENCES

1. Slagel,V., Flemington,E., Traina-Dorge,V., Bradshaw,H.,Jr. and Deininger,P.L. (1987) *Mol. Biol. Evol.* **4**, 19–29.
2. Willard,C., Nguyen,H.T. and Schmid,C.W. (1987) *J. Mol. Evol.* **26**, 180–186.
3. Britten,R.J., Baron,W.F., Stout,D.B. and Davidson,E.H.. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 4770–4774.
4. Jurka,J. and Smith,T. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 4775–4778.
5. Quentin,Y. (1988) *J. Mol. Evol.* **27**, 194–202.
6. Labuda,D. and Striker,G. (1989) *Nucl. Acids Res.* **17**, 2477–2491.
7. Matera,A.G., Hellmann,U. and Schmid,C.W. (1990) *Mol. Cell. Biol.* **10**, 5424–5432.
8. Matera,A.G., Hellmann,U., Hintz,M.F. and Schmid,C.W. (1990) *Nucl. Acids Res.* **18**, 6019–6023.
9. Batzer,M.A. and Deininger,P.L. (1991) *Genomics* **9**, 481–487.
10. Batzer,M.A., Kilroy,G.E., Richard,P.E., Shaikh,T.H., Desselle,T.D., Hoppens,C.L. and Deininger,P.L. (1990) *Nucl. Acids Res.* **18**, 6793–6798.
11. Schwartz,M.N., Trautner,T.A. and Kornberg,A. (1962) *J. Biol. Chem.* **237**, 1961–1967.
12. Bird,A.P. (1980) *Nucl. Acids Res.* **8**, 1499–1504.
13. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory Press.
14. Zhang,X.-Y., Loflin,P.T., Gehake,C.W., Andrews,P.A. and Ehrlich,M. (1987) *Nucl. Acids Res.* **15**, 9429–9449.
15. Hohjoh,H., Minakami,R. and Sakaki,Y. (1990) *Nucl. Acids Res.* **18**, 4099–4110.
16. Weiner,A.M., Deininger,P.L. and Estratiadis,A.E. (1986) *Annu. Rev. Biochem.* **55**, 631–662.
17. Besser,A., Götz,F., Schulze-Foster,K., Wagner,H., Kröger,H. and Simon,D. (1990) *FEBS Letters* **269**, 358–362.